KRONECKER STRUCTURED COVARIANCE MATRIX ESTIMATION

Karl Werner, Magnus Jansson[*]

Signal Processing,
School of Electrical Engineering,
Royal Institute of Technology (KTH),
SE-100 44 Stockholm, Sweden
{karl.werner,magnus.jansson}@ee.kth.se

Petre Stoica

Information Technology,
Department of Systems and Control,
Uppsala University,
P O Box 337, SE-751 05 Uppsala, Sweden
peter.stoica@it.uu.se

ABSTRACT

The estimation of signal covariance matrices is a crucial part of many signal processing algorithms. In some applications, the structure of the problem suggests that the underlying, true, covariance matrix is the Kronecker product of two matrices. Examples of such problems are channel modelling for MIMO communications and signal modelling of EEG data. In applications it may also be that the Kronecker factors in turn can be assumed to possess additional, linear, structure. The maximum likelihood (ML) estimator for the problem has been proposed previously. It is asymptotically efficient but has the drawback of requiring an iterative search. Two methods that are both non-iterative and asymptotically efficient are proposed in this paper. The first method is derived from a well-known iterative maximization technique for the likelihood function. It performs on par with ML in simulations, but has the drawback of not allowing for extra structure in addition to the Kronecker structure. The second method is based on covariance matching principles, and does not suffer from this drawback. However, while the large sample performance is shown to be identical to ML, it performs somewhat worse in small samples than the first estimator. In addition, the Cramér-Rao lower bound (CRB) for the problem is derived in a compact form.

Index Terms— Estimation, MIMO systems, Covariance matrices, Maximum likelihood estimation

1. INTRODUCTION

In statistical modelling of multiple input multiple output (MIMO) channels, Kronecker structured channel covariance matrices are often assumed [1], [2]. This assumption implies that

$$\mathrm{Cov}\,[\,\mathrm{vec}\{\mathbf{H}\}\,] = \mathbf{A} \otimes \mathbf{B} \tag{1}$$

where $\mathbf{H}$ is the stochastic $n \times m$ channel matrix, $\otimes$ denotes Kronecker matrix product, $\mathrm{vec}\{\,\}$ denotes the vectorization operator (see, e.g., [3]), $\mathbf{A}$ is an $m \times m$ transmit covariance matrix, and $\mathbf{B}$ is an $n \times n$ receive covariance matrix. Estimating such matrices is useful in the design and analysis of signal processing algorithms for MIMO communications. Imposing the structure implied by the Kronecker assumption gives the advantages of leading to more accurate estimators, of reducing the number of parameters needed when feeding back channel statistics, and of allowing for a reduced algorithm complexity. Models such as (1) also appear naturally when modelling spatio-temporal noise processes in MEG/EEG data [4]. In statistics, processes with covariance matrices that satisfy (1) are referred to as separable [5], [6]. The problem of estimating such a covariance matrix from a set of data naturally leads to the maximum likelihood (ML) method. As the optimization problem associated with the ML method lacks a known closed-form solution, an iterative search algorithm has to be used. The standard choice seems to be the minimization with respect to (w.r.t) $\mathbf{A}$ and $\mathbf{B}$ alternately, keeping the other matrix fixed at the previous estimate. The algorithm performs well in numerical studies [5]. However, it has the drawback of being iterative and it does not allow for imposing a general linear structure on the $\mathbf{A}$ and $\mathbf{B}$ matrices in addition to the positive definiteness implied by the problem formulation.

Another common approach is to simply calculate the (unstructured) sample covariance matrix of the data and then find the closest (in the Frobenius norm sense) Kronecker structured approximation. This approximation problem is treated in [7], [8], [2]. The corresponding method lacks the asymptotical efficiency of the ML approach but has the advantage of simplicity and low computational complexity. In this approach it is also possible to incorporate an additional linear structure on the $\mathbf{A}$ and $\mathbf{B}$ matrices (as will be demonstrated).

In this work we present a new method for the estimation of Kronecker structured covariance matrices based on a covariance matching criterion (see Section 6). The method is non-iterative and has a relatively low computational complexity. It is also shown to be asymptotically efficient. Similar to the Kronecker approximation method discussed above, it allows for linearly structured $\mathbf{A}$ and $\mathbf{B}$ matrices.

In addition, we propose a non-iterative version of the method mentioned above for ML estimation. The proposed method can be seen as terminating the iteration after three steps. It is shown analytically that the resulting estimate is asymptotically efficient, regardless of initialization, and numerical simulations indicate a very promising performance. However, the method has the drawback of not allowing for imposing additional linear structure.

Furthermore, the CRB for the problem is derived in Section 5.

In the following, $\mathbf{X}^\dagger$ and $|\mathbf{X}|$ denote the Moore-Penrose pseudo-inverse and determinant of the matrix $\mathbf{X}$, respectively. The permutation matrix $\mathbf{K}_{x,y}$ is defined such that $\mathbf{K}_{x,y}\mathrm{vec}\{\mathbf{X}\} = \mathrm{vec}\{\mathbf{X}^T\}$ if $\mathbf{X}$ is an $x \times y$ matrix. The notation $\mathbf{X}^{1/2}$ denotes a Hermitian square-root of the matrix $\mathbf{X}$. The i,jth element of the matrix $\mathbf{X}$ is denoted $[\mathbf{X}]_{ij}$. The superscript * denotes conjugate transpose and ^{T} denotes transpose. Also $\mathbf{X}^c = \mathbf{X}^{T*}$. The notation $\dot{\mathbf{X}}_j$ denotes the element-wise derivative of the matrix $\mathbf{X}$ w.r.t. the parameter at the jth position in the parameter vector in question. The notation $x_N = o_p(a_N)$ means that $\lim_{N\to\infty} \frac{x_N}{a_N} = 0$ in probability. In this work the asymptotic results hold when N tends to infinity.

This work was supported in part by the Swedish Research Council.
Corresponding author is Karl Werner, tel: +46 8 7907246.

2. PROBLEM FORMULATION

Let $\mathbf{x}(t)$ be a zero-mean, complex Gaussian, circularly symmetric random vector with Kronecker structured covariance matrix

$$\mathrm{E}\,[\,\mathbf{x}(k)\mathbf{x}^*(l)\,] = \mathbf{R}_0\delta(k,l),\ \ \mathbf{R}_0 = \mathbf{A}_0 \otimes \mathbf{B}_0 \tag{2}$$

where the $m \times m$ matrix $\mathbf{A}_0$ and the $n \times n$ matrix $\mathbf{B}_0$ are p.d. (positive definite) Hermitian matrices. The problem considered is to estimate $\mathbf{R}_0$ from the observed samples $[\mathbf{x}(t)]_{t=0}^{N-1}$.

For later use, define the $n_A \times 1-$vector $\boldsymbol{\theta}_\mathbf{A}$ and the $n_B \times 1-$vector $\boldsymbol{\theta}_\mathbf{B}$ as the real vectors necessary to parameterize $\mathbf{A}$ and $\mathbf{B}$, respectively. Furthermore, assume a linear structure:

$$\mathrm{vec}\{\mathbf{A}\} = \mathbf{P}_\mathbf{A}\boldsymbol{\theta}_\mathbf{A},\ \ \mathrm{vec}\{\mathbf{B}\} = \mathbf{P}_\mathbf{B}\boldsymbol{\theta}_\mathbf{B} \tag{3}$$

where $\mathbf{P}_\mathbf{A}$ and $\mathbf{P}_\mathbf{B}$ are data and parameter independent matrices of size $m^2 \times n_A$ and $n^2 \times n_B$ respectively. If the only structure imposed is that $\mathbf{A}$ and $\mathbf{B}$ are Hermitian matrices, then $n_A = m^2$ and $n_B = n^2$. Also introduce the concatenated parameter vector $\boldsymbol{\theta} = [\boldsymbol{\theta}_\mathbf{A}^T\ \boldsymbol{\theta}_\mathbf{B}^T]^T$. Note that this parameterization is ambiguous since $\mathbf{A}_0\alpha$ and $\mathbf{B}_0\alpha^{-1}$ give the same $\mathbf{R}_0$ for any $\alpha \neq 0$. Hence, $\mathbf{A}_0$ and $\mathbf{B}_0$ can only be estimated up to a scalar factor.

3. MAXIMUM LIKELIHOOD ESTIMATION

The ML estimator for the above problem has been proposed in, e.g., [5]. The associated maximization problem has no known closed-form solution. The negative log-likelihood function for the problem is (excluding constant terms)

$$l(\boldsymbol{\theta}_\mathbf{A}, \boldsymbol{\theta}_\mathbf{B}) = m\log|\mathbf{B}| + n\log|\mathbf{A}| + \mathrm{tr}\{\hat{\mathbf{R}}\,(\mathbf{A}\otimes\mathbf{B})^{-1}\} \tag{4}$$

where

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{t=0}^{N-1}\mathbf{x}(t)\mathbf{x}^*(t).$$

The last term of (4) can be rewritten as

$$\mathrm{tr}\{\hat{\mathbf{R}}\,(\mathbf{A}^{-1}\otimes\mathbf{B}^{-1})\} = \mathrm{tr}\{\sum_{k=1}^{m}\sum_{l=1}^{m}\hat{\mathbf{R}}^{kl}[\mathbf{A}^{-1}]_{lk}\mathbf{B}^{-1}\}$$

where $\hat{\mathbf{R}}^{kl}$ is the k,lth block of size $n \times n$ in the matrix $\hat{\mathbf{R}}$. Using a standard result on minimization of functions of the form (4) it is clear that given a fixed $\mathbf{A}$, the $\mathbf{B}$ minimizing (4) is given by

$$\hat{\mathbf{B}}(\mathbf{A}) = \frac{1}{m}\sum_{k=1}^{m}\sum_{l=1}^{m}\hat{\mathbf{R}}^{kl}[\mathbf{A}^{-1}]_{lk} \tag{5}$$

(see, e.g., [9]). It can be shown that $\hat{\mathbf{B}}(\mathbf{A})$ is positive definite when both $\mathbf{A}$ and $\hat{\mathbf{R}}$ are positive definite. Similarly, given a fixed $\mathbf{B}$, the minimizing $\mathbf{A}$ is

$$\hat{\mathbf{A}}(\mathbf{B}) = \frac{1}{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\bar{\mathbf{R}}^{kl}[\mathbf{B}^{-1}]_{lk} \tag{6}$$

where $\bar{\mathbf{R}}^{kl}$ is the k,lth $m \times m$ block of $\bar{\mathbf{R}} = \mathbf{K}_{m,n}^T\hat{\mathbf{R}}\mathbf{K}_{m,n}$. Furthermore, $\hat{\mathbf{A}}(\mathbf{B})$ is p.d. when $\mathbf{B}$ and $\hat{\mathbf{R}}$ are p.d.. The so-called flipflop algorithm [5] can be outlined as follows:

1. Select an initial estimate $\mathbf{A} = \mathbf{A}^0$.
2. Set i:=0. Using (5), find the $\mathbf{B}^0 = \hat{\mathbf{B}}(\mathbf{A}^0)$ that minimizes (4) w.r.t. $\mathbf{B}$ given $\mathbf{A} = \mathbf{A}^0$.
3. Set $i := i+1$. Using (6), find the $\mathbf{A}^i = \hat{\mathbf{A}}(\mathbf{B}^{i-1})$ that minimizes (4) given $\mathbf{B} = \mathbf{B}^{i-1}$.
4. Set $i := i+1$. Using (5), find the $\mathbf{B}^i = \hat{\mathbf{B}}(\mathbf{A}^{i-1})$ that minimizes (4) given $\mathbf{A} = \mathbf{A}^{i-1}$.
5. Iterate steps 3 and 4 until convergence.

An interesting alternative to the above procedure is to perform steps 1 to 4 without iterating. See Section 4.

Note that it is unclear how to incorporate a general linear structure of the form (3) into the flipflop algorithm.

Clearly the ML estimate can be found using other search techniques as well. Using either (5) or (6), the negative log-likelihood function can be concentrated so that the search is over n_A or n_B parameters, whichever is smallest. Based on numerical evidence, it is our experience that the flipflop algorithm converges faster than a Newton type search.

4. A NON-ITERATIVE FLIPFLOP APPROACH

The proposed estimate

$$\hat{\mathbf{R}}_{FF} = \hat{\mathbf{A}}(\hat{\mathbf{B}}(\mathbf{A}^0)) \otimes \hat{\mathbf{B}}(\hat{\mathbf{A}}(\hat{\mathbf{B}}(\mathbf{A}^0))) \tag{7}$$

is the result of steps 1-4 of the flipflop algorithm discussed above. The initial estimate $\mathbf{A}^0$ is an arbitrary p.d. matrix (and need not be data dependent). In the following it will be shown that $\hat{\mathbf{R}}_{FF}$ is an asymptotically efficient estimate of the covariance matrix independently of the initialization. In order to state the result, consider the rearrangement function [7]

$$R(\mathbf{R}) = \left[\mathrm{vec}\{\mathbf{R}^{11}\}\ \ldots\ \mathrm{vec}\{\mathbf{R}^{m1}\}\ \ldots\ \mathrm{vec}\{\mathbf{R}^{mm}\}\right]^T$$

where $\mathbf{R}^{kl}$ is the k,lth $n \times n$ block of $\mathbf{R}$. It will also be useful to introduce two other matrices that are obtained by rearranging the elements of the sample covariance matrix. They are

$$\hat{\mathbf{R}}_\mathbf{B} = \left[\mathrm{vec}\{\hat{\mathbf{R}}^{11}\}\ \ldots\ \mathrm{vec}\{\hat{\mathbf{R}}^{1m}\}\ \ldots\ \mathrm{vec}\{\hat{\mathbf{R}}^{mm}\}\right] \text{ and}$$

$$\hat{\mathbf{R}}_\mathbf{A} = \left[\mathrm{vec}\{\bar{\mathbf{R}}^{11}\}\ \ldots\ \mathrm{vec}\{\bar{\mathbf{R}}^{1n}\}\ \ldots\ \mathrm{vec}\{\bar{\mathbf{R}}^{nn}\}\right].$$

Also introduce the corresponding permutation matrices that satisfy

$$\mathrm{vec}\{\hat{\mathbf{R}}_\mathbf{A}\} = \mathbf{P}_{\mathbf{R}_\mathbf{A}}\mathrm{vec}\{\hat{\mathbf{R}}\},\ \ \mathrm{vec}\{\hat{\mathbf{R}}_\mathbf{B}\} = \mathbf{P}_{\mathbf{R}_\mathbf{B}}\mathrm{vec}\{\hat{\mathbf{R}}\},$$

$$\mathrm{vec}\{\hat{\mathbf{R}}\} = \mathbf{P}_R\mathrm{vec}\{R(\hat{\mathbf{R}})\}.$$

We are now ready to state the result.

Theorem 1 *Let $\hat{\mathbf{R}}_{FF}$ be the estimate of $\mathbf{R}_0$ given by (7). Then it has an asymptotic covariance given by*

$$\lim_{N\to\infty} N\mathrm{Cov}\left[\,\mathrm{vec}\{\hat{\mathbf{R}}_{FF}\}\,\right] = \boldsymbol{\Xi}(\mathbf{R}_0^T\otimes\mathbf{R}_0)\boldsymbol{\Xi}^*$$

where

$$\begin{aligned}\boldsymbol{\Xi} =& \mathbf{P}_R\Big[\frac{1}{n}\big(\mathrm{vec}\{\mathbf{B}_0\}\mathrm{vec}^T\{\mathbf{B}_0^{-1}\}\otimes\mathbf{I}_{m^2}\big)\mathbf{P}_{\mathbf{R}_\mathbf{A}}\\ &+\frac{1}{m}\big(\mathbf{I}_{n^2}\otimes\mathrm{vec}\{\mathbf{A}_0\}\mathrm{vec}^T\{\mathbf{A}_0^{-1}\}\big)\mathbf{K}_{n^2,m^2}\mathbf{P}_{\mathbf{R}_\mathbf{B}}\\ &-\frac{1}{mn}\big(\mathrm{vec}\{\mathbf{B}_0\}\mathrm{vec}^T\{\mathbf{A}_0^{-T}\}\\ &\qquad\otimes\mathrm{vec}\{\mathbf{A}_0\}\mathrm{vec}^T\{\mathbf{B}_0^{-1}\}\big)\mathbf{K}_{m^2,n^2}\mathbf{P}_{\mathbf{R}_\mathbf{A}}\Big].\end{aligned} \tag{8}$$

A proof is given in [10]. It is interesting to note that $\boldsymbol{\Xi}$ and the expression for the asymptotic covariance do not depend on the initial value $\mathbf{A}^0$! A similar result can be shown for the ML method.

Theorem 2 *Let $\hat{\mathbf{R}}_{ML}$ be the ML estimate of $\mathbf{R}_0$ in the model defined in Section 2. Then*

$$\lim_{N\to\infty} N\mathrm{Cov}\left[\,\mathrm{vec}\{\hat{\mathbf{R}}_{ML}\}\,\right] = \boldsymbol{\Xi}(\mathbf{R}_0^T\otimes\mathbf{R}_0)\boldsymbol{\Xi}^*$$

where $\boldsymbol{\Xi}$ is given by (8).

A proof is given in [10]. The somewhat surprising conclusion is that the asymptotic (in N) covariances of the ML estimate and the estimate $\hat{\mathbf{R}}_{FF}$ coincide *regardless of initialization* $\mathbf{A}^0$! Numerical studies in Section 7 also suggest very promising small sample performance for $\hat{\mathbf{R}}_{FF}$. Clearly this result together with the asymptotic efficiency of ML give us an expression for the Cramér-Rao lower bound for the problem in the special case when no linear structure is imposed. A more general and compact expression that can take linear structure into account is derived in Section 5.

5. THE CRAMÉR-RAO LOWER BOUND

In order to simplify derivations, note that

$$\mathrm{vec}\{\mathbf{A}\otimes\mathbf{B}\} = \mathbf{P}\boldsymbol{\eta}, \quad \boldsymbol{\eta} = \mathrm{vec}\{\boldsymbol{\theta}_{\mathbf{A}}\boldsymbol{\theta}_{\mathbf{B}}^T\}.$$

where $\mathbf{P} = \mathbf{P}_R(\mathbf{P}_{\mathbf{B}}\otimes\mathbf{P}_{\mathbf{A}})$. The i,jth element of the Fisher information matrix (FIM) is given by [9]

$$[\mathbf{I}(\boldsymbol{\theta})]_{i,j} = N\mathrm{tr}\{\mathbf{R}^{-1}\dot{\mathbf{R}}_i\mathbf{R}^{-1}\dot{\mathbf{R}}_j\}.$$

Construct a matrix $\boldsymbol{\Gamma}$ such that $[\boldsymbol{\Gamma}]_{j,i} = \frac{\partial[\boldsymbol{\eta}]_j}{\partial[\boldsymbol{\theta}]_i}$ which, when evaluated at $\boldsymbol{\theta}$, reads

$$\boldsymbol{\Gamma} = \left(\boldsymbol{\theta}_{\mathbf{B}}\otimes\mathbf{I}_{n_A}\ \mathbf{I}_{n_B}\otimes\boldsymbol{\theta}_{\mathbf{A}}\right).$$

Since $\mathrm{vec}\{\dot{\mathbf{R}}_j\} = \mathbf{P}\dot{\boldsymbol{\eta}}_j$, this immediately gives an expression for the FIM

$$\mathbf{I}(\boldsymbol{\theta}) = N\boldsymbol{\Gamma}^T\mathbf{P}^*\left(\mathbf{R}^{-T}\otimes\mathbf{R}^{-1}\right)\mathbf{P}\boldsymbol{\Gamma}.$$

Some care must be exercised when using this result to find the CRB for the elements of $\mathbf{R}$. The reason is that the mapping between the parameter vector $\boldsymbol{\theta}$ and the matrix $\mathbf{R}$ is many-to-one due to the ambiguous scaling of $\mathbf{A}$ and $\mathbf{B}$ mentioned above and possibly also due to the imposed linear structure. By using results proved in [11] we have that the sought CRB is given by $\boldsymbol{\Delta}\mathbf{I}^{\dagger}(\boldsymbol{\theta}_0)\boldsymbol{\Delta}^*$, where column i of $\boldsymbol{\Delta}$ is given by $\left.\frac{\partial\mathrm{vec}\{\mathbf{R}\}}{\partial[\boldsymbol{\theta}]_i}\right|_{\mathbf{R}=\mathbf{R}_0}$. It is then straightforward to conclude that $\boldsymbol{\Delta} = \mathbf{P}\boldsymbol{\Gamma}_0$, where $\boldsymbol{\Gamma}_0$ is equal to $\boldsymbol{\Gamma}$ evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. and thus that

$$CRB = \frac{1}{N}\mathbf{P}\boldsymbol{\Gamma}_0(\boldsymbol{\Gamma}_0^T\mathbf{P}^*\left(\mathbf{R}_0^{-T}\otimes\mathbf{R}_0^{-1}\right)\mathbf{P}\boldsymbol{\Gamma}_0)^{\dagger}\boldsymbol{\Gamma}_0^T\mathbf{P}^*. \quad (9)$$

6. A COVARIANCE MATCHING APPROACH

The CRB derived above can only be achieved by an estimator taking the imposed linear structure [see (3)] into account. It is not obvious how to incorporate such structure into the iterative and non-iterative flipflop algorithms described above. This section aims at developing a non-iterative algorithm that achieves the CRB also when a general linear structure is assumed.

A simple standard approach to the present estimation problem is to form the estimate of $\mathbf{R}_0$ from the minimizers of

$$\min_{\mathbf{A},\mathbf{B}}\ \|\hat{\mathbf{R}} - \mathbf{A}\otimes\mathbf{B}\|_F. \quad (10)$$

This minimization problem can be rewritten as [7]

$$\min_{\mathbf{A},\mathbf{B}}\ \|R(\hat{\mathbf{R}}) - \mathrm{vec}\{\mathbf{A}\}\mathrm{vec}^T\{\mathbf{B}\}\|_F \quad (11)$$

where $R(\hat{\mathbf{R}})$ is the rearrangement function introduced in Section 4. The reformulated minimization problem is a rank-one approximation problem that is easy to solve using the SVD. The resulting estimate is consistent since $\hat{\mathbf{R}}$ is consistent, but not asymptotically efficient. Incorporating a linear structure of the kind (3) can be done similar to what is shown at the end of this section.

It will be shown that, in constrast to the minimizers of (10), the estimate $\hat{\boldsymbol{\theta}}$ obtained by minimizing

$$\begin{aligned} V(\boldsymbol{\theta}) &= \|\hat{\mathbf{R}} - \mathbf{A}\otimes\mathbf{B}\|_{\mathbf{Q}} \\ &= \mathrm{vec}^*\{\hat{\mathbf{R}} - \mathbf{A}\otimes\mathbf{B}\}\mathbf{Q}\mathrm{vec}\{\hat{\mathbf{R}} - \mathbf{A}\otimes\mathbf{B}\} \end{aligned} \quad (12)$$

is asymptotically statistically efficient if $\mathbf{Q}$ is chosen as

$$\mathbf{Q} = \frac{1}{N}\left(\mathrm{Cov}\left[\mathrm{vec}\{\hat{\mathbf{R}}\}\right]\right)^{-1} = \mathbf{R}_0^T\otimes\mathbf{R}_0. \quad (13)$$

This result is not surprising, especially in the light of the *extended invariance principle* [12], [13]. Note that (13) depends on the unknown parameters. Replacing $\mathbf{Q}$ with a consistent estimate

$$\hat{\mathbf{Q}} = \mathbf{Q} + o_p(1) \quad (14)$$

does not affect the asymptotic efficiency (see Theorem 3 below). For a general structured $\hat{\mathbf{Q}}$, the minimization problem (12) lacks a simple closed form solution and iterative methods similar to the flipflop algorithm need to be used. Here, we will use a specially structured $\mathbf{Q}$ for which the minimization problem in (13) can be solved in closed form. We suggest using the weighting matrix

$$\hat{\mathbf{Q}} = (\hat{\mathbf{A}}^{-1}\otimes\hat{\mathbf{B}}^{-1})^T\otimes(\hat{\mathbf{A}}^{-1}\otimes\hat{\mathbf{B}}^{-1}) \quad (15)$$

where $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ are selected to be the closed form estimates given by (10). This choice ensures positive definiteness of $\hat{\mathbf{Q}}$ when $\hat{\mathbf{R}}$ is p.d., [7] and it also satisfies (14). With this choice of $\hat{\mathbf{Q}}$, the criterion function in (12) can be written as

$$\|\check{\mathbf{R}} - \mathrm{vec}\{\hat{\mathbf{A}}^{-1/2}\mathbf{A}\hat{\mathbf{A}}^{-1/2}\}\mathrm{vec}^T\{\hat{\mathbf{B}}^{-1/2}\mathbf{B}\hat{\mathbf{B}}^{-1/2}\}\|_F \quad (16)$$

where

$$\check{\mathbf{R}} = R((\hat{\mathbf{A}}^{-1/2}\otimes\hat{\mathbf{B}}^{-1/2})\hat{\mathbf{R}}(\hat{\mathbf{A}}^{-1/2}\otimes\hat{\mathbf{B}}^{-1/2})).$$

Using (3) in (16) gives

$$\|\hat{\mathbf{R}} - \mathbf{A}\otimes\mathbf{B}\|_{\hat{\mathbf{Q}}} = \|\check{\mathbf{R}} - \mathbf{Q}_{\mathbf{A}}\mathbf{T}_{\mathbf{A}}\boldsymbol{\theta}_{\mathbf{A}}\boldsymbol{\theta}_{\mathbf{B}}^T\mathbf{T}_{\mathbf{B}}^T\mathbf{Q}_{\mathbf{B}}^T\|_F \quad (17)$$

where $\mathbf{Q}_{\mathbf{A}}$ and $\mathbf{Q}_{\mathbf{B}}$ are orthonormal matrices and $\mathbf{T}_{\mathbf{B}}$ and $\mathbf{T}_{\mathbf{A}}$ are invertible matrices such that

$$\begin{aligned} \mathbf{Q}_{\mathbf{A}}\mathbf{T}_{\mathbf{A}} &= (\hat{\mathbf{A}}^{-T/2}\otimes\hat{\mathbf{A}}^{-1/2})\mathbf{P}_{\mathbf{A}}, \\ \mathbf{Q}_{\mathbf{B}}\mathbf{T}_{\mathbf{B}} &= (\hat{\mathbf{B}}^{-T/2}\otimes\hat{\mathbf{B}}^{-1/2})\mathbf{P}_{\mathbf{B}}. \end{aligned}$$

The criterion in (17) can be rewritten as

$$\|\hat{\mathbf{R}} - \mathbf{A}\otimes\mathbf{B}\|_{\hat{\mathbf{Q}}} = \|\mathbf{Q}_{\mathbf{A}}^*\check{\mathbf{R}}\mathbf{Q}_{\mathbf{B}}^c - \mathbf{T}_{\mathbf{A}}\boldsymbol{\theta}_{\mathbf{A}}\boldsymbol{\theta}_{\mathbf{B}}^T\mathbf{T}_{\mathbf{B}}^T\|_F. \quad (18)$$

The rank-one approximation problem in (18) is easily solved using SVD. The proposed estimator has a fixed computational complexity similar to that of the unweighted ad-hoc method, (10), and yet it achieves asymptotical efficiency, as will be shown in Theorem 3. The performance for finite sample sizes will be evaluated using simulations in Section 7.

One note is in place here. While the true covariance matrix is known to be p.d., this restriction is not imposed on (18). However, since the estimated covariance matrix is consistent, it will be p.d. for large enough N. The conclusion is that the asymptotic performance of the estimator is not affected by relaxing the positive definiteness constraint. We conclude by stating:

Theorem 3 *Let $\hat{\mathbf{R}}_C$ be an estimate of $\mathbf{R}_0$ constructed as*

$$\hat{\mathbf{R}}_C = \hat{\mathbf{A}}_C\otimes\hat{\mathbf{B}}_C$$

where $\hat{\mathbf{A}}_C$ and $\hat{\mathbf{B}}_C$ are minimizers of (12). Then

$$\begin{aligned} &\lim_{N\to\infty} N\mathrm{Cov}\left[\mathrm{vec}\{\hat{\mathbf{R}}_C\}\right] = \\ &\qquad \frac{1}{N}\mathbf{P}\boldsymbol{\Gamma}_0(\boldsymbol{\Gamma}_0^T\mathbf{P}^*\left(\mathbf{R}_0^{-T}\otimes\mathbf{R}_0^{-1}\right)\mathbf{P}\boldsymbol{\Gamma}_0)^{\dagger}\boldsymbol{\Gamma}_0^T\mathbf{P}^*. \end{aligned} \quad (19)$$

Furthermore, (19) still holds if $\mathbf{Q}$ in (12) is replaced by $\hat{\mathbf{Q}}$ given in (15).

A proof is given in [10]. The estimator is asymptotically efficient since (19) coincides with (9).

7. NUMERICAL STUDY

Monte Carlo simulations were used to evaluate small sample performance of the proposed methods. Two Hermitian, p.d. matrices $\mathbf{A}_0$ and $\mathbf{B}_0$ were randomly generated (and then fixed) and $\mathbf{R}_0$ was calculated. The fact that no additional structure was imposed allows all methods to be used. In each Monte Carlo trial, N independent samples were generated from a complex Gaussian distribution with covariance $\mathbf{R}_0$. Then each estimator was applied to the sample set and the normalized root-MSE was calculated as $\sqrt{\frac{1}{L}\sum_{k=1}^{L}\frac{\|\mathbf{R}_0-\hat{\mathbf{R}}_k\|_F}{\|\mathbf{R}_0\|_F}}$ where $\hat{\mathbf{R}}_k$ is the estimate produced by the estimator in question in Monte Carlo trial k and L is the number of Monte Carlo trials.

Five alternative estimators were tried: i) The unstructured sample covariance matrix, that does not utilize the known Kronecker structure of the problem; ii) The unweighted approximation of the sample covariance matrix by a Kronecker structured matrix [see (10)]; iii) The proposed method with the structured weighting matrix given by (15); iv) The ML method (implemented using the iterative flipflop algorithm) and v) The proposed non-iterative flipflop method with $\mathbf{A}^0 = \mathbf{I}_m$. The resulting normalized root-MSE as a function of the sample size is shown in Figure 1. The matrix dimensions used were $m = n = 4$. Our conclusion based on numerical evidence is that the global minimum is found in general in the ML problem, also when initialized far from the true value. A Newton search for the minimum gave exactly the same results in all experiments, regardless of initialization. For the non-iterative flipflop algorithm it was shown in Section 4 that the initialization does not affect the asymptotical results, but this does not rule out possible effects on performance for finite sample sizes. It is thus interesting to note that, in this example, the proposed non-iterative version of the flipflop algorithm performs as well as the ML. The weighted approximation method proposed in Section 6 performs worse than the ML based methods for the smallest sample sizes, but approaches the CRB for large N. As expected, the unweighted approximation method does not reach the CRB.

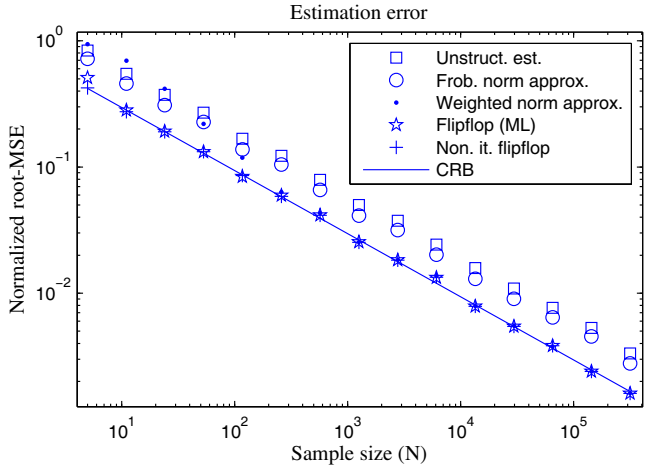

Fig. 1. Normalized root-MSE as a function of sample size for different alternative estimators. Simulations consisting of $L = 100$ Monte Carlo runs were used. The figure shows the results of an experiment where $\mathbf{A}_0$ and $\mathbf{B}_0$ are Hermitian but otherwise unstructured. The matrix dimensions were $m = n = 4$.

8. CONCLUSION

In this paper we have treated the problem of estimating the Kronecker structured covariance matrix from a set of N samples. Two cases were considered:

The Kronecker factors, $\mathbf{A}_0$ and $\mathbf{B}_0$, are Hermitian and p.d. but no other structure is assumed. It has previously been proposed that the ML estimate can be computed by the iterative so-called flipflop algorithm. In Section 4 we showed that a non-iterative algorithm can be derived that is asymptotically efficient. In an example, the proposed algorithm also showed small sample performance that is fully comparable to ML.

The second case is more general since it allows for linear structure (as defined in Section 2) of the Kronecker factors. For this case we suggested a method based on covariance matching. The proposed method is non-iterative and asymptotically efficient.

In Section 5, we derived the CRB for the estimation problem. Due to the asymptotical efficiency of the two proposed methods, the CRB also gives their asymptotical covariance.

9. REFERENCES

[1] J.P Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 6, pp. 1211–1226, Aug. 2002.

[2] K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, and P. Karlsson, "Modeling of wide-band MIMO radio channels based on NLoS indoor measurements," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 8, pp. 655–665, May 2004.

[3] Roger A. Horn and Charles R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, 1991.

[4] J. C. de Munck, H. M. Huizenga, L. J. Waldorp, and R. M. Heethaar, "Estimating stationary dipoles from MEG/EEG data contaminated with spatially and temporally correlated background noise," *IEEE Transactions on Signal Processing*, vol. 50, no. 7, pp. 1565–1572, Jul. 2002.

[5] N. Lu and D.L. Zimmerman, "On likelihood-based inference for a separable covariance matrix," Technical Report 337, 2004.

[6] 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.

[7] C. van Loan and N. Pitsianis, "Approximation with Kronecker products," in *Linear Algebra for Large Scale and Real Time Applications*, pp. 293–314. Kluwer Publications, 1993.

[8] P. Strobach, "Low-rank detection of multichannel gaussian signals using block matrix approximation," *IEEE Transactions on Signal Processing*, vol. 43, no. 1, pp. 233–242, Jan. 1995.

[9] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Prentice Hall, Upper Saddle River, NJ 07458, 2005.

[10] Karl Werner, Magnus Jansson, and Petre Stoica, "On estimation of Kronecker structured covariance matrices," *IEEE Transactions on Signal Processing (in preparation)*.

[11] P. Stoica and T.L. Marzetta, "Parameter estimation problems with singular information matrices," *IEEE Transactions on Signal Processing*, vol. 49, no. 1, pp. 87–90, Jan. 2001.

[12] P. Stoica and T. Söderström, "On reparameterization of loss functions used in estimation and the invariance principle," *Signal Processing*, vol. 17, pp. 383–387, 1989.

[13] B. Ottersten, P. Stoica, and R. Roy, "Covariance matching estimation techniques for array signal processing applications," in *Digital Signal Processing*, vol. 8, pp. 185–210. 1998.