ESTIMATION OF RANK DEFICIENT COVARIANCE MATRICES WITH KRONECKER STRUCTURE

Mario H. Castañeda and Josef A. Nossek

Institute for Circuit Theory and Signal Processing
Technische Universität München, Germany
{mario.castaneda,josef.a.nossek}@tum.de

ABSTRACT

Given a set of observations, the estimation of covariance matrices is required in the analysis of many applications. To this end, any know structure of the covariance matrix can be taken into account. For instance, in case of separable processes, the covariance matrix is given by the Kronecker product of two *factor* matrices. Assuming the covariance matrix is full rank, the *maximum likelihood* (ML) estimate in this case leads to an iterative algorithm known as the *flip-flop* algorithm in the literature. In this work, we first generalize the flip-flop algorithm to the case when the covariance matrix is rank deficient, which happens to be the case in several situations. In addition, we propose a non-iterative estimation approach which incurs in a performance loss compared to the ML estimate, but at the expense of less complexity.

Index Terms— Kronecker product, separable processes, covariance matrix estimation, flip-flop algorithm

1. INTRODUCTION

A simple approach for estimating the covariance matrix given a set of observations, consists in computing the unstructured sample estimate. Such an approach, however, ignores any inherent structure of the covariance matrix and is outperformed by approaches which consider the structure. In a wide range of applications, we encounter *separable* processes which result in covariance matrices of the data expressed by the Kronecker product [1] of two factor covariance matrices. Such processes can be found in communications, where *multiple-input multiple-output* (MIMO) channels resulting from multiple transmit and receive antennas can be modeled according to a Kronecker model [2, 3] due to the spatial correlations at the transmitter and receiver. Spatio-temporal correlations [4] such as the noise processes in the signal modeling of *magneto-* (MEG) and *electroencephalography* (EEG) data [5, 6] can also be described by the Kronecker model.

For Gaussian distributed data, the *maximum likelihood* (ML) estimate of full-rank covariance matrices with a Kronecker structure has been derived in [7, 8]. Nonetheless, the ML estimator is not given in closed-form and is computed iteratively via the flip-flop algorithm [7, 9, 10], which basically performs an alternating estimation of the two factor matrices of the Kronecker product which define the covariance matrix. The algorithm is repeated until convergence. At each iteration, nevertheless, the algorithm requires the inversion of a previous estimate of one of the factor matrices.

In this work, we generalize the flip-flop algorithm to the case when the covariance matrices are *not* full rank, i.e. the data lies in a subspace (for example rank deficient MIMO channels). In contrast to the iterative ML estimator, we also present a non-iterative estimation approach based on closed-form estimates of the factor matrices which define the covariance matrix. Simulation results show that the proposed algorithm incurs in a performance loss compared to the flip-flop algorithm, but at the expense of less complexity since it is non-iterative and no matrix inversions are required. Besides the Kronecker structure, we assume no further structure of the covariance matrix as for instance being Toeplitz or persymmetric [11, 12] or any sparsity of the factor matrices [13].

1.1. Notation

We define scalars, column vectors and matrices with lower and upper case letters, lower case bold letters and upper case bold letters, respectively. The determinant of matrix $\mathbf{G}$ is given by $\det(\mathbf{G})$, while the pseudo-determinant of $\mathbf{G}$ is given by $\mathrm{pdet}(\mathbf{G})$, which is equal to the product of all non-zero eigenvalues of $\mathbf{G}$. The pseudo-inverse of $\mathbf{G}$ is given by $\mathbf{G}^{+}$. With $\mathbf{G} \in \mathbb{C}^{m\times n}$, the operator $\mathrm{vec}(\mathbf{G})$ returns the mn-dimensional vector resulting from stacking the columns of $\mathbf{G}$. The $m \times m$ identity matrix is denoted by $\mathbf{1}_m$.

2. PROBLEM SETUP

The data has zero mean and is Gaussian distributed. We focus on separable processes and denote $\mathbf{x}_k \in \mathbb{C}^{MN}$ as the k-th data sample vector consisting of MN entries, such that

$$\mathbf{x}_k = \mathbf{w}_k \otimes \mathbf{y}_k, \tag{1}$$

The support of the German Federal Ministry of Economics and Technology, through grant 50NA1110 is kindly acknowledged. .

where $\mathbf{w}_k \in \mathbb{C}^M$ and $\mathbf{y}_k \in \mathbb{C}^M$ are two independent stochastic processes with zero mean and covariance matrix

$$\mathbf{R}_1 = \mathrm{E}\left[\mathbf{w}_k \mathbf{w}_k^{\mathrm{H}}\right] \in \mathbb{C}^{N \times N} \tag{2}$$

$$\mathbf{R}_2 = \mathrm{E}\left[\mathbf{y}_k \mathbf{y}_k^{\mathrm{H}}\right] \in \mathbb{C}^{M \times M}, \tag{3}$$

respectively. Without loss of generality we assume the data to be zero mean, so the covariance matrix $\mathbf{R}$ of the data is given by the correlation matrix, which from (1), (2) and (3) and applying a property of the Kronecker product we obtain

$$\mathbf{R} = \mathrm{E}\left[\mathbf{x}_k \mathbf{x}_k^{\mathrm{H}}\right] = \mathbf{R}_1 \otimes \mathbf{R}_2 \in \mathbb{C}^{MN \times MN}. \tag{4}$$

In contrast to the general assumption found in the literature, we assume the covariance matrix $\mathbf{R}$ or equivalently the factor matrices $\mathbf{R}_1$ and $\mathbf{R}_2$, not to be necessarily full rank, i.e.

$$P = \mathrm{rank}\left(\mathbf{R}_1\right) \leq N \tag{5}$$

$$Q = \mathrm{rank}\left(\mathbf{R}_2\right) \leq M \tag{6}$$

$$\mathrm{rank}\left(\mathbf{R}\right) = P \cdot Q \leq M \cdot N. \tag{7}$$

where (7) results from a property of the Kronecker product.

Given K independent samples $\mathbf{x}_k$ for $k = 1, \ldots, K$, a simple approach to estimate the correlation matrix is:

$$\hat{\mathbf{R}} = \frac{1}{K} \sum_{k=1}^{K} \mathbf{x}_k \mathbf{x}_k^{\mathrm{H}}. \tag{8}$$

This estimator does not take any structure into account and therefore, leads to a significant performance degradation compared to other approaches which consider the structure.

3. FLIP-FLOP ALGORITHM

Given K samples $\mathbf{x}_1, \cdots, \mathbf{x}_K$ and under the assumption that $\mathbf{R}_1$ and $\mathbf{R}_2$ are *full* rank, such that $\mathbf{R}$ is also full rank, the likelihood function of the factor matrices $\mathbf{R}_1$ and $\mathbf{R}_2$ reads as

$$L_1(\mathbf{R}_1, \mathbf{R}_2 \mid \mathbf{x}_1, \ldots, \mathbf{x}_K) = \frac{\exp\left(-\sum_{k=1}^{K} \mathbf{x}_k^{\mathrm{H}} (\mathbf{R}_1 \otimes \mathbf{R}_2)^{-1} \mathbf{x}_k\right)}{\pi^K \det\left(\mathbf{R}_1 \otimes \mathbf{R}_2\right)^K}, \tag{9}$$

where we have employed (4). By taking the derivative of the log-likelihood function with respect to $\mathbf{R}_1$ and $\mathbf{R}_2$ and applying some properties of the determinant and inverse with the Kronecker product, the maximum likelihood estimator of the factor matrices leads to an iterative alternating algorithm called the flip-flop algorithm, where the ML estimates are given by [7, 9]

$$\hat{\mathbf{R}}_1 = \frac{1}{KM} \cdot \sum_{k=1}^{K} \mathbf{X}_k^{\mathrm{H}} \mathbf{R}_2^{-1} \mathbf{X}_k \tag{10}$$

$$\hat{\mathbf{R}}_2 = \frac{1}{KN} \cdot \sum_{k=1}^{K} \mathbf{X}_k \mathbf{R}_1^{-1} \mathbf{X}_k^{\mathrm{H}}, \tag{11}$$

where the matrix $\mathbf{X}_k \in \mathbb{C}^{M \times N}$ results from rearranging the entries of the vector $\mathbf{x}_k$ into an $M \times N$ matrix:

$$\mathbf{X}_k = \begin{bmatrix} \mathbf{x}_{k,1} & \mathbf{x}_{k,2} & \cdots & \mathbf{x}_{k,N} \end{bmatrix} \in \mathbb{C}^{M \times N}, \tag{12}$$

where the columns $\mathbf{x}_{k,n} \in \mathbb{C}^M$ for $n = 1, \ldots, N$ are defined such that $\mathbf{x}_k = \mathrm{vec}\left(\mathbf{X}_k\right)$, i.e.

$$\mathbf{x}_k = \begin{bmatrix} \mathbf{x}_{k,1}^{\mathrm{T}} & \mathbf{x}_{k,2}^{\mathrm{T}} & \cdots & \mathbf{x}_{k,N}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^{MN}. \tag{13}$$

The iterative and alternating nature of the flip-flop algorithm can be observed in (10) and (11). The algorithm is initialized, for instance, with $\mathbf{R}_2 = \mathbf{1}_M$, and at each step an estimate of one of the factor matrices, i.e. $\mathbf{R}_1$ or $\mathbf{R}_2$, is obtained based on a previous estimate of the other factor matrix. It is repeated until convergence according to several possible criteria, like the Frobenius norm of the estimated covariance matrix or the likelihood function [7].

With $\hat{\mathbf{R}}_{1,\mathrm{ML}}$ and $\hat{\mathbf{R}}_{2,\mathrm{ML}}$ as the final estimates for $\mathbf{R}_1$ and $\mathbf{R}_2$, the ML estimate of $\mathbf{R}$ is thus obtained from

$$\hat{\mathbf{R}}_{\mathrm{ML}} = \hat{\mathbf{R}}_{1,\mathrm{ML}} \otimes \hat{\mathbf{R}}_{2,\mathrm{ML}}. \tag{14}$$

Observe that the estimation of the factor matrices $\hat{\mathbf{R}}_1$ and $\hat{\mathbf{R}}_2$ is valid up to a non-zero scalar factor, since given an arbitrary scalar $\alpha \neq 0$, $\alpha \hat{\mathbf{R}}_{1,\mathrm{ML}}$ and $\alpha^{-1} \hat{\mathbf{R}}_{2,\mathrm{ML}}$ lead to the same estimate $\hat{\mathbf{R}}_{\mathrm{ML}}$ given in (14). In case we are interested in estimating $\mathbf{R}_1$ and $\mathbf{R}_2$ and not only $\mathbf{R}$, this ambiguity can be resolved by assuming, for example, that $\mathrm{tr}(\mathbf{R}_1)$ is known.

3.1. Rank Deficient Case

The ML estimates (10) and (11) have been derived on the assumption that $\mathbf{R}_1$ and $\mathbf{R}_2$, and in turn $\mathbf{R}$, have full rank. When considering the general case where $\mathbf{R}_1$ and $\mathbf{R}_2$ are rank deficient, the likelihood function (9) and (10) and (11) are no longer valid, since $\mathbf{R}_1^{-1}$ and $\mathbf{R}_2^{-1}$ do not exist.

For the case that $\mathbf{R} = \mathbf{R}_1 \otimes \mathbf{R}_2$ is not full rank, the distribution of the data is *degenerate* and therefore, we cannot properly write a density in the MN-dimensional space. Assuming the ranks (5) and (6) are $P < N$ and $Q < M$, such that the Gaussian distributed data is supported in an PQ-dimensional subspace of the MN-dimensional space. For this case, we can write the likelihood function as follows

$$L_2(\mathbf{R}_1, \mathbf{R}_2 \mid \mathbf{x}_1, \ldots, \mathbf{x}_K) = \frac{\exp\left(-\sum_{k=1}^{K} \mathbf{x}_k^{\mathrm{H}} \left(\mathbf{R}_1 \otimes \mathbf{R}_2\right)^{+} \mathbf{x}_k\right)}{\pi^K \mathrm{pdet}\left(\mathbf{R}_1 \otimes \mathbf{R}_2\right)^K}, \tag{15}$$

where $\mathbf{G}^{+}$ and $\mathrm{pdet}\left(\mathbf{G}\right)$ are the pseudo-inverse and pseudo-determinant of $\mathbf{G}$. We now discuss the ML estimates of $\mathbf{R}_1$ and $\mathbf{R}_2$ when these matrices are rank deficient. To this end, denote the *eigenvalue decomposition* (EVD) of $\mathbf{R}_1$ and $\mathbf{R}_2$ as

$$\mathbf{R}_1 = \mathbf{V} \boldsymbol{\Lambda}_1 \mathbf{V}^{\mathrm{H}} \tag{16}$$

$$\mathbf{R}_2 = \mathbf{U} \boldsymbol{\Lambda}_2 \mathbf{U}^{\mathrm{H}}, \tag{17}$$

where $\mathbf{V} \in \mathbb{C}^{N\times P}$, $\boldsymbol{\Lambda}_1 \in \mathbb{R}_+^{P\times P}$, $\mathbf{U} \in \mathbb{C}^{M\times Q}$, $\boldsymbol{\Lambda}_2 \in \mathbb{R}_+^{Q\times Q}$ with $\mathbf{V}^{\mathrm{H}}\mathbf{V} = \mathbf{1}_P$ and $\mathbf{U}^{\mathrm{H}}\mathbf{U} = \mathbf{1}_Q$.

The log-likelihood function, i.e. $L_3 = \log L_2(\mathbf{R}_1, \mathbf{R}_2 \mid \mathbf{x}_1, \ldots, \mathbf{x}_K)$, without constant terms, can be written as

$$L_3 = -K\cdot Q\log\left(\mathrm{pdet}\left(\mathbf{R}_1\right)\right) - K\cdot P\log\left(\mathrm{pdet}\left(\mathbf{R}_2\right)\right) - K\cdot\mathrm{tr}\left(\left(\mathbf{R}_1^+\otimes\mathbf{R}_2^+\right)\hat{\mathbf{R}}\right) \tag{18}$$

which follows by rewriting

$$\sum_{k=1}^{K}\mathbf{x}_k^{\mathrm{H}}\left(\mathbf{R}_1\otimes\mathbf{R}_2\right)^+\mathbf{x}_k = \sum_{k=1}^{K}\mathrm{tr}\left(\left(\mathbf{R}_1\otimes\mathbf{R}_2\right)^+\mathbf{x}_k\mathbf{x}_k^{\mathrm{H}}\right), = K\cdot\mathrm{tr}\left(\left(\mathbf{R}_1\otimes\mathbf{R}_2\right)^+\hat{\mathbf{R}}\right),$$

in (15), where recall $\hat{\mathbf{R}}$ is given in (8). The expression in (18) also results from employing the following identities [14]

$$\left(\mathbf{R}_1\otimes\mathbf{R}_2\right)^+ = \mathbf{R}_1^+\otimes\mathbf{R}_2^+$$
$$\mathrm{pdet}\left(\mathbf{R}_1\otimes\mathbf{R}_2\right) = \mathrm{pdet}\left(\mathbf{R}_1\right)^Q\cdot\mathrm{pdet}\left(\mathbf{R}_2\right)^P.$$

Let us focus on finding the ML estimate of $\mathbf{R}_2$. To this end, let us rewrite the log likelihood function (without constant terms) from (18) as

$$L_3 = -K\cdot Q\log\left(\mathrm{pdet}\left(\mathbf{R}_1\right)\right) - K\cdot P\log\left(\prod_{\ell=1}^{Q}\lambda_{2,\ell}\right) - K\cdot\mathrm{tr}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\mathbf{R}}_{ji}\cdot R_{1,ij}^+\cdot\mathbf{U}\boldsymbol{\Lambda}_2^{-1}\mathbf{U}^{\mathrm{H}}\right), \tag{19}$$

where we used (17) with $\lambda_{1,\ell}$ for $\ell = 1, \ldots, Q$ as the eigenvalues of $\mathbf{R}_2$ and have also applied the definition of the pseudo-determinant and pseudo-inverse. We also employ the fact, which follows from the derivation for the full rank case, that for the last term in (18) we can rewrite using (17)

$$\mathrm{tr}\left(\left(\mathbf{R}_1^+\otimes\mathbf{R}_2^+\right)\hat{\mathbf{R}}\right) = \mathrm{tr}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\mathbf{R}}_{ji}\cdot R_{1,ij}^+\cdot\mathbf{U}\boldsymbol{\Lambda}_2^{-1}\mathbf{U}^{\mathrm{H}}\right),$$

where $R_{1,ij}^+$ is the i,j-th element of $\mathbf{R}_1^+ \in \mathbb{C}^{N\times N}$ and $\hat{\mathbf{R}}_{ji} \in \mathbb{C}^{M\times M}$ for $i, j = 1, \ldots, N$ are defined such that

$$\hat{\mathbf{R}} = \begin{bmatrix}\hat{\mathbf{R}}_{11} & \cdots & \hat{\mathbf{R}}_{1N}\\ \vdots & & \vdots\\ \hat{\mathbf{R}}_{N1} & \cdots & \hat{\mathbf{R}}_{NN}\end{bmatrix} \in \mathbb{C}^{MN\times MN}. \tag{20}$$

We can find the ML estimate of $\lambda_{2,\ell}$, by taking the partial derivative of the log-likelihood function L_3 in (19) with respect to $\lambda_{2,\ell}$ and setting it to zero, which results in

$$\frac{\partial L_3}{\partial\lambda_{2,\ell}} = -\frac{KP}{\lambda_{2,\ell}} + \frac{K}{\lambda_{2,\ell}^2}\cdot\mathrm{tr}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\mathbf{R}}_{ji}R_{1,ij}^+\mathbf{u}_\ell\mathbf{u}_\ell^{\mathrm{H}}\right) = 0 \tag{21}$$

where $\mathbf{u}_\ell$ is the the eigenvector of $\mathbf{R}_2$ corresponding to the eigenvalue $\lambda_{2,\ell}$. By applying the property of the trace, from (21) we can write

$$\lambda_{2,\ell} = \frac{1}{P}\cdot\mathbf{u}_\ell^{\mathrm{H}}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\mathbf{R}}_{ji}R_{1,ij}^+\right)\mathbf{u}_\ell. \tag{22}$$

Furthermore, based on the $\mathbf{x}_{k,n} \in \mathbb{C}^M$ for $n = 1, \ldots, N$ defined in (13) and with (8) and (20), notice that

$$\hat{\mathbf{R}}_{ji} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{x}_{k,j}\mathbf{x}_{k,i}^{\mathrm{H}}, \tag{23}$$

such that with (12), we can rewrite (22) as

$$\lambda_{2,\ell} = \mathbf{u}_\ell^{\mathrm{H}}\left(\frac{1}{KP}\cdot\sum_{k=1}^{K}\mathbf{X}_k\mathbf{R}_1^+\mathbf{X}_k^{\mathrm{H}}\right)\mathbf{u}_\ell. \tag{24}$$

Since the columns of $\mathbf{X}_k$ lie in the subspace span$(\mathbf{U})$ and $\mathbf{R}_1^+$ is positive semi-definite, the rank of $\sum_{k=1}^{K}\mathbf{X}_k\mathbf{R}_1^+\mathbf{X}_k^{\mathrm{H}}$ is also Q. Therefore, from (24), we can conclude that $\lambda_{2,\ell}$ and $\mathbf{u}_\ell$ for $\ell = 1, \ldots, Q$ are the Q eigenvalues and eigenvectors of $\left(\frac{1}{KP}\cdot\sum_{k=1}^{K}\mathbf{X}_k\mathbf{R}_1^+\mathbf{X}_k^{\mathrm{H}}\right)$. From (24) and (17), we have that the ML estimate of $\mathbf{R}_2$ for the rank deficient case is

$$\check{\mathbf{U}}\check{\boldsymbol{\Lambda}}_2\check{\mathbf{U}}^{\mathrm{H}} = \check{\mathbf{R}}_2 = \frac{1}{KP}\cdot\sum_{k=1}^{K}\mathbf{X}_k\mathbf{R}_1^+\mathbf{X}_k^{\mathrm{H}}. \tag{25}$$

In a similar fashion and following also the derivation for the full rank case [10, 15], we can show that

$$\check{\mathbf{R}}_1 = \frac{1}{KQ}\cdot\sum_{k=1}^{K}\mathbf{X}_k^{\mathrm{H}}\mathbf{R}_2^+\mathbf{X}_k. \tag{26}$$

Comparing the ML estimates for the full rank case, i.e. (10) and (11), with the general case, i.e. (26) and (25), we observe that basically we have replaced the inverse with the pseudo-inverse and the M and N with the ranks Q and P. As in the full rank case, the flip-flop algorithm for the rank deficient case is iterative and alternating, requiring several iterations for convergence. At each iteration, a pseudo-inverse of a previous estimate of one of the factor matrices is required.

4. NON-ITERATIVE APPROACH

We now present a non-iterative approach for computing the estimates of $\mathbf{R}_1$ and $\mathbf{R}_2$. Based on (16) and (17), we can express for the rearranged k-th sample (see (12)) as

$$\mathbf{X}_k = \mathbf{U}\boldsymbol{\Lambda}_2^{\frac{1}{2}}\mathbf{Z}_k\boldsymbol{\Lambda}_1^{\frac{1}{2}}\mathbf{V}^{\mathrm{H}}, \tag{27}$$

where $\mathbf{Z}_k \in \mathbb{C}^{Q\times P}$ contains i.i.d. complex Gaussian random variables with zero mean and unit variance. Hence, defining $\mathbf{Z}_k$ based on its columns $\mathbf{z}_{k,i} \in \mathbb{C}^Q$ for $i = 1, \ldots, P$

$$\mathbf{Z}_k = \begin{bmatrix}\mathbf{z}_{k,1} & \cdots & \mathbf{z}_{k,P}\end{bmatrix} \in \mathbb{C}^{Q\times P}, \tag{28}$$

we have that $\mathrm{E}\left[\mathbf{z}_{k,m}\mathbf{z}_{k,m}^{\mathrm{H}}\right] = \mathbf{1}_Q$ for $i = 1,\ldots,P$. Using (27), (28), the previous expectation and $\mathbf{V}^{\mathrm{H}}\mathbf{V} = \mathbf{1}_P$ we can show that

$$\begin{aligned}\mathrm{E}\left[\mathbf{X}_k\mathbf{X}_k^{\mathrm{H}}\right] &= \mathbf{U}\boldsymbol{\Lambda}_2^{\frac{1}{2}}\,\mathrm{E}\left[\mathbf{Z}_k\boldsymbol{\Lambda}_1\mathbf{Z}_k^{\mathrm{H}}\right]\boldsymbol{\Lambda}_2^{\frac{1}{2}}\mathbf{U}^{\mathrm{H}} \\ &= \mathbf{U}\boldsymbol{\Lambda}_2^{\frac{1}{2}}\left(\sum_{i=1}^{P}\lambda_{1,i}\,\mathrm{E}\left[\mathbf{z}_{k,i}\mathbf{z}_{k,i}^{\mathrm{H}}\right]\right)\boldsymbol{\Lambda}_2^{\frac{1}{2}}\mathbf{U}^{\mathrm{H}} \\ &= \mathrm{tr}\left(\boldsymbol{\Lambda}_1\right)\cdot\mathbf{U}\boldsymbol{\Lambda}_2\mathbf{U}^{\mathrm{H}} \\ &= \mathrm{tr}\left(\boldsymbol{R}_1\right)\cdot\mathbf{R}_2. \end{aligned} \qquad (29)$$

Furthermore, we can similarly show that

$$\mathrm{E}\left[\mathbf{X}_k^{\mathrm{H}}\mathbf{X}_k\right] = \mathrm{tr}\left(\boldsymbol{R}_2\right)\cdot\mathbf{R}_1. \qquad (30)$$

By approximating $\mathrm{E}\left[\mathbf{X}_k\mathbf{X}_k^{\mathrm{H}}\right]$ with $\frac{1}{K}\sum_{k=1}^{K}\mathbf{X}_k\mathbf{X}_k^{\mathrm{H}}$ and $\mathrm{E}\left[\mathbf{X}_k^{\mathrm{H}}\mathbf{X}_k\right]$ with $\frac{1}{K}\sum_{k=1}^{K}\mathbf{X}_k^{\mathrm{H}}\mathbf{X}_k$, we obtain *closed-form* estimates of $\hat{\hat{\mathbf{R}}}_1$ and $\hat{\hat{\mathbf{R}}}_2$ from (29) and (30) as follows

$$\hat{\hat{\mathbf{R}}}_2 = \frac{1}{K\cdot\mathrm{tr}\left(\mathbf{R}_1\right)}\sum_{k=1}^{K}\mathbf{X}_k\mathbf{X}_k^{\mathrm{H}} \qquad (31)$$

$$\hat{\hat{\mathbf{R}}}_1 = \frac{1}{K\cdot\mathrm{tr}\left(\hat{\hat{\mathbf{R}}}_2\right)}\sum_{k=1}^{K}\mathbf{X}_k^{\mathrm{H}}\mathbf{X}_k. \qquad (32)$$

given the K observations $\mathbf{X}_k$ and assuming $\mathrm{tr}\left(\mathbf{R}_1\right)$ is known to resolve a scalar ambiguity in the estimation of the factor matrices. The closed-form estimate of $\mathbf{R}$ results from $\hat{\hat{\mathbf{R}}} = \hat{\hat{\mathbf{R}}}_2 \otimes \hat{\hat{\mathbf{R}}}_2$. If we are only interested, however, in estimating $\mathbf{R}$, then $\mathrm{tr}\left(\mathbf{R}_1\right)$ can be set to an arbitrary value without influencing the estimation of $\mathbf{R}$. Note that the estimates (31) and (32) are independent of the rank of $\mathbf{R}_1$ and $\mathbf{R}_2$.

5. NUMERICAL RESULTS

To evaluate the performance of the presented algorithms, we generate two random rank deficient covariance matrices $\mathbf{R}_1$ and $\mathbf{R}_2$, which are kept fixed throughout the simulation, with parameters $N = 30$ and $M = 8$ and with rank $P = 15$ and $Q = 4$, respectively. Besides the fact that they are Hermitian and positive semi-definite, no additional structure is imposed on $\mathbf{R}_1$ and $\mathbf{R}_2$. The covariance matrix $\mathbf{R}$ is computed from the Kronecker product of the generated factor matrices $\mathbf{R}_1$ and $\mathbf{R}_2$ (see (4)). As a figure of merit, we consider the normalized root *mean square error* (MSE) of the coefficients of the covariance matrix $\mathbf{R}$ averaged over $T = 100$ realizations. For instance, for the case of the unstructured estimate (8), the average normalized root MSE is $\sqrt{\frac{1}{T}\sum_{t=1}^{T}\frac{\|\mathbf{R}-\hat{\mathbf{R}}_t\|_{\mathrm{F}}^2}{\|\mathbf{R}\|_{\mathrm{F}}^2}}$, where $\|\mathbf{R}\|_{\mathrm{F}}^2$ is the Frobenius norm of $\mathbf{R}$ and $\hat{\mathbf{R}}_t$ is the unstructured estimate given the K data samples of the t-th realization. The K independent samples which are available for the estimation of $\mathbf{R}$ are generated from a complex Gaussian distribution with zero mean and covariance matrix $\mathbf{R}$. In Fig. 1, we compare four different estimators as a function of the number of samples K: the unstructured estimate (4), the ML estimate obtained via the generalized flip-flop algorithm (25) and (26), the non-iterative estimate obtained with (31) and (32) and a structured estimate obtained by finding the factor matrices whose Kronecker product best matches the unstructured estimate $\hat{\mathbf{R}}$ in the Frobenius norm sense [16]. As expected the unstructured estimate performs the worse, since it does not take the structure into account. Taking into account the structure with the structured estimate [16] (requires a rank one approximation of a $N^2 \times M^2$ matrix) leads to a better estimation. The best performance is achieved, however, with the flip-flop algorithm which provides the ML estimate. Although the non-iterative proposed approach is not as good as the flip-flop algorithm, it has much less computational complexity (due to the iterative nature and the computation of the pseudo-inverse of a factor matrix at each iteration in the flip-flop algorithm) and thus, could be a better choice in some applications. Different randomly generated factor matrices $\mathbf{R}_1$ and $\mathbf{R}_2$ lead to the same qualitative behavior.

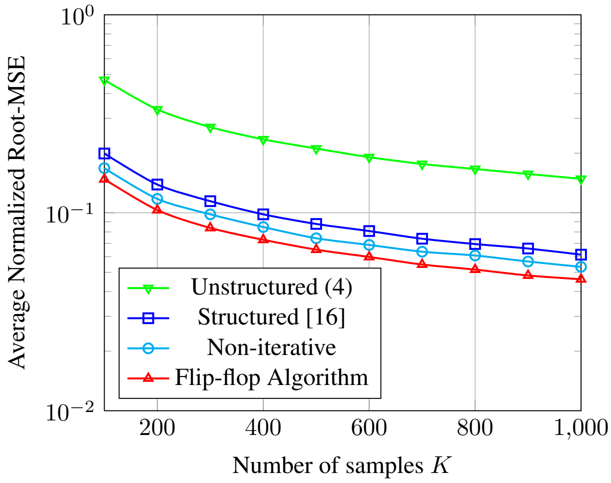

Fig. 1. Average normalized root-MSE as a function of the number of samples K for different estimators.

6. CONCLUSION

We have generalized the flip-flop algorithm for the case when the factor matrices are not full rank, and thus obtain, the ML estimator of rank deficient covariance matrices with a Kronecker structure. We have also presented a non-iterative approach which provides a closed-form estimate of the covariance matrix for the full rank or rank deficient case. Albeit simulation results show the flip-flop algorithm outperforms the proposed non-iterative approach, the performance gain might not be so significant when considering the higher complexity of the flip-flop algorithm, due to the iterative nature and a matrix pseudo-inversion required at each iteration.

7. REFERENCES

[1] C. F. van Loan, "The Ubiquitous Kronecker product," in *Journal of Computational and Applied Mathematics*, Vol. 123, pp. 85 – 100.

[2] J. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen, "A Stochastic MIMO Radio Channel Model with Experimental Validation," in *IEEE Journal on Sel. Areas in Comm.*, Vol. 20, No. 6, pp. 1211 – 1226, Aug. 2002.

[3] W. Weichselberger, M. Herdin, H. Ozcelik, and E. Bonek, "A Stochastic MIMO Channel Model with Joint Correlation of Both Link Ends," in *IEEE Trans. Wireless Commun.*, Vol.5, No.1, pp. 90 – 100, Jan. 2006.

[4] P. Wang, H. Li, B. Himed, "Knowledge-Aided Parametric Tests for Multichannel Adaptive Signal Detection," in *IEEE Trans. on Signal Proc.*, Vol. 59, No. 12, pp. 5970 – 5982 Dec. 2011.

[5] J. C. De Munck, H. M. Huizenaga, L. J. Waldorp and R. M. Heethaar, "Estimating Stationary Dipoles from MEG/EEG Data Contaminated with Spatially and Temporally Correlated Background Noise," in *IEEE Trans. on Signal Proc.*, Vol. 50, No. 7, pp. 1565 – 1572, Aug. 2002

[6] H. M. Huizenaga, J. C. de Munck, L. J. Waldorp, R. P. P. P. Grasman, "Spatiotemporal EEG/MEG Source Analysis based on a Parametric Noise Covariance Model," in *IEEE Trans. Biomed. Eng.*, Vol 49, No. 6, pp. 533 – 539, Jun. 2002.

[7] N. Lu and D. Zimmerman, "On the Likelihood-based Inference for a Separable Covariance Matrix," Technical Report 337, Dept. Of Statistics and Actuarial Science, Univ. of Iowa, Iowa City, Iowa, 2004.

[8] P. Dutilleul, "The MLE Algorithm for the Matrix Normal Distribution," in *Journal of Statistical Computation and Simulation*, vol. 64, pp. 105 – 123, 1999.

[9] F. Bijma, J. C de Munck, R. M. Heethaar, "The Spatiotemporal MEG Covariance Matrix Modeled as a Sum of Kronecker Products," in *NeuroImage*, Vol. 27, No. 2, pp. 402 – 415, Aug. 2005.

[10] K. Werner, M. Jansson, and P. Stoica, "On Estimation of Covariance Matrices with Kronecker Product Structure," in *IEEE Trans. on Signal Proc.*, Vol. 56, No. 2, pp. 478 – 491, Feb. 2008.

[11] M. Jansson, P. Wirfält, K. Werner, B. Ottersten, "ML Estimation of Covariance Matrices with Kronecker and Persymmetric Structure," in *Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop (DSP/SPE)*, pp. 298 – 301, Jan. 2009.

[12] P. Wirfalt and M. Jansson, "On Toeplitz and Kronecker structured covariance matrix estimation," in *IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pp. 185 – 188, Oct. 2010.

[13] T. Tsiligkaridis and A. O. Hero, "Sparse covariance estimation under Kronecker product structure," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3633 – 3636, Mar. 2012.

[14] K. B. Petersen and M. S. Pedersen, "The Matrix Cookbook," `http://matrixcookbook.com`, Nov. 2012.

[15] K. Werner, M. Jansson, and P. Stoica, "Kronecker structured covariance matrix estimation," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 825 – 828, Apr. 2007.

[16] C. van Loan and N. Pitsianis, "Approximation with Kronecker Products," from *Linear Algebra for Large Scale and Real- Time Applications*, Kluwer Publications, Netherlands, 1993.