CONDITIONAL MEAN ESTIMATOR FOR THE GRAMIAN MATRIX OF COMPLEX GAUSSIAN RANDOM VARIABLES

Frank A. Dietrich, Felix Hoffmann, 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

The problem of estimating and predicting the Gramian of a matrix (with rather general structure) of correlated complex Gaussian random variables is addressed. We propose its conditional mean estimator as the optimum Bayesian estimator for a quadratic risk function and present its mean square error (MSE) performance analysis. Numerical results for the example of linear pre-equalization in a wireless communications application show a significantly improved performance of the novel estimator compared to known approaches.

1. INTRODUCTION

Channel estimation based on training symbols is a key component in many communication systems. Very often the estimated channel parameters are not required themselves, but a non-linear transformation of them, e.g. channel power for adaptive modulation [1] or the Gramian of a block Toeplitz matrix in pre-equalization or equalizer design [2, 3]. Usually the *properties* of this *non-linear transformation* are not exploited for channel estimation, although the probability distribution of the parameters is changed significantly by a non-linear transformation. *Bayesian estimators*, which make use of this a priori information, can be improved significantly, when taking into account the transformation of the estimates, which is performed by the application.

To motivate the problem statement (Sec. 3), we describe a linear pre-equalizer in a multi-user system based on perfect channel knowledge [2, 3] as an application of our results in communications (Sec. 2). Here, the optimized filter depends on the Gramian of a block Toeplitz matrix of the channel coefficients. In a time-division duplex (TDD) system the channel coefficients can be predicted from the training symbols received in the uplink (reverse link). In Sec. 3 this example is *generalized* to the Gramian of a matrix, which is a general affine function in correlated complex Gaussian random variables. Two common estimation approaches are given as a reference, before deriving the *conditional mean* (CM) estimator of the Gramian without requiring the distribution of the Gramian matrix explicitly (Sec. 4.3). Estimating the squared magnitude of a channel coefficient (channel power) is a very special case of the general problem considered here: For this case an estimator equivalent to our solution was introduced in a heuristic fashion by Ekman et al. [1]. The additional computational complexity to design the CM estimator of the Gramian is small compared to a CM estimator of the channel coefficients.

The MSE of the novel and traditional estimator is derived explicitly (Sec. 5) and evaluated numerically for the example from Sec. 2. Results for different scenarios, values of SNR and system parameters show a significant performance gain for increasing number of users, low SNR, and increased time-variance of the channel parameters.

It has to be noted that the conditional mean estimator of the Gramian presented here allows further insights into the performance and interpretation of robust optimization of pre-equalization or precoding [4]: For example, it yields a *(structured) loading* of the inverse as obtained by *robust optimization* with the paradigm from stochastic programming.

Notation: Random vectors and matrices are denoted by lower and upper case sans serif bold letters (e.g. a, A), whereas the respective realizations or deterministic variables are italic (e.g. $\boldsymbol{a}$, $\boldsymbol{A}$). The operators $\mathrm{E}[\bullet]$, $(\bullet)^{\mathrm{T}}$, $(\bullet)^{\mathrm{H}}$, and $\mathrm{tr}(\bullet)$ stand for expectation, transpose, Hermitian transpose, and trace of a matrix, respectively. $*$ and $\otimes$ denote the convolution and Kronecker product. The $N \times N$ identity matrix is $\boldsymbol{I}_N$ and $\mathbf{0}_{M\times N}$ the $M \times N$ matrix of zeros.

2. EXAMPLE: DESIGN OF PRE-EQUALIZATION

To motivate the problem let us consider the design of a linear pre-equalization filter $\boldsymbol{P} \in \mathbb{C}^{MB\times KB}$ [3] for a wireless communication system with K users and M transmit antennas. A transmitted block of B data symbols for each user $\mathsf{s}_{\mathrm{d}} \in \mathbb{C}^{KB}$ is predistorted by $\boldsymbol{P}$. The received signal reads

$$\mathsf{y}_{\mathrm{d}} = \boldsymbol{\mathcal{H}}_w^{\mathrm{T}} \boldsymbol{P} \mathsf{s}_{\mathrm{d}} + \mathsf{n}_{\mathrm{d}} \in \mathbb{C}^{K(B+L)}, \tag{1}$$

where n_{d} is additive noise with variance σ_{n}^2 and

$$\boldsymbol{\mathcal{H}}_w = \sum_{\ell=0}^{L} \boldsymbol{J}_\ell \otimes \boldsymbol{H}_{w,\ell} \in \mathbb{C}^{MB\times K(B+L)} \tag{2}$$

the block Toeplitz channel matrix for a frequency selective channel $\boldsymbol{H}_w[n] = \sum_{\ell=0}^{L} \boldsymbol{H}_{w,\ell}\delta[n-\ell] \in \mathbb{C}^{M\times K}$ of order L with selection matrix $\boldsymbol{J}_\ell = [\mathbf{0}_{B\times\ell}, \boldsymbol{I}_B, \mathbf{0}_{B+L-\ell}]$ describing its structure. The transmit filter $\boldsymbol{P}$ is optimized together with a scalar receive filter β minimizing the mean square error (MSE) $\mathrm{E}[\|\beta^{-1}\mathbf{y}_\mathrm{d} - \boldsymbol{\Psi}\mathbf{s}_\mathrm{d}\|_2^2]$ under a constraint $\|\boldsymbol{P}\|_\mathrm{F}^2 \leq \mathrm{E}_\mathrm{Tx}$ on the transmit power (assuming uncorrelated symbols of variance one) [2, 3]. For complexity reasons the data-stream is split into blocks of size B [4]. Therefore, we have to suppress the interference generated for the following block described by $\boldsymbol{\Psi} = [\boldsymbol{I}_{BK}, \mathbf{0}_{BK\times LK}]^\mathrm{T}$.

The resulting transmit Wiener filter is

$$\boldsymbol{P} = \beta\left(\boldsymbol{\mathcal{H}}_w^*\boldsymbol{\mathcal{H}}_w^\mathrm{T} + \frac{\sigma_\mathbf{n}^2(B+L)K}{E_\mathrm{Tx}}\boldsymbol{I}_{MB}\right)^{-1}\boldsymbol{\mathcal{H}}_w^*\boldsymbol{\Psi} \quad (3)$$

with β chosen to satisfy the power constraint with equality. The matrix to be inverted depends on the *complex conjugate of the Gramian* of $\boldsymbol{\mathcal{H}}_w$, which leads us to the conclusion that it may be advantageous to *estimate the Gramian* $\boldsymbol{\mathcal{H}}_w\boldsymbol{\mathcal{H}}_w^\mathrm{H}$ directly additionally to estimating $\boldsymbol{\mathcal{H}}_w$ only.

Considering a TDD system, the channel $\boldsymbol{\mathcal{H}}_w$ for designing $\boldsymbol{P}$ can be estimated from the reverse link: N training symbols $\boldsymbol{s}[n] \in \mathbb{B}^K$ with $n \in \{-L,\cdots,N-1\}$ are time multiplexed including L guard symbols. The received training sequence $\mathbf{y}_w[n]$ in time-slot w is

$$\mathbf{y}_w[n] = \mathbf{H}_w[n] * \boldsymbol{s}[n] + \mathbf{n}_w[n] \in \mathbb{C}^M \quad (4)$$

assuming correlated block fading of the channel and can be rewritten as

$$\begin{aligned}\mathbf{y}_w &= \boldsymbol{S}\mathbf{h}_w + \mathbf{n}_w \in \mathbb{C}^{MN}, \text{ where} \quad (5)\\ \mathbf{y}_w &= [\mathbf{y}_w[0]^\mathrm{T},\cdots,\mathbf{y}_w[N-1]^\mathrm{T}]^\mathrm{T},\\ \mathbf{n}_w &= [\mathbf{n}_w[0]^\mathrm{T},\cdots,\mathbf{n}_w[N-1]^\mathrm{T}]^\mathrm{T},\end{aligned}$$

$\boldsymbol{S} = \boldsymbol{S}'^\mathrm{T} \otimes \boldsymbol{I}_M \in \mathbb{C}^{MN\times KM(L+1)}$, $\boldsymbol{S}' \in \mathbb{C}^{K(L+1)\times N}$ is block Toeplitz with first block row $[\boldsymbol{s}[0],\cdots,\boldsymbol{s}[N-1]]$ and first column $[\boldsymbol{s}[0]^\mathrm{T},\cdots,\boldsymbol{s}[-L]^\mathrm{T}]^\mathrm{T}$. The additive noise $\mathbf{n}_w$ is distributed as $\mathcal{N}_\mathrm{c}(\mathbf{0}, \boldsymbol{C}_{\mathbf{n}_w})$. $\mathbf{h}_w$=vec$([\mathbf{H}_{w,0},\cdots,\mathbf{H}_{w,L}]) \in \mathbb{C}^P$ contains all $P = MK(L+1)$ channel coefficients.

This example will be used for the numerical results in Sec. 6. For solving the problem of estimating $\boldsymbol{\mathcal{H}}_w\boldsymbol{\mathcal{H}}_w^\mathrm{H}$ we now switch to a more general notation.

3. PROBLEM STATEMENT

Given a matrix $\boldsymbol{A}$, whose elements are affine functions of

$$\boldsymbol{h} = \mathrm{vec}\left([\boldsymbol{H}_0, \boldsymbol{H}_1, \cdots, \boldsymbol{H}_L]\right) \in \mathbb{C}^P, \quad \boldsymbol{H}_\ell \in \mathbb{C}^{M\times K},$$

$$\text{i.e. } \boldsymbol{A}(\boldsymbol{h}) = \sum_{\ell=0}^{L} \boldsymbol{J}_\ell \otimes \boldsymbol{H}_\ell \quad (6)$$

with $\boldsymbol{J}_\ell$ (definition depends on the problem) describing its structure, we would like to estimate its Gramian matrix $\boldsymbol{G}$

$$\boldsymbol{G}(\boldsymbol{h}) = \boldsymbol{A}(\boldsymbol{h})\boldsymbol{A}(\boldsymbol{h})^\mathrm{H} = \sum_{\ell=0}^{L}\sum_{\ell'=0}^{L} \boldsymbol{J}_\ell\boldsymbol{J}_{\ell'}^\mathrm{H} \otimes \boldsymbol{H}_\ell\boldsymbol{H}_{\ell'}^\mathrm{H}. \quad (7)$$

We assume that $\boldsymbol{h}$ is a circular symmetric complex Gaussian random vector with mean $\mathrm{E}[\mathbf{h}] = \mathbf{0}$ (for simplicity) and covariance matrix $\boldsymbol{C}_\mathbf{h} = \mathrm{E}[\mathbf{h}\mathbf{h}^\mathrm{H}]$ denoted as $\mathbf{h} \sim \mathcal{N}_\mathrm{c}(\mathbf{0}, \boldsymbol{C}_\mathbf{h})$.

Estimation of $\boldsymbol{A}(\boldsymbol{h})\boldsymbol{A}(\boldsymbol{h})^\mathrm{H}$ is based on the (indirect) observation $\mathbf{y} = [\mathbf{y}_1^\mathrm{T}, \mathbf{y}_2^\mathrm{T}, \cdots, \mathbf{y}_W^\mathrm{T}]^\mathrm{T}$ of a sequence of W realizations of a zero-mean random vector $\mathbf{h}_w$ in the linear model

$$\mathbf{y} = \boldsymbol{S}_\mathrm{T}\mathbf{h}_\mathrm{T} + \mathbf{n} \in \mathbb{C}^{WNM}, \text{ where} \quad (8)$$

$\boldsymbol{S}_\mathrm{T} = \boldsymbol{I}_W \otimes \boldsymbol{S} \in \mathbb{C}^{WMN\times WP}$, $\mathbf{h}_\mathrm{T} = [\mathbf{h}_1^\mathrm{T}, \mathbf{h}_2^\mathrm{T}, \cdots, \mathbf{h}_W^\mathrm{T}]^\mathrm{T} \in \mathbb{C}^{WP}$. The cross-covariance matrix between $\mathbf{h}_\mathrm{T}$ and $\mathbf{h}$ is

$$\boldsymbol{C}_{\mathbf{h}_\mathrm{T}\mathbf{h}} = \mathrm{E}[\mathbf{h}_\mathrm{T}\mathbf{h}^\mathrm{H}] \in \mathbb{C}^{WP\times P}. \quad (9)$$

The measurement noise is also circular symmetric complex Gaussian: $\mathbf{n} \sim \mathcal{N}_\mathrm{c}(\mathbf{0}, \boldsymbol{C}_\mathbf{n})$. All first and second order statistics are assumed perfectly known in the sequel.

4. ESTIMATORS FOR THE GRAMIAN $\boldsymbol{G}(\boldsymbol{H})$

4.1. Maximum Likelihood Approach

Due the invariance property of the ML estimator [5], the ML estimate of $\boldsymbol{G}(\boldsymbol{h})$ is given by the transformation of the ML estimate $\hat{\boldsymbol{h}}_\mathrm{ML}$ of $\boldsymbol{h}$:

$$\hat{\boldsymbol{G}}_\mathrm{ML} = \boldsymbol{A}(\hat{\boldsymbol{h}}_\mathrm{ML})\boldsymbol{A}(\hat{\boldsymbol{h}}_\mathrm{ML})^\mathrm{H}. \quad (10)$$

If the probability density of the observation $\mathbf{y}$ (8) is not parameterized by $\boldsymbol{h}$, i.e. $\boldsymbol{h}$ is not contained in $\boldsymbol{h}_\mathrm{T}$, ML estimation is not possible based on this model, as for example in case of prediction (see our scenario in Sec. 6).

4.2. Heuristic Approach

The a priori information about the statistics of $\mathbf{h}$ is optimally exploited for estimation of $\boldsymbol{h}$ by the conditional mean (CM) estimator

$$\hat{\boldsymbol{h}}_\mathrm{CM} = \mathrm{E}[\mathbf{h}|\boldsymbol{y}] = \boldsymbol{W}_\mathrm{CM}\boldsymbol{y} \quad (11)$$

$$\boldsymbol{W}_\mathrm{CM} = \boldsymbol{C}_{\mathbf{h}_\mathrm{T}\mathbf{h}}^\mathrm{H}\boldsymbol{S}_\mathrm{T}^\mathrm{H}(\boldsymbol{S}_\mathrm{T}\boldsymbol{C}_{\mathbf{h}_\mathrm{T}}\boldsymbol{S}_\mathrm{T}^\mathrm{H} + \boldsymbol{C}_\mathbf{n})^{-1} \quad (12)$$

minimizing the average cost for a quadratic risk function [5]. Conventionally, when the Gramian matrix $\boldsymbol{G}(\boldsymbol{h})$ is desired (as in Eqn. 3) the estimate $\hat{\boldsymbol{h}}_\mathrm{CM}$ is simply plugged in the transformation to obtain an estimate:

$$\hat{\boldsymbol{G}}_\mathrm{H} = \boldsymbol{A}(\hat{\boldsymbol{h}}_\mathrm{CM})\boldsymbol{A}(\hat{\boldsymbol{h}}_\mathrm{CM})^\mathrm{H}. \quad (13)$$

This heuristic intuitively applies the ML invariance property to a case, where it does not hold: $\mathbf{h}$ is considered a random vector whose probability distribution is not invariant towards a non-linear transformation such as (7).

4.3. Conditional Mean Estimator

Thus, the estimate $\hat{G}$ can be improved considering the probability distribution of $\mathbf{G}$, which is a Wishart distribution [6]. The conditional mean estimate of $\mathbf{G} = \boldsymbol{A}(\mathbf{h})\boldsymbol{A}(\mathbf{h})^{\mathrm{H}}$

$$\hat{\boldsymbol{G}}_{\mathrm{CM}} = \mathrm{E}[\boldsymbol{G}(\mathbf{h})|\boldsymbol{y}] = \mathrm{E}[\boldsymbol{A}(\mathbf{h})\boldsymbol{A}(\mathbf{h})^{\mathrm{H}}|\boldsymbol{y}] = \sum_{\ell=0}^{L}\sum_{\ell'=0}^{L} \boldsymbol{J}_\ell \boldsymbol{J}_{\ell'}^{\mathrm{H}} \otimes \mathrm{E}[\mathbf{H}_\ell \mathbf{H}_{\ell'}^{\mathrm{H}}|\boldsymbol{y}] \quad (14)$$

minimizes the average cost for the risk function $\|\hat{\boldsymbol{G}} - \mathbf{G}\|_{\mathrm{F}}^2$, i.e. the squared Frobenius norm of the error. With $\boldsymbol{H}_\ell = [\boldsymbol{h}_{1,\ell}, \boldsymbol{h}_{2,\ell}, \cdots, \boldsymbol{h}_{K,\ell}]$ and $\boldsymbol{H}_\ell \boldsymbol{H}_{\ell'}^{\mathrm{H}} = \sum_{k=1}^{K} \boldsymbol{h}_{k,\ell}\boldsymbol{h}_{k,\ell'}^{\mathrm{H}}$ we can express the conditional mean estimate of the Gramian $\boldsymbol{H}_\ell \boldsymbol{H}_\ell^{\mathrm{H}}$ as

$$\mathrm{E}[\mathbf{H}_\ell \mathbf{H}_{\ell'}^{\mathrm{H}}|\boldsymbol{y}] = \sum_{k=1}^{K} \mathrm{E}[\mathbf{h}_{k,\ell}\mathbf{h}_{k,\ell'}^{\mathrm{H}}|\boldsymbol{y}], \quad (15)$$

i.e. the sum of conditional mean estimates of outer products $\boldsymbol{h}_{k,\ell}\boldsymbol{h}_{k,\ell}^{\mathrm{H}}$. Recognizing it as a conditional correlation matrix we do not require the distribution of $\mathbf{G}$ explicitly to derive the estimator, simplifying the derivation tremendously:

$$\mathrm{E}[\mathbf{h}_{k,\ell}\mathbf{h}_{k,\ell'}^{\mathrm{H}}|\boldsymbol{y}] = \mathrm{E}[\mathbf{h}_{k,\ell}|\boldsymbol{y}]\mathrm{E}[\mathbf{h}_{k,\ell'}|\boldsymbol{y}]^{\mathrm{H}} + \boldsymbol{C}_{\mathbf{h}_{k,\ell}\mathbf{h}_{k,\ell'}|\boldsymbol{y}} = \hat{\boldsymbol{h}}_{k,\ell,\mathrm{CM}}\hat{\boldsymbol{h}}_{k,\ell',\mathrm{CM}}^{\mathrm{H}} + \boldsymbol{C}_{\mathbf{h}_{k,\ell}\mathbf{h}_{k,\ell'}|\boldsymbol{y}}. \quad (16)$$

The first term is the outer product of the conditional mean estimate of $\mathbf{h}_{k,\ell}$ and $\mathbf{h}_{k,\ell'}$ as in (13). The conditional covariance matrix

$$\boldsymbol{C}_{\mathbf{h}_{k,\ell}\mathbf{h}_{k,\ell'}|\boldsymbol{y}} = \mathrm{E}[(\mathbf{h}_{k,\ell} - \hat{\boldsymbol{h}}_{k,\ell,\mathrm{CM}})(\mathbf{h}_{k,\ell'} - \hat{\boldsymbol{h}}_{k,\ell',\mathrm{CM}})^{\mathrm{H}}|\boldsymbol{y}]$$

is equal to the error covariance matrix of the conditional mean estimate $\hat{\boldsymbol{h}}_{k,\ell,\mathrm{CM}}$ due to the orthogonality property of the conditional mean estimator *and* jointly (complex) Gaussian random variables [5]. Thus, the error $\mathbf{h}_{k,\ell} - \hat{\mathbf{h}}_{k,\ell,\mathrm{CM}}$ is statistically independent from the observation $\mathbf{y}$, i.e.

$$\boldsymbol{C}_{\mathbf{h}_{k,\ell}\mathbf{h}_{k,\ell'}|\boldsymbol{y}} = \mathrm{E}[(\mathbf{h}_{k,\ell} - \hat{\mathbf{h}}_{k,\ell',\mathrm{CM}})(\mathbf{h}_{k,\ell} - \hat{\mathbf{h}}_{k,\ell',\mathrm{CM}})^{\mathrm{H}}],$$

which is a submatrix of

$$\boldsymbol{C}_{\mathbf{h}|\boldsymbol{y}} = \boldsymbol{C}_{\mathbf{h}} - \boldsymbol{W}_{\mathrm{CM}}\boldsymbol{S}_{\mathrm{T}}\boldsymbol{C}_{\mathbf{h}_{\mathrm{T}}\mathbf{h}}, \quad (17)$$

which directly depends on the CM estimator $\boldsymbol{W}_{\mathrm{CM}}$ (12) of $\boldsymbol{h}$. Summarizing the derivations, the CM estimator of $\boldsymbol{G}(\mathbf{h})$ (14) is given by

$$\hat{\boldsymbol{G}}_{\mathrm{CM}} = \mathrm{E}[\boldsymbol{G}(\mathbf{h})|\boldsymbol{y}] = \boldsymbol{A}(\hat{\boldsymbol{h}}_{\mathrm{CM}})\boldsymbol{A}(\hat{\boldsymbol{h}}_{\mathrm{CM}})^{\mathrm{H}} + \boldsymbol{C}_{\boldsymbol{A}(\mathbf{h})|\boldsymbol{y}}, \quad (18)$$

i.e. the Gramian based on the conditional mean estimate (11) plus the error covariance matrix for estimating $\boldsymbol{A}$ (6):

$$\boldsymbol{C}_{\boldsymbol{A}(\mathbf{h})|\boldsymbol{y}} = \mathrm{E}\left[\Big(\boldsymbol{A}(\mathbf{h}) - \boldsymbol{A}(\hat{\boldsymbol{h}}_{\mathrm{CM}})\Big)\Big(\boldsymbol{A}(\mathbf{h}) - \boldsymbol{A}(\hat{\boldsymbol{h}}_{\mathrm{CM}})\Big)^{\mathrm{H}}\middle|\,\boldsymbol{y}\right] = \sum_{\ell=0}^{L}\sum_{\ell'=0}^{L} \boldsymbol{J}_\ell \boldsymbol{J}_{\ell'}^{\mathrm{H}} \otimes \boldsymbol{C}_{\mathbf{h}_{k,\ell}\mathbf{h}_{k,\ell'}|\boldsymbol{y}}. \quad (19)$$

The additional *computational complexity* of the CM estimator (18) compared to the heuristic approach (13) is very small: some additional matrix multiplications – assuming the estimator in (11) is given – are needed to compute $\boldsymbol{C}_{\mathbf{h}|\boldsymbol{y}}$, whose submatrices are $\boldsymbol{C}_{\mathbf{h}_{k,\ell}\mathbf{h}_{k,\ell'}|\boldsymbol{y}}$.

5. PERFORMANCE ANALYSIS

An analytic expression for the mean square error (MSE) of the conditional mean estimator w.r.t. squared Frobenius norm $\|\bullet\|_{\mathrm{F}}^2$ of $\mathbf{E} = \hat{\mathbf{G}}_{\mathrm{CM}} - \boldsymbol{G}(\mathbf{h})$ (cf. Eqns. 14 and 18), i.e.

$$\sigma_{\mathbf{E}}^2 = \mathrm{E}[\|\hat{\mathbf{G}}_{\mathrm{CM}} - \boldsymbol{G}(\mathbf{h})\|_{\mathrm{F}}^2] = \mathrm{E}[\mathrm{tr}(\boldsymbol{G}(\mathbf{h})\boldsymbol{G}(\mathbf{h})^{\mathrm{H}})] - 2\mathrm{E}[\mathrm{tr}(\boldsymbol{G}(\mathbf{h})\hat{\mathbf{G}}_{\mathrm{CM}}))] + \mathrm{E}[\mathrm{tr}(\hat{\mathbf{G}}_{\mathrm{CM}}\hat{\mathbf{G}}_{\mathrm{CM}}^{\mathrm{H}})], \quad (20)$$

can be derived using results from [7] on higher order moments of complex Gaussian random variables. For zero mean $\mathbf{h}$ and $\mathbf{h}_{\mathrm{T}}$ we get

$$\begin{aligned}\sigma_{\mathbf{E}}^2 = \sum_{\ell,\ell'=0}^{L}\sum_{\alpha,\alpha'=0}^{L}\sum_{k,\gamma=1}^{K} &\mathrm{tr}(\boldsymbol{J}_\ell\boldsymbol{J}_{\ell'}\boldsymbol{J}_\alpha\boldsymbol{J}_{\alpha'})\times \quad (21)\\ \times[&\mathrm{tr}(\boldsymbol{C}_{\mathbf{h}_{\gamma,\alpha}\mathbf{h}_{k,\ell'}})\mathrm{tr}(\boldsymbol{C}_{\mathbf{h}_{\gamma,\alpha'}\mathbf{h}_{k,\ell}}) + \mathrm{tr}(\boldsymbol{C}_{\mathbf{h}_{k,\ell}\mathbf{h}_{k,\ell'}}\boldsymbol{C}_{\mathbf{h}_{\gamma,\alpha}\mathbf{h}_{\gamma,\alpha'}})\\ -2&\mathrm{tr}(\boldsymbol{C}_{\hat{\mathbf{h}}_{\gamma,\alpha}\mathbf{h}_{k,\ell'}})\mathrm{tr}(\boldsymbol{C}_{\mathbf{h}_{\gamma,\alpha'}\hat{\mathbf{h}}_{k,\ell}}) - 2\mathrm{tr}(\boldsymbol{C}_{\mathbf{h}_{k,\ell}\mathbf{h}_{k,\ell'}}\boldsymbol{C}_{\hat{\mathbf{h}}_{\gamma,\alpha}\hat{\mathbf{h}}_{\gamma,\alpha'}})\\ +&\mathrm{tr}(\boldsymbol{C}_{\hat{\mathbf{h}}_{\gamma,\alpha}\hat{\mathbf{h}}_{k,\ell'}})\mathrm{tr}(\boldsymbol{C}_{\hat{\mathbf{h}}_{\gamma,\alpha'}\hat{\mathbf{h}}_{k,\ell}}) + \mathrm{tr}(\boldsymbol{C}_{\hat{\mathbf{h}}_{k,\ell}\hat{\mathbf{h}}_{k,\ell'}}\boldsymbol{C}_{\hat{\mathbf{h}}_{\gamma,\alpha}\hat{\mathbf{h}}_{\gamma,\alpha'}})\\ +&\mathrm{tr}(\boldsymbol{C}_{\hat{\mathbf{h}}_{k,\ell}\hat{\mathbf{h}}_{k,\ell'}}\boldsymbol{C}_{\hat{\mathbf{h}}_{\gamma,\alpha}\hat{\mathbf{h}}_{\gamma,\alpha'}}) + \mathrm{tr}(\boldsymbol{C}_{\hat{\mathbf{h}}_{k,\ell}\hat{\mathbf{h}}_{k,\ell'}}\boldsymbol{C}_{\mathbf{h}_{\gamma,\alpha}\mathbf{h}_{\gamma,\alpha'}|\boldsymbol{y}})\\ +&\mathrm{tr}(\boldsymbol{C}_{\hat{\mathbf{h}}_{\gamma,\alpha}\hat{\mathbf{h}}_{\gamma,\alpha'}}\boldsymbol{C}_{\mathbf{h}_{k,\ell}\mathbf{h}_{k,\ell'}|\boldsymbol{y}}) + \mathrm{tr}(\boldsymbol{C}_{\mathbf{h}_{k,\ell}\mathbf{h}_{k,\ell'}|\boldsymbol{y}}\boldsymbol{C}_{\mathbf{h}_{\gamma,\alpha}\mathbf{h}_{\gamma,\alpha'}|\boldsymbol{y}})],\end{aligned}$$

where the necessary covariance matrices are submatrices of $\boldsymbol{C}_{\mathbf{h}}$, $\boldsymbol{C}_{\mathbf{h}|\boldsymbol{y}}$ from (17), $\boldsymbol{C}_{\hat{\mathbf{h}}\mathbf{h}} = \boldsymbol{W}_{\mathrm{CM}}\boldsymbol{S}_{\mathrm{T}}\boldsymbol{C}_{\mathbf{h}_{\mathrm{T}}\boldsymbol{h}}$ and $\boldsymbol{C}_{\hat{\mathbf{h}}} = \boldsymbol{W}_{\mathrm{CM}}\boldsymbol{S}_{\mathrm{T}}\boldsymbol{C}_{\mathbf{h}_{\mathrm{T}}\boldsymbol{h}}\boldsymbol{S}_{\mathrm{T}}^{\mathrm{H}}\boldsymbol{W}_{\mathrm{CM}}^{\mathrm{H}} + \boldsymbol{W}_{\mathrm{CM}}\boldsymbol{C}_{\mathbf{n}}\boldsymbol{W}_{\mathrm{CM}}^{\mathrm{H}}$. The MSE for the heuristic approach (13) is given from (21) for $\boldsymbol{C}_{\mathbf{h}|\boldsymbol{y}} = \mathbf{0}$.

6. NUMERICAL RESULTS FOR THE EXAMPLE

Returning to our example in Sec. 2, we now set $\mathbf{h} = \mathbf{h}_w$, $\mathbf{h}_{\mathrm{T}} = [\mathbf{h}_{w-3}^{\mathrm{T}}, \mathbf{h}_{w-5}^{\mathrm{T}}, \cdots, \mathbf{h}_{w-(2W+1)}^{\mathrm{T}}]^{\mathrm{T}}$, and $\boldsymbol{\mathcal{H}}_w = \boldsymbol{A}(\boldsymbol{h}_w)$ (Eqn. 2). Thus, we model a TDD system with alternating up-/downlink slots and a delay of 3 slots (due to processing the training sequence [4]) to the first slot available with a training sequence. $W = 5$ previous uplink slots are considered for prediction of the Gramian $\boldsymbol{G} = \boldsymbol{\mathcal{H}}_w\boldsymbol{\mathcal{H}}_w^{\mathrm{H}}$.

Model parameters – unless otherwise stated: A system with $K = 6$ users and an uniform linear array of $M = 8$ elements with half-wavelength spacing, $N = 30$ QPSK training symbols, a channel of order $L = 3$ with exponential power delay profile ($\propto \exp(-\ell/1.28)$) is considered. Channel coefficients of different users and delays are uncorrelated and spatial correlations are modeled by Laplace distributed angles of arrival with angular spread $\sigma_{\Delta\phi} = 10°$ around a uniformly distributed mean angle per tap and user [8]. Equal temporal correlations are assumed for all coefficients given by a Jakes power spectrum with a Doppler frequency $f_\mathrm{d} = 0.12$ normalized by the slot period, e.g. $1/1500\,\mathrm{s}$ as in TD-SCDMA. For pre-equalization (3) data is split into blocks of length $B = 9$, i.e. $\boldsymbol{\mathcal{H}}_w \in \mathbb{C}^{72\times 72}$ (2).

All figures show the MSE given by (21) for predicting the Gramian $\boldsymbol{G} = \boldsymbol{\mathcal{H}}_w \boldsymbol{\mathcal{H}}_w^\mathrm{H}$ in (3) *normalized* by the number of elements in $\boldsymbol{G}$. Due to its (implicit) a priori information about the distribution of the Gramian, the CM estimator (18) gains significantly over the heuristic approach (13), whenever predicting $\boldsymbol{G}$ is difficult: For low and moderate SNR (Fig. 1), large number of users K ($\Rightarrow$ high system load/more parameters, Fig. 2), and high Doppler frequency ($\Rightarrow$ low temporal correlations, Fig. 3) the CM estimator should be employed. Note, that it relies on the *same* information as the CM estimator of $\boldsymbol{h}$ (12) and requires only a small amount of additional complexity.

Moreover, the gain is larger for lower spatial correlation (Fig. 1) in the interesting SNR range, whereas for very low SNR the order is reversed.

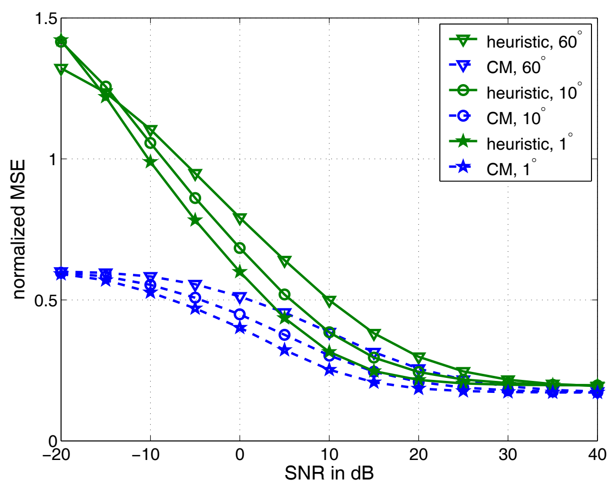

Fig. 1. MSE vs. SNR for different spatial correlations given by angular spread: $\{1°, 10°, 60°\}$ ($f_\mathrm{d} = 0.12$, $K = 6$).

7. REFERENCES

[1] T. Ekman, M. Sternad, and A. Athlén, “Unbiased power predicition of rayleigh fading channels,” in *Proc. VTC Fall 2002*, Vancouver, Canada, Sept. 2002.

[2] M. Joham, K. Kusume, M. H. Gzara, W. Utschick, and J. A. Nossek, “Transmit Wiener Filter for the Downlink of TDD DS-CDMA Systems,” in *Proc. ISSSTA 2002*, September 2002, vol. 1, pp. 9–13.

[3] M. Joham, *Optimization of Linear and Nonlinear Transmit Signal Processing*, Dissertation, Technische Universität München, June 2004.

[4] F. A. Dietrich, R. Hunger, M. Joham, and W. Utschick, “Robust Transmit Wiener Filter for Time Division Du-

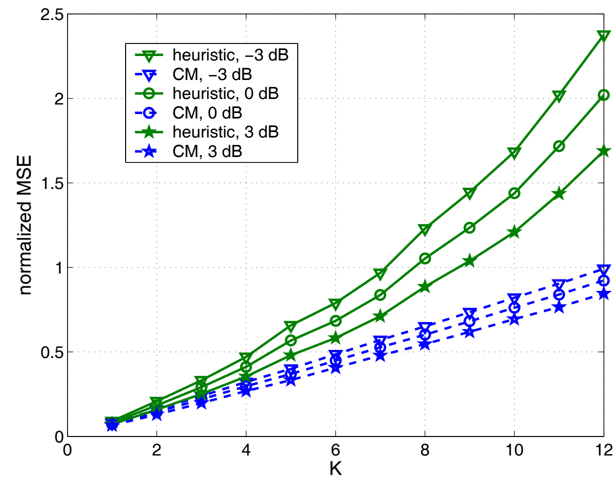

Fig. 2. MSE vs. K for different SNR (angular spread $10°$, $f_\mathrm{d} = 0.12$).

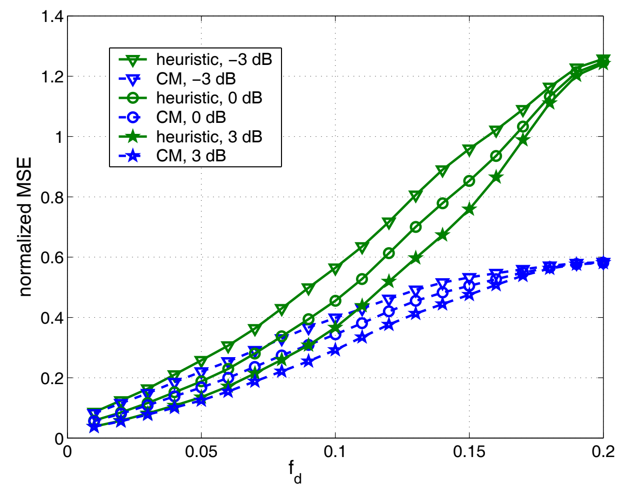

Fig. 3. MSE vs. max. Doppler frequency f_d for different SNR (angular spread $10°$, $K = 6$).

plex Systems,” in *Proc. of ISSPIT*, Darmstadt, Germany, Dec. 2003.

[5] Steven M. Kay, *Fundamentals of Statistical Signal Processing - Estimation Theory*, PTR Prentice Hall, 1st edition, 1993.

[6] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, 2nd edition, 1984.

[7] K. S. Miller, *Complex Stochastic Processes*, Addison-Wesley, 1st edition, 1974.

[8] P. E. Mogensen, K. I. Pedersen, B. Fleury, F. Frederiksen, P. Leth-Espensen, K. Olesen, and S. L. Larsen, “TSUNAMI II: 2D-Channel Characteristics,” Tech. Rep., Aalborg University, Dec. 1997.