On the sensitivity of transmit Wiener filter with respect to channel and noise second-order statistics estimation errors

Master Thesis

By

Despoina Tsipouridou

Submitted to the Department of Electronic & Computer Engineering in partial fulfillment of the requirements for the Master Degree.

Technical University of Crete

Advisor: Associate Professor Liavas Athanasios

Co-advisor: Professor Sidiropoulos Nikolaos

Co-advisor: Associate Professor Potamianos Alexandros

November 2006

Ευχαριστίες

Θα ήθελα να ευχαριστήσω ιδιαιτέρως τον επιβλέποντα καθηγητή μου, Αναπληρωτή Καθηγητή κ. Αθανάσιο Λιάβα, για την πολύτιμη υποστήριξή του, επιστημονική και ψυχολογική, την υπομονή, τις ευκαιρίες και τον χρόνο που μου αφιέρωσε τα δύο αυτά χρόνια. Χωρίς την διαρκή καθοδήγηση και βοήθειά του, δεν θα ήταν δυνατή η ολοκλήρωση της παρούσας διατριβής, και όχι μόνο.

NOTATION

TxWF	transmit Wiener filter
EMSE	excess mean-square error
MMSE	minimum mean-square error
SOS	second-order statistics
H	channel matrix
P	pre-coding (or pre-equalization) matrix
$\tilde{P}$	scaled pre-coding matrix
n_t	number of transmit antennas
n_r	number of receive antennas
σ^2	noise variance
R_s	input covariance matrix
R_n	noise covariance matrix
E_{tr}	transmit power
α_o	inverse SNR
ΔH	channel estimation error
N_{tr}	dimension of the training block
ΔR_n	noise SOS estimation error

CONTENTS

Abstract

We consider the behavior of the transmit Wiener filter under channel and noise second-order statistics (SOS) uncertainties. We study the influence of channel and noise SOS uncertainties separately, by assuming that only one quantity is estimated at a time, while the other is perfectly known. Using results from matrix perturbation theory, we derive second-order approximations to the excess mean-square error (EMSE) induced by using the channel and noise estimates as if they were the true quantities. In the high SNR cases, we develop simple and informative approximations to the EMSE for both cases. In the scenario we study, it turns out that channel estimation errors are much more significant than noise SOS estimation errors.

I. Introduction

Joint optimization of transmit and receive filters for combatting frequency selectivity and/or interstream interference in MIMO or multiuser systems has been extensively studied (see, for example, [1], [2] and the references therein).

If we want to keep the mobile units as simple as possible, then we may consider separate transmit or receive processing. The transmit matched filter (TxMF), the transmit zero-forcing filter (TxZF) and the transmit Wiener filter (TxWF) are three linear precoding (or pre-equalization) structures that combat frequency selectivity and/or inter-stream interference and keep the receivers simple, because the only assumed receiver processing is a scalar scaling (see [1], [2] and the references therein). This is particularly appealing in the broadcast scenario, where we want to keep the receivers of non-cooperative users as simple as possible.

The TxWF outperforms the two other structures in terms of mean-square error (MSE) and bit-error rate (BER). If the channel matrix and the input and noise second-order statistics (SOS) are perfectly known at the transmitter (due to, for example, TDD or feedback information channel), then the TxWF can be computed. If the channel and/or the noise SOS are unknown at the transmitter, as it is usually the case, then a common approach towards the design of the TxWF is to estimate the unknown quantities and use the estimates as if they were the true quantities.

We consider that only one quantity is estimated at a time, while the other is perfectly known, and develop second-order approximations to the excess MSE (EMSE) in terms of the

1) channel estimation error covariance matrix
2) noise SOS estimation error first and second-order statistics.

Then, we consider optimal training and derive simple and informative EMSE bounds in the high SNR cases. The bounds appear to be good approximations to the EMSE. It turns out that in case 1) the EMSE is proportional to the minimum MSE (MMSE), while in case 2) the EMSE is proportional to the squared noise variance, σ^4. We observe that the error induced by the channel estimation error is more significant than that induced by the noise SOS estimation error.

A. Notation and useful matrix results

We use ^{T}, * and ^{H} for the transpose, componentwise conjugate and conjugate transpose, respectively. $\mathbf{E}[\cdot]$ denotes expectation, $\mathbf{I}_M$ denotes the $M \times M$ identity matrix, $\mathrm{Re}\{\cdot\}$ extracts the real part of a complex number, $\mathrm{tr}(\cdot)$ and $||\cdot||_F$ denote, respectively, the trace and the Frobenius norm of the matrix argument. $A \otimes B$ denotes the Kronecker product of A and B and $\mathrm{vec}(\cdot)$ is the vectorization operator.

Next, we present some useful matrix results used in this work. If $A \in \mathbb{C}^{M \times N}$ and $\alpha \in \mathbb{C}$ then [1]

$$(A^H A + \alpha I_N)^{-1} A^H = A^H (A A^H + \alpha I_M)^{-1}. \tag{1}$$

If ΔA is a perturbation to matrix A, then a first-order approximation to the inverse of $A + \Delta A$ is given by [4, p. 131]

$$(A + \Delta A)^{-1} = A^{-1} - A^{-1} \Delta A A^{-1}. \tag{2}$$

For matrices with compatible dimensions [3, ch. 2, 4 and 9]

$$\mathrm{tr}(ABCD) = \mathrm{vec}^T \left(D^T\right) \left(C^T \otimes A\right) \mathrm{vec}(B) \tag{3}$$

$$\mathrm{vec}(ABC) = \left(C^T \otimes A\right) \mathrm{vec}(B) \tag{4}$$

$$AB \otimes CD = (A \otimes C)(B \otimes D) \tag{5}$$

$$\mathrm{vec}(A^T) = K_{mn} \mathrm{vec}(A) \tag{6}$$

$$K_{pm}(A \otimes B) K_{nq} = (B \otimes A), \quad A(m \times n),\ B(p \times q) \tag{7}$$

$$\mathrm{tr}(A \otimes B) = \mathrm{tr}(A)\mathrm{tr}(B) \tag{8}$$

where K_{mn} is the $mn \times mn$ commutation matrix [3, p. 9].

The trace of a $m \times m$ matrix A is defined as

$$\mathrm{tr}(A) = \sum_{i=1}^{m} A_{ii}.$$

If the eigenvalues of A are, λ_i, $i = 1, \ldots, m$, then it can be shown that

$$\mathrm{tr}(A) = \sum_{i=1}^{m} \lambda_i. \tag{9}$$

Some useful results, concerning the derivatives of matrix functions are [6]

$$\frac{\partial \mathrm{tr}\left(AX\right)}{\partial X} = A^T \tag{10}$$

$$\frac{\partial \mathrm{tr}\left(AX^H\right)}{\partial X} = \mathbf{0} \tag{11}$$

$$\frac{\partial \mathrm{tr}\left(XA_0X^HA_1\right)}{\partial X} = A_1^T X^* A_0^T \tag{12}$$

$$\frac{\partial AX^*}{\partial X^*} = A^T. \tag{13}$$

We also remind that for a function g of a complex-valued matrix X [7]

$$\nabla_X^2(g) \triangleq \nabla_{X^*}\left[\nabla_X(g)\right].$$

The structure of this report is as follows. In Section II, we present the derivation of the TxWF, assuming that the channel and the noise SOS are known at the transmitter [2]. In Section III, we develop second-order approximations to the excess MSE, assuming either channel or noise SOS estimation errors. We continue with Section IV, where we support our theoretical results with simulations.

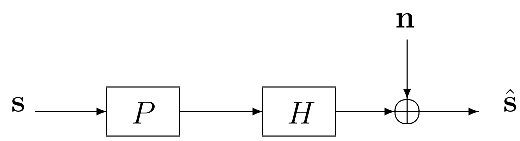

Fig. 1. System model

II. Transmit Wiener Filter

We consider the pre-equalized, baseband-equivalent, discrete-time MIMO system, with n_t transmit antennas and n_r receiver antennas (with $n_r \leq n_t$), depicted in Fig. 1 and described by the expression

$$\hat{\mathbf{s}} = HP\mathbf{s} + \mathbf{n} \tag{14}$$

where $\mathbf{s}$ is the $n_r \times 1$ input signal, P is the $n_t \times n_r$ pre-coding matrix, H is the $n_r \times n_t$ channel matrix and $\mathbf{n}$ is the $n_r \times 1$ additive channel noise. This model is particularly appealing in the broadcast scenario, when the users cannot cooperate in order to combat inter-symbol and/or inter-stream interference; thus, the need for pre-equalization is imperative. In this case, the i-th element of vector $\mathbf{s}$ is the symbol intended for the i-th receiver. Vectors $\mathbf{s}$ and $\mathbf{n}$ are complex-valued, circular, zero-mean, independent with covariance matrices R_s and R_n, respectively. Further, the noise $\mathbf{n}$ is assumed to be Gaussian.

Our aim is to find the transmit filter that minimizes the MSE, $\mathbf{E}\left[\|\mathbf{s} - \hat{\mathbf{s}}\|_2^2\right]$, under the transmit power constraint

$$\mathbf{E}\left[\|P\mathbf{s}\|_2^2\right] = E_{tr}. \tag{15}$$

It can be shown that in order to fulfill this power constraint, we should allow the transmit filter to generate a receive signal with possibly different amplitude from the original desired signal [1]. Therefore, we replace the estimate $\hat{\mathbf{s}}$ by the weighted version $\beta^{-1}\hat{\mathbf{s}}$ ($\beta \in \mathbb{R}^+$), and the MSE function is defined as [1]

$$\text{mse}(P, \beta) \triangleq \mathbf{E}\left[\|\mathbf{s} - \beta^{-1}\hat{\mathbf{s}}\|_2^2\right]. \tag{16}$$

We can write function $\text{mse}(\cdot)$ analytically as

$$\text{mse}(P, \beta) = \text{tr}(R_s) - \beta^{-1}\text{tr}(HPR_s) - \beta^{-1}\text{tr}(R_s P^H H^H) + \beta^{-2}\text{tr}(HPR_s P^H H^H) + \beta^{-2}\text{tr}(R_n). \tag{17}$$

Using (15) and (16), the optimization problem becomes [1]

$$(P_o, \beta_o) = \arg\min_{P,\beta}\ \text{mse}(P, \beta) \quad \text{s.t.} \quad \mathbf{E}\left[\|P\mathbf{s}\|_2^2\right] = E_{tr}. \tag{18}$$

In order to solve this problem, we construct the Langrangian function

$$L(P, \beta, \lambda) = \text{mse}(P, \beta) - \lambda\left(\text{tr}(PR_sP^H) - E_{tr}\right) \tag{19}$$

with $\lambda \in \mathbb{R}$, and set its derivatives with respect to P, β and λ equal to zero. By setting the derivative of L with respect to P equal to zero, and using (17) and (10)-(12), we obtain

$$\begin{aligned} \frac{\partial L}{\partial P} &= -\beta^{-1}H^T R_s^T + \beta^{-2}H^T H^* P^* R_s^T - \lambda P^* R_s^T \\ &= (-\beta^{-1}H^T + \beta^{-2}H^T H^* P^* - \lambda P^*)R_s^T = 0. \end{aligned} \tag{20}$$

Since R_s^T is invertible, solving (20) with respect to P, we get

$$P = \beta(H^H H - \lambda\beta^2 I_{n_t})^{-1} H^H. \tag{21}$$

From the constraint (15), using (21), we obtain

$$\begin{aligned} E_{tr} &= \text{tr}\left(\beta\left(H^H H - \lambda\beta^2 I_{n_t}\right)^{-1} H^H R_s H \left(H^H H - \lambda\beta^2 I_{n_t}\right)^{-1}\beta\right) \\ &= \beta^2 \text{tr}\left((H^H H - \lambda\beta^2 I_{n_t})^{-2} H^H R_s H\right). \end{aligned} \tag{22}$$

The weight β is

$$\beta = \sqrt{\frac{E_{tr}}{\text{tr}\left(\left(H^H H - \lambda\beta^2 I_{n_t}\right)^{-2} H^H R_s H\right)}}. \tag{23}$$

Unfortunately it is not possible to solve for β in (23). Thus, we shall proceed in a indirect way. If we put $\alpha = -\lambda\beta^2$, ($\alpha \in \mathbb{R}$), in (23) we obtain

$$\beta = \sqrt{\frac{E_{tr}}{\text{tr}\left(\left(H^H H + \alpha I_{n_t}\right)^{-2} H^H R_s H\right)}} \ .$$

Thus, β and P are functions of α

$$\beta(\alpha) = \sqrt{\frac{E_{tr}}{\text{tr}\left(\left(H^H H + \alpha I_{n_t}\right)^{-2} H^H R_s H\right)}}$$

and

$$P(\alpha) = \beta(\alpha)(H^H H + \alpha I_{n_t})^{-1} H^H.$$

Thus, the constrained optimization of (18) can be reduced to the following unconstrained optimization with respect to α, since the constraint is fulfilled with the choice of β [1]

$$\alpha_o = \arg\min_{\alpha} \ \varepsilon\left(\alpha\right) \tag{24}$$

where

$$\begin{aligned}\varepsilon(\alpha) \triangleq \text{mse}\left(P\left(\alpha\right), \beta\left(\alpha\right)\right) = \text{tr}(R_s) - 2\text{tr}\left(H(H^H H + \alpha I_{n_t})^{-1} H^H R_s\right) \\ + \text{tr}\big(H(H^H H + \alpha I_{n_t})^{-1} H^H R_s H (H^H H + \alpha I_{n_t})^{-1} H^H\big) \\ + \text{tr}\big((H^H H + \alpha I_{n_t})^{-2} H^H R_s H\big) \frac{\text{tr}(R_n)}{E_{tr}} \ .\end{aligned} \tag{25}$$

Using (1), we rewrite the term of the second line of (25) as

$$\begin{aligned}&\text{tr}\big(H(H^H H + \alpha I_{n_t})^{-1} H^H R_s H (H^H H + \alpha I_{n_t})^{-1} H^H\big) \\ &\qquad = \text{tr}\big((H^H H + \alpha I_{n_t})^{-2} H^H H H^H R_s H\big).\end{aligned} \tag{26}$$

Using the expression of (26), function $\varepsilon(\cdot)$ from (25) becomes

$$\begin{aligned}\varepsilon(\alpha) = \text{tr}(R_s) - 2\text{tr}\big((H^H H + \alpha I_{n_t})^{-1} H^H R_s H\big) \\ + \text{tr}\big((H^H H + \alpha I_{n_t})^{-2} H^H H H^H R_s H\big) \\ + \text{tr}\big((H^H H + \alpha I_{n_t})^{-2} H^H R_s H\big) \frac{\text{tr}(R_n)}{E_{tr}}.\end{aligned} \tag{27}$$

At this point, we can find the optimal α by setting the derivative of $\varepsilon(\alpha)$, with respect to α, equal to zero

$$\begin{aligned}\frac{\partial \varepsilon(\alpha)}{\partial \alpha} &= 2\text{tr}\big((H^H H + \alpha I_{n_t})^{-2} H^H R_s H\big) \\ &\quad - 2\text{tr}\big((H^H H + \alpha I_{n_t})^{-3} H^H H H^H R_s H\big) \\ &\quad - 2\text{tr}\big((H^H H + \alpha I_{n_t})^{-3} H^H R_s H\big) \frac{\text{tr}(R_n)}{E_{tr}} = 0.\end{aligned} \tag{28}$$

Because of the linearity of the trace, the above equation can be written as

$$\text{tr}\left(\left(H^H H + \alpha I_{n_t}\right)^{-3} \left(\alpha I_{n_t} - \frac{\text{tr}(R_n)}{E_{tr}} I_{n_t}\right) H^H R_s H\right) = 0 \tag{29}$$

which is equivalent to

$$\left(\alpha - \frac{\text{tr}(R_n)}{E_{tr}}\right) \text{tr}\left(\left(H^H H + \alpha I_{n_t}\right)^{-3} H^H R_s H\right) = 0.$$

Using the fact that $\text{tr}\left(\left(H^H H + \alpha I_{n_t}\right)^{-3} H^H R_s H\right) > 0$, we obtain

$$\alpha_o = \frac{\text{tr}(R_n)}{E_{tr}}. \tag{30}$$

Term α_o has been used in [2, eq. 5] as a measure of inverse SNR. Thus, as high SNR cases we consider the cases that lead to $\alpha_o \ll 1$.

Thus, using (21), (23) and (30), the closed form solution of the optimization problem (18) is

$$\begin{aligned}P_o = \beta_o \tilde{P}_o \quad \text{and} \quad \beta_o = \sqrt{\frac{E_{tr}}{\text{tr}(\tilde{P}_o R_s \tilde{P}_o^H)}} \\ \text{with} \quad \tilde{P}_o = \left(H^H H + \alpha_o I_{n_t}\right)^{-1} H^H.\end{aligned} \tag{31}$$

In the sequel we assume that the covariance matrix of s is

$$R_s = I_{n_r}. \tag{32}$$

Using that $P_o = \beta_o \tilde{P}_o$, we express the minimum MSE (MMSE) as

$$\text{MMSE} \triangleq \text{mse}(P_o, \beta_o) = \text{tr}(I_{n_r}) - 2\text{tr}(\tilde{P}_o H) + \text{tr}(H \tilde{P}_o \tilde{P}_o^H H^H) + \alpha_o \, \text{tr}(\tilde{P}_o \tilde{P}_o^H). \tag{33}$$

III. Computation of the excess MSE

We continue with the computation of the excess MSE assuming channel and noise SOS estimation errors. We start by giving a general expression for the excess MSE. We denote with $\hat{\tilde{P}}$ the estimate for the scaled pre-coding matrix $\tilde{P}_o$, for the cases where we have estimated either the channel or the noise SOS. We will consider separately each case in detail further on.

We return to (16) and use the pre-coding matrix estimate $\hat{P} = \hat{\beta}\hat{\tilde{P}}$ as if it were the true pre-coding matrix. Then, the MSE achieved by $\hat{\tilde{P}}$ is

$$\mathrm{MSE}(\hat{\tilde{P}}) \triangleq \mathrm{mse}(\hat{P}, \hat{\beta}) = \mathrm{tr}(I_{n_t}) - 2\mathrm{Re}\left(\mathrm{tr}(\hat{\tilde{P}}H)\right) + \mathrm{tr}(H\hat{\tilde{P}}\hat{\tilde{P}}^H H^H) + \frac{\mathrm{tr}(\hat{\tilde{P}}\hat{\tilde{P}}^H)}{E_{tr}}\mathrm{tr}(R_n). \tag{34}$$

Expanding function $\mathrm{MSE}(\cdot)$ around $\tilde{P}_o$, we obtain the second-order expansion

$$\mathrm{MSE}(\hat{\tilde{P}}) = \mathrm{MSE}(\tilde{P}_o + \Delta\tilde{P}) = \mathrm{MSE}(\tilde{P}_o) + \mathrm{tr}(\Delta\tilde{P}^H \mathrm{MSE}''(\tilde{P}_o)\Delta\tilde{P}) \tag{35}$$

where $\mathrm{MSE}''(\tilde{P}_o)$ is the second derivative of the MSE evaluated at the point $\tilde{P}_o$ and $\Delta\tilde{P} \triangleq \hat{\tilde{P}} - \tilde{P}_o$. Using (12) and (13) we obtain

$$\mathrm{MSE}''(\tilde{P}_o) = H^H H + \alpha_o I_{n_t}. \tag{36}$$

A general expression for the excess MSE (EMSE) can be obtained by taking expectation in (35) with respect to estimation errors

$$\begin{aligned}
\mathrm{EMSE}(\hat{\tilde{P}}) &\triangleq \mathbf{E}\left[\mathrm{MSE}(\hat{\tilde{P}}) - \mathrm{MSE}(\tilde{P}_o)\right] \\
&= \mathbf{E}\left[\mathrm{tr}\Big(\Delta\tilde{P}^H \mathrm{MSE}''(\tilde{P}_o)\Delta\tilde{P}\Big)\right] \\
&\overset{(36)}{=} \mathbf{E}\left[\mathrm{tr}\Big(\Delta\tilde{P}^H (H^H H + \alpha_o I_{n_t})\Delta\tilde{P}\Big)\right].
\end{aligned} \tag{37}$$

From (37), we can derive the EMSE for the above mentioned cases, by expressing the term $\Delta\tilde{P}$ in the appropriate form.

A. Channel estimation errors

Starting with the channel estimation errors, we express the estimate of the filtering matrix $\hat{H}$ as

$$\hat{H} \triangleq H + \Delta H \tag{38}$$

where ΔH denotes the channel estimation error. ΔH is assumed to be zero-mean, complex-valued, circular, with

$$\mathbf{R}_{\mathrm{vec}(\Delta H)} \triangleq \mathbf{E}\left[\mathrm{vec}(\Delta H)\mathrm{vec}^H(\Delta H)\right] = \Sigma.$$

If we use $\hat{H}$ as if it were the true channel matrix in (31), we compute the scaled pre-coding matrix

$$\hat{\tilde{P}} = \left(\hat{H}^H\hat{H} + \alpha_o I_{n_t}\right)^{-1}\hat{H}^H. \tag{39}$$

We continue by applying (2) to (39), taking into consideration (38) and ignoring products of error terms

$$\begin{aligned}
\hat{\tilde{P}} &= \left(\hat{H}^H\hat{H} + \alpha_o I_{n_t}\right)^{-1}\hat{H}^H \\
&= \left((H^H + \Delta H^H)(H + \Delta H) + \alpha_o I_{n_t}\right)^{-1}(H^H + \Delta H^H) \\
&= \Big(\underbrace{H^H H + \alpha_o I_{n_t}}_{A} + \underbrace{H^H \Delta H + \Delta H^H H}_{K_\Delta}\Big)^{-1}(H^H + \Delta H^H) \\
&\overset{(2)}{=} \left[(H^H H + \alpha_o I_{n_t})^{-1} - (H^H H + \alpha_o I_{n_t})^{-1} K_\Delta (H^H H + \alpha_o I_{n_t})^{-1}\right](H^H + \Delta H^H) \\
&= \tilde{P}_o - (H^H H + \alpha_o I_{n_t})^{-1}(K_\Delta \tilde{P}_o - \Delta H^H).
\end{aligned} \tag{40}$$

Thus, a first-order approximation to term $\Delta\tilde{P}$ is

$$\Delta\tilde{P} = -(H^H H + \alpha_o I_{n_t})^{-1}(K_\Delta \tilde{P}_o - \Delta H^H). \tag{41}$$

Having expressed $\Delta\tilde{P}$ as a function of the channel estimation error ΔH we return to (37) and write for the EMSE, using (3)

$$\begin{aligned}
\mathrm{EMSE}(\hat{\tilde{P}}) &= \mathbf{E}\left[\mathrm{tr}\left(\Delta\tilde{P}^H(H^H H + \alpha_o I_{n_t})\Delta\tilde{P}\right)\right] \\
&= \mathbf{E}\left[\mathrm{tr}\Big((H^H H + \alpha_o I_{n_t})^{-1}(K_\Delta\tilde{P}_o - \Delta H^H)\ I_{n_r}\ (K_\Delta\tilde{P}_o - \Delta H^H)^H\Big)\right] \\
&\overset{(3)}{=} \mathbf{E}\left[\mathrm{tr}\Big(\mathrm{vec}^H\big(K_\Delta\tilde{P}_o - \Delta H^H\big)\big(I_{n_r}\otimes(H^H H + \alpha_o I_{n_t})^{-1}\big)\mathrm{vec}(K_\Delta\tilde{P}_o - \Delta H^H)\Big)\right] \\
&= \mathrm{tr}\left(\big(I_{n_r}\otimes(H^H H + \alpha_o I_{n_t})^{-1}\big)\mathbf{E}\left[\mathrm{vec}(K_\Delta\tilde{P}_o - \Delta H^H)\mathrm{vec}^H\big(K_\Delta\tilde{P}_o - \Delta H^H\big)\right]\right).
\end{aligned} \tag{42}$$

Before taking the expectation in (42), we express term $\mathcal{T}_1 \triangleq \mathrm{vec}(K_\Delta\tilde{P}_o - \Delta H^H)$, using (4) and the definition of K_Δ from (40), as

$$\begin{aligned}
\mathcal{T}_1 &\triangleq \mathrm{vec}\left(K_\Delta\tilde{P}_o - \Delta H^H\right) = \mathrm{vec}\left(H^H\Delta H\tilde{P}_o\right) + \mathrm{vec}\left(\Delta H^H H\tilde{P}_o\right) - \mathrm{vec}(\Delta H^H) \\
&= \mathrm{vec}(H^H\Delta H\tilde{P}_o) + \mathrm{vec}\left(I_{n_t}\Delta H^H\left(H\tilde{P}_o - I_{n_r}\right)\right) \\
&\overset{(4)}{=} (\tilde{P}_o^T\otimes H^H)\mathrm{vec}(\Delta H) + \big((\tilde{P}_o^T H^T - I_{n_r})\otimes I_{n_t}\big)\mathrm{vec}(\Delta H^H).
\end{aligned} \tag{43}$$

Using (6) and noting that the expectation of the cross terms of the product $\mathcal{T}_1\mathcal{T}_1^H$ vanishes, due to the circular symmetry of $\mathrm{vec}(\Delta H)$, we can go back to (42) and write for the EMSE

$$\begin{aligned}
\mathrm{EMSE}(\hat{\tilde{P}}) &= \mathrm{tr}\Big(\big(I_{n_r}\otimes(H^H H + \alpha_o I_{n_t})^{-1}\big)\ \mathbf{E}\left[\mathcal{T}_1\mathcal{T}_1^H\right]\Big) \\
&= \mathrm{tr}\Big(\big(I_{n_r}\otimes(H^H H + \alpha I_{n_t})^{-1}\big)(\tilde{P}_o^T\otimes H^H)\mathbf{R}_{\mathrm{vec}(\Delta H)}(\tilde{P}_o^*\otimes H)\Big) \\
&+ \mathrm{tr}\Big(\big(I_{n_r}\otimes(H^H H + \alpha_o I_{n_t})^{-1}\big)\big((\tilde{P}_o^T H^T - I_{n_r})\otimes I_{n_t}\big)K_{n_t n_r}\mathbf{R}^*_{\mathrm{vec}(\Delta H)}K^T_{n_t n_r}\big((H^*\tilde{P}_o^* - I_{n_r})\otimes I_{n_t}\big)\Big) \\
&= \mathcal{A}_1 + \mathcal{A}_2
\end{aligned} \tag{44}$$

where

$$\begin{aligned}
\mathcal{A}_1 &\triangleq \mathrm{tr}\Big(\big(I_{n_r}\otimes(H^H H + \alpha_o I_{n_t})^{-1}\big)(\tilde{P}_o^T\otimes H^H)\mathbf{R}_{\mathrm{vec}(\Delta H)}(\tilde{P}_o^*\otimes H)\Big) \\
&= \mathrm{tr}\Big((\tilde{P}_o^*\otimes H)\big(I_{n_r}\otimes(H^H H + \alpha_o I_{n_t})^{-1}\big)(\tilde{P}_o^T\otimes H^H)\mathbf{R}_{\mathrm{vec}(\Delta H)}\Big) \\
&\overset{(5)}{=} \mathrm{tr}\Big(\big(\tilde{P}_o^*\tilde{P}_o^T\otimes H(H^H H + \alpha_o I_{n_t})^{-1}H^H\big)\mathbf{R}_{\mathrm{vec}(\Delta H)}\Big).
\end{aligned} \tag{45}$$

and

$$
\begin{aligned}
\mathcal{A}_2 &\triangleq \mathrm{tr}\Big(\big(I_{n_r} \otimes (H^H H + \alpha_o I_{n_t})^{-1}\big)\big((\tilde{P}_o^T H^T - I_{n_r}) \otimes I_{n_t}\big) K_{n_t n_r} \mathbf{R}^*_{\mathrm{vec}(\Delta H)} K^T_{n_t n_r} \big((H^* \tilde{P}_o^* - I_{n_r}) \otimes I_{n_t}\big)\Big) \\
&= \mathrm{tr}\Big(K^T_{n_t n_r}\big((H^* \tilde{P}_o^* - I_{n_r}) \otimes I_{n_t}\big)\big(I_{n_r} \otimes (H^H H + \alpha_o I_{n_t})^{-1}\big)\big((\tilde{P}_o^T H^T - I_{n_r}) \otimes I_{n_t}\big) K_{n_t n_r} \mathbf{R}^*_{\mathrm{vec}(\Delta H)}\Big) \\
&= \mathrm{tr}\Big(K^T_{n_t n_r}\big((H^* \tilde{P}_o^* - I_{n_r})(\tilde{P}_o^T H^T - I_{n_r}) \otimes (H^H H + \alpha_o I_{n_t})^{-1}\big) K_{n_t n_r} \mathbf{R}^*_{\mathrm{vec}(\Delta H)}\Big) \\
&\overset{(7)}{=} \mathrm{tr}\Big(\big((H^H H + \alpha_o I_{n_t})^{-1} \otimes (H^* \tilde{P}_o^* - I_{n_r})(\tilde{P}_o^T H^T - I_{n_r})\big) \mathbf{R}^*_{\mathrm{vec}(\Delta H)}\Big).
\end{aligned}
\tag{46}
$$

We have expressed the EMSE in terms of $\mathbf{R}_{\mathrm{vec}(\Delta H)}$. Expressions (44)-(46) are admittedly complicated and do not provide significant insight. In the sequel, we shall consider the high SNR cases and derive a simple and informative expression for the EMSE.

1) Simplifications in the high SNR cases: For sufficiently high SNR, we can derive a simple approximation for the EMSE. First, we determine the covariance matrix of $\mathrm{vec}(\Delta H)$, assuming that we have estimated the channel in an optimal way.

Assuming that we have used training-based estimation with optimal training block and the noise is spatially and temporally white, circularly symmetric complex Gaussian with variance σ^2 (see Appendix), the covariance matrix of $\mathrm{vec}(\Delta H)$ is given by [5, p.175]

$$\mathbf{R}_{\mathrm{vec}(\Delta H)} = \frac{\sigma^2}{N_{tr}} I_{n_t n_r} \tag{47}$$

where N_{tr} is the number of independent columns of the training block we used for estimating H.

We continue by writing terms $\mathcal{A}_1$ and $\mathcal{A}_2$, from (45) and (46) respectively, in a more simple form. We start with $\mathcal{A}_1$, which becomes

$$\mathcal{A}_1 = \frac{\sigma^2}{N_{tr}} \mathrm{tr}\Big(\tilde{P}_o^* \tilde{P}_o^T \otimes H(H^H H + \alpha_o I_{n_t})^{-1} H^H \Big). \tag{48}$$

At this point we will examine term $H(H^H H + \alpha_o I_{n_t})^{-1} H^H$. It can be shown that ([4, p. 138])

$$\lambda_i \left(H \left(H^H H + \alpha_o I_{n_t} \right)^{-1} H^H \right) = \frac{\lambda_i \left(H^H H \right)}{\lambda_i \left(H^H H \right) + \alpha_o} \leq 1 \tag{49}$$

where $\lambda_i(\cdot)$ denotes the i-th eigenvalue of the matrix argument. Thus, for the trace of this term we can write that

$$\mathrm{tr} \left(H \left(H^H H + \alpha_o I_{n_t} \right)^{-1} H^H \right) \overset{(49)}{=} \sum_{i=1}^{n_r} \frac{\lambda_i \left(H^H H \right)}{\lambda_i \left(H^H H \right) + \alpha_o} \overset{(\alpha_o \ll 1)}{\approx} \mathrm{tr} \left(I_{n_r} \right). \tag{50}$$

In the high SNR cases ($\alpha_o \ll 1$), matrix $H(H^H H + \alpha_o I_{n_t})^{-1} H^H$ is very close to the identity matrix I_{n_r}. We return to term $\mathcal{A}_1$, and using (8) and (50) we get

$$\begin{aligned} \mathcal{A}_1 &= \frac{\sigma^2}{N_{tr}} \mathrm{tr} \left(\tilde{P}_o^* \tilde{P}_o^T \otimes H(H^H H + \alpha_o I_{n_t})^{-1} H^H \right) \\ &\overset{(8),(50)}{\approx} \frac{\sigma^2}{N_{tr}} \mathrm{tr} \left(\tilde{P}_o^* \tilde{P}_o^T \right) \mathrm{tr} \left(I_{n_r} \right) \\ &= \frac{n_r \sigma^2}{N_{tr}} \mathrm{tr} \left(\tilde{P}_o^* \tilde{P}_o^T \right). \end{aligned} \tag{51}$$

We continue with term $\mathcal{A}_2$, and using (47) we obtain

$$\mathcal{A}_2 = \frac{\sigma^2}{N_{tr}} \text{tr}\Big((H^H H + \alpha_o I_{n_t})^{-1} \otimes (H^* \tilde{P}_o^* - I_{n_r})(\tilde{P}_o^T H^T - I_{n_r})\Big). \tag{52}$$

At this point, we will examine term $\Big((H^* \tilde{P}_o^* - I_{n_r})(\tilde{P}_o^T H^T - I_{n_r})\Big)$. Using the expression of $\tilde{P}_o$ from (31) it can be shown that the eigenvalues of this term are

$$\lambda_i\left((H^* \tilde{P}_o^* - I_{n_r})(\tilde{P}_o^T H^T - I_{n_r})\right) = \frac{\alpha_o^2}{\left(\lambda_i\left(H^H H\right) + \alpha_o\right)^2}. \tag{53}$$

For the high SNR case we study ($\alpha_o \ll 1$)

$$\lambda_i\left((H^* \tilde{P}_o^* - I_{n_r})(\tilde{P}_o^T H^T - I_{n_r})\right) = \frac{\alpha_o^2}{\left(\lambda_i\left(H^H H\right) + \alpha_o\right)^2} \approx 0. \tag{54}$$

Thus, for the trace of this term we have that

$$\text{tr}\left((H^* \tilde{P}_o^* - I_{n_r})(\tilde{P}_o^T H^T - I_{n_r})\right) \overset{(54)}{\approx} 0. \tag{55}$$

Then, using (8), term $\mathcal{A}_2$ becomes

$$\mathcal{A}_2 = \frac{\sigma^2}{N_{tr}} \text{tr}\left(\left(H^H H + \alpha_o I_{n_t}\right)^{-1}\right) \text{tr}\left(\left(H^* \tilde{P}_o^* - I_{n_r}\right)\left(\tilde{P}_o^T H^T - I_{n_r}\right)\right) \overset{(55)}{\approx} 0. \tag{56}$$

Thus, we conclude that term $\mathcal{A}_2$ is negligible compared with $\mathcal{A}_1$ (this claim is confirmed in the simulations section). We return to the EMSE expression and using (51) we get

$$\text{EMSE}(\hat{\tilde{P}}) \approx \mathcal{A}_1 \approx \frac{n_r \sigma^2}{N_{tr}} \text{tr}\left(\tilde{P}_o^* \tilde{P}_o^T\right) = \frac{n_r \sigma^2}{N_{tr}} ||\tilde{P}_o||_F^2. \tag{57}$$

We can simplify (57) further, if we examine term $\text{tr}(\tilde{P}_o^* \tilde{P}_o^T)$ which is equal to $\text{tr}(\tilde{P}_o \tilde{P}_o^H)$. Using (33) and (50), we obtain that for the high SNR cases

$$\begin{aligned}
\text{MMSE} &= \text{tr}(I_{n_r}) - 2\text{tr}(\tilde{P}_o H) + \text{tr}(\tilde{P}_o \tilde{P}_o^H H^H H) + \alpha_o\, \text{tr}(\tilde{P}_o \tilde{P}_o^H) \\
&= \text{tr}(I_{n_r}) - 2\text{tr}\left(\left(H^H H + \alpha_o I_{n_t}\right)^{-1} H^H H\right) \\
&+ \text{tr}\left(\left(H^H H + \alpha_o I_{n_t}\right)^{-1} H^H H \left(H^H H + \alpha_o I_{n_t}\right)^{-1} H^H H\right) + \alpha_o\, \text{tr}(\tilde{P}_o \tilde{P}_o^H) \\
&\overset{(\alpha_o \ll 1)}{\approx} \text{tr}(I_{n_r}) - 2\text{tr}(I_{n_r}) + \text{tr}(I_{n_r}) + \alpha_o\, \text{tr}(\tilde{P}_o \tilde{P}_o^H) \\
&= \alpha_o\, \text{tr}\left(\tilde{P}_o \tilde{P}_o^H\right).
\end{aligned} \tag{58}$$

Thus, from (58) we get

$$\mathrm{tr}\left(\tilde{P}_o\tilde{P}_o^H\right) \approx \frac{1}{\alpha_o}\mathrm{MMSE}\ . \tag{59}$$

We return to the EMSE expression (57), which becomes

$$\begin{aligned}\mathrm{EMSE}(\hat{\tilde{P}}) &\approx \frac{n_r\sigma^2}{N_{tr}}\frac{1}{\alpha_o}\mathrm{MMSE}\\ &= \frac{E_{tr}}{N_{tr}}\mathrm{MMSE}.\end{aligned}$$

Thus, in the high SNR cases we have that

$$\boxed{\mathrm{EMSE}(\hat{\tilde{P}}) \approx \frac{E_{tr}}{N_{tr}}\mathrm{MMSE}.} \tag{60}$$

As we see from (60), we derived a simple and informative bound for the EMSE. In extensive simulation studies we have observed that this is a very good approximation to the true EMSE. We observe that the EMSE is proportional to the MMSE, which decreases for increasing SNR. The proportionality factor is the ratio of transmit power E_{tr} and the dimension of the training block N_{tr}.

Expression (60) can be used as a criterion for the choice of the length of the training block N_{tr}, and/or the total transmit power E_{tr}.

B. Noise estimation errors

In this section, we assume that the channel is perfectly known at the transmitter and we estimate the SOS of the received noise. The estimate of the noise covariance matrix is expressed as

$$\hat{R}_n \triangleq R_n + \Delta R_n. \tag{61}$$

The scaled pre-coding matrix becomes

$$\hat{\tilde{P}} = \left(H^H H + \frac{\text{tr}(R_n + \Delta R_n)}{E_{tr}} I_{n_t}\right)^{-1} H^H. \tag{62}$$

Next, we apply (2) to (62) in order to compute $\Delta\tilde{P}$

$$\begin{aligned}\hat{\tilde{P}} &= \Big(\underbrace{H^H H + \alpha_o I_{n_t}}_{A} + \underbrace{\frac{\text{tr}(\Delta R_n)}{E_{tr}} I_{n_t}}_{\Delta A}\Big)^{-1} H^H \\ &= \tilde{P}_o - \frac{\text{tr}(\Delta R_n)}{E_{tr}}\left(H^H H + \alpha_o I_{n_t}\right)^{-1} \tilde{P}_o.\end{aligned} \tag{63}$$

Thus, a first-order approximation to $\Delta\tilde{P}$ is

$$\Delta\tilde{P} = -\frac{\text{tr}(\Delta R_n)}{E_{tr}}\left(H^H H + \alpha_o I_{n_t}\right)^{-1} \tilde{P}_o. \tag{64}$$

Having expressed $\Delta\tilde{P}$ as a function of the noise covariance matrix estimation error ΔR_n, we return to (37) and we write for the EMSE using (64)

$$\begin{aligned}\text{EMSE}(\hat{\tilde{P}}) &= \mathbf{E}\left[\text{tr}\Big(\Delta\tilde{P}^H (H^H H + \alpha_o I_{n_t})\Delta\tilde{P}\Big)\right] \\ &\overset{(64)}{=} \frac{\mathbf{E}\left[\text{tr}^2(\Delta R_n)\right]}{E_{tr}^2}\,\text{tr}\Big(\tilde{P}_o\tilde{P}_o^H (H^H H + \alpha_o I_{n_t})^{-1}\Big)\end{aligned}$$

$$\boxed{\text{EMSE}(\hat{\tilde{P}}) = \frac{\mathbf{E}\left[\text{tr}^2(\Delta R_n)\right]}{E_{tr}^2}\,\text{tr}\Big(\tilde{P}_o\tilde{P}_o^H (H^H H + \alpha_o I_{n_t})^{-1}\Big).} \tag{65}$$

We have expressed the EMSE as a function of $\mathbf{E}\left[\text{tr}^2(\Delta R_n)\right]$. In the sequel, we consider the high SNR cases and derive a simple and informative expression for the EMSE.

1) Simplifications in the high SNR cases: In the high SNR cases, we can simplify expression (65) and give a simple approximation to the EMSE.

We start by making specific assumptions for the noise SOS estimate. Considering that we have estimated the channel, we can estimate the noise covariance matrix, using the channel estimate. More specifically, we prove in the Appendix that for the spatially and temporally white Gaussian noise case we consider, the unbiased estimator of the noise variance given by [5, p. 174], has variance

$$\mathbf{E}\left[\left(\sigma^2 - \hat{\hat{\sigma}}^2\right)^2\right] = \frac{\sigma^4}{n_r(N_{tr} - n_t)}. \tag{66}$$

Using the assumptions we made for the noise, we return to (65) and using (66) we obtain

$$\begin{aligned} \text{EMSE}(\hat{\hat{P}}) &= \frac{\mathbf{E}\left[\text{tr}^2(\Delta R_n)\right]}{E_{tr}^2} \text{ tr}\left(\tilde{P}_o \tilde{P}_o^H (H^H H + \alpha_o I_{n_t})^{-1}\right) \\ &= \frac{n_r \mathbf{E}\left[\left(\sigma^2 - \hat{\hat{\sigma}}^2\right)^2\right]}{E_{tr}^2} \text{ tr}\left(\tilde{P}_o \tilde{P}_o^H (H^H H + \alpha_o I_{n_t})^{-1}\right) \\ &= \frac{\sigma^4}{E_{tr}^2 \left(N_{tr} - n_t\right)} \text{ tr}\left(\tilde{P}_o \tilde{P}_o^H (H^H H + \alpha_o I_{n_t})^{-1}\right). \end{aligned} \tag{67}$$

Next, we examine term $\text{tr}\left(\tilde{P}_o \tilde{P}_o^H (H^H H + \alpha_o I_{n_t})^{-1}\right)$. Using (31), we obtain

$$\text{tr}\left(\tilde{P}_o \tilde{P}_o^H (H^H H + \alpha_o I_{n_t})^{-1}\right) = \text{tr}\left(H \left(H^H H + \alpha_o I_{n_t}\right)^{-3} H^H\right). \tag{68}$$

It can be shown that the eigenvalues of the matrix inside the trace are ([4, p. 138])

$$\lambda_i\left(H \left(H^H H + \alpha_o I_{n_t}\right)^{-3} H^H\right) = \frac{\lambda_i\left(H^H H\right)}{\left(\lambda_i\left(H^H H\right) + \alpha_o\right)^3}$$

and for the high SNR case ($\alpha_o \ll 1$)

$$\lambda_i\left(H \left(H^H H + \alpha_o I_{n_t}\right)^{-3} H^H\right) \approx \frac{1}{\lambda_i^2\left(H^H H\right)} \overset{(*)}{=} \frac{1}{\lambda_i^2\left(H H^H\right)}$$

where at point (*) we used that if $\lambda_1, \ldots, \lambda_{n_r}$ are the eigenvalues of HH^H, then $\lambda_1, \ldots, \lambda_{n_r}, 0, \ldots, 0$

are the eigenvalues of $H^H H$. Thus, for the trace of (68) we get

$$\begin{aligned}
\text{tr}\left(\tilde{P}_o \tilde{P}_o^H (H^H H + \alpha_o I_{n_t})^{-1}\right) &= \text{tr}\left(H\left(H^H H + \alpha_o I_{n_t}\right)^{-3} H^H\right) \\
&\approx \sum_{i=1}^{n_r} \frac{1}{\lambda_i^2\left(HH^H\right)} \\
&= \sum_{i=1}^{n_r} \lambda_i\left(\left(HH^H\right)^{-2}\right) \\
&= \text{tr}\left(\left(HH^H\right)^{-2}\right) \\
&= ||\left(HH^H\right)^{-1}||_F^2.
\end{aligned} \tag{69}$$

Thus, combining expressions (67) and (69), we obtain

$$\boxed{\text{EMSE}(\hat{\tilde{P}}) \approx \frac{\sigma^4}{E_{tr}^2\left(N_{tr} - n_t\right)} ||\left(HH^H\right)^{-1}||_F^2.} \tag{70}$$

This approximation states that the EMSE is proportional to the squared noise variance, σ^4, which decreases “fast” enough for increasing SNR. The proportionality factor is determined by the transmit power, E_{tr}, the length of the training block used for channel estimation, N_{tr}, and the number of the transmit antennas, n_t. The Frobenius norm $||\left(HH^H\right)^{-1}||_F^2$ is also a constant which depends on the specific realization of the channel matrix. In the simulations section we will see that this bound is a good approximation to the EMSE, especially at high SNR.

Table I

Elements of channel matrix H

-0.0648+0.0388*j	-0.0547+0.2974*j	0.2588-0.0954*j
0.4186+0.2072*j	-0.5157-0.3955*j	0.3897-0.1856*j

IV. Simulations

In this section, we present simulations which support our theoretical results. We consider a broadcast system with $n_t = 3$ transmit antennas and $n_r = 2$ non-cooperative receivers.

The filtering matrix H is a realization of a 2×3 random matrix, with elements i.i.d. complex, circular, zero-mean Gaussian random variables, normalized so that $\|H\|_F^2 = 1$. Its elements are given in Table I.

The noise is spatially and temporally white, circularly symmetric complex Gaussian. We assume that the noise variance is σ^2, the same for all receivers.

We set the transmit power $\mathrm{E}_{tr} = n_t$. We assume that the training block is composed of $N_{tr} = 20$ columns.

Simulation 1. *Channel estimation errors.*

In Fig. 2, we plot the theoretical second-order approximation (44), the corresponding experimentally computed EMSE and the bound (60). We observe that the experimental and theoretical EMSE values practically coincide for SNR higher than 5 dB, while expression (60) is a good approximation to the EMSE, especially at high SNR. Analogous results have been observed in extensive simulations.

In Fig. 3, we plot terms $\mathcal{A}_1$ and $\mathcal{A}_2$ of the theoretical EMSE of (44). We observe that, for SNR higher than 15 dB, the contribution of term $\mathcal{A}_2$ to the EMSE is negligible compared to the contribution of term $\mathcal{A}_1$. Thus, our claim that the EMSE is approximately equal to term $\mathcal{A}_1$ for

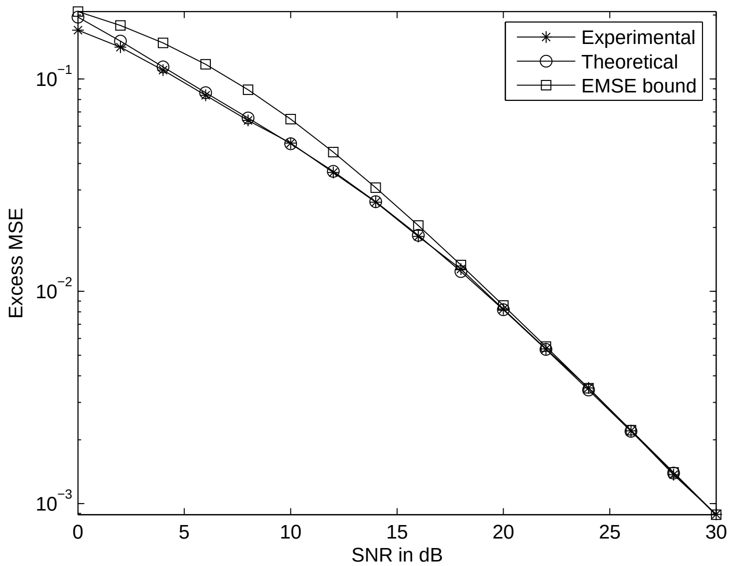

Fig. 2. Theoretical EMSE (second-order approximation), experimentally computed EMSE and bound (60) for the case of channel estimation errors.

the high SNR cases (57), is confirmed.

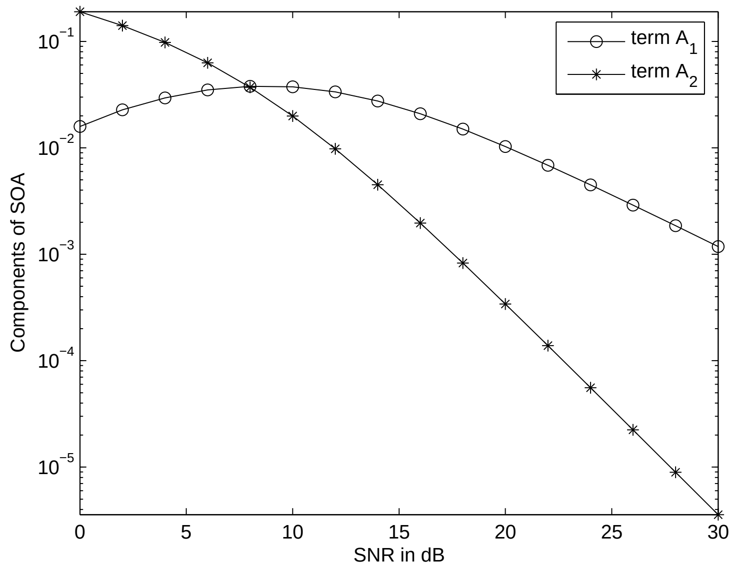

Fig. 3. Terms $\mathcal{A}_1$ and $\mathcal{A}_2$ of the second-order approximation EMSE (44) for the case of channel estimation errors.

Simulation 2. *Noise estimation errors.*

In Fig. 4, we present the theoretical second-order approximation (67), the corresponding experimentally computed EMSE and the bound (70). We observe that the first two quantities coincide. Regarding the bound, we observe that it is a very good approximation to the EMSE for SNR higher than 15 dB.

Comparing the EMSEs for the two cases we study (see Fig. 2 and Fig. 4), we observe that the error induced by the channel estimate is much more significant than that induced by the noise SOS estimate.

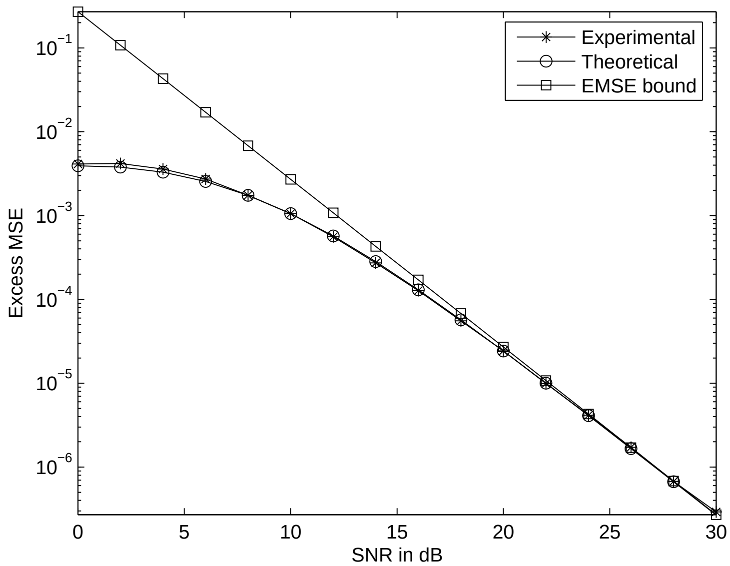

Fig. 4. Theoretical EMSE (second-order approximation), experimentally computed EMSE and bound (70) for the case of noise SOS estimation errors.

V. Conclusion

In this work, we considered the behavior of the *transmit Wiener filter* (TxWF), under channel and noise SOS uncertainties. Using matrix perturbation theory, we developed second-order approximations to the EMSE in terms of channel and noise SOS estimation errors. We derived simple EMSE bounds in the high SNR cases. In particular, for the case of channel estimation errors we concluded that the EMSE is proportional to the MMSE, with the proportionality factor determined by the transmit power E_{tr} and the length of the training block N_{tr}. For the case of noise SOS estimation errors, we showed that the EMSE is proportional to the squared noise variance, σ^4. A comparison of the EMSEs for the two cases we study, shows that the error induced by the channel estimate is much more significant than that induced by the noise SOS estimate.

APPENDIX

CHANNEL AND NOISE VARIANCE ML ESTIMATES

The ML estimate of the channel gain matrix H and the noise variance estimate can be derived from training-based estimation [5]. Using the training block X_t of dimension $n_t \times N_{tr}$, the received block of dimension $n_r \times N_{tr}$ is given by [5]

$$Y_t = HX_t + E_t$$

where E_t is the corresponding $n_r \times N_{tr}$ noise matrix. The additive noise is assumed to be spatially and temporally white Gaussian.

A. ML estimate of the channel matrix

The ML estimate of the channel H based on the received training block Y_t is given by [5, p. 174]

$$\hat{H} = Y_t X_t^H \left(X_t X_t^H \right)^{-1}.$$

This estimate is unbiased and the covariance matrix of $\mathrm{vec}(\Delta H)$ is given by [5, p. 175]

$$\Sigma \triangleq \mathbf{E}\left[\mathrm{vec}(\Delta H)\mathrm{vec}^H(\Delta H)\right] = \sigma^2 \left(\left(X_t X_t^H \right)^{-T} \otimes I_{n_r} \right).$$

As shown in [5, p. 176], the optimal training block X_t should satisfy

$$X_t X_t^H \propto I_{n_t}.$$

B. ML noise variance estimate

Having estimated the channel matrix H, the ML noise variance estimate is [5, p. 174]

$$\hat{\sigma}^2 = \frac{1}{N_{tr} n_r} \mathrm{tr}\left(Y_t \mathbf{P}_{X_t^H}^{\perp} Y_t^H \right)$$

where $\mathbf{P}^{\perp}_{X_t^H}$ is the orthogonal projector onto the orthogonal complement of the column space of X_t^H. It can be shown that this estimate is biased. More specifically,

$$\begin{aligned}\mathrm{tr}\left(Y_t\mathbf{P}^{\perp}_{X_t^H}Y_t^H\right) &= \mathrm{tr}\left(\left(HX_t+E_t\right)\mathbf{P}^{\perp}_{X_t^H}\left(X_t^H H^H + E_t^H\right)\right)\\ &= \mathrm{tr}\left(E_t^H E_t \mathbf{P}^{\perp}_{X_t^H}\right)\end{aligned}$$

giving that

$$\mathbf{E}\,\mathrm{tr}\left(Y_t\mathbf{P}^{\perp}_{X_t^H}Y_t^H\right) = \left(N_{tr}-n_t\right)n_r\sigma^2.$$

Thus, an unbiased estimate of σ^2 is given by

$$\hat{\hat{\sigma}}^2 = \frac{1}{n_r(N_{tr}-n_t)}\,\mathrm{tr}\left(Y_t\mathbf{P}^{\perp}_{X_t^H}Y_t^H\right) = \frac{1}{c}\,\mathrm{tr}\left(Y_t\mathbf{P}^{\perp}_{X_t^H}Y_t^H\right).$$

where

$$c \triangleq n_r\left(N_{tr}-n_t\right).$$

We continue with the computation of the variance of the unbiased noise variance estimator

$$\begin{aligned}\mathrm{var}\hat{\hat{\sigma}}^2 = \mathbf{E}\left[\left|\sigma^2-\hat{\hat{\sigma}}^2\right|^2\right] &= \mathbf{E}\left[\left|\sigma^2-\frac{1}{c}\mathrm{tr}\left(\mathbf{P}^{\perp}_{X_t^H}E_t^H E_t\right)\right|^2\right]\\ &= \sigma^4-\frac{2}{c}\,\sigma^2\mathrm{Re}\left(\underbrace{\mathbf{E}\left[\left|\mathrm{tr}\left(\mathbf{P}^{\perp}_{X_t^H}E_t^H E_t\right)\right|\right]}_{=c\,\sigma^2}\right)+\frac{1}{c^2}\,\mathbf{E}\left[\left|\mathrm{tr}\left(\mathbf{P}^{\perp}_{X_t^H}E_t^H E_t\right)\right|^2\right]\\ &= -\sigma^4+\frac{1}{c^2}\underbrace{\mathbf{E}\left[\left|\mathrm{tr}\left(\mathbf{P}^{\perp}_{X_t^H}E_t^H E_t\right)\right|^2\right]}_{\mathcal{B}}.\end{aligned}$$

In order to compute term $\mathcal{B}$, we examine $\mathrm{tr}\left(\mathbf{P}^{\perp}_{X_t^H}E_t^H E_t\right)$

$$\begin{aligned}\mathrm{tr}\left(\mathbf{P}^{\perp}_{X_t^H}E_t^H E_t\right) &= \mathrm{tr}\left(\mathbf{P}^{\perp}_{X_t^H}E_t^H I_{n_r} E_t\right)\\ &\overset{(3)}{=} \underbrace{\mathrm{vec}^T\left(E_t^T\right)}_{\triangleq\mathbf{e}^T}\left(I_{n_r}\otimes\mathbf{P}^{\perp}_{X_t^H}\right)\underbrace{\mathrm{vec}(E_t^H)}_{=\mathbf{e}^*}\\ &= \mathbf{e}^T\left(I_{n_r}\otimes\mathbf{P}^{\perp}_{X_t^H}\right)\mathbf{e}^*.\end{aligned}$$

Thus,

$$\begin{aligned}
\mathcal{B} &= \mathbf{E}\left[\left|\operatorname{tr}\left(\mathbf{P}^{\perp}_{X_t^H} E_t^H E_t\right)\right|^2\right] = \mathbf{E}\left[\operatorname{tr}\left(\mathbf{P}^{\perp}_{X_t^H} E_t^H E_t\right)\operatorname{tr}^H\left(\mathbf{P}^{\perp}_{X_t^H} E_t^H E_t\right)\right] \\
&= \mathbf{E}\left[\mathbf{e}^T\left(I_{n_r}\otimes\mathbf{P}^{\perp}_{X_t^H}\right)\mathbf{e}^*\,\mathbf{e}^T\left(I_{n_r}\otimes\mathbf{P}^{\perp}_{X_t^H}\right)\mathbf{e}^*\right] \\
&= \mathbf{E}\left[\operatorname{tr}\left(\left(I_{n_r}\otimes\mathbf{P}^{\perp}_{X_t^H}\right)^T\mathbf{e}^*\mathbf{e}^T\left(I_{n_r}\otimes\mathbf{P}^{\perp}_{X_t^H}\right)\mathbf{e}^*\mathbf{e}^T\right)\right] \\
&\overset{(3)}{=} \mathbf{E}\left[\operatorname{vec}^T\left(\mathbf{e}\mathbf{e}^H\right)\left(I_{n_r}\otimes\mathbf{P}^{\perp}_{X_t^H}\right)^T\otimes\left(I_{n_r}\otimes\mathbf{P}^{\perp}_{X_t^H}\right)\operatorname{vec}\left(\mathbf{e}^*\mathbf{e}^T\right)\right] \\
&= \mathbf{E}\left[\operatorname{tr}\left(\left(I_{n_r}\otimes\mathbf{P}^{\perp}_{X_t^H}\right)^T\otimes\left(I_{n_r}\otimes\mathbf{P}^{\perp}_{X_t^H}\right)\operatorname{vec}\left(\mathbf{e}^*\mathbf{e}^T\right)\operatorname{vec}^T\left(\mathbf{e}\mathbf{e}^H\right)\right)\right] \\
&= \operatorname{tr}\left(\left(I_{n_r}\otimes\mathbf{P}^{\perp}_{X_t^H}\right)^T\otimes\left(I_{n_r}\otimes\mathbf{P}^{\perp}_{X_t^H}\right)\mathbf{E}\left[\operatorname{vec}\left(\mathbf{e}^*\mathbf{e}^T\right)\operatorname{vec}^T\left(\mathbf{e}\mathbf{e}^H\right)\right]\right).
\end{aligned}$$

Terms $\operatorname{vec}\left(\mathbf{e}^*\mathbf{e}^H\right)$ and $\operatorname{vec}^T\left(\mathbf{e}\mathbf{e}^T\right)$ can be computed analytically, by writing down the exact form of each vector. Thus, having computed these terms, we take expectation, using [8, p. 508]

$$\mathbf{E}\left(x_i^* x_j x_k^* x_l\right) = \mathbf{E}\left(x_i^* x_j\right)\mathbf{E}\left(x_k^* x_l\right) + \mathbf{E}\left(x_i^* x_l\right)\mathbf{E}\left(x_j x_k^*\right).$$

Applying this property to the last term of $\mathcal{B}$, we obtain

$$\mathbf{E}\left[\operatorname{vec}\left(\mathbf{e}^*\mathbf{e}^H\right)\operatorname{vec}^T\left(\mathbf{e}\mathbf{e}^T\right)\right] = \sigma^4 I_{n_r N_{tr}} + \sigma^4\operatorname{vec}\left(I_{n_r N_{tr}}\right)\operatorname{vec}^H\left(I_{n_r N_{tr}}\right).$$

Thus, term $\mathcal{B}$ becomes

$$\begin{aligned}
\mathcal{B} &= \operatorname{tr}\left(\left(I_{n_r}\otimes\mathbf{P}^{\perp}_{X_t^H}\right)^T\otimes\left(I_{n_r}\otimes\mathbf{P}^{\perp}_{X_t^H}\right)\left(\sigma^4 I_{n_r N_{tr}} + \sigma^4\operatorname{vec}\left(I_{n_r N_{tr}}\right)\operatorname{vec}^H\left(I_{n_r N_{tr}}\right)\right)\right) \\
&= \sigma^4\operatorname{tr}\left(\left(I_{n_r}\otimes\mathbf{P}^{\perp}_{X_t^H}\right)^T\otimes\left(I_{n_r}\otimes\mathbf{P}^{\perp}_{X_t^H}\right)\right) \\
&+ \sigma^4\operatorname{tr}\left(\operatorname{vec}^H\left(I_{n_r N_{tr}}\right)\left(\left(I_{n_r}\otimes\mathbf{P}^{\perp}_{X_t^H}\right)^T\otimes\left(I_{n_r}\otimes\mathbf{P}^{\perp}_{X_t^H}\right)\right)\operatorname{vec}\left(I_{n_r N_{tr}}\right)\right) \\
&\overset{(3),(8)}{=} \sigma^4 n_r^2\left(N_{tr}-n_t\right)^2 + \sigma^4\operatorname{tr}\left(\left(I_{n_r}\otimes\mathbf{P}^{\perp}_{X_t^H}\right)\left(I_{n_r}\otimes\mathbf{P}^{\perp}_{X_t^H}\right)\right) \\
&= \sigma^4 n_r^2\left(N_{tr}-n_t\right)^2 + \sigma^4\operatorname{tr}\left(I_{n_r}\otimes\mathbf{P}^{\perp}_{X_t^H}\right) \\
&= c^2\sigma^4 + c\,\sigma^4.
\end{aligned}$$

Now, we return to $\operatorname{var}\hat{\hat{\sigma}}^2$, which becomes

$$\operatorname{var}\hat{\hat{\sigma}}^2 = -\sigma^4 + \frac{1}{c^2}\left(c^2\sigma^4 + c\,\sigma^4\right) = \frac{1}{c}\,\sigma^4.$$

Thus, for the unbiased case of the noise variance estimate, the variance is given by

$$\mathrm{var}\hat{\hat{\sigma}}^2 = \mathbf{E}\left[\left(\sigma^2 - \hat{\hat{\sigma}}^2\right)^2\right] = \frac{\sigma^4}{n_r(N_{tr} - n_t)}.$$

REFERENCES

[1] M. Joham, K. Kusume, M. H. Gzara, W. Utschick, and J. A. Nossek, "Transmit Wiener Filter," Tech. Rep. TUM-LNS-TR-02-1, Munich University of Technology, January 2002.

[2] M. Joham, W. Utschick, and J. A. Nossek, "Linear Transmit Processing in MIMO Communications Systems," *IEEE Trans. Signal Processing*, Vol. 53, no. 11, pp. 2700-2712, August 2005.

[3] H. Luetkepohl. *Handbook of matrices*. John Wiley, 1996.

[4] G. Stewart and J.-G, Sun. *Matrix perturbation theory.* Academic Press, 1990.

[5] E. G. Larsson and P. Stoica. *Space-Time Block Coding for Wireless Communications.* Cambridge University Press, 2003.

[6] http://www.unik.no/personer/arehj/publications/matrixdiff.pdf

[7] A. Sayed. *Fundamentals of adaptive filtering.* IEEE-Wiley, 2003.

[8] Steven M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory.* Prentice-Hall, 1993.