ESTIMATION OF CHANNEL AND NOISE CORRELATIONS FOR MIMO CHANNEL ESTIMATION

Frank A. Dietrich, Tzvetan Ivanov, and Wolfgang Utschick

Institute for Circuit Theory and Signal Processing, Munich University of Technology,
80290 Munich, Germany, E-Mail: {Dietrich, Utschick}@nws.ei.tum.de

ABSTRACT

A (linear) minimum mean square error (MMSE) channel estimator leads to a significantly improved quality of the channel estimates compared to a least squares approach. To realize these performance gains, it requires knowledge about the statistical second order moments of the channel parameters and noise. Previously proposed estimators for the second order moments are either heuristic or do not ensure positive definite estimates of the channel covariance matrix. We present novel estimators, which are either guaranteed to be positive definite or ensure numerical stability when applied for MMSE estimation of MIMO channels. They ensure these gains already for a few noisy and statistically independent observations of the channel.

1. INTRODUCTION

The increased number of degrees of freedom in multiple-input-multiple-output (MIMO) systems leads to a larger number of channel parameters, which need to be estimated accurately to realize the envisioned capacity gains in a communication system.

A (linear) minimum mean square error (MMSE) channel estimator yields a significantly improved quality of the channel estimates compared to a least squares approach. To realize these performance gains, it requires knowledge about the statistical second order moments of the channel parameters and noise. For example, consider an 8×8 MIMO system with 64 channel parameters, which requires estimation of a channel covariance matrix with $64^2 = 4096$ independent real-valued parameters.

The estimation of the channel and noise statistics based on a received training sequence over B independently fading blocks can be formulated as the problem of estimating a covariance matrix with linear structure [1, 2]. Previously proposed estimators for the second order moments are either heuristic [3] or do not ensure positive semidefinite (psd) estimates of the channel covariance matrix [1, 2]. If the estimated channel covariance matrix is not positive semidefinite, we typically obtain an ill-conditioned system of linear equations ($\Rightarrow$ numerically unstable) for computing the MMSE channel estimator. Typically ill-conditioning arises, if less than 100 independent observations of the channel are available, which is often the case in practical wireless MIMO channel as indicated by channel measurements [4].

In Sec. 2 the signal model and estimation problem are stated and in Sec. 3 the importance of having a positive semidefinite estimate of the channel covariance matrix for MMSE channel estimation is illustrated. The unbiased least-squares (LS) estimator for the channel covariance matrix and noise variance is reviewed in Sec. 4.1 and a new derivation is given. We present two novel estimators, which are either guaranteed to be positive semidefinite (Sec. 4.2) or ensure numerical stability (Sec. 4.3) when applied for MMSE estimation of MIMO channels. They ensure the performance gains already for a *few* statistically independent observations of the channel. Moreover, positive semidefinite estimates are guaranteed and, thus, numerical stability is ensured for small B. These algorithms are also applicable in frequency-selective channels, e.g., to estimate the power-delay profile. Furthermore an optimization problem for the heuristic in [3] is given in Sec. 4.4. The estimators in Sec. 4 are derived assuming white noise. In Sec. 5 we generalize these estimators to spatially correlated noise and interference.

Notation: Random vectors and matrices are denoted by lower and upper case sans serif bold letters (e.g. b, B), whereas the realizations or deterministic variables are, e.g., $\boldsymbol{b}$, $\boldsymbol{B}$. The operators $\mathrm{E}[\bullet]$, $(\bullet)^{\mathrm{T}}$, $(\bullet)^{\mathrm{H}}$, $\boldsymbol{A}^{\dagger}$, and $\mathrm{tr}(\bullet)$ stand for expectation, transpose, Hermitian transpose, pseudo-inverse, and trace of a matrix, respectively. $\mathrm{E}_{\mathsf{a}}[\bullet]$ denotes the expectation w.r.t. to random vector a. $\otimes$ and $\delta_{k,k'}$ denote the Kronecker product and function, $\mathbf{vec}[\boldsymbol{B}]$ stacks the columns of $\boldsymbol{B}$ in a vector. $\boldsymbol{e}_i$ is the ith column of an $N \times N$ identity matrix $\boldsymbol{I}_N$. The squared Frobenius norm of a matrix $\boldsymbol{A}$ is $\|\boldsymbol{A}\|_{\mathrm{F}}^2 = \mathrm{tr}\left[\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\right]$.

2. PROBLEM STATEMENT

In a system with K transmit and M receive antennas, N training symbols $\boldsymbol{s}[n] \in \mathbb{C}^K$, $n \in \{1, 2, \ldots, N\}$, per transmitter are observed in the qth time slot. All N received training symbols in time slot q

$$\boldsymbol{y}[n,q] = \boldsymbol{H}[q]\boldsymbol{s}[n] + \boldsymbol{n}[n,q] \quad (1)$$

are collected in

$$\boldsymbol{Y}[q] = [\boldsymbol{y}[1,q], \boldsymbol{y}[2,q], \ldots, \boldsymbol{y}[N,q]]. \quad (2)$$

Thus, we obtain

$$\boldsymbol{Y}[q] = \boldsymbol{H}[q]\bar{\boldsymbol{S}} + \boldsymbol{N}[q] \in \mathbb{C}^{M\times N}, \quad (3)$$

where $\boldsymbol{H}[q] \in \mathbb{C}^{M\times K}$ describes the frequency flat MIMO channel, $\boldsymbol{N}[q] = [\boldsymbol{n}[1,q], \boldsymbol{n}[2,q], \ldots, \boldsymbol{n}[N,q]]$ is additive noise and interference, and $\bar{\boldsymbol{S}} = [\boldsymbol{s}[1], \boldsymbol{s}[2], \ldots, \boldsymbol{s}[N]] \in \mathbb{C}^{K\times N}$. With $\boldsymbol{S} = \bar{\boldsymbol{S}}^{\mathrm{T}} \otimes \boldsymbol{I}_M$, $\boldsymbol{y}[q] = \mathrm{vec}[\boldsymbol{Y}[q]]$, and $\boldsymbol{h}[q] = \mathrm{vec}[\boldsymbol{H}[q]]$ the system model can be rewritten as

$$\boldsymbol{y}[q] = \boldsymbol{S}\boldsymbol{h}[q] + \boldsymbol{n}[q] \in \mathbb{C}^{MN}. \quad (4)$$

The stationary and zero-mean channel and noise with covariance matrices $\boldsymbol{C}_{\boldsymbol{h}} = \mathrm{E}[\boldsymbol{h}[q]\boldsymbol{h}[q]^{\mathrm{H}}]$ and $\boldsymbol{C}_{\boldsymbol{n}} = \mathrm{E}[\boldsymbol{n}[q]\boldsymbol{n}[q]^{\mathrm{H}}]$ are mutually uncorrelated. In Sec. 5 the case of spatially correlated noise is addressed where $\boldsymbol{C}_{\boldsymbol{n}} = \boldsymbol{I}_N \otimes \boldsymbol{C}_{\boldsymbol{n},\mathrm{S}}$ with $\boldsymbol{C}_{\boldsymbol{n},\mathrm{S}} = \mathrm{E}\left[\boldsymbol{n}[n,q]\boldsymbol{n}[n,q]^{\mathrm{H}}\right]$, but at first we assume $\boldsymbol{C}_{\boldsymbol{n}} = c_{\boldsymbol{n}}\boldsymbol{I}_{MN}$, for simplicity.

This yields the covariance matrix of the received signal $\boldsymbol{y}[q]$

$$\boldsymbol{C}_{\boldsymbol{y}} = \mathrm{E}[\boldsymbol{y}[q]\boldsymbol{y}[q]^{\mathrm{H}}] = \boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}} + c_{\boldsymbol{n}}\boldsymbol{I}_{MN}, \quad (5)$$

which depends on the channel covariance matrix $\boldsymbol{C}_{\boldsymbol{h}}$ and noise variance $c_{\boldsymbol{n}}$ to be estimated. Thus, the problem of estimating $\boldsymbol{C}_{\boldsymbol{h}}$ and $c_{\boldsymbol{n}}$ can be formulated as estimation of $\boldsymbol{C}_{\boldsymbol{y}}$ taking into account its structure.

Problem statement: Estimate $\boldsymbol{C}_{\boldsymbol{h}}$ and c_n (or $\boldsymbol{C}_{\boldsymbol{n},\mathrm{S}}$) based on B observations $\{\boldsymbol{y}[q]\}_{q=1}^{B}$ of the random vector $\boldsymbol{y}[q]$.

3. MMSE CHANNEL ESTIMATION

The MMSE channel estimator based on the observation $\boldsymbol{y}[q]$ is given by

$$\underset{\boldsymbol{W}}{\mathrm{argmin}}\, \mathrm{E}_{\boldsymbol{h},\boldsymbol{y}}[\|\boldsymbol{h} - \boldsymbol{W}\boldsymbol{y}[q]\|_2^2] = \boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}}(\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}} + c_{\boldsymbol{n}}\boldsymbol{I}_{MN})^{-1} \quad (6)$$

With the eigenvalue decomposition (EVD) of $\boldsymbol{C}_{\boldsymbol{h}} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}}$, where $\boldsymbol{\Lambda}$ is diagonal with λ_i ($\lambda_i \geq \lambda_{i+1}$) on the diagonal, and assuming $\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S} = N\boldsymbol{I}_{MK}$ Eq. (6) can be written as

$$\boldsymbol{W} = \boldsymbol{U}\boldsymbol{\Lambda}(\boldsymbol{\Lambda}N + c_{\boldsymbol{n}}\boldsymbol{I}_{MK})^{-1}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{S}^{\mathrm{H}}. \quad (7)$$

If $\lambda_i < 0$ and $\lambda_i N + c_{\boldsymbol{n}} \approx 0$ for some i due to an indefinite estimate of $\boldsymbol{C}_{\boldsymbol{h}}$, the system of equations is highly ill-conditioned, which leads to numerical problems in implementation. Thus, for indefinite estimates of $\boldsymbol{C}_{\boldsymbol{h}}$ numerical stability is not guaranteed in this application.

4. ESTIMATION APPROACHES

All estimation approaches in this Section aim at a structured least squares approximation of the sufficient[1] statistic w.r.t. $\boldsymbol{C}_{\boldsymbol{h}}$ and c_{n}: $\tilde{\boldsymbol{C}}_{\boldsymbol{y}} = \frac{1}{B}\sum_{q=1}^{B}\boldsymbol{y}[q]\boldsymbol{y}[q]^{\mathrm{H}}$ (sample covariance matrix).

4.1. Unbiased Least Squares Estimator

As indicated in [2] the structured least squares (LS) approximation of $\tilde{\boldsymbol{C}}_{\boldsymbol{y}}$

$$\min_{\boldsymbol{C}_{\boldsymbol{h}},c_{\boldsymbol{n}}} \|\tilde{\boldsymbol{C}}_{\boldsymbol{y}} - \boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}} - c_{\boldsymbol{n}}\boldsymbol{I}_{MN}\|_{\mathrm{F}}^2 \quad (8)$$

performs close to maximum likelihood approaches as $B \to \infty$, which is referred to as "extended invariance principle" in [5].[2] The solution is obtained using the gradient of the cost function [1] (see also [6])

$$\hat{c}_{\boldsymbol{n}}^{\mathrm{LS}} = \frac{1}{M(N-K)}\mathrm{trace}\left[\tilde{\boldsymbol{C}}_{\boldsymbol{y}} - \boldsymbol{P}\tilde{\boldsymbol{C}}_{\boldsymbol{y}}\boldsymbol{P}\right] \quad (9)$$

$$= \frac{1}{M(N-K)}\mathrm{trace}\left[\boldsymbol{P}^{\perp}\tilde{\boldsymbol{C}}_{\boldsymbol{y}}\boldsymbol{P}^{\perp}\right] \quad (10)$$

$$\hat{\boldsymbol{C}}_{\boldsymbol{h}}^{\mathrm{LS}} = \boldsymbol{S}^{\dagger}\tilde{\boldsymbol{C}}_{\boldsymbol{y}}\boldsymbol{S}^{\dagger,\mathrm{H}} - (\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S})^{-1}\hat{c}_{\boldsymbol{n}}^{\mathrm{LS}} \quad (11)$$

with projector $\boldsymbol{P} = \boldsymbol{S}\boldsymbol{S}^{\dagger}$ on the column space of $\boldsymbol{S}$ and orthogonal projector $\boldsymbol{P}^{\perp} = (\boldsymbol{I}_{MN} - \boldsymbol{S}\boldsymbol{S}^{\dagger})$. As shown in Fig. 1 this leads to an indefinite estimate for moderate to low B, as it is only an asymptotically optimum approach.

The estimate of the noise variance is obtained projecting the observations $\boldsymbol{y}[q]$ on the noise subspace, which does not contain the (training) signal. The estimate of the channel covariance matrix is unbiased, as the second term in (11) compensates for the bias in the first term, which is identical to the heuristic estimator in Sec. 4.4 [1].

For further insights in the least squares approximation we rewrite the cost function introducing $\tilde{\boldsymbol{C}}_{\boldsymbol{y}} = (\boldsymbol{P}^{\perp} + \boldsymbol{P})\tilde{\boldsymbol{C}}_{\boldsymbol{y}}(\boldsymbol{P}^{\perp} + \boldsymbol{P})$ and $\boldsymbol{I}_{MN} = \boldsymbol{P} + \boldsymbol{P}^{\perp}$

$$\|\tilde{\boldsymbol{C}}_{\boldsymbol{y}} - \boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}} - c_{\boldsymbol{n}}\boldsymbol{I}_{MN}\|_{\mathrm{F}}^2 = \|\boldsymbol{P}\tilde{\boldsymbol{C}}_{\boldsymbol{y}}\boldsymbol{P} - \boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}} - c_{\boldsymbol{n}}\boldsymbol{P}\|_{\mathrm{F}}^2 + \|\boldsymbol{P}^{\perp}\tilde{\boldsymbol{C}}_{\boldsymbol{y}}\boldsymbol{P}^{\perp} - c_{\boldsymbol{n}}\boldsymbol{P}^{\perp}\|_{\mathrm{F}}^2, \quad (12)$$

where we also used the property $\|\boldsymbol{A}+\boldsymbol{B}\|_{\mathrm{F}}^2 = \|\boldsymbol{A}\|_{\mathrm{F}}^2 + \|\boldsymbol{B}\|_{\mathrm{F}}^2$ of two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ with inner product $\mathrm{tr}\left[\boldsymbol{A}^{\mathrm{H}}\boldsymbol{B}\right] = 0$. The first term in (12) can be forced to zero for any c_n choosing $\boldsymbol{C}_{\boldsymbol{h}} = \boldsymbol{S}^{\dagger}\tilde{\boldsymbol{C}}_{\boldsymbol{y}}\boldsymbol{S}^{\dagger,\mathrm{H}} - (\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S})^{-1}c_n$, which yields (11). With this choice problem (8) is equivalent to

$$\min_{c_{\boldsymbol{n}}} \|\boldsymbol{P}^{\perp}\tilde{\boldsymbol{C}}_{\boldsymbol{y}}\boldsymbol{P}^{\perp} - c_{\boldsymbol{n}}\boldsymbol{P}^{\perp}\|_{\mathrm{F}}^2, \quad (13)$$

which leads to (10).

[1] In case of complex Gaussian distributed observations.

[2] If a weighted Frobenius norm is used ML performance is achieved asymptotically.

4.2. Positive Semidefinite Least Squares Estimator

In (8) the optimization problem was solved for all quadratic matrices $\boldsymbol{C_h} \in \mathbb{C}^{MK\times MK}$. We propose to include constraints on $\boldsymbol{C_h}$ and $c_{\boldsymbol{n}}$ to be positive semidefinite (psd) into the optimization:

$$\min_{\boldsymbol{C_h}, c_{\boldsymbol{n}}} \|\tilde{\boldsymbol{C}}_{\boldsymbol{y}} - \boldsymbol{S}\boldsymbol{C_h}\boldsymbol{S}^{\mathrm{H}} - c_{\boldsymbol{n}}\boldsymbol{I}_{MN}\|_{\mathrm{F}}^2 \text{ s.t. } \boldsymbol{C_h} \succeq 0, c_{\boldsymbol{n}} \geq 0 \tag{14}$$

This can be solved analytically using Karush-Kuhn-Tucker conditions [7] and the parameterization of $\boldsymbol{C_h}$ via its EVD $\boldsymbol{C_h} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}}$ and corresponding constraints on the eigenvectors $\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U} = \boldsymbol{I}_{MK}$ and eigenvalues $\lambda_i \geq 0$.

Here, we proceed based on (12) with a more intuitive reasoning. We define $\boldsymbol{S}' = \boldsymbol{S}(\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S})^{-1/2}$ with orthonormalized columns, i.e., $\boldsymbol{S}'^{\mathrm{H}}\boldsymbol{S}' = \boldsymbol{I}_{MK}$, which results in $\boldsymbol{P} = \boldsymbol{S}'\boldsymbol{S}'^{\mathrm{H}}$. We rewrite (12) using the EVD of

$$\boldsymbol{S}'^{,\mathrm{H}}\tilde{\boldsymbol{C}}_{\boldsymbol{y}}\boldsymbol{S}' = \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{H}} \tag{15}$$

with $\boldsymbol{\Sigma} = \mathbf{diag}[\sigma_1, \sigma_2, \ldots, \sigma_{MK}]$, $\sigma_i \geq \sigma_{i+1} \geq 0$, and of

$$\boldsymbol{X} = (\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S})^{1/2}\boldsymbol{C_h}(\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S})^{1/2} = \boldsymbol{U}_{\mathrm{x}}\boldsymbol{D}_{\mathrm{x}}\boldsymbol{U}_{\mathrm{x}}^{\mathrm{H}} \tag{16}$$

with $\boldsymbol{D}_{\mathrm{x}} = \mathbf{diag}[d_1, d_2, \ldots, d_{MK}]$

$$\begin{aligned}
&\|\tilde{\boldsymbol{C}}_{\boldsymbol{y}} - \boldsymbol{S}\boldsymbol{C_h}\boldsymbol{S}^{\mathrm{H}} - c_{\boldsymbol{n}}\boldsymbol{I}_{MN}\|_{\mathrm{F}}^2 \\
&\quad = \|\boldsymbol{S}'(\boldsymbol{S}'^{\mathrm{H}}\tilde{\boldsymbol{C}}_{\boldsymbol{y}}\boldsymbol{S}' - \boldsymbol{X} - c_{\boldsymbol{n}}\boldsymbol{I}_{MK})\boldsymbol{S}'^{\mathrm{H}}\|_{\mathrm{F}}^2 + \\
&\qquad + \|\boldsymbol{P}^{\perp}\tilde{\boldsymbol{C}}_{\boldsymbol{y}}\boldsymbol{P}^{\perp} - c_{\boldsymbol{n}}\boldsymbol{P}^{\perp}\|_{\mathrm{F}}^2 \\
&= \|\boldsymbol{S}'^{\mathrm{H}}\tilde{\boldsymbol{C}}_{\boldsymbol{y}}\boldsymbol{S}' - \boldsymbol{X} - c_{\boldsymbol{n}}\boldsymbol{I}_{MK}\|_{\mathrm{F}}^2 + \|\boldsymbol{P}^{\perp}\tilde{\boldsymbol{C}}_{\boldsymbol{y}}\boldsymbol{P}^{\perp} - c_{\boldsymbol{n}}\boldsymbol{P}^{\perp}\|_{\mathrm{F}}^2 \\
&= \sum_{i=1}^{MK}(\sigma_i - d_i - c_{\boldsymbol{n}})^2 + \|\boldsymbol{P}^{\perp}\tilde{\boldsymbol{C}}_{\boldsymbol{y}}\boldsymbol{P}^{\perp} - c_{\boldsymbol{n}}\boldsymbol{P}^{\perp}\|_{\mathrm{F}}^2.
\end{aligned} \tag{17}$$

The last step is due to the definition of the Frobenius norm and $\boldsymbol{U}_{\mathrm{x}} = \boldsymbol{V}$, which is the optimum choice of eigenvectors. Completing the squares the optimization problem (14) reads

$$\begin{aligned}
\min_{\{d_i\}_{i=1}^{MK}, c_{\boldsymbol{n}}} \; & \sum_{i=1}^{MK}(\sigma_i - d_i - c_{\boldsymbol{n}})^2 + \\
& + \mathrm{tr}\left[\boldsymbol{P}^{\perp}\right]\left(c_{\boldsymbol{n}} - \mathrm{tr}\left[\boldsymbol{P}^{\perp}\tilde{\boldsymbol{C}}_{\boldsymbol{y}}\boldsymbol{P}^{\perp}\right]/\mathrm{tr}\left[\boldsymbol{P}^{\perp}\right]\right)^2 \\
& \text{s.t. } d_i \geq 0, c_{\boldsymbol{n}} \geq 0.
\end{aligned} \tag{18}$$

The term in the second line is equivalent to the cost function in (13). Problem (18) can be solved based on the following observations:

1. If the unconstrained LS solution (11) is positive semidefinite, i.e., $d_i \geq 0$ for all i, it is the solution to (18).

2. The constraints on d_i are either active or inactive corresponding to d_i being zero or positive, i.e., 2^{MK} possibilities should be checked in general. These can be reduced exploiting the order of σ_i, which is our focus in the sequel.

3. As a first step we could set $d_j = 0$ for all j with $\sigma_j - \hat{c}_{\boldsymbol{n}}^{\mathrm{LS}} < 0$. All indices j with $\sigma_j - \hat{c}_{\boldsymbol{n}}^{\mathrm{LS}} < 0$ are collected in the set $\mathcal{Z} \subseteq \{1, 2, \ldots, MK\}$ with cardinality $Z = |\mathcal{Z}|$. For the remaining indices we choose $d_i = \sigma_j - c_{\boldsymbol{n}}$. Thus the cost function in (18) is

$$\sum_{i\in\mathcal{Z}}(\sigma_i - c_{\boldsymbol{n}})^2 + + \mathrm{tr}\left[\boldsymbol{P}^{\perp}\right]\left(c_{\boldsymbol{n}} - \mathrm{tr}\left[\boldsymbol{P}^{\perp}\tilde{\boldsymbol{C}}_{\boldsymbol{y}}\boldsymbol{P}^{\perp}\right]/\mathrm{tr}\left[\boldsymbol{P}^{\perp}\right]\right)^2. \tag{19}$$

Minimization w.r.t. $c_{\boldsymbol{n}}$ yields

$$\hat{c}_{\boldsymbol{n}}^{\mathrm{psd}} = \frac{1}{\mathrm{tr}\left[\boldsymbol{P}^{\perp}\right] + Z}\left(\mathrm{tr}\left[\boldsymbol{P}^{\perp}\tilde{\boldsymbol{C}}_{\boldsymbol{y}}\boldsymbol{P}^{\perp}\right] + \sum_{i\in\mathcal{Z}}\sigma_i\right) \tag{20}$$

$$= \frac{1}{M(N-K)+Z}\left(M(N-K)\,\hat{c}_{\boldsymbol{n}}^{\mathrm{LS}} + \sum_{i\in\mathcal{Z}}\sigma_i\right). \tag{21}$$

We have $\hat{c}_{\boldsymbol{n}}^{\mathrm{psd}} \leq \hat{c}_{\boldsymbol{n}}^{\mathrm{LS}}$ with strict inequality in case $Z > 0$. Thus, also a part of the signal subspace is considered as noise subspace.

4. But the cost function (19) could be reduced further, if only fewer d_i are chosen zero. This is possible, if we start with the smallest σ_i, i.e., $i = MK$, and check whether $\sigma_i - \hat{c}_{\boldsymbol{n}}^{\mathrm{LS}} < 0$. If negative we set $d_{MK} = 0$. This decreases $\hat{c}_{\boldsymbol{n}}^{\mathrm{psd}}$ based on $\mathcal{Z} = \{MK\}$. We continue with $j = MK-1$ and check if $\sigma_{MK-1} - \hat{c}_{\boldsymbol{n}}^{\mathrm{psd}} < 0$. $\hat{c}_{\boldsymbol{n}}^{\mathrm{psd}}$ is recomputed based on (21) in every step. As $\hat{c}_{\boldsymbol{n}}^{\mathrm{psd}}$ decreases, generally fewer $\sigma_i - \hat{c}_{\boldsymbol{n}}^{\mathrm{psd}}$ are negative than for $\sigma_i - \hat{c}_{\boldsymbol{n}}^{\mathrm{LS}}$. We continue with decreasing i until $\sigma_i - \hat{c}_{\boldsymbol{n}}^{\mathrm{psd}} \geq 0$, for which we choose $d_i = \sigma_i - \hat{c}_{\boldsymbol{n}}^{\mathrm{psd}}$.

5. Thus, the number $Z = |\mathcal{Z}|$ of zero d_i is minimized, i.e., fewer terms appear in the cost function (19).

These observations yield Algorithm 1, which provides a computationally efficient way to deal with optimization problem (18). The estimates $\hat{c}_{\boldsymbol{n}}^{\mathrm{psd}}$ and $\hat{\boldsymbol{C}}_{\boldsymbol{h}}^{\mathrm{psd}}$ are biased as long as the probability for any $\sigma_i - \hat{c}_{\boldsymbol{n}}^{\mathrm{LS}} < 0$ is non-zero, which is more likely for small B, large noise variance $c_{\boldsymbol{n}}$, and higher correlations in $\boldsymbol{C_h}$.

As $\hat{c}_{\boldsymbol{n}}^{\mathrm{psd}}$ underestimates the noise variance $c_{\boldsymbol{n}}$, the following heuristic achieves better results, when applied to MMSE channel estimation:

Algorithm 1 Positive semidefinite estimate of channel and noise covariance matrix.

$Z = 0$, $\mathcal{Z} = \{\}$
$\hat{c}_{\boldsymbol{n}} = \hat{c}_{\boldsymbol{n}}^{\mathrm{LS}}$
$\boldsymbol{S}' = \boldsymbol{S}(\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S})^{-1/2}$
compute EVD $\boldsymbol{S}'^{\mathrm{H}}\tilde{\boldsymbol{C}}_{\boldsymbol{y}}\boldsymbol{S}' = \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{H}}$
$\boldsymbol{\Sigma} = \mathbf{diag}\,[\sigma_1, \sigma_2, \ldots, \sigma_{MK}]$, $\sigma_j \geq \sigma_{j+1} \forall j$
for $i = MK, MK-1, \ldots, 1$ **do**
  if $\sigma_i - \hat{c}_{\boldsymbol{n}} < 0$ **then**
    $d_i = 0$
    $Z \leftarrow Z + 1$
    $\mathcal{Z} \leftarrow \mathcal{Z} \cup \{i\}$
    $\hat{c}_{\boldsymbol{n}} = \frac{1}{M(N-K)+Z}\left(M(N-K)\,\hat{c}_{\boldsymbol{n}}^{\mathrm{LS}} + \sum_{i \in \mathcal{Z}} \sigma_i\right)$
  else
    $d_i = \sigma_i - \hat{c}_{\boldsymbol{n}}$
  end if
end for
$\boldsymbol{D} = \mathbf{diag}\,[d_1, d_2, \ldots, d_{MK}]$
$\hat{c}_{\boldsymbol{n}}^{\mathrm{psd}} = \hat{c}_{\boldsymbol{n}}$
$\hat{\boldsymbol{C}}_{\boldsymbol{h}}^{\mathrm{psd}} = (\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S})^{-1/2}\boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^{\mathrm{H}}(\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S})^{-1/2}$

1. Estimate $c_{\boldsymbol{n}}$ using $\hat{c}_{\boldsymbol{n}}^{\mathrm{LS}}$ (10), which is unbiased and based on the true noise subspace.

2. Solve the optimization problem

$$\min_{\{d_i\}_{i=1}^{MK}} \sum_{i=1}^{MK} (\sigma_i - d_i - \hat{c}_{\boldsymbol{n}}^{\mathrm{LS}})^2 \text{s.t. } d_i \geq 0, \quad (22)$$

i.e., set all $d_i = 0$ for $\sigma_i - \hat{c}_{\boldsymbol{n}}^{\mathrm{LS}} < 0$. The solution reads

$$\hat{\boldsymbol{C}}_{\boldsymbol{h}}^{\mathrm{psd2}} = (\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S})^{-1/2}\boldsymbol{V}\boldsymbol{D}^{+}\boldsymbol{V}^{\mathrm{H}}(\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S})^{-1/2}, \quad (23)$$

where $\boldsymbol{D}^{+}$ performs $\max(0, d_i)$ for all elements in $\boldsymbol{D}$ from the EVD of

$$(\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S})^{1/2}\hat{\boldsymbol{C}}_{\boldsymbol{h}}^{\mathrm{LS}}(\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S})^{1/2} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^{\mathrm{H}}. \quad (24)$$

For $\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S} \propto \boldsymbol{I}_{MK}$ this is equivalent to discarding the negative definite part of the estimate $\hat{\boldsymbol{C}}_{\boldsymbol{h}}^{\mathrm{LS}}$ (11) similar to [8].

The additional complexity for computing the positive definite solution compared to (10) and (11) results from the EVD of $\boldsymbol{S}'^{\mathrm{H}}\tilde{\boldsymbol{C}}_{\boldsymbol{y}}\boldsymbol{S}'$ and computation of $(\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S})^{1/2}$ and $(\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S})^{-1/2}$. For the indefinite least-squares estimate (11) a tracking-algorithm of very low-complexity was presented by [9], whereas tracking of eigenvalues and eigenvectors is more difficult and complex.

4.3. Biased Estimator using Tikhonov Regularization

In a third approach we propose to employ Tikhonov regularization [10, 7] with regularization parameters α and β. First we constrain the norm of $c_{\boldsymbol{n}}$ and, thus, decreases the negative eigenvalues of $\hat{\boldsymbol{C}}_{\boldsymbol{h}}$ in magnitude:

$$\min_{\boldsymbol{C}_{\boldsymbol{h}}, c_{\boldsymbol{n}}} \|\tilde{\boldsymbol{C}}_{\boldsymbol{y}} - \boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}} - c_{\boldsymbol{n}}\boldsymbol{I}_{MN}\|_{\mathrm{F}}^2 + \\ + \alpha\,\|\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}}\|_{\mathrm{F}}^2 + \beta\,|c_{\boldsymbol{n}}|^2. \quad (25)$$

Additionally, as the MMSE estimator (6) is invariant to a common scaling in $\boldsymbol{C}_{\boldsymbol{h}}$ and $c_{\boldsymbol{n}}$, the first regularization term with α is introduced to balance the norm of $\hat{\boldsymbol{C}}_{\boldsymbol{h}}$ and $\hat{c}_{\boldsymbol{n}}$. The estimator is biased and the solution is computationally as simple as the unbiased LS (8):

$$\hat{c}_{\boldsymbol{n}}^{\mathrm{R}} = \frac{1+\alpha}{(1+\alpha)(MN+\beta) - MK} \times \\ \times \operatorname{trace}\left[\tilde{\boldsymbol{C}}_{\boldsymbol{y}} - \frac{1}{1+\alpha}\boldsymbol{P}\tilde{\boldsymbol{C}}_{\boldsymbol{y}}\boldsymbol{P}\right] \quad (26)$$

$$\hat{\boldsymbol{C}}_{\boldsymbol{h}}^{\mathrm{R}} = \frac{1}{1+\alpha}\left(\boldsymbol{S}^{\dagger}\tilde{\boldsymbol{C}}_{\boldsymbol{y}}\boldsymbol{S}^{\dagger,\mathrm{H}} - (\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S})^{-1}\hat{c}_{\boldsymbol{n}}\right). \quad (27)$$

4.4. Heuristic Estimator

For $\beta \to \infty$ and $\alpha = 0$ we obtain a justification for the heuristic estimator proposed by [3], i.e., it minimizes

$$\min_{\boldsymbol{C}_{\boldsymbol{h}}} \|\tilde{\boldsymbol{C}}_{\boldsymbol{y}} - \boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}}\|_{\mathrm{F}}^2. \quad (28)$$

Implicitly the noise variance is assumed small and neglected. Its solution is the sample-mean of the least-square channel estimates $\hat{\boldsymbol{h}}^{\mathrm{LS}}[q] = \boldsymbol{S}^{\dagger}\boldsymbol{y}[q]$

$$\hat{\boldsymbol{C}}_{\boldsymbol{h}}^{\mathrm{Heur}} = \boldsymbol{S}^{\dagger}\tilde{\boldsymbol{C}}_{\boldsymbol{y}}\boldsymbol{S}^{\dagger,\mathrm{H}} = \frac{1}{B}\sum_{q=1}^{B}\hat{\boldsymbol{h}}^{\mathrm{LS}}[q]\hat{\boldsymbol{h}}^{\mathrm{LS}}[q]^{\mathrm{H}}. \quad (29)$$

and is positive semidefinite and biased.

5. GENERALIZATION OF ESTIMATORS TO CORRELATED NOISE

The algorithms presented in the previous section can be easily extended to correlated noise. Here, we present the results for spatially correlated and temporally uncorrelated noise with covariance matrix $\boldsymbol{C}_{\boldsymbol{n}} = \boldsymbol{I}_N \otimes \boldsymbol{C}_{\boldsymbol{n},\mathrm{S}}$.

We rewrite (12) for this noise covariance matrix

$$\|\tilde{\boldsymbol{C}}_{\boldsymbol{y}} - \boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}} - \boldsymbol{I}_{MN} \otimes \boldsymbol{C}_{\boldsymbol{n},\mathrm{S}}\|_{\mathrm{F}}^2 = \\ = \|\boldsymbol{P}\tilde{\boldsymbol{C}}_{\boldsymbol{y}}\boldsymbol{P} - \boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}}\boldsymbol{S}^{\mathrm{H}} - \boldsymbol{P}(\boldsymbol{I}_{MN} \otimes \boldsymbol{C}_{\boldsymbol{n},\mathrm{S}})\boldsymbol{P}\|_{\mathrm{F}}^2 \\ + \|\boldsymbol{P}^{\perp}\tilde{\boldsymbol{C}}_{\boldsymbol{y}}\boldsymbol{P}^{\perp} - \boldsymbol{P}^{\perp}(\boldsymbol{I}_{MN} \otimes \boldsymbol{C}_{\boldsymbol{n},\mathrm{S}})\boldsymbol{P}^{\perp}\|_{\mathrm{F}}^2. \quad (30)$$

As before the first term is zero choosing $\boldsymbol{C_h}$ from the space of general quadratic MK-dimensional matrices as

$$\hat{\boldsymbol{C}}_{\boldsymbol{h}}^{\mathrm{LS}} = \boldsymbol{S}^{\dagger}(\tilde{\boldsymbol{C}}_{\boldsymbol{y}} - \boldsymbol{I}_{MN} \otimes \hat{\boldsymbol{C}}_{\boldsymbol{n},\mathrm{S}}^{\mathrm{LS}})\boldsymbol{S}^{\dagger,\mathrm{H}} \tag{31}$$

given an estimate $\hat{\boldsymbol{C}}_{\boldsymbol{n},\mathrm{S}}^{\mathrm{LS}}$ as described below. Generally, $\hat{\boldsymbol{C}}_{\boldsymbol{h}}^{\mathrm{LS}}$ is indefinite. Now, we can minimize the second term in (30)

$$\min_{\boldsymbol{C}_{\boldsymbol{n},\mathrm{S}}} \|\boldsymbol{P}^{\perp}\tilde{\boldsymbol{C}}_{\boldsymbol{y}}\boldsymbol{P}^{\perp} - \boldsymbol{P}^{\perp}(\boldsymbol{I}_{MN} \otimes \boldsymbol{C}_{\boldsymbol{n},\mathrm{S}})\boldsymbol{P}^{\perp}\|_{\mathrm{F}}^{2}. \tag{32}$$

This yields the estimate of the noise covariance matrix $\boldsymbol{C}_{\boldsymbol{n},\mathrm{S}}$

$$\hat{\boldsymbol{C}}_{\boldsymbol{n},\mathrm{S}}^{\mathrm{LS}} = \frac{1}{N-K}\sum_{n=1}^{N}(\boldsymbol{e}_n^{\mathrm{T}} \otimes \boldsymbol{I}_M)\boldsymbol{P}^{\perp}\tilde{\boldsymbol{C}}_{\boldsymbol{y}}\boldsymbol{P}^{\perp}(\boldsymbol{e}_n \otimes \boldsymbol{I}_M), \tag{33}$$

which is performed in the noise subspace. It is equivalent to the sample mean of the estimated noise

$$\hat{\boldsymbol{n}}[q,n] = (\boldsymbol{e}_n^{\mathrm{T}} \otimes \boldsymbol{I}_M)\boldsymbol{P}^{\perp}\boldsymbol{y}[q] = \boldsymbol{y}[q,n] - \hat{\boldsymbol{H}}^{\mathrm{LS}}[q]\boldsymbol{s}[n] \tag{34}$$

based on the least-squares channel estimate

$$\hat{\boldsymbol{h}}^{\mathrm{LS}}[q] = \mathbf{vec}\left[\hat{\boldsymbol{H}}^{\mathrm{LS}}[q]\right] = \boldsymbol{S}^{\dagger}\boldsymbol{y}[q] \tag{35}$$

which leads to

$$\hat{\boldsymbol{C}}_{\boldsymbol{n},\mathrm{S}}^{\mathrm{LS}} = \frac{1}{B(N-K)}\sum_{n=1}^{N}\sum_{q=1}^{B}\hat{\boldsymbol{n}}[q,n]\hat{\boldsymbol{n}}[q,n]^{\mathrm{H}}. \tag{36}$$

This estimate of the noise covariance matrix is very similar to the well-known ML estimate [11] of the noise covariance matrix, when estimated jointly with the channel $\boldsymbol{h}[q]$. The difference is in the scaling by $N-K$ instead of N, which yields an improved estimate.

A positive semi-definite solution of $\boldsymbol{C_h}$ can be obtained similar to the heuristic introduced at the end of Sec. 4.2.

6. PERFORMANCE EVALUATION

In *scenario 1* a statistically independent zero-mean complex Gaussian channel $\boldsymbol{h}[q] \sim \mathcal{N}_{\mathrm{c}}(\mathbf{0}, \boldsymbol{C_h})$ and noise $\boldsymbol{n}[q] \sim \mathcal{N}_{\mathrm{c}}(\mathbf{0}, c_n\boldsymbol{I}_{MN})$, with $M = K = 8$ and $N = 16$ training symbols is considered. The channel covariance is modeled as $\boldsymbol{C_h} = \boldsymbol{C} \otimes \boldsymbol{C}$ with elements $[\boldsymbol{C}]_{i,j} = \rho^{|i-j|}$ and $\rho = 0.9$ as in [1] for comparison. For $B = 100$ and $c_n = 1$ a significant number of eigenvalues (normalized to the maximum eigenvalue λ_1) of the unbiased LS are negative (Figure 1) and very close in magnitude to the estimated (and scaled) noise-variance (size of loading in MMSE; solid horizontal line). This leads to eigenvalues very close to zero in the inverse of (6) ($\Rightarrow$ Ill-conditioned matrix).

The MSE of the MMSE channel estimator based on the proposed and previous estimators of $\boldsymbol{C_h}$ and c_n is shown in Fig. 2 for $c_n = 1$. The maximum likelihood (ML) channel estimator $\boldsymbol{W} = \boldsymbol{S}^{\dagger}$ (least squares channel estimator) serves as a reference. The MSE is averaged over 100 independent estimates. For the unbiased LS estimator (8) performance already degrades severely for moderate B. The heuristic estimator (28) saturates for large B. The positive definite extensions in (14) and the biased estimator (25) show a performance better than the LS channel estimator for $B \geq 5$ and converge to the unbiased LS approach for high B.

Parameterization of the biased LS estimator (Sec. 4.3) is easier at low SNR—as in this example. A good choice is $\alpha = 10^4$ and $\beta = 10^6$ with a small deviation from the results shown here, but in general adaptation to SNR and B is necessary and not straight forward as for many robust methods. In the results shown here we chose the best α and β for this application via a brute-force grid search.

For $B = 100$ the MSE vs. SNR is shown in Fig. 3. Whenever the MMSE channel estimator shows a significant performance gain over the ML estimator (low SNR and high channel correlations), already a few independent observations are sufficient to achieve performance superior to the ML estimator. If the gap between MMSE and ML estimator is small more realizations are necessary: In Fig. 3 the MMSE estimator is worse than the ML estimator at high SNR. Moreover, the heuristic estimator (Sec. 4.4) is more robust in this case, as the biased estimate can be interpreted as a diagonally dominant loading.

In *scenario 2* we have $M = 8$ and $K = 3$, where the users' mean angles of arrival are at $[-15°, 0°, 15°]$ relatively to the bore side with $3°$ uniform angular spread each. A uniform linear array with half wavelength element spacing is assumed. The spatially correlated interference is modeled as $\boldsymbol{C}_{\boldsymbol{n},\mathrm{S}} = \boldsymbol{C}_{\mathrm{i}} + c_{\mathrm{n}}\boldsymbol{I}_M$ with $c_{\mathrm{n}} = 0.01$ and $\mathrm{tr}\,[\boldsymbol{C}_{\mathrm{i}}] = Mc_{\mathrm{i}}$. $\boldsymbol{C}_{\mathrm{i}}$ is given by 100 random angles of arrival uniformly distributed within $30°$ around the mean of $30°$.

Due to the spatially correlated noise the gains of our new estimators over the heuristic (Sec. 4.4) are substantial, where Fig. 4 shows the MSE versus B at $c_{\mathrm{i}} = 1$ and Fig. 5 versus c_{i} for $B = 100$. $B > 10$ observations are necessary to be superior to ML channel estimation. Unbiased LS refers to (31) and (36), positive definite LS takes only the positive semidefinite part of (31) as discussed in (23).

7. CONCLUSIONS

Knowledge of the second order moments of the channel parameters and noise is important for a wide class of signal processing algorithms. We presented estimators for these moments based on least-squares approximation of the sample covariance matrix of the observed signal, which exploit the linear structure of this covariance matrix. The most

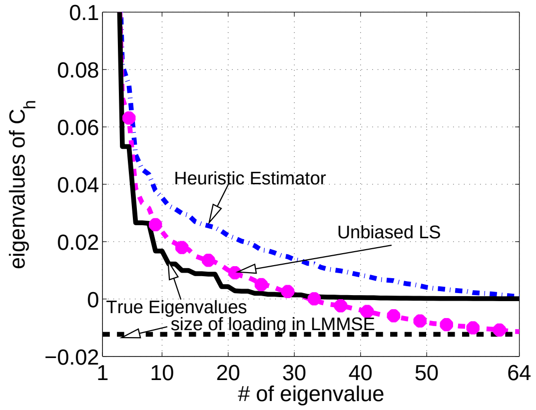

Figure 1. Eigenvalues λ_i/λ_1 of true channel covariance matrix $\boldsymbol{C_h}$ and different estimates $\hat{\boldsymbol{C}}_{\boldsymbol{h}}$ ($B = 100$). Solid horizontal line shows negative value of loading, i.e., $-\hat{c}_n^{\mathrm{LS}}/(N\lambda_1)$ (cf. Eq. 7), in the MMSE estimator to illustrate problem of ill-conditioned inverse in (6).

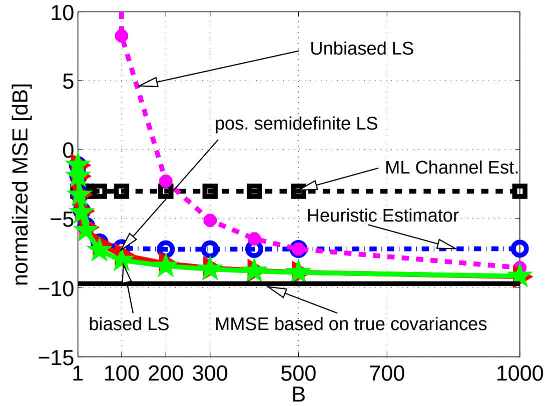

Figure 2. Normalized MSE of MMSE channel estimator (6) for different estimators of $\boldsymbol{C_h}$ and c_n compared to ML channel estimation (Scenario 1, $c_{\mathrm{n}} = 1$).

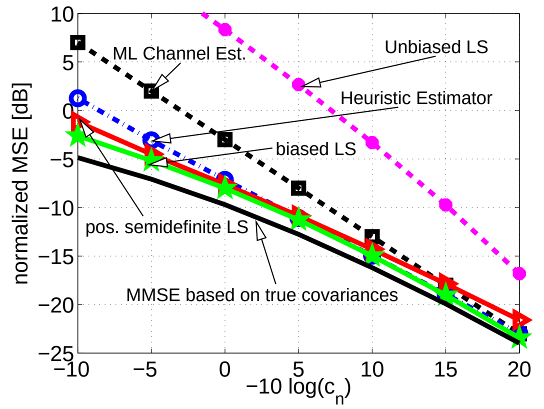

Figure 3. Normalized MSE of MMSE channel estimator (6) for different estimators of $\boldsymbol{C_h}$ and c_n compared to ML channel estimation (Scenario 1, $B = 100$).

promising solution ensures that all estimates are positive semidefinite. Moreover, we introduced extensions for spatially correlated noise. Applying the estimates to MMSE channel estimation, performance gains over conventional ML (least-squares) channel estimation can already be obtained for $B = 5 - 10$ observations at low SNR.

8. REFERENCES

[1] 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*, New York City, U.S.A., June 2005.

[2] T. Ivanov and F. Dietrich, "Estimation of Second Order Channel and Noise Statistics in Wireless Communications," Technical report, Munich University of Technology, April 2005.

[3] F. A. Dietrich and W. Utschick, "Pilot-Assisted Channel Estimation Based on Second-Order Statistics," *IEEE Trans. Signal Processing*, vol. 53, no. 3, pp. 1178–1193, March 2005.

[4] I. Viering, H. Hofstetter, and W. Utschick, "Validity of Spatial Covariance Matrices over Time and Frequency," in *Proc. IEEE Globecom*. Nov. 2002, Taipeh, Taiwan.

[5] P. Stoica and T. Söderström, "On reparametrization of loss functions used in estimation and the invariance principle," *Signal Processing*, vol. 17, no. 4, pp. 383–387, Aug. 1989.

[6] Louis L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, Addison-Wesley, 1991.

[7] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[8] N. J. Higham, *Applications of Matrix Theory*, chapter Matrix Nearness Problems and Applications, pp. 1–27, Oxford University Press, 1989.

[9] G. Matz, "Recursive MMSE Estimation of Wireless Channels based on Training Data and Structured Correlation Learning," in *Proceedings of the 13th Statistical Signal Processing Workshop (SSP)*, Bordeaux, France, July 2005.

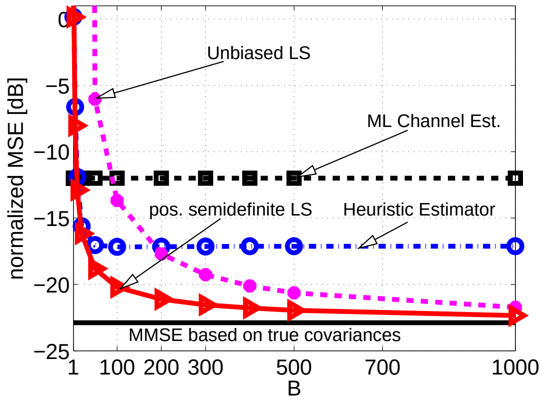

Figure 4. Normalized MSE of MMSE channel estimator (6) for different estimators of $\boldsymbol{C_h}$ and $\boldsymbol{C}_{\boldsymbol{n},\mathrm{S}}$ compared to ML channel estimation (Scenario 2, $c_\mathrm{i} = 1$).

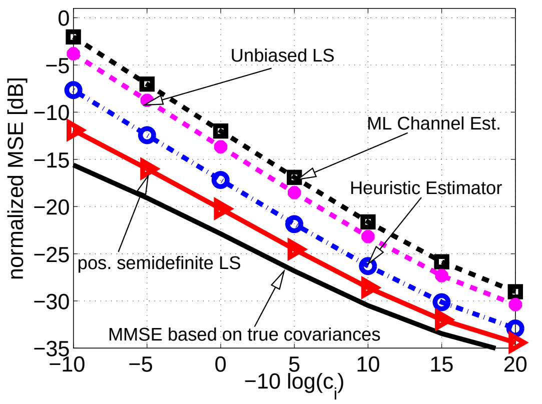

Figure 5. Normalized MSE of MMSE channel estimator (6) for different estimators of $\boldsymbol{C_h}$ and $\boldsymbol{C}_{\boldsymbol{n},\mathrm{S}}$ compared to ML channel estimation (Scenario 2, $B = 100$).

[10] Andrey N. Tikhonov and Vasiliy Y. Arsenin, *Solutions of Ill-Posed Problems*, V. H. Winston & Sons, 1977.

[11] M. Nicoli, M. Sternad, U. Spagnolini, and A. Ahlén, “Reduced-Rank Channel Estimation and Tracking in Time-Slotted CDMA Systems,” in *Proc. IEEE Int. Conference on Communications*. April 2002, pp. 533–537, New York City.