ML ESTIMATION OF COVARIANCE MATRIX FOR TENSOR VALUED SIGNALS IN NOISE

Andreas Richter, Jussi Salmi, and Visa Koivunen

Helsinki University of Technology,
SMARAD CoE, Signal Processing Laboratory,
PO. Box 3000, FIN-02015 TKK, Finland

ABSTRACT

In many signal processing algorithms the estimation of signal covariance matrices is a key task. In many applications using tensor representation for the signals provides significant benefits in deriving new algorithms and revealing interesting signal properties. It is natural to model many signals in MIMO communications, physics, principal component analysis, or medical imaging using tensors. It is of high interest to develop signal processing algorithms for such problems. For some tensor-valued signals the covariance matrix may be approximated by a structured covariance with a Kronecker-product structure. This type of signals are referred to as separable. When the observed signals are contaminated by additive Gaussian noise, the separability property is lost and one ends up with shifted Kronecker-structured covariance matrices. In this paper, an iterative Maximum Likelihood (ML) estimator for covariance matrices of tensor-valued signals where covariance matrices have a shifted Kronecker-structure is proposed. The proposed algorithm is applied to wideband MIMO channel sounding measurements needed in realistic MIMO channel modeling.

Index Terms— Tensor, Stochastic Signals, Maximum Likelihood Estimation, Channel Estimation, Covariance Matrix

1. INTRODUCTION

Statistical methods to analyze vector-valued random variables are well established. Recently the analysis of tensor-valued signals has attracted a lot of interest, see e.g. the articles in [1]. But how does one characterize *tensor*-valued random variables?

A lot of research has been focused on the estimation of first order moments of tensor-valued signals. Many techniques for tensor-valued signal analysis using tensor decompositions like PARAFAC (Parallel Factor Model) [2], HOSVD (Higher Order SVD) [3], or PACA (PARAFAC/CANDECOMP) [4] has been developed [1, 5].

Yet, there are only few papers on the second order statistics of tensor valued signals. In general the second order statistics of a N-dimensional tensor-valued random variable has to be represented by a $2N$-dimensional tensor. Fortunately, the covariance matrix of tensor-valued signals may have a structure, which is related to the underlying physical problem. This leads to a constrained covariance model with significantly lower dimensionality [1] (pp. 220-236), [6].

This work is focused on the estimation of covariance matrices of separable stochastic processes [7]. For example, the model used to describe sampled wideband MIMO radio channels consists of two components. One component is a superposition of dominant specular-like propagation paths $\mathbf{S}(\boldsymbol{\theta}_p)$, having deterministic parameters $\boldsymbol{\theta}_p$. Another component is so called dense-multipath (distributed scattering), modeled as a separable stochastic process. Let M_1, M_2, M_3, denote the number of frequency samples, number of transmit antennas, and number of receive antennas. Then, a realization of the wideband MIMO radio channel is assumed to be distributed according to

$$\sim \mathcal{CN}\left(\mathrm{vec}\left\{\mathbf{S}(\boldsymbol{\theta}_\mathrm{p})\right\}, \mathbf{R}_\mathrm{d}(\boldsymbol{\theta}_\mathrm{d})\right) \in \mathbb{C}^{M_1 \times M_2 \times M_3} \tag{1}$$

where

$$\mathbf{R}_d(\boldsymbol{\theta}_d) = \mathbf{R}_f\left(\boldsymbol{\theta}_d\right) \otimes \mathbf{R}_T\left(\boldsymbol{\theta}_d\right) \otimes \mathbf{R}_R\left(\boldsymbol{\theta}_d\right),$$

and $\mathbf{R}_f\left(\boldsymbol{\theta}_d\right) \in \mathbb{C}^{M_1 \times M_1}$, $\mathbf{R}_T\left(\boldsymbol{\theta}_d\right) \in \mathbb{C}^{M_2 \times M_2}$, and $\mathbf{R}_R\left(\boldsymbol{\theta}_d\right) \in \mathbb{C}^{M_3 \times M_3}$ are covariance matrices describing the distribution of dense-multipath in frequency domain, at the transmit-antenna ports, and the receive-antenna ports. The objective is to obtain an unstructured estimate of these covariance matrices ($\hat{\mathbf{R}}_f, \hat{\mathbf{R}}_T, \hat{\mathbf{R}}_R$) or, if a model for the respective covariance matrix is available, an structured (constrained) estimate $\hat{\boldsymbol{\theta}}_d$.

The estimation of the parameters of the specular component ($\mathbf{S}(\boldsymbol{\theta}_p)$) in (1) is not considered in this paper. Appropriate algorithms as well as in depth discussion of the wideband MIMO channel model can be found in e.g. [8, 9].

Since an observation of the radio channel is *always* contaminated by additive white Gaussian (AWGN) receiver noise having variance σ^2, an observation of the wideband MIMO radio channel is distributed as

$$\sim \mathcal{CN}\left(\mathrm{vec}\left\{\mathbf{S}(\boldsymbol{\theta}_\mathrm{p})\right\}, \mathbf{R}_\mathrm{d}(\boldsymbol{\theta}_\mathrm{d}) + \sigma^2\mathbf{I}\right).$$

The AWGN can not be neglected in physical sensory measurements made in channel sounding, radio communications or in any application. In case of tensor-valued signals, it makes the estimation problem challenging since the convenient separability property is lost and the covariance matrix will have a shifted Kronecker structure. Consequently, the algorithms developed for separable signals can not be applied, except in the case of high SNR. In this paper, a Maximum Likelihood (ML) estimator of the covariance matrix with a shifted Kronecker-structure is proposed. An iterative algorithm for optimizing the likelihood function is introduced.

The paper is structured as follows. In Section 2, the data model employed in deriving the new estimator is summarized. In Section 3, a new covariance matrix estimator for the shifted Kronecker-structure case based on the ML-criterion is derived. An iterative algorithm for optimizing the likelihood function is presented. In Section 4, estimation examples from wideband MIMO channel measurements needed for channel modeling are provided. Finally, in Section 5, a summary of the paper is given.

Thanks to TEKES, and the Finish Academy agency for funding this research.

2. DATA MODEL

Suppose the observed signal is represented using a tensor structured model $\mathbf{Y} \in \mathbb{C}^{M_1 \times M_2 \times M_3}$. In this paper, it is assumed the signal is complex circular Gaussian distributed, i.e., the[1]

$$\operatorname{vec}\{\mathbf{Y}\} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{R}_3(\boldsymbol{\theta}) \otimes \mathbf{R}_2(\boldsymbol{\theta}) \otimes \mathbf{R}_1(\boldsymbol{\theta})\right). \tag{2}$$

Furthermore, it is assumed that there is also additive complex second order circular Gaussian observation noise $\mathbf{N} \in \mathbb{C}^{M_1 \times M_2 \times M_3}$ present (3), which is independent of $\mathbf{Y}$.

$$\operatorname{vec}\{\mathbf{N}\} \sim \mathcal{CN}\left(\mathbf{0}, \sigma^2 \mathbf{I}\right). \tag{3}$$

An observation $\mathbf{X} \in \mathbb{C}^{M_1 \times M_2 \times M_3}$, is a superposition of the two signals, $\mathbf{X} = \mathbf{Y} + \mathbf{N}$, and is distributed as

$$\begin{aligned} \operatorname{vec}\{\mathbf{X}\} &= \mathbf{x} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{R}(\boldsymbol{\theta})\right), \text{with} && (4)\\ \mathbf{R}(\boldsymbol{\theta}) &= \mathbf{R}_3(\boldsymbol{\theta}) \otimes \mathbf{R}_2(\boldsymbol{\theta}) \otimes \mathbf{R}_1(\boldsymbol{\theta}) + \sigma^2 \mathbf{I}. && (5) \end{aligned}$$

See also Fig. 1 for a visualization of the signal model. The focus of this paper is on the estimation of the second order statistics of this signal, i.e. the structured covariance matrix (5). An estimator of $\mathbf{R}(\boldsymbol{\theta})$ in (5), based on the ML-criteria, is derived in the following.

3. MAXIMUM LIKELIHOOD ESTIMATION

The model introduced in the previous section cannot be identified unambiguously, since

$$\begin{aligned} \mathbf{R}(\boldsymbol{\theta}) &= \mathbf{R}_3(\boldsymbol{\theta}) \otimes \mathbf{R}_2(\boldsymbol{\theta}) \otimes \mathbf{R}_1(\boldsymbol{\theta}) \\ &= (a\mathbf{R}_3(\boldsymbol{\theta})) \otimes (b\mathbf{R}_2(\boldsymbol{\theta})) \otimes (\frac{1}{ab}\mathbf{R}_1(\boldsymbol{\theta})), \forall a, b \in \mathbb{C} \\ &= \mathbf{R}'_3(\boldsymbol{\theta}) \otimes \mathbf{R}'_2(\boldsymbol{\theta}) \otimes \mathbf{R}'_1(\boldsymbol{\theta}), \end{aligned}$$

where both sets of covariance matrices $\mathbf{R}_1(\boldsymbol{\theta})$, $\mathbf{R}_2(\boldsymbol{\theta})$, $\mathbf{R}_3(\boldsymbol{\theta})$ and $\mathbf{R}'_1(\boldsymbol{\theta})$, $\mathbf{R}'_2(\boldsymbol{\theta})$, $\mathbf{R}'_3(\boldsymbol{\theta})$ are valid Kronecker-factorizations of the covariance matrix $\mathbf{R}(\boldsymbol{\theta})$. Therefore, a constraint has to be imposed on two of the three covariance matrices $\mathbf{R}_1(\boldsymbol{\theta})$, $\mathbf{R}_2(\boldsymbol{\theta})$, and $\mathbf{R}_3(\boldsymbol{\theta})$ in order to ensure identifiability. A reasonable constraint is the trace of the covariance matrices. A natural choice for the constraints is $\operatorname{trace}(\mathbf{R}_2(\boldsymbol{\theta})) = \mathrm{M}_2$ and $\operatorname{trace}(\mathbf{R}_3(\boldsymbol{\theta})) = \mathrm{M}_3$.

Altogether, the probability density function of an observation given the parameters $\boldsymbol{\theta}$ is

$$p(\mathbf{x}|\boldsymbol{\theta}) = |\pi\mathbf{R}(\boldsymbol{\theta})|^{-1} \mathrm{e}^{-\operatorname{trace}\left(\mathbf{x}^{\mathrm{H}}\mathbf{R}^{-1}(\boldsymbol{\theta})\mathbf{x}\right)}. \tag{6}$$

In the following section a MLE is derived to estimated the components of $\mathbf{R}(\boldsymbol{\theta})$.

3.1. Algorithm Outline

The log-likelihood function of (6) dropping constant terms is

$$\mathcal{L}(\mathbf{x}|\boldsymbol{\theta}) = -\log(|\mathbf{R}(\boldsymbol{\theta})|) - \operatorname{trace}\{\mathbf{x}^{\mathrm{H}}\mathbf{R}^{-1}(\boldsymbol{\theta})\mathbf{x}\}. \tag{7}$$

Note that in contrast to separable processes, $\log(|\mathbf{R}(\boldsymbol{\theta})|)$ cannot be separated into $M\sum_{i=1}^{3}(M_i^{-1}\log(|\mathbf{R}_i(\boldsymbol{\theta})|))$, with $M = \prod_{i=1}^{M} M_i$ due to the shifted Kronecker-structure in (5). A closed form solution of the MLE problem

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\mathbf{x}|\boldsymbol{\theta}) \tag{8}$$

[1]The symbol $\otimes$ denotes the Kronecker product and vec$\{\bullet\}$ the vector operator, which stacks the elements of a tensor into a vector beginning with the lowest dimension of the tensor, see e.g. [10].

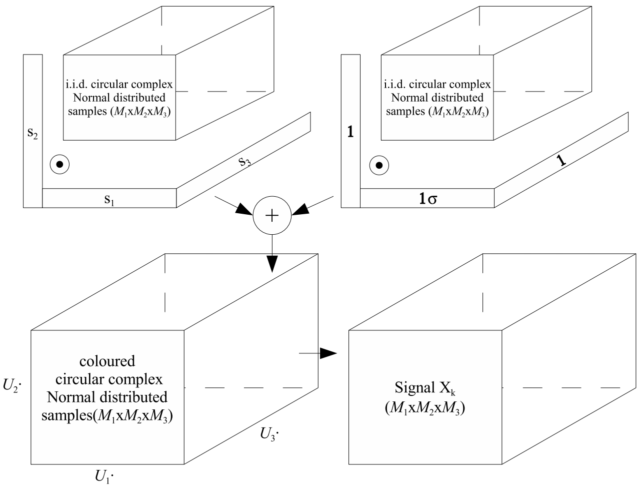

Fig. 1. Tensor Signal Model: The observed signal is a superposition of two tensor valued signals. The observed signal (lhs) is first colored in the three dimensions by the singular values (13) of the related covariance matrices $\mathbf{R}_i$. In the next step this signal is combined with the i.i.d. measurement noise (rhs) scaled by the standard deviation σ, and in the last step a correlation is introduced in the three dimensions by the eigenvectors $\mathbf{U}_i$ of the related covariance matrices (9).

is, to the best knowledge of the authors, not available. Therefore, an iterative MLE is proposed in the next section.

For notational convenience the dependencies on the parameter vector $\boldsymbol{\theta}$ is dropped for the moment. In the description of the proposed algorithm the following definitions are used

$$\begin{aligned} \hat{\mathbf{R}}_i &= \hat{\mathbf{U}}_i \hat{\mathbf{\Lambda}}_i \hat{\mathbf{U}}_i^{\mathrm{H}} && (9)\\ \hat{\mathbf{\Lambda}} &= \hat{\mathbf{\Lambda}}_3 \otimes \hat{\mathbf{\Lambda}}_2 \otimes \hat{\mathbf{\Lambda}}_1 + \hat{\sigma}^2 \mathbf{I} && (10)\\ \hat{\boldsymbol{\iota}} &= \operatorname{diag}\{\hat{\mathbf{\Lambda}}^{-1}\} && (11)\\ \hat{\boldsymbol{\lambda}}_i &= \operatorname{diag}\{\hat{\mathbf{\Lambda}}_{\mathrm{i}}\} && (12)\\ \hat{\mathbf{s}}_i &= \operatorname{diag}\{\hat{\mathbf{\Lambda}}_{\mathrm{i}}^{\frac{1}{2}}\}. && (13) \end{aligned}$$

The symbols $\odot$ and $\oslash$ denote element-wise multiplication and element-wise division, respectively. Furthermore, the operator $\operatorname{vec}\{\bullet\}$ is used to reshape a tensor into a vector, and the operator $\operatorname{mat}\{\bullet, \mathrm{M}, \mathrm{N}\}$ reshapes a vector into a matrix of dimensions $M \times N$.

3.2. Iterative Maximum Likelihood Estimation

The proposed algorithm is based on the iterative ML technique [11]. Let $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})$ be the Fisher Information matrix for the parameter-vector and $\boldsymbol{\theta}$ and $\mathbf{q}(\boldsymbol{\theta}|\mathbf{X})$ the score-function (gradient of the Log-Likelihood function) given the data $\mathbf{X}$, at point $\boldsymbol{\theta}$. Then an iterative algorithm to maximize the log-likelihood function is given by

$$\boldsymbol{\theta}^{\{k+1\}} = \boldsymbol{\theta}^{\{k\}} + \mu\Delta\boldsymbol{\theta}^{\{k\}},$$

with the Gauss-Newton-type update

$$\Delta\boldsymbol{\theta}^{\{k\}} = \boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\theta}^{\{k\}})\mathbf{q}(\boldsymbol{\theta}^{\{k\}}|\mathbf{X}). \tag{14}$$

One should note that the full Fisher Information matrix for the problem at hand may have large dimensions. In MIMO radio channel sounding applications the dimensions of the tensor are typically

$M_1 = 8 \dots 400$, $M_2 = 8 \dots 400$, and $M_3 = 100 \dots 4000$. The measurement data acquired by the authors have, e.g. dimensions of $M_1 = 16, M_2 = 16, M_3 = 193$. As an example, this parameterization leads to a Fisher information matrix of dimensions $\approx 37000 \times 37000$, since in the unconstrained case each covariance matrix $\mathbf{R}_i$ is parameterized by M_i^2 real-valued parameters. Consequently, the direct implementation of (14) is not feasible.

In order to reduce the dimensionality of the Hessian, it may be approximated by a block diagonal matrix. To this end the correlation between parameters of different covariance matrices $\mathbf{R}_i, \mathbf{R}_l, \forall i \neq l, i, l = 1, 2, 3$ as well as the observation noise variance σ^2 is neglected in the Fisher Information matrix. With this approach, the update (14) can be broken down into four separate updates, i.e., the individual updates of $\mathbf{R}_i, i = 1, 2, 3$ and an update of σ.

Due to the limited space available, the derivation of the expressions for the score-function as well as the approximation of the Hessian are omitted in this paper. In the following, only the final expressions to compute the updates $\Delta\mathbf{R}_i^{\{k\}}, i = 1, 2, 3$, and $\Delta\sigma^{\{k\}}$ are given. For notational convenience the dependency on the iteration k is dropped in the following expressions, i.e. all quantities except $\mathbf{x}$ are understood as dependent on k.

$$
\begin{aligned}
\tilde{\mathbf{x}} &= \hat{\boldsymbol{\imath}} \odot \Big((\hat{\mathbf{U}}_3 \otimes \hat{\mathbf{U}}_2 \otimes \hat{\mathbf{U}}_1)^{\mathrm{H}} \mathrm{vec}\{\mathbf{X}\} \Big) && (15)\\
\mathbf{Y}_1 &= \mathrm{mat}\{\tilde{\mathbf{x}}, \mathrm{M}_1, \mathrm{M}_2\mathrm{M}_3\} \odot \Big(\mathbf{1}\, (\hat{\mathbf{s}}_3 \otimes \hat{\mathbf{s}}_2)^{\mathrm{T}} \Big) && (16)\\
\mathbf{q}_1 &= \mathrm{mat}\{\hat{\boldsymbol{\imath}}, \mathrm{M}_1, \mathrm{M}_2\mathrm{M}_3\} \Big(\hat{\boldsymbol{\lambda}}_3 \otimes \hat{\boldsymbol{\lambda}}_2 \Big) && (17)\\
\mathbf{D}_1 &= \mathrm{mat}\{\hat{\boldsymbol{\imath}}, \mathrm{M}_1, \mathrm{M}_2\mathrm{M}_3\} \odot \Big(\mathbf{1} (\hat{\boldsymbol{\lambda}}_3 \otimes \hat{\boldsymbol{\lambda}}_2)^{\mathrm{T}} \Big) && (18)\\
\Delta\mathbf{E}_1 &= \Big(\mathbf{Y}_1 \mathbf{Y}_1^{\mathrm{H}} - \mathrm{diag}\{\mathbf{q}_1\} \Big) \oslash \Big(\mathbf{D}_1 \mathbf{D}_1^{\mathrm{T}} \Big) && (19)\\
\Delta\mathbf{R}_1 &= \hat{\mathbf{U}}_1 \Delta\mathbf{E}_1 \hat{\mathbf{U}}_1^{\mathrm{H}} && (20)
\end{aligned}
$$

The expressions for $\Delta\mathbf{R}_2, \Delta\mathbf{R}_3$ follow directly from the expressions for $\Delta\mathbf{R}_1$ by cyclic permuting the tensor dimensions. Note that cyclic permutation of the tensor dimensions requires only re-indexing the values in $\mathbf{X}$, such that Dim. 1 $\rightarrow$ Dim. 3, Dim. 3 $\rightarrow$ Dim. 2, and Dim. 2 $\rightarrow$ Dim. 1, et cetera.

One should note that the block-diagonal approximation of the Hessian used, and the chosen parameterization lead to a diagonal approximation of the Hessian matrix. This reduces the computational complexity of the algorithm significantly. Instead of solving a linear system of equations, as it is generally required by (14), an element-wise division of the gradient $(\mathbf{Y}_1\mathbf{Y}_1^{\mathrm{H}} - \mathrm{diag}\{\mathbf{q}_1\})$ and the diagonal elements of the approximated Hessian computed by $\mathbf{D}_1\mathbf{D}_1^{\mathrm{T}}$ are used. The Gauss-Newton step for $\hat{\sigma}$ is

$$
\Delta\sigma = \Big(2\hat{\sigma}\hat{\boldsymbol{\imath}}^{\mathrm{T}}\hat{\boldsymbol{\imath}} \Big)^{-1} \hat{\boldsymbol{\imath}}^{\mathrm{T}} \Big((\tilde{\mathbf{x}} \odot \tilde{\mathbf{x}}^*) \hat{\boldsymbol{\Lambda}} - 1 \Big). \qquad (21)
$$

An iteration of the proposed algorithm consists of the following steps:

1. Compute the matrices $\Delta\mathbf{R}_1^{\{k\}}$, $\Delta\mathbf{R}_2^{\{k\}}$, $\Delta\mathbf{R}_3^{\{k\}}$ (Eqs. (15-20), and $\Delta\sigma^{\{k\}}$ (Eqn. (21)).
2. Update the covariance matrices $\mathbf{R}_1^{\{k\}}$, $\mathbf{R}_2^{\{k\}}$, $\mathbf{R}_3^{\{k\}}$, and $\sigma^{\{k\}}$ using:
$$
\begin{aligned}
\tilde{\mathbf{R}}_i^{\{k+1\}} &= \mathbf{R}_i^{\{k\}} + \mu\Delta\mathbf{R}_i^{\{k\}}, \forall i, i = 1, 2, 3\\
\hat{\sigma}^{\{k+1\}} &= \hat{\sigma}^{\{k\}} + \mu\Delta\sigma^{\{k\}}\\
\hat{\mathbf{R}}_1^{\{k+1\}} &= \tilde{\mathbf{R}}_1^{\{k+1\}}\\
\hat{\mathbf{R}}_i^{\{k+1\}} &= \tilde{\mathbf{R}}_i^{\{k+1\}} M_i / \mathrm{trace}\Big(\tilde{\mathbf{R}}_i^{\{k+1\}} \Big), \forall i, i = 2, 3
\end{aligned}
$$
with $\mu \in \langle 0 \cdots 1 \rangle$. A step size of $\mu = 0.7$ has proved to be a good trade-off in simulations and for measured data. It ensures convergence and has a small impact on the rate of convergence of the algorithm.
3. Compute the eigenvalue decomposition: $\hat{\mathbf{R}}_i^{\{k+1\}} = \hat{\mathbf{U}}_i^{\{k+1\}} \hat{\boldsymbol{\Lambda}}_i^{\{k+1\}} (\hat{\mathbf{U}}_i^{\{k+1\}})^{\mathrm{H}}$
4. Check algorithm convergence $|\Delta\sigma^{\{k\}}|/|\hat{\sigma}^{\{k+1\}}| < \epsilon$, and $||\Delta\mathbf{R}_i^{\{k\}}||_F / ||\hat{\mathbf{R}}_i^{\{k+1\}}||_F < \epsilon, \forall i, i = 1, 2, 3$. The algorithm is stopped when the rate of change is below the chosen threshold ϵ.

The third step in the algorithm is computationally costly. It requires a full eigenvalue decomposition of the three updated covariance matrices in each iteration. In the next subsection this step is replaced by an QR-decomposition based update of the eigenvectors and eigenvalues of the covariance matrices.

3.3. Reduction of Computational Complexity

Considering that the derived algorithm is iterative by design, a full diagonalization of the intermediate covariance matrices is not needed, since the diagonality is partially destroyed in the next update anyway. Note that an update of $\hat{\mathbf{R}}_i^{\{k\}}$ can also be written as

$$
\begin{aligned}
\hat{\mathbf{R}}_i^{\{k+1\}} &= \hat{\mathbf{U}}_i^{\{k\}} \mathbf{E}_i^{\{k\}} (\hat{\mathbf{U}}_i^{\{k\}})^{\mathrm{H}}, \text{with} && (22)\\
\mathbf{E}_i^{\{k\}} &= \hat{\Lambda}_i^{\{k\}} + \mu\Delta\mathbf{E}_i^{\{k\}}. && (23)
\end{aligned}
$$

Now recalling the QR-decomposition based iterative eigenvalue decomposition algorithm [12], one can directly update the estimated eigenvalues and eigenvectors of the estimated covariance matrices using the QR-decomposition of $\mathbf{E}_i^{\{k\}}$. The QR-decomposition (24) factorizes $\mathbf{E}_i^{\{k\}}$ into a unitary matrix $\mathbf{Q}_i^{\{k\}}$ and an upper-triangular matrix $\mathbf{Z}_i^{\{k\}}$. The new update steps are

$$
\begin{aligned}
\mathbf{E}_i^{\{k\}} &= \mathbf{Q}_i^{\{k\}} \mathbf{Z}_i^{\{k\}} && (24)\\
\hat{\mathbf{U}}_i^{\{k+1\}} &= \hat{\mathbf{U}}_i^{\{k\}} \mathbf{Q}_i^{\{k\}}\\
\hat{\boldsymbol{\lambda}}_i^{\{k+1\}} &= \mathrm{diag}\{\mathbf{Z}_\mathrm{i}^{\{\mathrm{k}\}} (\mathbf{Q}_\mathrm{i}^{\{\mathrm{k}\}})^{\mathrm{H}}\}.
\end{aligned}
$$

The resulting algorithm does not need to form the covariance matrices explicitly. Instead, it is updating its eigenvalue decomposition.

3.4. Covariance Matrices with Additional Structure

As mentioned in Section 1 additional information about the structure of some of the covariance matrices $\mathbf{R}_i$ may be available. For wideband MIMO radio channels exists a model for the frequency correlation of dense-multipath [8, 13]. The covariance matrix of diffuse scattering in frequency domain has Toeplitz structure[2]

$$
\mathbf{R}_1(\boldsymbol{\theta}_1) = \mathrm{toep}\big(\boldsymbol{\kappa}(\boldsymbol{\theta}_1), \boldsymbol{\kappa}^{\mathrm{H}}(\boldsymbol{\theta}_1) \big), \qquad (25)
$$

where

$$
\boldsymbol{\kappa}(\boldsymbol{\theta}_1) = \frac{\alpha_1}{M_1} \left[\frac{1}{\beta_d} \quad \cdots \quad \frac{\mathrm{e}^{-\mathrm{j}2\pi(\mathrm{M}_1-1)\tau_\mathrm{d}}}{\beta_d + j2\pi\frac{M_1-1}{M_1}} \right] \qquad (26)
$$

models the correlation between samples at different frequencies. The parameter vector $\boldsymbol{\theta}_1$ contains three parameters, the normalized base

[2] The operator $\mathrm{toep}(\mathbf{c}, \mathbf{r})$ forms a Toeplitz matrix with the first column $\mathbf{c}$, and the first row $\mathbf{r}$.

delay $\tau_d \in [0,1)$, the normalized coherence bandwidth β_d, and the variance of dense multipath α_d. The ML-estimator described in [8] is applied in every iteration to estimate the parameters $\hat{\boldsymbol{\theta}}_1^{\{k+1\}}$ from the unstructured estimate $\hat{\mathbf{R}}_1^{\{k+1\}}$. Then an eigenvalue decomposition is used to estimate the updated eigenvectors and eigenvalues, i.e.

$$\hat{\mathbf{R}}_1(\boldsymbol{\theta}_1^{\{k+1\}}) = \hat{\mathbf{U}}_i^{\{k+1\}} \mathrm{diag}\{\boldsymbol{\lambda}_1^{\{\mathrm{k}+1\}}\} \left(\hat{\mathbf{U}}_{\mathrm{i}}^{\{\mathrm{k}+1\}}\right)^{\mathrm{H}}.$$

In the next section, the proposed MLE will be applied to wideband MIMO radio channel sounding measurements.

4. ESTIMATION EXAMPLE

For a description of the measurement setup as well as the measured scenario see [14]. As discussed in Section 1, a radio channel observation contains two components, contributions from dominant specular like propagation path and dense multipath. Therefore, the Extended Kalman filter based estimator described in [9] has been employed together with the derived algorithm in order to estimate both components. The Extended Kalman filter estimates $\hat{\boldsymbol{\theta}}_p$, while the proposed iterative MLE estimates $\hat{\mathbf{R}}_d(\boldsymbol{\theta}_\mathbf{d})$. Note that the estimators depend on each other. Having estimates, $\hat{\boldsymbol{\theta}}_p$ and $\hat{\boldsymbol{\theta}}_d$ one can first remove the contribution of the estimated dominant propagation paths from the observations, and then use $\hat{\mathbf{R}}_d^{\frac{1}{2}}(\boldsymbol{\theta}_\mathbf{d})$ to whiten the remaining signal. Figure 2.(a), and 2.(b) show the transmit-angle-delay-power spectrum and the receive-angle-delay-power spectrum before and after whitening.

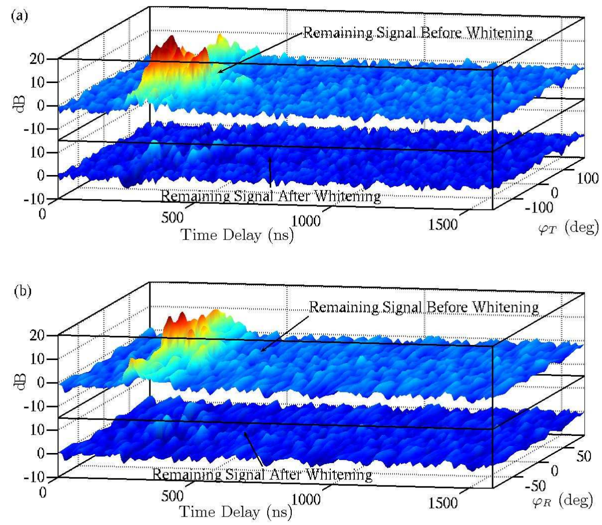

Fig. 2. Power-delay-transmit-azimuth (a) and Power-delay-receive-azimuth (b) of the remaining signal after removing estimated dominant propagation paths from a wideband MIMO radio channel measurement. The remaining signal consists of dense multipath and observation noise. In each plot the remaining signal before and after whitening with $\hat{\mathbf{R}}_d^{\frac{1}{2}}(\boldsymbol{\theta}_\mathbf{d})$ is shown.

5. CONCLUSION

In this paper a maximum likelihood estimator has been derived to estimate the covariance matrix of tensor-valued signal observations having a shifted Kronecker-structure.

The proposed algorithm is based on iterative MLE. The direct implementation of the iterative MLE algorithm is not feasible due to the large dimensionality of the Hessian involved. However, by exploiting the tensor-structure of the signal an approximation of the Hessian has been found. This leads to a significant reduction of computational complexity. The general structure of the algorithm allows to impose additional model constraints on the Kronecker-factors.

The estimator performs well for wideband MIMO radio channel sounding measurements data (The authors want to thank MEDAV GmbH and TU Ilmenau for providing the measurement data).

6. REFERENCES

[1] J. Ruiz-Alzola and C.-F. Westin, Eds., *Special Issue: Tensor Signal Processing*, vol. 87 of *Signal Processing*, ELSEVIER, February 2007, ISSN 0165-1684.

[2] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrica*, vol. 31, pp. 279–311, 1966.

[3] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, pp. 1253–1278, April 2000.

[4] J. B. Kruskal, "Three-way arrays: rank and uniqueness of trilinear decomposition, with applications to arithmetic complexity and statistics," *Ann. Statistics*, vol. 18, pp. 95–138, 1977.

[5] N. D. Sidropoulus, R. Bro, and G. B. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Transactions on Signal Processing*, vol. 48, no. 8, pp. 2377–2387, August 2000.

[6] K. Werner, M. Jansson, and P. Stoica, "Kronecker structured covariance matrix estimation," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2007.

[7] N. Lu and D. L. Zimmerman, "The likelihood ratio test for a separable covariance matrix," *Statistics and Probability Letters*, vol. 73, no. 5, pp. 449–457, May 2005.

[8] A. Richter, *Estimation of Radio Channel Parameters: Models and Algorithms*, Ph.d. dissertation, Technische Universität Ilmenau, Ilmenau, Germany, 2005, urn:nbn:de:gbv:ilm1-2005000111.

[9] J. Salmi, A. Richter, and V. Koivunen, "Enhanced tracking of radio propagation path parameters using state-space modeling," in *14th European Signal Processing Conference*, Florence, Italy, July 2006.

[10] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, 1991.

[11] L. L. Scharf, *Statistical Signal Processing*, Addison-Wesley Publishing Comp., Reading, MA, 1991.

[12] G. H. Golub and C. F. van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, 3rd edition, 1996.

[13] C. B. Ribeiro, A. Richter, and V. Koivunen, "Joint angular and delay domain mimo propagation parameter estimation using an approximate ml method," *IEEE Transactions on Signal Processing*, vol. 55, no. 10, pp. 4775–4790, October 2007.

[14] U. Trautwein, M. Landmann, G. Sommerkorn, and R. Thomä, "System-oriented measurement and analysis of mimo channels," in *COST 273 TD(05) 063*, Bologna, Italy, Jan. 2005, COST 273.