15th European Signal Processing Conference (EUSIPCO 2007), Poznan, Poland, September 3-7, 2007, copyright by EURASIP

ESTIMATION OF KRONECKER STRUCTURED CHANNEL COVARIANCES USING TRAINING DATA

Karl Werner and Magnus Jansson

School of Electrical Engineering, KTH
SE-100 44, Stockholm, Sweden
email: firstname.lastname@ee.kth.se

ABSTRACT

The problem of estimating second order statistics for MIMO channels is treated. It is assumed that the so called Kronecker-model holds. This implies that the channel covariance is the Kronecker product of two covariance matrices associated with the transmit and receive array, respectively. The proposed estimator uses training data from a number of signal blocks to compute the estimate. This is in contrast to methods that assume that the channel realizations are directly available, or possible to estimate almost without error. It is also demonstrated how methods that make use of the training data indirectly via channel estimates can be biased.

An estimator is derived that can, in an asymptotically optimal way, use, not only the structure implied by the Kronecker assumption, but also linear structure on the transmit- and receive covariance matrices. The performance of the proposed estimator is analyzed and numerical simulations illustrate the results and also provide insight into the small sample behavior of the proposed method.

I. INTRODUCTION

In the modelling of frequency flat multiple input multiple output (MIMO) channels, Kronecker structured channel covariance matrices are often assumed [6], [13], [2]. The underlying assumption is that, while the channel changes between the signal blocks, the second order statistics of the channel are valid over a longer period of time. The Kronecker model also assumes that the channel covariance is structured according to

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

where $\mathbf{H}_t$ is the stochastic $m \times n$ channel matrix, $\otimes$ denotes Kronecker matrix product, $\mathrm{vec}\{\}$ denotes the vectorization operator (see, e.g., [4]), $\mathbf{A}$ is an $n \times n$ transmit covariance matrix, and $\mathbf{B}$ is an $m \times m$ receive covariance matrix. Estimating such covariance 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.

In a typical communication system, the receiver estimates the channel statistics based on training data received in a number of signal blocks. If the amount of training data available in each block is very large, or the signal to noise ratio (SNR) is high, then the channel estimates can be assumed to be identical to the true underlying channel and methods such as those of [12] or [11] can be used to calculate the covariance matrix estimate.

If the channel estimates cannot be assumed perfect, this should be taken into account in the design of the estimator. The input to the estimator should be the training data, not the channel estimates. In this work we present a new method for the estimation problem based on a covariance matching criterion. The method is non-iterative and thus has a fixed computational complexity. It is also asymptotically statistically efficient. Furthermore, it allows for linearly structured $\mathbf{A}$ and $\mathbf{B}$ matrices. Such structure can, e.g., be due to certain array geometries.

In the following, $\mathbf{X}^{\dagger}$ denotes the Moore-Penrose pseudo-inverse 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 $\mathrm{conj}\{\mathbf{X}\} = \mathbf{X}^{T*}$. 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. Finally, the notation $x_N = o_p(a_N)$ means that $\lim_{N\to\infty} x_N/a_N = 0$ in probability. In this work the asymptotic results hold when the number of signal blocks, N, tends to infinity.

II. DATA MODEL

Denote the MIMO channel during signal block t by $\mathbf{H}_t$. Assume that the same sequence of p training symbols $[\mathbf{x}(i)]_{i=0}^{p-1}$ is sent once as part of each signal block. The corresponding received data in signal block t can then be modelled as

$$\mathbf{y}_t(i) = \mathbf{H}_t\mathbf{x}(i) + \mathbf{e}_t(i),\ i = 0, \dots, p-1. \tag{2}$$

The noise $\mathbf{e}_t(i)$ is assumed to be zero mean complex Gaussian with covariance matrix

$$\mathrm{E}\left[\mathbf{e}_s(k)\mathbf{e}_t^*(l)\right] = \sigma_0^2 \mathbf{I}_m \delta_{s,t}\delta_{k,l}. \tag{3}$$

The noise variance σ_0^2 is assumed unknown. Assume that the vectorized MIMO channel, $\mathrm{vec}\{\mathbf{H}_t\}$, in each signal block is a realization of a complex Gaussian, zero mean, vector valued random variable that is uncorrelated with the noise and has covariance

$$\mathrm{E}\left[\mathrm{vec}\{\mathbf{H}_k\}\mathrm{vec}^*\{\mathbf{H}_l\}\right] = \mathbf{R}_{\mathbf{H}}\delta_{k,l} = (\mathbf{A}_0 \otimes \mathbf{B}_0)\,\delta_{k,l}. \tag{4}$$

This paper treats the problem of estimating $\mathbf{R}_{\mathbf{H}}$ based on the received data in N blocks,

$$\mathbf{y}_t(i),\ i = 0, \dots, p-1, \quad t = 0, \dots, N-1. \tag{5}$$

Let the $n_A \times 1-$vector $\theta_{\mathbf{A}}$ and the $n_B \times 1-$vector $\theta_{\mathbf{B}}$ be the real vectors used to parameterize the Kronecker factors (the transmit and receive covariance matrices), $\mathbf{A}$ and $\mathbf{B}$, respectively. Furthermore, assume a linear dependence on $\theta_{\mathbf{A}}$ and $\theta_{\mathbf{B}}$:

$$\mathrm{vec}\{\mathbf{A}\} = \mathbf{P}_{\mathbf{A}}\theta_{\mathbf{A}}, \quad \mathrm{vec}\{\mathbf{B}\} = \mathbf{P}_{\mathbf{B}}\theta_{\mathbf{B}} \tag{6}$$

where $\mathbf{P}_{\mathbf{A}}$ and $\mathbf{P}_{\mathbf{B}}$ are data and parameter independent matrices of size $n^2 \times n_A$ and $m^2 \times n_B$, 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 $n_A = n^2$ and $n_B = m^2$. Also introduce the concatenated parameter vector $\theta = [\theta_{\mathbf{A}}^T\ \theta_{\mathbf{B}}^T]^T$. For later use, denote the parameter vector that corresponds to $\mathbf{A}_0$ and $\mathbf{B}_0$ by θ^0. Note that this parameterization is ambiguous since $\mathbf{A}\alpha \otimes \mathbf{B}\alpha^{-1} = \mathbf{A} \otimes \mathbf{B}$ for any $\alpha \neq 0$. Hence, we can only estimate $\mathbf{A}_0$ and $\mathbf{B}_0$ up to a scalar factor.

Collect the available data in signal block t into the matrix $\mathbf{Y}_t = [\mathbf{y}_t(1) \dots \mathbf{y}_t(p)]$. It is then straightforward to show that the data is complex Gaussian, zero mean with covariance matrix

$$\mathrm{Cov}\left[\mathrm{vec}\{\mathbf{Y}_t\}\right] = \mathbf{R}_0 = \Psi\mathbf{R}_{\mathbf{H}}(\theta^0)\Psi^* + \sigma_0^2\mathbf{I}_{mp} \tag{7}$$

where

$$\Psi = \left(\mathbf{X}^T \otimes \mathbf{I}_m\right) \tag{8}$$

and $\mathbf{X}$ is constructed from $[\mathbf{x}(i)]_{i=0}^{p-1}$ similar to $\mathbf{Y}_t$.

The Gaussian distribution of the received signals implies that the sample covariance matrix

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{t=0}^{N-1} \text{vec}\{\mathbf{Y}_t\}\text{vec}^*\{\mathbf{Y}_t\} \tag{9}$$

is a sufficient statistic, i.e. it contains all relevant information for the estimation problem.

III. INITIAL APPROACHES

The channel is typically estimated from the training data similar to

$$\hat{\mathbf{H}}_t = \mathbf{Y}_t \mathbf{X}^\dagger. \tag{10}$$

When channel statistics are available, a Wiener estimate can also be used. In a scenario with high SNR or when the amount of training data is large, the estimates $\hat{\mathbf{H}}_t$ can be assumed to be identical to $\mathbf{H}_t$ and the methods in [12] and [11] can be applied to estimate $\mathbf{R_H}$. If the SNR is low or the amount of training data is limited, then some care must be taken. It is, e.g., easy to see that the channel covariance estimate

$$\hat{\hat{\mathbf{R}}}_{\mathbf{H}} = \frac{1}{N}\sum_{t=0}^{N-1} \text{vec}\{\hat{\mathbf{H}}_t\}\text{vec}^*\{\hat{\mathbf{H}}_t\} = \Psi^\dagger \hat{\mathbf{R}} \Psi^{\dagger *} \tag{11}$$

is biased and inconsistent. In order to devise an unbiased estimator, assume for a moment that the channel covariance matrix $\text{Cov}[\text{vec}\{\mathbf{H}_t\}] = \mathbf{R_H}$ has no structure except the structure inherent to all covariance matrices. Then the derivations for the ML estimator of the signal covariance matrix in array processing in [8] and [5] carry over directly to the present data model. It can be concluded that the ML estimators (for the unstructured case) of $\mathbf{R_H}$ and σ^2 is

$$\hat{\mathbf{R}}_{\mathbf{H}}^U = \Psi^\dagger \hat{\mathbf{R}} \Psi^{\dagger *} - \hat{\sigma}^{2U}\mathbf{I}_{mp}, \quad \hat{\sigma}^{2U} = \frac{\text{tr}\{\Pi_\Psi^\perp \hat{\mathbf{R}}\}}{mp - mn}. \tag{12}$$

A similar (but not identical) approach for the unstructured estimation of $\mathbf{R_H}$ for the purpose of channel estimation, is proposed in [3]. A simple strategy for imposing the structure of $\mathbf{R_H}$ is then to calculate

$$\begin{aligned} \hat{\mathbf{R}}_{\mathbf{H}}^F &= \mathbf{A}(\hat{\theta}_{\mathbf{A}}^F) \otimes \mathbf{B}(\hat{\theta}_{\mathbf{B}}^F), \\ \hat{\theta}_{\mathbf{A}}^F,\ \hat{\theta}_{\mathbf{B}}^F &= \underset{\theta_{\mathbf{A}},\, \theta_{\mathbf{B}}}{\text{argmin}}\, \|\hat{\mathbf{R}}_{\mathbf{H}}^U - \mathbf{A}(\theta_{\mathbf{A}}) \otimes \mathbf{B}(\theta_{\mathbf{B}})\|_F. \end{aligned} \tag{13}$$

It can be shown that the resulting estimate is unbiased but not statistically efficient. In order to calculate the minimizers, consider the permutation

$$R(\hat{\mathbf{R}}_{\mathbf{H}}^U) = \left[\text{vec}\{\hat{\mathbf{R}}_{\mathbf{H}}^{11}\} \cdots \text{vec}\{\hat{\mathbf{R}}_{\mathbf{H}}^{n1}\}\ \text{vec}\{\hat{\mathbf{R}}_{\mathbf{H}}^{21}\} \cdots \text{vec}\{\hat{\mathbf{R}}_{\mathbf{H}}^{nn}\}\right]^T$$

where $\hat{\mathbf{R}}_{\mathbf{H}}^{kl}$ is the k,lth $m \times m$ block of $\hat{\mathbf{R}}_{\mathbf{H}}^U$. This permutation has the property $R(\mathbf{A} \otimes \mathbf{B}) = \text{vec}\{\mathbf{A}\}\text{vec}^T\{\mathbf{B}\}$. When $\mathbf{A}$ and $\mathbf{B}$ are general Hermitian matrices the solution to (13) is given by a rank one approximation of the permuted matrix $R(\hat{\mathbf{R}}_{\mathbf{H}}^U)$ [9]. A solution that can incorporate linear structure of $\mathbf{A}$ and $\mathbf{B}$ is given in [11]. For later use, introduce the permutation matrix, $\mathbf{P}_R$, such that

$$\text{vec}\{\mathbf{R_H}\} = \mathbf{P}_R \text{vec}\{R(\mathbf{R_H})\} \tag{14}$$

for any matrix $\mathbf{R_H}$ of compatible dimensions.

IV. AN IMPROVED ALGORITHM

The estimator proposed in this paper is based on minimizing

$$\begin{aligned} \bar{V}_C(\theta, \sigma^2) =& \text{vec}^*\{\hat{\mathbf{R}} - \sigma^2\mathbf{I}_{mp} - \Psi\mathbf{R_H}(\theta)\Psi^*\}\mathbf{W} \\ &\times \text{vec}\{\hat{\mathbf{R}} - \sigma^2\mathbf{I}_{mp} - \Psi\mathbf{R_H}(\theta)\Psi^*\} \end{aligned} \tag{15}$$

w.r.t. $\theta_{\mathbf{A}}$, $\theta_{\mathbf{B}}$ and σ^2. Based on covariance matching principles [7], the weighting matrix $\mathbf{W}$ should be chosen to be

$$\frac{1}{N}\left(\text{Cov}\left[\text{vec}\{\hat{\mathbf{R}}\}\right]\right)^{-1} = \left(\mathbf{R}_0^{-T} \otimes \mathbf{R}_0^{-1}\right). \tag{16}$$

Clearly this estimator is not feasible since the weighting matrix depends on unknown quantities. We propose using

$$\mathbf{W} = \hat{\mathbf{Q}}^T \otimes \hat{\mathbf{Q}}, \qquad \hat{\mathbf{Q}} = [\Psi\mathbf{R_H}(\bar{\theta})\Psi^* + \bar{\sigma}^2\mathbf{I}_{mp}]^{-1} \tag{17}$$

where $\bar{\theta}$ is such that $\mathbf{R_H}(\bar{\theta})$ is a consistent estimate of $\mathbf{R_H}(\theta^0)$ and $\bar{\sigma}^2$ is a consistent estimate of σ_0^2. As will be shown later, the proposed choice of weighting does not degrade asymptotic performance.

The minimization w.r.t. σ^2 of (15) using the weighting matrix given in (17) gives

$$\hat{\sigma}^{2C} = \frac{\text{tr}\{\hat{\mathbf{Q}}\hat{\mathbf{R}}\hat{\mathbf{Q}}\}}{\text{tr}\{\hat{\mathbf{Q}}^2\}} - \frac{\text{tr}\{\Psi^*\hat{\mathbf{Q}}^2\Psi\mathbf{R_H}(\theta)\}}{\text{tr}\{\hat{\mathbf{Q}}^2\}}. \tag{18}$$

Note that $\hat{\sigma}^{2C}$ is not constrained to be positive. However, as a consequence of consistency of the estimates, for large enough N it will be positive, and thus this relaxation does not affect asymptotic performance. Inserting (18) and (17) into (15) gives

$$\bar{V}_C(\theta) = [\mathbf{r} - \mathbf{F}\text{vec}\{\mathbf{R_H}\}]^*(\hat{\mathbf{Q}}^T \otimes \hat{\mathbf{Q}})[\mathbf{r} - \mathbf{F}\text{vec}\{\mathbf{R_H}\}] \tag{19}$$

where

$$\begin{aligned} \mathbf{r} &= \text{vec}\{\hat{\mathbf{R}} - \mathbf{I}_{mp}\frac{\text{tr}\{\hat{\mathbf{Q}}\hat{\mathbf{R}}\hat{\mathbf{Q}}\}}{\text{tr}\{\hat{\mathbf{Q}}^2\}}\}, \\ \mathbf{F} &= \left[\text{conj}\{\Psi\} \otimes \Psi - \frac{1}{\text{tr}\{\hat{\mathbf{Q}}^2\}}\text{vec}\{\mathbf{I}_{mp}\}\text{vec}^*\{\Psi^*\hat{\mathbf{Q}}^2\Psi\}\right]. \end{aligned} \tag{20}$$

Next, note that

$$\text{vec}\{\mathbf{R_H}\} = \mathbf{P}_\theta \text{vec}\{\theta_{\mathbf{A}}\theta_{\mathbf{B}}^T\}, \quad \mathbf{P}_\theta = \mathbf{P}_R(\mathbf{P_B} \otimes \mathbf{P_A}). \tag{21}$$

By introducing $\Phi = \theta_{\mathbf{A}}\theta_{\mathbf{B}}^T$, (19) can be written

$$\bar{V}_C(\Phi) = [\mathbf{r} - \mathbf{F}\mathbf{P}_\theta\text{vec}\{\Phi\}]^*(\hat{\mathbf{Q}}^T \otimes \hat{\mathbf{Q}})[\mathbf{r} - \mathbf{F}\mathbf{P}_\theta\text{vec}\{\Phi\}]. \tag{22}$$

The matrix Φ is thus constrained to be a rank one matrix. If, for the moment, we consider minimizing (22) w.r.t. Φ while disregarding the rank constraint we get the minimum norm solution

$$\text{vec}\{\hat{\Phi}\} = [\mathbf{P}_\theta^*\mathbf{F}^*(\hat{\mathbf{Q}}^T \otimes \hat{\mathbf{Q}})\mathbf{F}\mathbf{P}_\theta]^\dagger \mathbf{P}_\theta^*\mathbf{F}^*(\hat{\mathbf{Q}}^T \otimes \hat{\mathbf{Q}})\mathbf{r}. \tag{23}$$

Since $\bar{V}_C(\Phi)$ is quadratic in Φ it follows by Taylor expansion that

$$\bar{V}_C(\Phi) = \bar{V}_C(\hat{\Phi}) + \frac{1}{2}\text{vec}^*\{\hat{\Phi} - \Phi\}\ddot{\bar{V}}_C(\hat{\Phi})\text{vec}\{\hat{\Phi} - \Phi\}. \tag{24}$$

where the derivatives are w.r.t. the elements of vec$\{\Phi\}$. The first term is constant w.r.t. Φ. The conclusion is that the minimization of the weighted low rank approximation problem (WLRA)

$$\begin{aligned} &\underset{\theta_{\mathbf{A}},\theta_{\mathbf{B}}}{\text{argmin}}\, V_C(\theta_{\mathbf{A}}\theta_{\mathbf{B}}^T) \\ V_C(\theta_{\mathbf{A}}\theta_{\mathbf{B}}^T) &= \text{vec}^*\{\hat{\Phi} - \theta_{\mathbf{A}}\theta_{\mathbf{B}}^T\}\Omega\text{vec}\{\hat{\Phi} - \theta_{\mathbf{A}}\theta_{\mathbf{B}}^T\}, \\ \Omega &= \ddot{\bar{V}}_C(\hat{\Phi}) = \mathbf{P}_\theta^*\mathbf{F}^*(\hat{\mathbf{Q}}^T \otimes \hat{\mathbf{Q}})\mathbf{F}\mathbf{P}_\theta \end{aligned} \tag{25}$$

gives the estimates of $\theta_{\mathbf{A}}$ and $\theta_{\mathbf{B}}$. It is interesting to note that $\hat{\Phi}$ is a real matrix. This result is given in the following lemma.

Lemma 1 *Under the assumptions of Section II, the matrices $\hat{\Phi}$ and Ω in (25) have the following properties:*

1) *The matrix $\hat{\Phi}$ defined in (23) is real.*
2) *The matrix $\hat{\Phi}$ is consistent in the sense that*

$$\lim_{N\to\infty} \hat{\Phi} = \theta_{\mathbf{A}}\theta_{\mathbf{B}}^T \tag{26}$$

(w.p.1) where $\theta_{\mathbf{A}}$ and $\theta_{\mathbf{B}}$ are parameter vectors that yield the true channel covariance matrix.

3) *The matrix Ω is full rank and real when $\hat{\mathbf{Q}}$ is positive definite*
4) *The limiting matrix $\Omega_0 = \lim_{N\to\infty}\Omega$ is the normalized inverse of the asymptotic covariance of* $\text{vec}\{\hat{\Phi}\}$,

$$\lim_{N\to\infty} N\text{Cov}\left[\text{vec}\{\hat{\Phi}\}\right] = \Omega_0^{-1}. \tag{27}$$

Proof: The matrices $\mathbf{F}$ and $\mathbf{r}$ in (23) can be expressed as

$$\mathbf{F} = \mathbf{D}(\text{conj}\{\Psi\}\otimes\Psi), \quad \mathbf{r} = \mathbf{D}\text{vec}\{\hat{\mathbf{R}}\} \tag{28}$$

where

$$\mathbf{D} = \left[\mathbf{I}_{m^2p^2} - \text{vec}\{\mathbf{I}_{mp}\}\text{vec}^*\{\hat{\mathbf{Q}}^2\}\frac{1}{\text{tr}\{\hat{\mathbf{Q}}^2\}}\right]. \tag{29}$$

Assuming no structure of the matrix $\hat{\Phi}$, it can be written

$$\text{vec}\{\hat{\Phi}\} = \sum_{k=1}^{\min(n_A,n_B)} (\phi_k + j\bar{\phi}_k) \tag{30}$$

where ϕ_k and $\bar{\phi}_k$ are vectorized real rank one matrices. Then

$$\begin{aligned}&(\text{conj}\{\Psi\}\otimes\Psi)\mathbf{P}_\theta\text{vec}\{\hat{\Phi}\} \\ &= \sum_{k=1}^{\min(n_A,n_B)}\left(\text{vec}\{\Psi(\mathbf{A}_k\otimes\mathbf{B}_k)\Psi^*\} + j\text{vec}\{\Psi(\bar{\mathbf{A}}_k\otimes\bar{\mathbf{B}}_k)\Psi^*\}\right)\end{aligned} \tag{31}$$

where $\mathbf{A}_k$, $\mathbf{B}_k$, $\bar{\mathbf{A}}_k$ and $\bar{\mathbf{B}}_k$ are Hermitian matrices (that are zero if and only if ϕ_k and $\bar{\phi}_k$ are zero). Before proceeding with the proof, introduce a matrix $\mathbf{J}$ such that $\upsilon = \mathbf{J}^{-1}\text{vec}\{\Upsilon\}$ is real if Υ is Hermitian. Then introduce the three real vectors

$$\begin{aligned}\upsilon_{\mathbf{A},\mathbf{B}} &= \mathbf{J}^{-1}\sum_{k=1}^{\min(n_A,n_B)}\text{vec}\{\Psi(\mathbf{A}_k\otimes\mathbf{B}_k)\Psi^*\},\\ \bar{\upsilon}_{\mathbf{A},\mathbf{B}} &= \mathbf{J}^{-1}\sum_{k=1}^{\min(n_A,n_B)}\text{vec}\{\Psi(\bar{\mathbf{A}}_k\otimes\bar{\mathbf{B}}_k)\Psi^*\},\\ \upsilon_{\hat{\mathbf{R}}} &= \mathbf{J}^{-1}\text{vec}\{\hat{\mathbf{R}}\}.\end{aligned} \tag{32}$$

The criterion function (22) can then be written

$$\begin{aligned}\text{vec}^*\{\upsilon_{\hat{\mathbf{R}}} - \upsilon_{\mathbf{A},\mathbf{B}} - j\bar{\upsilon}_{\mathbf{A},\mathbf{B}}\}\mathbf{J}^*\mathbf{D}^*(\hat{\mathbf{Q}}^T\otimes\hat{\mathbf{Q}})\mathbf{D}\mathbf{J}\\ \times\text{vec}\{\upsilon_{\hat{\mathbf{R}}} - \upsilon_{\mathbf{A},\mathbf{B}} - j\bar{\upsilon}_{\mathbf{A},\mathbf{B}}\}.\end{aligned} \tag{33}$$

Clearly, minimizing w.r.t. $\bar{\upsilon}_{\mathbf{A},\mathbf{B}}$ yields, as one possible solution, $\bar{\upsilon}_{\mathbf{A},\mathbf{B}} = 0$ which gives that, since $\hat{\Phi}$ is the minimum norm solution, $\bar{\phi}_k = 0$ and thus $\hat{\Phi}$ is real. The consistency result 2) follows by noting that $\mathbf{D}\text{vec}\{\mathbf{I}_{mp}\} = \mathbf{0}_{m^2p^2\times 1}$.

In order to prove the non-singularity of Ω, write

$$\Omega = \mathbf{P}_\theta^*(\Psi^T\otimes\Psi^*)\mathbf{D}^*(\hat{\mathbf{Q}}^T\otimes\hat{\mathbf{Q}})\mathbf{D}(\text{conj}\{\Psi\}\otimes\Psi)\mathbf{P}_\theta. \tag{34}$$

It is straightforward to show that

$$\mathbf{D}^*(\hat{\mathbf{Q}}^T\otimes\hat{\mathbf{Q}})\mathbf{D} = (\hat{\mathbf{Q}}^T\otimes\hat{\mathbf{Q}}) - \text{vec}\{\hat{\mathbf{Q}}^2\}\text{vec}^*\{\hat{\mathbf{Q}}^2\}\frac{1}{\text{tr}\{\hat{\mathbf{Q}}^2\}}. \tag{35}$$

Hence it can be seen that the result of multiplication of Ω from the right with any real vector υ can be written

$$\begin{aligned}\mathbf{P}_\theta^*(\Psi^T\otimes\Psi^*)\Bigg(\text{vec}\{\hat{\mathbf{Q}}\big[\Psi\big(\sum_{k=1}^{\min(n_A,n_B)}(\mathbf{A}_k\otimes\mathbf{B}_k)\big)\Psi^*\\ -\mathbf{I}_{mp}\sum_{k=1}^{\min(n_A,n_B)}\frac{\text{tr}\{\hat{\mathbf{Q}}\Psi(\mathbf{A}_k\otimes\mathbf{B}_k)\Psi^*\hat{\mathbf{Q}}\}}{\text{tr}\{\hat{\mathbf{Q}}^2\}}\big]\hat{\mathbf{Q}}\}\Bigg).\end{aligned} \tag{36}$$

By considering the dimensions of Ψ it is clear that the matrix within square brackets above can be zero only if $\sum_{k=1}^{\min(n_A,n_B)}\mathbf{A}_k\otimes\mathbf{B}_k = \mathbf{0}_{mn\times mn}$. This implies that $\upsilon = \mathbf{0}$. Hence Ω is full rank when $\hat{\mathbf{Q}}$ is full rank. The expression (36) can also be used to prove the realness of Ω. Note that the result of multiplying Ω from the right and left with any real vectors (of suitable dimensions) can be written

$$[\upsilon^T\mathbf{P}_\theta^*(\Psi^T\otimes\Psi^*)\mathbf{J}^{-*}][(\mathbf{J}^{-1}(\hat{\mathbf{Q}}^{-T}\otimes\hat{\mathbf{Q}}^{-1})\mathbf{J}^{-*})^{-1}][\mathbf{J}^{-1}\text{vec}\{\mathbf{Y}\}]$$

with υ a real vector, $\mathbf{Y}$ an Hermitian matrix and $\mathbf{J}$ defined above. The bracketed expressions can all be shown to be real, hence Ω is real. The final part of the proof is straightforward and is omitted due to space limitations. ■

It is not known how to solve the WLRA problem (25) in closed form. However, the statistical properties of the estimation problem at hand allows us to approximate the solution without degrading the performance asymptotically. The procedure is similar to the one proposed for reduced rank linear regression in [10].

Theorem 1 *Let θ^0 be a parameter vector that corresponds to $\mathbf{A}_0\otimes\mathbf{B}_0$. Let $\bar{\theta}$ be such that*

$$\lim_{N\to\infty}\sqrt{N}\|\bar{\theta}-\theta^0\|_2 \tag{37}$$

is bounded in probability. The estimate

$$\hat{\mathbf{R}}_{\mathbf{H}} = \mathbf{A}(\hat{\theta}_{\mathbf{A}})\otimes\mathbf{B}(\hat{\theta}_{\mathbf{B}}) \tag{38}$$

where

$$\begin{aligned}\begin{pmatrix}\hat{\theta}_{\mathbf{A}}\\ \hat{\theta}_{\mathbf{B}}\end{pmatrix} &= \begin{pmatrix}\bar{\theta}_{\mathbf{A}}\\ \bar{\theta}_{\mathbf{B}}\end{pmatrix} + 2\mathbf{K}^\dagger\begin{pmatrix}\bar{\theta}_{\mathbf{B}}^T\otimes\mathbf{I}_{n_A}\\ \mathbf{I}_{n_B}\otimes\bar{\theta}_{\mathbf{A}}^T\end{pmatrix}\Omega\text{vec}\{\hat{\Phi}-\bar{\theta}_{\mathbf{A}}\bar{\theta}_{\mathbf{B}}^T\},\\ \mathbf{K} &= 2\begin{pmatrix}\bar{\theta}_{\mathbf{B}}^T\otimes\mathbf{I}_{n_A}\\ \mathbf{I}_{n_B}\otimes\bar{\theta}_{\mathbf{A}}^T\end{pmatrix}\Omega\left(\bar{\theta}_{\mathbf{B}}\otimes\mathbf{I}_{n_A}\ \mathbf{I}_{n_B}\otimes\bar{\theta}_{\mathbf{A}}\right)\end{aligned} \tag{39}$$

is asymptotically equivalent to the estimate obtained by exactly minimizing $V_C(\theta_{\mathbf{A}},\theta_{\mathbf{B}})$ in (25).

Proof: The estimates $\hat{\theta}_{\mathbf{A}}$ and $\hat{\theta}_{\mathbf{B}}$ are minimizers of

$$V(\theta) = \dot{V}_C(\bar{\theta})(\theta-\bar{\theta}) + \frac{1}{2}(\theta-\bar{\theta})^T\mathbf{K}(\theta-\bar{\theta}). \tag{40}$$

Thus

$$\dot{V}(\theta^0) = \dot{V}_C(\bar{\theta}) + \mathbf{K}(\theta^0-\bar{\theta}). \tag{41}$$

Next, a Taylor expansion of $\dot{V}_C(\theta)$ around $\bar{\theta}$ gives

$$\begin{aligned}\dot{V}_C(\theta^0) &= \dot{V}_C(\bar{\theta}) + \ddot{V}_C(\bar{\theta})(\theta^0-\bar{\theta}) + O(\|\bar{\theta}-\theta^0\|^2)\\ &= \dot{V}_C(\bar{\theta}) + \mathbf{K}(\theta^0-\bar{\theta}) + o_p(\frac{1}{\sqrt{N}}).\end{aligned} \tag{42}$$

Note that we have that $\mathbf{K} = \ddot{V}_C(\theta^0) + o_p(1)$. The relations

$$\begin{aligned}\dot{V}(\theta^0) &= \dot{V}_C(\theta^0) + o_p(\frac{1}{\sqrt{N}}),\\ \ddot{V}(\theta^0) &= \ddot{V}_C(\theta^0) + o_p(1)\end{aligned} \tag{43}$$

now lead us to conclude that the theorem holds. ■

Theorem 1 gives a practical way of calculating the estimates. Initial estimates $\bar{\theta}$ can be obtained by for example calculating a rank one approximation (in the Frobenius norm sense) of $\hat{\Phi}$, or by using $\hat{\theta}_{\mathbf{A}}^F$ and $\hat{\theta}_{\mathbf{A}}^F$. Note that if the latter choice is made, the same estimates can be used for the calculation of $\hat{\mathbf{Q}}$ in (17).

V. PERFORMANCE ANALYSIS

Theorem 2 *Under the assumptions of Section II, the estimate $\hat{\mathbf{R}}_{\mathbf{H}}$ proposed in Theorem 1 is consistent and has an asymptotic normal distribution with zero mean and covariance*

$$\lim_{N\to\infty} N\text{Cov}\left[\text{vec}\{\hat{\mathbf{R}}_{\mathbf{H}}\}\right] = \mathbf{P}_\theta \Gamma_0 [\Gamma_0^T \Omega_0 \Gamma_0]^\dagger \Gamma_0^T \mathbf{P}_\theta^* \tag{44}$$

where

$$\Omega_0 = \mathbf{P}_\theta^* \mathbf{F}_0^* (\mathbf{R}_0^{-T} \otimes \mathbf{R}_0^{-1}) \mathbf{F}_0 \mathbf{P}_\theta,$$
$$\mathbf{F}_0 = \left[\text{conj}\{\Psi\} \otimes \Psi - \frac{1}{\text{tr}\{(\mathbf{R}_0^{-1})^2\}} \text{vec}\{\mathbf{I}_{mp}\} \text{vec}^*\{\Psi^*(\mathbf{R}_0^{-1})^2 \Psi\}\right],$$
$$\Gamma_0 = \left(\theta_{\mathbf{B}}^0 \otimes \mathbf{I}_{n_A} \; \mathbf{I}_{n_B} \otimes \theta_{\mathbf{A}}^0\right). \tag{45}$$

Proof: We begin by proving the theorem assuming that $\hat{\theta}_{\mathbf{A}}$ and $\hat{\theta}_{\mathbf{B}}$ really are the minimizers of the WLRA problem. For simplicity, assume that the weighting matrix Ω of $V_C(\theta)$ in (25) is replaced by Ω_0 and denote the resulting criterion function by $V(\theta)$. The consistency of the estimate (in the sense that $\mathbf{R}_{\mathbf{H}}(\hat{\theta})$ converges to $\mathbf{R}_{\mathbf{H}}(\theta^0)$) is a direct consequence of the consistency of $\hat{\mathbf{R}}$. Also, for brevity of notation, assume that θ^0 is the limit of $\hat{\theta}$. In order to derive the asymptotic covariance, note that a Taylor series expansion gives

$$\text{vec}\{\mathbf{R}_{\mathbf{H}}(\theta^0) - \mathbf{A} \otimes \mathbf{B}\} = \mathbf{P}_\theta \Gamma_0 (\theta^0 - \theta) + O(\|\theta^0 - \theta\|_2^2). \tag{46}$$

Let $\hat{\theta}$ be the parameter vector that minimizes the criterion function. A Taylor series expansion of $V(\theta)$ gives

$$\mathbf{0} = \dot{V}(\hat{\theta}) = \dot{V}(\theta^0) - \mathbf{G}(\theta^0 - \hat{\theta}) + o_p(\|\theta^0 - \hat{\theta}\|_2) \tag{47}$$

where the gradient vector and Hessian used are defined as

$$[\dot{V}(\theta)]_i = \frac{\partial V(\theta)}{\partial [\theta]_i}, \quad [\mathbf{G}]_{i,j} = \lim_{N\to\infty} \left.\frac{\partial^2 V(\theta)}{\partial [\theta]_i \partial [\theta]_j}\right|_{\theta=\theta^0}. \tag{48}$$

Now,

$$\dot{V}(\theta^0) = \mathbf{G}(\theta^0 - \hat{\theta}) + o_p(\|\theta^0 - \hat{\theta}\|_2) \tag{49}$$

and therefore

$$(\theta^0 - \hat{\theta}) - \left[\mathbf{I}_{n_A+n_B} - \mathbf{G}^\dagger \mathbf{G}\right](\theta^0 - \hat{\theta}) = \mathbf{G}^\dagger \dot{V}(\theta^0) + o_p(\|\hat{\theta} - \theta^0\|_2).$$

Note that the matrix within square brackets above is a projection matrix on the null-space of $\mathbf{G} = 2\Gamma_0^T \Omega_0 \Gamma_0$. Thus multiplication from the left with $\mathbf{P}_\theta \Gamma_0$ gives

$$\mathbf{P}_\theta \Gamma_0 (\theta^0 - \hat{\theta}) = \mathbf{P}_\theta \Gamma_0 \mathbf{G}^\dagger \dot{V}(\theta^0) + o_p(\|\theta^0 - \hat{\theta}\|). \tag{50}$$

This gives, making use of (46) and the consistency of the estimate,

$$\lim_{N\to\infty} N\text{Cov}\left[\text{vec}\{\mathbf{R}_{\mathbf{H}}(\theta_0) - \mathbf{A}(\hat{\theta}_{\mathbf{A}}) \otimes \mathbf{B}(\hat{\theta}_{\mathbf{B}})\}\right]$$
$$= \mathbf{P}_\theta \Gamma_0 \mathbf{G}^\dagger \left[\lim_{N\to\infty} N\text{E}\left[\dot{V}(\theta^0) \dot{V}^*(\theta^0)\right]\right] \mathbf{G}^\dagger \Gamma_0^T \mathbf{P}_\theta^*. \tag{51}$$

Finally, the relation (which makes use of Lemma 1)

$$\lim_{N\to\infty} N\text{E}\left[\dot{V}(\theta^0) \dot{V}^*(\theta^0)\right]$$
$$= \lim_{N\to\infty} 4\Gamma_0^T \Omega_0 \text{E}\left[N \text{vec}\{\hat{\Phi} - \theta_{\mathbf{A}}^0 \theta_{\mathbf{B}}^{0T}\} \text{vec}^*\{\Phi - \theta_{\mathbf{A}}^0 \theta_{\mathbf{B}}^{0T}\}\right] \Omega_0 \Gamma_0$$
$$= 2\mathbf{G} \tag{52}$$

shows the covariance part of the proof. Also note that the asymptotic performance is unaffected by using a consistent estimate (such as Ω) of the weighting matrix instead of the true Ω_0. The asymptotic normality of the estimate follows from the asymptotic normality of the elements of the sample covariance matrix, see, e.g., [1]. Theorem 1 finally shows that it is not necessary to find the exact solution to the WLRA, and that $\hat{\theta}$ gives the same asymptotic performance. ∎

Having derived an expression for the covariance of the asymptotic distribution of the proposed estimate we now turn to the problem of determining if the estimate is asymptotically efficient. The answer is given in the following theorem.

Theorem 3 *Let $\hat{\mathbf{R}}_{\mathbf{H}}^C$ be an estimate of $\mathbf{A}_0 \otimes \mathbf{B}_0$ constructed as*

$$\hat{\mathbf{R}}_{\mathbf{H}}^C = \mathbf{A}(\hat{\theta}_{\mathbf{A}}^C) \otimes \mathbf{B}(\hat{\theta}_{\mathbf{B}}^C)$$

where $\hat{\theta}_{\mathbf{A}}^C$ and $\hat{\theta}_{\mathbf{B}}^C$ are given by

$$\underset{\theta}{\text{argmin}} \min_{\sigma^2 \geqslant 0} \bar{V}_C(\theta, \sigma^2) \tag{53}$$

and where $\bar{V}_C(\theta, \sigma^2)$ is given by (15). Then, under the data model described in Section II, it has an asymptotic complex Gaussian distribution with

$$\lim_{N\to\infty} N\text{Cov}\left[\text{vec}\{\hat{\mathbf{R}}_{\mathbf{H}}^C\}\right] = N\,CRB(\theta^0, \sigma_0^2, N) \tag{54}$$

where $CRB(\theta^0, \sigma_0^2, N)$ is the Cramér-Rao lower bound for the data model. The estimate $\hat{\mathbf{R}}_{\mathbf{H}}^C$ is a consistent estimate of $\mathbf{A}_0 \otimes \mathbf{B}_0$. Furthermore, (54) still holds if $\mathbf{W}$ is replaced by any consistent estimate.

Proof: The proof is omitted due to space limitations. ∎

The implication of the above is that the covariance matching estimate is efficient and hence that the asymptotic covariance (44) gives a compact expression for the CRB. In the next section, numerical results will be given that illustrate the small sample performance of the estimator.

VI. NUMERICAL RESULTS

Monte Carlo simulations are used to evaluate the small sample performance of the proposed estimator. Two matrices $\mathbf{A}_0$ and $\mathbf{B}_0$ are generated (and then fixed) and the corresponding $\mathbf{R}_{\mathbf{H}}(\theta^0)$ is calculated. The matrix $\mathbf{X}$ was randomly generated and then fixed throughout the simulations. The SNR is given by

$$\text{SNR} = \frac{\|\mathbf{X}\|_F^2}{p}, \; \sigma_0^2 = 1, \; \text{tr}\{\mathbf{R}_{\mathbf{H}}(\theta^0)\} = mn. \tag{55}$$

Note that the fixed trace of the channel covariance matrix and the fixed noise variance is introduced in order to give a consistent definition of SNR, the information is *not* used by any of the estimators. In each Monte Carlo trial, N samples are generated from a complex Gaussian distribution with covariance given by $\mathbf{R}_0$ in (7). Each estimator was applied to the sample set and the normalized root-MSE was defined as

$$\sqrt{\frac{1}{L}\sum_{k=1}^{L} \frac{\|\mathbf{R}_{\mathbf{H}}(\theta^0) - \hat{\mathbf{R}}_{\mathbf{H}}^k\|_F^2}{\|\mathbf{R}_{\mathbf{H}}(\theta^0)\|_F^2}} \tag{56}$$

where $\hat{\mathbf{R}}_{\mathbf{H}}^k$ is the estimate produced by the estimator in question in Monte Carlo trial k and L is the number of Monte Carlo trials. Here $L = 500$. The methods evaluated are: 1) The unstructured estimate $\hat{\mathbf{R}}_{\mathbf{H}}^U$, (12); 2) The structured approximation of the latter estimate $\hat{\mathbf{R}}_{\mathbf{H}}^F$, (13); 3) The unstructured estimate $\hat{\hat{\mathbf{R}}}_{\mathbf{H}}$, (11) that can be seen as a high SNR (HSNR) approximation of $\hat{\mathbf{R}}_{\mathbf{H}}^U$; 4) The structured approximation (similar to (13)) of the latter; 5) An estimate obtained by separately estimating the transmit and receive covariances (with proper normalization)

$$\check{\mathbf{B}} = \frac{1}{N}\sum_{t=1}^{N} \hat{\mathbf{H}}_t \hat{\mathbf{H}}_t^*, \quad \check{\mathbf{A}} = \frac{1}{N}\sum_{t=1}^{N} \hat{\mathbf{H}}_t^* \hat{\mathbf{H}}_t. \tag{57}$$

This estimate is marked by Trans./Rec. in the figures; 6) The covariance matching estimator proposed in [12], applied to the channel estimates $\hat{\mathbf{H}}_t$ (this is motivated when SNR is high). Finally, the estimate proposed in Section IV is included. The CRB is also included as is the asymptotic covariance of the estimate $\hat{\mathbf{R}}^F_{\mathbf{H}}$ in 2) above. In all simulations presented in this section the number of training symbols per block is fixed to $p = 4$. If p is significantly larger, then the methods of [12] are more suitable than those proposed in this work.

Figure 1 shows an example where $\mathbf{A}$ and $\mathbf{B}$ are Toeplitz structured, SNR= 3dB and the number of signal blocks, N, is varied. The proposed estimator and estimators 2), 4) and 6) above take advantage of the Toeplitz structure and this naturally gives them a better performance. At about $N = 150$, the bias in the high SNR estimators becomes a significant part of the error. The proposed estimator exhibits promising threshold behavior and attains the CRB also for small sample sizes. In Figure 2, the sample size was fixed, $N = 100$, and the effect of varying the SNR is illustrated. In this example the only structure assumed on the Kronecker factors $\mathbf{A}$ and $\mathbf{B}$ is that they are Hermitian matrices. The conclusion is that the proposed method outperforms the other methods for both high and low SNR. As expected, the biased methods perform equivalent to their unbiased counterparts when SNR is high but significantly worse when the SNR is low. The proposed method reaches the asymptotical performance expressions for quite small number of signal blocks. Note that the expressions are not valid for the high SNR/small sample size case.

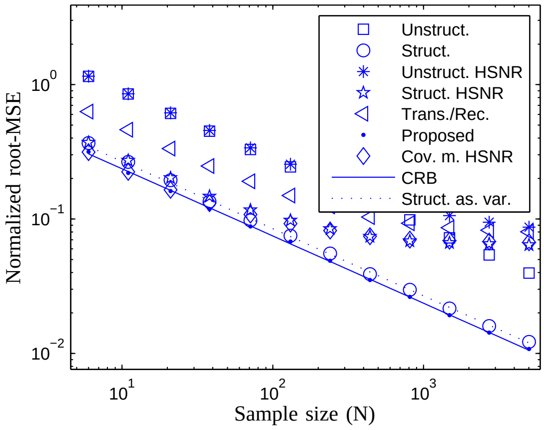

Fig. 1. Normalized root-MSE as a function of the number of signal blocks for different alternative estimators. In the example $\mathbf{A}$ and $\mathbf{B}$ are both Toeplitz matrices, $m = n = 3$, $p = 4$ and SNR= 3dB.

VII. CONCLUSION

This paper considers the problem of estimating second order channel statistics for MIMO channels. It is assumed that training data from a number of signal blocks is available (as opposed to perfect observations of the channel realizations). A first approach is the ML estimate that disregards the known structure of the channel covariance. Imposing the known structure can improve the estimate. The performance can be further improved if a statistically optimal weighting is applied when the structure is imposed. The resulting estimator is asymptotically optimal as shown in Theorem 3 but requires solving a weighted low rank approximation problem, which is a non-convex optimization problem. This issue is addressed in Theorem 1 that shows how an approximate solution can be used without compromising asymptotic performance. The covariance of asymptotic distribution of the proposed estimator is derived in Theorem 2. Numerical studies that indicate promising performance conclude the paper.

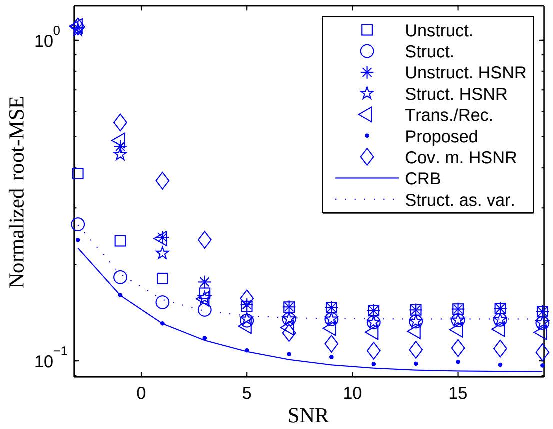

Fig. 2. Normalized root-MSE as a function of SNR for different estimators. In the example $\mathbf{A}$ and $\mathbf{B}$ are Hermitian matrices, $m = n = 3$, $p = 4$ and $N = 100$.

REFERENCES

[1] T. W. Andersson, *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons Inc., 1958.

[2] M. Bengtsson and P. Zetterberg, “Some notes on the Kronecker model,” *EURASIP Journal on Wireless Communications and Networking*, submitted, April 2006.

[3] N. Czink, G. Matz, D. Seethaler, and F. Hlawatsch, “Improved MMSE estimation of correlated MIMO channels using a structured correlation estimator,” in *Proceedings of SPAWC*, June 2005.

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

[5] A. Jaffer, “Maximum likelihood direction finding of stochastic sources: a separable solution,” in *Proceedings of ICASSP 88*, April 1988.

[6] J. 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.

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

[8] P. Stoica and A. Nehorai, “On the concentrated stochastic likelihood function in array signal processing,” *Circuits, Systems, and Signal Processing*, vol. 14, no. 5, pp. 669–674, May 1992.

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

[10] K. Werner and M. Jansson, “Reduced rank linear regression and weighted low rank approximations,” *IEEE Transactions on Signal Processing*, no. 6, pp. 2063–2075, June 2006.

[11] K. Werner, M. Jansson, and P. Stoica, “On estimation of covariance matrices with kronecker product structure,” *IEEE Transactions on Signal Processing*, submitted, October 2006.

[12] ——, “Kronecker structured covariance matrix estimation,” in *Proceedings of ICASSP 07*, April 2007.

[13] 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.