Jussi Salmi, Andreas Richter, and Visa Koivunen. Sequential Unfolding SVD for tensors with applications in array signal processing. IEEE Transactions on Signal Processing, accepted for publication.

Sequential Unfolding SVD for Tensors with Applications in Array Signal Processing

Jussi Salmi, *Student Member, IEEE,* Andreas Richter, *Senior Member, IEEE,* Visa Koivunen, *Senior Member, IEEE*

***Abstract*—This paper contributes to the field of higher order ($N > 2$) tensor decompositions in signal processing. A novel PARATREE tensor model is introduced, accompanied with Sequential Unfolding SVD (SUSVD) algorithm. SUSVD, as the name indicates, applies a matrix singular value decomposition sequentially on the unfolded tensor reshaped from the right hand basis vectors of the SVD of the previous mode. The consequent PARATREE model is related to the well known family of PARAFAC [1] tensor decomposition models. Both of them describe a tensor as a sum of rank-1 tensors, but PARATREE has several advantages over PARAFAC, when it is applied as a lower rank approximation technique. PARATREE is orthogonal (due to SUSVD), fast and reliable to compute, and the order (or rank) of the decomposition can be adaptively adjusted. The low rank PARATREE approximation can be applied for, e.g., reducing computational complexity in inverse problems, measurement noise suppression as well as data compression. The benefits of the proposed algorithm are illustrated through application examples in signal processing in comparison to PARAFAC and HOSVD.**

***Index Terms*—array signal processing, channel modeling, low rank approximation, MIMO, SVD, tensor decompositions**

I. Introduction

A tensor is any N-dimensional collection of data (a second order tensor $N = 2$ is a matrix). In many signal processing applications, instrumental data contains information in more than two dimensions. Recently, researchers in several application areas have contributed to extending well established matrix operations to their tensor equivalents. Unfortunately, these extensions from their matrix counterparts are not trivial. For instance, even though the Singular Value Decomposition (SVD) has proven to be a powerful tool for analyzing second order tensors (matrices), its generalization to higher order tensors is not straightforward. There are several approaches for doing this, and none of them is superior in all aspects. In practice there are two major classes of models for higher order tensor decomposition, namely Tucker-model [2] and PARAFAC (parallel factorization [1], [3]). The latter is also known as CANDECOMP (canonical decomposition [4]). PARAFAC-based tensor decomposition stem from multilinear analysis in the fields of psychometrics [2], [5], sociology, chromatography and chemometrics [6]. It has been applied in many signal processing applications, such as image recognition, acoustics, wireless channel estimation [7] and array signal processing [8], [9]. Recently, also a Tucker-model based HOSVD (Higher Order SVD) [10] tensor decomposition subspace technique has been formulated to improve multidimensional harmonic retrieval problems [11]. In addition, the PARAFAC and HOSVD have been represented in an unified manner in a general framework in [12].

In this paper a novel tensor model is introduced, which belongs to the class of PARAFAC techniques. The new model, referred to as PARATREE, has a distinct hierarchical tree structure. The key idea is to sequentially unfold the tensor (reshape into a matrix), and to apply the singular value decomposition (SVD) on this matrix. This procedure is repeated for the right-hand singular vectors until no more data dimensions remain encompassed in them. As a result, a hierarchical tree structure for the factors is formed (see Section II-D for details). In the following, this decomposition method will be referred to as Sequential Unfolding SVD (SUSVD).

The formulation of a tensor decomposition as a sum of rank-1 tensors (as in PARAFAC) is suitable for several applications. By additionally imposing the rank-1 terms of the decomposition to be orthogonal — a property which is inherent in the SUSVD — the PARATREE model can be efficiently applied to approximate higher-order ($N > 2$) tensors. One example of such application involves interpreting the vector of eigenvalues of a large covariance matrix as a tensor, which is then used in a linear algebraic expression for finding the Fisher Information Matrix. Approximating this tensor using PARATREE decomposition allows for a significant reduction in computational complexity over a straight-forward matrix multiplication or any other exact solution. PARATREE also achieves a significant complexity reduction against HOSVD and PARAFAC. However, the use of PARATREE in practice is far more convenient than PARAFAC since the SUSVD does not suffer from convergence problems. Also the order of the PARATREE decomposition can be easily controlled, and the corresponding approximation error is well defined.

In a second novel application the PARATREE model is applied to suppress measurement noise in multidimensional MIMO radio channel measurements. This is performed by identifying the PARATREE components spanning the noise subspace, and removing their contribution from the channel

The research is partially funded by NORDITE WILATI project. The first author would like to thank Finnish Technology Promotion Foundation (TES), Emil Aaltonen Foundation, Finnish Society of Electronics Engineers (EIS), HPY Research Foundation, and Nokia Foundation for financial support.

J. Salmi and V. Koivunen are with Department of Signal Processing and Acoustics, Helsinki University of Technology/SMARAD CoE, Espoo, Finland (e-mail: {firstname.lastname}@tkk.fi).

A. Richter was with Department of Signal Processing and Acoustics, Helsinki University of Technology/SMARAD CoE, Espoo, Finland. He is now with Nokia Research Center, Helsinki, Finland (e-mail: andreas.richter@nokia.com).

observation.

To summarize, the benefits of the proposed PARATREE method include:

- Reduced computational complexity in high dimensional inverse problems
- Measurement noise suppression (subspace filtering)
- Compression of data (similar to low rank matrix approximation)
- Fast and reliable computation and adaptive order (rank) selection
- Revealing of hidden structures and dependencies in data.

The paper is structured as follows. In Section II a brief introduction to tensor modeling is provided along with the PARATREE description. Section III introduces the SUSVD and the PARATREE approximation for tensors. In Section IV two example applications are introduced in the field of array signal processing. Section V contains results for applying the algorithm on real world data and comparing the performance against PARAFAC and HOSVD approaches.

The notation used throughout the paper is as follows:

- Calligraphic uppercase letters ($\mathcal{A}$) denote higher order ($N > 2$) tensors.
- Boldface upper case letters (Roman $\mathbf{A}$ or Greek $\mathbf{\Sigma}$) denote matrices and lower case ($\mathbf{a}$, $\boldsymbol{\sigma}$) denote (column) vectors.
- The vector $\mathbf{a}_i = (\mathbf{A})_i$ denotes the i^{th} column of a matrix $\mathbf{A}$ and the scalar $a_j = (\mathbf{a})_j$ denotes the j^{th} element of a vector $\mathbf{a}$.
- Non-boldface upper case letters (N) denote constants, and lower case (a) denote scalar variables.
- Superscripts *, $^{\mathrm{T}}$, $^{\mathrm{H}}$, and $^+$ denote complex conjugate, matrix transpose, Hermitian (complex conjugate) transpose, and Moore-Penrose pseudo inverse, respectively.
- Different multiplication operators are defined for Kronecker $\otimes$, Schur (elementwise) $\odot$, Khatri-Rao $\lozenge$, outer $\circ$, and n-mode $\times_n$ products, respectively.
- Symbol $\hat{\mathcal{A}}$ denotes an estimate of the tensor $\mathcal{A}$.
- Operation $\mathrm{vec}(\bullet)$ stacks all the elements of the input tensor into a column vector. Operations $\mathrm{diag}(\bullet)$, $\mathrm{reshape}(\bullet, \{M_1, \ldots, M_N\})$ (inverse of vec), $\mathrm{permute}(\bullet, \{j_1, \ldots, j_N\})$, and $\mathrm{ipermute}(\bullet, \{j_1, \ldots, j_N\})$ (inverse of permute) are defined as in MATLAB computing software [13].
- The Frobenius norm of a tensor is defined as

$$||\mathcal{A}||_F = \left(\sum_i \left| \Big(\mathrm{vec}(\mathcal{A}) \Big)_i \right|^2 \right)^{\frac{1}{2}} = \sqrt{\mathrm{vec}(\mathcal{A})^{\mathrm{H}} \mathrm{vec}(\mathcal{A})}.$$

II. Tensor Decompositions

There are two major families of approaches to form a tensor decomposition, namely the PARAFAC [1], [3] (CANDECOMP [4]) and the TUCKER [2] models. PARAFAC is based on modeling the N-mode tensor as a sum of R rank-1 tensors, whereas the TUCKER model decomposes a tensor using a (smaller dimensional) core tensor and (possibly orthonormal) basis matrices for each mode. A good description of the properties and differences of the two approaches can be found in e.g. [14], [15]. In general, PARAFAC modeling has a more intuitive interpretation with common instrumental data, as the data can be often uniquely decomposed into individual contributions. Therefore, owing to its uniqueness properties [1], [3], [8], [16], [17], PARAFAC modeling is commonly used for signal modeling and estimation purposes, whereas orthogonal models such as HOSVD are better suited for tensor approximation, data compression, and filtering applications.

A. Basic Tensor Operations

In order to ensure the clarity of the notation, some tensor terminology is introduced in the following. The term N-mode (or N-way) tensor can be used to describe any N-dimensional data structure. A factor is an individual rank-1 contribution used in forming the tensor decomposition. It is a successive outer product of basis vectors (one from each mode), yielding a rank-1 contribution to the tensor. The term rank refers to the minimum number of rank-1 components yielding the tensor in linear combination. Further discussion on tensor rank can be found in [18], [19].

In the following some basic operations for an N-dimensional tensor $\mathcal{X} \in \mathbb{C}^{M_1 \times \ldots \times M_n \times \ldots \times M_N}$ are defined.

Definition 1 (The n-mode matrix unfolding): The n-mode matrix unfolding $\mathbf{X}_{(n)}$ of a tensor $\mathcal{X}$ comprises of:

1) Permutation of the tensor dimensions into an order $\{n, n+1, \ldots, N, 1, \ldots, n-1\}$.
2) Reshaping the permuted tensor into a matrix $\mathbf{X}_{(n)} \in \mathbb{C}^{M_n \times \prod_{i \neq n} M_i}$, i.e.,

$$\mathbf{X}_{(n)} = \mathrm{reshape}\big(\mathrm{permute}(\mathcal{X}, \{n, n+1, \ldots, N, 1, \ldots, n-1\}) , \{M_n, \prod_{i \neq n} M_i\}\big). \quad (1)$$

The order in which the columns of the matrix after unfolding are chosen in the latter step is not important, as long as the order is known and remains constant throughout the calculations. A more general treatment of the unfolding (or *matricization*), including nesting of several modes in the matrix rows, is given in [20].

Definition 2 (The n-mode product): The n-mode product $\mathcal{X} \times_n \mathbf{U} \in \mathbb{C}^{M_1 \times \ldots \times R_n \times \ldots \times M_N}$ of a tensor $\mathcal{X}$ and a matrix $\mathbf{U} \in \mathbb{C}^{R_n \times M_n}$ is defined as

$$\mathcal{X} \times_n \mathbf{U} = \mathrm{ipermute}\left(\mathcal{X}_U, \{n, n+1, \ldots, N, 1, \ldots, n-1\}\right), \quad (2)$$

where

$$\mathcal{X}_U = \mathrm{reshape}\left(\mathbf{U}\mathbf{X}_{(n)}, \{R_n, M_{n+1}, \ldots, M_1, M_N \ldots, M_{n-1}\}\right).$$

Definition 3 (Rank-1 tensors and vector outer product): A tensor $\mathcal{X}$ is rank-1 if it can be expressed as an outer product of N vectors as

$$\mathcal{X} = \mathbf{a}^{(1)} \circ \ldots \circ \mathbf{a}^{(N)}. \quad (3)$$

The elements of $\mathcal{X}$ are then defined as

$$x_{m_1 m_2 \ldots m_N} = \prod_{n=1}^{N} (\mathbf{a}^{(n)})_{m_n} = a^{(1)}_{m_1} \cdot a^{(2)}_{m_2} \cdot \ldots \cdot a^{(N)}_{m_N} \quad (4)$$

B. PARAFAC Model

The PARAFAC model is essentially a description of the tensor as a sum of R rank-1 tensors. There are a number of ways to express a PARAFAC decomposition [14], [15]. Consider an N-mode tensor $\mathcal{X} \in \mathbb{C}^{M_1 \times M_2 \times \cdots \times M_N}$ and N matrices $\mathbf{A}^{(n)} \in \mathbb{C}^{M_n \times R}$, where R is the number of factors — ideally equal to the rank of the tensor. Then the matrices $\mathbf{A}^{(n)}$, $n \in [1, \ldots, N]$, with columns $\mathbf{a}_r^{(n)}$, $r \in [1, \ldots, R]$ can be formed such that the tensor $\mathcal{X}$ is the sum of outer products

$$\mathcal{X} = \sum_{r=1}^{R} \mathbf{a}_r^{(1)} \circ \mathbf{a}_r^{(2)} \circ \cdots \circ \mathbf{a}_r^{(N)}, \tag{5}$$

where each outer product of the vectors $\mathbf{a}_r^{(n)}$ is a rank-1 tensor. Equivalently, the PARAFAC model can be expressed element-wise as

$$x_{m_1, m_2, \cdots, m_N} = \sum_{r=1}^{R} a_{r,m_1}^{(1)} \cdot a_{r,m_2}^{(2)} \cdot \cdots \cdot a_{r,m_N}^{(N)}, \tag{6}$$

where m_i denotes the index in i^{th} mode. A vectorized definition is given by

$$\text{vec}\left(\mathcal{X}\right) = \left(\mathbf{A}^{(N)} \lozenge \cdots \lozenge \mathbf{A}^{(1)}\right) \mathbf{1}_R = \sum_{r=1}^{R} \mathbf{a}_r^{(N)} \otimes \cdots \otimes \mathbf{a}_r^{(1)}, \tag{7}$$

where $\mathbf{1}_R$ is a column vector of R ones.

An illustration of the (N=3) PARAFAC model is shown in Fig. 1, where the relation to (5)–(7) is given by $\mathbf{a}_r^{(1)} = \mathbf{a}_r$, $\mathbf{a}_r^{(2)} = \mathbf{b}_r$ and $\mathbf{a}_r^{(3)} = \mathbf{c}_r$. The following properties of the PARAFAC decomposition should be pointed out:

- Finding the correct rank of a tensor may be tedious, and it typically involves evaluating several decompositions with different number of components. An example of such approach is the *core consistency diagnostics* (CORCONDIA) [21], which is based on comparing the weights of a PARAFAC decomposition mapped on a superdiagonally constrained TUCKER3 core tensor against an unconstrained TUCKER3 core tensor having the basis vectors of the PARAFAC model. Another approach was introduced in [19] based on evaluating the rank of the Jacobian, i.e., the matrix of partial derivatives of the tensor w.r.t. the parameters of the decomposition. However, as the tensor dimensions and the number of factors grow, the computational complexity and memory requirements involved in evaluating the Jacobian with dimensions $R(M_1 + \ldots + M_N) \times \prod_{n=1}^{N} M_n$ limit the applicability of this approach.
- The PARAFAC decomposition can not be deflated while still maintaining optimality in the LS (least-squares) sense for the reduced rank (as opposed to the matrix SVD). The best rank $R-1$ approximation of a tensor does not consist of the same rank-1 components as the rank R approximation [18], [22], [23]. Consequently, the PARAFAC decomposition has to be evaluated for each $R = 1 \ldots R_{max}$ separately to obtain the best fit.

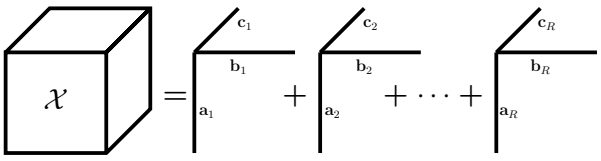


Fig. 1: Illustration of the PARAFAC decomposition — a sum of R rank-1 tensors. The relation to (5)–(7) is established by setting the factors $\mathbf{a}_r^{(1)} = \mathbf{a}_r$, $\mathbf{a}_r^{(2)} = \mathbf{b}_r$ and $\mathbf{a}_r^{(3)} = \mathbf{c}_r$.

C. Tucker (HOSVD) Model

Tucker models [2], [14], [15] are another common way to express a tensor decomposition. The idea is to form a limited set of basis vectors for each mode, and express the tensor as a linear combination of the outer products of the basis vectors of the different modes. A tensor $\mathcal{X}$ can be expressed using the Tucker model as

$$\mathcal{X} = \mathcal{S} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \ldots \times_N \mathbf{U}^{(N)}, \tag{8}$$

where $\mathcal{S} \in \mathbb{C}^{R_1 \times \ldots \times R_N}$ is called *the core tensor*, and the matrices $\mathbf{U}^{(n)} \in \mathbb{C}^{M_n \times R_n}$ contain the basis vectors. The Tucker decomposition ($N = 3$) is illustrated in Fig. 2.

In this paper the discussion is limited to a special case of the Tucker3 model, commonly known as the HOSVD [10]. In HOSVD, the unit norm columns of each of the matrices $\mathbf{U}^{(n)}$ form an orthonormal basis, and the core tensor $\mathcal{S}$ is *all-orthogonal* [10], [12]. Both of these properties result from the computational strategy of the HOSVD, which is briefly described in Section III-B.

D. PARATREE Model

We introduce a novel hierarchical formulation for a PARAFAC-type model having not only different number of factors in different modes (as in block-PARAFAC [7] or PARALIND [24]), but additionally the number of factors in each mode can vary for each *branch* in the hierarchical tree structure. The term branch refers to a set of factors having common factor(s) in previous mode(s) (see Fig. 3).

The PARATREE model (for an N-mode tensor) can be

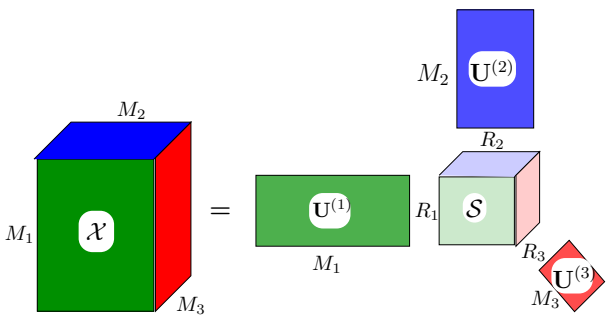


Fig. 2: Illustration of the Tucker3 decomposition. The tensor is decomposed as a linear combination of basis vectors in different modes according to (8).

expressed as the sum of outer products as follows

$$\mathcal{X} = \sum_{r_1=1}^{R_1} \mathbf{a}_{r_1}^{(1)} \circ \left(\sum_{r_2=1}^{R_2} \mathbf{a}_{r_1,r_2}^{(2)} \circ \cdots \right.$$
$$\left. \circ \sum_{r_{N-1}=1}^{R_{N-1}} \left(\mathbf{a}_{r_1,\ldots,r_{N-2},r_{N-1}}^{(N-1)} \circ \mathbf{a}_{r_1,\ldots,r_{N-2},r_{N-1}}^{(N)} \right) \right). \quad (9)$$

The vector $\mathbf{a}_{r_1,\ldots,r_n}^{(n)}$ in (9) denotes the r_n^{th} column of the n^{th} mode matrix of basis vectors $\mathbf{A}_{r_1,\ldots,r_{n-1}}^{(n)}$. The subscript $r_1,\ldots,r_{n-1}$ indicates the dependency of these matrices on the indexes of the previous factors of that branch in the decomposition tree. Also the number of factors R_n within each mode n can vary over different branches, i.e. R_n in (9) is actually a shorthand notation for $R_{r_1,\ldots,r_{n-1}}^{(n)}$.

An example of the tree structure of a full PARATREE model for a four-way ($N = 4$) tensor $\mathcal{X} \in \mathbb{C}^{2\times2\times2\times2}$ is given in Fig. 3. Depending on the true rank of the tensor, some of the branches may not contribute to the decomposition and can be ignored.

The PARATREE model is illustrated for a three-way tensor in Fig. 4. Compared to the PARAFAC model in Fig. 1, the r_a^{th} basis vector $\mathbf{a}_{r_a}$ in the first mode may be common for several factors in the remaining modes. To clarify the illustration in Fig. 4, the notation of (9) can be simplified (for N=3) to

$$\mathcal{X} = \sum_{r_a=1}^{R_a} \mathbf{a}_{r_a} \circ \sum_{r_b=1}^{R_b} \left(\mathbf{b}_{r_a,r_b} \circ \mathbf{c}_{r_a,r_b} \right), \quad (10)$$

where the relation to (9) is obtained by setting $\{\mathbf{a}_{r_a}, \mathbf{b}_{r_a,r_b}, \mathbf{c}_{r_a,r_b}\} \equiv \{\mathbf{a}_{r_1}^{(1)}, \mathbf{a}_{r_1,r_2}^{(2)}, \mathbf{a}_{r_1,r_2}^{(3)}\}$. Note that the number of factors $R_b = R_{r_a}^{(b)}$ for the second and third mode (vectors $\mathbf{b}_{r_a,r_b}$ and $\mathbf{c}_{r_a,r_b}$) may depend on the factor index r_a of the first mode (vectors $\mathbf{a}$). In addition, the number of factors in the last two modes is equal. For $N = 2$, the PARATREE reverts to the regular matrix SVD model.

We would like to point out that connections among different tensor decomposition models have been established. A

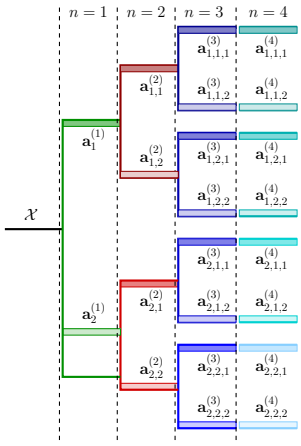


Fig. 3: An example of a full PARATREE for a four-way, ($2 \times 2\times2\times2$) tensor. Different colors represent different modes $n = \{1, 2, 3, 4\} \equiv \{green, red, blue, cyan\}$, and the tone within each mode represents decreasing magnitude.

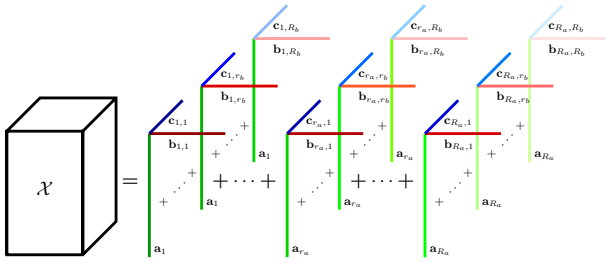


Fig. 4: Illustration of the three-way PARATREE decomposition — a hierarchical sum of R rank-1 tensors. The basis vector $\mathbf{a}_{r_a}$ of the r_a^{th} factor in the first mode may be common for multiple factors in the remaining modes.

PARAFAC model can be written as a Tucker model with a superdiagonal core tensor. On the other hand a Tucker model can be written as a PARAFAC model (with R equal to the number of elements in the core tensor). Hence, it would be straightforward to write the PARATREE model in terms of PARAFAC or Tucker models as well. A general framework unifying the different decompositions has been recently introduced in [12]. Appendix A provides links between the PARATREE model and tensor decomposition in block terms [12], as well as PARATREE and the recently introduced CONFAC (constrained factor decomposition) [9].

III. Algorithms

This section describes methods for computing the tensor decompositions introduced in Section II. The basic algorithms for computing the PARAFAC and HOSVD tensor decompositions are outlined in subsections III-A and III-B. Subsection III-C introduces a novel Sequential Unfolding SVD (SUSVD) decomposition proposed in this paper, and Subsection III-D describes its use for low rank approximation of a tensor.

A. PARAFAC - Alternating Least Squares

The alternating least squares (ALS) [1] is the most common algorithm for fitting a PARAFAC model. The basic idea is to have the number of factors R fixed, and obtain an update of the n^{th} mode basis vectors $\mathbf{A}^{(n)}$ as

$$\hat{\mathbf{A}}_{ALS}^{(n)} = \quad (11)$$
$$\mathbf{X}_{(n)} \cdot \left(\left(\mathbf{A}^{(N)} \lozenge \ldots \lozenge \mathbf{A}^{(n+1)} \lozenge \mathbf{A}^{(n-1)} \lozenge \ldots \lozenge \mathbf{A}^{(1)} \right)^{+} \right)^{\mathrm{T}},$$

while keeping the basis vectors of the other modes fixed. An iterative update of the matrices $\mathbf{A}^{(n)}$ is obtained by altering $n \in [1, \ldots, N]$ until a convergence is reached. The improvement of fit is monotonic. However, depending on the initial values for the matrices $\mathbf{A}^{(n)}$, a local optimum may be reached instead of the global one or the convergence may be very slow. Therefore, $\mathbf{A}^{(n)}$ are typically initialized by either using multiple random initial values, or so called *rational start* (based on either generalized rank annihilation or direct trilinear decomposition (DTLD)), or a *semi-rational start* (based on SVD/EVD) [15]. A discussion on different algorithms for fitting a PARAFAC model may be found in [6], [25].

B. Higher Order SVD (HOSVD)

The HOSVD [10] is obtained by computing the matrix SVD for each 1-mode unfolding of the tensor $\mathcal{X}$ and selecting the left singular vectors as the orthonormal basis of each mode, respectively. For the *complete* HOSVD, the basis matrices $\mathbf{U}^{(n)} \in \mathbb{C}^{M_n \times R_n}$ are hence given by the first $R_n = \text{rank}(\mathbf{X}_{(n)})$ left-hand singular vectors of the SVD of $\mathbf{X}_{(n)}$, defined as

$$\mathbf{X}_{(n)} = \mathbf{U}^{(n)} \mathbf{\Sigma}^{(n)} \mathbf{V}^{(n)^{\mathrm{H}}}. \tag{12}$$

Having computed the matrices $\mathbf{U}^{(n)}$, $n \in [1, \ldots, N]$, the core tensor $\mathcal{S} \in \mathbb{C}^{R_1 \times \ldots \times R_N}$ is given in closed form as

$$\mathcal{S} = \mathcal{X} \times_1 \mathbf{U}^{(1)^{\mathrm{H}}} \times_2 \mathbf{U}^{(2)^{\mathrm{H}}} \ldots \times_N \mathbf{U}^{(N)^{\mathrm{H}}}. \tag{13}$$

C. Sequential Unfolding SVD (SUSVD)

We introduce Sequential Unfolding SVD algorithm for estimating the PARATREE model for given tensor-valued data. The power of the algorithm lies in its recursive and inherently orthogonal structure. Furthermore, it allows adaptive selection of the order of the decomposition. SUSVD can be applied for any N-dimensional (real or complex) tensor, and for $N = 2$ it equals the matrix SVD.

The full SUSVD decomposition for a N-way tensor $\mathcal{X}^{M_1 \times \ldots \times M_N}$ ($M_1 \geq \ldots \geq M_N$) is described in Table I, and visualized for a $2 \times 2 \times 2$ three-way tensor in Fig. 5. The core idea of the algorithm is to apply the matrix SVD on the 1-mode matrix unfolding of the tensor (see Definition 1) to form the basis vectors of the first mode. Then each of the conjugated right-hand singular vectors $\mathbf{v}_{r_1}^{(1)^*}$ are reshaped into tensors, and the matrix SVD is applied on the 1-mode unfoldings of these tensors. This is repeated to construct the PARATREE model, until there are only the elements of the last mode contained in the right-hand singular vectors. Note that for a full SUSVD (all possible factors included) described in Table I, the number of basis vectors within each mode is the same for all branches and is given by

$$R_n = \min(M_n, \prod_{j=n+1}^{N-1} M_j). \tag{14}$$

Hence, the total number of orthogonal components in the decomposition (tips of the branches in Fig. 3) is given by

$$R = \prod_{n=1}^{N-1} R_n. \tag{15}$$

The $2 \times 2 \times 2$ tensor in Fig. 5 can be reconstructed with the PARATREE model as

$$\mathcal{X} = \sum_{r_1=1}^{R_1} \sigma_{r_1}^{(1)} \cdot \mathbf{u}_{r_1}^{(1)} \circ \sum_{r_2=1}^{R_2} \sigma_{r_1,r_2}^{(2)} \cdot \mathbf{u}_{r_1,r_2}^{(2)} \circ \mathbf{u}_{r_1,r_2}^{(3)}. \tag{16}$$

The full ($R_1 = 2$, $R_2 = 2$) reconstruction is illustrated in Fig. 6. The relation of the values in (16) to the ones in the

TABLE I: Description of the SUSVD algorithm for a complete tensor decomposition. The output variables $\{S\}$, and $\{U\}$ denote abstract tree structures to store the elements of the decomposition (see also Fig. 5).

$[\{S\}, \{U\}] = \text{SUSVD}\,(\mathcal{X})$

- Set $\mathcal{T}_0^{(1)} = \mathcal{X}$.
- Set $R^{(0)} = 1$
- For each $n = \{1, \ldots, N-1\}$:
  - For each $r_{n-1} = \{1, \ldots, R_{n-1}\}$:
  1) Unfold the tensor $\mathbf{T}_{r_1,\ldots,r_{n-1}}^{(n)} = \left(\mathcal{T}_{r_1,\ldots,r_{n-1}}^{(n)}\right)_{(1)}$,
  2) Compute the SVD $\mathbf{T}_{r_1,\ldots,r_{n-1}}^{(n)} = \mathbf{U}^{(n)} \mathbf{\Sigma}^{(n)} \mathbf{V}^{(n)^{\mathrm{H}}}$,
  3) For each $r_n \in \{1, \ldots, R^{(n)}\}$, with $R^{(n)} = \text{rank}\left(\mathbf{T}_{r_1,\ldots,r_{n-1}}^{(n)}\right)$:
     a) Store $\sigma_{r_1,\ldots,r_n}^{(n)} = \left(\mathbf{\Sigma}^{(n)}\right)_{r_n r_n}$ in $\{S\}$, and $\mathbf{u}_{r_1,\ldots,r_n}^{(n)} = \left(\mathbf{U}^{(n)}\right)_{r_n}$ in $\{U\}$.
     b) Then, if $n < N-1$,
        - Reshape $\left(\mathbf{V}^{(n)^*}\right)_{r_n}$ into a tensor $\mathcal{T}_{r_1,\ldots,r_n}^{(n+1)} \in \mathbb{C}^{M_{n+1} \times \cdots \times M_N}$, or else,
        - Store the vector $\mathbf{u}_{r_1,\ldots,r_{N-1}}^{(N)} = \left(\mathbf{V}^{(N-1)^*}\right)_{r_{N-1}}$.

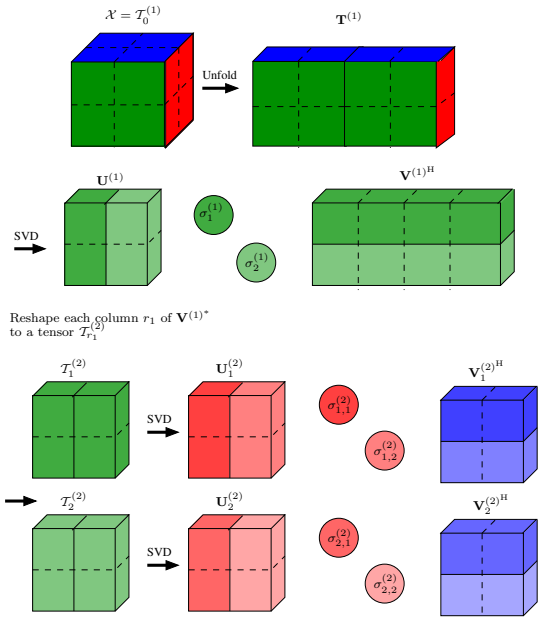


Fig. 5: The SUSVD decomposition for an arbitrary $2 \times 2 \times 2$ tensor. Different colors refer to different dimensions of the tensor. A circled σ denotes a singular value, dashed blocks are elements of the tensors, and solid lines are used to separate the column vectors. The tensor is first unfolded to a matrix $\mathbf{T}_0^{(1)}$. After applying SVD on this matrix, each of the right-hand singular vectors are reshaped (into $\mathbf{T}_{r_1,\ldots,r_{n-1}}^{(n)} \in \mathbb{C}^{M_n \times \prod_{q=n+1}^{N} M_q}$) and another SVD is applied on them. The procedure is repeated for each "branch" and "sub-branch", until no additional dimensions remain in the right hand basis vectors, i.e., the matrix $\mathbf{V}^{(N-1)}$ has only M_N rows (see also Table I).

3D-PARATREE formulation (10) or the general form (9) is given by

$$\begin{aligned}\mathbf{a}_{r_a} &= \mathbf{a}^{(1)}_{r_1} = \sigma^{(1)}_{r_1}\mathbf{u}^{(1)}_{r_1}\\ \mathbf{b}_{r_a,r_b} &= \mathbf{a}^{(2)}_{r_1,r_2} = \sigma_{r_1,r_2}\mathbf{u}^{(2)}_{r_1,r_2}\\ \mathbf{c}_{r_a,r_b} &= \mathbf{a}^{(3)}_{r_1,r_2} = \mathbf{u}^{(3)}_{r_1,r_2}.\end{aligned}$$

Note that the basis vectors of the SUSVD are exactly the same for the first mode as those of the HOSVD ($\mathbf{U}^{(1)}_{SUSVD} = \mathbf{U}^{(1)}_{HOSVD}$). However, the number of basis vectors of the latter modes is limited to $R_n = \text{rank}(\mathbf{X}_{(n)})$ for HOSVD, whereas in SUSVD the basis is formed independently for each branch. The result is that the total number R of individual rank-1 contributions, e.g., as if the decomposition would be written in PARAFAC form (5), is typically much less for the SUSVD than for the HOSVD. Another difference between the two decompositions is the fact that the HOSVD is unique, whereas for the SUSVD the solution depends on the order of the modes. This is further demonstrated in the application examples in Section V.

D. Reduced Rank Approximations

The individual rank-1 contributions of the HOSVD and SUSVD are orthogonal to each other. The practical implication of this property is that for a reduced rank approximation $\mathcal{X}_A$ of a tensor $\mathcal{X}$, the squared magnitudes of individual terms directly contribute to the squared magnitude of the approximated tensor. Hence, the squared Frobenius norm of the tensor approximation is given for the SUSVD by

$$||\mathcal{X}_{A,SU}||^2_F = \sum_{r_A} ||\mathbf{a}^{(1)}_{r_A} \circ \ldots \circ \mathbf{a}^{(N)}_{r_A}||^2_F = \sum_{r_A} \sigma^{(1)}_{r_A} \cdot \ldots \cdot \sigma^{(N-1)}_{r_A}, \tag{17}$$

where $\sigma^{(n)}_{r_A}$ denotes the n^{th} mode singular value, and r_A denotes an index of a rank-1 component included in the reduced rank decomposition. Equivalently, the squared Frobenius norm of the HOSVD approximation is given by

$$||\mathcal{X}_{A,HO}||^2_F = \sum_{r_A} |\,(\mathcal{S})_{r_A}\,|^2, \tag{18}$$

where $(\mathcal{S})_{r_A}$ denotes an element of the HOSVD core tensor, and index r_A denotes the indexes contributing to the approximation.

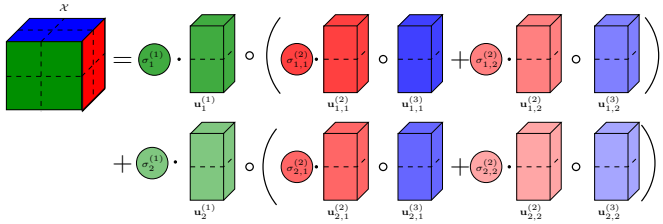


Fig. 6: A PARATREE tensor is reconstructed as a sum of outer products of weighted (by σ_{r_1}, σ_{r_1,r_2}) unitary basis vectors $\mathbf{u}^{(1)}_{r_1}$, $\mathbf{u}^{(2)}_{r_1,r_2}$ and $\mathbf{u}^{(3)}_{r_1,r_2}$. The tree structure allows common basis vectors in the previous dimensions (main branch). The symbol $\circ$ denotes outer product of vectors.

1) Deflating the full SUSVD: The PARATREE model built with SUSVD can be deflated to form a reduced rank approximation of a tensor. This can be done either offline after building the full SUSVD, or online during the computation of the decomposition. Here the discussion is limited to the offline approach.

Due to orthogonality of the decomposition (see Appendix for the proof), the approximation error can be equivalently expressed in terms of the sum of the product of the singular values related to each factor. These can be interpreted as the magnitudes of the single rank-1 components in the PARATREE, and are given by

$$\tilde{\sigma}_{r_1,\ldots,r_{N-1}} = \sigma^{(1)}_{r_1} \cdot \sigma^{(2)}_{r_1,r_2} \cdot \ldots \cdot \sigma^{(n-1)}_{r_1,\ldots,r_{n-1}}. \tag{19}$$

By stacking all the R (14)–(15) magnitude values (19) in descending order to a vector $\tilde{\boldsymbol{\sigma}} \in \mathbb{R}^{R\times 1}$, the normalized SUSVD approximation error can be expressed as

$$\epsilon_{r,SU} = \frac{||\mathcal{X} - \mathcal{X}_A||_F}{||\mathcal{X}||_F} = \sqrt{1 - \frac{\sum_{r_A=1}^{R_A} \tilde{\sigma}^2_{r_A}}{\sum_{r=1}^{R} \tilde{\sigma}^2_r}}, \tag{20}$$

or equivalently

$$\epsilon^2_{r,SU} = 1 - \frac{||\tilde{\boldsymbol{\sigma}}_A||^2_F}{||\tilde{\boldsymbol{\sigma}}||^2_F}. \tag{21}$$

Table II describes the offline PARATREE approximation method. The input data to the approximation function are the abstract tree structures $\{S\}$ and $\{U\}$ from the (full) SUSVD (see Table I), and a threshold ϵ_r for the target normalized approximation error (20). The output data consist of similar structures, but with reduced number of factors to approximate the tensor.

The rank of the approximation (total number R_A of rank-1 components) is given by

$$R_A = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R^{(2)}_{r_1}} \cdots \sum_{r_{N-2}=1}^{R^{(N-2)}_{r_1,\ldots,r_{N-3}}} R^{(N-1)}_{r_1,\ldots,r_{N-2}}. \tag{22}$$

The number of factors $R^{(n)}_{r_1,\ldots,r_{n-1}}$ in each mode n of each branch $r_1,\ldots,r_{n-1}$, depends on how the factor magnitudes are distributed among different branches.

The described offline approach allows for defining the achieved relative approximation error (20) precisely at the

TABLE II: Description of the *offline* SUSVD tensor approximation.

$[\{S_A\},\{U_A\}] = \text{PACK\_SUSVD\_OFFLINE}\,(\{S\},\{U\},\epsilon_r)$

1) Compute the products of the singular values (19) for each of the R possible branches (14)–(15) of the full PARATREE structure.
2) Sort all the products in descending order to a vector $\tilde{\boldsymbol{\sigma}} \in \mathbb{R}^{R\times 1}$.
3) Pick the minimum number R_A of singular values $\tilde{\boldsymbol{\sigma}}_A \in \mathbb{R}^{R_A\times 1}$, fulfilling the criterion (20)
$$||\tilde{\boldsymbol{\sigma}}_A||^2_F \geq (1-\epsilon^2_r)||\tilde{\boldsymbol{\sigma}}||^2_F.$$
4) Construct $\{S_A\},\{U_A\}$ based on the selected singular values $\tilde{\boldsymbol{\sigma}}_A$.

price of having to form the full SUSVD. This is most useful when saving in the computational complexity of the SUSVD decomposition itself is not crucial for the application at hand. This is the case in both of the example applications in Section IV. Further improvements in computational complexity could be achieved by truncating the decomposition in an online fashion, but then controlling the normalized error of the obtained approximation would not be as straightforward.

2) Deflating the full HOSVD: The HOSVD could be deflated either by reducing the least significant basis vectors from the matrices $\mathbf{U}^{(n)}$, or by considering the full decomposition, and selecting the elements from the core tensor that one wishes to include in the approximation. Here, the discussion is limited to the latter approach as that strategy for controlling the approximation error is similar to the one applied for the SUSVD deflation. Hence, the deflation of the HOSVD is obtained by setting the undesired contributions in the core tensor to zero, yielding a sparse core tensor $\mathcal{S}_A$. The obtained HOSVD approximation error is given by

$$\epsilon^2_{r,HO} = 1 - \frac{||\mathcal{S}_A||^2_F}{||\mathcal{S}||^2_F}. \tag{23}$$

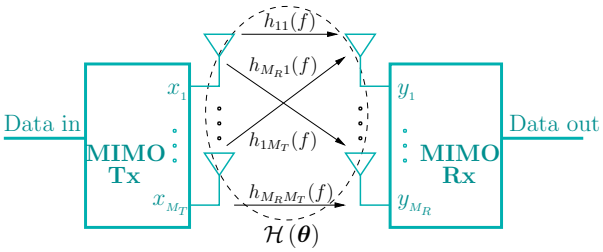
It should be noted that the low rank tensor approximations obtained by deflation are suboptimal both for the SUSVD and the HOSVD. However, in the proposed application example in Section IV-B, the optimality of the approximation in LS sense may be sacrificed to the benefit of a computationally efficient method yielding an approximation with a low number of rank-1 factors in the decomposition (nonzero core tensor elements for HOSVD). For HOSVD, the truncation could be performed also by reducing the rank of each mode, which would allow to further optimize the solution using Tucker3-ALS [14], [15], [26]. However, this method is time consuming, the achievable approximation error would not be as precisely controllable, and the obtained decomposition may still contain numerous insignificant rank-1 factors.

Applying a similar target approximation error requirement for a general PARAFAC model would require a trial and error approach for finding a proper rank (as well as for determining the tensor-rank in general). Also the convergence of the alternating least squares (ALS [26]) algorithms used for PARAFAC is very slow for high dimensional or ill-conditioned problems [27], [28].

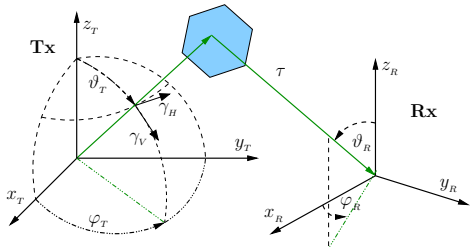
IV. Application Examples

A. MIMO Propagation Channel Modeling

The concept of the SUSVD algorithm and the PARATREE tensor model emerged while solving the problem of restoring the Kronecker structure of a data model appearing in multiantenna (MIMO) communications. Fig. 7a illustrates the MIMO principle. Measurement based modeling of wireless MIMO channels ($\mathcal{H}$ in Fig. 7a) is increasingly important as MIMO technology is being employed in many recent and emerging wireless transmission standards such as WLAN, 3GPP LTE and WiMAX. The measurements are crucial in order to analyze the potential of MIMO communications, develop advanced channel models, and to parameterize and verify existing models. These are necessary for transceiver development and network planning. In the following the measurement model for a wideband multiple-input-multiple-output (MIMO) radio channel measurement (also known as *radio channel sounding*) [29]–[31] is introduced in tensor notation.



(a) MIMO principle.



(b) Propagation path parameters.

Fig. 7: Illustration of the MIMO principle (Fig. 7a). The MIMO channel $\mathbf{H}$ is modeled as a superposition of propagation paths, whose parameters $\boldsymbol{\theta}$ (25) are illustrated in Fig. 7b.

The tensor valued measured radio channel $\mathcal{H} \in \mathbb{C}^{M_f \times M_T \times M_R}$ contains the complex coefficients describing the channel transfer function of the radio channel between the M_T transmitter antenna ports and M_R receiver antenna ports sampled at M_f frequencies. Realizations of such MIMO channels are obtained from MIMO channel sounding measurements [29]–[31]. The measurement model for the radio wave propagation is assumed to be comprised of two model components: the dominant (specular) propagation paths $\mathcal{H}_S$, and the diffuse scattering $\mathcal{H}_D$ (also called Dense Multipath Component, DMC). In the following, the polarization effects are ignored to simplify the description.

The propagation paths can be expressed as a constrained[1] PARAFAC model

$$\mathcal{H}_S(\boldsymbol{\theta}) = \sum_{p=1}^{P} \gamma_p \left(\mathbf{b}^{(f)}(\tau_p) \circ \mathbf{b}^{(T)}(\varphi_{T,p}, \vartheta_{T,p}) \circ \mathbf{b}^{(R)}(\varphi_{R,p}, \vartheta_{R,p}) \right), \tag{24}$$

where P is the number of individual propagation paths. Each rank-1 component (a path) in the model (24) is parameterized with L parameters illustrated in Fig. 7b. The parameter vector $\boldsymbol{\theta} \in \mathbb{R}^{LP \times 1}$ is defined as

$$\boldsymbol{\theta} = \left[\boldsymbol{\tau}^{\mathrm{T}} \ \boldsymbol{\varphi_T}^{\mathrm{T}} \ \boldsymbol{\vartheta_T}^{\mathrm{T}} \ \boldsymbol{\varphi_R}^{\mathrm{T}} \ \boldsymbol{\vartheta_R}^{\mathrm{T}} \ \Re\left\{\log\left(\boldsymbol{\gamma}^{\mathrm{T}}\right)\right\} \ \Im\left\{\log\left(\boldsymbol{\gamma}^{\mathrm{T}}\right)\right\} \right]^{\mathrm{T}}. \tag{25}$$

[1]The model is constrained on a limited set of basis functions $\mathbf{b}$ determined by the (calibrated) measurement system.

TABLE III: System dimensions in the complexity examples.

	Label	Value
Frequency samples	M_f	193
Transmitter ports	M_T	30
Receiver ports	M_R	31
Total Number of Parameters	L'	800

The parameters are delay (τ), azimuth (φ) and elevation (ϑ) angles of departure (at Tx) and arrival (at Rx), and complex path weights γ for each path. The vectors $\mathbf{b}^{(i)}(\theta_p) \in \mathbb{C}^{M_i \times 1}$, $i \in \{f, T, R\}$, $\theta \in \{\tau, \{\varphi_T, \vartheta_T\}, \{\varphi_R, \vartheta_R\}\}$ in (24) denote the basis functions for the frequency, transmit array, and receive array modes, respectively. The description of the mapping of the parameters within the basis functions is rather lengthy, and hence omitted here. The interested reader is directed to [29], [32] for additional details on how the parameters characterize the channel. Parametric estimation of (24) is discussed in [29], [32]–[34].

The diffuse scattering component is defined as a tensor valued complex circular symmetric normal distributed random variable

$$\mathcal{H}_D \sim \mathcal{N}_C\left(\mathbf{0}, \mathcal{R}_D\right), \tag{26}$$

with a covariance tensor $\mathcal{R} \in \mathbb{C}^{M_f \times M_T \times M_R \times M_f \times M_T \times M_R}$. The following Kronecker structure is assumed for the covariance tensor reshaped into a ($M \times M$, with $M = M_f M_T M_R$) matrix

$$\mathbf{R}_D = \mathrm{E}\left\{\mathrm{vec}(\mathcal{H}_D)\mathrm{vec}(\mathcal{H}_D)^{\mathrm{H}}\right\} = \mathbf{R}_f \otimes \mathbf{R}_T \otimes \mathbf{R}_R \in \mathbb{C}^{M \times M}. \tag{27}$$

The matrices $\mathbf{R}_f \in \mathbb{C}^{M_f \times M_f}$, $\mathbf{R}_T \in \mathbb{C}^{M_T \times M_T}$, and $\mathbf{R}_R \in \mathbb{C}^{M_R \times M_R}$ are the covariance matrices of the frequency, the transmit array, and the receive array modes, respectively. Estimation of these covariance matrices is discussed in [35].

Furthermore, a tensor $\mathcal{H}_N(k)$ is defined, denoting zero mean i.i.d. normal distributed complex circular symmetric (measurement) noise with covariance

$$\mathbf{R}_N = \mathrm{E}\{\mathrm{vec}(\mathcal{H}_N)\mathrm{vec}(\mathcal{H}_N)^{\mathrm{H}}\} = \sigma_N^2 \mathbf{I}. \tag{28}$$

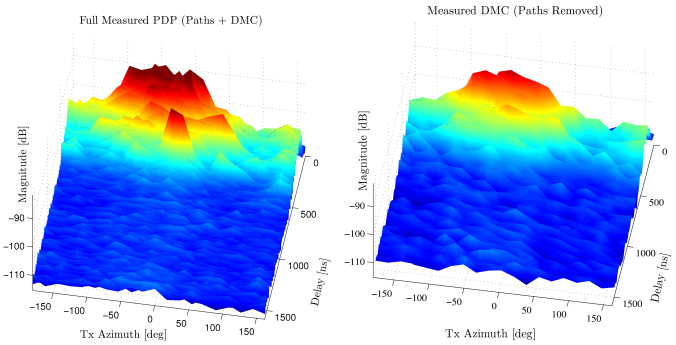
Using (24), (26), and (28) the model for the full measured complex transfer function of the radio channel tensor (a snapshot) at time k is defined as

$$\mathcal{H}(k) = \mathcal{H}_S(k) + \mathcal{H}_D(k) + \mathcal{H}_N(k) \sim \mathcal{N}_C\left(\mathcal{H}_S, \mathcal{R}\right), \tag{29}$$

where the covariance tensor is defined as $\mathcal{R} = \mathcal{R}_D + \mathcal{R}_N$. The covariance tensor of (29) can be written in matrix form as

$$\mathbf{R} = \mathrm{E}\{\mathrm{vec}(\mathcal{H})\mathrm{vec}(\mathcal{H})^{\mathrm{H}}\} = \mathbf{R}_f \otimes \mathbf{R}_T \otimes \mathbf{R}_R + \sigma_N^2 \mathbf{I}. \tag{30}$$

The contribution of the model components $\mathcal{H}_S$, $\mathcal{H}_D$, and $\mathcal{H}_N$ in the Power-Angular-Delay-Profile (PADP) of a MIMO radio channel measurement is visualized in Fig. 8. The dimensions of the measurements used e.g. in [31], are listed in Table III. These values will be used in the complexity evaluations in the following sections, where $\mathcal{O}(M)$ is used for denoting that M is the order of multiplications and additions required by an algebraic operation in question. The total number of parameters $L' = LP$ in Table III is a typical number of paths ($P \approx 40$) times a suitable number of estimated parameters $L \approx 20$ per path[2].



(a) Full measured channel $\mathcal{H}$ (29). (b) DMC and measurement noise $\mathcal{H} - \mathcal{H}_S = \mathcal{H}_D + \mathcal{H}_N$.

Fig. 8: Illustration of the influence of dominant paths $\mathcal{H}_S$ (24) and the DMC $\mathcal{H}_D$ (26) in a power-Tx azimuth-delay profile of a measured channel. The data are averaged over all receiver elements.

B. The Fisher Information Matrix - Key Quantity in Parameter Estimation

1) Computational Challenges of the FIM in Propagation Parameter Estimation: The model for the propagation paths (24) in vectorized form is defined as

$$\mathbf{h}_S(\boldsymbol{\theta}) = \mathrm{vec}\left(\mathcal{H}_S(\boldsymbol{\theta})\right). \tag{31}$$

The parameters ($\boldsymbol{\theta}$) used for identifying the model $\mathbf{s}(\boldsymbol{\theta})$ may be estimated using, e.g., iterative Maximum Likelihood (ML) [29], or Extended Kalman Filter (EKF) [32], [33]. Both of these estimation methods[3] rely on the evaluation of the expression

$$\mathbf{J} = 2\Re\{\mathbf{D}^{\mathrm{H}}\mathbf{R}^{-1}\mathbf{D}\} \tag{32}$$

which is commonly known as the Fisher Information Matrix (FIM) — a measure of the amount of information about $\boldsymbol{\theta}$ carried in $\mathbf{h}_S(\boldsymbol{\theta})$. This expression contains an inverse of the (full rank) measurement covariance matrix $\mathbf{R}$ (30), as well as a Jacobian matrix

$$\mathbf{D} = \frac{\partial \mathbf{h}_S(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \in \mathbb{C}^{M \times L'}. \tag{33}$$

Due to the structure of the data model in our example application (24), the expression of the FIM (32) can be expanded as

$$\mathbf{J} = 2\Re\{(\mathbf{D}_f \Diamond \mathbf{D}_T \Diamond \mathbf{D}_R)^{\mathrm{H}} \cdot \left(\mathbf{R}_f \otimes \mathbf{R}_T \otimes \mathbf{R}_R + \sigma^2 \mathbf{I}\right)^{-1} \cdot (\mathbf{D}_f \Diamond \mathbf{D}_T \Diamond \mathbf{D}_R)\}. \tag{34}$$

Straight-forward computation of (34) has very high computational complexity $\mathcal{O}\left(\prod_i M_i^3 = M^3 \approx 10^{16}\right)$. Expression (34)

[2]The number of parameters L may include also dynamic parameters which are not explicitly defined in the static data model (24)–(25), such as the rates of change (first order derivatives). Also polarization modeling increases the total number of parameters.

[3]A computationally efficient alternative form of the EKF can be formulated in terms of the FIM [32], [33].

also requires memory for storing the full matrix $\mathbf{R} \in \mathbb{C}^{M\times M}$ (and $\mathbf{D} \in \mathbb{C}^{M\times L'}$), resulting in the order of 600 GB (IEEE double precision) assuming the dimensions in Table III ($M = M_f M_T M_R = 197632$, and $L' = 800$).

To facilitate feasible computation of the FIM (32), the positive definite covariance matrix $\mathbf{R}$ (30) can be expressed in terms of its eigenvalue decomposition as

$$\begin{aligned}\mathbf{R}=&\mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathrm{H}} + \sigma^2\mathbf{I}_M\\ =&(\mathbf{U}_R \otimes \mathbf{U}_T \otimes \mathbf{U}_f)(\mathbf{\Lambda}_R \otimes \mathbf{\Lambda}_T \otimes \mathbf{\Lambda}_f + \sigma^2\mathbf{I}_M)\\ &\cdot(\mathbf{U}_R \otimes \mathbf{U}_T \otimes \mathbf{U}_f)^{\mathrm{H}}.\end{aligned} \tag{35}$$

The FIM can thus be expressed as

$$\mathbf{J} = 2\Re\Big\{ \left(\mathbf{D}'_R \lozenge \mathbf{D}'_T \lozenge \mathbf{D}'_f\right)^{\mathrm{H}} \mathbf{\Lambda}^{-1} \left(\mathbf{D}'_R \lozenge \mathbf{D}'_T \lozenge \mathbf{D}'_f\right) \Big\}, \tag{36}$$

where

$$\mathbf{D}'_i = \mathbf{U}_i^{\mathrm{H}}\mathbf{D}_i,\ i \in \{f, T, R\}, \tag{37}$$

and

$$\mathbf{\Lambda} = (\mathbf{\Lambda}_R \otimes \mathbf{\Lambda}_T \otimes \mathbf{\Lambda}_f + \sigma^2\mathbf{I}_M). \tag{38}$$

This form has computational complexity $\mathcal{O}\left(L'^2 \prod_i M_i \approx 10^{11}\right)$ (with dimensions in Table III), and it requires storing the full matrix $\mathbf{D}$ (33). Let us reshape the diagonal elements of $\mathbf{\Lambda}^{-1}$ in (38) into a tensor $\mathcal{L} \in \mathbb{C}^{M_f\times M_T\times M_R}$ as

$$\mathcal{L} = \text{reshape}\left(\text{diag}\left(\mathbf{\Lambda}^{-1}\right), \{M_f, M_T, M_R\}\right). \tag{39}$$

This tensor is, in general, of full rank[4]. One feasible solution for computing the FIM (36) is then given by

$$\begin{aligned}\mathbf{J}_o =&2\Re\Big\{ \sum_{m_f=1}^{M_f} \Bigg[\mathbf{d}'^{\mathrm{H}}_{f_{m_f}} \mathbf{d}'_{f_{m_f}} \odot \sum_{m_T=1}^{M_T} \Big[\mathbf{d}'^{\mathrm{H}}_{T_{m_T}} \mathbf{d}'_{T_{m_T}}\\ &\odot \sum_{m_R=1}^{M_R} \left[l_{m_f,m_T,m_R} \cdot \mathbf{d}'^{\mathrm{H}}_{R_{m_R}} \mathbf{d}'_{R_{m_R}}\right]\Big]\Bigg]\Big\},\end{aligned} \tag{40}$$

where $\mathbf{d}'_{i_{m_i}}$, $i \in \{f, T, R\}$, denotes the m_i^{th} row of the matrix $\mathbf{D}'_i$ in (37). Expression (40) is exact and does not require storing the full matrix $\mathbf{D}$ in (33), but has the same (high) computational complexity

$$\mathcal{O}\left(L'^2 \prod_i M_i \approx 10^{11}\right) \tag{41}$$

as (36).

2) Applying Tensor Decompositions for Solving the FIM: The PARATREE model can be applied to reduce the computational complexity of (40). The tensor $\mathcal{L}$ in (39) is decomposed into a PARATREE model with a single matrix of basis vectors $\mathbf{L}^{(f)} \in \mathbb{R}^{M_f\times R_f}$ for the f-mode, and R_f matrices $\mathbf{L}_{r_f}^{(T)} \in \mathbb{R}^{M_T\times R_T}$ and $\mathbf{L}_{r_f}^{(R)} \in \mathbb{R}^{M_R\times R_T}$ for the T- and R-modes (see Table I for details). Then a PARATREE approximation $\mathbf{J}_{PT}$ for the FIM can be expressed as

$$\begin{aligned}\mathbf{J}_{PT} = 2\Re\Bigg\{ &\sum_{r_f=1}^{R_f} \Bigg[\left(\mathbf{D}'^{\mathrm{H}}_f \mathbf{\Lambda}^{(f)}_{r_f} \mathbf{D}'_f\right)\\ \odot &\sum_{r_T}^{R_T} \Big[\left(\mathbf{D}'^{\mathrm{H}}_T \mathbf{\Lambda}^{(T)}_{r_f,r_T} \mathbf{D}'_T\right) \odot \left(\mathbf{D}'^{\mathrm{H}}_R \text{diag}\mathbf{\Lambda}^{(R)}_{r_f,r_T} \mathbf{D}'_R\right) \Big]\Bigg]\Bigg\},\end{aligned} \tag{42}$$

where $\mathbf{\Lambda}^{(R)}_{r_f,r_T} = \text{diag}((\mathbf{L}^{(R)}_{r_f})_{r_T})$ denotes a diagonal matrix formed from the r_T^{th} column of $\mathbf{L}^{(R)}_{r_f}$ etc. This solution has computational complexity

$$\mathcal{O}\left(L'^2 \cdot 2R_f(M_f + R_T(M_T + M_R)) \approx 8 \cdot 10^9\right), \tag{43}$$

where values $R_f = 12$ and $R_T = 5$ were used. These values correspond to $\epsilon_r \approx 10^{-5}$ to compare with PARAFAC and HOSVD, see Fig. 10. In practice $R_f << M_f$ and $R_T << M_T$, see [36], which provides a significant reduction in computational complexity compared to (41).

Similar to (42), an expression for evaluating the FIM using PARAFAC is given by

$$\begin{aligned}\mathbf{J}_{PF} = 2\Re\Bigg\{ &\sum_{r}^{R} \Bigg[\left(\mathbf{D}'^{\mathrm{H}}_f \mathbf{\Lambda}^{(f)}_{r} \mathbf{D}'_f\right)\\ &\odot \left(\mathbf{D}'^{\mathrm{H}}_T \mathbf{\Lambda}^{(T)}_{r} \mathbf{D}'_T\right) \odot \left(\mathbf{D}'^{\mathrm{H}}_R \mathbf{\Lambda}^{(R)}_{r} \mathbf{D}'_R\right) \Bigg]\Bigg\},\end{aligned} \tag{44}$$

which has computational complexity in the order of

$$\mathcal{O}\left(L'^2 \cdot 2R(M_T + M_R + M_f) \approx 2 \cdot 10^{10}\right), \tag{45}$$

where $R = 50 \sim \epsilon_r \approx 10^{-5}$ was used (see Fig. 10).

Furthermore, a computational strategy for evaluating the FIM using HOSVD is given by

$$\begin{aligned}\mathbf{J}_{HO} = 2\Re\Bigg\{ &\sum_{r_f}^{R_f} \Bigg[\left(\mathbf{D}'^{\mathrm{H}}_f \mathbf{\Lambda}^{(f)}_{r_f} \mathbf{D}'_f\right)\\ &\odot \sum_{r_T}^{R_T} \Big[\left(\mathbf{D}'^{\mathrm{H}}_T \mathbf{\Lambda}^{(T)}_{r_T} \mathbf{D}'_T\right)\\ &\odot \sum_{r_R}^{R_R} \left[s_{r_f,r_T,r_R} \cdot \left(\mathbf{D}'^{\mathrm{H}}_R \mathbf{\Lambda}^{(R)}_{r_R} \mathbf{D}'_R\right)\right]\Big]\Bigg]\Bigg\},\end{aligned} \tag{46}$$

where s_{r_f,r_T,r_R} denotes an element of the core tensor. In (46) only the terms corresponding to a nonzero core tensor value need to be evaluated. Hence, the computational complexity of (46) is given by

$$\begin{aligned}\mathcal{O}\Big(L'^2 \cdot 2\Big(&R_f M_f + \sum_{r_f} \left[R^{(T)}_{r_f} M_T + M_R \cdot \sum_{r_T} R^{(R)}_{r_f,r_T}\right]\Big)\\ &\approx 3 \cdot 10^{10}\Big),\end{aligned} \tag{47}$$

where the numerical value is again based on the decomposition yielding $\epsilon_r \approx 10^{-5}$ (see also Fig. 9).

It should be mentioned that further reduction in computational complexity using HOSVD could be achieved, at the cost of very high memory consumption, if all the terms

[4]In [36] it is shown that if the term $\sigma^2\mathbf{I}$ would not be present, $\mathcal{L}$ would be a rank-1 tensor, and solving (36) becomes computationally attractive.

$\mathbf{D'}_i^{\mathrm{H}} \mathbf{\Lambda}_{r_i}^{(i)} \mathbf{D'}_i$ would be stored while computed for the first time. However, with the current system dimensions in Table III (and also considering that a much higher value for L' is possible), the memory requirements for such strategy become prohibitive. Given the proposed approaches to approximate the FIM using PARATREE (42), PARAFAC (44) and HOSVD (46), the PARATREE/SUSVD provides the best performance in terms of computational complexity. The difference is even more evident for smaller ϵ_r, as illustrated in the results in Section V.

C. Noise Suppression of Multidimensional Radio Channel Measurements

Another novel application to utilize the PARATREE/SUSVD is noise suppression for MIMO channel sounding measurement data [37], [38]. These data are often directly used, e.g., in link-level simulations of a wireless communication system (as opposed to drawing channel realizations based on measurement-based parametric modeling [39]). The tensor decomposition based filtering is very useful for enhancing the SNR (Signal-to-Noise Ratio) of the measured channel data to be used in the simulator. This allows a wider range of noise power (or other interfering signals) to be defined within the simulation.

For convenience, the time index k in (29) is dropped and a single snapshot $\mathcal{H}$ of a channel sounding measurement is considered. The nominal SNR of the measurement is defined as

$$S_{dB}\left(\mathcal{H}\right) = 10 \cdot \log_{10}\left(\frac{P_H - P_N}{P_N}\right), \tag{48}$$

where $P_H = ||\mathcal{H}||_F^2$ is the total power in the measurement, and $P_N = ||\mathcal{H}_N||_F^2$ is the power of the measurement noise. These quantities are assumed to be known, which is a valid assumption in channel sounding[5]. The suppression of the measurement noise is achieved by the following procedure:

1) Compute the SUSVD of $\mathcal{H}$, as described in Table I.
2) Define a threshold ϵ_r (20) for selecting the factors, i.e. the signal subspace, for the approximation.
   - Here $\epsilon_r = \sqrt{\frac{P_N}{P_H}}$ is chosen, i.e., only the factors whose cumulative power exceeds the noise power are included in the decomposition.
3) Approximate $\mathcal{H}$ by $\mathcal{H}_A$, with the procedure described in Table II.

This filter can be equivalently expressed using a projector matrix to the signal subspace [37] defined as

$$\mathbf{\Pi}_A = \left(\mathbf{U}_A^{(R)} \lozenge \mathbf{U}_A^{(T)} \lozenge \mathbf{U}_A^{(f)}\right)\left(\mathbf{U}_A^{(R)} \lozenge \mathbf{U}_A^{(T)} \lozenge \mathbf{U}_A^{(f)}\right)^{\mathrm{H}}, \tag{49}$$

where the matrices $\mathbf{U}_A^{(i)} \in \mathbb{C}^{M_i \times R_A}$ contain all the R_A (22) factor combinations expanded in PARAFAC fashion (redundancy in columns possible). The filtered channel estimate is then given by

$$\mathrm{vec}\left(\mathcal{H}_A\right) = \mathbf{\Pi}_A \mathrm{vec}\left(\mathcal{H}\right). \tag{50}$$

[5]The noise power P_N may be assessed, e.g., by sampling while Tx is off, or by estimating it from excess delay samples.

This method effectively suppresses the measurement noise as is shown in Section V-B. It will also be shown that the approximation is beneficial in terms of data compression.

V. Results and Validation using Real Data

A. Computational Gain in FIM Computation

As discussed in Section IV-B, the PARATREE approximation can be used to reduce computational complexity of calculating the FIM in certain applications. Fig. 9 illustrates the performance of different tensor decompositions in terms of computational complexity while computing the Fisher information matrix (excluding the complexity of computing the tensor decomposition). The complexities are given by the equations (41) (exact), (43) (PARATREE), and (45) (PARAFAC), and are normalized by the complexity of the exact solution (41). The complexities are averaged by approximating 1500 realizations of the tensor $\mathcal{L}$ (39) with different relative approximation errors (20). Additionally, the influence of varying the dimension order is studied for PARATREE. It can be seen that the complexity is clearly lowest while applying PARATREE with the largest dimension decomposed first. The comparison results for PARAFAC-ALS are obtained using random initialization of the factors, and for two R values ($R = \{10, 50\}$) only. For higher rank approximations ($R = \{100, 200\}$), the algorithm failed to converge in a reasonable time. The considered HOSVD strategy (46) has the highest complexity in this task, and the complexity can even grow higher than that of the exact solution.

Fig. 10 shows the average number of basis vectors for the first mode (R_f), as well as the total number of factors (R_{tot}) of the PARATREE and HOSVD decompositions as a function of the normalized approximation error. Also the rank of the PARAFAC decomposition ($R_{PARAFAC}$) is shown. The total number of factors is given by (22) for PARATREE, whereas for HOSVD it equals the number of nonzero elements in the core tensor of the approximation. The numbers corresponding

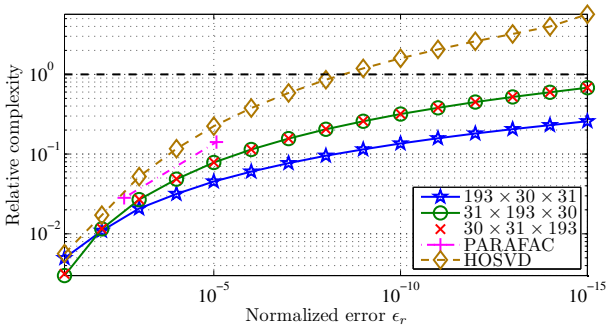


Fig. 9: Complexity relative to an exact solution (41) for computing the Fisher information matrix with PARATREE (43), PARAFAC (45), and HOSVD (47). Having the largest dimension first yields the most reduction in computational complexity with PARATREE. For the PARAFAC-ALS, only the values corresponding to $R = 10$ and $R = 50$ are shown for comparison. The considered HOSVD strategy (46) has the highest complexity in this task, and the complexity can even grow higher than that of the exact solution.

to $\epsilon_r = 10^{-5}$ were used in the complexity evaluation in Section IV-B ((43), (45), and (47)).

B. Filtering of MIMO Channel Sounding Measurements

Fig. 11 shows the Power-Delay Profile (PDP) of a MIMO channel sounding measurement before and after applying a PARATREE model for subspace-based noise suppression (see Section IV-C). It can be observed that the SNR is significantly improved (noise is suppressed) after filtering, especially in the beginning of the measurement, where the initial SNR is also very low.

Fig. 12 shows the PDP for three time instants of the same measurement as Fig. 11. The two filtered PARATREE estimates of the measurement $\hat{\mathcal{H}}_1$ and $\hat{\mathcal{H}}_2$ refer to different ordering of the tensor modes as:

1) Set the significantly largest dimension as the first ($M_1 = M_f = 193$ for $\hat{\mathcal{H}}_1$)
2) Set a lower dimension as the first ($M_1 = M_T = 31$ for $\hat{\mathcal{H}}_2$).

It can be observed that the first approach yields consistently better performance, although both approaches provide excellent performance while applied to low SNR measurements. The HOSVD ($\mathcal{H}_{HO}$) also provides good results, the improvement in SNR being slightly less than for $\mathcal{H}_1$ (PARATREE).

The upper part of Fig. 13 shows the SNR improvement achieved with the PARATREE as well as HOSVD filtering as a function of the original measurement SNR (48). It can be observed that $\hat{\mathcal{H}}_1$ approximation achieves better results than $\hat{\mathcal{H}}_2$, and it also outperforms HOSVD by a couple of dBs in the low SNR regime.

The lower part of Fig. 13 illustrates the data compression percentage (one minus the number of bits required by the stored PARATREE or HOSVD structures divided by the number of bits in the original data tensor) as a function of initial SNR. In poor SNR conditions ($SNR_{dB}(\mathcal{H}) < 0$ dB), the compression of data is close to 95 % and even in good SNR ($SNR_{dB}(\mathcal{H}) > 25$ dB), the compression is in the order of 60 %. The HOSVD outperforms PARATREE in terms of compression in the high SNR regime. This is because the increase in the amount of data results mainly from including more elements in the sparse core tensor instead of increasing the number of actual basis vectors, whereas in PARATREE the number of basis vectors has to increase to obtain better fit.

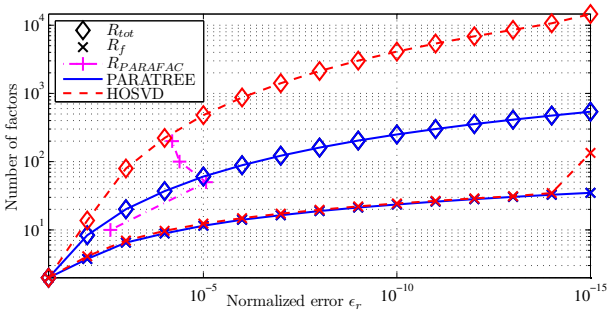
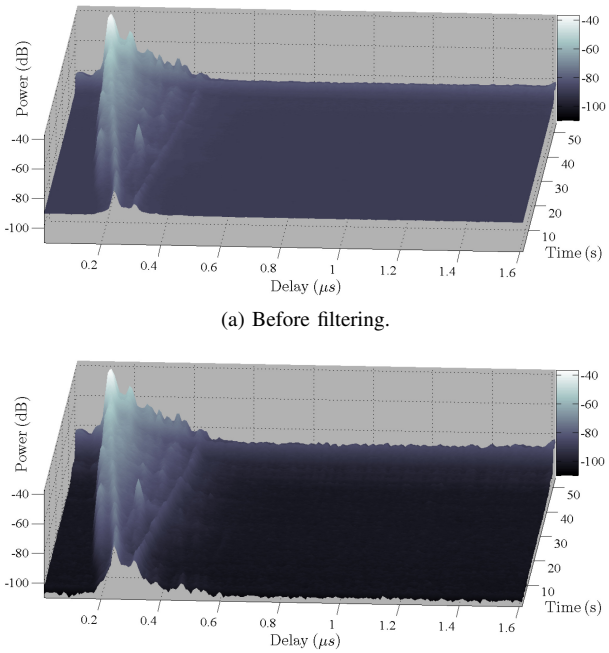


Fig. 10: The average of the total number of factors (R_{tot}) for different decompositions as a function of normalized error, as well as the number of basis vectors in the first mode (R_f) for PARATREE and HOSVD. The value of R_f is similar for PARATREE and HOSVD, whereas R_{tot} is higher (by a factor of R_R) for HOSVD. PARAFAC-ALS often failed to converge with $R = 100$ and $R = 200$.



(a) Before filtering.

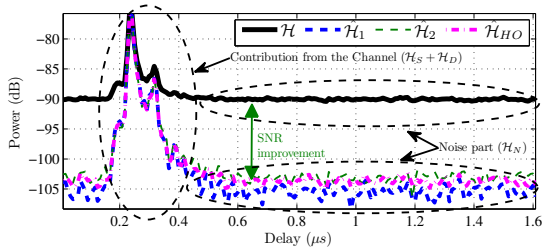
(b) After filtering.

Fig. 11: Power-Delay Profile over measurement time before and after filtering. The PDPs are averaged over all Tx-Rx channels. After filtering the SNR is improved by $\sim$ 15 dB in the low SNR regime.

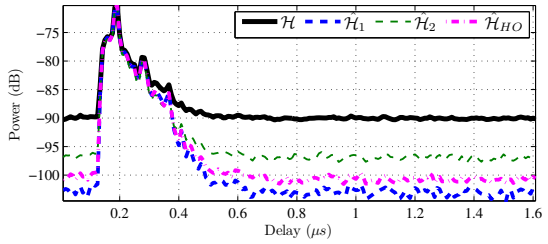
VI. Conclusions

This paper introduces Sequential Unfolding SVD (SUSVD) — a novel orthogonal, non-iterative tensor decomposition technique, which is scalable to arbitrary high dimensional tensors. The SUSVD provides a tensor model with hierarchical tree structure between the factors in different dimensions. The new model, named as PARATREE, is related to the family of PARAFAC tensor models. Links between PARATREE and other existing models are established in the paper as well.

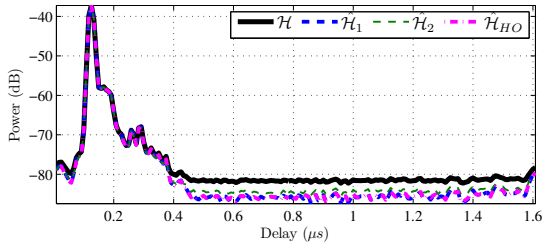
The PARATREE model can be used for flexible low rank tensor approximation with precisely defined approximation error level through deflation of the orthogonal rank-1 components. The low rank PARATREE approximation can be used for reducing the computational complexity in high dimensional problems, measurement noise suppression, data compression, as well as providing insight on structures and dependencies in the data. Whereas similar strategy is obtainable for Tucker3 based HOSVD, the PARATREE clearly outperforms HOSVD (as well as PARAFAC) in terms of computational complexity while solving the Fisher Information matrix in a given application example. This is due to rich structure of the basis vectors resulting from the SUSVD, yielding more precise approximation with smaller tensor rank. Another novel



(a) Poor SNR ($t = 6$ s).



(b) Low SNR ($t = 26$ s).



(c) High SNR ($t = 50$ s).

Fig. 12: Power-delay profiles from the measurement in Fig. 11 at three time instants with different SNR conditions. $\mathcal{H}$ denotes measured data, $\hat{\mathcal{H}}_1$ and $\hat{\mathcal{H}}_2$ are PARATREE estimates with $M_1 = M_f = 193$, and $M_1 = M_T = 31$, respectively, and $\hat{\mathcal{H}}_{HO}$ denotes HOSVD estimate. Note that the power (y-)axis have different scales. Filtering provides the most improvement for measurements with low SNR (see also Fig. 13). PARATREE provides slightly better SNR improvement than HOSVD.

application example involves noise suppression for tensor valued MIMO channel sounding measurements. The results for this application also indicate the superiority of PARATREE over HOSVD.

Appendix A
Connections Between Models

A. Relation between PARATREE and Decompositions in Block Terms

The PARATREE-model can also be expressed in terms of the *block term decomposition* notation introduced recently in [12]. Hence, the expression (10) for the $N = 3$ PARATREE is equivalent to a rank-$(1, R_{b_{|r_a}}, R_{b_{|r_a}})$ block term decomposition [12]

$$\mathcal{X} = \sum_{r_a=1}^{R_a} \mathbf{a}_{r_a} \circ \left(\mathbf{B}_{r_a} \cdot \mathbf{C}_{r_a}^{\mathrm{T}} \right), \tag{51}$$

where the R_a pairs of matrices $\mathbf{B}_{r_a}$ and $\mathbf{C}_{r_a}$ may have different number of columns ($R_{r_a}^{(b)}$). The uniqueness properties (in the generic sense, without considering orthogonality) of this decomposition were also addressed in [12]. In this paper, the notation using the vector outer products (10) facilitates better the description of the SUSVD algorithm (see Section III-C) as well as the derivation of a low rank approximation strategy based on the product of singular values related to each branch (see Section III-D).

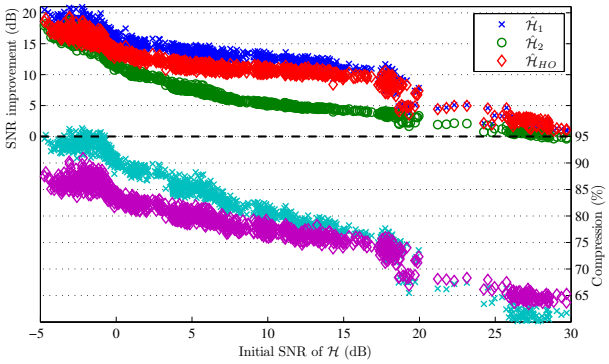


Fig. 13: SNR improvement (upper part) and data compression ratio (lower part) as a function of the initial measured SNR (48). At low SNR values, the PARATREE $\mathcal{H}_1$ provides the best performance both in terms of SNR improvement and data compression.

B. Relation between PARATREE and CONFAC

The relationship between the PARATREE (see Section II-D) and CONFAC [9] models (for $N = 3$) can be shown by defining the *factor matrices* [9] as

$$\begin{aligned} \mathbf{A}^{(1)} &= [\mathbf{a}_{r_1}^{(1)} \ \ldots \mathbf{a}_{R_1}^{(1)}] \in \mathbb{C}^{M_1 \times R_1}, \\ \mathbf{A}^{(2)} &= [\mathbf{A}_{r_1}^{(2)} \ \ldots \ \mathbf{A}_{R_1}^{(2)}] \in \mathbb{C}^{M_2 \times F}, \\ \mathbf{A}^{(3)} &= [\mathbf{A}_{r_1}^{(3)} \ \ldots \ \mathbf{A}_{R_1}^{(3)}] \in \mathbb{C}^{M_3 \times F}, \end{aligned} \tag{52}$$

with $F = R_1 \sum_{r_1} R_{r_1}^{(2)}$. The *constraint matrices* [9] are given by $\mathbf{\Phi} = \mathbf{\Omega} = \mathbf{I}_F$, and $\mathbf{\Psi} = [\mathbf{E}_{r_1} \ \ldots \ \mathbf{E}_{R_1}]$, where the $R_1 \times R_{r_1}^{(2)}$ matrices $\mathbf{E}_{r_1}$ consist of repeating $R_{r_1}^{(2)}$ times the canonical vectors $\mathbf{e}_{r_1}$[6].

Appendix B
Orthogonality of SUSVD

The orthogonality of the SUSVD decomposition is proven as follows.

Lemma 1: Let $\mathbf{A} \in \mathbb{C}^{M_a \times R_a}$, $\mathbf{B} \in \mathbb{C}^{M_b \times R_b}$, and $\mathbf{C} \in \mathbb{C}^{M_c \times R_c}$, have orthogonal columns $\mathbf{a}_i \perp \mathbf{a}_j$, $\mathbf{b}_k \perp \mathbf{b}_l$, and $\mathbf{c}_m \perp \mathbf{c}_n$ (for all $i \neq j$, $k \neq l$, and $m \neq n$). Then the Kronecker products of these columns are orthogonal

$$(\mathbf{a}_i \otimes \mathbf{b}_k \otimes \mathbf{c}_m) \perp (\mathbf{a}_j \otimes \mathbf{b}_l \otimes \mathbf{c}_n), \tag{53}$$

if at least one of the conditions $i \neq j$, or $k \neq l$, or $m \neq n$ holds.

[6] A canonical vector $\mathbf{e}_n$ is a unitary vector containing a 1 at n^{th} position and zeros elsewhere.

Proof: The proof is given by evaluating the inner product

$$\begin{aligned}
&\left(\mathbf{a}_i \otimes \mathbf{b}_k \otimes \mathbf{c}_m\right)^{\mathrm{H}} \left(\mathbf{a}_j \otimes \mathbf{b}_l \otimes \mathbf{c}_n\right) \\
&= \left(\mathbf{a}_i^{\mathrm{H}} \otimes \left[\mathbf{b}_k \otimes \mathbf{c}_m\right]^{\mathrm{H}}\right) \left(\mathbf{a}_j \otimes \left[\mathbf{b}_l \otimes \mathbf{c}_n\right]\right) \\
&= \left(\mathbf{a}_i^{\mathrm{H}} \mathbf{a}_j\right) \otimes \left(\left[\mathbf{b}_k \otimes \mathbf{c}_m\right]^{\mathrm{H}} \otimes \left[\mathbf{b}_l \otimes \mathbf{c}_n\right]\right) \\
&= \left(\mathbf{a}_i^{\mathrm{H}} \mathbf{a}_j\right) \otimes \left(\mathbf{b}_k^{\mathrm{H}} \mathbf{b}_l\right) \otimes \left(\mathbf{c}_m^{\mathrm{H}} \mathbf{c}_n\right). \qquad (54)
\end{aligned}$$

The expression (54) is zero if any of the three terms in the Kronecker product of the last row are zero. Hence, the vectors $(\mathbf{a}_i \otimes \mathbf{b}_k \otimes \mathbf{c}_m)$ and $(\mathbf{a}_j \otimes \mathbf{b}_l \otimes \mathbf{c}_n)$ are always orthogonal ((54) is nonzero), except when $i = j$, and $k = l$, and $m = n$. ∎

Now let us consider the outer product form of PARATREE for $N = 3$

$$\begin{aligned}
\mathcal{X} &= \sum_{r_1=1}^{R_1} \sigma_{r_1}^{(1)} \cdot \mathbf{u}_{r_1}^{(1)} \circ \sum_{r_2=1}^{R_2} \left(\sigma_{r_1,r_2}^{(2)} \cdot \mathbf{u}_{r_1,r_2}^{(2)} \circ \mathbf{u}_{r_1,r_2}^{(3)}\right) \qquad (55) \\
&= \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \left(\sigma_{r_1}^{(1)} \sigma_{r_1,r_2}^{(2)} \cdot \mathbf{u}_{r_1}^{(1)} \circ \mathbf{u}_{r_1,r_2}^{(2)} \circ \mathbf{u}_{r_1,r_2}^{(3)}\right).
\end{aligned}$$

The orthogonality of the conventional 2-D SVD decomposition within the SUSVD computation yields

$$\left(\mathbf{u}_{r_1,\ldots,r_n}^{(n)}\right)^{\mathrm{H}} \left(\mathbf{u}_{r_1,\ldots,r_n}^{(n)}\right) = \qquad (56)$$

$$\begin{cases} = 0, & r_n \neq q_n,\ \{r_1,\ldots,r_{n-1}\} = \{q_1,\ldots,q_{n-1}\}, \\ = 1, & \{r_1,\ldots,r_n\} = \{q_1,\ldots,q_n\}, \\ \neq 0, & \{r_1,\ldots,r_{n-1}\} \neq \{q_1,\ldots,q_{n-1}\}. \end{cases}$$

Given the above conditions (56) along with Lemma 1, yields, for the three-way example in (55),

$$\left(\mathbf{u}_{r_1}^{(1)} \otimes \mathbf{u}_{r_1,r_2}^{(2)} \otimes \mathbf{u}_{r_1,r_2}^{(3)}\right)^{\mathrm{H}} \left(\mathbf{u}_{q_1}^{(1)} \otimes \mathbf{u}_{q_1,q_2}^{(2)} \otimes \mathbf{u}_{q_1,q_2}^{(3)}\right) = 0, \qquad (57)$$

for all values except if $\{r_1, r_2\} = \{q_1, q_2\}$. The summation indexes in (55) result only in values for which the basis vectors are orthogonal in at least one of the dimensions. Hence, all the complete Kronecker (or outer) product terms in the decomposition are orthogonal. The proof applies eqivalently to higher order tensors.

ACKNOWLEDGMENT

The authors would like to acknowledge the effort of Veli-Matti Kolmonen, Jukka Koivunen, Katsuyuki Haneda, Peter Almers, and Mário Costa who were involved in performing the channel sounding measurements used in this paper in September 2007 at Espoo, Finland.

REFERENCES

[1] 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, 1970.

[2] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 36, pp. 279–311, 1966.

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

[4] J. D. Carroll and J. J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart-Young' decomposition," *Psychometrika*, vol. 35, pp. 283–319, 1970.

[5] J. D. Carroll, S. Pruzansky, and J. Kruskal, "Candelinc: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters," *Psychometrika*, vol. 45, no. 1, pp. 3–24, Mar. 1980.

[6] N. K. M. Faber, R. Bro, and P. K. Hopke, "Recent developments in CANDECOMP/PARAFAC algorithms: a critical review," *Chemometrics and Intelligent Laboratory Systems*, vol. 65, no. 1, pp. 119–137, Jan. 2003.

[7] A. L. F. de Almeida, G. Favier, J. C. M. Mota, and R. L. de Lacerda, "Estimation of frequency-selective block-fading MIMO channels using PARAFAC modeling and alternating least squares," in *The 40th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Oct.–Nov. 2006, pp. 1630–1634.

[8] N. Sidiropoulos, R. Bro, and G. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Transactions on Signal Processing*, vol. 48, no. 8, pp. 2377–2388, Aug. 2000.

[9] A. de Almeida, G. Favier, and J. Mota, "A constrained factor decomposition with application to MIMO antenna systems," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2429–2442, Jun. 2008.

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

[11] 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," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3198–3213, Jul. 2008.

[12] L. De Lathauwer, "Decompositions of a higher-order tensor in block terms — Part II: Definitions and uniqueness," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1033–1066, 2008. [Online]. Available: http://publi-etis.ensea.fr/2008/De 08f

[13] Mathworks. Matlab. [Online]. Available: http://www.mathworks.com

[14] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, Sept. 2009, to appear.

[15] A. Smilde, R. Bro, and P. Geladi, *Multiway Analysis with Applications in the Chemical Sciences*. Chichester, England: John Wiley and Sons, Ltd, 2004, 381 p.

[16] N. D. Sidiropoulos and R. Bro, "On the uniqueness of multilinear decomposition of N-way arrays," *Journal of Chemometrics*, vol. 14, no. 3, pp. 229–239, May 2000.

[17] A. Stegeman, J. Berge, and L. Lathauwer, "Sufficient conditions for uniqueness in candecomp/parafac and indscal with random component matrices," *Psychometrika*, vol. 71, no. 2, pp. 219–229, Jun. 2006.

[18] L. De Lathauwer, B. De Moor, and J. Vandewalle, "On the best rank-1 and rank-$(r_1, r_2, ..., r_n)$ approximation of higher-order tensors," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1324–1342, 2000. [Online]. Available: http://link.aip.org/link/?SML/21/1324/1

[19] P. Comon and J. ten Berge, "Generic and typical ranks of three-way arrays," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008)*, Las Vegas, NV, Apr. 2008, pp. 3313–3316.

[20] T. G. Kolda, "Multilinear operators for higher-order decompositions," Sandia National Laboratories, Albuquerque, NM and Livermore, CA, Technical Report SAND2006-2081, Apr. 2006. [Online]. Available: http://csmr.ca.sandia.gov/~tgkolda/pubs/

[21] R. Bro and H. A. L. Kiers, "A new efficient method for determining the number of components in PARAFAC models," *Journal of Chemometrics*, vol. 17, no. 5, pp. 274–286, 2003.

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

[23] T. G. Kolda, "A counterexample to the possibility of an extension of the Eckart-Young low-rank approximation theorem for the orthogonal rank tensor decomposition," *SIAM Journal on Matrix Analysis and Applications*, vol. 24, no. 3, pp. 762–767, January 2003.

[24] R. Bro, R. A. Harshman, and N. D. Sidiropoulos, "Modeling multi-way data with linearly dependent loadings," Dept. of Dairy and Food Science, The Royal Veterinary and Agricultural University, Copenhagen, Denmark, Technical Report 2005-176, 2005.

[25] L. De Lathauwer, B. De Moor, and J. Vandewalle, "Computation of the canonical decomposition by means of a simultaneous generalized Schur decompositition," *SIAM J. Matrix Anal. Appl.*, vol. 26, no. 2, pp. 295–327, 2004. [Online]. Available: http://publi-etis.ensea.fr/2004/DDV04

[26] P. 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.

[27] G. Tomasi and R. Bro, "A comparison of algorithms for fitting the PARAFAC model," *Computational Statistics & Data Analysis*, vol. 50, no. 7, pp. 1700–1734, Apr. 2006.
[28] P. K. Hopke, P. Paatero, H. Jia, R. T. Ross, and R. A. Harshman, "Three-way (PARAFAC) factor analysis: examination and comparison of alternative computational methods as applied to ill-conditioned data," *Chemometrics and Intelligent Laboratory Systems*, vol. 43, pp. 25–42, Sept. 1998.
[29] A. Richter, "Estimation of radio channel parameters: Models and algorithms," Ph. D. dissertation, Technischen Universität Ilmenau, Germany, May 2005, ISBN 3-938843-02-0. [Online]. Available: www.db-thueringen.de
[30] V.-M. Kolmonen, J. Kivinen, L. Vuokko, and P. Vainikainen, "5.3-GHz MIMO radio channel sounder," *IEEE Transactions on Instrumentation and Measurement*, vol. 55, no. 4, pp. 1263–1269, Aug. 2006.
[31] J. Koivunen, P. Almers, V.-M. Kolmonen, J. Salmi, A. Richter, F. Tufvesson, P. Suvikunnas, A. Molisch, and P. Vainikanen, "Dynamic multilink indoor MIMO measurements at 5.3 GHz," in *The 2nd European Conference on Antennas and Propagation (EuCAP 2007)*, Edinburgh, UK, Nov. 2007, pp. 1–6.
[32] J. Salmi, A. Richter, and V. Koivunen, "Detection and tracking of MIMO propagation path parameters using state-space approach," *IEEE Transactions on Signal Processing*, vol. 57, no. 4, pp. 1538–1550, April 2009.
[33] ——, "Enhanced tracking of radio propagation path parameters using state-space modeling," in *The 14th European Signal Processing Conference (EUSIPCO 2006)*, Florence, Italy, Sept. 4–8 2006.
[34] M. Landmann, "Limitations of experimental channel characterization," Ph. D. dissertation, Technischen Universität Ilmenau, Ilmenau, Germany, Mar. 2008. [Online]. Available: http://www.db-thueringen.de
[35] A. Richter, J. Salmi, and V. Koivunen, "ML estimation of covariance matrix for tensor valued signals in noise," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2008)*, Las Vegas, USA, Mar. 31–Apr. 4 2008, pp. 2349–2352.
[36] J. Salmi, A. Richter, and V. Koivunen, "Tracking of MIMO propagation parameters under spatio-temporal scattering model," in *The 41st Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2007, pp. 666–670.
[37] A. Richter, J. Salmi, and V. Koivunen, "Tensor decomposition of MIMO channel sounding measurements and its applications," in *URSI General Assembly*, Chicago, USA, Aug. 7–16 2008.
[38] J. Salmi, A. Richter, and V. Koivunen, "Sequential unfolding SVD for low rank orthogonal tensor approximation," in *The 42nd Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Oct. 2008, pp. 1713–1717.
[39] U. Trautwein, C. Schneider, and R. Thomä, "Measurement-based performance evaluation of advanced MIMO transceiver designs," *EURASIP Journal on Applied Signal Processing*, vol. 2005, no. 1, pp. 1712–1724, 2005.

Jussi Salmi (S'05) was born in Finland in 1981. He received his M.Sc. degree with honors from Helsinki University of Technology, Dept. of Electrical and Communications Engineering, Espoo, Finland, in 2005, and is currently finalizing his Ph.D. degree. From 2004 to 2005, he worked as a Research Assistant at Radio Laboratory, Helsinki University of Technology. Since 2005 he has held a Researcher position at Dept. of Signal Processing and Acoustics, Helsinki University of Technology. Since 2007 he has been a member of Graduate School in Electronics, Telecommunications and Automation (GETA). His current research interests include measurement based MIMO channel modeling, parameter estimation, analysis of interference limited multiuser MIMO measurements as well as tensor modeling and decomposition techniques. He has authored a paper receiving the Best Student Paper Award in (EUSIPCO'06), and co-authored a paper receiving the Best Paper Award in Propagation (EuCAP'06).

Andreas Richter (M04, SM08) received the Dipl.-Ing. (M.Sc.) degree in electrical engineering and the Dr.-Ing. (Ph.D.) degree (summa cum laude) from Technische Universität Ilmenau, Ilmenau, Germany, in 1995 and 2005, respectively. From 1995 to 2004, he worked as a Research Assistant at Electronic Measurement Laboratory of Technische Universität Ilmenau. From July to October 2001, he was a Guest Researcher at NTT DoCoMo's, Wireless Laboratories, Yokosuka, Japan. From 2004 to 2008, he has been working as a Senior Research Fellow in the Statistical Signal Processing Laboratory at Helsinki University of Technology, Finland. Since August 2008, he is Principal Member of Research Staff at Nokia Research Center, Helsinki. His research interests are in the fields of digital communication, sensor-array-, and statistical signal processing. He has published more than 80 peer-reviewed papers in international scientific conferences and journals. Dr. Richter has co-authored or authored of five papers receiving a Best Paper Award (EPMCC01, ISAP04, PIMRC05, EUSIPCO06, and EuCAP06). In 2005, he received the Siemens Communications Academic Award. He and his former colleagues at Technische Universität Ilmenau received the Thuringian Research Award for Applied Research in 2007 for their work on MIMO channel sounding.

Visa Koivunen (S'87-M'93-SM'98) received his D.Sc. (Tech) degree with honors from the University of Oulu, Dept. of Electrical Engineering. He received the primus doctor (best graduate) award among the doctoral graduates in years 1989-1994. From 1992 to 1995 he was a visiting researcher at the University of Pennsylvania, Philadelphia, USA. From 1997 to 1999 he was an Associate Professor at the Signal Processing Labroratory, Tampere University of Technology. Since 1999 he has been a Professor of Signal Processing at Helsinki University of Technology (HUT), Finland. He is one of the Principal Investigators in SMARAD (Smart Radios and Wireless Systems) Center of Excellence in Radio and Communications Engineering nominated by the Academy of Finland. He has been also adjunct full professor at the University of Pennsylvania, Philadelphia, USA. During his sabbatical leave in 2006–2007 he was Visiting Fellow at Nokia Research Center as well as at Princeton University.

Dr. Koivunen's research interest include statistical, communications and sensor array signal processing. He has published more than 280 papers in international scientific conferences and journals. He co-authored the papers receiving the best paper award in IEEE PIMRC 2005, EUSIPCO 2006 and EuCAP 2006. He has been awarded the IEEE Signal Processing Society best paper award for the year 2007 (co-authored with J. Eriksson). He served as an associate editor for IEEE Signal Processing Letters. He is a member of the editorial board for the Signal Processing journal and Journal of Wireless Communication and Networking. He is also a member of the IEEE Signal Processing for Communication and Networking Technical Committee (SPCOM-TC) and Sensor Array and Multichannel Technical Committee (SAM-TC). He was the general chair of the IEEE SPAWC (Signal Processing Advances in Wireless Communication) 2007 conference in Helsinki, June 2007.