On Estimation of Covariance Matrices With Kronecker Product Structure

Karl Werner, *Student Member, IEEE*, Magnus Jansson, *Member, IEEE*, and Petre Stoica, *Fellow, IEEE*

***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 valid covariance matrices. Examples of such problems are channel modeling for multiple-input multiple-output (MIMO) communications and signal modeling 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) method for the associated estimation problem has been proposed previously. It is asymptotically efficient but has the drawback of requiring an iterative search for the maximum of the likelihood function. Two methods that are fast and noniterative are proposed in this paper. Both methods are shown to be asymptotically efficient. The first method is a noniterative variant of a well-known alternating 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 the same, it performs somewhat worse than the first estimator in small samples. In addition, the Cramér–Rao lower bound for the problem is derived in a compact form. The problem of estimating the Kronecker factors and the problem of detecting if the Kronecker structure is a good model for the covariance matrix of a set of samples are related. Therefore, the problem of detecting the dimensions of the Kronecker factors based on the minimum values of the criterion functions corresponding to the two proposed estimation methods is also treated in this work.**

***Index Terms*—covariance matching, Cramér–Rao bound, Kronecker model, multiple-input multiple-output (MIMO) channel modeling , structured covariance matrix estimation .**

I. Introduction

In the statistical modeling of multiple-input multiple-output (MIMO) wireless communications channels, covariance matrices with a Kronecker product structure are often assumed [8], [19], [3]. This assumption implies that

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

Manuscript received October 18, 2006; revised June 5, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Franz Hlawatsch. This work was supported in part by the Swedish Research Council.

K. Werner and M. Jansson are with the ACCESS Linnaeus Center, Electrical Engineering, KTH—Royal Institute of Technology, SE-100 44 Stockholm, Sweden (e-mail: karl.werner@ee.kth.se;magnus.jansson@ee.kth.se).

P. Stoica is with the Systems and Control Division, Information Technology, Department of Information Technology, Uppsala University, SE-751 05 Uppsala, Sweden (e-mail: peter.stoica@it.uu.se).

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.2007.907834

where $\mathbf{M}$ is the stochastic $n \times m$ channel matrix, $\mathbf{A}$ is an $m \times m$ transmit covariance matrix, and $\mathbf{B}$ is an $n \times n$ receive covariance matrix. The vector $\mathrm{vec}\{\mathbf{M}\}$ is obtained by stacking the columns of $\mathbf{M}$ on top of each other and $\otimes$ denotes the Kronecker matrix product (see e.g., [6]). Estimating such matrices is useful in the design and analysis of signal processing algorithms for MIMO communications. Imposing the structure implied by the Kronecker product 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 modeling spatio-temporal noise processes in MEG/EEG data [4], [17].

In statistics, processes with covariance matrices that satisfy (1) are referred to as separable [9], [10]. They appear when variables in the data set can be cross-classified by two vector valued factors. The process

$$\mathbf{x}(t) = \mathbf{a}(t) \otimes \mathbf{b}(t) \tag{2}$$

contains products of all possible pairs of an element from $\mathbf{a}(t)$ and an element from $\mathbf{b}(t)$, where $\mathbf{a}(t)$ and $\mathbf{b}(t)$ are two stochastic processes. If $\mathbf{a}(t)$ and $\mathbf{b}(t)$ have zero means and are mutually uncorrelated with covariance matrices

$$\begin{aligned} \mathrm{Cov}[\mathbf{a}(t)] &= \mathbf{A} \\ \mathrm{Cov}[\mathbf{b}(t)] &= \mathbf{B} \end{aligned} \tag{3}$$

then $\mathbf{x}(t)$ has a covariance matrix of the form (1). Some examples of problems where separable stochastic processes appear are given in [9]. A framework for estimation of a related class of structured covariance matrices is developed in [2], together with examples of applications where they appear.

In practical scenarios, the properties of the underlying problem sometimes imply that the matrices $\mathbf{A}$ and $\mathbf{B}$ in (1) have additional structure. In the MIMO communications scenario for example, uniform linear arrays (ULAs) at the receiver or transmitter side imply that $\mathbf{A}$ or $\mathbf{B}$ has Toeplitz structure. If the estimator is able to take such additional structure into account, the performance might be improved. Such performance gains have been reported for array processing when the signal covariance matrix has Toeplitz structure [5].

The intimately connected problems of estimating such a covariance matrix from a set of data and evaluating whether (1)

is an accurate model for the underlying true covariance matrix naturally lead 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. This algorithm is called the flip-flop algorithm in [9]. The algorithm performs well in numerical studies [9]. Numerical experience indicates that the flip-flop algorithm converges faster than a Newton-type search [9]; the global minimum was found in general in our experiments. The flip-flop algorithm, however, has the drawback of being iterative and it does not allow for 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) approximation with a Kronecker product structure. This approximation problem is treated in [17]–[19]. The corresponding method lacks the asymptotic efficiency of the ML approach but has the advantage of simplicity and low computational complexity. It is also possible to incorporate an additional linear structure of the $\mathbf{A}$ and $\mathbf{B}$ matrices in this approach (as will be demonstrated).

In this paper, we derive a new method for the estimation of covariance matrices with Kronecker product structure based on a covariance matching criterion (see Section V). The method is noniterative and has a relatively low computational complexity. It is also shown to be asymptotically efficient. Similarly to the Kronecker approximation method discussed above, it allows for linearly structured $\mathbf{A}$ and $\mathbf{B}$ matrices.

In addition, we propose a noniterative version of the flip-flop method for ML estimation. The proposed method can be seen as the flip-flop algorithm terminated after three iterations. It is shown analytically that the resulting estimate is asymptotically efficient, regardless of initialization (see Section IV), and numerical simulations indicate a very promising small-sample performance. However, the method has the drawback of not allowing for additional linear structure.

Furthermore, the Cramér–Rao lower bound (CRB) for the problem is derived in Section VI.

The problem of determining if (1) is an appropriate model for the covariance matrix of a set of samples is treated in Section VIII. For the noniterative version of the flip-flop algorithm, the generalized-likelihood ratio test (GLRT) is proposed. For the covariance matching approach, the minimum value of the criterion function can be used. Its asymptotic distribution is therefore derived in Section VIII.

Computer simulations are used to compare the asymptotic, theoretical results to empirical results in Section IX.

In the following, $\mathbf{X}^{\dagger}$ and $|\mathbf{X}|$ denote the Moore–Penrose pseudoinverse and the determinant of the matrix $\mathbf{X}$, respectively. For a positive semidefinite (p.s.d.) matrix $\mathbf{X}$, the notation $\mathbf{X}^{1/2}$ denotes the unique p.s.d. matrix that satisfies $\mathbf{X}^{1/2}\mathbf{X}^{1/2} = \mathbf{X}$. The i,jth element of the matrix $\mathbf{X}$ is denoted $[\mathbf{X}]_{ij}$. The superscript * denotes the conjugate transpose and ^{T} denotes the transpose. Also $\mathbf{X}^c = \mathbf{X}^{T*}$. The notation $\dot{\mathbf{X}}_j$ or $(\dot{\mathbf{X}})_j$ denotes the elementwise derivative of the matrix $\mathbf{X}$ w.r.t. the parameter at the jth position in the parameter vector in question. Finally, the notation $x_N = o_p(a_N)$ means that

$$\lim_{N\to\infty} \frac{x_N}{a_N} = 0 \tag{4}$$

in probability. In this paper, the asymptotic results hold when the number of samples N tends to infinity.

II. Problem Formulation

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

$$\mathrm{E}[\mathbf{x}(k)\mathbf{x}^*(l)] = \mathbf{R}_0\delta(k,l). \tag{5}$$

The covariance matrix $\mathbf{R}_0$ is assumed to have a Kronecker product structure, i.e.,

$$\mathbf{R}_0 = \mathbf{A}_0 \otimes \mathbf{B}_0 \tag{6}$$

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

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 used to parameterize $\mathbf{A}$ and $\mathbf{B}$, respectively. Furthermore, assume that $\mathbf{A}$ and $\mathbf{B}$ depend linearly on $\boldsymbol{\theta}_{\mathbf{A}}$ and $\boldsymbol{\theta}_{\mathbf{B}}$:

$$\begin{aligned} \mathrm{vec}\{\mathbf{A}\} &= \mathbf{P}_{\mathbf{A}}\boldsymbol{\theta}_{\mathbf{A}} \\ \mathrm{vec}\{\mathbf{B}\} &= \mathbf{P}_{\mathbf{B}}\boldsymbol{\theta}_{\mathbf{B}} \end{aligned} \tag{7}$$

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. The matrices $\mathbf{P}_{\mathbf{A}}$ and $\mathbf{P}_{\mathbf{B}}$ are required to have full rank. 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 $\boldsymbol{\theta} = [\boldsymbol{\theta}_{\mathbf{A}}^T \boldsymbol{\theta}_{\mathbf{B}}^T]^T$. Denote the parameter vector that corresponds to $\mathbf{A}_0$ and $\mathbf{B}_0$ by $\boldsymbol{\theta}_0$. Note that the Kronecker product parameterization is ambiguous since

$$\mathbf{A}\gamma \otimes \mathbf{B}\gamma^{-1} = \mathbf{A} \otimes \mathbf{B} \tag{8}$$

for any $\gamma \neq 0$. Hence, we can only estimate $\mathbf{A}_0$ and $\mathbf{B}_0$ up to a scalar factor.

III. Maximum-Likelihood Estimation

The ML estimator for the above estimation problem has been proposed, e.g., in [9] and [11]. The associated maximization problem has no known closed-form solution. In this section,

methods for iteratively calculating the ML estimate will be reviewed. The negative log-likelihood function for the problem is (excluding constant terms)

$$f(\boldsymbol{\theta}_{\mathbf{A}},\boldsymbol{\theta}_{\mathbf{B}}) = m\log|\mathbf{B}|+n\log|\mathbf{A}|+\mathrm{tr}\{\hat{\mathbf{R}}(\mathbf{A}^{-1}\otimes\mathbf{B}^{-1})\} \quad (9)$$

where

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{t=0}^{N-1}\mathbf{x}(t)\mathbf{x}^{*}(t). \quad (10)$$

The last term in (9) can be rewritten as

$$\mathrm{tr}\{\hat{\mathbf{R}}(\mathbf{A}^{-1}\otimes\mathbf{B}^{-1})\} = m\mathrm{tr}\left\{\frac{1}{m}\sum_{k=1}^{m}\sum_{l=1}^{m}\hat{\mathbf{R}}^{kl}[\mathbf{A}^{-1}]_{lk}\mathbf{B}^{-1}\right\} \quad (11)$$

where $\hat{\mathbf{R}}^{kl}$ is the k,lth block of size $n\times n$ in the matrix $\hat{\mathbf{R}}$. It is a standard result that the p.d. minimizer, $\mathbf{X}$, of

$$\log|\mathbf{X}| + \mathrm{tr}\{\mathbf{R}\mathbf{X}^{-1}\} \quad (12)$$

where $\mathbf{R}$ is p.d., is $\mathbf{X}=\mathbf{R}$ (see, e.g., [1], [14]). Hence, for a fixed $\mathbf{A}$, the $\mathbf{B}$ minimizing (9) is given by

$$\hat{\mathbf{B}}(\mathbf{A}) = \frac{1}{m}\sum_{k=1}^{m}\sum_{l=1}^{m}\hat{\mathbf{R}}^{kl}[\mathbf{A}^{-1}]_{lk} \quad (13)$$

assuming it is p.d. Note that $\hat{\mathbf{B}}(\mathbf{A})$ is Hermitian by construction. However, a structure of the kind of (7) is in general hard to impose. If such a structure is to be imposed, it is unclear how to express the minimizer in closed form, except for special cases. In order to show that $\hat{\mathbf{B}}(\mathbf{A})$ is p.d. when $\hat{\mathbf{R}}$ is p.d., let $\mathbf{y}$ be an arbitrary $n\times 1$ complex vector and consider

$$\mathbf{y}^{*}\hat{\mathbf{B}}(\mathbf{A})\mathbf{y} = \frac{1}{m}\sum_{k=1}^{m}\sum_{l=1}^{m}\mathbf{y}^{*}\hat{\mathbf{R}}^{kl}\mathbf{y}[\mathbf{A}^{-1}]_{lk}. \quad (14)$$

By defining a matrix $\mathbf{W}$ such that

$$[\mathbf{W}]_{kl} = \mathbf{y}^{*}\hat{\mathbf{R}}^{kl}\mathbf{y} \quad (15)$$

we have $\mathbf{W} = (\mathbf{I}_m\otimes\mathbf{y}^{*})\hat{\mathbf{R}}(\mathbf{I}_m\otimes\mathbf{y})$ and

$$\begin{aligned}\mathbf{y}^{*}\hat{\mathbf{B}}(\mathbf{A})\mathbf{y} &= \frac{1}{m}\mathrm{tr}\{\mathbf{W}\mathbf{A}^{-1}\}\\ &= \frac{1}{m}\mathrm{tr}\{(\mathbf{I}_m\otimes\mathbf{y}^{*})\hat{\mathbf{R}}(\mathbf{I}_m\otimes\mathbf{y})\mathbf{A}^{-1}\}. \quad (16)\end{aligned}$$

Clearly, $\mathbf{y}^{*}\hat{\mathbf{B}}(\mathbf{A})\mathbf{y}>0$ for any $\mathbf{y}\neq\mathbf{0}$ if both $\hat{\mathbf{R}}$ and $\mathbf{A}$ are p.d. In fact, this result is conservative in the sense that $\hat{\mathbf{B}}(\mathbf{A})$ can be p.d. even if $\hat{\mathbf{R}}$ is only p.s.d. In the real case, at least, it is possible to relax the condition on $\hat{\mathbf{R}}$ significantly [10].

After inserting $\hat{\mathbf{B}}(\mathbf{A})$ into (9) and after removing constant terms, the concentrated criterion function becomes

$$f(\mathbf{A}) = m\log|\hat{\mathbf{B}}(\mathbf{A})| + n\log|\mathbf{A}|. \quad (17)$$

A Newton-type search procedure can be used to search for the minimum of $f(\mathbf{A})$ w.r.t. $\mathbf{A}$. It is necessary to make sure that the estimate is p.s.d., for example by setting $\mathbf{A}=\mathbf{C}\mathbf{C}^{*}$ and minimizing w.r.t. $\mathbf{C}$.

If $m>n$, it is more advantageous to concentrate $\mathbf{A}$ out instead, and then minimize the resulting criterion function w.r.t. $\mathbf{B}$. Similar to (13), for a fixed $\mathbf{B}$, the $\mathbf{A}$ that minimizes (9) is

$$\hat{\mathbf{A}}(\mathbf{B}) = \frac{1}{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\bar{\mathbf{R}}^{kl}[\mathbf{B}^{-1}]_{lk} \quad (18)$$

where $\bar{\mathbf{R}}^{kl}$ is the k,lth $m\times m$ block of

$$\bar{\mathbf{R}} = \mathbf{K}_{m,n}^{T}\hat{\mathbf{R}}\mathbf{K}_{m,n}. \quad (19)$$

The permutation matrix $\mathbf{K}_{x,y}$ is defined such that

$$\mathbf{K}_{x,y}\mathrm{vec}\{\mathbf{X}\} = \mathrm{vec}\{\mathbf{X}^{T}\} \quad (20)$$

for any $x\times y$ matrix $\mathbf{X}$. A similar argument as above guarantees that $\hat{\mathbf{A}}(\mathbf{B})$ is p.d. as long as $\mathbf{B}$ and $\hat{\mathbf{R}}$ are p.d.

The flip-flop algorithm is obtained by alternately minimizing w.r.t. $\mathbf{B}$ and $\mathbf{A}$, keeping the last available estimate of $\mathbf{A}$ fixed while minimizing w.r.t. $\mathbf{B}$ and vice versa [9]–[11]. This algorithm can be outlined as follows:

1) Select an initial estimate $\mathbf{A}=\mathbf{A}_{\mathrm{init}}$.
2) Set $i:=0$. Using (13), find the $\mathbf{B}^{0}=\hat{\mathbf{B}}(\mathbf{A}_{\mathrm{init}})$ that minimizes (9) w.r.t. $\mathbf{B}$ given $\mathbf{A}=\mathbf{A}_{\mathrm{init}}$.
3) Set $i:=i+1$. Using (18), find the $\mathbf{A}^{i}=\hat{\mathbf{A}}(\mathbf{B}^{i-1})$ that minimizes (9) given $\mathbf{B}=\mathbf{B}^{i-1}$.
4) Set $i:=i+1$. Using (13), find the $\mathbf{B}^{i}=\hat{\mathbf{B}}(\mathbf{A}^{i-1})$ that minimizes (9) given $\mathbf{A}=\mathbf{A}^{i-1}$.
5) Iterate steps 3) and 4) until convergence.

Numerical experiments have been reported which indicate that the flip-flop algorithm can converge faster than a Newton-type search [9].

Imposing a linear structure of the form (7) into the flip-flop algorithm can be done by imposing (7) in each step. However, as pointed out above, this typically would require running an iterative search in steps 2), 3), and 4) above.

An interesting alternative to the iterative search for the minimum of the negative likelihood function is to perform only steps 1)–4), without iterating. This method can be shown to inherit the asymptotic efficiency of the ML approach. See the next section.

IV. Noniterative Flip-Flop Approach

The proposed estimate

$$\hat{\mathbf{R}}_{FF} = \hat{\mathbf{A}}(\hat{\mathbf{B}}(\mathbf{A}_{\text{init}})) \otimes \hat{\mathbf{B}}(\hat{\mathbf{A}}(\hat{\mathbf{B}}(\mathbf{A}_{\text{init}}))) \tag{21}$$

is the result of steps 1) to 4) of the flip-flop algorithm discussed above. The initial estimate $\mathbf{A}_{\text{init}}$ 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 $\mathbf{R}_0$ independently of the initialization. In order to state the result, consider the rearrangement function [18]

$$R(\mathbf{R}) = \begin{bmatrix} \text{vec}^T\{\mathbf{R}^{11}\} \\ \vdots \\ \text{vec}^T\{\mathbf{R}^{m1}\} \\ \text{vec}^T\{\mathbf{R}^{21}\} \\ \vdots \\ \text{vec}^T\{\mathbf{R}^{mm}\} \end{bmatrix} \tag{22}$$

where $\mathbf{R}^{kl}$ is the k,lth $n \times n$ block of $\mathbf{R}$. This rearrangement function has the property that

$$R(\mathbf{A} \otimes \mathbf{B}) = \text{vec}\{\mathbf{A}\}\text{vec}^T\{\mathbf{B}\}. \tag{23}$$

It is easy to see that a permutation matrix $\mathbf{P}_R$ can be defined such that

$$\text{vec}\{\mathbf{R}\} = \mathbf{P}_R \text{vec}\{R(\mathbf{R})\} \tag{24}$$

for any matrix $\mathbf{R}$ of compatible dimensions. It will also be useful to introduce two other matrices that are obtained by rearranging the elements of the sample covariance matrix. They are

$$\hat{\mathbf{R}}_{\mathbf{B}} = [\text{vec}\{\hat{\mathbf{R}}^{11}\} \cdots \text{vec}\{\hat{\mathbf{R}}^{1m}\} \cdots \text{vec}\{\hat{\mathbf{R}}^{mm}\}] \tag{25}$$

and

$$\hat{\mathbf{R}}_{\mathbf{A}} = [\text{vec}\{\bar{\mathbf{R}}^{11}\} \cdots \text{vec}\{\bar{\mathbf{R}}^{1n}\} \cdots \text{vec}\{\bar{\mathbf{R}}^{nn}\}]. \tag{26}$$

Also introduce the corresponding permutation matrices that satisfy

$$\begin{aligned} \text{vec}\{\hat{\mathbf{R}}_{\mathbf{A}}\} &= \mathbf{P}_{\mathbf{R}_{\mathbf{A}}} \text{vec}\{\hat{\mathbf{R}}\} \\ \text{vec}\{\hat{\mathbf{R}}_{\mathbf{B}}\} &= \mathbf{P}_{\mathbf{R}_{\mathbf{B}}} \text{vec}\{\hat{\mathbf{R}}\}. \end{aligned} \tag{27}$$

We are now ready to state the result.

Theorem 1: Let $\hat{\mathbf{R}}_{FF}$ be the estimate of $\mathbf{R}_0$ given by (21). Then, under the data model described in Section II, $\hat{\mathbf{R}}_{FF}$ has an asymptotic complex Gaussian distribution with covariance given by

$$\lim_{N\to\infty} N\text{Cov}[\text{vec}\{\hat{\mathbf{R}}_{FF}\}] = \mathbf{\Xi}\left(\mathbf{R}_0^T \otimes \mathbf{R}_0\right)\mathbf{\Xi}^* \tag{28}$$

where

$$\begin{aligned} \mathbf{\Xi} = \mathbf{P}_R \Bigg[& \frac{1}{n}\left(\text{vec}\{\mathbf{B}_0\}\text{vec}^T\left\{\mathbf{B}_0^{-1}\right\} \otimes \mathbf{I}_{m^2}\right)\mathbf{P}_{\mathbf{R}_{\mathbf{A}}} \\ & + \frac{1}{m}\left(\mathbf{I}_{n^2} \otimes \text{vec}\{\mathbf{A}_0\}\text{vec}^T\left\{\mathbf{A}_0^{-1}\right\}\right)\mathbf{K}_{n^2,m^2}\mathbf{P}_{\mathbf{R}_{\mathbf{B}}} \\ & - \frac{1}{mn}\left(\text{vec}\{\mathbf{B}_0\}\text{vec}^T\left\{\mathbf{A}_0^{-T}\right\}\right. \\ & \left. \otimes \text{vec}\{\mathbf{A}_0\}\text{vec}^T\left\{\mathbf{B}_0^{-1}\right\}\right)\mathbf{K}_{m^2,n^2}\mathbf{P}_{\mathbf{R}_{\mathbf{A}}}\Bigg]. \end{aligned} \tag{29}$$

The matrix $\mathbf{K}_{n^2,m^2}$ is defined in (20). Furthermore, $\hat{\mathbf{R}}_{FF}$ is a consistent estimate of $\mathbf{R}_0$.

Proof: See Appendix I. ∎

It is interesting to note that the expression for the asymptotic covariance does not depend on the initial value $\mathbf{A}_{\text{init}}$. This implies that the dominating part (for large N) of the estimation error is the same as if the initialization was at the true value itself.

A similar result can be shown for the ML method.

Theorem 2: Let $\mathbf{x}(t)$ be defined as in Section II, where $\mathbf{A}_0$ and $\mathbf{B}_0$ are p.d. but otherwise assumed unstructured. Let

$$\hat{\mathbf{R}}_{\text{ML}} = \hat{\mathbf{A}}_{\text{ML}} \otimes \hat{\mathbf{B}}_{\text{ML}}. \tag{30}$$

be an ML estimate of $\mathbf{R}_0$, given $\mathbf{x}(t), t = 0, \ldots, N-1$. Then, $\hat{\mathbf{R}}_{\text{ML}}$ has an asymptotic complex Gaussian distribution, and

$$\lim_{N\to\infty} N\text{Cov}[\text{vec}\{\hat{\mathbf{R}}_{\text{ML}}\}] = \mathbf{\Xi}\left(\mathbf{R}_0^T \otimes \mathbf{R}_0\right)\mathbf{\Xi}^* \tag{31}$$

where $\mathbf{\Xi}$ is given by (29). Furthermore, $\hat{\mathbf{R}}_{\text{ML}}$ is a consistent estimate of $\mathbf{R}_0$.

Proof: See Appendix II. ∎

Clearly, this result together with the asymptotic efficiency of ML gives us an expression for the asymptotic Cramér–Rao lower bound for 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 VI.

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 the initialization* $\mathbf{A}_{\text{init}}$. Both estimates are consistent. Hence, $\hat{\mathbf{R}}_{FF}$ is an asymptotically efficient estimate, when no structure is imposed on $\mathbf{A}$ and $\mathbf{B}$ except that they are Hermitian matrices. Numerical studies in Section IX also suggest a promising small sample performance for $\hat{\mathbf{R}}_{FF}$.

It is interesting to note that the estimate

$$\hat{\mathbf{A}}(\hat{\mathbf{B}}(\mathbf{A}_{\text{init}})) \otimes \hat{\mathbf{B}}(\mathbf{A}_{\text{init}}) \tag{32}$$

which is obtained by terminating the algorithm after step 3 lacks the asymptotic efficiency of $\hat{\mathbf{R}}_{FF}$. This can be shown using techniques similar to those used in the proof of Theorem 1 (the asymptotic covariance of (32) will depend on $\mathbf{A}_{\text{init}}$).

V. Covariance Matching Approach

It is not obvious how to modify the noniterative version of the flip-flop algorithm proposed above to take linear structure as in (7) into account without making the algorithm iterative. This section aims at developing a noniterative method that achieves the CRB for large N also when a general linear structure is imposed.

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

$$\min_{\mathbf{A},\mathbf{B}} \|\hat{\mathbf{R}} - \mathbf{A} \otimes \mathbf{B}\|_F^2. \tag{33}$$

This minimization problem can be rewritten [18] as

$$\min_{\mathbf{A},\mathbf{B}} \|R(\hat{\mathbf{R}}) - \text{vec}\{\mathbf{A}\}\text{vec}^T\{\mathbf{B}\}\|_F^2 \tag{34}$$

where $R(\hat{\mathbf{R}})$ is the rearrangement function introduced in Section IV. The reformulated minimization problem is a rank-one approximation problem that is easy to solve using the singular value decomposition (SVD). The resulting estimate is consistent since $\hat{\mathbf{R}}$ is consistent, but it is not asymptotically efficient. Incorporating a linear structure as in (7) can be done similarly to what is shown at the end of this section. It can also be proven [18], that the estimates of $\mathbf{A}$ and $\mathbf{B}$ obtained from (34) are guaranteed to be p.d. if $\hat{\mathbf{R}}$ is p.d.

In the following, it will be shown that the estimate $\hat{\boldsymbol{\theta}}$ obtained by minimizing

$$\begin{aligned} V(\boldsymbol{\theta}) &= \|\hat{\mathbf{R}} - \mathbf{A} \otimes \mathbf{B}\|_{\mathbf{Q}}^2 \\ &= \text{vec}^*\{\hat{\mathbf{R}} - \mathbf{A} \otimes \mathbf{B}\}\mathbf{Q}\text{vec}\{\hat{\mathbf{R}} - \mathbf{A} \otimes \mathbf{B}\} \end{aligned} \tag{35}$$

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

$$\mathbf{Q} = \frac{1}{N}(\text{Cov}[\text{vec}\{\hat{\mathbf{R}}\}])^{-1}. \tag{36}$$

This result is not very surprising, especially in the light of the *extended invariance principle* [12], [15]. It can be observed that the minimization problem in (33) coincides with (35) if $\mathbf{Q} = \mathbf{I}$. That choice would not yield an asymptotically efficient estimator, however. It is well known that the sample covariance matrix has a covariance given by

$$\begin{aligned} \text{Cov}[\text{vec}\{\hat{\mathbf{R}}\}] &= \frac{1}{N}\mathbf{R}_0^T \otimes \mathbf{R}_0 \\ &= \frac{1}{N}(\mathbf{A}_0 \otimes \mathbf{B}_0)^T \otimes (\mathbf{A}_0 \otimes \mathbf{B}_0). \end{aligned} \tag{37}$$

Note that (37) depends on the unknown parameters. It will be shown in Section VII that replacing $\mathbf{Q}$ with a consistent estimate

$$\hat{\mathbf{Q}} = \mathbf{Q} + o_p(1) \tag{38}$$

does not affect the asymptotic efficiency.

For a general $\hat{\mathbf{Q}}$, the minimization problem (35) lacks a simple, closed form solution and iterative methods similar to the flip-flop algorithm have to be used. Here, we will use a specially structured $\hat{\mathbf{Q}}$ for which the minimization problem in (35) can be solved in closed form. We suggest using the weighting matrix

$$\hat{\mathbf{Q}} = (\hat{\mathbf{A}}^{-1} \otimes \hat{\mathbf{B}}^{-1})^T \otimes (\hat{\mathbf{A}}^{-1} \otimes \hat{\mathbf{B}}^{-1}) \tag{39}$$

where $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ are selected such that

$$\hat{\mathbf{A}} \otimes \hat{\mathbf{B}} = \mathbf{R}_0 + o_p(1). \tag{40}$$

This condition is satisfied, e.g., by the closed form estimates given by the minimizers of (33). This choice also ensures positive definiteness of $\hat{\mathbf{Q}}$ when $\hat{\mathbf{R}}$ is p.d. [18]. With the choice of $\hat{\mathbf{Q}}$ in (39), (35) can be written as

$$\begin{aligned} &\|\hat{\mathbf{R}} - \mathbf{A} \otimes \mathbf{B}\|_{\hat{\mathbf{Q}}}^2 \\ &\quad = \|\check{\mathbf{R}} - (\hat{\mathbf{A}}^{-1/2}\mathbf{A}\hat{\mathbf{A}}^{-1/2}) \otimes (\hat{\mathbf{B}}^{-1/2}\mathbf{B}\hat{\mathbf{B}}^{-1/2})\|_{\mathbf{I}_{m^2n^2}}^2 \\ &\quad = \|R(\check{\mathbf{R}}) - \text{vec}\{\hat{\mathbf{A}}^{-1/2}\mathbf{A}\hat{\mathbf{A}}^{-1/2}\} \\ &\qquad \times \text{vec}^T\{\hat{\mathbf{B}}^{-1/2}\mathbf{B}\hat{\mathbf{B}}^{-1/2}\}\|_F^2 \end{aligned} \tag{41}$$

where

$$\check{\mathbf{R}} = (\hat{\mathbf{A}}^{-1/2} \otimes \hat{\mathbf{B}}^{-1/2})\hat{\mathbf{R}}(\hat{\mathbf{A}}^{-1/2} \otimes \hat{\mathbf{B}}^{-1/2}). \tag{42}$$

Next, using (7), we have that

$$\text{vec}\{\hat{\mathbf{A}}^{-1/2}\mathbf{A}\hat{\mathbf{A}}^{-1/2}\} = (\hat{\mathbf{A}}^{-T/2} \otimes \hat{\mathbf{A}}^{-1/2})\mathbf{P}_{\mathbf{A}}\boldsymbol{\theta}_{\mathbf{A}} \tag{43}$$

and

$$\text{vec}\{\hat{\mathbf{B}}^{-1/2}\mathbf{B}\hat{\mathbf{B}}^{-1/2}\} = (\hat{\mathbf{B}}^{-T/2} \otimes \hat{\mathbf{B}}^{-1/2})\mathbf{P}_{\mathbf{B}}\boldsymbol{\theta}_{\mathbf{B}}. \tag{44}$$

Hence, (41) can be expressed as

$$\|\hat{\mathbf{R}}-\mathbf{A}\otimes\mathbf{B}\|_{\mathbf{Q}}^2=\left\|R(\check{\mathbf{R}})-\mathbf{Q}_{\mathbf{A}}\mathbf{T}_{\mathbf{A}}\boldsymbol{\theta}_{\mathbf{A}}\boldsymbol{\theta}_{\mathbf{B}}^T\mathbf{T}_{\mathbf{B}}^T\mathbf{Q}_{\mathbf{B}}^T\right\|_F^2 \quad (45)$$

where $\mathbf{Q}_{\mathbf{A}}$ and $\mathbf{Q}_{\mathbf{B}}$ are matrices (with dimensions $m^2\times n_A$ and $n^2\times n_B$, respectively) with orthonormal columns and $\mathbf{T}_{\mathbf{A}}$ and $\mathbf{T}_{\mathbf{B}}$ are invertible matrices (with dimensions $n_A\times n_A$ and $n_B\times n_B$, respectively) such that

$$\begin{aligned}\mathbf{Q}_{\mathbf{A}}\mathbf{T}_{\mathbf{A}}&=(\hat{\mathbf{A}}^{-T/2}\otimes\hat{\mathbf{A}}^{-1/2})\mathbf{P}_{\mathbf{A}}\\ \mathbf{Q}_{\mathbf{B}}\mathbf{T}_{\mathbf{B}}&=(\hat{\mathbf{B}}^{-T/2}\otimes\hat{\mathbf{B}}^{-1/2})\mathbf{P}_{\mathbf{B}}.\end{aligned} \quad (46)$$

The criterion in (45) can be rewritten as

$$\begin{aligned}&\left\|\mathrm{vec}\{R(\check{\mathbf{R}})\}-(\mathbf{Q}_{\mathbf{B}}\otimes\mathbf{Q}_{\mathbf{A}})\mathrm{vec}\left\{\mathbf{T}_{\mathbf{A}}\boldsymbol{\theta}_{\mathbf{A}}\boldsymbol{\theta}_{\mathbf{B}}^T\mathbf{T}_{\mathbf{B}}^T\right\}\right\|_2^2\\ &\quad=\left\|(\mathbf{Q}_{\mathbf{B}}^*\otimes\mathbf{Q}_{\mathbf{A}}^*)\mathrm{vec}\{R(\check{\mathbf{R}})\}-\mathrm{vec}\left\{\mathbf{T}_{\mathbf{A}}\boldsymbol{\theta}_{\mathbf{A}}\boldsymbol{\theta}_{\mathbf{B}}^T\mathbf{T}_{\mathbf{B}}^T\right\}\right\|_2^2\\ &\quad=\left\|\mathbf{Q}_{\mathbf{A}}^*R(\check{\mathbf{R}})\mathbf{Q}_{\mathbf{B}}^c-\mathbf{T}_{\mathbf{A}}\boldsymbol{\theta}_{\mathbf{A}}\boldsymbol{\theta}_{\mathbf{B}}^T\mathbf{T}_{\mathbf{B}}^T\right\|_F^2.\end{aligned} \quad (47)$$

The rank-one approximation problem in (47) is easily solved using SVD. The proposed estimator has a fixed computational complexity similar to that of the unweighted ad hoc method, (33), and yet it achieves asymptotic efficiency as will be shown in Section VII. The performance for finite sample sizes will be evaluated using simulations in Section IX.

One remark is in place here. While the true covariance matrix is known to be p.d., this constraint is not imposed on the estimate obtained from (47). However, the sample covariance matrix is a consistent estimate of the true covariance matrix and it will be p.d. with probability one (w.p.1) as $N\to\infty$ due to the strong law of large numbers. This implies that the covariance matching estimate will also be p.d. w.p.1 as $N\to\infty$. The conclusion is that the asymptotic performance of the estimator is not affected by relaxing the positive definiteness constraint. See also [16].

VI. Cramér–Rao Lower Bound

The CRB gives a lower bound on the covariance matrix of any unbiased estimator. In this section the CRB will be derived for the estimation problem described in Section II.

The elements of the matrix $(\mathbf{A}\otimes\mathbf{B})$ are linear combinations of products of the form $[\boldsymbol{\theta}_{\mathbf{A}}]_i[\boldsymbol{\theta}_{\mathbf{B}}]_j$. Therefore, it is possible to construct a constant, parameter independent, $m^2n^2\times n_An_B-$matrix $\mathbf{P}$ such that

$$\mathrm{vec}\{\mathbf{A}\otimes\mathbf{B}\}=\mathbf{P}\boldsymbol{\eta},\quad \boldsymbol{\eta}=\mathrm{vec}\left\{\boldsymbol{\theta}_{\mathbf{A}}\boldsymbol{\theta}_{\mathbf{B}}^T\right\}. \quad (48)$$

In order to find an expression for $\mathbf{P}$, note that by definition

$$R(\mathbf{A}\otimes\mathbf{B})=\mathrm{vec}\{\mathbf{A}\}\mathrm{vec}^T\{\mathbf{B}\}=\mathbf{P}_{\mathbf{A}}\boldsymbol{\theta}_{\mathbf{A}}\boldsymbol{\theta}_{\mathbf{B}}^T\mathbf{P}_{\mathbf{B}}^T. \quad (49)$$

Then, by using the previously defined permutation matrix $\mathbf{P}_R$ it follows immediately that

$$\mathrm{vec}\{\mathbf{A}\otimes\mathbf{B}\}=\mathbf{P}_R(\mathbf{P}_{\mathbf{B}}\otimes\mathbf{P}_{\mathbf{A}})\mathrm{vec}\left\{\boldsymbol{\theta}_{\mathbf{A}}\boldsymbol{\theta}_{\mathbf{B}}^T\right\}. \quad (50)$$

This gives

$$\mathbf{P}=\mathbf{P}_R(\mathbf{P}_{\mathbf{B}}\otimes\mathbf{P}_{\mathbf{A}}). \quad (51)$$

The main result of this section can now be stated in the following theorem.

Theorem 3: The covariance matrix of any unbiased estimator $\hat{\mathbf{R}}_E$ of $\mathbf{R}_0$ in the data model described in Section II must satisfy

$$\begin{aligned}&N\mathrm{Cov}[\mathrm{vec}\{\hat{\mathbf{R}}_E\}]\\ &\qquad\geq\mathbf{P}\boldsymbol{\Gamma}_0\left(\boldsymbol{\Gamma}_0^T\mathbf{P}^*\left(\mathbf{R}_0^{-T}\otimes\mathbf{R}_0^{-1}\right)\mathbf{P}\boldsymbol{\Gamma}_0\right)^{\dagger}\boldsymbol{\Gamma}_0^T\mathbf{P}^*\end{aligned} \quad (52)$$

where $\boldsymbol{\Gamma}_0$ is given by

$$\boldsymbol{\Gamma}_0=\left(\boldsymbol{\theta}_{\mathbf{B}}\otimes\mathbf{I}_{n_A}\quad\mathbf{I}_{n_B}\otimes\boldsymbol{\theta}_{\mathbf{A}}\right). \quad (53)$$

evaluated for parameters $\boldsymbol{\theta}_{\mathbf{A}}$ and $\boldsymbol{\theta}_{\mathbf{B}}$ that give the true covariance matrix $\mathbf{R}_0$.

Proof: The i,jth element of the Fisher information matrix (FIM) is given by [7], [14]

$$[\mathbf{I}(\boldsymbol{\theta})]_{i,j}=N\mathrm{tr}\{\mathbf{R}^{-1}\dot{\mathbf{R}}_i\mathbf{R}^{-1}\dot{\mathbf{R}}_j\}. \quad (54)$$

It follows that

$$\begin{aligned}[\mathbf{I}(\boldsymbol{\theta})]_{i,j}&=N\mathrm{vec}^*\{\dot{\mathbf{R}}_i\}\left(\mathbf{R}^{-T}\otimes\mathbf{R}^{-1}\right)\mathrm{vec}\{\dot{\mathbf{R}}_j\}\\ &=N\dot{\boldsymbol{\eta}}_i^T\mathbf{P}^*\left(\mathbf{R}^{-T}\otimes\mathbf{R}^{-1}\right)\mathbf{P}\dot{\boldsymbol{\eta}}_j.\end{aligned} \quad (55)$$

Construct a matrix $\boldsymbol{\Gamma}$ such that

$$[\boldsymbol{\Gamma}]_{j,i}=\frac{\partial[\boldsymbol{\eta}]_j}{\partial[\boldsymbol{\theta}]_i} \quad (56)$$

which, when evaluated at $\boldsymbol{\theta}$, reads

$$\boldsymbol{\Gamma}=\left(\boldsymbol{\theta}_{\mathbf{B}}\otimes\mathbf{I}_{n_A}\quad\mathbf{I}_{n_B}\otimes\boldsymbol{\theta}_{\mathbf{A}}\right). \quad (57)$$

This immediately gives the following expression for the FIM

$$\mathbf{I}(\boldsymbol{\theta})=N\boldsymbol{\Gamma}^T\mathbf{P}^*(\mathbf{R}^{-T}\otimes\mathbf{R}^{-1})\mathbf{P}\boldsymbol{\Gamma}. \quad (58)$$

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 [see (8)] and possibly also due to the imposed linear structure. By using results proved in [13] we have that the desired CRB is given by

$$\boldsymbol{\Delta}\mathbf{I}^{\dagger}(\boldsymbol{\theta}_0)\boldsymbol{\Delta}^* \tag{59}$$

where column i of $\boldsymbol{\Delta}$ is given by

$$\left.\frac{\partial \mathrm{vec}\{\mathbf{R}\}}{\partial [\boldsymbol{\theta}]_i}\right|_{\mathbf{R}=\mathbf{R}_0}. \tag{60}$$

It is then straightforward to conclude that $\boldsymbol{\Delta} = \mathbf{P}\boldsymbol{\Gamma}_0$ and thus that the CRB is given by (52). The matrix $\boldsymbol{\Gamma}_0$ is equal to $\boldsymbol{\Gamma}$ evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. ■

VII. Asymptotic Performance of the Covariance Matching Estimator

The asymptotic performance of the estimator proposed in Section V is derived in this section. The result is summarized in the following theorem.

Theorem 4: 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 (35). Then, under the data model described in Section II, $\hat{\mathbf{R}}_C$ has an asymptotic complex Gaussian distribution with

$$\lim_{N\to\infty} N\mathrm{Cov}[\mathrm{vec}\{\hat{\mathbf{R}}_C\}] = \mathbf{P}\boldsymbol{\Gamma}_0 \left(\boldsymbol{\Gamma}_0^T\mathbf{P}^* \left(\mathbf{R}_0^{-T}\otimes\mathbf{R}_0^{-1}\right)\mathbf{P}\boldsymbol{\Gamma}_0\right)^{\dagger}\boldsymbol{\Gamma}_0^T\mathbf{P}^*. \tag{61}$$

Furthermore, (61) still holds if $\mathbf{Q}$ in (35) is replaced by any consistent estimate $\hat{\mathbf{Q}}$ of $\mathbf{Q}$ [e.g., (39)]. The matrices $\boldsymbol{\Gamma}_0$ and $\mathbf{P}$ are defined in Section VI. The estimate $\hat{\mathbf{R}}_C$ is a consistent estimate of $\mathbf{R}_0$ for any p.d. $\mathbf{Q}$ (or, for any consistent $\hat{\mathbf{Q}}$).

Proof: The consistency of the estimate is a direct consequence of the consistency of $\hat{\mathbf{R}}$. In order to derive the asymptotic covariance, first note that by using the definitions from the previous section and a Taylor series expansion one obtains

$$\mathrm{vec}\{\mathbf{R}_0 - \mathbf{A}\otimes\mathbf{B}\} = \mathbf{P}\boldsymbol{\Gamma}_0(\boldsymbol{\theta}_0 - \boldsymbol{\theta}) + O(\|\boldsymbol{\theta}_0 - \boldsymbol{\theta}\|^2). \tag{62}$$

Let $\hat{\boldsymbol{\theta}}$ be the minimizer of the covariance matching criterion function in (35). A Taylor series expansion of (35) gives

$$\mathbf{0} = \dot{V}(\hat{\boldsymbol{\theta}}) = \dot{V}(\boldsymbol{\theta}_0) - \mathbf{H}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}) + o_p(\|\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}\|) \tag{63}$$

where

$$[\dot{V}(\boldsymbol{\theta})]_i = \frac{\partial V(\boldsymbol{\theta})}{\partial [\boldsymbol{\theta}]_i} \tag{64}$$

and

$$[\mathbf{H}]_{i,j} = \lim_{N\to\infty} \left.\frac{\partial^2 V(\boldsymbol{\theta})}{\partial[\boldsymbol{\theta}]_i \partial[\boldsymbol{\theta}]_j}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}. \tag{65}$$

In order to derive an expression for $\mathbf{H}$, note that the criterion function (35) can be written

$$V(\boldsymbol{\theta}) = (\mathrm{vec}\{\hat{\mathbf{R}}\} - \mathbf{P}\boldsymbol{\eta})^*\mathbf{Q}(\mathrm{vec}\{\hat{\mathbf{R}}\} - \mathbf{P}\boldsymbol{\eta}) \tag{66}$$

where $\mathbf{P}$ and $\boldsymbol{\eta}$ were introduced in (51) and (48), respectively. From the consistency of the sample covariance matrix it follows that:

$$\lim_{N\to\infty} \mathrm{vec}\{\hat{\mathbf{R}}\} = \mathbf{P}\boldsymbol{\eta}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}. \tag{67}$$

Hence, we have that

$$[\mathbf{H}]_{i,j} = 2\boldsymbol{\eta}_i^*\mathbf{P}^*\mathbf{Q}\mathbf{P}\boldsymbol{\eta}_j\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}. \tag{68}$$

Making use of (56), this gives

$$\mathbf{H} = 2\boldsymbol{\Gamma}_0^T\mathbf{P}^*\mathbf{Q}\mathbf{P}\boldsymbol{\Gamma}_0 \tag{69}$$

where $\boldsymbol{\Gamma}_0$ is $\boldsymbol{\Gamma}$ evaluated at $\boldsymbol{\theta}_0$. Now, it follows from (63) that

$$\dot{V}(\boldsymbol{\theta}_0) = \mathbf{H}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}) + o_p(\|\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}\|) \tag{70}$$

and therefore

$$(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}) - \left[\mathbf{I}_{n_A+n_B} - \mathbf{H}^{\dagger}\mathbf{H}\right](\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}) = \mathbf{H}^{\dagger}\dot{V}(\boldsymbol{\theta}_0) + o_p(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|). \tag{71}$$

Note that the matrix within square brackets above is a projection matrix upon the null-space of $\mathbf{H}$. Thus, multiplying (71) from the left with $\mathbf{P}\boldsymbol{\Gamma}_0$ gives

$$\mathbf{P}\boldsymbol{\Gamma}_0(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}) = \mathbf{P}\boldsymbol{\Gamma}_0\mathbf{H}^{\dagger}\dot{V}(\boldsymbol{\theta}_0) + o_p(\|\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}\|). \tag{72}$$

This gives, making use of (62) and the consistency of the estimate

$$\lim_{N\to\infty} N\mathrm{Cov}[\mathrm{vec}\{\mathbf{R}_0 - \hat{\mathbf{A}}\otimes\hat{\mathbf{B}}\}] = \mathbf{P}\boldsymbol{\Gamma}_0\mathbf{H}^{\dagger}\left[\lim_{N\to\infty} N\mathrm{E}[\dot{V}(\boldsymbol{\theta}_0)\dot{V}^*(\boldsymbol{\theta}_0)]\right]\mathbf{H}^{\dagger}\boldsymbol{\Gamma}_0^T\mathbf{P}^*. \tag{73}$$

Finally, the relation

$$\begin{aligned}\lim_{N\to\infty} N\mathrm{E}[\dot{V}(\boldsymbol{\theta}_0)\dot{V}^*(\boldsymbol{\theta}_0)] &= 4\boldsymbol{\Gamma}_0^T\mathbf{P}^*\mathbf{Q}\left[\lim_{N\to\infty} N\mathrm{E}[\mathrm{vec}\{\hat{\mathbf{R}}-\mathbf{R}_0\}\mathrm{vec}^*\{\hat{\mathbf{R}}-\mathbf{R}_0\}]\right] \\ &\quad\times\mathbf{Q}\mathbf{P}\boldsymbol{\Gamma}_0 \\ &= 2\mathbf{H} \end{aligned} \tag{74}$$

shows that the asymptotic covariance matrix in (73) is given by (61). The first equality of (74) follows by making use of (66) and the last equality depends on (36). Also note that the asymptotic performance is unaffected by using a consistent estimate of the weighting matrix, such as $\hat{\mathbf{Q}}$, instead of the true $\mathbf{Q}$. The asymptotic normality of the estimate follows from the asymptotic normality of the elements of the sample covariance matrix; see, e.g., [1]. ■

Comparing Theorem 3 with Theorem 4 shows that the proposed covariance matching method is asymptotically efficient in the statistical sense.

VIII. Detection

The problem we study in this section is as follows: Given a set of data $[\mathbf{x}(t)]_{t=0}^{N-1}$, test whether the covariance matrix of $\mathbf{x}(t)$ is the Kronecker product of two matrices $\mathbf{A}$ and $\mathbf{B}$ with (given) dimensions m and n, and with linear structure of the type (7). This detection problem is thus closely related to the estimation problem treated in the previous sections.

In the MIMO communications example, m and n are known from the number of antennas on the transmitter and receiver, respectively, and the test is then used to accept or reject the Kronecker product structure of the covariance matrix, and possibly additional linear structure. In an application where the dimensions m and n are unknown, the test can be used to detect them by successively testing hypotheses with different values of m and n.

The generalized-likelihood ratio test (GLRT) has been previously proposed for the considered detection problem [10]. It can be shown (using standard theory) that twice the difference of the minimum values of the negative log-likelihood functions for the Kronecker structured model and the unstructured model, namely

$$\begin{aligned}\xi_{\mathrm{GLRT}} = 2N(&\log|(\hat{\mathbf{A}}_{\mathrm{ML}}\otimes\hat{\mathbf{B}}_{\mathrm{ML}})| \\ &+\mathrm{tr}\left\{\hat{\mathbf{R}}(\hat{\mathbf{A}}_{\mathrm{ML}}^{-1}\otimes\hat{\mathbf{B}}_{\mathrm{ML}}^{-1})\right\}-\log|\hat{\mathbf{R}}|-mn)\end{aligned} \tag{75}$$

has an asymptotic distribution given by

$$\xi_{\mathrm{GLRT}} \sim \chi^2(m^2n^2-m^2-n^2+1). \tag{76}$$

Thus, this quantity can be used to accept or reject the hypothesis that the covariance matrix for the data has a Kronecker product structure (with given dimensions of the Kronecker factors). With the results of Section IV in mind, the estimate $\hat{\mathbf{R}}_{FF}$ from (21) can be used in lieu of the true ML estimates, at least asymptotically, when calculating the GLRT test statistic. This makes it possible to construct the detection algorithm without using iterative estimation algorithms.

When structure of the kind (7) is to be included in the test, the GLRT can still be used. However, it is then necessary to calculate the ML estimates given such structure. In order to avoid the associated iterative search, it is natural to use the covariance matching estimator instead. The minimum value of the criterion function from Section V can then be used instead of the GLRT statistic for designing the detection algorithm. Note that this minimum value of the criterion function is given immediately by the SVD solution to the minimization of (47). The test statistic is thus readily available.

Establishing the statistical distribution of the minimum value of the criterion function is then an important aspect. This distribution is given in the following theorem.

Theorem 5: Let $\hat{\boldsymbol{\theta}}$ be the minimizer of the criterion function in (35). Then, under the data model described in Section II, the corresponding minimum function value is distributed asymptotically in N according to

$$NV(\hat{\boldsymbol{\theta}}) \sim \chi^2(m^2n^2-n_B-n_A+1) \tag{77}$$

provided that $m^2n^2-n_B-n_A+1>0$.

Proof: Using the notation introduced in the previous section, it follows from (70) that

$$\mathbf{H}\mathbf{H}^\dagger\dot{V}(\boldsymbol{\theta}_0) = \dot{V}(\boldsymbol{\theta}_0) + o_p(\|\boldsymbol{\theta}_0-\hat{\boldsymbol{\theta}}\|) \tag{78}$$

where $\hat{\boldsymbol{\theta}}$ is the minimizer of $V(\boldsymbol{\theta})$. A series expansion then gives

$$\begin{aligned} NV(\hat{\boldsymbol{\theta}}) &= NV(\boldsymbol{\theta}_0) - N\dot{V}^*(\boldsymbol{\theta}_0)(\boldsymbol{\theta}_0-\hat{\boldsymbol{\theta}}) \\ &\quad + N\frac{1}{2}(\boldsymbol{\theta}_0-\hat{\boldsymbol{\theta}})^*\mathbf{H}(\boldsymbol{\theta}_0-\hat{\boldsymbol{\theta}}) + o_p(N\|\boldsymbol{\theta}_0-\hat{\boldsymbol{\theta}}\|^2) \\ &= NV(\boldsymbol{\theta}_0) - N\left[\dot{V}^*(\boldsymbol{\theta}_0)\mathbf{H}^\dagger - \frac{1}{2}(\boldsymbol{\theta}_0-\hat{\boldsymbol{\theta}})^*\mathbf{H}\mathbf{H}^\dagger\right] \\ &\quad\times\mathbf{H}(\boldsymbol{\theta}_0-\hat{\boldsymbol{\theta}}) + o_p(N\|\boldsymbol{\theta}_0-\hat{\boldsymbol{\theta}}\|^2). \end{aligned} \tag{79}$$

Making use of (72), it can be shown that

$$\mathbf{H}(\boldsymbol{\theta}_0-\hat{\boldsymbol{\theta}}) = \mathbf{H}\mathbf{H}^\dagger\dot{V}(\boldsymbol{\theta}_0) + o_p(\|\boldsymbol{\theta}_0-\hat{\boldsymbol{\theta}}\|). \tag{80}$$

This gives

$$NV(\hat{\boldsymbol{\theta}}) = NV(\boldsymbol{\theta}_0) - N\frac{1}{2}\dot{V}^*(\boldsymbol{\theta}_0)\mathbf{H}^\dagger\dot{V}(\boldsymbol{\theta}_0) + o_p(1). \tag{81}$$

It is possible to construct an invertible $m^2n^2\times m^2n^2$ matrix $\mathbf{J}$ such that $\mathbf{J}\mathrm{vec}\{\mathbf{R}\}$ is real for any Hermitian matrix $\mathbf{R}$. Next note that

$$\begin{aligned}\dot{V}(\boldsymbol{\theta}_0) &= -2\frac{1}{N}\boldsymbol{\Gamma}_0^T\mathbf{P}^*\mathbf{Q}\mathrm{vec}\{\hat{\mathbf{R}}-\mathbf{R}_0\} \\ &= -2\frac{1}{N}\boldsymbol{\Gamma}_0^T\mathbf{P}^*\mathbf{J}^*\mathbf{J}^{-*}\mathbf{Q}\mathbf{J}^{-1}\mathbf{J}\mathrm{vec}\{\hat{\mathbf{R}}-\mathbf{R}_0\}. \end{aligned} \tag{82}$$

The real random vector $\mathbf{J}\text{vec}\{\hat{\mathbf{R}}-\mathbf{R}_0\}$ has the real covariance matrix

$$N\text{Cov}[\mathbf{J}\text{vec}\{\hat{\mathbf{R}}-\mathbf{R}_0\}]=\mathbf{J}\mathbf{Q}^{-1}\mathbf{J}^*. \tag{83}$$

It is therefore possible to introduce a real whitening matrix $\mathbf{S}$ such that

$$\mathbf{J}^{-*}\mathbf{Q}\mathbf{J}^{-1}=\mathbf{S}\mathbf{S}^T. \tag{84}$$

The gradient in (82) can then be written

$$\begin{aligned}\dot{V}(\boldsymbol{\theta}_0)&=-2\frac{1}{\sqrt{N}}\boldsymbol{\Gamma}_0^T\mathbf{P}^*\mathbf{J}^*\mathbf{S}\mathbf{g},\\ \mathbf{g}&=\sqrt{N}\mathbf{S}^T\mathbf{J}\text{vec}\{\hat{\mathbf{R}}-\mathbf{R}_0\}.\end{aligned} \tag{85}$$

Here, $\mathbf{g}$ is a zero-mean real Gaussian-distributed random vector with covariance $\mathbf{I}_{m^2n^2}$ for large N. The test statistic can now be written as

$$\begin{aligned}&NV(\hat{\boldsymbol{\theta}})\\ &\quad=\mathbf{g}^T\Big[\mathbf{I}_{m^2n^2}-\mathbf{S}^T\mathbf{J}\mathbf{P}\boldsymbol{\Gamma}_0\\ &\qquad\times\left(\boldsymbol{\Gamma}_0^T\mathbf{P}^*\mathbf{J}^*\mathbf{J}^{-*}\mathbf{Q}\mathbf{J}^{-1}\mathbf{J}\mathbf{P}\boldsymbol{\Gamma}_0\right)^{\dagger}\boldsymbol{\Gamma}_0^T\mathbf{P}^*\mathbf{J}^*\mathbf{S}\Big]\mathbf{g}\\ &\qquad+o_p(1).\end{aligned} \tag{86}$$

Next, note that the matrix within square brackets is idempotent with rank equal to $m^2n^2-\text{rank}\{\boldsymbol{\Gamma}_0\}$. Using a standard procedure, the rank of $\boldsymbol{\Gamma}_0$ can be found as follows:

$$\begin{aligned}\text{rank}\{\boldsymbol{\Gamma}_0\}&=\text{rank}\begin{bmatrix}\|\boldsymbol{\theta}_\mathbf{B}\|^2\mathbf{I}_{n_A} & \boldsymbol{\theta}_\mathbf{B}^T\otimes\boldsymbol{\theta}_\mathbf{A}\\ \boldsymbol{\theta}_\mathbf{B}\otimes\boldsymbol{\theta}_\mathbf{A}^T & \|\boldsymbol{\theta}_\mathbf{A}\|^2\mathbf{I}_{n_B}\end{bmatrix}\\ &=\text{rank}\{\|\boldsymbol{\theta}_\mathbf{B}\|^2\mathbf{I}_{n_A}\}\\ &\quad+\text{rank}\Big\{\|\boldsymbol{\theta}_\mathbf{A}\|^2\mathbf{I}_{n_B}\\ &\quad-\left(\boldsymbol{\theta}_\mathbf{B}\otimes\boldsymbol{\theta}_\mathbf{A}^T\right)\left(\boldsymbol{\theta}_\mathbf{B}^T\otimes\boldsymbol{\theta}_\mathbf{A}\right)\|\boldsymbol{\theta}_\mathbf{B}\|^{-2}\Big\}\\ &=n_A+n_B-1.\end{aligned} \tag{87}$$

The rank deficiency of $\boldsymbol{\Gamma}_0$ is due to the above-mentioned ambiguity in the parameterization. The matrix $\mathbf{JP}$ is real since any postmultiplication of it by a real vector gives a real result. It can thus be verified that the matrix within square brackets in (86) is real. Then, the stated result follows. ■

IX. Numerical Study

A. Estimation

Monte Carlo simulations were used to evaluate the small sample performance of the proposed estimators. Two matrices $\mathbf{A}_0$ and $\mathbf{B}_0$ were generated (and then fixed) and $\mathbf{R}_0$ was calculated. In each Monte Carlo trial, N 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 defined as

$$\sqrt{\frac{1}{L}\sum_{k=1}^{L}\frac{\|\mathbf{R}_0-\hat{\mathbf{R}}_k\|_F^2}{\|\mathbf{R}_0\|_F^2}} \tag{88}$$

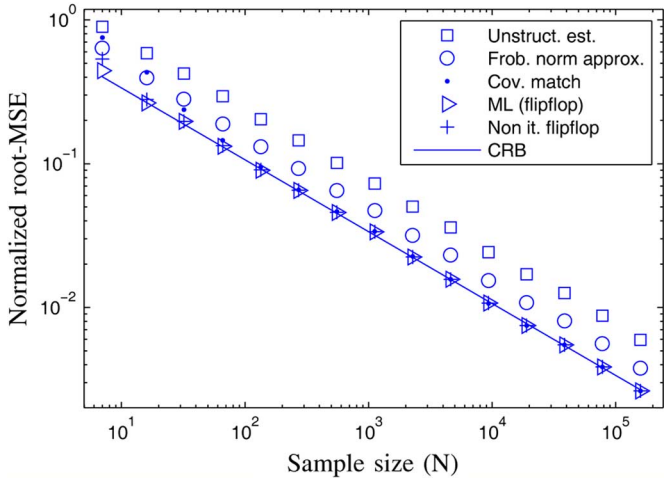

Fig. 1. Normalized root-MSE as a function of the sample size for five different estimators. Simulations consisting of 100 Monte Carlo runs were used. The figure shows the results of an experiment, where $\mathbf{A}_0$ and $\mathbf{B}_0$ are Hermitian but otherwise unstructured. The matrix dimensions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

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.

The resulting normalized root-MSE as a function of the sample size is shown in Fig. 1. In this example, the true matrices $\mathbf{A}_0$ and $\mathbf{B}_0$ were unstructured randomly generated p.d. matrices. This allows all considered methods to be used. The matrix dimensions used were $m=n=4$. 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 Frobenius norm approximation of the sample covariance matrix by a Kronecker structured matrix [see (33)]; iii) the proposed covariance matching method with the structured weighting matrix given by (39); iv) the ML method (implemented using the iterative flip-flop algorithm); and v) the proposed noniterative flip-flop method discussed in Section IV. In the presented results for this algorithm, the identity matrix was used for initialization, $\mathbf{A}_{\text{init}}=\mathbf{I}_m$.

After trying different initializations and different search methods, our conclusion based on numerical evidence is that the global minimum was found in general in the ML problem. A Newton search for the minimum gave exactly the same results in all experiments, regardless of initialization. For the noniterative flip-flop algorithm, it was shown in Section IV that the initialization does not affect the asymptotic 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 noniterative version of the flip-flop algorithm performs as well as the ML method. The covariance matching method proposed in Section V performs worse than the ML-based methods for the smaller sample sizes, but approaches the CRB for larger N. The unweighted approximation method, based on (33), does not reach the CRB (as expected).

In the example used to produce Fig. 1, the dimension of the sample covariance matrix $\hat{\mathbf{R}}$ is $mn\times mn=16\times 16$. This

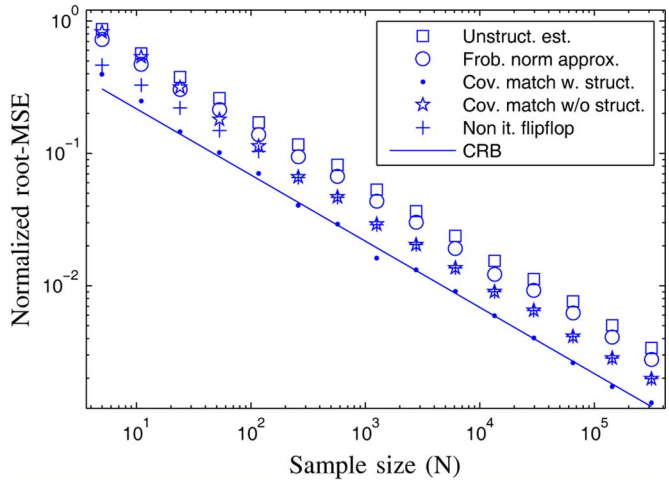

Fig. 2. Normalized root-MSE as a function of the sample size for five different estimators. Monte Carlo simulations were used with 100 Monte Carlo runs. The figure shows the results of an experiment were the $\mathbf{A}_0$ and $\mathbf{B}_0$ matrices were Toeplitz structured. The matrix dimensions used were $m = n = 3$.

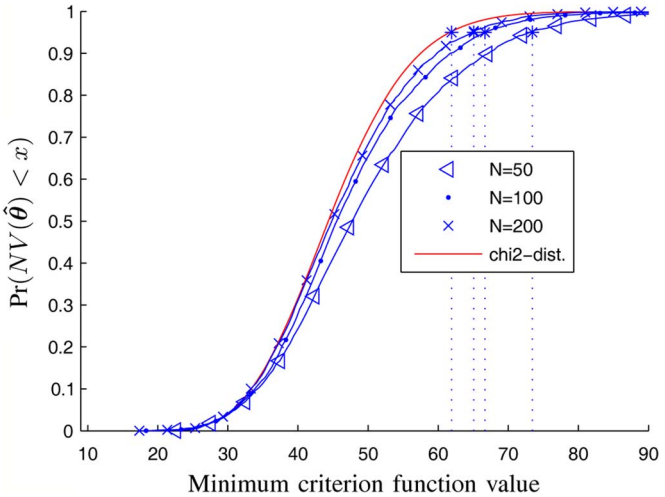

Fig. 3. Asymptotical (in N) theoretical cumulative distribution function (cdf) of the minimum criterion function value (unmarked line) and empirically estimated cdfs for $N = 50, 100, 200$. The detection thresholds corresponding to a 5% probability of incorrectly rejecting the null hypothesis are marked for each cdf. In this experiment $m = 4, n = 2$.

implies that $\hat{\mathbf{R}}$ is singular when $N < 16$. Thus, there is no guarantee that the estimates at each iteration of the two flip-flop algorithms $\mathbf{A}^i$ and $\mathbf{B}^i, i > 0$ are p.d. However, in the numerical experiments, they have always been p.d. (also for sample sizes as small as $N = 2$). The same observation applies to the matrices $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ that are used to form the weighting matrix $\hat{\mathbf{Q}}$ in the covariance matching method.

Fig. 2 shows the results of Monte Carlo simulations when a Toeplitz structure was imposed on the matrices $\mathbf{A}_0$ and $\mathbf{B}_0$. In the wireless communications application, this corresponds to using uniform linear arrays at both the receiver and the transmitter. In this example, the dimensions were set to $m = n = 3$. In this case, not all methods can take advantage of the structure. The included methods are as follows: i) the unstructured covariance estimate (that does not take the Kronecker structure into account); ii) the method based on Frobenius norm approximation (not taking the Toeplitz structure into account); iii) the proposed covariance matching method taking the full structure into account; iv) the same method but without taking the Toeplitz structure into account; and v) the noniterative flip-flop algorithm (not taking the Toeplitz structure into account).

The results presented in Fig. 2 confirm that making use of knowledge of the structure improves the estimation performance. As expected, the proposed method based on covariance matching outperforms the other methods since it is asymptotically efficient also in a scenario with additional structure. The figure also illustrates that, in order to achieve the CRB, it is necessary to use the correct weighting, and also to take the structure into account.

B. Detection

For the detection problem, an important question is for which sample sizes can the asymptotic result in Theorem 5 and the corresponding result for the GLRT test statistic (75) be used for calculating the detection thresholds. In order to compare the asymptotic results with empirical distributions, experiments

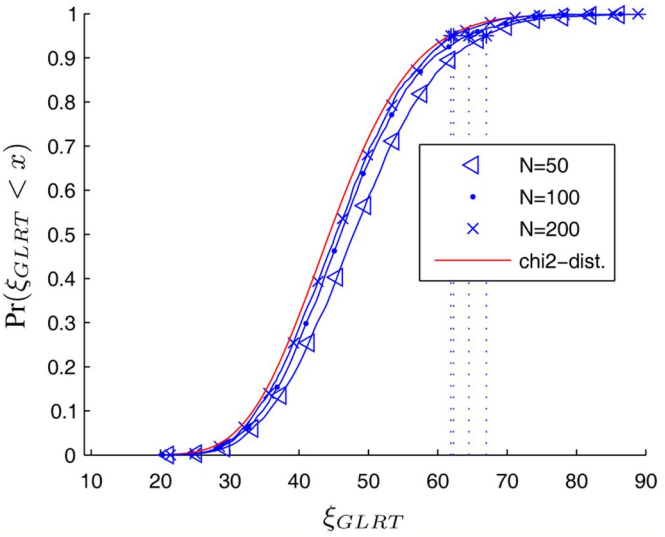

Fig. 4. Asymptotical (in N) theoretical cumulative distribution function (cdf) of the GLRT test statistic, ξ_{GLRT} (unmarked line) and empirically estimated cdfs for $N = 50, 100, 200$. The detection thresholds corresponding to a 5% probability of incorrectly rejecting the null hypothesis are marked for each cdf. In this experiment, $m = 4, n = 2$.

were conducted where a large number of test statistics were generated based on independent sample sets. The matrices $\mathbf{A}_0$ and $\mathbf{B}_0$ were randomly generated p.d. matrices with dimensions $m = 4, n = 2$. Different sample sizes were tested. Fig. 3 shows empirical cumulative distribution functions for the minimum value of the covariance matching criterion function together with the theoretical asymptotic result of Theorem 5. The corresponding results for the GLRT statistic in (75) are shown in Fig. 4. The match appears to be good for $N \geq 200$ for both methods.

Two detection algorithms are compared in Fig. 5. The true Kronecker factors $\mathbf{A}_0$ and $\mathbf{B}_0$ were unstructured with dimensions $m = n = 4$ (the chosen matrices were the same as those

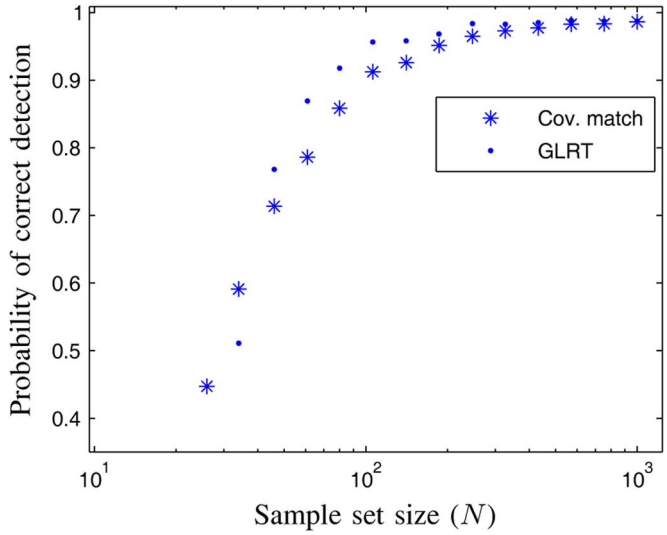

Fig. 5. Empirical probability of correct detection using 2000 independent realizations for each sample size. The probability of falsely rejecting the null hypothesis was set to $\alpha = 0.01$. The performance of the algorithm based on the minimum value of the covariance matching criterion function is compared to the performance of the algorithm based on the GLRT test statistic; both algorithms are implemented as described in Section IX.

used for the example in Fig. 1). The detection algorithms were implemented using the GLRT test statistic (75) and the minimum value of the covariance matching criterion function (35), respectively. The asymptotic distributions of these statistics are given in Section VIII.

The null hypothesis, that the underlying covariance matrix is a Kronecker product of matrices with dimensions $n = \bar{n}$ and $m = \bar{m}$, was accepted if

$$F_{\bar{m},\bar{n}}(\psi) < (1-\alpha) \tag{89}$$

where ψ is the test statistic (GLRT or covariance matching) calculated under the hypothesis and where $F_{m,n}(\psi)$ is the theoretical cumulative distribution function under the hypothesis (as derived in Section VIII). The parameter α, which is the probability of falsely rejecting the null hypothesis, was set to $\alpha = 0.01$. In this example, the hypotheses tested were (in the order tested): $\bar{m} = 2, \bar{n} = 8; \bar{m} = 4, \bar{n} = 4; \bar{m} = 8, \bar{n} = 2$; and $\bar{m} = 16, \bar{n} = 1$. In order to make a fair comparison, no extra linear structure of $\mathbf{A}$ and $\mathbf{B}$ was assumed. It is possible to perform the tests in a different order, which may give a different result.

The sample size was varied and the probability of correct detection (the detected m and n equal the true m and n) was estimated and plotted in Fig. 5 for both methods. It can be seen that both methods approach the desired 99% of correct detections as N increases, but the GLRT performs significantly better in small samples.

X. Conclusion

The problem of estimating a covariance matrix with Kronecker product structure from a set of N samples and that of detecting the dimensions and structure of the Kronecker factors have been treated. The focus has been on developing fast, noniterative methods. Two cases were considered: In the first case, 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 shown that the ML estimate can be computed by an iterative so-called flip-flop algorithm. It was shown in Section IV that a noniterative version of the flip-flop algorithm can be derived that is asymptotically efficient. In a numerical example, the proposed algorithm also showed a small sample performance that is fully comparable to that of ML. For the detection problem, it was natural to use the GLRT in this case. Simulations were used to investigate the performance of the GLRT; these results were presented in Section IX.

The second case is more general since it allows for linear structure of the Kronecker factors (as defined in Section II). If such a structure can be assumed, a different approach is needed. A method based on covariance matching was suggested. The proposed method is noniterative and also asymptotically efficient. The minimum value of the criterion function can be used as a statistic for a detection algorithm. The asymptotic distribution of this statistic was derived in Section VIII. Numerical evaluations of a detection procedure based on this quantity were presented in Section IX.

The Cramér–Rao lower bound for the considered estimation problem was derived in Section VI. Due to the asymptotic efficiency of the two proposed estimation methods, the resulting expression also gives their asymptotic covariance matrix. Expressions for the asymptotic performance of these methods were also derived more directly in Section IV for the methods based on the ML criterion and in Section V for the covariance matching method.

Appendix I
Proof of Theorem 1

In order to simplify the notation, note that (13) gives

$$\begin{aligned}\mathrm{vec}\{\hat{\mathbf{B}}(\mathbf{A})\} &= \frac{1}{m}\sum_{k=1}^{m}\sum_{l=1}^{m}\mathrm{vec}\{\hat{\mathbf{R}}^{kl}\}[\mathbf{A}^{-1}]_{lk} \\ &= \frac{1}{m}\hat{\mathbf{R}}_{\mathbf{B}}\mathrm{vec}\{\mathbf{A}^{-1}\}\end{aligned} \tag{90}$$

where $\hat{\mathbf{R}}_{\mathbf{B}}$ is defined in (25). Similarly

$$\mathrm{vec}\{\hat{\mathbf{A}}(\mathbf{B})\} = \frac{1}{n}\hat{\mathbf{R}}_{\mathbf{A}}\mathrm{vec}\{\mathbf{B}^{-1}\}$$

where $\hat{\mathbf{R}}_{\mathbf{A}}$ is defined in (26). By introducing

$$\begin{aligned}\mathbf{R}_{\mathbf{B}} &= \lim_{N\to\infty}\hat{\mathbf{R}}_{\mathbf{B}} = \mathrm{vec}\{\mathbf{B}_0\}\mathrm{vec}^T\{\mathbf{A}_0^T\} \quad \text{(w.p.1)} \\ \mathbf{R}_{\mathbf{A}} &= \lim_{N\to\infty}\hat{\mathbf{R}}_{\mathbf{A}} = \mathrm{vec}\{\mathbf{A}_0\}\mathrm{vec}^T\{\mathbf{B}_0^T\} \quad \text{(w.p.1)}\end{aligned} \tag{91}$$

it follows that (w.p.1)

$$\begin{aligned}\mathbf{B}_* &= \lim_{N\to\infty} \hat{\mathbf{B}}(\mathbf{A}_{\text{init}}) = \frac{\text{tr}\{\mathbf{A}_0(\mathbf{A}_{\text{init}})^{-1}\}}{m}\mathbf{B}_0 \\ \mathbf{A}_* &= \lim_{N\to\infty} \hat{\mathbf{A}}(\hat{\mathbf{B}}(\mathbf{A}_{\text{init}})) = \frac{m}{\text{tr}\{\mathbf{A}_0(\mathbf{A}_{\text{init}})^{-1}\}}\mathbf{A}_0 \\ &\lim_{N\to\infty} \hat{\mathbf{B}}(\hat{\mathbf{A}}(\hat{\mathbf{B}}(\mathbf{A}_{\text{init}}))) = \mathbf{B}_* \end{aligned} \tag{92}$$

where $\mathbf{B}_*$ and $\mathbf{A}_*$ are defined as the limiting estimates as shown above. Thus, $\hat{\mathbf{R}}_{FF}$ is a consistent estimate:

$$\lim_{N\to\infty} \hat{\mathbf{R}}_{FF} = \mathbf{R}_0 \quad \text{(w.p.1)}. \tag{93}$$

Now, consider a first-order perturbation analysis of $\hat{\mathbf{R}}_{FF}$. We proceed by investigating the error in the estimate of $\mathbf{B}$ after the initial iteration (step 2) in the algorithm outlined in Section III. To that end, define

$$\text{vec}\{\tilde{\mathbf{B}}^0\} = \text{vec}\{\hat{\mathbf{B}}(\mathbf{A}_{\text{init}}) - \mathbf{B}_*\} = \frac{1}{m}\tilde{\mathbf{R}}_{\mathbf{B}}\text{vec}\{(\mathbf{A}_{\text{init}})^{-1}\} \tag{94}$$

where

$$\tilde{\mathbf{R}}_{\mathbf{B}} = \hat{\mathbf{R}}_{\mathbf{B}} - \mathbf{R}_{\mathbf{B}}. \tag{95}$$

In order to proceed, note that the matrix inversion lemma gives

$$\begin{aligned}(\mathbf{B}_* + \tilde{\mathbf{B}}^0)^{-1} &= \mathbf{B}_*^{-1} - \mathbf{B}_*^{-1}\tilde{\mathbf{B}}^0(\mathbf{I} + \mathbf{B}_*^{-1}\tilde{\mathbf{B}}^0)^{-1}\mathbf{B}_*^{-1} \\ &\simeq \mathbf{B}_*^{-1} - \mathbf{B}_*^{-1}\tilde{\mathbf{B}}^0\mathbf{B}_*^{-1}\end{aligned} \tag{96}$$

where the symbol $\simeq$ denotes a first-order approximate equality where terms that have higher order effects on the asymptotics (in N) compared to the retained terms are removed:

$$\mathbf{X} \simeq \mathbf{Y} \to \mathbf{X} = \mathbf{Y} + o_p(\|\hat{\mathbf{R}} - \mathbf{R}_0\|_2). \tag{97}$$

In analogy with (94), define the error in the $\mathbf{A}$ estimate after the first iteration (step 3) as

$$\begin{aligned}\text{vec}\{\tilde{\mathbf{A}}^1\} &= \text{vec}\{\hat{\mathbf{A}}(\hat{\mathbf{B}}(\mathbf{A}_{\text{init}})) - \mathbf{A}_*\} \\ &= \frac{1}{n}\hat{\mathbf{R}}_{\mathbf{A}}\text{vec}\{(\mathbf{B}_* + \tilde{\mathbf{B}}^0)^{-1}\} - \text{vec}\{\mathbf{A}_*\} \\ &\simeq \frac{1}{n}\tilde{\mathbf{R}}_{\mathbf{A}}\text{vec}\{\mathbf{B}_*^{-1}\} - \frac{1}{n}\mathbf{R}_{\mathbf{A}}\text{vec}\{\mathbf{B}_*^{-1}\tilde{\mathbf{B}}^0\mathbf{B}_*^{-1}\}\end{aligned} \tag{98}$$

where use was made of (96) in the last step and where the definition

$$\tilde{\mathbf{R}}_{\mathbf{A}} = \hat{\mathbf{R}}_{\mathbf{A}} - \mathbf{R}_{\mathbf{B}} \tag{99}$$

was introduced similar to $\tilde{\mathbf{R}}_{\mathbf{B}}$ in (95). By inserting (94) into (98), we obtain

$$\begin{aligned}\text{vec}\{\tilde{\mathbf{A}}^1\} \simeq{}& \frac{1}{n}\tilde{\mathbf{R}}_{\mathbf{A}}\text{vec}\{\mathbf{B}_*^{-1}\} \\ &-\frac{1}{mn}\mathbf{R}_{\mathbf{A}}(\mathbf{B}_*^{-T}\otimes\mathbf{B}_*^{-1})\tilde{\mathbf{R}}_{\mathbf{B}}\text{vec}\{(\mathbf{A}_{\text{init}})^{-1}\}.\end{aligned} \tag{100}$$

Next, define the error in the estimate of $\mathbf{B}$ at step 4 as

$$\begin{aligned}&\text{vec}\{\tilde{\mathbf{B}}^2\} \\ &\quad= \text{vec}\{\hat{\mathbf{B}}(\hat{\mathbf{A}}(\hat{\mathbf{B}}(\mathbf{A}_{\text{init}}))) - \mathbf{B}_*\} \\ &\quad= \hat{\mathbf{R}}_{\mathbf{B}}\text{vec}\{(\tilde{\mathbf{A}}^1 + \mathbf{A}_*)^{-1}\} - \text{vec}\{\mathbf{B}_*\} \\ &\quad\simeq \frac{1}{m}\tilde{\mathbf{R}}_{\mathbf{B}}\text{vec}\{\mathbf{A}_*^{-1}\} - \frac{1}{m}\mathbf{R}_{\mathbf{B}}\text{vec}\{\mathbf{A}_*^{-1}\tilde{\mathbf{A}}^1\mathbf{A}_*^{-1}\}.\end{aligned} \tag{101}$$

The second step was obtained similar to (98). Using (100) in (101) then gives

$$\begin{aligned}&\text{vec}\{\tilde{\mathbf{B}}^2\} \\ &\quad\simeq \frac{1}{m}\tilde{\mathbf{R}}_{\mathbf{B}}\text{vec}\{\mathbf{A}_*^{-1}\} \\ &\qquad - \frac{1}{mn}\mathbf{R}_{\mathbf{B}}(\mathbf{A}_*^{-T}\otimes\mathbf{A}_*^{-1})\tilde{\mathbf{R}}_{\mathbf{A}}\text{vec}\{\mathbf{B}_*^{-1}\} \\ &\qquad + \frac{1}{m^2n}\mathbf{R}_{\mathbf{B}}(\mathbf{A}_*^{-T}\otimes\mathbf{A}_*^{-1})\mathbf{R}_{\mathbf{A}} \\ &\qquad \times (\mathbf{B}_*^{-T}\otimes\mathbf{B}_*^{-1})\tilde{\mathbf{R}}_{\mathbf{B}}\text{vec}\{(\mathbf{A}_{\text{init}})^{-1}\}.\end{aligned} \tag{102}$$

In order to relate these results to the error in $\hat{\mathbf{R}}_{FF}$, write

$$(\hat{\mathbf{R}}_{FF} - \mathbf{R}_0) \simeq (\tilde{\mathbf{A}}^1\otimes\mathbf{B}_*) + (\mathbf{A}_*\otimes\tilde{\mathbf{B}}^2). \tag{103}$$

At this stage, recall the rearrangement function defined in (22). Applying it to (103) gives

$$\begin{aligned}&R(\hat{\mathbf{R}}_{FF} - \mathbf{R}_0) \\ &\quad\simeq \text{vec}\{\tilde{\mathbf{A}}^1\}\text{vec}^T\{\mathbf{B}_*\} + \text{vec}\{\mathbf{A}_*\}\text{vec}^T\{\tilde{\mathbf{B}}^2\}.\end{aligned} \tag{104}$$

Inserting (100) and (102) in (104) yields

$$\begin{aligned}&\text{vec}\{R(\hat{\mathbf{R}}_{FF} - \mathbf{R}_0)\} \\ &\quad\simeq \frac{1}{n}(\text{vec}\{\mathbf{B}_*\}\text{vec}^T\{\mathbf{B}_*^{-1}\}\otimes\mathbf{I}_{m^2})\text{vec}\{\tilde{\mathbf{R}}_{\mathbf{A}}\} \\ &\qquad - \frac{1}{mn}(\text{vec}\{\mathbf{B}_*\}\text{vec}^T\{(\mathbf{A}_{\text{init}})^{-1}\} \\ &\qquad \otimes \mathbf{R}_{\mathbf{A}}(\mathbf{B}_*^{-T}\otimes\mathbf{B}_*^{-1}))\text{vec}\{\tilde{\mathbf{R}}_{\mathbf{B}}\} \\ &\qquad + \frac{1}{m}(\mathbf{I}_{n^2}\otimes\text{vec}\{\mathbf{A}_*\}\text{vec}^T\{\mathbf{A}_*^{-1}\})\text{vec}\left\{\tilde{\mathbf{R}}_{\mathbf{B}}^T\right\} \\ &\qquad - \frac{1}{mn}(\mathbf{R}_{\mathbf{B}}(\mathbf{A}_*^{-T}\otimes\mathbf{A}_*^{-1}) \\ &\qquad \otimes \text{vec}\{\mathbf{A}_*\}\text{vec}^T\{\mathbf{B}_*^{-1}\})\text{vec}\left\{\tilde{\mathbf{R}}_{\mathbf{A}}^T\right\} \\ &\qquad + \frac{1}{nm^2}(\mathbf{R}_{\mathbf{B}}(\mathbf{A}_*^{-T}\otimes\mathbf{A}_*^{-1})\mathbf{R}_{\mathbf{A}}(\mathbf{B}_*^{-T}\otimes\mathbf{B}_*^{-1}) \\ &\qquad \otimes \text{vec}\{\mathbf{A}_*\}\text{vec}^T\{(\mathbf{A}_{\text{init}})^{-1}\})\text{vec}\left\{\tilde{\mathbf{R}}_{\mathbf{B}}^T\right\}.\end{aligned} \tag{105}$$

Next, note that

$$\begin{aligned}\mathbf{R}_{\mathbf{B}}(\mathbf{A}_*^{-T}\otimes\mathbf{A}_*^{-1}) &= \text{vec}\{\mathbf{B}_*\}\text{vec}^T\{\mathbf{A}_*^{-T}\} \\ \mathbf{R}_{\mathbf{A}}(\mathbf{B}_*^{-T}\otimes\mathbf{B}_*^{-1}) &= \text{vec}\{\mathbf{A}_*\}\text{vec}^T\{\mathbf{B}_*^{-T}\}.\end{aligned} \tag{106}$$

The last term in (105) can then be simplified to

$$\begin{aligned}&\frac{\text{tr}\{\mathbf{A}_*\mathbf{A}_*^{-1}\}}{nm^2}(\text{vec}\{\mathbf{B}_*\}\text{vec}^T\{\mathbf{B}_*^{-T}\} \\ &\qquad \otimes\text{vec}\{\mathbf{A}_*\}\text{vec}^T\{(\mathbf{A}_{\text{init}})^{-1}\})\text{vec}\left\{\tilde{\mathbf{R}}_{\mathbf{B}}^T\right\}.\end{aligned} \tag{107}$$

This shows that the second and the last terms of (105) cancel each other. Thus

$$\mathrm{vec}\{\hat{\mathbf{R}}_{FF}-\mathbf{R}_0\}\simeq\boldsymbol{\Xi}\mathrm{vec}\{\tilde{\mathbf{R}}\} \tag{108}$$

where $\boldsymbol{\Xi}$ is given by (29).

Finally, making use of the standard result (see, e.g., [12])

$$\mathrm{Cov}[\mathrm{vec}\{\tilde{\mathbf{R}}\}]=\frac{1}{N}\mathbf{R}_0^T\otimes\mathbf{R}_0 \tag{109}$$

leads to the conclusion that the asymptotic covariance of the noniterative flip-flop estimate is given by

$$\lim_{N\to\infty}N\mathrm{Cov}[\mathrm{vec}\{\hat{\mathbf{R}}_{FF}\}]=\boldsymbol{\Xi}(\mathbf{R}_0^T\otimes\mathbf{R}_0)\boldsymbol{\Xi}^*. \tag{110}$$

The asymptotic normality of the estimate follows from the asymptotic normality of the elements of the sample covariance matrix; see, e.g., [1]. This concludes the proof.

Appendix II
Proof of Theorem 2

The proof will be along the same lines as the proof of Theorem 1. First assume that $\hat{\mathbf{B}}_{\mathrm{ML}}$ is such that

$$\hat{\mathbf{R}}_{\mathrm{ML}}=\hat{\mathbf{A}}_{\mathrm{ML}}\otimes\hat{\mathbf{B}}_{\mathrm{ML}} \tag{111}$$

where $\hat{\mathbf{R}}_{\mathrm{ML}}$ is the ML estimate of the sought covariance. This implies that, with the definition (18) in mind

$$\hat{\mathbf{A}}_{\mathrm{ML}}=\hat{\mathbf{A}}(\hat{\mathbf{B}}_{\mathrm{ML}}). \tag{112}$$

Also define [analogously to (92)]

$$\begin{aligned}\mathbf{B}_*^{\mathrm{ML}}&=\lim_{N\to\infty}\hat{\mathbf{B}}_{\mathrm{ML}}=\gamma\mathbf{B}_*\\ \mathbf{A}_*^{\mathrm{ML}}&=\lim_{N\to\infty}\hat{\mathbf{A}}_{\mathrm{ML}}=\gamma^{-1}\mathbf{A}_*\end{aligned} \tag{113}$$

for some constant γ. This allows us to define

$$\tilde{\mathbf{B}}_{\mathrm{ML}}=\hat{\mathbf{B}}_{\mathrm{ML}}-\mathbf{B}_*^{\mathrm{ML}}. \tag{114}$$

Now, define, similarly to (98), the error in $\hat{\mathbf{A}}_{\mathrm{ML}}$ as

$$\begin{aligned}&\mathrm{vec}\{\tilde{\mathbf{A}}_{\mathrm{ML}}\}\\ &\quad=\mathrm{vec}\{\hat{\mathbf{A}}_{\mathrm{ML}}-\mathbf{A}_*^{\mathrm{ML}}\}\\ &\quad=\frac{1}{n}\hat{\mathbf{R}}_{\mathbf{A}}\mathrm{vec}\{(\hat{\mathbf{B}}_{\mathrm{ML}})^{-1}\}-\mathrm{vec}\{\mathbf{A}_*^{\mathrm{ML}}\}\\ &\quad\simeq\frac{1}{n}\tilde{\mathbf{R}}_{\mathbf{A}}\mathrm{vec}\{(\mathbf{B}_*^{\mathrm{ML}})^{-1}\}\\ &\qquad-\frac{1}{n}\mathrm{vec}\{\mathbf{A}_*^{\mathrm{ML}}\}\mathrm{vec}^T\{(\mathbf{B}_*^{\mathrm{ML}})^{-T}\}\mathrm{vec}\{\tilde{\mathbf{B}}_{\mathrm{ML}}\}\end{aligned} \tag{115}$$

where use was made of (106). In the same way, note that since $\mathbf{B}_*^{\mathrm{ML}}=\hat{\mathbf{B}}(\mathbf{A}_*^{\mathrm{ML}})$

$$\begin{aligned}&\mathrm{vec}\{\tilde{\mathbf{B}}_{\mathrm{ML}}\}\\ &\quad=\mathrm{vec}\{\hat{\mathbf{B}}(\tilde{\mathbf{A}}^{\mathrm{ML}}+\mathbf{A}_*^{\mathrm{ML}})-\mathbf{B}_*^{\mathrm{ML}}\}\\ &\quad=\frac{1}{m}\hat{\mathbf{R}}_{\mathbf{B}}\mathrm{vec}\{(\tilde{\mathbf{A}}^{\mathrm{ML}}+\mathbf{A}_*^{\mathrm{ML}})^{-1}\}-\mathrm{vec}\{\mathbf{B}_*^{\mathrm{ML}}\}\\ &\quad\simeq\frac{1}{m}\tilde{\mathbf{R}}_{\mathbf{B}}\mathrm{vec}\{(\mathbf{A}_*^{\mathrm{ML}})^{-1}\}\\ &\qquad-\frac{1}{mn}\mathrm{vec}\{\mathbf{B}_*^{\mathrm{ML}}\}\mathrm{vec}^T\{(\mathbf{A}_*^{\mathrm{ML}})^{-T}\}\\ &\qquad\times\tilde{\mathbf{R}}_{\mathbf{A}}\mathrm{vec}\{(\mathbf{B}_*^{\mathrm{ML}})^{-1}\}\\ &\qquad+\frac{1}{n}\mathrm{vec}\{\mathbf{B}_*^{\mathrm{ML}}\}\mathrm{vec}^T\{(\mathbf{B}_*^{\mathrm{ML}})^{-T}\}\mathrm{vec}\{\tilde{\mathbf{B}}_{\mathrm{ML}}\}.\end{aligned} \tag{116}$$

The last term was simplified using

$$\mathrm{vec}^T\{(\mathbf{A}_*^{\mathrm{ML}})^{-T}\}\mathrm{vec}\{\mathbf{A}_*^{\mathrm{ML}}\}=m. \tag{117}$$

Now, similar to (104), write

$$\begin{aligned}&R(\hat{\mathbf{R}}_{\mathrm{ML}}-\mathbf{R}_0)\\ &\quad\simeq\mathrm{vec}\{\tilde{\mathbf{A}}_{\mathrm{ML}}\}\mathrm{vec}^T\{\mathbf{B}_*^{\mathrm{ML}}\}\\ &\qquad+\mathrm{vec}\{\mathbf{A}_*^{\mathrm{ML}}\}\mathrm{vec}^T\{\tilde{\mathbf{B}}_{\mathrm{ML}}\}\\ &\quad\simeq\frac{1}{n}\tilde{\mathbf{R}}_{\mathbf{A}}\mathrm{vec}\{\mathbf{B}_*^{-1}\}\mathrm{vec}^T\{\mathbf{B}_*\}\\ &\qquad-\frac{1}{n}\mathrm{vec}\{\mathbf{A}_*^{\mathrm{ML}}\}\mathrm{vec}^T\{(\mathbf{B}_*^{\mathrm{ML}})^{-T}\}\\ &\qquad\times\mathrm{vec}\{\tilde{\mathbf{B}}_{\mathrm{ML}}\}\mathrm{vec}^T\{\mathbf{B}_*^{\mathrm{ML}}\}\\ &\qquad+\frac{1}{m}\mathrm{vec}\{\mathbf{A}_*\}\mathrm{vec}^T\{\mathbf{A}_*^{-1}\}\tilde{\mathbf{R}}_{\mathbf{B}}^T\\ &\qquad-\frac{1}{mn}\mathrm{vec}\{\mathbf{A}_*\}\mathrm{vec}^T\{\mathbf{B}_*^{-1}\}\\ &\qquad\times\tilde{\mathbf{R}}_{\mathbf{A}}^T\mathrm{vec}\{\mathbf{A}_*^{-T}\}\mathrm{vec}^T\{\mathbf{B}_*\}\\ &\qquad+\frac{1}{n}\mathrm{vec}\{\mathbf{A}_*^{\mathrm{ML}}\}\mathrm{vec}^T\{\tilde{\mathbf{B}}_{\mathrm{ML}}\}\\ &\qquad\times\mathrm{vec}\{(\mathbf{B}_*^{\mathrm{ML}})^{-T}\}\mathrm{vec}^T\{\mathbf{B}_*^{\mathrm{ML}}\}.\end{aligned} \tag{118}$$

Here, the second and last terms cancel each other and after vectorizing (precisely in parallel to the proof of Theorem 1), we have that

$$\mathrm{vec}\{\hat{\mathbf{R}}_{\mathrm{ML}}-\mathbf{R}_0\}\simeq\boldsymbol{\Xi}\mathrm{vec}\{\tilde{\mathbf{R}}\} \tag{119}$$

with $\boldsymbol{\Xi}$ given by (29). The asymptotic normality follows similar to the proof of Theorem 1. By combining (109) and (119), the result of Theorem 2 follows.

Acknowledgment

The authors would like to thank the reviewers for their many constructive comments that helped improving the presentation of the paper.

REFERENCES

[1] T. W. Andersson, *An Introduction to Multivariate Statistical Analysis*. New York: Wiley, 1958.

[2] T. A. Barton and D. R. Fuhrmann, "Covariance structures for multidimensional data," *Multidimension. Syst. Signal Process.*, vol. V4, no. 2, pp. 111–123, Apr. 1993.

[3] M. Bengtsson and P. Zetterberg, "Some notes on the Kronecker model," *EURASIP J. Wireless Commun. Netw.*, Apr. 2006, submitted for publication.

[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 Trans. Signal Process.*, vol. 50, no. 7, pp. 1565–1572, Jul. 2002.

[5] D. R. Fuhrmann, "Application of Toeplitz covariance estimation to adaptive beamforming and detection," *IEEE Trans. Signal Process.*, vol. 39, no. 10, pp. 2194–2198, Oct. 1991.

[6] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1991.

[7] S. Kay, *Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice-Hall, 1993.

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

[9] N. Lu and D. Zimmerman, "On likelihood-based inference for a separable covariance matrix," Statistics and Actuarial Science Dept., Univ. of Iowa, Iowa City, IA, Tech. Rep. 337, 2004.

[10] N. Lu and D. Zimmerman, "The likelihood ratio test for a separable covariance matrix," *Statist. Probab. Lett.*, vol. 73, no. 5, pp. 449–457, May 2005.

[11] K. Mardia and C. Goodall, "Spatial-temporal analysis of multivariate environmental data," in *Multivariate Environmental Statistics*. Amsterdam, The Netherlands: Elsevier, 1993, pp. 347–386.

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

[13] P. Stoica and T. Marzetta, "Parameter estimation problems with singular information matrices," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 87–90, Jan. 2001.

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

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

[16] P. Stoica, B. Ottersten, M. Viberg, and R. L. Moses, "Maximum likelihood array processing for stochastic coherent sources," *IEEE Trans. Signal Process.*, vol. 44, no. 1, pp. 96–105, Jan. 1996.

[17] P. Strobach, "Low-rank detection of multichannel Gaussian signals using block matrix approximation," *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 233–242, Jan. 1995.

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

[19] K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, and P. Karlsson, "Modeling of wideband MIMO radio channels based on NLoS indoor measurements," *IEEE Trans. Veh. Technol.*, vol. 53, no. 8, pp. 655–665, May 2004.

Karl Werner (S'03) was born in Halmstad, Sweden, on November 26, 1978. In 2001–2002, he was an undergraduate student at the University of Waterloo, ON, Canada. He received the M.Sc. degree in computer engineering from Lund University, Lund, Sweden, in 2002 and the Tech. Lic. and the Ph.D. degrees from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 2005 and 2007, respectively.

Currently, he is at Ericsson AB, Stockholm, Sweden. His research interests include statistical signal processing, time series analysis, and estimation theory.

Magnus Jansson (S'93–M'98) was born in Enköping, Sweden, in 1968. He received the M.S., Tech. Lic., and Ph.D. degrees in electrical engineering from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 1992, 1995, and 1997, respectively.

In January 2002, he was appointed Docent in Signal Processing at KTH. Since 1993, he has held various research positions at the Department of Signals, Sensors, and Systems, KTH, where he currently is an Associate Professor. From 1998 to 1999, he visited the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis. His research interests include statistical signal processing, time series analysis, and system identification.

Peter Stoica (F'94) is a Professor of Systems Modelling at Uppsala University, Uppsala, Sweden. More details about him can be found at http://user.it.uu.se/~ps/ps.html