Stochastic Transceiver Design in Point-to-Point MIMO Channels with Imperfect CSI

Andreas Gründinger, Michael Joham, and Wolfgang Utschick

International ITG Workshop on Smart Antennas

February 24-25th, 2011

Technische Universität München
Associate Institute for Signal Processing
http://www.msv.ei.tum.de

Stochastic Transceiver Design in Point-to-Point MIMO Channels with Imperfect CSI

Andreas Gründinger, Michael Joham, and Wolfgang Utschick
Associate Institute for Signal Processing, Technische Universität München, 80290 Munich, Germany
{gruendinger,joham,utschick}@tum.de

***Abstract*—The problem of stochastically robust minimum mean square error (MMSE) transceiver design is addressed for multiple-input multiple-output (MIMO) point-to-point channels with different imperfect channel state information (CSI) at the receiver and the transmitter. While the receiver has distribution knowledge of the doubly correlated Gaussian channel that is conditioned on pilot-based training observations (partial CSIRx), the transmitter has either conditional distribution knowledge about the receiver's observations based on feedback (partial CSITx), or only unconditioned distribution knowledge (statistical CSITx). In case of partial CSITx, the design is based on an alternating optimization of the transmit and receive filter. For statistical CSITx, a novel closed-form expression for the expected MMSE is calculated and the structure of the optimal precoder is determined. This enables us to employ an efficient gradient projection method for the robust precoder design.**

***Index Terms*—MIMO point-to-point channel, MSE minimization, statistical transmitter CSI, partial receiver CSI**

I. Introduction

We examine the stochastically robust linear MMSE transceiver design problem in the MIMO point-to-point channel, where one multi-antenna transmitter serves one multi-antenna receiver. A comprehensive study of this problem in the case of complete CSI—perfect knowledge of each channel realization—is presented in [1]. However, the joint design is still a difficult problem when only imperfect *transmitter CSI* (CSITx) and *receiver CSI* (CSIRx) is available, even though a tractable cost as the *mean square error* (MSE) is given.

For imperfect CSI, two scenarios are common in the literature. The first scenario describes a system with complete CSIRx, while CSITx is modeled with an estimated channel mean and covariance (e.g., see [2], [3], and references therein). In the second scenario, equal CSIRx and CSITx is assumed that is expressed by the conditional *probability density function* (PDF) of the channel (cf., [4], [5]). However, both scenarios lack in an accurate description of mobile communication systems. Whereas the first scenario does not take into account limitations in training and estimation for the receiver, the second scenario neglects the fact that CSIRx is generally more accurate than CSITx, e.g., for limited feedback of CSI.

In contrast to above models, we study the scenario, where both ends—the transmitter and the receiver—have *different* imperfect CSI. CSIRx is obtained via pilot aided conditional channel estimation and CSITx is acquired via perfect feedback of quantized receiver's observations. Based on this channel knowledge model, we differentiate between *partial CSITx* and *statistical CSITx*: partial CSITx denotes the case where we can model the receiver's observations via their PDF conditioned on the transmitter's observations, whereas statistical CSITx means that the transmitter's observations give only knowledge about the long-term distribution of the receiver's observations, i.e., the unconditioned PDF of the receiver's observations.

For above model, we formulate a proper average MSE minimization problem and discuss the joint robust transceiver design for partial and statistical CSITx. Unfortunately, the case of partial CSITx (and partial CSIRx) appears to be difficult. So far, there exists no closed form expression for the expected MMSE at the transmitter side. An alternating optimization (AO) strategy, which is common in literature, for determining the precoder and equalizer will, therefore, be presented.

The focus and the main contributions of this work are for statistical CSITx. Contrary to partial CSITx, we are able to derive a novel closed form expression for the ergodic MMSE. The key point of this derivation is a first order derivative of an ergodic rate expression w.r.t. the effective *signal to noise ratio* (SNR) (similar to those in [6]). The derived expression provides some structure for the ergodic MMSE minimizing transmit covariance matrix (and the precoder). This structure, together with the ergodic MMSE expression, can be exploited for the transmit covariance matrix design with efficient gradient methods. To this end, we present closed-form expressions for the derivative of the ergodic MMSE expression. In this context, first order derivatives of eigenvalues w.r.t. scalar real-valued parameters are applied. Note that the presented analyses and designs for point-to-point MIMO systems are the basis for those of multi-user MIMO systems in [7].

We present numerical results to show the correctness of the proposed ergodic MMSE expression. Then, we compare the achievable average performance for the partial CSITx and the statistical CSITx scenarios with the performance of the method in [5], where the receiver uses the same CSI as the transmitter, and the setup where no CSI is available at the transmitter side.

Notation: We say that a random matrix $\boldsymbol{X} \in \mathbb{C}^{p\times q}$ has a matrix variate complex Gaussian distribution with mean $\boldsymbol{M} \in \mathbb{C}^{p\times q}$ and covariance matrix $\boldsymbol{\Psi}^{\mathrm{T}} \otimes \boldsymbol{\Phi}$ (where $\boldsymbol{\Psi} \in \mathbb{C}^{q\times q}$ and $\boldsymbol{\Phi} \in \mathbb{C}^{p\times p}$ are positive definite), denoted as $\boldsymbol{X} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{M}, \boldsymbol{\Psi}, \boldsymbol{\Phi})$, if $\mathrm{vec}(\boldsymbol{X}) \sim \mathcal{N}_{\mathbb{C}}\big(\mathrm{vec}(\boldsymbol{M}), \boldsymbol{\Psi}^{\mathrm{T}} \otimes \boldsymbol{\Phi}\big)$ [8]. For $\boldsymbol{X} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{M}, \boldsymbol{\Psi}, \boldsymbol{\Phi})$, we have $\boldsymbol{X}^{\mathrm{H}} \sim \mathcal{N}_{\mathbb{C}}\big(\boldsymbol{M}^{\mathrm{H}}, \boldsymbol{\Phi}, \boldsymbol{\Psi}\big)$ and

$$\mathrm{E}_{\boldsymbol{X}}[\boldsymbol{X}\boldsymbol{B}\boldsymbol{X}^{\mathrm{H}}] = \boldsymbol{M}\boldsymbol{B}\boldsymbol{M}^{\mathrm{H}} + \mathrm{tr}(\boldsymbol{\Psi}\boldsymbol{B})\boldsymbol{\Phi}. \tag{1}$$

II. System Model and Channel Statistics

We consider the block model of a time-discrete MIMO point-to-point channel with n transmit and m receive antennas.

The transmitter forms the transmitted signal by spatially filtering the $l \leq \min(m,n)$ zero-mean and mutually uncorrelated data entries of $\boldsymbol{d} \in \mathbb{C}^l$ with the linear precoder $\boldsymbol{P} \in \mathbb{C}^{n\times l}$, i.e., $\boldsymbol{x} = \boldsymbol{Pd} \in \mathbb{C}^n$. This transmitted signal propagates over the channel $\boldsymbol{H} \in \mathbb{C}^{m\times n}$ to the receiver and is perturbed by the additive circularly symmetric complex Gaussian noise vector $\boldsymbol{n} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \boldsymbol{C_n})$. The resulting received signal vector $\boldsymbol{y} = \boldsymbol{HPx} + \boldsymbol{n} \in \mathbb{C}^m$ is linearly filtered with the equalizer $\boldsymbol{G} \in \mathbb{C}^{l\times m}$ to obtain the estimate for the intended data vector

$$\hat{\boldsymbol{d}} = \boldsymbol{GHPd} + \boldsymbol{Gn} \ \in \mathbb{C}^l, \tag{2}$$

which is completely described by the statistics of the noise, the data symbols, and the channel.

For the channel, we use a narrow-band flat Rayleigh fading model with Kronecker product covariance structure. That is, the channel has a matrix variate complex Gaussian distribution, i.e., $\boldsymbol{H} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \boldsymbol{R}_{\text{Tx}}, \boldsymbol{R}_{\text{Rx}})$, where $\boldsymbol{R}_{\text{Tx}} \in \mathbb{C}^{n\times n}$ and $\boldsymbol{R}_{\text{Rx}} \in \mathbb{C}^{m\times m}$ comprise the channel correlations at the transmitter and receiver side, respectively. This model is sufficiently accurate in many practical scenarios, where the distance between transmitter and receiver is large compared to the distance between the elements of the same linear antenna pattern [9]. We remark that the channel matrix is equivalently distributed as $\boldsymbol{H} \simeq \boldsymbol{R}_{\text{Rx}}^{1/2}\boldsymbol{H}_{\text{w}}\boldsymbol{R}_{\text{Tx}}^{1/2,\text{H}}$, where we obtain $\boldsymbol{R}_{\text{Rx}}^{1/2}$ and $\boldsymbol{R}_{\text{Tx}}^{1/2}$ from $\boldsymbol{R}_{\text{Rx}}$ and $\boldsymbol{R}_{\text{Tx}}$ via the Cholesky-factorization [10] and $\boldsymbol{H}_{\text{w}} \in \mathbb{C}^{m\times n}$ consists of i.i.d. zero-mean and unit-variance complex Gaussian entries.

III. Channel Training and Parameter Estimation

A key issue regarding the stochastic robust transceiver design methodology for above MIMO point-to-point channel is the available degree of CSI at the transmitter and the receiver side, and the way it is acquired. Current approaches try to combine the advantages of pilot-based channel estimation relying on the channel reciprocity in TDD systems and the use of a dedicated feedback link with limited capacity [11]. Here, we simply assume that CSIRx is obtained via the transmission of a training sequence and MMSE channel estimation according to the orthogonal training method in [5]. CSITx is acquired via perfect feedback of the quantized receiver's observations.

The receiver obtains the observation matrix $\boldsymbol{Y}_{\text{Rx}} = \boldsymbol{HX} + \boldsymbol{N}_{\text{Rx}} \ \in \mathbb{C}^{m\times n}$ in the training period, where we choose the deterministic training signal matrix $\boldsymbol{X} = \alpha_{\text{Rx}}\boldsymbol{R}_{\text{Tx}}^{-1/2,\text{H}}\boldsymbol{T} \in \mathbb{C}^{n\times n}$, with unitary $\boldsymbol{T} \in \mathbb{C}^{n\times n}$, and $\boldsymbol{N}_{\text{Rx}} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \mathbf{I}_n, \boldsymbol{C}_{\boldsymbol{n}_{\text{Rx}}})$. The scalar α_{Rx} is introduced to meet the power requirements of the transmitter during training, i.e., $\alpha_{\text{Rx}} = P_{\text{Tx}}\big/\sqrt{\text{tr}(\boldsymbol{R}_{\text{Tx}})}$ for a maximal total average transmit power P_{Tx}. Performing MMSE estimation of $\boldsymbol{H}$, we can state the conditional channel estimate and the distribution of the estimation error in matrix variate form, i.e., (cf. [5, Section II.*B*.])

$$\boldsymbol{H} = \hat{\boldsymbol{H}} + \boldsymbol{E}_{\text{Rx}}. \tag{3}$$

The conditional channel estimate is given by

$$\hat{\boldsymbol{H}} = \boldsymbol{A}_{\text{Rx}}\boldsymbol{Y}_{\text{Rx}}\boldsymbol{X}^{-1} \tag{4}$$

and the estimation error is distributed as $\boldsymbol{E}_{\text{Rx}} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \boldsymbol{R}_{\text{Tx}}, \boldsymbol{B}_{\text{Rx}})$, where $\boldsymbol{A}_{\text{Rx}} = \alpha_{\text{Rx}}^2\boldsymbol{R}_{\text{Rx}}(\alpha_{\text{Rx}}^2\boldsymbol{R}_{\text{Rx}} + \boldsymbol{C}_{\boldsymbol{n}_{\text{Rx}}})^{-1}$ and $\boldsymbol{B}_{\text{Rx}} = \left(\boldsymbol{R}_{\text{Rx}}^{-1} - \alpha_{\text{Rx}}^2\boldsymbol{C}_{\boldsymbol{n}_{\text{Rx}}}^{-1}\right)^{-1}$ stem from the estimation process (cf. [5, Section II.*B*.]). Note that complete CSIRx is a special case of this channel knowledge model. For very low noise during training, $\boldsymbol{C}_{\boldsymbol{n}_{\text{Rx}}} \to \boldsymbol{0}$, it follows that $\hat{\boldsymbol{H}} \to \boldsymbol{H}$.

The transmitter's observation matrix $\boldsymbol{Y}_{\text{Tx}}$ is obtained via perfect feedback of quantized receiver observations and modeled as $\boldsymbol{Y}_{\text{Tx}} = \boldsymbol{Y}_{\text{Rx}} + \boldsymbol{N}_{\text{q}}$, with noise $\boldsymbol{N}_{\text{q}} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \mathbf{I}_m, \boldsymbol{C}_{\text{q}})$. Note that Gaussian noise is improper for correctly describing quantization noise and is merely introduced for remaining mathematically tractable. Applying MMSE estimation, the transmitter models the receiver's observations as

$$\boldsymbol{Y}_{\text{Rx}} = \hat{\boldsymbol{Y}}_{\text{Rx}} + \boldsymbol{E}_{\text{Tx}}, \tag{5}$$

with estimate $\hat{\boldsymbol{Y}}_{\text{Rx}} = \boldsymbol{A}_{\text{Tx}}\boldsymbol{Y}_{\text{Tx}}$ and estimation error $\boldsymbol{E}_{\text{Tx}} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \mathbf{I}_n, \boldsymbol{B}_{\text{Tx}})$. Here, the matrices $\boldsymbol{A}_{\text{Tx}} = \left(\alpha_{\text{Rx}}^2\boldsymbol{R}_{\text{Rx}} + \boldsymbol{C}_{\boldsymbol{n}_{\text{Rx}}}\right)\left(\alpha_{\text{Rx}}^2\boldsymbol{R}_{\text{Rx}} + \boldsymbol{C}_{\boldsymbol{n}_{\text{Rx}}} + \boldsymbol{C}_{\text{q}}\right)^{-1}$ and $\boldsymbol{B}_{\text{Tx}} = \left((\alpha_{\text{Rx}}^2\boldsymbol{R}_{\text{Rx}} + \boldsymbol{C}_{\boldsymbol{n}_{\text{Rx}}})^{-1} + \boldsymbol{C}_{\text{q}}^{-1}\right)^{-1}$ are defined via the MMSE estimation process. Note that we assume perfect feedback, that is, we imply that the receiver is aware of $\boldsymbol{E}_{\text{Tx}}$ and $\boldsymbol{Y}_{\text{Tx}}$. The important statistical CSITx scenario, where the transmitter has only access to the unconditioned distribution of $\boldsymbol{Y}_{\text{Rx}}$, can be seen as the extreme case for very high quantization noise, i.e., $\boldsymbol{C}_{\text{q}}^{-1} \to \boldsymbol{0}$. In this case, $\boldsymbol{Y}_{\text{Tx}}$ becomes statistically independent of $\boldsymbol{Y}_{\text{Rx}}$ and, therefore, $\hat{\boldsymbol{Y}}_{\text{Rx}} \to \boldsymbol{0}$ and $\boldsymbol{B}_{\text{Tx}} \to \alpha_{\text{Rx}}^2\boldsymbol{R}_{\text{Rx}} + \boldsymbol{C}_{\boldsymbol{n}_{\text{Rx}}}$.

IV. Mean Square Error Optimization

The MSE $\varepsilon(\boldsymbol{G}, \boldsymbol{P}) \triangleq \text{E}_{\boldsymbol{d},\boldsymbol{n}}[\|\boldsymbol{d} - \hat{\boldsymbol{d}}\|_2^2]$ of the considered data transmission model in Section II is given by

$$\varepsilon(\boldsymbol{G}, \boldsymbol{P}) = l + \text{tr}\big(\boldsymbol{G}(\boldsymbol{HQH}^{\text{H}} + \boldsymbol{C_n})\boldsymbol{G}^{\text{H}}\big) - 2\text{Re}\left\{\text{tr}(\boldsymbol{GHP})\right\}, \tag{6}$$

where $\boldsymbol{Q} = \boldsymbol{PP}^{\text{H}}$ is the transmit covariance matrix. For perfect CSI at both sides of the communication link, common objectives for the joint design of the precoder $\boldsymbol{P}$ and the equalizer $\boldsymbol{G}$ are strongly related to the MMSE—the achieved MSE with MMSE receiver. Popular examples are the throughput maximization, the BER minimization, and the maximization of the individual signal-to-interference-plus-noise ratios (SINR) [12], [13], [14].

Unfortunately, these objectives are not that closely connected to the MMSE for imperfect CSI. Moreover, as the influence of partial CSIRx is best understood with the MSE, we focus on minimizing its average value. In this context, note that the imperfect CSIRx is considerably more accurate than the CSITx which is acquired via feedback. Therefore, the average MSE can be minimized w.r.t. $\boldsymbol{G}$ taking into account the better CSIRx. Specifically, we focus on

$$\min_{\boldsymbol{P}} \text{E}_{\boldsymbol{Y}_{\text{Rx}}|\boldsymbol{Y}_{\text{Tx}}}\big[\min_{\boldsymbol{G}} \text{E}_{\boldsymbol{H}|\boldsymbol{Y}_{\text{Rx}}}[\varepsilon(\boldsymbol{G}, \boldsymbol{P})|\boldsymbol{Y}_{\text{Tx}}]\big] \quad \text{s.t.: } \|\boldsymbol{P}\|_{\text{F}}^2 \leq P_{\text{Tx}}, \tag{7}$$

where the constraint reflects limitations in the total average transmit power. The inner optimization w.r.t. $\boldsymbol{G}$ is based on the better conditional channel knowledge $\boldsymbol{H}|\boldsymbol{Y}_{\text{Rx}}$ as desired, and an outer optimization w.r.t. $\boldsymbol{P}$ is based on the less accurate training/feedback observations $\boldsymbol{Y}_{\text{Tx}}$. Note that this optimization order also takes into account the two extreme cases of complete CSIRx and statistical CSITx. For complete CSIRx, $\boldsymbol{H}$ is completely attained via $\boldsymbol{Y}_{\text{Rx}}$ and the inner expectation disappears. In the statistical CSITx scenario, the observation matrix $\boldsymbol{Y}_{\text{Tx}}$ becomes independent of $\boldsymbol{Y}_{\text{Rx}}$ and the outer averaging is only based on the unconditioned PDF $\mathrm{f}_{\boldsymbol{Y}_{\text{Rx}}}(\boldsymbol{Y}_{\text{Rx}})$. Furthermore, when the receiver neglects $\boldsymbol{Y}_{\text{Rx}}$ and designs $\boldsymbol{G}$ w.r.t. $\boldsymbol{Y}_{\text{Tx}}$, we have equal CSI as in [5].

A. Equalizer Design

As the receiver is aware of $\boldsymbol{Y}_{\text{Tx}}$ (output of quantizer at the receiver), it can perfectly determine $\boldsymbol{P}$ by doing the same computations as the transmitter. Hence, the expected MSE at the receiver side $\mathrm{E}_{\boldsymbol{H}|\boldsymbol{Y}_{\text{Rx}}}[\varepsilon(\boldsymbol{P},\boldsymbol{G})|\boldsymbol{Y}_{\text{Tx}}]$ is given by

$$\bar{\varepsilon}_{\text{Rx}}(\boldsymbol{G},\boldsymbol{P})=l+\operatorname{tr}\big(\boldsymbol{G}(\hat{\boldsymbol{H}}\boldsymbol{Q}\hat{\boldsymbol{H}}^{\mathrm{H}}+\boldsymbol{C}_{\text{est}})\boldsymbol{G}^{\mathrm{H}}\big)-2\operatorname{Re}\big\{\operatorname{tr}\big(\boldsymbol{G}\hat{\boldsymbol{H}}\boldsymbol{P}\big)\big\}, \tag{8}$$

where the effective noise covariance matrix is [cf. (1)]

$$\boldsymbol{C}_{\text{est}}\triangleq\mathrm{E}_{\boldsymbol{E}_{\text{Rx}}}\big[\boldsymbol{E}_{\text{Rx}}\boldsymbol{Q}\boldsymbol{E}_{\text{Rx}}^{\mathrm{H}}\big]+\boldsymbol{C}_{\boldsymbol{n}}=\operatorname{tr}(\boldsymbol{Q}\boldsymbol{R}_{\text{Tx}})\boldsymbol{B}_{\text{Rx}}+\boldsymbol{C}_{\boldsymbol{n}} \tag{9}$$

and $\hat{\boldsymbol{H}}$ is given by (4).[1] With (8), the inner optimization of (7) is readily solved via the MMSE filter function

$$\boldsymbol{G}(\boldsymbol{P})=\boldsymbol{P}^{\mathrm{H}}\hat{\boldsymbol{H}}^{\mathrm{H}}\big(\hat{\boldsymbol{H}}\boldsymbol{Q}\hat{\boldsymbol{H}}^{\mathrm{H}}+\boldsymbol{C}_{\text{est}}\big)^{-1}. \tag{10}$$

Inserting (10) into (8), the minimum expected MSE $\mathrm{E}_{\boldsymbol{H}|\boldsymbol{Y}_{\text{Rx}}}[\varepsilon(\boldsymbol{G}(\boldsymbol{P}),\boldsymbol{P})|\boldsymbol{Y}_{\text{Tx}}]$ results in

$$\begin{aligned}\bar{\varepsilon}_{\text{Rx}}(\boldsymbol{G}(\boldsymbol{P}),\boldsymbol{P})&=l-\operatorname{tr}\big(\boldsymbol{P}^{\mathrm{H}}\hat{\boldsymbol{H}}^{\mathrm{H}}(\boldsymbol{C}_{\text{est}}+\hat{\boldsymbol{H}}\boldsymbol{Q}\hat{\boldsymbol{H}}^{\mathrm{H}})^{-1}\hat{\boldsymbol{H}}\boldsymbol{P}\big)\\&=\operatorname{tr}\big((\mathbf{I}_l+\boldsymbol{P}^{\mathrm{H}}\hat{\boldsymbol{H}}^{\mathrm{H}}\boldsymbol{C}_{\text{est}}^{-1}\hat{\boldsymbol{H}}\boldsymbol{P})^{-1}\big),\end{aligned} \tag{11}$$

where the second line follows via the *matrix-inversion-lemma*.

B. Precoder Design

With (11), the outer optimization, that has to be performed at the transmitter and the receiver, is equivalent to [cf. (7)]

$$\min_{\boldsymbol{P}}\ \mathrm{E}_{\boldsymbol{Y}_{\text{Rx}}|\boldsymbol{Y}_{\text{Tx}}}[\bar{\varepsilon}_{\text{Rx}}(\boldsymbol{G}(\boldsymbol{P}),\boldsymbol{P})]\quad \text{s.t.: } \|\boldsymbol{P}\|_{\mathrm{F}}^2\le P_{\text{Tx}}. \tag{12}$$

Although having a simple cost function as the MSE, the resulting expected minimization problem for the precoder design in (12) is difficult to solve, especially, when partial CSITx is available. The reason is, that we are not aware of an analytic representation of the objective in this case.

To overcome this difficulty, one might think of a precoder design based on an approximation of $\mathrm{E}_{\boldsymbol{Y}_{\text{Rx}}|\boldsymbol{Y}_{\text{Tx}}}[\bar{\varepsilon}_{\text{Rx}}(\boldsymbol{G}(\boldsymbol{P}),\boldsymbol{P})]$. A lower bound of the objective, that has a simple analytic expression, is obtained via Jensen's inequality [4]:

$$\mathrm{E}_{\boldsymbol{Y}_{\text{Rx}}|\boldsymbol{Y}_{\text{Tx}}}[\bar{\varepsilon}_{\text{Rx}}]\ge\operatorname{tr}\big((\mathbf{I}_l+\boldsymbol{P}^{\mathrm{H}}\boldsymbol{\Psi}_{\boldsymbol{H}}\boldsymbol{P})^{-1}\big), \tag{13}$$

where the effective channel covariance matrix is [cf. (1)]

$$\boldsymbol{\Psi}_{\boldsymbol{H}}\triangleq\bar{\boldsymbol{H}}^{\mathrm{H}}\boldsymbol{C}_{\text{est}}^{-1}\bar{\boldsymbol{H}}+\operatorname{tr}\big(\boldsymbol{A}_{\text{Rx}}^{\mathrm{H}}\boldsymbol{C}_{\text{est}}^{-1}\boldsymbol{A}_{\text{Rx}}\boldsymbol{B}_{\text{Tx}}\big)\boldsymbol{X}^{-\mathrm{H}}\boldsymbol{X}^{-1}, \tag{14}$$

with $\bar{\boldsymbol{H}}=\boldsymbol{A}_{\text{Rx}}\hat{\boldsymbol{Y}}_{\text{Rx}}\boldsymbol{X}^{-1}$. In the extreme case of complete CSIRx, where $\boldsymbol{C}_{\text{est}}=\boldsymbol{C}_{\boldsymbol{n}}$, above approximation is known to have a bounded asymptotic error and its minimizer has a closed form expression [4]. However, these results cannot easily be extended to the case of partial CSIRx, where the effective noise $\boldsymbol{C}_{\text{est}}$ is linear in $\operatorname{tr}(\boldsymbol{Q}\boldsymbol{R}_{\text{Tx}})$ [cf. (9)].

Moreover, minimizing the lower bound is of minor interest as its optimal value gives no accurate information about the actual achievable minimum average MMSE. In the next section, we devise a second approach, which is based on AO, and use Monte-Carlo simulations for the expectation evaluations.

The focus and the main contributions of this work, however, are on the extreme case of statistical CSITx, which is detailed in Section VI. For this case, we are able to derive novel closed form expressions of the expected MMSE and the basic structure of the ergodic MMSE minimizing precoders. This structure and the ergodic MMSE expression can be exploited for approaching the average MMSE minimization via an efficient gradient projection method.

V. Robust Precoder Design for Partial CSITx

The original precoder design in (12) is difficult due to the lack of a closed form expressions for the average MSE. Instead, we employ an AO procedure which iteratively switches between updating the precoder and the equalizer function $\boldsymbol{G}(\boldsymbol{P})$ to arrive at an MSE minimizing transceiver pair. In every step of the algorithm, the equalizer function is first found for the better partial CSIRx. Second, the precoder is optimized for the worse partial CSITx and for above equalizer function.

In the i-th iteration, given an initial $\boldsymbol{P}^{(i-1)}$, we update $\boldsymbol{G}^{(i)}=\boldsymbol{G}(\boldsymbol{P}^{(i-1)})$ according to (10)—as minimizer to the inner optimization of problem (7). Then, we insert this fixed filter function into the MSE (8). The precoder $\boldsymbol{P}^{(i)}$ is found as a minimizer to the resulting outer optimization:

$$\boldsymbol{P}^{(i)}=\underset{\boldsymbol{P}}{\operatorname{argmin}}\,\mathrm{E}_{\boldsymbol{Y}_{\text{Rx}}|\boldsymbol{Y}_{\text{Tx}}}\big[\bar{\varepsilon}_{\text{Rx}}(\boldsymbol{G}^{(i)},\boldsymbol{P})\big]\ \text{ s.t.: } \|\boldsymbol{P}\|_{\mathrm{F}}^2\le P_{\text{Tx}}. \tag{15}$$

Calculating the *Karush-Kuhn-Tucker* (KKT) conditions of (15), the (i)-th precoder update is readily found to be

$$\begin{aligned}\boldsymbol{P}^{(i)}=\big(&\mathrm{E}_{\boldsymbol{Y}_{\text{Rx}}|\boldsymbol{Y}_{\text{Tx}}}\big[\hat{\boldsymbol{H}}^{\mathrm{H}}\boldsymbol{D}^{(i)}\hat{\boldsymbol{H}}\big]+\mathrm{E}_{\boldsymbol{Y}_{\text{Rx}}|\boldsymbol{Y}_{\text{Tx}}}\big[\operatorname{tr}\big(\boldsymbol{B}_{\text{Rx}}\boldsymbol{D}^{(i)}\big)\big]\boldsymbol{R}_{\text{Tx}}\\&+\varrho^{(i)}\mathbf{I}_n\big)^{-1}\,\mathrm{E}_{\boldsymbol{Y}_{\text{Rx}}|\boldsymbol{Y}_{\text{Tx}}}\big[\hat{\boldsymbol{H}}^{\mathrm{H}}\boldsymbol{G}^{(i),\mathrm{H}}\big],\end{aligned} \tag{16}$$

where $\boldsymbol{D}^{(i)}=\boldsymbol{G}^{(i),\mathrm{H}}\boldsymbol{G}^{(i)}$. The non-negative Lagrangian multiplier is $\varrho^{(i)}=\max\{0,\bar{\varrho}^{(i)}\}$ and $\bar{\varrho}^{(i)}$ is the largest (real-valued) solution of

$$\sum_{k=1}^{n}\frac{\big\|\mathrm{E}_{\boldsymbol{Y}_{\text{Rx}}|\boldsymbol{Y}_{\text{Tx}}}[\boldsymbol{G}^{(i)}\hat{\boldsymbol{H}}]\boldsymbol{v}_k^{(i)}\big\|_2^2}{(\delta_k^{(i)}+\bar{\varrho}^{(i)})^2}=P_{\text{Tx}}. \tag{17}$$

The real-valued scalars $\delta_k^{(i)}$ and the vectors $\boldsymbol{v}_k^{(i)}\in\mathbb{C}^n$, $k\in\{1,\ldots,n\}$, are the decreasingly ordered eigenvalues and the corresponding eigenvectors of the positive definite matrix

$$\mathrm{E}_{\boldsymbol{Y}_{\text{Rx}}|\boldsymbol{Y}_{\text{Tx}}}\big[\hat{\boldsymbol{H}}^{\mathrm{H}}\boldsymbol{D}^{(i)}\hat{\boldsymbol{H}}\big]+\mathrm{E}_{\boldsymbol{Y}_{\text{Rx}}|\boldsymbol{Y}_{\text{Tx}}}\big[\operatorname{tr}\big(\boldsymbol{B}_{\text{Rx}}\boldsymbol{D}^{(i)}\big)\big]\boldsymbol{R}_{\text{Tx}}. \tag{18}$$

[1]For complete CSIRx, $\hat{\boldsymbol{H}}$ and $\boldsymbol{C}_{\text{est}}$ can be replaced with $\boldsymbol{H}$ and $\boldsymbol{C}_{\boldsymbol{n}}$, respectively.

Since $\boldsymbol{D}^{(i)}$ and $\boldsymbol{G}^{(i)}$ are functions of $\boldsymbol{P}^{(i-1)}$, we essentially arrived at a *fixed point* algorithm for the precoder. Unfortunately, no closed form expressions of the expectations in (16)–(18) are available. Therefore, we apply Monte-Carlo simulations for the numerical results.

Note that this iteration scheme converges, i.e., in each iteration the MSE is reduced and, therefore, smaller (or equal) than in previous iterations. Moreover, as the KKT conditions of (7) are satisfied in the convergence point, this procedure ensures local optimality of the solution. As initialization $\boldsymbol{P}^{(0)}$, we choose the l dominant eigenvectors of $\boldsymbol{R}_{\text{Tx}}$.

VI. Robust Precoder Design for Statistical CSITx

Now, we focus on the statistical CSITx case, where the conditional PDF is given by $\mathrm{f}_{\boldsymbol{Y}_{\text{Rx}}|\boldsymbol{Y}_{\text{Tx}}}(\boldsymbol{Y}_{\text{Rx}}|\boldsymbol{Y}_{\text{Tx}}) = \mathrm{f}_{\boldsymbol{Y}_{\text{Rx}}}(\boldsymbol{Y}_{\text{Rx}})$ with $\boldsymbol{Y}_{\text{Rx}} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I}_n, \boldsymbol{C}_{\text{Rx}})$ and $\boldsymbol{C}_{\text{Rx}} = \alpha_{\text{Rx}}^2 \boldsymbol{R}_{\text{Rx}} + \boldsymbol{C}_{\boldsymbol{n}_{\text{Rx}}}$ (see Section III). Statistical CSITx is beneficial in the sense that we are able to: *A.)* propose a closed-form expression for the ergodic MMSE and *B.)* determine the basic structure of the MMSE minimizing precoder. The basis is a very close connection between the MMSE and a virtual average rate expression. In *C.)*, the results are exploited to employ an efficient gradient projection method of [15].

A. Ergodic Minimum Mean Square Error Derivation

A direct derivation of the ergodic MMSE, given by [cf. (11)]

$$\bar{\varepsilon}(\gamma, \boldsymbol{Q}) \triangleq l - \mathrm{E}_{\boldsymbol{Y}_{\text{Rx}}}\left[\mathrm{tr}\left((\boldsymbol{C}_{\text{est}} + \gamma \hat{\boldsymbol{H}}\boldsymbol{Q}\hat{\boldsymbol{H}}^{\mathrm{H}})^{-1} \gamma \hat{\boldsymbol{H}}\boldsymbol{Q}\hat{\boldsymbol{H}}^{\mathrm{H}}\right)\right], \quad (19)$$

appears to be difficult. However, similar to [6, Theorem 2], we devise to calculate the expected MMSE through a simple derivation of the ergodic rate

$$\bar{R}(\gamma, \boldsymbol{Q}) \triangleq \mathrm{E}_{\boldsymbol{Y}_{\text{Rx}}}\left[\log\left|\mathbf{I}_m + \gamma \boldsymbol{C}_{\text{est}}^{-1} \hat{\boldsymbol{H}}\boldsymbol{Q}\hat{\boldsymbol{H}}^{\mathrm{H}}\right|\right] \quad (20)$$

w.r.t. the slack variable γ, that can be seen as the normalized SNR. For $\gamma = 1$, (19) meets exactly the ergodic value of (11).

Lemma 1. *Let the ergodic MMSE and the ergodic rate be defined by* (19) *and* (20)*, respectively. Then,*

$$\bar{\varepsilon}(\gamma, \boldsymbol{Q}) = l - \gamma \tfrac{\partial}{\partial \gamma} \bar{R}(\gamma, \boldsymbol{Q}).$$

A verification of Lemma 1 is straightforward using derivatives of matrix valued functions and Leibniz's rule for differentiation under the integral [16]. Fortunately, we can state a closed form expression for the ergodic rate expression in (20).

Theorem 1 ([17, Theorem 3.6])**.** *The ergodic value of* $R(\gamma, \boldsymbol{\Sigma}, \boldsymbol{\Omega}) = \log\left|\mathbf{I}_\mu + \gamma \boldsymbol{\Sigma}\boldsymbol{W}\boldsymbol{\Omega}\boldsymbol{W}^{\mathrm{H}}\right|$ *w.r.t.* $\boldsymbol{W} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_{\mu\times\nu}, \mathbf{I}_\nu, \mathbf{I}_\mu)$, $\mu \le \nu$ *is given by*

$$\bar{R}(\gamma, \boldsymbol{\Sigma}, \boldsymbol{\Omega}) = c\chi(\gamma, \boldsymbol{\Sigma}, \boldsymbol{\Omega}) \sum_{p=1}^{\mu} |\boldsymbol{\Xi}_p(\gamma, \boldsymbol{\Sigma}, \boldsymbol{\Omega})|,$$

with the constant c *and the substitute* $\chi(\gamma, \boldsymbol{\Sigma}, \boldsymbol{\Omega})$ *given as*

$$c = \frac{\Gamma(\nu) \prod_{i=\mu+1}^{\nu} [1-\nu]_{i-\mu-1}}{\prod_{i=1}^{\nu-1} i^i} \quad (21)$$

$$\chi(\gamma, \boldsymbol{\Sigma}, \boldsymbol{\Omega}) = \frac{\gamma^{\frac{(\nu-\mu)(\nu-\mu-1)-\nu(\nu-1)}{2}}}{|\boldsymbol{\Sigma}|^{\nu-\mu} \, \mathrm{v}_\mu(\boldsymbol{\Sigma}) \, \mathrm{v}_\nu(\boldsymbol{\Omega})}, \quad (22)$$

respectively, and the $\nu \times \nu$ *matrix* $\boldsymbol{\Xi}_p(\gamma, \boldsymbol{\Sigma}, \boldsymbol{\Omega})$ *of elements*

$$[\boldsymbol{\Xi}_p]_{i,j} = \begin{cases} (z_{i,j})^\nu f\left(\frac{1}{z_{i,j}}\right) & p = i \le \mu \\ \sum_{k=\nu-\mu}^{\nu-1} (z_{i,j})^k [\nu-k]_k & p \ne i \le \mu \\ \omega_j^{i-\mu-1} & i > \mu \end{cases}, \quad (23)$$

where $z_{i,j} \triangleq \gamma \sigma_i \omega_j$, $i, j \in \{1, \ldots, \nu\}$ *and the scalars* σ_i *and* ω_j *are the distinct and decreasingly ordered eigenvalues of* $\boldsymbol{\Sigma}$ *and* $\boldsymbol{\Omega}$*, respectively. The expression* $\Gamma(k)$ *denotes the Gamma-function,* $f(x) = x\,\mathrm{e}^x\,\mathrm{E}_1(x)$*, where* $\mathrm{E}_1(x)$ *is the exponential integral function,* $[x]_k$ *is the Pochhammer symbol, i.e.,* $[x]_k = \prod_{i=1}^{k-1}(x+i)$ *and* $[x]_0 = 1$*, and* $\mathrm{v}_k(\boldsymbol{X})$ *is the Vandermonde determinant of the decreasingly ordered eigenvalues* $\{x_i\}_{i=1}^k$ *of* $\boldsymbol{X} \in \mathbb{C}^{k\times k}$*, i.e.,* $\mathrm{v}_k(\boldsymbol{X}) = \prod_{i<j}^k (x_i - x_j)$.

To arrive at the unifying mutual information expression in this theorem, we show that

$$R \triangleq \log\left|\mathbf{I}_m + \gamma \boldsymbol{C}_{\text{est}}^{-1} \hat{\boldsymbol{H}}\boldsymbol{Q}\hat{\boldsymbol{H}}^{\mathrm{H}}\right| \simeq R(\gamma, \boldsymbol{\Sigma}, \boldsymbol{\Omega}) \quad (24)$$

for some suitable choice of eigenvalue matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$. To this end, we remark that the receiver's channel estimate $\hat{\boldsymbol{H}}$ is linear in $\boldsymbol{Y}_{\text{Rx}} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I}_n, \boldsymbol{C}_{\text{Rx}})$ (for statistical CSITx), which is equivalently distributed as $\boldsymbol{Y}_{\text{Rx}} \simeq \boldsymbol{C}_{\text{Rx}}^{1/2} \boldsymbol{Y}_{\text{w}}$, where $\boldsymbol{Y}_{\text{w}}$ consists of i.i.d. complex Gaussian elements, i.e., $[\boldsymbol{Y}_{\text{w}}]_{i,j} \sim \mathcal{N}_{\mathbb{C}}(0, 1)$, and is of the same size as $\boldsymbol{Y}_{\text{Rx}}$. Inserting $\boldsymbol{Y}_{\text{Rx}}$ into (20), we find

$$R = \log\left|\mathbf{I}_m + \gamma \boldsymbol{S}\boldsymbol{Y}_{\text{w}}\boldsymbol{T}^{\mathrm{H}}\boldsymbol{O}\boldsymbol{T}\boldsymbol{Y}_{\text{w}}^{\mathrm{H}}\right|, \quad (25)$$

where we introduced the substitutes

$$\boldsymbol{O} \triangleq \tfrac{1}{\alpha_{\text{Rx}}^2} \boldsymbol{R}_{\text{Tx}}^{1/2,\mathrm{H}} \boldsymbol{Q} \boldsymbol{R}_{\text{Tx}}^{1/2}, \quad r_{\boldsymbol{O}} = \mathrm{rank}\{\boldsymbol{O}\} \quad (26a)$$

$$\boldsymbol{S} \triangleq \boldsymbol{C}_{\text{Rx}}^{1/2,\mathrm{H}} \boldsymbol{A}_{\text{Rx}}^{\mathrm{H}} \boldsymbol{C}_{\text{est}}^{-1} \boldsymbol{A}_{\text{Rx}} \boldsymbol{C}_{\text{Rx}}^{1/2}, \quad r_{\boldsymbol{S}} = \mathrm{rank}\{\boldsymbol{S}\}. \quad (26b)$$

By noticing that the i.i.d. complex Gaussian distributed $m \times n$ matrix $\boldsymbol{Y}_{\text{w}}$ is invariant to left and right multiplications with unitary matrices $\boldsymbol{U}$ and $\boldsymbol{V}$, i.e., $\boldsymbol{Y}_{\text{w}} \simeq \boldsymbol{U}\boldsymbol{Y}_{\text{w}}\boldsymbol{V}$ (e.g., see [8]), it can easily be deduced that the expected value of R only depends on the non-zero eigenvalues of $\boldsymbol{O}$ and $\boldsymbol{S}$. We combine these positive decreasingly ordered eigenvalues $\{\sigma_i\}_{i=1}^{\mu}$ and $\{\omega_j\}_{j=1}^{\nu}$, with $\{\boldsymbol{u}_i\}_{i=1}^{\mu}$ and $\{\boldsymbol{v}_j\}_{j=1}^{\nu}$ being the corresponding eigenvectors, in the diagonal matrices $\boldsymbol{\Sigma} = \mathrm{diag}\{\sigma_1, \ldots, \sigma_\mu\}$ and $\boldsymbol{\Omega} = \mathrm{diag}\{\omega_1, \ldots, \omega_\nu\}$, respectively. As Theorem 1 only considers the case where $\nu \ge \mu$, the mapping between $(\boldsymbol{O}, \boldsymbol{S})$ and $(\boldsymbol{\Sigma}, \boldsymbol{\Omega})$ is given as follows: if we define $\mu \triangleq \min(r_{\boldsymbol{S}}, r_{\boldsymbol{O}})$ and $\nu \triangleq \max(r_{\boldsymbol{S}}, r_{\boldsymbol{O}})$, then

$$\begin{aligned} \tilde{\boldsymbol{U}}\boldsymbol{\Sigma}\tilde{\boldsymbol{U}}^{\mathrm{H}} &= \begin{cases} \boldsymbol{S} & \text{for } \mu = r_{\boldsymbol{S}} \le r_{\boldsymbol{O}} \\ \boldsymbol{O} & \text{for } \mu = r_{\boldsymbol{O}} < r_{\boldsymbol{S}} \end{cases} \quad \tilde{\boldsymbol{U}} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_\mu] \\ \tilde{\boldsymbol{V}}\boldsymbol{\Omega}\tilde{\boldsymbol{V}}^{\mathrm{H}} &= \begin{cases} \boldsymbol{O} & \text{for } \nu = r_{\boldsymbol{O}} \ge r_{\boldsymbol{S}} \\ \boldsymbol{S} & \text{for } \nu = r_{\boldsymbol{S}} > r_{\boldsymbol{O}} \end{cases} \quad \tilde{\boldsymbol{V}} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_\nu]. \end{aligned} \quad (27)$$

Note that both eigenvalue matrices, $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$, are matrix valued functions of $\boldsymbol{Q}$ as $\boldsymbol{C}_{\text{est}}$ is affine in $\mathrm{tr}(\boldsymbol{Q}\boldsymbol{R}_{\text{Tx}})$. Furthermore, we remark that Theorem 1 is only valid for pairwise distinct eigenvalues $\{\sigma_i\}_{i=1}^{\mu}$ and $\{\omega_j\}_{j=1}^{\nu}$. Cases with equal eigenvalues can be deduced either with suitable limiting processes [17], or from [18]. Here, we restrict to the case of distinct eigenvalues and $\mu = r_{\boldsymbol{S}} \leq r_{\boldsymbol{O}} = \nu$ for simplicity.

According to Lemma 1, we differentiate $\bar{R}$ w.r.t. γ as

$$\frac{\partial}{\partial\gamma}\bar{R} = c\bigg(\frac{\partial\chi}{\partial\gamma}\sum_{p=1}^{\mu}|\boldsymbol{\varXi}_p| + \chi\sum_{p=1}^{\mu}\frac{\partial|\boldsymbol{\varXi}_p|}{\partial\gamma}\bigg), \tag{28}$$

where the derivative of χ w.r.t. γ is simply

$$\tfrac{\partial}{\partial\gamma}\chi = \tfrac{d}{\gamma}\chi, \quad d = \tfrac{1}{2}[(\nu-\mu)(\nu-\mu-1) - \nu(\nu-1)]. \tag{29}$$

The derivative of $|\boldsymbol{\varXi}_p|$ is found to be

$$\tfrac{\partial}{\partial\gamma}|\boldsymbol{\varXi}_p| = \sum\nolimits_{o=1}^{\mu}\big|\tilde{\boldsymbol{\varXi}}_{p,o}\big|, \tag{30}$$

where $\tilde{\boldsymbol{\varXi}}_{p,o}$ is equal to $\boldsymbol{\varXi}_p$, except for its o-th row (or alternatively column) which is differentiated w.r.t. γ, i.e.,

$$\big[\tilde{\boldsymbol{\varXi}}_{p,o}\big]_{i,j} = \begin{cases} [\boldsymbol{\varXi}_p]_{i,j} & i \neq o \\ g_{\nu-1}(z_{o,j})\frac{\partial z_{o,j}}{\partial\gamma} & i = o = p \leq \mu \\ \sum\limits_{k=\nu-\mu}^{\nu-1} k z_{o,j}^{k-1}[\nu-k]_k\frac{\partial z_{o,j}}{\partial\gamma} & p \neq i = o \leq \mu. \end{cases} \tag{31}$$

Here, we defined $g_y(x)$ for $x \in \mathbb{R}_+$ and $y \in \mathbb{N}_+$ as

$$g_y(x) \triangleq \tfrac{\partial}{\partial x}x^{y+1}f\left(\tfrac{1}{x}\right) = x^y\left[\left(y - \tfrac{1}{x}\right)f\left(\tfrac{1}{x}\right) + \tfrac{1}{x}\right] \tag{32}$$

and $\frac{\partial z_{i,j}}{\partial\gamma} = \sigma_i\omega_j$. With (28)–(32) and Lemma 1, we can directly state the following proposition.

Proposition 1. *The ergodic MMSE $\bar{\varepsilon}(\gamma, \boldsymbol{Q})$ is calculated as*

$$\bar{\varepsilon}(\gamma, \boldsymbol{Q}) = l - c\chi\Bigg(d\sum_{p=1}^{\mu}|\boldsymbol{\varXi}_p| + \gamma\sum_{p,o=1}^{\mu}\big|\tilde{\boldsymbol{\varXi}}_{p,o}\big|\Bigg),$$

where c, χ, d, $\boldsymbol{\varXi}_p$, and $\tilde{\boldsymbol{\varXi}}_{p,o}$, $p, o \in \{1, \ldots, \mu\}$, are defined by (21), (22), (29), (23), *and* (31), *respectively.*

B. Ergodic MMSE Minimizing Precoder Structure

With above ergodic MMSE expression, we can derive the basic structure of the optimal (MMSE minimizing) transmit covariance matrix. To this end, we recast problem (12) as

$$\begin{aligned} &\min_{\boldsymbol{Q}} \bar{\varepsilon}(\gamma, \boldsymbol{Q}) \\ &\text{s.t.: } \boldsymbol{Q} \succeq \boldsymbol{0},\ \operatorname{tr}(\boldsymbol{Q}) \leq P_{\text{Tx}},\ \operatorname{rank}\{\boldsymbol{Q}\} \leq l, \end{aligned} \tag{33}$$

i.e., in terms of the transmit covariance matrix $\boldsymbol{Q} = \boldsymbol{P}\boldsymbol{P}^{\text{H}}$. The trace constraint in (33) is due to $\|\boldsymbol{P}\|_{\text{F}}^2 = \operatorname{tr}(\boldsymbol{P}\boldsymbol{P}^{\text{H}}) = \operatorname{tr}(\boldsymbol{Q})$ and the positive semidefiniteness constraint as well as the rank constraint follow from the definition of $\boldsymbol{Q}$ as the Gramian product of the precoder $\boldsymbol{P}$ with dimension $n \times l$, $l \leq \min(m, n)$ (see Section II). Introducing the EVD $\boldsymbol{Q} = \boldsymbol{F}\boldsymbol{\varPhi}\boldsymbol{F}^{\text{H}}$, with (sub-)unitary matrix $\boldsymbol{F}$ and diagonal matrix $\boldsymbol{\varPhi}$, we can equivalently reformulate the optimization as

$$\begin{aligned} &\min_{\boldsymbol{F},\boldsymbol{\varPhi}} \bar{\varepsilon}(\gamma, \boldsymbol{F}\boldsymbol{\varPhi}\boldsymbol{F}^{\text{H}}) \\ &\text{s.t.: } \boldsymbol{\varPhi} \succeq \boldsymbol{0},\ \operatorname{tr}(\boldsymbol{\varPhi}) \leq P_{\text{Tx}},\ \operatorname{rank}\{\boldsymbol{\varPhi}\} \leq l, \end{aligned} \tag{34}$$

because the positive semidefiniteness constraint, the trace operator, and the rank operator are independent w.r.t. the transform with $\boldsymbol{F}$. That is, we can interpret the ergodic MMSE minimization as two problems: first, find the optimal eigendirections $\boldsymbol{F}$ of $\boldsymbol{Q}$ and then, compute the optimal power allocation matrix $\boldsymbol{\varPhi}$. Now, we focus on the first problem.

From the ergodic MMSE expression in Proposition 1, we can infer that $\bar{\varepsilon}(\gamma, \boldsymbol{Q})$ only depends on the positive and decreasingly-ordered eigenvalues of $\alpha_{\text{Rx}}^2\boldsymbol{O} = \boldsymbol{R}_{\text{Tx}}^{1/2,\text{H}}\boldsymbol{Q}\boldsymbol{R}_{\text{Tx}}^{1/2}$, that are given by $\{\alpha_{\text{Rx}}^2\omega_j\}_{j=1}^{\nu}$ if $\nu = r_{\boldsymbol{O}} \geq r_{\boldsymbol{S}}$ [cf. (26)]. The eigenvalues of $\boldsymbol{S}$, $\{\sigma_i\}_{i=1}^{\mu}$, are functions of $\operatorname{tr}(\boldsymbol{R}_{\text{Tx}}\boldsymbol{Q}) = \alpha_{\text{Rx}}^2\sum_{j=1}^{\nu}\omega_j$ as $\boldsymbol{C}_{\text{est}}$ is given by (9).[2] Moreover, we remark that the objective is monotonically decreasing in the used transmit power $\operatorname{tr}(\boldsymbol{Q})$. Following a similar argumentation as in [19, Proof of Theorem 1], we are able to prove that the ergodic MMSE minimizing transmit covariance matrices satisfy the following proposition.

Proposition 2. *Ergodic MMSE minimizing transmit covariance matrices must be constructed as $\boldsymbol{Q} = \boldsymbol{F}\boldsymbol{\varPhi}\boldsymbol{F}^{\text{H}}$, where $\boldsymbol{F} = [\boldsymbol{f}_1, \ldots, \boldsymbol{f}_{r_{\text{Tx}}}] \in \mathbb{C}^{n \times r_{\text{Tx}}}$ is (sub-)unitary and comprises the eigenvectors to the $r_{\text{Tx}} = \operatorname{rank}\{\boldsymbol{R}_{\text{Tx}}\}$ strictly positive eigenvalues of $\boldsymbol{R}_{\text{Tx}}$, i.e., $\boldsymbol{R}_{\text{Tx}} = \boldsymbol{F}\boldsymbol{\varPsi}\boldsymbol{F}^{\text{H}}$ with diagonal $\boldsymbol{\varPsi} = \operatorname{diag}\{\psi_1, \ldots, \psi_{r_{\text{Tx}}}\}$ and $\psi_1 \geq \ldots \geq \psi_{r_{\text{Tx}}} > 0$ for convenience. The diagonal power allocation matrix $\boldsymbol{\varPhi} = \operatorname{diag}\{\phi_1, \ldots, \phi_{r_{\text{Tx}}}\}$ satisfies $\|\boldsymbol{\varPhi}\|_{\text{F}}^2 = \sum_{i=1}^{r_{\text{Tx}}}\phi_i = P_{\text{Tx}}$.*

In the considered case where $\boldsymbol{Q}$ is defined as the Gramian product of $\boldsymbol{P} \in \mathbb{C}^{n \times l}$ and $l \leq \min(m, n)$, $\boldsymbol{\varPhi}$ is additionally constrained to have at most $r_l = \min(l, r_{\text{Tx}})$ strictly positive entries. Precisely, only those r_l eigenvalues in $\boldsymbol{\varPhi}$ are of interest which correspond to the dominant eigendirections of $\boldsymbol{R}_{\text{Tx}}$, i.e., $\{\phi_k\}_{k=1}^{r_l}$ for non-increasingly ordered eigenvalues $\{\psi_k\}_{k=1}^{r_{\text{Tx}}}$ of $\boldsymbol{R}_{\text{Tx}}$. This claim is based on the observation that the dominant eigenvalues of $\boldsymbol{O}$ correspond to the dominant eigenvalues of $\boldsymbol{R}_{\text{Tx}}$, i.e., $\alpha_{\text{Rx}}^2\omega_j = \psi_j\phi_j$ with $\omega_1 \geq \ldots \geq \omega_\nu$ and $\psi_1 \geq \ldots \geq \psi_\nu$ (cf. [19, Proof of Theorem 1]).

Using above claim (for $r_{\text{Tx}} \geq l$), we recast the structure of the minimizer for problem (33) as

$$\tilde{\boldsymbol{Q}} = \tilde{\boldsymbol{F}}\tilde{\boldsymbol{\varPhi}}\tilde{\boldsymbol{F}}^{\text{H}}, \tag{35}$$

where $\tilde{\boldsymbol{F}} = [\boldsymbol{f}_1, \ldots, \boldsymbol{f}_l] \in \mathbb{C}^{n \times l}$ comprises the l dominant eigenvectors of $\boldsymbol{R}_{\text{Tx}}$ and $\tilde{\boldsymbol{\varPhi}} = \operatorname{diag}\{\phi_1, \ldots, \phi_l\}$. Finally, note that the transmit covariance matrix $\tilde{\boldsymbol{Q}}$ defines the precoder $\boldsymbol{P}$ up to a unitary transform from the right. That is,

$$\boldsymbol{P} = \tilde{\boldsymbol{F}}\tilde{\boldsymbol{\varPhi}}^{1/2}\boldsymbol{J}^{\text{H}}, \tag{36}$$

where $\tilde{\boldsymbol{F}}$ is defined in (35), $\tilde{\boldsymbol{\varPhi}}^{1/2} = \operatorname{diag}\{\phi_1^{1/2}, \ldots, \phi_l^{1/2}\}$, and $\boldsymbol{J} \in \mathbb{C}^{r_l \times r_l}$ is an arbitrary unitary matrix.

C. Gradient Projection Based Precoder Design

The structure of $\tilde{\boldsymbol{Q}}$ in (35) can be exploited for efficient (gradient based) MMSE minimization approaches which work directly on the eigenvalues $\{\phi_k\}_{k=1}^{l}$. Precisely, inserting (35) into (33), we consider the following problem [cf. (34)]:

$$\min_{\tilde{\boldsymbol{\varPhi}}} \bar{\varepsilon}(\gamma, \tilde{\boldsymbol{F}}\tilde{\boldsymbol{\varPhi}}\tilde{\boldsymbol{F}}^{\text{H}}) \ \ \text{s.t.: } \tilde{\boldsymbol{\varPhi}} \succeq \boldsymbol{0},\ \operatorname{tr}\big(\tilde{\boldsymbol{\varPhi}}\big) \leq P_{\text{Tx}}. \tag{37}$$

[2] In the special case of complete CSITx, $\boldsymbol{S}$ is independent of $\boldsymbol{Q}$.

Contrarily to (33), the rank constraint $\mathrm{rank}\{\tilde{\boldsymbol{Q}}\} \leq l$ is always satisfied in this formulation and is, therefore, superfluous.

Here, we suggest to customize the covariance based projected gradient method in [15] for problem (37).[3] The algorithm performs a preconditioned gradient descent step in each iteration, followed by an *orthogonal* projection onto the constraint set. The gradient descent step for the eigenvalue matrix $\tilde{\boldsymbol{\Phi}}$ can be expressed as

$$\bar{\boldsymbol{\Phi}} = \tilde{\boldsymbol{\Phi}} - \theta \cdot \beta \cdot \tfrac{\partial}{\partial \tilde{\boldsymbol{\Phi}}} \bar{\varepsilon}(\gamma, \tilde{\boldsymbol{F}} \tilde{\boldsymbol{\Phi}} \tilde{\boldsymbol{F}}^{\mathrm{H}}), \tag{38}$$

where $\beta \in \mathbb{R}_+$ denotes the (iteration-dependent) step size. The iteration-dependent preconditioning scalar θ can be chosen to normalize the Frobenius norm of the gradient to the transmit power P_{Tx} for increasing the convergence speed. Then, an orthogonal projection onto the constraint set in (37) is applied. The projection can be formulated as

$$\tilde{\boldsymbol{\Phi}}' = \operatorname*{argmin}_{\boldsymbol{\Lambda}} \left\| \bar{\boldsymbol{\Phi}} - \boldsymbol{\Lambda} \right\|_{\mathrm{F}}^2 \ \text{ s.t.: } \mathrm{tr}(\boldsymbol{\Lambda}) \leq P_{\mathrm{Tx}}, \boldsymbol{\Lambda} \succeq \mathbf{0}, \tag{39}$$

which is a convex optimization problem and can be solved in closed form. The iteration-dependent step-size β can be determined iteratively in order to ensure that the objective decreases. For details, we refer to [15].

In order to apply gradient based optimization procedures, like above projected gradient method for example, it is crucial to derive the gradient of the ergodic MMSE, i.e.,

$$\tfrac{\partial}{\partial \tilde{\boldsymbol{\Phi}}} \bar{\varepsilon}(\gamma, \tilde{\boldsymbol{F}} \tilde{\boldsymbol{\Phi}} \tilde{\boldsymbol{F}}^{\mathrm{H}}) = \mathrm{diag}\left\{ \tfrac{\partial \bar{\varepsilon}}{\partial \phi_1}, \ldots, \tfrac{\partial \bar{\varepsilon}}{\partial \phi_l} \right\}. \tag{40}$$

The problem is, that the ergodic MMSE is a function of the eigenvalues $\{\sigma_i\}_{i=1}^{\mu}$ and $\{\omega_j\}_{j=1}^{\nu}$ which are continuous functions in the power allocations $\{\phi_k\}_{k=1}^{l}$. Hence, we have to employ derivatives of eigenvalues w.r.t. real scalar parameters. A short review is given in Appendix A.

With (47), we can explicitly calculate the necessary eigenvalue derivatives for evaluating (40) as we assume pairwise distinct parameters $\{\sigma_i\}_{i=1}^{\mu}$ and $\{\omega_j\}_{j=1}^{\nu}$. To simplify expositions, we restrict here to the case where $\nu = r_{\boldsymbol{O}} \geq r_{\boldsymbol{S}} = \mu$, such that $\boldsymbol{\Sigma}$ comprises the eigenvalues of $\boldsymbol{S}$ while $\boldsymbol{\Omega}$ comprises the eigenvalues of $\boldsymbol{O}$. The other case, where $\mu = r_{\boldsymbol{O}} \leq r_{\boldsymbol{S}} = \nu$ is analogously obtained.

The first order derivatives in (40) can be recast as

$$\frac{\partial \bar{\varepsilon}}{\partial \phi_k} = \frac{\partial \chi}{\partial \phi_k} \frac{\bar{\varepsilon} - l}{\chi} - c\chi \left[d \sum_{p=1}^{\mu} \frac{\partial |\boldsymbol{\Xi}_p|}{\partial \phi_k} + \gamma \sum_{p,o=1}^{\mu} \frac{\partial |\tilde{\boldsymbol{\Xi}}_{p,o}|}{\partial \phi_k} \right], \tag{41}$$

as a function of the parameter γ, the (distinct) eigenvalues $\{\sigma_i\}_{i=1}^{\mu_k}$ and $\{\omega_j\}_{j=1}^{\nu_k}$, and its derivatives w.r.t. ϕ_k. The details of the derivation are presented in Appendix B and the first order eigenvalue derivatives are given by [cf. (47)]

$$\tfrac{\partial \sigma_i}{\partial \phi_k} = \boldsymbol{u}_i^{\mathrm{H}} \tfrac{\partial \boldsymbol{S}}{\partial \phi_k} \boldsymbol{u}_i \quad i \in \{1, \ldots, \mu\} \tag{42a}$$

$$\tfrac{\partial \omega_j}{\partial \phi_k} = \boldsymbol{v}_j^{\mathrm{H}} \tfrac{\partial \boldsymbol{O}}{\partial \phi_k} \boldsymbol{v}_j \quad j \in \{1, \ldots \nu\}, \tag{42b}$$

where the unit-norm eigenvectors $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ are defined via the eigenvalue decompositions of $\boldsymbol{S}$ and $\boldsymbol{O}$, respectively. With (26) and Proposition 2, above matrix derivatives are

$$\tfrac{\partial \boldsymbol{S}}{\partial \phi_k} = -\psi_k \boldsymbol{C}_{\mathrm{Rx}}^{1/2,\mathrm{H}} \boldsymbol{A}_{\mathrm{Rx}}^{\mathrm{H}} \boldsymbol{C}_{\mathrm{est}}^{-1} \boldsymbol{B}_{\mathrm{Rx}} \boldsymbol{C}_{\mathrm{est}}^{-1} \boldsymbol{A}_{\mathrm{Rx}} \boldsymbol{C}_{\mathrm{Rx}}^{1/2} \tag{43a}$$

$$\tfrac{\partial \boldsymbol{O}}{\partial \phi_k} = \tfrac{\psi_k}{\alpha_{\mathrm{Rx}}^2} \tilde{\boldsymbol{U}} \boldsymbol{e}_k \boldsymbol{e}_k^{\mathrm{T}} \tilde{\boldsymbol{U}}^{\mathrm{H}}, \tag{43b}$$

respectively, where $\boldsymbol{e}_k$ denotes the canonical unit norm vector with a one at the k-th position and zeros elsewhere.

[3] Alternatively, we could exploit (36) for a precoder based gradient projection method (e.g., see [20])

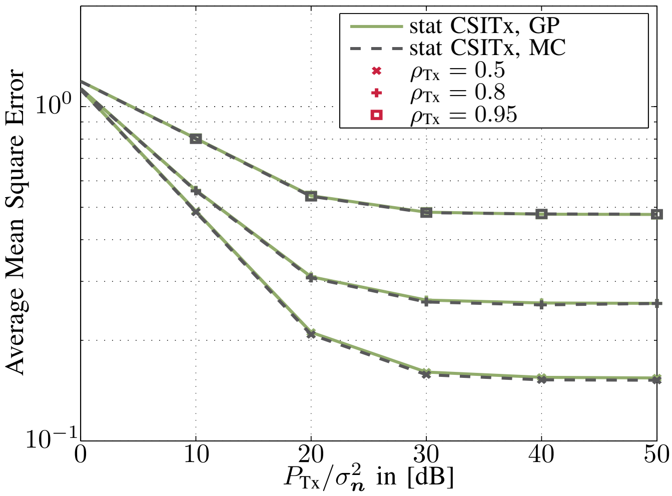

Figure 1. Average achievable ergodic MMSE (statistical CSITx) over $P_{\mathrm{Tx}}/\sigma_{\boldsymbol{n}}^2$ for $n = 4$, $m = l = 2$, and values for $\rho_{\mathrm{Tx}} \in \{0.5, 0.8, 0.95\}$.

VII. Numerical Results

In this section, we first depict the dependence of the proposed ergodic MMSE expression in Proposition 1 w.r.t. the correlations at the transmitter side. Then, we compare the achieved minimum average MMSE for partial and statistical CSITx with the obtained performance for no CSITx and the performance for designing both filters based on equal CSI [5].

We consider a system setup with $n = 4$ transmit antennas, $m = 2$ receive antennas, and a maximum of $l = 2$ data streams. The correlation matrices of the channel are constructed with the exponential correlation matrix model, i.e., $[\boldsymbol{R}_{\mathrm{Tx}}]_{i,j} = \rho_{\mathrm{Tx}}^{|i-j|}$ and $[\boldsymbol{R}_{\mathrm{Rx}}]_{i,j} = \rho_{\mathrm{Rx}}^{|i-j|}$, with $\rho_{\mathrm{Tx}}, \rho_{\mathrm{Rx}} \in [0, 1]$. The noise covariance matrices of the data link and the training models are $\boldsymbol{C}_{\boldsymbol{n}} = \sigma_{\boldsymbol{n}}^2 \mathbf{I}_m$, $\boldsymbol{C}_{\mathrm{q}} = \sigma_{\mathrm{q}}^2 \mathbf{I}_m$, and $\boldsymbol{C}_{\boldsymbol{n}_{\mathrm{Rx}}} = \sigma_{\boldsymbol{n}_{\mathrm{Rx}}}^2 \mathbf{I}_m$, respectively. For the plots in this section, we used $\sigma_{\mathrm{q}}, \sigma_{\boldsymbol{n}_{\mathrm{Rx}}} = 0.1$ and calculated the precoder and the equalizer for 1000 channel realizations and averaged over the resulting MSEs.

In Fig. 1, we depict the achievable minimum average MMSE over the SNR $P_{\mathrm{Tx}}/\sigma_{\boldsymbol{n}}^2$ for statistical CSITx, partial CSIRx, weak correlation at the receiver, i.e., $\rho_{\mathrm{Rx}} = 0.5$, and three different values for the channel correlations at the transmitter's side, $\rho_{\mathrm{Tx}} \in \{0.5, 0.8, 0.95\}$. Besides the proposed average MMSE minimization method using gradient projections (GP), we used a Monte-Carlo (MC) approach to illustrate correctness of the resulting curves. We see that the 'theoretical' results and the curves of the Monte-Carlo approach intersect. All three minimum average MMSE curves

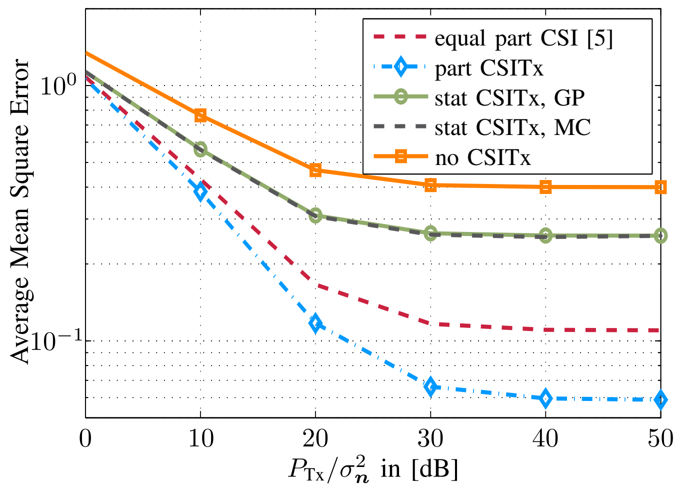

Figure 2. Average MSE versus $P_{\text{Tx}}/\sigma_{\boldsymbol{n}}^2$ for a MIMO point-to-point system with $n = 4$ Tx antennas, $m = 2$ Rx antennas, $l = 2$, and partial CSIRx.

saturate in the high SNR regime. Due to partial CSIRx, the receiver can only partially compensate for the imperfect precoding strategy. Interestingly, we can observe that the minimum average MMSE for statistical CSITx increases with the correlation coefficient ρ_{Tx} in the depicted SNR interval. From these results we draw the conclusion that the statistical CSITx performance strongly depends on the receiver's CSI and the experienced channel correlations at the transmitter side.

Now, we compare the minimum average MMSE for statistical CSITx and partial CSITx with the results for random precoding and the filter design based on equal partial CSITx and CSIRx, respectively. In Fig. 2, the obtained curves are plotted over the SNR for $\rho_{\text{Tx}} = 0.8$ and $\rho_{\text{Rx}} = 0.5$, i.e., weak receiver and medium transmitter correlation. In this context, *no CSITx* means that the transmitter has no information about the channel PDF $\mathrm{f}_{\boldsymbol{H}}(\boldsymbol{H})$ and the precoder is a randomly created scaled sub-unitary $n \times l$ matrix. Already in this medium correlation model, we achieve a remarkable gain if we exploit statistical CSITx instead of randomly creating the precoder. Statistical CSITx provides knowledge about the on average dominant right eigendirections of the channel. We expect that this gain increases with the correlation coefficient ρ_{Tx}. When $\rho_{\text{Tx}} \to 1$, the transmitter obtains almost full knowledge about the single dominant eigendirection of the channel, as $\boldsymbol{R}_{\text{Tx}}$ becomes a rank one matrix, whereas $\rho_{\text{Tx}} \to 0$ means that the transmitter has no CSITx as $\boldsymbol{R}_{\text{Tx}} \to \mathbf{I}_n$—all eigendirections of the channel become equally probable.

Regarding partial CSITx, Fig. 2 shows one curve for the AO approach in Section IV and one curve for the case when the better receiver's channel knowledge is neglected and both ends of the data transmission link use the worse partial CSITx for the joint filter design [5]. Comparing these two curves, we see that exploiting the better CSIRx can be highly beneficial, depending on the imposed variance of the quantization noise and the correlations at the transmitter's side.[4]

[4]In symmetric system setups with $l = m = n$, similar conclusions can be drawn. However, simulations have shown that the obtained gains are smaller.

Acknowledgments

The authors would like to thank the Deutsche Forschungsgemeinschaft (DFG), who supported this work under fund Jo 724/1-1.

Appendix A
Derivatives of Eigenvalues

The authors in [21] give a comparison of the methods for calculating first (and second) order derivatives of eigenvalues with algebraic multiplicity one. Calculating derivatives for eigenvalues with algebraic multiplicity larger than one is slightly more challenging (cf. [22], [23]). Next, we show that the first order derivatives of eigenvalues from Hermitian matrices can always be determined.

To this end, let the algebraic multiplicity of the l-th eigenvalue of the Hermitian matrix $\boldsymbol{Z}(t) \in \mathbb{C}^{M\times M}$, with parameter $t \in \mathbb{R}$, be m_l, where $1 \leq m_l \leq M$. Furthermore, let the diagonal matrix $\boldsymbol{\Lambda}_l \in \mathbb{R}^{m_l\times m_l}$ comprise these m_l equal eigenvalues, i.e., $\boldsymbol{\Lambda}_l \triangleq \lambda_l \mathbf{I}_{m_l}$, and $\boldsymbol{U}_l \in \mathbb{C}^{M\times m_l}$ denotes the corresponding (sub-)unitary eigenbasis. Then,

$$\boldsymbol{Z}\boldsymbol{U}_l = \boldsymbol{U}_l\boldsymbol{\Lambda}_l. \quad (44)$$

Note that $\boldsymbol{U}_l \triangleq \boldsymbol{U}_l'\boldsymbol{B}$ is also a (sub-)unitary basis for any unitary $\boldsymbol{B} \in \mathbb{C}^{m_l\times m_l}$, where $\boldsymbol{U}_l'$ is numerically determined.

Differentiating (44) w.r.t. t, multiplying the result with $\boldsymbol{U}_l^{\mathrm{H}}$ from the left, and reformulating the term, we obtain

$$\tfrac{\partial \boldsymbol{\Lambda}_l}{\partial t} = \boldsymbol{U}_l^{\mathrm{H}}\tfrac{\partial \boldsymbol{Z}}{\partial t}\boldsymbol{U}_l + \boldsymbol{U}_l^{\mathrm{H}}\boldsymbol{Z}\tfrac{\partial \boldsymbol{U}_l}{\partial t} - \boldsymbol{U}_l^{\mathrm{H}}\tfrac{\partial \boldsymbol{U}_l}{\partial t}\boldsymbol{\Lambda}_l. \quad (45)$$

Since $\boldsymbol{\Lambda}_l = \lambda_l\mathbf{I}_{m_l}$ at the current position t, it commutes with any other matrix. Moreover, due to (44), the last two summands in (45) cancel each other out, such that

$$\tfrac{\partial \boldsymbol{\Lambda}_l}{\partial t} = \boldsymbol{U}_l^{\mathrm{H}}\tfrac{\partial \boldsymbol{Z}}{\partial t}\boldsymbol{U}_l. \quad (46)$$

Note that the result must be diagonal. That is, given a numerically determined eigenvector matrix $\boldsymbol{U}_l'$, we calculate $\boldsymbol{U}_l = \boldsymbol{U}_l'\boldsymbol{B}_l$, where $\boldsymbol{B}_l$ must be a unitary eigenbasis of $\boldsymbol{U}_l'^{,\mathrm{H}}\tfrac{\partial \boldsymbol{Z}}{\partial t}\boldsymbol{U}_l'$. Then, the partial derivatives of the m_l equal eigenvalues are the diagonal elements of (46). When the algebraic multiplicity of λ_l is $m_l = 1$, (46) simplifies to

$$\tfrac{\partial \lambda_l}{\partial t} = \boldsymbol{u}_l^{\mathrm{H}}\tfrac{\partial \boldsymbol{Z}}{\partial t}\boldsymbol{u}_l. \quad (47)$$

Appendix B
Ergodic MMSE Derivation

Here, we determine the first order derivative of the ergodic MMSE in Proposition 1 w.r.t. $s \in \mathbb{R}$, where we assume that $\{\sigma_i\}_{i=1}^{\mu}$ and $\{\omega_j\}_{j=1}^{\mu}$ depend on s. With the product rule for differentiation, the derivative reads as

$$\frac{\partial\bar{\varepsilon}}{\partial s} = \tfrac{\partial\chi}{\partial s}\tfrac{\bar{\varepsilon}-l}{\chi} - c\chi\left[d\sum_{p=1}^{\mu}\tfrac{\partial|\boldsymbol{\Xi}_p|}{\partial s} + \gamma\sum_{p,o=1}^{\mu}\tfrac{\partial|\tilde{\boldsymbol{\Xi}}_{p,o}|}{\partial s}\right]. \quad (50)$$

$$
\left[\hat{\tilde{\boldsymbol{\Xi}}}_{p,o,q}\right]_{i,j} = \begin{cases} [\tilde{\boldsymbol{\Xi}}_{p,o}]_{i,j} & q\neq i \\ [\check{\boldsymbol{\Xi}}_{p,q}]_{i,j} & o\neq q=i \\ g_{\nu-1}(z_{i,j})\frac{\partial^2 z_{i,j}}{\partial\gamma\partial s} + \frac{\partial g_{\nu-1}(z_{i,j})}{\partial z_{i,j}}\frac{\partial z_{i,j}}{\partial\gamma}\frac{\partial z_{i,j}}{\partial s} & q=o=p=i\leq\mu \\ \sum_{k=\nu-\mu}^{\nu-1} k z_{i,j}^{k-2}[\nu-k]_k\big((k-1)\frac{\partial z_{i,j}}{\partial\gamma}\frac{\partial z_{i,j}}{\partial s}+z_{i,j}\frac{\partial^2 z_{i,j}}{\partial\gamma\partial s}\big) & p\neq o=q=i\leq\mu \end{cases} \tag{48}
$$

$$
\frac{\partial}{\partial x} g_y(x) = \frac{y}{x} g_y(x) + x^{y-2}\left[\left(1-\left(y-\tfrac{1}{x}\right)(1+x)\right) f\left(\tfrac{1}{x}\right) + \left(y-1-\tfrac{1}{x}\right)\right] \tag{49}
$$

In order to differentiate χ w.r.t. s, we remark that

$$
\frac{\partial}{\partial s}|\boldsymbol{\Sigma}|^{-(\nu-\mu)} = -\frac{\nu-\mu}{|\boldsymbol{\Sigma}|^{\nu-\mu}}\sum_{i=1}^{\mu}\frac{\frac{\partial}{\partial s}\sigma_i}{\sigma_i} \tag{51a}
$$

$$
\frac{\partial}{\partial s}\,\mathrm{v}_\mu(\boldsymbol{\Sigma})^{-1} = -\,\mathrm{v}_\mu(\boldsymbol{\Sigma})^{-1}\sum_{i<j}^{\mu}\frac{\frac{\partial}{\partial s}(\sigma_i-\sigma_j)}{\sigma_i-\sigma_j} \tag{51b}
$$

$$
\frac{\partial}{\partial s}\,\mathrm{v}_{\nu_k}(\boldsymbol{\Omega})^{-1} = -\,\mathrm{v}_\nu(\boldsymbol{\Omega})^{-1}\sum_{i<j}^{\nu}\frac{\frac{\partial}{\partial s}(\omega_i-\omega_j)}{\omega_i-\omega_j}, \tag{51c}
$$

respectively, to arrive at $\frac{\partial}{\partial s}\chi = \kappa_s\chi$, where

$$
\kappa_s = -(\nu-\mu)\sum_{i=1}^{\mu}\frac{\frac{\partial}{\partial s}\sigma_i}{\sigma_i} - \sum_{i<j}^{\mu}\frac{\frac{\partial}{\partial s}(\sigma_i-\sigma_j)}{\sigma_i-\sigma_j} - \sum_{i<j}^{\nu}\frac{\frac{\partial}{\partial s}(\omega_i-\omega_j)}{\omega_i-\omega_j}. \tag{52}
$$

The derivative $\frac{\partial|\boldsymbol{\Xi}_p|}{\partial s}$ is similarly determined as in (30), i.e.,

$$
\frac{\partial}{\partial s}|\boldsymbol{\Xi}_p| = \sum_{q=1}^{\nu}\left|\check{\boldsymbol{\Xi}}_{p,q}\right|, \tag{53}
$$

where the first μ rows in $\check{\boldsymbol{\Xi}}_{p,q}$ are equal to those of $\tilde{\boldsymbol{\Xi}}_{p,o}$ in (31), except for replacing $\frac{\partial z_{i,j}}{\partial\gamma}$ with $\frac{\partial z_{i,j}}{\partial s}$, i.e.,

$$
\left[\check{\boldsymbol{\Xi}}_{p,q}\right]_{i,j} = \begin{cases} [\boldsymbol{\Xi}_p]_{i,j} & i\neq q \\ g_{\nu-1}(z_{i,j})\frac{\partial z_{i,j}}{\partial s} & i=q=p\leq\mu \\ \sum\limits_{k=\nu-\mu}^{\nu-1} k z_{i,j}^{k-1}[\nu-k]_k\frac{\partial z_{i,j}}{\partial s} & p\neq i=q\leq\mu \\ (i-\mu-1)\omega_j^{i-\mu-2}\frac{\partial\omega_j}{\partial s} & q=i>\mu. \end{cases} \tag{54}
$$

The derivation of $\left|\tilde{\boldsymbol{\Xi}}_{p,o}\right|$, $i,o\in\{1,\ldots,\mu\}$, w.r.t. s reads as

$$
\frac{\partial}{\partial s}\left|\tilde{\boldsymbol{\Xi}}_{p,o}\right| = \sum_{q=1}^{\nu}\left|\hat{\tilde{\boldsymbol{\Xi}}}_{p,o,q}\right| \tag{55}
$$

where $\hat{\tilde{\boldsymbol{\Xi}}}_{p,o,q}$ is equal to $\tilde{\boldsymbol{\Xi}}_{p,o}$, except for its q-th row which is differentiated w.r.t. s. $\hat{\tilde{\boldsymbol{\Xi}}}_{p,o,q}$ is defined in (48), and the derivative of $g_y(x)$ w.r.t. x is given by (49). Note that $z_{i,j} = \gamma\sigma_i\omega_j$, for $i\in\{1,\ldots,\mu\}$ and $j\in\{1,\ldots,\nu\}$, such that

$$
\frac{\partial z_{i,j}}{\partial s} = \gamma\big(\frac{\partial\sigma_i}{\partial s}\omega_j + \sigma_i\frac{\partial\omega_j}{\partial s}\big). \tag{56}
$$

The second order derivative of $z_{i,j}$ is calculated as

$$
\frac{\partial^2 z_{i,j}}{\partial s\partial\gamma} = \frac{\partial\sigma_i}{\partial s}\omega_j + \sigma_i\frac{\partial\omega_j}{\partial s}. \tag{57}
$$

Inserting (55)–(51) into (50) yields

$$
\frac{\partial\bar{\varepsilon}}{\partial s} = \kappa_s(\bar{\varepsilon}-l) - c\chi\sum_{q=1}^{\nu}\left[d\sum_{p=1}^{\mu}\left|\check{\boldsymbol{\Xi}}_{p,q}\right| + \gamma\sum_{p,o=1}^{\mu}\left|\hat{\tilde{\boldsymbol{\Xi}}}_{p,o,q}\right|\right]. \tag{58}
$$

REFERENCES

[1] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx Beamforming Design for Multicarrier MIMO Channels: a Unified Framework for Convex Optimization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381–2401, Sept. 2003.

[2] Xi Zhang, D. P. Palomar, and B. Ottersten, "Robust Design of Linear Mimo Transceivers Under Channel Uncertainty," in *Proc. ICASSP 2006*, Toulouse, May 2006, vol. 4.

[3] M. Joham, M. Vonbun, and W. Utschick, "MIMO BC/MAC MSE Duality with Imperfect Transmitter and Perfect Receiver CSI," in *Proc. SPAWC 2010*, May 2010.

[4] Xi Zhang, D. P. Palomar, and B. Ottersten, "Statistically Robust Design of Linear MIMO Transceivers," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3678–3689, Aug. 2008.

[5] M. Ding and S. D. Blostein, "MIMO Minimum Total MSE Transceiver Design with Imperfect CSI at Both Ends," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1141–1150, Mar. 2009.

[6] D. Guo, S. Shamai, and S. Verdu, "Mutual Information and Minimum Mean-square Error in Gaussian Channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

[7] A. Gründinger, M. Joham, and W. Utschick, "Stochastic Transceiver Design in Multi-Antenna Channels with Statistical Channel State Information," Accepted for publication at ICASSP 2011.

[8] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*, Chapman & Hall, Boca Raton, FL, USA, 2000.

[9] Da-Shan Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading Correlation and its Effect on the Capacity of Multielement Antenna Systems," *IEEE Trans. Com.*, vol. 48, no. 3, pp. 502–513, Mar. 2000.

[10] G. H. Golub and C. F. van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, NY, USA, 3rd edition, 1996.

[11] U. Salim, D. Gesbert, and D. Slock, "Combining Training and Quantized Feedback in Multi-Antenna Reciprocal Channels," *IEEE Trans. Signal Process.*, 2010, submitted.

[12] D. P. Palomar and Y. Jiang, "MIMO Transceiver Design via Majorization Theory," *Foundations and Trends in Communications and Information Theory*, vol. 3, no. 4, pp. 331–551, 2006.

[13] A. Wiesel, Y. C. Eldar, and S. Shamai, "Linear Precoding via Conic Optimization for Fixed MIMO Receivers," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 161–176, Jan. 2006.

[14] S. Shi, M. Schubert, and H. Boche, "Rate Optimization for Multiuser MIMO Systems With Linear Processing," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 4020–4030, Aug. 2008.

[15] R. Hunger, D. A. Schmidt, M. Joham, and W. Utschick, "A General Covariance-Based Optimization Framework Using Orthogonal Projections," in *Proc. SPAWC 2008*, May 2008, pp. 76–80.

[16] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover Publications Inc., 1st edition, 1964.

[17] M. Kiessling, *Statistical Analysis and Transmit Prefiltering for MIMO Wireless Systems in Correlated Fading Environments*, Dissertation, University of Stuttgart, June 2004.

[18] M. Chiani, M. Z. Win, and Hyundong Shin, "MIMO Networks: The Effects of Interference," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 336–349, Jan. 2010.

[19] E. A. Jorswieck and H. Boche, "Channel Capacity and Capacity-Range of Beamforming in MIMO Wireless Systems Under Correlated Fading with Covariance Feedback," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1543–1553, Sept. 2004.

[20] M. Mezghani, M. Joham, R. Hunger, and W. Utschick, "Transceiver Design for Multiuser MIMO Systems," in *Proc. ITG WSA*, Mar. 2006.

[21] D. V. Murthy and R. T. Haftka, "Derivatives of Eigenvalues and Eigenvectors of a General Complex Matrix," *Int. Journ. Numer. Meth. Eng.*, vol. 26, no. 2, pp. 293–311, June 2005.

[22] A. L. Andrew, K.-W. Eric Chu, and P. Lancaster, "Derivatives of Eigenvalues and Eigenvectors of Matrix Functions," *SIAM JMAA*, vol. 14, no. 4, pp. 903–926, May 1993.

[23] N. P. van der Aa, H. G. ter Morsche, and R. R. M. Mattheij, "Computation of Eigenvalue and Eigenvector Derivatives for a General Complex-Valued Eigensystem," *Electronic Journal of Linear Algebra*, vol. 16, no. 1, pp. 300–314, Oct. 2007.