Higher-Order SVD-Based Subspace Estimation to Improve the Parameter Estimation Accuracy in Multidimensional Harmonic Retrieval Problems

Martin Haardt, *Senior Member, IEEE*, Florian Roemer, *Student Member, IEEE*, and Giovanni Del Galdo

***Abstract*—Multidimensional harmonic retrieval problems are encountered in a variety of signal processing applications including radar, sonar, communications, medical imaging, and the estimation of the parameters of the dominant multipath components from MIMO channel measurements. R-dimensional subspace-based methods, such as R-D Unitary ESPRIT, R-D RARE, or R-D MUSIC, are frequently used for this task. Since the measurement data is multidimensional, current approaches require stacking the dimensions into one highly structured matrix. However, in the conventional subspace estimation step, e.g., via an SVD of the latter matrix, this structure is not exploited. In this paper, we define a measurement tensor and estimate the signal subspace through a higher-order SVD. This allows us to exploit the structure inherent in the measurement data already in the first step of the algorithm which leads to better estimates of the signal subspace. We show how the concepts of forward-backward averaging and the mapping of centro-Hermitian matrices to real-valued matrices of the same size can be extended to tensors. As examples, we develop the R-D standard Tensor-ESPRIT and the R-D Unitary Tensor-ESPRIT algorithms. However, these new concepts can be applied to any multidimensional subspace-based parameter estimation scheme. Significant improvements of the resulting parameter estimation accuracy are achieved if there is at least one of the R dimensions, which possesses a number of sensors that is larger than the number of sources. This can already be observed in the two-dimensional case.**

***Index Terms*—Antenna arrays, array signal processing, direction of arrival estimation, harmonic analysis, HOSVD, multidimensional signal processsing, parameter estimation, subspace estimation, Tensor-ESPRIT.**

I. Introduction

HIGH-resolution parameter estimation from R-dimensional (R-D) signals is a task required for a variety of applications, such as estimating the multidimensional parameters of the dominant multipath components from MIMO channel measurements [11], which may be used for geometry-based channel modeling. In this case, the dominant multipath components may be parametrized in terms of their azimuth and elevation angles at the transmitter (directions of departure), their azimuth and elevation angles at the receiver (directions of arrival), as well as the corresponding propagation delays and Doppler shifts, leading to an $R = 6$-dimensional harmonic retrieval problem [11]. Other applications include radar, wireless communications [21], sonar, seismology, and medical imaging. Numerous multidimensional harmonic retrieval techniques have been developed, ranging from Fourier-based methods to parametric high resolution techniques, cf. [20] for an overview. Efficient solutions to this problem are given by subspace-based algorithms like ESPRIT- or MUSIC-based techniques [26] and their multidimensional extensions such as 2-D Unitary ESPRIT [33], R-D Unitary ESPRIT [10], R-D MUSIC [32], R-D multidimensional folding (MDF) [22], or the R-D rank reduction estimator (RARE) [23].

In the existing approaches to subspace-based parameter estimation, the R-D signals are stored in matrices by means of a stacking operation. Obviously, this representation does not account for the R-D grid structure inherent in the data. A more natural approach to store and manipulate multidimensional data is given by tensors. Tensors have already been used in parallel factor (PARAFAC) analysis techniques to obtain important identifiability results for the multidimensional harmonic retrieval problem [13], [19]. Parameter estimates based on the PARAFAC model are often obtained via iterative techniques such as alternating least squares (ALS) [16] that might require many iterations and do not guarantee convergence to the global optimum [24]. Therefore, it has been proposed to use ESPRIT-type methods to initialize these iterative techniques [28]. In contrast to existing tensor approaches using PARAFAC [28], we focus on a direct analogy to the matrix case by using higher-order extensions of the SVD, i.e., the higher-order SVD (HOSVD) [5], and their low-rank approximations [6]. The HOSVD can be viewed as a Tucker3 model [31], which has a long history in tensor analysis [16], [15], [14]. Note that the COMFAC algorithm [4], [29], which is a fast implementation of trilinear ALS, also uses a low-rank approximation based on the Tucker3 model as a preprocessing step. This "Tucker3 compression" is used to speed up the iterative least squares fitting procedure of the PARAFAC model (without the Vandermonde structure that is specific to the harmonic retrieval problem), and thereby avoids a brute force implementation of ALS in the raw data space.

Manuscript received March 12, 2007; revised December 19, 2007. This work was supported in part by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) under Contract HA 2239/1-2. Parts of this paper have been presented at the 40th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, November 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Daniel Fuhrmann.

M. Haardt and F. Roemer are with the Communications Research Laboratory, Ilmenau University of Technology, D-98684 Ilmenau, Germany (e-mail: martin.haardt@tu-ilmenau.de; florian.roemer@tu-ilmenau.de).

G. Del Galdo was with the Communications Research Laboratory, Ilmenau University of Technology, D-98684 Ilmenau, Germany. He is now with Fraunhofer IIS, Erlangen, Germany (e-mail: giovanni@delgaldo.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2008.917929

In this paper, we show that the tensor representation allows us to exploit the structure inherent in the data further. We demonstrate how existing concepts like forward-backward averaging [18], the mapping of complex centro-Hermitian covariance matrices to real-valued matrices of the same size [17], [18], and spatial smoothing [27] can be generalized to tensors. We also discuss how an HOSVD-based low-rank approximation leads to an improved estimate of the signal subspace which can be used to improve any multidimensional subspace-based parameter estimation scheme, e.g., R-D Unitary ESPRIT, R-D MUSIC, or R-D RARE. As examples, we derive the R-D standard Tensor-ESPRIT and the R-D Unitary Tensor-ESPRIT algorithms explicitly.

This paper is organized as follows: Section II introduces the tensor and matrix notation, important tensor operations, and the data model for multidimensional harmonic retrieval problems. Then, HOSVD-based signal subspace estimation techniques are presented in Section III. After the direct data approach in Section III-A, an alternative covariance approach is developed in Section III-B. To this end, covariance tensors, Hermitian tensors, and Hermitian unfoldings are defined. Then, the multidimensional invariance equations required for ESPRIT-type algorithms are described in tensor notation, and the R-D standard Tensor-ESPRIT algorithm is developed in Section IV as an extension of the well-known standard ESPRIT algorithm [26]. After that, concepts like forward-backward averaging and the mapping of complex centro-Hermitian covariance matrices to real-valued matrices of the same size are extended to tensors in Section V. This facilitates the development of the R-D Unitary Tensor-ESPRIT algorithm at the end of this section. Spatial smoothing concepts are extended to the tensor case in Section VI to deal with more than two coherent sources if forward-backward averaging is used (and more than one coherent source without forward-backward averaging) or with situations where not enough snapshots are available. The benefits of the new HOSVD-based parameter estimation algorithms are demonstrated through numerical simulations in Section VII, before the conclusions are drawn in Section VIII. In Appendix I, we compare the HOSVD-based signal subspace estimate with its matrix-valued counterpart. In particular, we identify conditions for which both yield the same signal subspace estimate (as, for example, in the noiseless case). The equivalence of the direct data approach and the covariance approach is shown in Appendix II. Finally, the real-valued invariance equations in tensor notation are derived in Appendix III. They are used in Section V for the development of R-D Unitary Tensor-ESPRIT.

II. Notation and Data Model

A. Tensor and Matrix Notation

In order to facilitate the distinction between scalars, matrices, and tensors, the following notation is used: Scalars are denoted as italic letters $(a, b, \ldots, A, B, \ldots, \alpha, \beta, \ldots)$, column vectors as lower-case bold-face letters $(\boldsymbol{a}, \boldsymbol{b}, \ldots)$, matrices as bold-face capitals $(\boldsymbol{A}, \boldsymbol{B}, \ldots)$, and tensors are written as bold-face calligraphic letters $(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}, \ldots)$. Lower-order parts are consistently named: the (i, j)-element of the matrix $\boldsymbol{A}$ is denoted as $a_{i,j}$ and the (i, j, k)-element of a third-order tensor $\boldsymbol{\mathcal{B}}$ as $b_{i,j,k}$.

We use the superscripts $^T, ^H, ^*, ^{-1}, ^+$ for transposition, Hermitian transposition, complex conjugation, matrix inversion, and the Moore–Penrose pseudo-inverse of a matrix, respectively. Moreover, the Kronecker product of two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ is denoted as $\boldsymbol{A} \otimes \boldsymbol{B}$ and the Khatri–Rao product (column-wise Kronecker product) as $\boldsymbol{A} \diamond \boldsymbol{B}$. An n-mode vector of an $(I_1 \times I_2 \times \ldots \times I_N)$-dimensional tensor $\boldsymbol{\mathcal{A}}$ is an I_n-dimensional vector obtained from $\boldsymbol{\mathcal{A}}$ by varying the index i_n and keeping the other indices fixed.

The tensor operations we use are consistent with [5].

- **The outer product** of the tensors $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$ and $\boldsymbol{\mathcal{B}} \in \mathbb{C}^{J_1 \times J_2 \times \ldots \times J_M}$ is given by

$$\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{A}} \circ \boldsymbol{\mathcal{B}} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N \times J_1 \times J_2 \times \cdots \times J_M}, \quad \text{where}$$
$$c_{i_1,i_2,\ldots,i_N,j_1,j_2,\ldots,j_M} = a_{i_1,i_2,\ldots,i_N} \cdot b_{j_1,j_2,\ldots,j_M}. \tag{1}$$

  In other words, the tensor $\boldsymbol{\mathcal{C}}$ contains all possible combinations of pairwise products between the elements of $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$. This operator is very closely related to the Kronecker product defined for matrices.
- **The scalar product** of two tensors $\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$ is denoted by $\langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \rangle$ and computed by summing the element-wise product of $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}^*$ over all indices, i.e.,

$$c = \langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \rangle \in \mathbb{C}$$
$$c = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} b^*_{i_1,i_2,\ldots,i_N} \cdot a_{i_1,i_2,\ldots,i_N}. \tag{2}$$

  The scalar product allows us to define the higher-order norm of a tensor $\boldsymbol{\mathcal{A}}$ as $\|\boldsymbol{\mathcal{A}}\|_{\mathrm{H}} \doteq \sqrt{\langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{A}} \rangle}$.
- **The n-mode product** of a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$ and a matrix $\boldsymbol{U} \in \mathbb{C}^{J_n \times I_n}$ along the nth mode is denoted as

$$\boldsymbol{\mathcal{B}} = \boldsymbol{\mathcal{A}} \times_n \boldsymbol{U} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_{n-1} \times J_n \times I_{n+1} \times \cdots \times I_N}$$
$$b_{i_1,i_2,\ldots,i_{n-1},j_n,i_{n+1},\ldots,i_N} = \sum_{i_n=1}^{I_n} a_{i_1,i_2,\ldots,i_N} \cdot u_{j_n,i_n}. \tag{3}$$

  It may be visualized by multiplying all n-mode vectors of $\boldsymbol{\mathcal{A}}$ from the left-hand side by the matrix $\boldsymbol{U}$.
- **The n-mode contraction product** or inner product of a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ and a tensor $\boldsymbol{\mathcal{B}} \in \mathbb{C}^{J_1 \times J_2 \times \cdots \times J_M}$ is defined if $I_n = J_n$ and $n \leq \min\{M, N\}$. In this case, it is given by

$$\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{A}} \bullet_n \boldsymbol{\mathcal{B}}, \quad \text{where}$$
$$c_{i_1,i_2,\ldots,i_{n-1},i_{n+1},\ldots,i_N,j_1,j_2,\ldots,j_{n-1},j_{n+1},\ldots,j_M} = \sum_{i_n=1}^{I_n} a_{i_1,i_2,\ldots,i_N} \cdot b_{j_1,j_2,\ldots,j_{n-1},i_n,j_{n+1},\ldots,j_M}. \tag{4}$$

  Here, $\boldsymbol{\mathcal{C}}$ is of size $I_1 \times I_2 \times \cdots \times I_{n-1} \times I_{n+1} \times \cdots \times I_N \times J_1 \times J_2 \times \cdots \times J_{n-1} \times J_{n+1} \times \cdots \times J_M$.
  Therefore, the inner product is similar to the outer product, but it additionally involves a sum over the nth index. This

operation facilitates the definition of covariance tensor estimates.

- **The higher-order SVD** (HOSVD) of a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ is given by

$$\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{S}} \times_1 \boldsymbol{U}_1 \times_2 \boldsymbol{U}_2 \cdots \times_N \boldsymbol{U}_N \tag{5}$$

where $\boldsymbol{\mathcal{S}} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ is the core tensor which satisfies the all-orthogonality conditions [5] and $\boldsymbol{U}_n \in \mathbb{C}^{I_n \times I_n}, n = 1, 2, \ldots, N$, are the unitary matrices of n-mode singular vectors.

A subtensor of the tensor $\boldsymbol{\mathcal{A}}$, denoted by $\boldsymbol{\mathcal{A}}_{i_n=k}$, is obtained by fixing the nth index to some value k. Moreover, a matrix unfolding of the tensor $\boldsymbol{\mathcal{A}}$ along the nth mode is denoted by $[\boldsymbol{\mathcal{A}}]_{(n)}$ and can be understood as a matrix containing all the n-mode vectors of the tensor $\boldsymbol{\mathcal{A}}$. The order of the columns is chosen in accordance with [5]. We also define the concatenation of two tensors along the nth mode via the operator $[\boldsymbol{\mathcal{A}} \sqcup_n \boldsymbol{\mathcal{B}}]$.

The operations we have defined so far satisfy the following properties that can easily be verified:

$$\begin{aligned} &\boldsymbol{\mathcal{A}} \times_1 \boldsymbol{X}_1 \times_2 \boldsymbol{X}_2 \\ &\quad = \boldsymbol{\mathcal{A}} \times_2 \boldsymbol{X}_2 \times_1 \boldsymbol{X}_1 \qquad (6) \\ &(\boldsymbol{\mathcal{A}} \times_1 \boldsymbol{X}_1) \times_1 \boldsymbol{Y}_1 \\ &\quad = \boldsymbol{\mathcal{A}} \times_1 (\boldsymbol{Y}_1 \cdot \boldsymbol{X}_1) \qquad (7) \\ &[\boldsymbol{\mathcal{A}} \times_1 \boldsymbol{X}_1 \times_2 \boldsymbol{X}_2 \cdots \times_R \boldsymbol{X}_R]_{(n)} \\ &\quad = \boldsymbol{X}_n \cdot [\boldsymbol{\mathcal{A}}]_{(n)} \cdot (\boldsymbol{X}_{n+1} \otimes \boldsymbol{X}_{n+2} \cdots \\ &\qquad \otimes \boldsymbol{X}_R \otimes \boldsymbol{X}_1 \cdots \otimes \boldsymbol{X}_{n-1})^T \qquad (8) \end{aligned}$$

where $\otimes$ denotes the Kronecker product. It is easily verified that for every tensor $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{I_1 \times I_2 \cdots \times I_N}$

$$\|\boldsymbol{\mathcal{A}}\|_{\mathrm{H}} = \left\|[\boldsymbol{\mathcal{A}}]_{(n)}\right\|_{\mathrm{F}}, \quad n = 1, 2, \ldots, N. \tag{9}$$

B. Data Model for Multidimensional Harmonic Retrieval

The algorithms discussed in this paper are applicable to R-dimensional harmonic retrieval problems if the data is sampled on an R-dimensional lattice. A typical example is the parameter estimation step which follows channel sounding [11] in which parameters like directions of arrival (DOAs), directions of departure (DODs), Doppler shifts, and time delays of arrival (TDOAs) of the dominant paths are estimated from a multidimensional measurement. The latter consists of sampling the channel in several dimensions, such as time, frequency, and space (at both link ends). The observations are modeled as a superposition of d undamped exponentials[1] sampled on an R-dimensional grid of size $M_1 \times M_2 \times \cdots \times M_R$ at N subsequent time instants [10]. The measurement samples are given by

$$y_{m_1,m_2,\ldots,m_R,t_n} = \sum_{i=1}^{d} s_i(t_n) \prod_{r=1}^{R} e^{j \cdot (m_r - 1) \cdot \mu_i^{(r)}} + n_{m_1,m_2,\ldots,m_R,t_n} \tag{10}$$

where $m_r = 1, 2, \ldots, M_R$, $n = 1, 2, \ldots, N$, $s_i(t_n)$ denotes the complex amplitude of the ith exponential at time instant t_n, $\mu_i^{(r)}$ symbolizes the spatial frequency of the ith exponential in the rth mode, and $n_{m_1,m_2,\ldots,m_R,t_n}$ represents the additive noise component inherent in the measurement process, which is modeled as a zero-mean random process that is uncorrelated in all dimensions. Note that (10) assumes a uniform sampling in the spatial domain. However, this assumption can be relaxed to more generic geometries as long as they feature shift invariances, e.g., [25].

In the context of array signal processing, each of the R-dimensional exponentials represents one planar wavefront and the complex amplitudes $s_i(t_n)$ are the symbols. It is our goal to estimate the d spatial frequency vectors

$$\boldsymbol{\mu}_i = \begin{bmatrix} \mu_i^{(1)} & \mu_i^{(2)} & \cdots & \mu_i^{(R)} \end{bmatrix}^T, \quad i = 1, 2, \ldots, d$$

that correspond to the d R-dimensional signals and their correct pairing since they are directly related to the physical parameters we want to estimate, e.g., the directions of arrival and the directions of departure [11].

In the classical matrix approach, (10) is transformed into a matrix-vector equation by defining an array steering matrix [10]

$$\begin{aligned} \boldsymbol{A} &= \boldsymbol{A}^{(1)} \diamond \boldsymbol{A}^{(2)} \ldots \diamond \boldsymbol{A}^{(R)} \in \mathbb{C}^{M \times d} \\ \boldsymbol{A}^{(r)} &= \left[\boldsymbol{a}^{(r)}\left(\mu_1^{(r)}\right), \quad \boldsymbol{a}^{(r)}\left(\mu_2^{(r)}\right), \quad \ldots, \quad \boldsymbol{a}^{(r)}\left(\mu_d^{(r)}\right)\right] \\ &\quad \in \mathbb{C}^{M_r \times d} \end{aligned} \tag{11}$$

$r = 1, 2, \ldots, R$, where $M = \prod_{r=1}^{R} M_r$ and the vector

$$\begin{aligned} &\boldsymbol{a}^{(r)}\left(\mu_i^{(r)}\right) \\ &\quad = \begin{bmatrix} 1 & e^{j \cdot \mu_i^{(r)}} & e^{j \cdot 2\mu_i^{(r)}} & \ldots & e^{j \cdot (M_r - 1)\mu_i^{(r)}} \end{bmatrix}^T \in \mathbb{C}^{M_r \times 1} \end{aligned}$$

denotes the array response in the rth dimension for the ith source. All the spatial dimensions are stacked into column vectors. This stacking operation allows us to write the measurement equation in matrix form

$$\boldsymbol{Y} = \boldsymbol{A} \cdot \boldsymbol{S} + \boldsymbol{N} \tag{12}$$

where $\boldsymbol{Y} \in \mathbb{C}^{M \times N}$ contains the measurements stacked in a similar fashion as in $\boldsymbol{A}$, the matrix $\boldsymbol{S} \in \mathbb{C}^{d \times N}$ is composed of the symbols $s_i(t_n)$ and the noise samples are collected in the matrix $\boldsymbol{N} \in \mathbb{C}^{M \times N}$. It is obvious that the stacking operation does not capture the structure inherent in the lattice that is used to sample the data.

We, therefore, replace the measurement matrix $\boldsymbol{Y}$ by a measurement tensor $\boldsymbol{\mathcal{Y}} \in \mathbb{C}^{M_1 \times M_2 \times \cdots \times M_R \times N}$. Its elements are given by (10). Similarly to (12), $\boldsymbol{\mathcal{Y}}$ can be modeled as

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{A}} \times_{R+1} \boldsymbol{S}^T + \boldsymbol{\mathcal{N}}. \tag{13}$$

[1]For the sake of notational simplicity, we restrict our formulas to the case of *undamped exponentials*. If, however, forward-backward averaging is not used, as in the case of R-D standard Tensor-ESPRIT (derived in Section IV), the parameters of a superposition of d R-dimensional *damped exponentials* can be estimated, and (8) may be generalized to

$$y_{m_1,m_2,\ldots,m_R,t_n} = \sum_{i=1}^{d} s_i(t_n) \prod_{r=1}^{R} e^{(m_r - 1) \cdot [\zeta_i^{(r)} + j \cdot \mu_i^{(r)}]} + n_{m_1,m_2,\ldots,m_R,t_n}$$

where $\zeta_i^{(r)} \leq 0$ denotes the damping factor of the ith signal in the rth mode.

Here the matrix $\boldsymbol{S}$ is the same as in (12), the tensor $\boldsymbol{\mathcal{N}}$ contains the noise samples as defined in (10), and $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{M_1 \times M_2 \times \cdots \times M_R \times d}$ is the array steering tensor, constructed in the following fashion:

$$\boldsymbol{\mathcal{A}} = [\boldsymbol{\mathcal{A}}_1 \quad \sqcup_{R+1} \quad \boldsymbol{\mathcal{A}}_2 \quad \ldots \quad \sqcup_{R+1} \quad \boldsymbol{\mathcal{A}}_d] \tag{14}$$

where $\sqcup_n$ represents the concatenation operation along mode n. The R-dimensional array steering tensor of the ith signal, denoted by $\boldsymbol{\mathcal{A}}_i \in \mathbb{C}^{M_1 \times M_2 \times \cdots \times M_R}$, is obtained from the array response vectors in the rth dimension $\boldsymbol{a}^{(r)}\left(\mu_i^{(r)}\right)$ through the outer product operator

$$\boldsymbol{\mathcal{A}}_i = \boldsymbol{a}^{(1)}\left(\mu_i^{(1)}\right) \circ \boldsymbol{a}^{(2)}\left(\mu_i^{(2)}\right) \circ \ldots \circ \boldsymbol{a}^{(R)}\left(\mu_i^{(R)}\right) \tag{15}$$

$i = 1, 2, \ldots, d$. Using (14) and (15) in (13), we obtain

$$\boldsymbol{\mathcal{Y}} = \sum_{i=1}^{d} \boldsymbol{a}^{(1)}\left(\mu_i^{(1)}\right) \circ \boldsymbol{a}^{(2)}\left(\mu_i^{(2)}\right) \circ \ldots \circ \boldsymbol{a}^{(R)}\left(\mu_i^{(R)}\right) \circ \boldsymbol{s}_i^T + \boldsymbol{\mathcal{N}} \tag{16}$$

where $\boldsymbol{s}_i^T$ is the ith row of $\boldsymbol{S}$. An important consequence we can draw from (16) is that in the absence of noise the tensor $\boldsymbol{\mathcal{Y}}$ has rank d since it is computed through a sum of d rank-1 terms[2] [6]. Consequently, all the n-ranks of $\boldsymbol{\mathcal{Y}}$, i.e., the ranks of all the unfoldings $[\boldsymbol{\mathcal{Y}}]_{(n)}$, are less than or equal to d.

A strong connection between the matrix and tensor data model is given by the identities

$$\begin{aligned} \boldsymbol{A} &= [\boldsymbol{\mathcal{A}}]_{(R+1)}^T \\ \boldsymbol{N} &= [\boldsymbol{\mathcal{N}}]_{(R+1)}^T \\ \boldsymbol{Y} &= [\boldsymbol{\mathcal{Y}}]_{(R+1)}^T \end{aligned} \tag{17}$$

i.e., the measurement matrix $\boldsymbol{Y}$ is equal to the transpose of the unfolding of the measurement tensor $\boldsymbol{\mathcal{Y}}$ along the last dimension.

III. HOSVD-Based Signal Subspace Estimation

A. Direct Data Approach

In this section, we generalize the signal subspace estimation concept to the tensor case. We therefore introduce the HOSVD of the measurement tensor $\boldsymbol{\mathcal{Y}} \in \mathbb{C}^{M_1 \times M_2 \cdots \times M_R \times N}$ given by

$$\begin{aligned} &\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{S}} \times_1 \boldsymbol{U}_1 \times_2 \boldsymbol{U}_2 \cdots \times_{R+1} \boldsymbol{U}_{R+1}, \quad \text{where} \\ &\boldsymbol{U}_r \in \mathbb{C}^{M_r \times M_r}, \quad r = 1, 2, \ldots, R \\ &\boldsymbol{U}_{R+1} \in \mathbb{C}^{N \times N} \end{aligned} \tag{18}$$

and the core tensor $\boldsymbol{\mathcal{S}}$ is of same size as $\boldsymbol{\mathcal{Y}}$.

[2]The rank cannot be larger than d, but it might be smaller. This occurs only in degenerate cases which are not relevant for our discussion, e.g., two coherent sources at exactly the same position. A rank-1 tensor of order $R+1$ is a tensor that consists of the outer product of $R+1$ vectors. If and only if a tensor has rank d, it can be decomposed into a sum of d, but not less than d, rank-1 terms.

As aforementioned, in the noiseless case, the tensor $\boldsymbol{\mathcal{Y}}$ is of rank d and, therefore, all the n-ranks are at most equal to d [5]. Thus, we can express the noiseless $\boldsymbol{\mathcal{Y}}$ in terms of an "economy size" HOSVD in the following fashion:

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{S}}^{[s]} \times_1 \boldsymbol{U}_1^{[s]} \times_2 \boldsymbol{U}_2^{[s]} \cdots \times_{R+1} \boldsymbol{U}_{R+1}^{[s]} \tag{19}$$

where $\boldsymbol{U}_r^{[s]} \in \mathbb{C}^{M_r \times p_r}$, $p_r = \min\{M_r, d\}$ for $r = 1, 2, \ldots, R$, $\boldsymbol{U}_{R+1}^{[s]} \in \mathbb{C}^{N \times d}$, and $\boldsymbol{\mathcal{S}}^{[s]} \in \mathbb{C}^{p_1 \times p_2 \cdots \times p_R \times d}$. Note that it is assumed that the number of wavefronts d is known. We also require that $N \geq d$. If less snapshots are available, preprocessing schemes such as forward-backward averaging or spatial smoothing have to be applied. They will be covered in Sections V and VI, respectively.

Equation (19) holds exactly if the n-ranks (i.e., the ranks of the n-mode unfoldings) of $\boldsymbol{\mathcal{Y}}$ are less or equal than $(p_1, p_2, \ldots, p_R, d)$ which is true in the absence of noise.[3] In the presence of noise, (19) represents an HOSVD-based low-rank approximation of $\boldsymbol{\mathcal{Y}}$. This low-rank approximation can be obtained by truncating the core tensor to p_r elements in the rth mode and the corresponding matrices $\boldsymbol{U}_r$ to p_r columns for $r = 1, 2, \ldots, R$. In the last dimension $r = R+1$ we need to truncate the core tensor to d elements and $\boldsymbol{U}_{R+1}$ to d columns. However, in [6] it is shown that the best rank-$(p_1, p_2, \ldots, p_R, d)$ approximation in the least squares sense can only be computed through an iterative procedure. Nevertheless, for practical cases, the parameter estimates resulting from the two low-rank approximations (i.e., the truncation of the HOSVD and the one described in [6]) are very similar. There are, however, some applications (as the estimation of damping factors in low SNR regions) for which this is not true [3].

In the following, we compare the tensor-based signal subspace estimate to its matrix-based counterpart. In the matrix case, $\boldsymbol{U}_s \in \mathbb{C}^{M \times d}$ represents a basis for the estimated signal subspace and may be determined by truncating the SVD of $\boldsymbol{Y}$ in the following fashion:

$$\boldsymbol{Y} = \boldsymbol{U} \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{V}^H \approx \boldsymbol{U}_s \cdot \boldsymbol{\Sigma}_s \cdot \boldsymbol{V}_s^H \tag{20}$$

where $\boldsymbol{\Sigma}_s \in \mathbb{C}^{d \times d}$ and $\boldsymbol{V}_s \in \mathbb{C}^{N \times d}$. As a multidimensional extension of the basis $\boldsymbol{U}_s$ we define a tensor $\boldsymbol{\mathcal{U}}^{[s]}$ in the following way:

$$\begin{aligned} \boldsymbol{\mathcal{U}}^{[s]} &= \boldsymbol{\mathcal{S}}^{[s]} \times_1 \boldsymbol{U}_1^{[s]} \times_2 \boldsymbol{U}_2^{[s]} \cdots \times_R \boldsymbol{U}_R^{[s]} \\ &\in \mathbb{C}^{M_1 \times M_2 \cdots \times d}. \end{aligned} \tag{21}$$

Note that $\boldsymbol{\mathcal{U}}^{[s]}$ can be based on any HOSVD-based low-rank approximation of $\boldsymbol{\mathcal{Y}}$, e.g., the iterative solution described in [6] or the truncated HOSVD outlined above. The relationship between the subspace spanned by the columns of $\boldsymbol{U}_s$ and the columns of $[\boldsymbol{\mathcal{U}}^{[s]}]_{(R+1)}$ is explored further in Appendix I. In particular, we identify conditions for which both yield exactly the same signal subspace estimate as, for instance, in the noiseless case or if $d \geq \max\{M_1, M_2, \ldots, M_R\}$.

[3]Note that without additive noise and under the assumption that $d < M_r$, the r-rank of $\boldsymbol{\mathcal{Y}}$ is less than $p_r = d$ if there are at least two common frequencies in the rth mode.

TABLE I
COMPARISON OF THE REQUIRED NUMBER OF MULTIPLICATIONS FOR SIGNAL SUBSPACE ESTIMATION IN CASE OF THE MATRIX-BASED APPROACH AND THE TENSOR-BASED APPROACH

	Matrix-based approach	**Tensor-based approach**
$d \leq M_r,\ \forall r$	$k_t \cdot M \cdot N \cdot d$	$k_t \cdot M \cdot N \cdot d \cdot (R+1)$ $+M \cdot N \cdot d \cdot R + M \cdot d^2 \cdot R$
$d > M_r,\ \forall r$	$k_t \cdot M \cdot N \cdot d$	$k_t \cdot M \cdot N \cdot (d+R)$ $+M \cdot N \cdot \left(d + \sum_{r=1}^{R} M_r\right) + M \cdot d \cdot \sum_{r=1}^{R} M_r$

In the matrix case, it can be shown that in the absence of noise or with an infinite number of measurements,[4] the array steering matrix $\boldsymbol{A}$ and the matrix $\boldsymbol{U}_s$ span the same column space. Similarly, the r-spaces (i.e., the vector space spanned by the r-mode vectors) of $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{U}}^{[s]}$ are equal for $r = 1, 2, \ldots, R$. This implies that

$$\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{U}}^{[s]} \times_{R+1} \boldsymbol{T} \tag{22}$$

for some nonsingular transform matrix $\boldsymbol{T} \in \mathbb{C}^{d \times d}$. In the presence of noise and with a finite number of measurements N, (22) holds only approximately.

Computational Complexity: In order to compare the required number of multiplications for the computation of the signal subspaces, we have to know the complexity of the SVD algorithm. There is a large variety of methods to compute the SVD and their complexities differ. An efficient solution is given by the method of Orthogonal Iterations [8, pp. 410–411] which for an $M \times N$ matrix truncated to rank r has a complexity in terms of the required number of multiplications of $k_t \cdot M \cdot N \cdot r$, where k_t is a constant that depends on the design of the algorithm. To compute $\boldsymbol{U}_s$ in the matrix approach, a single SVD of the measurement matrix $\boldsymbol{Y} \in \mathbb{C}^{M \times N}$ truncated to rank d has to be computed. In the tensor case, we need the HOSVD of the measurement tensor which is equivalent to SVDs of all its unfoldings. Additional multiplications are required to compute the core tensor as well as the tensor $\boldsymbol{\mathcal{U}}^{[s]}$ from the HOSVD. The total number of required multiplications is compared in Table I. It shows that the computational complexity is higher in the tensor approach than in the matrix approach but of the same order. Nevertheless, the performance improvement demonstrated in Section VII justifies this increase of the computational complexity.

B. Covariance Approach

The method discussed in the previous subsection operates directly on the measurement data and is therefore termed the "direct data approach." Similarly to the matrix case [26], [9], [10], we define a covariance approach operating on a covariance tensor as well. This covariance approach is particularly efficient if the number of independent snapshots N is very large (i.e., $N > M$) since then the covariance matrix/tensor represents a sufficient statistic of the measurement data with less elements. If a multidimensional stochastic process (with constant parameters) is observed during a long period of time, it requires less memory to store all the data in such a covariance tensor than to store the measurement tensor itself. This covariance tensor can be updated incrementally with every time snapshot that becomes available. Moreover, some windowing may be included in this procedure. To this end, we define a covariance tensor in the following fashion.

Definition 1: Let $\boldsymbol{\mathcal{V}} \in \mathbb{C}^{M_1 \times M_2 \cdots \times M_R}$ denote a random process that describes the measurements at one time instant. Then, the corresponding covariance tensor $\boldsymbol{\mathcal{R}} \in \mathbb{C}^{M_1 \times M_2 \cdots \times M_R \times M_1 \times M_2 \cdots \times M_R}$ is defined as

$$\boldsymbol{\mathcal{R}} = \mathrm{E}\{\boldsymbol{\mathcal{V}} \circ \boldsymbol{\mathcal{V}}^*\}. \tag{23}$$

Given a set of N subsequent time snapshots in a measurement tensor $\boldsymbol{\mathcal{Y}} \in \mathbb{C}^{M_1 \times M_2 \cdots \times M_R \times N}$, the corresponding sample covariance tensor $\hat{\boldsymbol{\mathcal{R}}} \in \mathbb{C}^{M_1 \times M_2 \times \cdots \times M_R \times M_1 \times M_2 \times \cdots \times M_R}$ can by computed through

$$\hat{\boldsymbol{\mathcal{R}}} = \frac{1}{N} \cdot \boldsymbol{\mathcal{Y}} \bullet_{R+1} \boldsymbol{\mathcal{Y}}^*. \tag{24}$$

Obviously, this rectangular window can also be replaced by other windows.

For the derivation of the covariance approach, we introduce the following definitions. As an extension to the concept of Hermitian matrices, we first define a Hermitian tensor.[5]

Definition 2: A tensor $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N \times I_1 \times I_2 \times \cdots \times I_N}$ is Hermitian, iff $\forall i_n, j_n \in [1, 2, \ldots, I_n], n = 1, 2, \ldots, N$

$$x_{i_1, i_2, \ldots, i_N, j_1, j_2, \ldots, j_N} = x^*_{j_1, j_2, \ldots, j_N, i_1, i_2, \ldots i_N}. \tag{25}$$

It is easy to show that $\boldsymbol{\mathcal{R}}$ and $\hat{\boldsymbol{\mathcal{R}}}$ are always Hermitian tensors according to this definition.

Next, we introduce a matrix representation of a tensor which differs from the n-mode unfoldings defined before, in that it rearranges the elements of the tensor in a Hermitian matrix.

Lemma 1: For every Hermitian tensor $\boldsymbol{\mathcal{X}}$ there exists a matrix unfolding $\boldsymbol{\mathcal{X}}_{(\mathrm{H})}$, such that $\boldsymbol{\mathcal{X}}_{(\mathrm{H})} = \boldsymbol{\mathcal{X}}^H_{(\mathrm{H})}$.

This matrix unfolding is not unique. The Hermitian unfolding that is used in this paper is obtained by letting the first N indices vary along the columns (the Nth index is the fastest, then index $N-1$, then index $N-2$, and so on) and the last N indices along the rows (in the same order) of $[\boldsymbol{\mathcal{X}}]_{(\mathrm{H})}$. For example, for $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{I \times J \times K \times I \times J \times K}$ the unfolding $[\boldsymbol{\mathcal{X}}]_{(\mathrm{H})}$ takes the following

[4] In these special cases, the matrix approach and the tensor approach result in the same error-free estimate of the signal subspace.

[5] Note that this is not the only possible definition of the term "Hermitian tensor."

form shown in (26) at the bottom of the page. Definition 2 and Lemma 1 facilitate the computation of a square-root factor of a Hermitian tensor in the following sense.

Lemma 2: For every Hermitian tensor $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_R \times I_1 \times I_2 \times \cdots \times I_R}$ there exists a square-root factor $\boldsymbol{\mathcal{P}} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_R \times K}$ such that

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{P}} \bullet_{R+1} \boldsymbol{\mathcal{P}}^*. \tag{27}$$

Here, K is equal to the rank of $\boldsymbol{\mathcal{X}}_{(\mathrm{H})}$.

In order to compute the square-root factor, we exploit the following identity:

$$[\boldsymbol{\mathcal{X}}]_{(\mathrm{H})} = [\boldsymbol{\mathcal{P}} \bullet_{R+1} \boldsymbol{\mathcal{P}}^*]_{(\mathrm{H})} = [\boldsymbol{\mathcal{P}}]^T_{(R+1)} \cdot [\boldsymbol{\mathcal{P}}]^*_{(R+1)} \tag{28}$$

which is verified using straightforward calculations. Equation (28) shows how the tensor factorization problem (27) can be solved by transforming it into a matrix factorization problem which has a solution since $\boldsymbol{\mathcal{X}}_{(\mathrm{H})}$ is Hermitian.

Consequently, from the "economy size" version of the SVD of $\boldsymbol{\mathcal{X}}_{(\mathrm{H})}$ given by

$$\boldsymbol{\mathcal{X}}_{(\mathrm{H})} = \underbrace{\boldsymbol{U}_{s,X}}_{M \times p} \cdot \underbrace{\boldsymbol{\Sigma}_{s,X}}_{p \times p} \cdot \boldsymbol{U}^H_{s,X} \tag{29}$$

where p is the rank of $\boldsymbol{\mathcal{X}}_{(\mathrm{H})}$, we find $\boldsymbol{\mathcal{P}}$ defined through its $(R+1)$-mode unfolding

$$[\boldsymbol{\mathcal{P}}]_{(R+1)} = \left(\boldsymbol{U}_{s,X} \cdot \boldsymbol{\Sigma}^{1/2}_{s,X}\right)^T. \tag{30}$$

The size of $\boldsymbol{\mathcal{P}}$ along its last mode (which we denoted as K in Lemma 2) is equal to p. We can also choose p to be smaller than the actual rank of $\boldsymbol{\mathcal{X}}_{(\mathrm{H})}$ which results in an additional low-rank approximation. For our tasks, $p = d$ is a reasonable choice.

Recall that every covariance tensor estimate $\hat{\boldsymbol{\mathcal{R}}}$ defined through (24) is Hermitian. This allows us to compute a signal subspace estimate from $\hat{\boldsymbol{\mathcal{R}}}$ by factorizing it into $\hat{\boldsymbol{\mathcal{R}}} = \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{R}}} \bullet_{R+1} \boldsymbol{\mathcal{P}}^*_{\boldsymbol{\mathcal{R}}}$. Then, $\boldsymbol{\mathcal{U}}^{[s]}$ can be computed by replacing $\boldsymbol{\mathcal{Y}}$ by $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{R}}}$ in the HOSVD-based signal subspace estimation technique discussed in Section III-A. In Appendix II, we show that this covariance approach is equivalent to the direct data approach covered before.

IV. Shift Invariance Equations and R-D Standard Tensor-ESPRIT

In order to apply R-D ESPRIT-type methods, the R-dimensional array must feature shift invariances in each of its modes [10]. We can therefore express R shift invariance equations in tensor notation in the following fashion:

$$\begin{aligned}
\boldsymbol{\mathcal{A}} \times_1 \boldsymbol{J}_1^{(1)} \times_{R+1} \boldsymbol{\Phi}^{(1)} &= \boldsymbol{\mathcal{A}} \times_1 \boldsymbol{J}_2^{(1)} \\
\boldsymbol{\mathcal{A}} \times_2 \boldsymbol{J}_1^{(2)} \times_{R+1} \boldsymbol{\Phi}^{(2)} &= \boldsymbol{\mathcal{A}} \times_2 \boldsymbol{J}_2^{(2)} \\
&\vdots \\
\boldsymbol{\mathcal{A}} \times_R \boldsymbol{J}_1^{(R)} \times_{R+1} \boldsymbol{\Phi}^{(R)} &= \boldsymbol{\mathcal{A}} \times_R \boldsymbol{J}_2^{(R)}
\end{aligned} \tag{31}$$

where

$$\boldsymbol{\Phi}^{(r)} = \mathrm{diag}\left\{\left[e^{j\mu_1^{(r)}}, \ldots, e^{j\mu_d^{(r)}}\right]\right\}, \quad r = 1, 2, \ldots, R.$$

Here, $\boldsymbol{J}_i^{(r)} \in \mathbb{R}^{M_r^{(\mathrm{sel})} \times M_r}, i = 1, 2$, represent the selection matrices for the rth mode, $r = 1, 2, \ldots, R$, that select $M_r^{(\mathrm{sel})}$ out of M_r elements for the first $(i = 1)$ and the second $(i = 2)$ subarray, between which the shift invariance is defined. If the array is uniformly spaced in the rth mode, $M_r^{(\mathrm{sel})} = M_r - 1$ and we can use

$$\begin{aligned}
\boldsymbol{J}_1^{(r)} &= \left[\boldsymbol{I}_{M_r - 1} \quad \boldsymbol{0}_{(M_r-1)\times 1}\right] \\
\boldsymbol{J}_2^{(r)} &= \left[\boldsymbol{0}_{(M_r-1)\times 1} \quad \boldsymbol{I}_{M_r - 1}\right]
\end{aligned}$$

in case of maximum overlap. The unknown array steering tensor is eliminated from (31) by applying (22) which yields

$$\begin{aligned}
\boldsymbol{\mathcal{U}}^{[s]} \times_1 \boldsymbol{J}_1^{(1)} \times_{R+1} \boldsymbol{\Psi}^{(1)} &\approx \boldsymbol{\mathcal{U}}^{[s]} \times_1 \boldsymbol{J}_2^{(1)} \\
\boldsymbol{\mathcal{U}}^{[s]} \times_2 \boldsymbol{J}_1^{(2)} \times_{R+1} \boldsymbol{\Psi}^{(2)} &\approx \boldsymbol{\mathcal{U}}^{[s]} \times_2 \boldsymbol{J}_2^{(2)} \\
&\vdots \\
\boldsymbol{\mathcal{U}}^{[s]} \times_R \boldsymbol{J}_1^{(R)} \times_{R+1} \boldsymbol{\Psi}^{(R)} &\approx \boldsymbol{\mathcal{U}}^{[s]} \times_R \boldsymbol{J}_2^{(R)}.
\end{aligned} \tag{32}$$

Here $\boldsymbol{\Psi}^{(r)}$ and $\boldsymbol{\Phi}^{(r)}$ are related through the eigenvalue-preserving transformation $\boldsymbol{\Psi}^{(r)} = \boldsymbol{T}^{-1} \cdot \boldsymbol{\Phi}^{(r)} \cdot \boldsymbol{T}$. Consequently, the eigenvalues of $\boldsymbol{\Psi}^{(r)}$ are estimates of $e^{j\mu_i^{(r)}}, i = 1, 2, \ldots, d$.

In order to solve the shift invariance equations, we generalize the concept of matrix least squares to the tensor case. The

$$[\boldsymbol{\mathcal{X}}]_{(\mathrm{H})} = \begin{bmatrix}
x_{1,1,1,1,1,1} & x_{1,1,1,1,1,2} & \cdots & x_{1,1,1,1,1,K} & x_{1,1,1,1,2,1} & \cdots & x_{1,1,1,I,J,K} \\
x_{1,1,2,1,1,1} & x_{1,1,2,1,1,2} & \cdots & x_{1,1,2,1,1,K} & x_{1,1,2,1,2,1} & \cdots & x_{1,1,2,I,J,K} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
x_{1,1,K,1,1,1} & x_{1,1,K,1,1,2} & \cdots & x_{1,1,K,1,1,K} & x_{1,1,K,1,2,1} & \cdots & x_{1,1,K,I,J,K} \\
x_{1,2,1,1,1,1} & x_{1,2,1,1,1,2} & \cdots & x_{1,2,1,1,1,K} & x_{1,2,1,1,2,1} & \cdots & x_{1,2,1,I,J,K} \\
x_{1,2,2,1,1,1} & x_{1,2,2,1,1,2} & \cdots & x_{1,2,2,1,1,K} & x_{1,2,2,1,2,1} & \cdots & x_{1,2,2,I,J,K} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
x_{I,J,K,1,1,1} & x_{I,J,K,1,1,2} & \cdots & x_{I,J,K,1,1,K} & x_{I,J,K,1,2,1} & \cdots & x_{I,J,K,I,J,K}
\end{bmatrix}. \tag{26}$$

solution of (32) is formulated as a tensor-valued least squares problem in the following fashion:

$$\boldsymbol{\Psi}^{(r)} = \arg\min_{\boldsymbol{\Psi}^{(r)}} \left\| \boldsymbol{\mathcal{U}}^{[s]} \times_r \boldsymbol{J}_1^{(r)} \times_{R+1} \boldsymbol{\Psi}^{(r)} - \boldsymbol{\mathcal{U}}^{[s]} \times_r \boldsymbol{J}_2^{(r)} \right\|_{\mathrm{H}} \tag{33}$$

for $r = 1, 2, \ldots, R$. This tensor least squares problem can be rewritten as a matrix least squares problem by applying (9) for $n = R+1$. The latter yields

$$\begin{aligned} \boldsymbol{\Psi}^{(r)} = \arg\min_{\boldsymbol{\Psi}^{(r)}} \Bigg\| & \boldsymbol{\Psi}^{(r)} \cdot \left[\boldsymbol{\mathcal{U}}^{[s]}\right]_{(R+1)} \cdot \left(\tilde{\boldsymbol{J}}_1^{(r)}\right)^T \\ & - \left[\boldsymbol{\mathcal{U}}^{[s]}\right]_{(R+1)} \cdot \left(\tilde{\boldsymbol{J}}_2^{(r)}\right)^T \Bigg\|_{\mathrm{F}} \end{aligned} \tag{34}$$

$$\text{where} \quad \tilde{\boldsymbol{J}}_i^{(r)} = \boldsymbol{I}_{\Gamma_1^{(r)}} \otimes \boldsymbol{J}_i^{(r)} \otimes \boldsymbol{I}_{\Gamma_2^{(r)}}, \quad i = 1, 2$$

$$\text{and} \quad \Gamma_1^{(r)} = \prod_{q=1}^{r-1} M_q, \quad \Gamma_2^{(r)} = \prod_{q=r+1}^{R} M_q. \tag{35}$$

The solution to the matrix least squares problem in (34) is given by

$$\boldsymbol{\Psi}^{(r)T} = \left(\tilde{\boldsymbol{J}}_1^{(r)} \cdot \left[\boldsymbol{\mathcal{U}}^{[s]}\right]_{(R+1)}^T \right)^+ \cdot \tilde{\boldsymbol{J}}_2^{(r)} \cdot \left[\boldsymbol{\mathcal{U}}^{[s]}\right]_{(R+1)}^T. \tag{36}$$

If we compare this solution to the least squares solution obtained by solving the shift invariance equations from the classical matrix approach

$$\boldsymbol{\Psi}^{(r)} = \left(\tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{U}_s \right)^+ \cdot \tilde{\boldsymbol{J}}_2^{(r)} \cdot \boldsymbol{U}_s \tag{37}$$

we notice strong similarities. Neglecting the transpose operation for the $\boldsymbol{\Psi}^{(r)}$ in (36) (which does not change its eigenvalues), the matrix and the tensor least squares solution are essentially identical except for the fact that the estimated basis for the signal subspace $\boldsymbol{U}_s$ is replaced by $[\boldsymbol{\mathcal{U}}^{[s]}]_{(R+1)}^T$. The latter represents an improved signal subspace estimate. This improvement results from the fact that in the tensor approach we take into account the special structure of the R-dimensional lattice while computing a low-rank approximation based on the HOSVD of the measurement tensor. This allows us to "denoise" the measurements more efficiently.

However, there is a limitation to this improvement. If $d \geq \max\{M_1, M_2, \ldots, M_R\}$, it is shown in Appendix I that $\boldsymbol{U}_s$ and $[\boldsymbol{\mathcal{U}}^{[s]}]_{(R+1)}^T$ span exactly the same subspace. Conversely, the signal subspace estimate obtained through the HOSVD is only better than in the matrix case if the number of signals d is strictly less than the number of sensors M_r in *at least one* of the modes r, $r = 1, 2, \ldots, R$. Otherwise, the tensor approach yields the same result as the matrix approach.

The final step of the R-D standard Tensor-ESPRIT algorithm is the joint eigenvalue estimation step. After computing the matrices $\boldsymbol{\Psi}^{(r)}$ via (36), their joint eigenvalues $\lambda_i^{(r)}$ need to be estimated through a joint Schur decomposition (e.g., [1], [2]) or a simultaneous diagonalization algorithm (e.g., [7]). This guarantees the correct pairing of the sources over the modes. The spatial frequencies are related to these eigenvalues through $\lambda_i^{(r)} = e^{j\mu_i^{(r)}}$.

We, therefore, conclude that the entire R-D standard ESPRIT algorithm can be formulated in terms of tensors and that the least squares solution to the tensor-valued shift invariance equations leads to a similar solution as in the classical matrix approach, now being based on an improved signal subspace estimate.

V. Forward-Backward Averaging and R-D Unitary Tensor-ESPRIT

Many high-resolution parameter estimation schemes use forward-backward averaging as a preprocessing step in order to enhance the estimation accuracy [18]. Thereby, the number of available snapshots is virtually doubled without sacrificing array aperture. If forward-backward averaging is used, the spatial covariance matrix or its square-root factor can efficiently be transformed into a real-valued matrix, which significantly reduces the computational complexity of the subsequent signal subspace estimation step [18], [9]. If such a transformation is used for Unitary ESPRIT, real-valued computations can be maintained for all steps of the algorithm [9], [10].

In this section, the concepts of forward-backward averaging and real-valued subspace estimation are extended to the tensor case. Then, R-D Unitary Tensor-ESPRIT is derived.

Definition 3: The forward-backward averaged version of the measurement tensor $\boldsymbol{\mathcal{Y}}$ is computed through

$$\boldsymbol{\mathcal{Z}} \doteq \left[\boldsymbol{\mathcal{Y}} \sqcup_{R+1} \left(\boldsymbol{\mathcal{Y}}^* \times_1 \boldsymbol{\Pi}_{M_1} \times_2 \boldsymbol{\Pi}_{M_2} \cdots \times_{R+1} \boldsymbol{\Pi}_N\right)\right]. \tag{38}$$

Here $[\boldsymbol{\mathcal{A}} \ \sqcup_n \ \boldsymbol{\mathcal{B}}]$ denotes the concatenation of $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$ along the nth mode and $\boldsymbol{\Pi}_p$ is the $p \times p$ exchange matrix having ones on its antidiagonal and zeros elsewhere.

As an analogy to centro-Hermitian matrices [17], [12], we also define centro-Hermitian tensors.

Definition 4: A tensor $\boldsymbol{\mathcal{C}} \in \mathbb{C}^{I_1 \times I_2 \cdots \times I_N}$ is called centro-Hermitian if it satisfies

$$\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{C}}^* \times_1 \boldsymbol{\Pi}_{I_1} \times_2 \boldsymbol{\Pi}_{I_2} \cdots \times_N \boldsymbol{\Pi}_{I_N}. \tag{39}$$

Lemma 3: For each measurement tensor $\boldsymbol{\mathcal{Y}}$, the forward-backward averaged tensor $\boldsymbol{\mathcal{Z}}$, according to (36), is centro-Hermitian.

Lemma 4: The set of centro-Hermitian tensors $\boldsymbol{\mathcal{C}} \in \mathbb{C}^{I_1 \times I_2 \cdots \times I_N}$ can be mapped onto the set of real-valued tensors using the transformation

$$\varphi(\boldsymbol{\mathcal{C}}) = \boldsymbol{\mathcal{C}} \times_1 \boldsymbol{Q}_{I_1}^H \times_2 \boldsymbol{Q}_{I_2}^H \cdots \times_N \boldsymbol{Q}_{I_N}^H \tag{40}$$

where $\boldsymbol{Q}_p \in \mathbb{C}^{p \times p}$ are left-$\boldsymbol{\Pi}$-real matrices, i.e., they satisfy $\boldsymbol{\Pi}_p \boldsymbol{Q}_p^* = \boldsymbol{Q}_p$, and they are also assumed to be unitary.

Proof: First assume that $\boldsymbol{\mathcal{C}} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ is centro-Hermitian, i.e., it satisfies (40). Therefore

$$\begin{aligned} \varphi(\boldsymbol{\mathcal{C}})^* &= \boldsymbol{\mathcal{C}}^* \times_1 \boldsymbol{Q}_{I_1}^T \times_2 \boldsymbol{Q}_{I_2}^T \cdots \times_N \boldsymbol{Q}_{I_N}^T \\ &= (\boldsymbol{\mathcal{C}} \times_1 \boldsymbol{\Pi}_{I_1} \times_2 \boldsymbol{\Pi}_{I_2} \cdots \times_N \boldsymbol{\Pi}_{I_N}) \\ &\quad \times_1 \boldsymbol{Q}_{I_1}^T \times_2 \boldsymbol{Q}_{I_2}^T \cdots \times_N \boldsymbol{Q}_{I_N}^T \\ &= \boldsymbol{\mathcal{C}} \times_1 \left(\boldsymbol{Q}_{I_1}^T \cdot \boldsymbol{\Pi}_{I_1}\right) \times_2 \left(\boldsymbol{Q}_{I_2}^T \cdot \boldsymbol{\Pi}_{I_2}\right) \ldots \\ &\quad \times_N \left(\boldsymbol{Q}_{I_N}^T \cdot \boldsymbol{\Pi}_{I_N}\right). \end{aligned} \tag{41}$$

Next use the fact that the matrices $\boldsymbol{Q}_{I_n}$, $n = 1,2,\ldots,N$, are left-$\boldsymbol{\Pi}$-real, i.e.,

$$\begin{aligned} \boldsymbol{Q}_{I_n} &= \boldsymbol{\Pi}_{I_n} \cdot \boldsymbol{Q}_{I_n}^* \\ \Rightarrow \boldsymbol{Q}_{I_n}^T &= \boldsymbol{Q}_{I_n}^H \cdot \boldsymbol{\Pi}_{I_n} \\ \Rightarrow \boldsymbol{Q}_{I_n}^T \cdot \boldsymbol{\Pi}_{I_n} &= \boldsymbol{Q}_{I_n}^H. \end{aligned} \tag{42}$$

Substituting (42) into (41) yields

$$\varphi(\boldsymbol{\mathcal{C}})^* = \boldsymbol{\mathcal{C}} \times_1 \boldsymbol{Q}_{I_1}^H \times_2 \boldsymbol{Q}_{I_2}^H \cdots \times_N \boldsymbol{Q}_{I_N}^H = \varphi(\boldsymbol{\mathcal{C}}). \tag{43}$$

Therefore, $\varphi(\boldsymbol{\mathcal{C}})$ is real-valued.

In the same manner it can be shown that the inverse mapping, i.e., multiplying by $\boldsymbol{Q}_{I_n}$ in each of the modes $n = 1,2,\ldots,N$, transforms every real-valued tensor into a centro-Hermitian tensor. □

Consequently, as in the classical matrix version of Unitary ESPRIT, we first apply forward-backward averaging and then use the real-valued transformation to obtain a real-valued measurement tensor $\varphi(\boldsymbol{\mathcal{Z}})$. Its HOSVD is then also real-valued and given by

$$\begin{aligned} \varphi(\boldsymbol{\mathcal{Z}}) &= \boldsymbol{\mathcal{S}}_Z \times_1 \boldsymbol{E}_1 \times_2 \boldsymbol{E}_2 \cdots \times_R \boldsymbol{E}_R \times_{R+1} \boldsymbol{E}_{R+1} \\ &\approx \underbrace{\boldsymbol{\mathcal{S}}_Z^{[s]} \times_1 \boldsymbol{E}_1^{[s]} \times_2 \boldsymbol{E}_2^{[s]} \cdots \times_R \boldsymbol{E}_R^{[s]}}_{\boldsymbol{\mathcal{E}}^{[s]}} \times_{R+1} \boldsymbol{E}_{R+1}^{[s]} \end{aligned} \tag{44}$$

where $\boldsymbol{\mathcal{E}}^{[s]}$ is the multidimensional extension of the real-valued basis $\boldsymbol{E}_s$ known from the matrix approach. This definition enables us to establish the real-valued shift invariance equations given by

$$\begin{aligned} \boldsymbol{\mathcal{E}}^{[s]} \times_1 \boldsymbol{K}_1^{(1)} \times_{R+1} \boldsymbol{\Upsilon}^{(1)} &\approx \boldsymbol{\mathcal{E}}^{[s]} \times_1 \boldsymbol{K}_2^{(1)} \\ \boldsymbol{\mathcal{E}}^{[s]} \times_2 \boldsymbol{K}_1^{(2)} \times_{R+1} \boldsymbol{\Upsilon}^{(2)} &\approx \boldsymbol{\mathcal{E}}^{[s]} \times_2 \boldsymbol{K}_2^{(2)} \\ &\vdots \\ \boldsymbol{\mathcal{E}}^{[s]} \times_R \boldsymbol{K}_1^{(R)} \times_{R+1} \boldsymbol{\Upsilon}^{(R)} &\approx \boldsymbol{\mathcal{E}}^{[s]} \times_R \boldsymbol{K}_2^{(R)} \end{aligned} \tag{45}$$

where

$$\begin{aligned} \boldsymbol{K}_1^{(r)} &= 2 \cdot \mathrm{Re}\left\{ \boldsymbol{Q}_{M_r^{(\mathrm{sel})}}^H \cdot \boldsymbol{J}_2^{(r)} \cdot \boldsymbol{Q}_{M_r} \right\} \\ \boldsymbol{K}_2^{(r)} &= 2 \cdot \mathrm{Im}\left\{ \boldsymbol{Q}_{M_r^{(\mathrm{sel})}}^H \cdot \boldsymbol{J}_2^{(r)} \cdot \boldsymbol{Q}_{M_r} \right\} \end{aligned}$$

and the eigenvalues of $\boldsymbol{\Upsilon}^{(r)}$ are asymptotically equal to $\tan(\mu_i^{(r)}/2)$, $i = 1,2,\ldots,d$, $r = 1,2,\ldots,R$, as shown in Appendix III.

As before, (45) represents a tensor least squares problem that is solved in a similar manner as discussed in Section IV. The solution is given by

$$\begin{aligned} \boldsymbol{\Upsilon}^{(r)T} &= \left(\tilde{\boldsymbol{K}}_1^{(r)} \cdot \left[\boldsymbol{\mathcal{E}}^{[s]} \right]_{(R+1)}^T \right)^+ \cdot \tilde{\boldsymbol{K}}_2^{(r)} \cdot \left[\boldsymbol{\mathcal{E}}^{[s]} \right]_{(R+1)}^T \\ \tilde{\boldsymbol{K}}_1^{(r)} &= \boldsymbol{I}_{\Gamma_1^{(r)}} \otimes \boldsymbol{K}_1^{(r)} \otimes \boldsymbol{I}_{\Gamma_2^{(r)}} \\ \tilde{\boldsymbol{K}}_2^{(r)} &= \boldsymbol{I}_{\Gamma_1^{(r)}} \otimes \boldsymbol{K}_2^{(r)} \otimes \boldsymbol{I}_{\Gamma_2^{(r)}} \end{aligned} \tag{46}$$

where $\Gamma_1^{(r)}$ and $\Gamma_2^{(r)}$ are defined in (35). As in the complex-valued case, this solution is similar to the matrix approach if the real-valued basis for the signal subspace $\boldsymbol{E}_s$ is replaced by $[\boldsymbol{\mathcal{E}}^{[s]}]_{(R+1)}^T$ and additionally if for the real-valued transformation in the matrix case we use left-$\boldsymbol{\Pi}$-real matrices of the form

$$\boldsymbol{Q}_M = \boldsymbol{Q}_{M_1} \otimes \boldsymbol{Q}_{M_2} \ldots \otimes \boldsymbol{Q}_{M_R}. \tag{47}$$

The final step of the R-D Unitary Tensor-ESPRIT algorithm is the simultaneous Schur decomposition (SSD) of the matrices $\boldsymbol{\Upsilon}^{(r)}$ to estimate their joint eigenvalues $\omega_i^{(r)}$ [10]. This guarantees the correct pairing of the sources over the modes. The spatial frequencies are related to these eigenvalues through $\omega_i^{(r)} = \tan(\mu_i^{(r)}/2)$.

VI. Coherent Sources

In signal subspace based algorithms (like standard ESPRIT or Unitary ESPRIT), the rank of the signal subspace has to be equal to the number of sources d. In case of standard ESPRIT, this condition is not fulfilled if two or more sources are coherent, i.e., the correlation coefficient has magnitude one (in case of Unitary ESPRIT up to two coherent sources are allowed due to the forward-backward averaging inherent in the algorithm) or if not enough snapshots N are available.[6] A solution to this problem is given by the spatial smoothing preprocessing scheme [27], which leads to a decorrelation of the sources and an increase in the number of available snapshots. The key idea is to split the array into smaller subarrays and to average the covariance matrix over all these subarrays. Consequently, array aperture is sacrificed.

Spatial smoothing is readily formulated in terms of tensors. The smoothed measurements with L_r subarrays in mode $r = 1,2,\ldots,R$ are given by the equation at the bottom of the page, where the subtensor $\boldsymbol{\mathcal{Y}}_{\ell_1,\ell_2,\ldots,\ell_R}$, for $\ell_r = 1,2,\ldots,L_r$, is of size $M_{\mathrm{Sub},1} \times M_{\mathrm{Sub},2} \cdots \times M_{\mathrm{Sub},R} \times N \cdot L$ with $M_{\mathrm{Sub},r} = M_r - L_r + 1$ and $L = \prod_{r=1}^R L_r$. It can be computed through

$$\boldsymbol{\mathcal{Y}}_{\ell_1,\ell_2,\ldots,\ell_R} = \boldsymbol{\mathcal{Y}} \times_1 \boldsymbol{J}_{\ell_1}^{(L_1)} \times_2 \boldsymbol{J}_{\ell_2}^{(L_2)} \cdots \times_R \boldsymbol{J}_{\ell_R}^{(L_R)}$$

where the selection matrix in the rth dimension is defined as

$$\boldsymbol{J}_{\ell_r}^{(L_r)} = \left[\boldsymbol{0}_{M_{\mathrm{Sub},r} \times (\ell_r - 1)} \quad \boldsymbol{I}_{M_{\mathrm{Sub},r}} \quad \boldsymbol{0}_{M_{\mathrm{Sub},r} \times (L_r - \ell_r)} \right].$$

The smoothed tensor $\boldsymbol{\mathcal{Y}}_{ss}$ replaces the measurement tensor $\boldsymbol{\mathcal{Y}}$ in all subsequent steps of the corresponding algorithms.

An overview of the conditions when the different versions of R-D standard Tensor-ESPRIT and R-D Unitary Tensor-ESPRIT should be used is given in Table II. Here it is assumed that

[6] $N \geq d$ is required without forward-backward averaging, and $N \geq (d/2)$ is required with forward-backward averaging, e.g., in case of Unitary ESPRIT.

$$\begin{aligned} \boldsymbol{\mathcal{Y}}_{ss} \doteq [& \boldsymbol{\mathcal{Y}}_{1,1,\ldots,1} \quad \sqcup_{R+1} \quad \boldsymbol{\mathcal{Y}}_{2,1,\ldots,1} \quad \ldots \quad \sqcup_{R+1} \quad \boldsymbol{\mathcal{Y}}_{L_1,1,\ldots,1} \quad \sqcup_{R+1} \\ & \boldsymbol{\mathcal{Y}}_{1,2,\ldots,1} \quad \sqcup_{R+1} \quad \boldsymbol{\mathcal{Y}}_{2,2,\ldots,1} \quad \ldots \quad \sqcup_{R+1} \quad \boldsymbol{\mathcal{Y}}_{L_1,2,\ldots,1} \quad \sqcup_{R+1} \quad \ldots \\ & \boldsymbol{\mathcal{Y}}_{1,L_2,\ldots,L_R} \quad \sqcup_{R+1} \quad \boldsymbol{\mathcal{Y}}_{2,L_2,\ldots,L_R} \quad \ldots \quad \sqcup_{R+1} \quad \boldsymbol{\mathcal{Y}}_{L_1,L_2,\ldots,L_R}] \end{aligned}$$

TABLE II
OVERVIEW OF THE CONDITIONS FOR WHICH THE DIFFERENT VERSIONS OF R-D STANDARD TENSOR-ESPRIT AND R-D UNITARY TENSOR-ESPRIT SHOULD BE USED. HERE IT IS ASSUMED THAT c OF THE d IMPINGING WAVEFRONTS ARE COHERENT. IF THERE ARE NO COHERENT WAVEFRONTS, $c = 1$

Algorithm	R-D standard Tensor-ESPRIT	R-D Unitary Tensor-ESPRIT
Data model	superposition of damped exponentials sampled on an R-dimensional grid	superposition of undamped exponentials sampled on an R-dimensional grid
R-D smoothing is required if ...	$N < d$ or $c > 1$	$2N < d$ or $c > 2$
The covariance approach should be preferred to the direct data approach if ...	$M > N$	$M > 2N$

among the d impinging wavefronts, c of them are coherent. If there are no coherent wavefronts, $c = 1$.

VII. SIMULATION RESULTS

In this section, we demonstrate the improvements introduced by the tensor approach through numerical computer simulations. This evaluation is based on two criteria: First, we compare the true signal subspace (given by the span of the columns of the array steering matrix) with the estimated signal subspace obtained through different approaches. The comparison is based on the largest principal angle (LPA, cf. [8]) which measures the "distance" between two subspaces. The LPA between two matrices $\boldsymbol{U}_1, \boldsymbol{U}_2 \in \mathbb{C}^{M \times d}$ containing a basis for the signal subspace can be computed through

$$\mathrm{LPA} = \cos^{-1}(\sigma_{\min}\{\mathrm{orth}\{\boldsymbol{U}_1\}^H \cdot \mathrm{orth}\{\boldsymbol{U}_2\}\}) \qquad (48)$$

where $\mathrm{orth}\{\boldsymbol{U}_i\}, i = 1, 2$, is an orthonormal basis for the linear vector space spanned by the columns of $\boldsymbol{U}_i$ and $\sigma_{\min}\{\boldsymbol{Z}\}$ denotes the smallest singular value of the matrix $\boldsymbol{Z}$. The LPA is usually given in degrees.

The second criterion is the root mean square estimation error (RMSE) for the standard ESPRIT (SE) algorithm, the Unitary ESPRIT algorithm, and their corresponding tensor versions in various scenarios. The *total* RMSE in the spatial frequency domain is defined as

$$\mathrm{RMSE}_{\mathrm{tot}} = \sqrt{\mathrm{E}\left\{\sum_{r=1}^{R}\sum_{i=1}^{d}\left(\mu_i^{(r)} - \hat{\mu}_i^{(r)}\right)^2\right\}}$$

where $\hat{\mu}_i^{(r)}$ represents an estimate of $\mu_i^{(r)}$.

In the sequel, we assume d sources in the far field of the receiving antenna array. If not stated otherwise, the source symbols are drawn from a circular complex Gaussian distribution without any correlation, the number of temporal snapshots N is set to 10, and the SNR at the receiver equals 10 dB.

The scenario for the simulations shown in Fig. 1 consists of two uncorrelated sources emitting BPSK symbols that are located at $\boldsymbol{\mu}_1 = [1, 1]^T$ and $\boldsymbol{\mu}_2 = [-1, -1]^T$, impinging on a uniform rectangular array (URA) that consists of 8×8 elements. The SNR is varied between -5 and 20 dB. We observe an improvement in the LPA between SE and STE by a factor of two. A similar improvement can be seen by comparing Unitary ESPRIT (UE) and Unitary Tensor-ESPRIT (UTE).

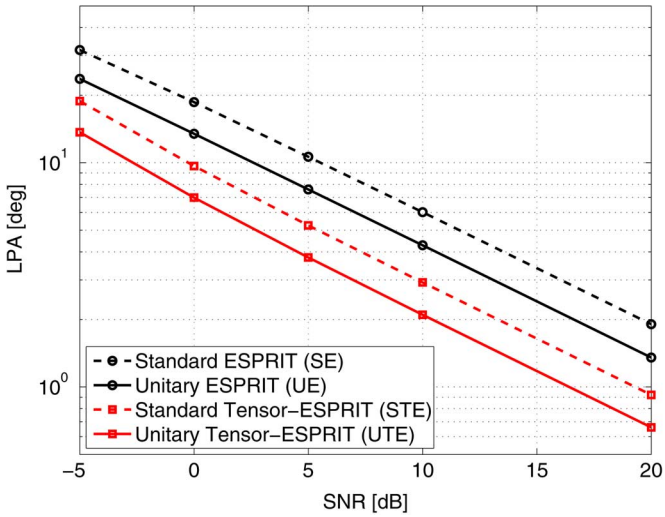


Fig. 1. LPA versus SNR for two uncorrelated sources positioned at $\boldsymbol{\mu}_1 = [1, 1]^T$ and $\boldsymbol{\mu}_2 = [-1, -1]^T$ for a URA of size 8×8 and $N = 10$ snapshots comparing SE, UE, STE, and UTE.

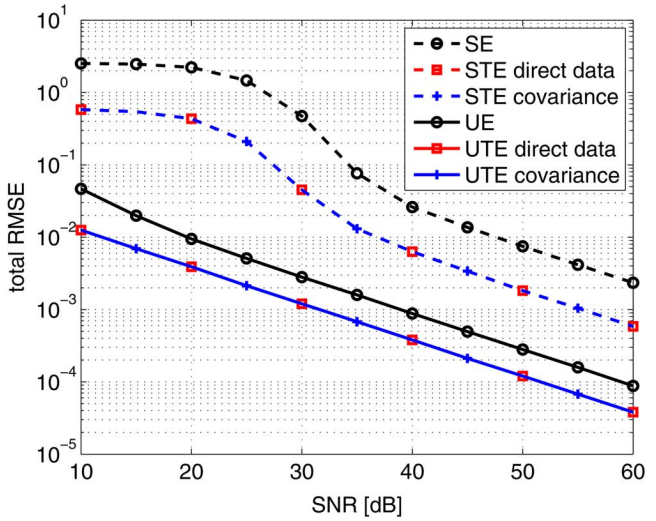


Fig. 2. Total RMSE of spatial frequencies versus the SNR for $d = 2$ highly correlated sources. Along with the direct data approach, the covariance approach is shown. The source positions are $\boldsymbol{\mu}_1 = [1, -0.5]^T$ and $\boldsymbol{\mu}_2 = [-0.5, 1]^T$, the array is a URA of size 8×8, and the number of snapshots N is set to 10.

To demonstrate the covariance approach introduced in Section III-B, we present simulation results in Fig. 2. Here, $d = 2$ sources emitting complex Gaussian distributed symbols

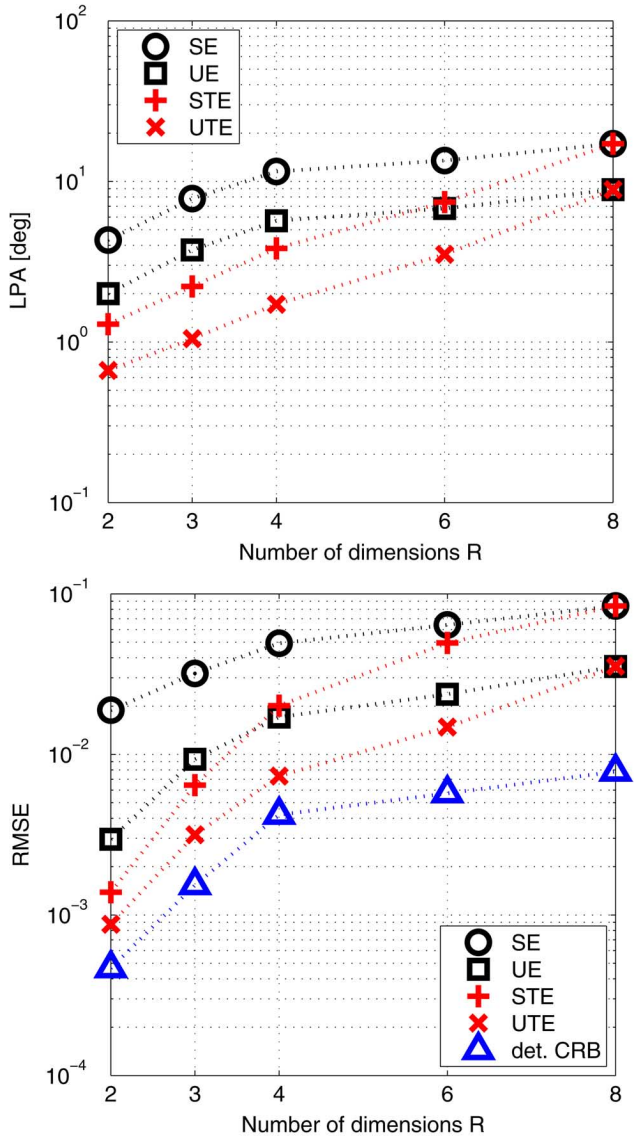


Fig. 3. Performance of the algorithms versus the number of dimensions R: The number of sensors is fixed to 256 and distributed among R dimensions with uniform spacing in all modes ($R = 2$: 16×16, $R = 3$: $8 \times 8 \times 4$, $R = 4$: $4 \times 4 \times 4 \times 4$, $R = 6$: $4 \times 4 \times 2 \times 2 \times 2 \times 2$, and $R = 8$: $2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2$). Two sources are positioned at $\mu_1^{(r)} = 1$, $\mu_2^{(r)} = 0.95$, $\forall r = 1, 2, \ldots, R$. The SNR is fixed to 40 dB, $N = 10$ snapshots are used. Top: LPA of the estimated subspaces, bott om: RMSE of the estimated spatial frequencies.

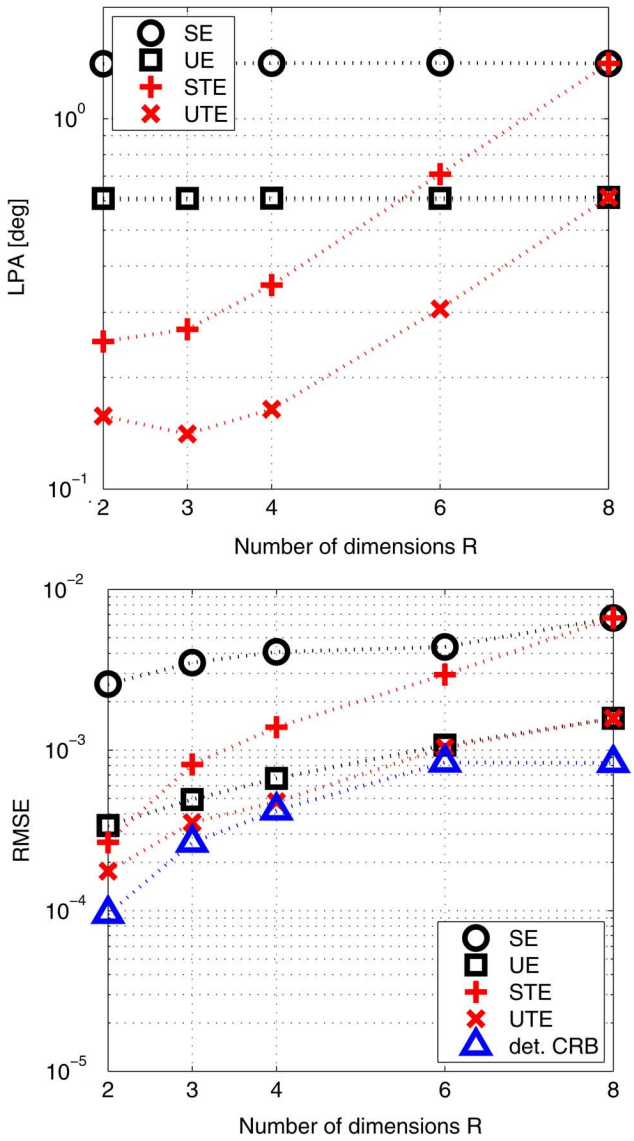


Fig. 4. Performance of the algorithms versus the number of dimensions R: The number of sensors is fixed to 256 and distributed among R dimensions with uniform spacing in all modes ($R = 2$: 16×16, $R = 3$: $8 \times 8 \times 4$, $R = 4$: $4 \times 4 \times 4 \times 4$, $R = 6$: $4 \times 4 \times 2 \times 2 \times 2 \times 2$, and $R = 8$: $2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2$). Two sources are positioned at $\mu_1^{(r)} = 1$, $\mu_2^{(r)} = -0.5$, $\forall r = 1, 2, \ldots, R$. The SNR is fixed to 40 dB, $N = 10$ snapshots are used. Top: LPA of the estimated subspaces, bottom: RMSE of the estimated spatial frequencies.

are captured by a URA of size 8×8, and the number of temporal snapshots is equal to $N = 10$. The sources are correlated with a correlation coefficient $\rho = 0.9999$, and the source positions are fixed to $\boldsymbol{\mu}_1 = [1, -0.5]^T$ and $\boldsymbol{\mu}_2 = [-0.5, 1]^T$ (the performance is very similar if the source positions are set to the same values as in Fig. 1). Since the forward-backward averaging decorrelates the sources, the RMSE for UE is significantly lower than the RMSE for SE. The corresponding tensor versions, i.e., STE and UTE, are clearly advantageous compared to standard ESPRIT and Unitary ESPRIT. In all cases, the covariance approach delivers exactly the same results as the direct data approach, which confirms the equivalence of the direct data approach and the covariance approach shown in Appendix II.

The simulations results depicted in Figs. 3 and 4 provide a comparison of the performance as a function of the number of dimensions R. In both simulations, a total of $M = 256$ sensors is distributed among a varying number of dimensions R. The point $R = 2$ corresponds to a 16×16 URA, for $R = 3$, a $8 \times 8 \times 4$ dimensional harmonic retrieval problem is simulated, and $R = 4, 6$, and 8 correspond to R-dimensional harmonic retrieval problems of size $4 \times 4 \times 4 \times 4$, $4 \times 4 \times 2 \times 2 \times 2 \times 2$, and $2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2$, respectively. The SNR is fixed to 40 dB and $N = 2$ snapshots are used. In Fig. 3 the $d = 2$ sources are relatively close to each other since there $\mu_1^{(r)} = 1$ and $\mu_2^{(r)} = 0.95$ for $r = 1, 2, \ldots, R$. On the other hand, in Fig. 4, the source spacing is increased by changing

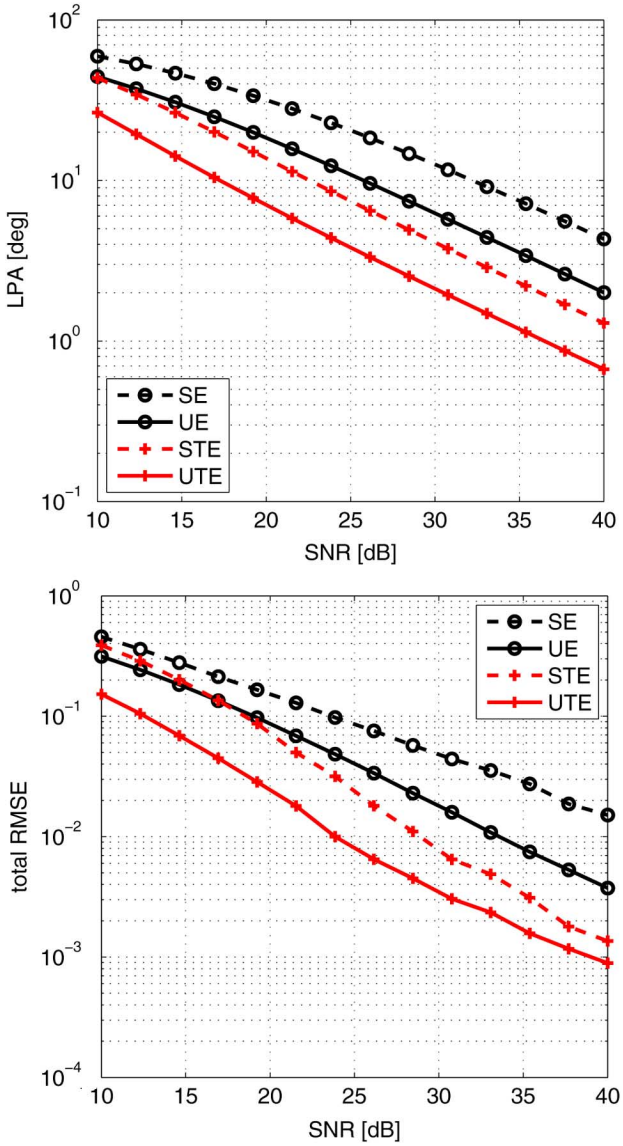


Fig. 5. Performance of the algorithms versus SNR in dB. Two sources are positioned at $\boldsymbol{\mu}_1 = [1, 1]^T$ and $\boldsymbol{\mu}_2 = [0.95, 0.95]^T$, a 16×16 URA and Top: LPA of the estimated subspaces, bottom: RMSE of the estimated spatial frequencies.

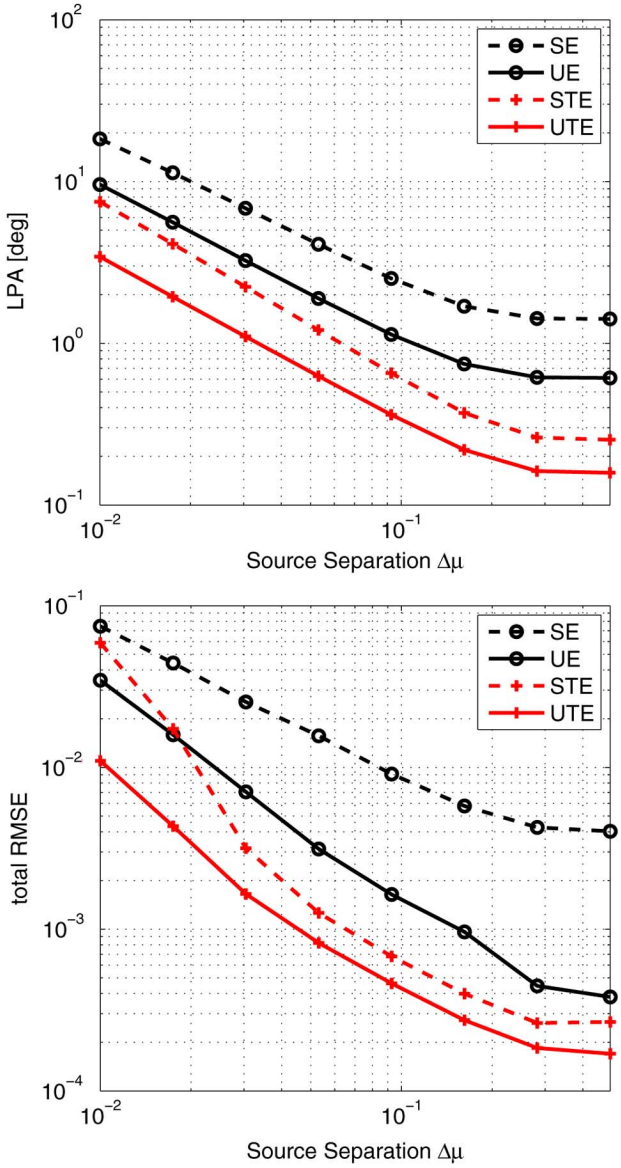


Fig. 6. Performance of the algorithms versus the separation of the two sources. The first source is fixed at $\boldsymbol{\mu}_1 = [1, 1]^T$ and the second source is moved to the positions $\boldsymbol{\mu}_2 = [1 - \Delta\mu, 1 - \Delta\mu]^T$, where $\Delta\mu$ is the separation shown on the horizontal axis. The SNR is fixed to 40 dB, a 16×16 URA, and $N = 2$ snapshots are used. Top: LPA of the estimated subspaces, bottom: RMSE of the estimated spatial frequencies.

$\mu_2^{(r)}$ to -0.5 for $r = 1, 2, \ldots, R$. Both, the LPA of the estimated subspaces as well as the RMSE of the estimated spatial frequencies are shown. The latter is defined as

$$\text{RMSE} = \sqrt{\frac{1}{R} \cdot \text{E}\left\{\sum_{r=1}^{R}\sum_{i=1}^{d}\left(\mu_i^{(r)} - \hat{\mu}_i^{(r)}\right)^2\right\}}$$

to facilitate a fair comparison for different values of R. The results show that there is a significant improvement both in terms of the LPA as well as the RMSE. We observe the largest improvement in the 2-D case ($R = 2$). Note that the improvement of the tensor approach diminishes as R increases. This gain vanishes completely if $R = 8$. In this case, we have $M_r = 2$ sensors in each of the dimensions and, therefore, the condition $d \geq \max\{M_1, M_2, \ldots, M_R\}$ in (52) is fulfilled. Consequently, the HOSVD-based signal subspace estimate is equal to the one obtained through the matrix approach as demonstrated in Appendix I. As a comparison, we have also plotted the corresponding deterministic Cramér-Rao bounds (det. CRB) [30].

In Fig. 5 we vary the SNR for closely spaced sources at $\boldsymbol{\mu}_1 = [1, 1]^T$ and $\boldsymbol{\mu}_2 = [0.95, 0.95]^T$. Again, only $N = 2$ snapshots are used, and a URA of size 16×16 is employed. As before, the improvements in terms of the LPA and the RMSE are clearly visible.

Finally, we study the effects of varying the source separation in the simulation results shown in Fig. 6. The simulation setup is similar to the previous case. However, now the SNR is fixed to 40 dB and the source positions are varied in the following fashion: The first source is fixed at $\boldsymbol{\mu}_1 = [1, 1]^T$, while the second source is moved to the positions $\boldsymbol{\mu}_2 = [1 - \Delta\mu, 1 - \Delta\mu]^T$ for various values of the separation parameter $\Delta\mu$, which is also shown

on the horizontal axis. We can see that increasing the separation above 0.5 does not alter the performance any further. The improvements of standard Tensor- ESPRIT and Unitary Tensor-ESPRIT in terms of the LPA and the RMSE are significant.

VIII. Conclusion

In this contribution a tensor approach to subspace-based high resolution parameter estimation is developed. It is shown how the HOSVD-based low-rank approximation leads to an improved signal subspace estimate which can be used for any multidimensional subspace-based parameter estimation scheme. This improvement results from the fact that in the tensor case we take the special structure of the R-dimensional lattice into account while computing a low-rank approximation based on the HOSVD of the measurement tensor. This allows us to "denoise" the measurements more efficiently. Moreover, it is demonstrated how standard ESPRIT and Unitary ESPRIT can be generalized to the tensor case, leading to the R-D standard Tensor-ESPRIT algorithm and the R-D Unitary Tensor-ESPRIT algorithm.

In addition to the square-root version operating directly on the measurement tensor, a covariance version that uses the covariance tensor is described. The forward-backward averaging and the spatial smoothing preprocessing schemes are also expressed in terms of tensors.

Significant improvements of the resulting parameter estimation accuracy can be achieved if there is at least one dimension r, $r = 1, 2, \ldots R$, where the number of sensors in this dimension M_r is larger than the number of sources d. This is particularly relevant in critical scenarios, e.g., if there is only a small number of snapshots, if there are highly correlated sources, or if some of the sources are closely spaced. Note that these improvements can already be observed in the two-dimensional case ($R = 2$).

Appendix I
Properties of the HOSVD-Based Signal Subspace Estimate

Here, we compare the HOSVD-based signal subspace estimate with its matrix-valued counterpart. In particular, we identify conditions for which both yield the same signal subspace estimate. In order to gain more insight into the significance of the tensor

$$\begin{aligned}\boldsymbol{\mathcal{U}}^{[s]} &= \boldsymbol{\mathcal{S}}^{[s]} \times_1 \boldsymbol{U}_1^{[s]} \times_2 \boldsymbol{U}_2^{[s]} \cdots \times_R \boldsymbol{U}_R^{[s]} \\ &\in \mathbb{C}^{M_1 \times M_2 \cdots \times M_R \times d}\end{aligned} \tag{49}$$

that represents the HOSVD-based signal subspace estimate and is defined in (21), consider the following cases:

- **Noiseless case**: In the noiseless case, the rank of the measurement matrix and the rank of the measurement tensor are equal to d and, therefore, (19) and (20) hold exactly. Substituting these two equations into (17) we obtain, using the definition of $\boldsymbol{\mathcal{U}}^{[s]}$ in (49) and property (8)

$$\boldsymbol{U}_s \cdot \underbrace{\boldsymbol{\Sigma}_s \cdot \boldsymbol{V}_s^H}_{\boldsymbol{T}_1} = \left[\boldsymbol{\mathcal{U}}^{[s]} \times_{R+1} \boldsymbol{U}_{R+1}^{[s]}\right]_{(R+1)}^T \tag{50}$$

$$\boldsymbol{U}_s \cdot \boldsymbol{T}_1 = \left[\boldsymbol{\mathcal{U}}^{[s]}\right]_{(R+1)}^T \cdot {\boldsymbol{U}_{R+1}^{[s]}}^T. \tag{51}$$

Notice that $[\boldsymbol{\mathcal{U}}^{[s]}]_{(R+1)}^T$ is of size $M \times d$. This leads to the conclusion that $[\boldsymbol{\mathcal{U}}^{[s]}]_{(R+1)}^T$ contains a basis for the signal subspace exactly in the same way $\boldsymbol{U}_s$ does, i.e., both matrices deliver the same estimate for the signal subspace.

- **Large number of wavefronts**: Next, consider the case where

$$d \geq \max\{M_1, M_2, \ldots, M_R\}. \tag{52}$$

Note that if there are not more than two coherent sources, the maximum number of sources which can be estimated jointly using Unitary Tensor-ESPRIT is significantly larger and is given by

$$\begin{aligned}d_{\max} &= \min_{r=1,2,\ldots,R} \left\{ M_r^{(\mathrm{sel})} \cdot \prod_{\substack{p=1 \\ p \neq r}}^{R} M_p \right\} \\ &= \min_{r=1,2,\ldots,R} \left\{ \frac{M}{M_r} \cdot M_r^{(\mathrm{sel})} \right\}\end{aligned} \tag{53}$$

if $N \geq d_{\max}/2$ snapshots are available.
In the case described in (52), the HOSVD-based low-rank approximation of $\boldsymbol{\mathcal{Y}}$ is given by

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{S}}^{[s]} \times_1 \boldsymbol{U}_1 \times_2 \boldsymbol{U}_2 \cdots \times_R \boldsymbol{U}_R \times_{R+1} \boldsymbol{U}_{R+1}^{[s]},$$

where now

$$\begin{aligned}&\boldsymbol{\mathcal{S}}^{[s]} \in \mathbb{C}^{M_1 \times M_2 \cdots \times M_R \times d} \\ &\boldsymbol{U}_r \in \mathbb{C}^{M_r \times M_r}, r = 1, 2, \ldots, R, \quad \text{and} \\ &\boldsymbol{U}_{R+1}^{[s]} \in \mathbb{C}^{N \times d}.\end{aligned} \tag{54}$$

Applying the $(R+1)$-mode unfolding and taking its transpose, we obtain

$$\boldsymbol{\mathcal{Y}}_{(R+1)}^T = (\boldsymbol{U}_1 \otimes \boldsymbol{U}_2 \ldots \otimes \boldsymbol{U}_R) \cdot \left[\boldsymbol{\mathcal{S}}^{[s]}\right]_{(R+1)}^T \cdot \left(\boldsymbol{U}_{R+1}^{[s]}\right)^T \tag{55}$$

by using (8). Let us define a matrix $\tilde{\boldsymbol{S}}_{R+1}^{[s]} \in \mathbb{C}^{d \times M}$, such that

$$\begin{aligned}&\tilde{\boldsymbol{S}}_{R+1}^{[s]} \tilde{\boldsymbol{S}}_{R+1}^{[s]H} = \boldsymbol{I}_d, \quad \text{and} \\ &[\boldsymbol{\mathcal{S}}^{[s]}]_{(R+1)} = \boldsymbol{\Sigma}_{R+1}^{[s]} \cdot \tilde{\boldsymbol{S}}_{R+1}^{[s]}\end{aligned} \tag{56}$$

where $\boldsymbol{\Sigma}_{R+1}^{[s]}$ is a $d \times d$ diagonal matrix with the d dominant $(R+1)$-mode singular values on its main diagonal. The matrix $\tilde{\boldsymbol{S}}_{R+1}^{[s]}$ is, therefore, obtained from the matrix $[\boldsymbol{\mathcal{S}}^{[s]}]_{(R+1)}$ by normalizing its rows to unit norm. Substituting (56) into (55) yields

$$\begin{aligned}\boldsymbol{\mathcal{Y}}_{(R+1)}^T &= \underbrace{(\boldsymbol{U}_1 \otimes \boldsymbol{U}_2 \ldots \otimes \boldsymbol{U}_R) \cdot \tilde{\boldsymbol{S}}_{R+1}^{[s]^T}}_{\boldsymbol{U}_{s,Y}} \cdot \boldsymbol{\Sigma}_{R+1}^{[s]} \\ &\quad \cdot \left(\boldsymbol{U}_{R+1}^{[s]}\right)^T \\ &= \boldsymbol{U}_{s,Y} \cdot \boldsymbol{\Sigma}_{R+1}^{[s]} \cdot \left(\boldsymbol{U}_{R+1}^{[s]}\right)^T.\end{aligned} \tag{57}$$

The last equation shows how $\boldsymbol{\mathcal{Y}}_{(R+1)}^T$ is decomposed into a product of an $M \times d$ unitary matrix, a $d \times d$ diagonal matrix with positive entries and a $d \times N$ unitary matrix. Consequently, it represents an SVD of $\boldsymbol{\mathcal{Y}}_{(R+1)}^T$ truncated

to d singular values. Remembering that $\boldsymbol{\mathcal{Y}}_{(R+1)}^T = \boldsymbol{Y}$ leads to the conclusion that $\boldsymbol{U}_s$ and $\boldsymbol{U}_{s,Y}$ span exactly the same subspace even under an arbitrary additive noise influence. Finally, applying the $(R+1)$-mode unfolding and the transpose operation to the definition of $\boldsymbol{\mathcal{U}}^{[s]}$ in (49) results in

$$\begin{aligned}\left[\boldsymbol{\mathcal{U}}^{[s]}\right]_{(R+1)}^T &= (\boldsymbol{U}_1 \otimes \boldsymbol{U}_2 \ldots \otimes \boldsymbol{U}_R) \cdot [\boldsymbol{\mathcal{S}}^{[s]}]_{(R+1)}^T \\ &= (\boldsymbol{U}_1 \otimes \boldsymbol{U}_2 \ldots \otimes \boldsymbol{U}_R) \cdot \tilde{\boldsymbol{S}}_{R+1}^{[s]^T} \cdot \boldsymbol{\Sigma}_{R+1}^{[s]} \\ &= \boldsymbol{U}_{s,Y} \cdot \boldsymbol{\Sigma}_{R+1}^{[s]}. \end{aligned} \tag{58}$$

Therefore, in this case $[\boldsymbol{\mathcal{U}}^{[s]}]_{(R+1)}^T$ estimated through the HOSVD of $\boldsymbol{\mathcal{Y}}$ spans exactly the same subspace as $\boldsymbol{U}_s$ estimated through the SVD of $\boldsymbol{Y}$.

- **All other cases**: In the remaining cases, i.e., in the presence of noise and if the number of signals d is strictly less than the number of sensors M_r in at least one of the modes r, $r = 1, 2, \ldots, R$, the estimated basis for the signal subspace given by $\boldsymbol{U}_s$ differs from the one given by $[\boldsymbol{\mathcal{U}}^{[s]}]_{(R+1)}^T$ if a finite number of snapshots N is available. The "quality" of this subspace estimate can be assessed, for instance, via the largest principal angle between the estimated signal subspaces and the true signal subspace given by the span of the columns of $\boldsymbol{A}$ which is demonstrated in Section VII. As a result, a significant improvement in the estimation of the signal subspace is observed. The reason for this improvement is that through the HOSVD of the tensor, the structure inherent in the R dimensional measurement data is already exploited in the subspace estimation step. Using the matrix approach, this structure cannot be captured by the measurement matrix. The tensor approach allows us to filter out noise in each of the modes of $\boldsymbol{\mathcal{U}}^{[s]}$ separately which results in an improved noise suppression. It is important to note that the improved signal subspace estimation can be used to enhance any multidimensional subspace-based parameter estimation scheme.

Appendix II
Equivalence of the Direct Data Approach and the Covariance Approach

In this Appendix, we show the equivalence of the direct data approach described in Section III-A and the covariance approach covered in Section III-B. To this end, consider a tensor $\boldsymbol{\mathcal{Y}} \in \mathbb{C}^{M_1 \times M_2 \cdots \times M_R \times N}$. Its HOSVD is given by

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{S}} \times_1 \boldsymbol{U}_1 \cdots \times_R \boldsymbol{U}_R \times_{R+1} \boldsymbol{U}_{R+1}. \tag{59}$$

Let $M = \prod_{r=1}^{R} M_r$ and $p = \operatorname{rank}\{[\boldsymbol{\mathcal{Y}}]_{(R+1)}\} \leq \min\{M, N\}$. The $(R+1)$-mode unfolding of $\boldsymbol{\mathcal{Y}}$ can be expressed in the following fashion:

$$\begin{aligned}[\boldsymbol{\mathcal{Y}}]_{(R+1)} &= \boldsymbol{U}_{R+1} \cdot [\boldsymbol{\mathcal{S}}]_{(R+1)} \cdot (\boldsymbol{U}_1 \otimes \boldsymbol{U}_2 \ldots \otimes \boldsymbol{U}_R)^T \\ &= \boldsymbol{U}_{R+1} \cdot \boldsymbol{\Sigma}_{R+1} \cdot \tilde{\boldsymbol{S}}_{R+1} \cdot (\boldsymbol{U}_1 \otimes \boldsymbol{U}_2 \ldots \otimes \boldsymbol{U}_R)^T \\ &= \boldsymbol{U}_{R+1} \cdot \boldsymbol{\Sigma}_{R+1} \cdot \boldsymbol{Q}^T \end{aligned} \tag{60}$$

where we have defined $\boldsymbol{Q} \doteq (\boldsymbol{U}_1 \otimes \boldsymbol{U}_2 \ldots \otimes \boldsymbol{U}_R) \cdot \tilde{\boldsymbol{S}}_{R+1}^T$. Moreover, the matrix $\tilde{\boldsymbol{S}}_{R+1} \in \mathbb{C}^{N \times M}$ is obtained from $[\boldsymbol{\mathcal{S}}]_{(R+1)}$ by normalizing its rows to unit norm, i.e., $[\boldsymbol{\mathcal{S}}]_{(R+1)} = \boldsymbol{\Sigma}_{R+1} \cdot \tilde{\boldsymbol{S}}_{R+1}$. Here, $\boldsymbol{\Sigma}_{R+1}$ is a diagonal matrix of size $N \times N$ that has the $(R+1)$-mode singular values $\sigma_i^{(R+1)}$, $i = 1, 2, \ldots, N$, of $\boldsymbol{\mathcal{Y}}$ on its main diagonal. From the all-orthogonality conditions of the HOSVD it is clear that $[\boldsymbol{\mathcal{S}}]_{(R+1)}$ has orthogonal rows. Hence, $\tilde{\boldsymbol{S}}_{R+1}$ has unitary rows. Let us define the following "economy-size" matrices and tensors:

$$\begin{aligned}\boldsymbol{U}_{R+1}^{[\mathrm{e}]} &= [\boldsymbol{U}_{R+1}]_{:,1:p} \in \mathbb{C}^{N \times p} \\ \boldsymbol{\Sigma}_{R+1}^{[\mathrm{e}]} &= [\boldsymbol{\Sigma}_{R+1}]_{1:p,1:p} \in \mathbb{R}^{p \times p} \\ \boldsymbol{Q}^{[\mathrm{e}]} &= [\boldsymbol{Q}]_{:,1:p} \in \mathbb{C}^{M \times p} \\ \boldsymbol{\mathcal{S}}^{[\mathrm{e}]} &= \boldsymbol{\mathcal{S}} \times_{R+1} \begin{bmatrix} \boldsymbol{I}_p & \boldsymbol{0}_{p \times (N-p)} \end{bmatrix}^T \\ &\quad \in \mathbb{C}^{M_1 \times \cdots \times M_R \times p} \\ [\boldsymbol{\mathcal{S}}^{[\mathrm{e}]}]_{(R+1)} &= \boldsymbol{\Sigma}_{R+1}^{[\mathrm{e}]} \cdot \tilde{\boldsymbol{S}}_{R+1}^{[\mathrm{e}]} \in \mathbb{C}^{p \times M}. \end{aligned}$$

Here, the notation $[\boldsymbol{X}]_{:,i:j}$ represents a matrix containing the columns i to j of the matrix $\boldsymbol{X}$. Similarly, the symbol $[\boldsymbol{X}]_{i:j,:}$ denotes a matrix containing the rows i to j of the matrix $\boldsymbol{X}$. Note that the tensor $\boldsymbol{\mathcal{S}}^{[\mathrm{e}]}$ is obtained by truncating $\boldsymbol{\mathcal{S}}$ to p elements in the $(R+1)$th mode. Since $p = \operatorname{rank}\{[\boldsymbol{\mathcal{Y}}]_{(R+1)}\}$ we conclude that (60) can also be expressed as

$$[\boldsymbol{\mathcal{Y}}]_{(R+1)} = \boldsymbol{U}_{R+1}^{[\mathrm{e}]} \cdot \boldsymbol{\Sigma}_{R+1}^{[\mathrm{e}]} \cdot \boldsymbol{Q}^{[\mathrm{e}]^T} \tag{61}$$

which represents an "economy-size" singular value decomposition, because $\boldsymbol{U}_{R+1}^{[\mathrm{e}]}$ as well as $\boldsymbol{Q}^{[\mathrm{e}]}$ have unitary columns.

According to (24), the sample covariance tensor $\hat{\boldsymbol{\mathcal{R}}}$ is defined as $\hat{\boldsymbol{\mathcal{R}}} = (1/N) \cdot \boldsymbol{\mathcal{Y}} \bullet_{R+1} \boldsymbol{\mathcal{Y}}^*$. Using property (28), its Hermitian unfolding can be expressed in terms of $\boldsymbol{\mathcal{Y}}$ in the following fashion:

$$[\hat{\boldsymbol{\mathcal{R}}}]_{(\mathrm{H})} = \frac{1}{N} \cdot \boldsymbol{\mathcal{Y}}_{(R+1)}^T \cdot \boldsymbol{\mathcal{Y}}_{(R+1)}^*. \tag{62}$$

Inserting (61) into (62) we obtain

$$\begin{aligned}[\hat{\boldsymbol{\mathcal{R}}}]_{(\mathrm{H})} &= \frac{1}{N} \cdot \boldsymbol{Q}^{[\mathrm{e}]} \cdot \boldsymbol{\Sigma}_{R+1}^{[\mathrm{e}]^T} \cdot \underbrace{\boldsymbol{U}_{R+1}^{[\mathrm{e}]^T} \cdot \boldsymbol{U}_{R+1}^{[\mathrm{e}]^*}}_{\boldsymbol{I}_p} \\ &\quad \cdot \boldsymbol{\Sigma}_{R+1}^{[\mathrm{e}]^*} \cdot \boldsymbol{Q}^{[\mathrm{e}]^H} \\ &= \frac{1}{N} \cdot \boldsymbol{Q}^{[\mathrm{e}]} \cdot \left(\boldsymbol{\Sigma}_{R+1}^{[\mathrm{e}]}\right)^2 \cdot \boldsymbol{Q}^{[\mathrm{e}]^H}. \end{aligned} \tag{63}$$

It is easy to see that (63) represents a singular value decomposition of the Hermitian matrix $\left[\hat{\boldsymbol{\mathcal{R}}}\right]_{(\mathrm{H})}$, where the ith singular value is given by $[\sigma_i^{(R+1)}]^2/N$ for $i = 1, 2, \ldots, p$ and zero for $i > p$. We conclude that a square-root factor of $\hat{\boldsymbol{\mathcal{R}}}$ is defined through

$$[\boldsymbol{\mathcal{P}}]_{(R+1)}^T \doteq \frac{1}{\sqrt{N}} \cdot \boldsymbol{Q}^{[\mathrm{e}]} \cdot \boldsymbol{\Sigma}_{R+1}^{[\mathrm{e}]}. \tag{64}$$

The corresponding tensor $\mathcal{P} \in \mathbb{C}^{M_1 \times M_2 \cdots \times M_R \times p}$ that satisfies this condition is given by

$$\begin{aligned}\mathcal{P} &= \frac{1}{\sqrt{N}} \cdot \mathcal{S}^{[\mathrm{e}]} \times_1 \boldsymbol{U}_1 \cdots \times_R \boldsymbol{U}_R \\ &= \frac{1}{\sqrt{N}} \cdot \mathcal{S} \times_1 \boldsymbol{U}_1 \cdots \times_R \boldsymbol{U}_R \; {}_{R+1} \left[\boldsymbol{I}_p \quad \boldsymbol{0}_{p\times(N-p)}\right]^T\end{aligned} \tag{65}$$

which can be verified by computing the $(R+1)$-mode unfolding of $\mathcal{P}$.

$$\begin{aligned}[\mathcal{P}]_{(R+1)} &= \frac{1}{\sqrt{N}} \cdot \boldsymbol{I}_N \cdot \left[\mathcal{S}^{[\mathrm{e}]}\right]_{(R+1)} (\boldsymbol{U}_1 \otimes \boldsymbol{U}_2 \ldots \otimes \boldsymbol{U}_R)^T \\ &= \frac{1}{\sqrt{N}} \cdot \boldsymbol{\Sigma}^{[\mathrm{e}]}_{R+1} \cdot \tilde{\boldsymbol{S}}^{[\mathrm{e}]}_{R+1} \cdot (\boldsymbol{U}_1 \otimes \boldsymbol{U}_2 \ldots \otimes \boldsymbol{U}_R)^T \\ &= \frac{1}{\sqrt{N}} \cdot \boldsymbol{\Sigma}^{[\mathrm{e}]}_{R+1} \cdot \boldsymbol{Q}^{[\mathrm{e}]^T}.\end{aligned} \tag{66}$$

Comparing (59) and (65) we conclude that an HOSVD of the measurement tensor and an HOSVD of the square-root factor of the corresponding covariance tensor agree in the core tensor as well as the first R singular vector matrices. Since the subspace tensor $\mathcal{U}^{[s]}$, defined in (21), is computed from these quantities, the corresponding subspace tensors must also span the same r-spaces, $\forall\, r = 1, 2, \ldots, R$.

Appendix III
Real-Valued Invariance Equations

In this Appendix, we derive the real-valued invariance equations (45) from their complex-valued counterparts in (32) using the real-valued HOSVD of $\varphi(\mathcal{Z})$ in (44).

In the noiseless case, we obtain from the HOSVD of $\varphi(\mathcal{Z})$ in (44)

$$\varphi(\mathcal{Z}) = \mathcal{E}^{[s]} \times_{R+1} \boldsymbol{E}^{[s]}_{R+1}. \tag{67}$$

In this case, $\varphi(\mathcal{Z})$ can be replaced using its definition in (40) which yields

$$\begin{aligned}\mathcal{Z} \times_1 \boldsymbol{Q}^H_{M_1} \times_2 \boldsymbol{Q}^H_{M_2} \cdots \times_R \boldsymbol{Q}^H_{M_R} \times_{R+1} \boldsymbol{Q}^H_{2N} \\ = \mathcal{E}^{[s]} \times_{R+1} \boldsymbol{E}^{[s]}_{R+1}.\end{aligned} \tag{68}$$

The HOSVD of $\mathcal{Z}$ can be expressed as

$$\begin{aligned}\mathcal{Z} &= \underbrace{\mathcal{S}_Z \times_1 \boldsymbol{U}^{[s]}_1 \times_2 \boldsymbol{U}^{[s]}_2 \cdots \times_R \boldsymbol{U}^{[s]}_R}_{\mathcal{U}^{[s]}} \times_{R+1} \boldsymbol{U}^{[s]}_{R+1} \\ &= \mathcal{U}^{[s]} \times_{R+1} \boldsymbol{U}^{[s]}_{R+1}\end{aligned} \tag{69}$$

where we have used the fact that $\mathcal{U}^{[s]}$ can be obtained from $\mathcal{Z}$ in a similar fashion as from $\mathcal{Y}$. Using (69) in (68) leads to a relation between the complex-valued signal subspace tensor $\mathcal{U}^{[s]}$ and its real-valued counterpart $\mathcal{E}^{[s]}$

$$\begin{aligned}&\left(\mathcal{U}^{[s]} \times_{R+1} \boldsymbol{U}^{[s]}_{R+1}\right) \times_1 \boldsymbol{Q}^H_{M_1} \times_2 \boldsymbol{Q}^H_{M_2} \cdots \\ &\quad \times_R \boldsymbol{Q}^H_{M_R} \times_{R+1} \boldsymbol{Q}^H_{2N} = \mathcal{E}^{[s]} \times_{R+1} \boldsymbol{E}^{[s]}_{R+1} \\ &\Rightarrow \mathcal{U}^{[s]} = \mathcal{E}^{[s]} \times_1 \boldsymbol{Q}_{M_1} \times_2 \boldsymbol{Q}_{M_2} \cdots \times_R \boldsymbol{Q}_{M_R} \\ &\quad \times_{R+1} \underbrace{\left(\boldsymbol{U}^{[\mathrm{s}]^H}_{R+1} \boldsymbol{Q}_{2N} \boldsymbol{E}^{[s]}_{R+1}\right)}_{\boldsymbol{T}_E}.\end{aligned} \tag{70}$$

This relationship is used to express the shift invariance equations in terms of the real-valued signal subspace tensor $\mathcal{E}^{[s]}$. In the absence of noise, the rth shift invariance equation is given by

$$\mathcal{U}^{[s]} \times_r \boldsymbol{J}^{(r)}_1 \times_{R+1} \boldsymbol{\Psi}^{(r)} = \mathcal{U}^{[s]} \times_r \boldsymbol{J}^{(r)}_2. \tag{71}$$

The tensor $\mathcal{U}^{[s]}$ is eliminated by applying (70)

$$\begin{aligned}&\left(\mathcal{E}^{[s]} \times_1 \boldsymbol{Q}_{M_1} \times_2 \boldsymbol{Q}_{M_2} \cdots \times_R \boldsymbol{Q}_{M_R} \times_{R+1} \boldsymbol{T}_E\right) \\ &\quad \times_r \boldsymbol{J}^{(r)}_1 \times_{R+1} \boldsymbol{\Psi}^{(r)} = \left(\mathcal{E}^{[s]} \times_1 \boldsymbol{Q}_{M_1} \times_2 \boldsymbol{Q}_{M_2} \cdots \right. \\ &\quad \left. \times_R \boldsymbol{Q}_{M_R} \times_{R+1} \boldsymbol{T}_E\right) \times_r \boldsymbol{J}^{(r)}_2.\end{aligned} \tag{72}$$

Equation (72) can be multiplied with $\boldsymbol{Q}^H_{M_p}$ along the modes $p = 1, 2, \ldots, r-1, r+1, \ldots, R$ to eliminate the transformations along the unused modes and with $\boldsymbol{Q}^H_{M^{(\mathrm{sel})}_r}$ along the rth mode.

$$\begin{aligned}&\mathcal{E}^{[s]} \times_r \left(\boldsymbol{Q}^H_{M^{(\mathrm{sel})}_r} \cdot \boldsymbol{J}^{(r)}_1 \cdot \boldsymbol{Q}_{M_r}\right) \times_{R+1} \left(\boldsymbol{\Psi}^{(r)} \cdot \boldsymbol{T}_E\right) \\ &\quad = \mathcal{E}^{[s]} \times_r \left(\boldsymbol{Q}^H_{M^{(\mathrm{sel})}_r} \cdot \boldsymbol{J}^{(r)}_2 \cdot \boldsymbol{Q}_{M_r}\right) \times_{R+1} \boldsymbol{T}_E \\ &\mathcal{E}^{[s]} \times_r \left(\boldsymbol{Q}^H_{M^{(\mathrm{sel})}_r} \cdot \boldsymbol{J}^{(r)}_1 \cdot \boldsymbol{Q}_{M_r}\right) \times_{R+1} \underbrace{\left(\boldsymbol{T}_E^{-1} \cdot \boldsymbol{\Psi}^{(r)} \cdot \boldsymbol{T}_E\right)}_{\tilde{\boldsymbol{\Psi}}^{(r)}} \\ &\quad = \mathcal{E}^{[s]} \times_r \left(\boldsymbol{Q}^H_{M^{(\mathrm{sel})}_r} \cdot \boldsymbol{J}^{(r)}_2 \cdot \boldsymbol{Q}_{M_r}\right).\end{aligned} \tag{73}$$

Note that $\tilde{\boldsymbol{\Psi}}^{(r)}$ has the same eigenvalues as $\boldsymbol{\Psi}^{(r)}$. Let us define transformed selection matrices $\boldsymbol{K}^{(r)}_1$ and $\boldsymbol{K}^{(r)}_2$ such that

$$\begin{aligned}\boldsymbol{K}^{(r)}_1 &= 2 \cdot \mathrm{Re}\left\{\boldsymbol{Q}^H_{M^{(\mathrm{sel})}_r} \cdot \boldsymbol{J}^{(r)}_2 \cdot \boldsymbol{Q}_{M_r}\right\} \\ \boldsymbol{K}^{(r)}_2 &= 2 \cdot \mathrm{Im}\left\{\boldsymbol{Q}^H_{M^{(\mathrm{sel})}_r} \cdot \boldsymbol{J}^{(r)}_2 \cdot \boldsymbol{Q}_{M_r}\right\}.\end{aligned} \tag{74}$$

These matrices satisfy

$$\begin{aligned}&\boldsymbol{K}^{(r)}_1 + j\boldsymbol{K}^{(r)}_2 = 2 \cdot \boldsymbol{Q}^H_{M^{(\mathrm{sel})}_r} \cdot \boldsymbol{J}^{(r)}_2 \cdot \boldsymbol{Q}_{M_r} \quad \text{and} \\ &\boldsymbol{K}^{(r)}_1 - j\boldsymbol{K}^{(r)}_2 = 2 \cdot \boldsymbol{Q}^H_{M^{(\mathrm{sel})}_r} \cdot \boldsymbol{J}^{(r)}_1 \cdot \boldsymbol{Q}_{M_r}.\end{aligned} \tag{75}$$

Using (75), (73) can be rewritten in the following fashion:

$$\begin{aligned}
&\boldsymbol{\mathcal{E}}^{[s]} \times_r \left(\boldsymbol{K}_1^{(r)} - j\boldsymbol{K}_2^{(r)}\right) \times_{R+1} \tilde{\boldsymbol{\Psi}}^{(r)} \\
&\quad = \boldsymbol{\mathcal{E}}^{[s]} \times_r \left(\boldsymbol{K}_1^{(r)} + j\boldsymbol{K}_2^{(r)}\right) \\
&\boldsymbol{\mathcal{E}}^{[s]} \times_r \boldsymbol{K}_1^{(r)} \times_{R+1} \tilde{\boldsymbol{\Psi}}^{(r)} - j \cdot \boldsymbol{\mathcal{E}}^{[s]} \times_r \boldsymbol{K}_2^{(r)} \times_{R+1} \tilde{\boldsymbol{\Psi}}^{(r)} \\
&\quad = \boldsymbol{\mathcal{E}}^{[s]} \times_r \boldsymbol{K}_1^{(r)} + j \cdot \boldsymbol{\mathcal{E}}^{[s]} \times_r \boldsymbol{K}_2^{(r)} \\
&\boldsymbol{\mathcal{E}}^{[s]} \times_r \boldsymbol{K}_1^{(r)} \times_{R+1} \left(\tilde{\boldsymbol{\Psi}}^{(r)} - \boldsymbol{I}_d\right) \\
&\quad = +j \cdot \boldsymbol{\mathcal{E}}^{[s]} \times_r \boldsymbol{K}_2^{(r)} \times_{R+1} \left(\tilde{\boldsymbol{\Psi}}^{(r)} + \boldsymbol{I}_d\right) \\
&\boldsymbol{\mathcal{E}}^{[s]} \times_r \boldsymbol{K}_1^{(r)} \times_{R+1} \underbrace{\left((-j) \cdot \left(\tilde{\boldsymbol{\Psi}}^{(r)} + \boldsymbol{I}_d\right)^{-1} \cdot \left(\tilde{\boldsymbol{\Psi}}^{(r)} - \boldsymbol{I}_d\right)\right)}_{\boldsymbol{\Upsilon}^{(r)}} \\
&\quad = \boldsymbol{\mathcal{E}}^{[s]} \times_r \boldsymbol{K}_2^{(r)} \\
&\boldsymbol{\mathcal{E}}^{[s]} \times_r \boldsymbol{K}_1^{(r)} \times_{R+1} \boldsymbol{\Upsilon}^{(r)} \\
&\quad = \boldsymbol{\mathcal{E}}^{[s]} \times_r \boldsymbol{K}_2^{(r)}. \qquad (76)
\end{aligned}$$

Similar to the matrix approach [9], it can be shown that the eigenvalues $\omega_i^{(r)}$ of $\boldsymbol{\Upsilon}^{(r)}$ are equal to $\tan(\mu_i^{(r)}/2)$, for $i = 1, 2, \ldots, d$ and $r = 1, 2, \ldots, R$, i.e., from an estimate of $\boldsymbol{\Upsilon}^{(r)}$, estimates for the spatial frequencies are readily obtained. Moreover, since $\boldsymbol{\mathcal{E}}^{[s]}$ and $\boldsymbol{K}_i^{(r)}$, $i = 1, 2$, are real-valued, $\boldsymbol{\Upsilon}^{(r)}$ must be real-valued as well.

REFERENCES

[1] K. Abed-Meraim and Y. Hua, "A least-squares approach to joint Schur decomposition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP 1998)*, May 1998, vol. IV, pp. 2541–2544.

[2] K. Abed-Meraim, Y. Hua, and M. Haardt, "Joint schur decomposition: Algorithms and applications," in *Defense Applications of Signal Processing*, D. Cochran, W. Moran, and L. B. White, Eds. New York: Elsevier, 2002.

[3] R. Boyer, L. De Lathauwer, and K. Abed-Meraim, "Delayed exponential fitting by best tensor rank-(R_1, R_2, R_3) approximation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP 2005)*, Philadelphia, PA, Mar. 2005, vol. IV, pp. 269–272.

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

[5] 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.

[6] L. de Lathauwer, B. de Moor, and J. Vanderwalle, "On the best rank-1 and rank-$(r_1, r_2, \ldots, r_n)$ approximation of higher-order tensors," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1324–1342, 2000.

[7] T. Fu and X. Gao, "Simultaneous diagonalization with similarity transformation for non-defective matrices," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP 2006)*, Toulouse, France, May 2006, vol. IV, pp. 1137–1140.

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

[9] M. Haardt and J. A. Nossek, "Unitary ESPRIT: How to obtain increased estimation accuracy with a reduced computational burden," *IEEE Trans. Signal Process.*, vol. 43, no. 5, pp. 1232–1242, May 1995.

[10] M. Haardt and J. A. Nossek, "Simultaneous Schur decomposition of several non-symmetric matrices to achieve automatic pairing in multidimensional harmonic retrieval problems," *IEEE Trans. Signal Process.*, vol. 46, no. 1, pp. 161–169, Jan. 1998.

[11] M. Haardt, R. S. Thomä, and A. Richter, "Multidimensional high-resolution parameter estimation with applications to channel sounding," in *High-Resolution and Robust Signal Processing*, Y. Hua, A. Gershman, and Q. Chen, Eds. New York: Marcel Dekker, 2004, ch. 5, pp. 255–338.

[12] R. D. Hill, R. G. Bates, and S. R. Waters, "On centrohermitian matrices," *SIAM J. Matrix Anal. Appl.*, vol. 11, pp. 128–133, Jan. 1990.

[13] T. Jiang, N. D. Sidiropoulos, and J. M. F. ten Berge, "Almost sure identifiability of multidimensional harmonic retrieval," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 1849–1859, Sep. 2002.

[14] H. A. L. Kiers, P. M. Kroonenberg, and J. M. F. ten Berge, "An efficient algorithm for TUCKALS3 on data with large numbers of observation units," *Psychometrika*, vol. 57, no. 3, pp. 415–422, 1992.

[15] P. M. Kroonenberg, *Three-Mode Principle Component Analysis: Theory and Applications*. Leiden: DSWO , 1983.

[16] 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.

[17] A. Lee, "Centrohermitian and skew-centrohermitian matrices," *Linear Algebra Its Appl.*, vol. 29, pp. 205–210, 1980.

[18] D. A. Linebarger, R. D. DeGroat, and E. M. Dowling, "Efficient direction finding methods employing forward/backward averaging," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2136–2145, Aug. 1994.

[19] X. Liu and N. D. Sidiropoulos, "Almost sure identifiability of constant modulus multidimensional harmonic retrieval," *IEEE Trans. Signal Process.*, vol. 50, no. 9, pp. 2366–2368, Sep. 2002.

[20] X. Liu, N. D. Sidiropoulos, and T. Jiang, "Multidimensional harmonic retrieval with applications in MIMO wireless channel sounding," in *Space-Time Processing for MIMO Communications*, A. Gershman and N. Sidiropoulos, Eds. New York: Wiley, 2005, ch. 2, pp. 41–75.

[21] X. Liu, N. D. Sidiropoulos, and A. Swami, "Blind high-resolution localization and tracking of multiple frequency hopped signals," *IEEE Trans. Signal Process.*, vol. 50, no. 4, pp. 889–901, Apr. 2002.

[22] K. N. Mokios, N. D. Sidiropoulos, M. Pesavento, and C. E. Mecklenbrauker, "On 3-D harmonic retrieval for wireless channel sounding," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP 2004)*, May 2004, vol. 2, pp. 89–92.

[23] M. Pesavento, C. F. Mecklenbrauker, and J. F. Böhme, "Multidimensional rank reduction estimator for parametric MIMO channel models," *EURASIP J. Appl. Signal Process.*, pp. 1354–1363, Sep. 2004.

[24] M. Rajih, "Blind identification of underdetermined mixtures based on the characteristic function," Ph.D. dissertation, Univ. Nice at Sophia Antipolis, Nice, France, Dec. 2006.

[25] F. Roemer and M. Haardt, "Using 3-D unitary ESPRIT on a hexagonal shaped ESPAR antenna for 1-D and 2-D direction of arrival estimation," presented at the Proc. ITG/IEEE Int. Workshop on Smart Antennas (WSA'05), Duisburg, Germany, Apr. 2005.

[26] R. Roy and T. Kailath, "ESPRIT—Estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-37, pp. 984–995, Jul. 1989.

[27] T. J. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. AASP-33, pp. 806–811, Aug. 1985.

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

[29] N. D. Sidiropoulos, G. B. Giannakis, and R. Bro, "Blind PARAFAC receivers for DS-CDMA systems," *IEEE Trans. Signal Process.*, vol. 48, no. 3, pp. 810–823, Mar. 2000.

[30] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-37, pp. 720–741, May 1989.

[31] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, Sep. 1966.

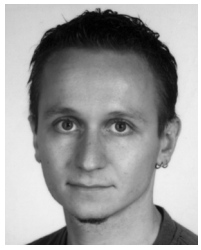
[32] H. L. Van Trees, *Optimum Array Processing: Detection, Estimation, and Modulation Theory*. New York: Wiley, 2002, pt. IV.

[33] M. D. Zoltowski, M. Haardt, and C. P. Mathews, "Closed-form 2D angle estimation with rectangular arrays in element space or beamspace via Unitary ESPRIT," *IEEE Trans. Signal Process.*, vol. 44, no. 2, pp. 316–328, Feb. 1996.

Martin Haardt (S'90–M'98–SM'99) received the Diplom-Ingenieur (M.S.) degree from the Ruhr-University Bochum, Germany, in 1991 and the Doktor-Ingenieur (Ph.D.) degree from Munich University of Technology, Munich, Germany, in 1996.

He has been a Full Professor with the Department of Electrical Engineering and Information Technology, as well as Head of the Communications Research Laboratory at Ilmenau University of Technology, Ilmenau, Germany, since 2001. In 1997, he joined Siemens Mobile Networks, Munich, where he was responsible for strategic research for third-generation mobile radio systems. From 1998 to 2001, he was the Director for International Projects and University Cooperations in the mobile infrastructure business of Siemens Mobile Networks, where his work focused on mobile communications beyond the third generation. During his time with Siemens, he also taught in the international Master of Science in Communications Engineering program at the Munich University of Technology.

Dr. Haardt received the Vodafone (formerly Mannesmann Mobilfunk) Innovations-Award for Outstanding Research in Mobile Communications, the ITG Best Paper award from the Association of Electrical Engineering, Electronics, and Information Technology (VDE), and the Rohde and Schwarz Outstanding Dissertation Award. In the fall of 2006 and the fall of 2007, he was a visiting professor at the University of Nice in Sophia-Antipolis, France, and at the University of York, U.K., respectively. His research interests include wireless communications, array signal processing, high-resolution parameter estimation, as well as numerical linear and multilinear algebra. He has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2002–2006), the IEEE SIGNAL PROCESSING LETTERS (since 2006), and the *Research Letters in Signal Processing* (since 2007). He also served as the Technical Co-Chair of the IEEE International Symposiums on Personal Indoor and Mobile Radio Communications (PIMRC) 2005 in Berlin, Germany.



Florian Roemer (S'04) studied computer engineering at Ilmenau University of Technology, Ilmenau, Germany, and McMaster University, Canada. He received the Diplom-Ingenieur (M.S.) degree with a major in communications engineering in October 2006. For his diploma thesis, he received the Siemens Communications Academic Award 2006.

Since December 2006, he has been a Research Assistant with the Communications Research Laboratory, Ilmenau University of Technology. His research interests include multidimensional signal processing, high-resolution parameter estimation, as well as multiuser MIMO precoding.

Giovanni Del Galdo was born in Italy on April 15, 1977. He received the M.Sc. degree in telecommunications engineering from Politecnico di Milano, Milan, Italy, in 2002. He received the Doktor-Ingenieur (Ph.D.) degree with highest honors in 2007 from the Ilmenau University of Technology, Ilmenau, Germany.

He then joined the Communications Research Laboratory, Ilmenau University of Technology, as a Teaching and Research Assistant. He is currently a Researcher with the Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany. His research interests include array signal processing for communications and audio applications, harmonic retrieval, antenna and propagation modelling, and multidimensional linear algebra.