Performance degradation of telecommunication transceivers due to parameter estimation errors

Despoina Tsipouridou

Advisor: Professor Athanasios P. Liavas

PhD Thesis

TECHNICAL UNIVERSITY OF CRETE
DEPARTMENT OF ELECTRONIC & COMPUTER ENGINEERING

September 2010

To my parents

Acknowledgments

I am deeply indebted to my advisor Prof. Athanasios Liavas for his unfailing support all these years, for his encouragement to proceed and for being always available for uncountable discussions. His patience and support helped me finish this thesis. Thank you very much!

I am grateful to Prof. George Karystinos for his support, the high-quality lectures and the valuable comments and suggestions on this manuscript. I am also grateful for having the opportunity to enjoy the inspirational lectures of Prof. Nikos Sidiropoulos.

I would like to thank my lab-friends, Giannis, Tasos, Vaios, Lefteris, Alexandros, Dimitris and Stamatis, for creating a pleasant and friendly working environment and for helping me in many ways to complete this thesis. I would also like to thank all my friends, and, specially, Georgia and Mary.

Last, but not least, I would like to thank my family and Kleon, for being a constant source of support and strength all these years.

Abstract

In this thesis we consider the performance degradation of three telecommunication transceivers with respect to parameters uncertainties. The major contribution of our analysis is the derivation of simple and informative approximations to the excess mean square error, which uncover the basic factors that determine the performance of each transceiver under the presence of parameter uncertainties.

In Chapter 1, we briefly present the channel models and the transceiver structures used in this thesis.

In Chapters 2 and 3, we consider the case of the Multiple-Input Multiple-Output (MIMO) flat fading channel and two widely known transceiver schemes, i.e., the transmit MIMO Wiener filtering and the Tomlinson-Harashima (TH) precoding. For the case of the transmit MIMO Wiener filtering, degradation is due to channel estimation errors and time-variations, and noise second-order statistics estimation errors. For the case of the TH precoder, degradation is due to channel estimation errors and time-variations. For both cases, our final expressions uncover the factors that determine the performance degradation in practice, and the relative importance between different error sources.

In Chapter 4, we consider the case of the Single-Input Single-Output (SISO) frequency selective channel and the minimum mean square error (MMSE) linear equalizer. The error sources that cause the system degradation is the channel and CFO estimation errors. Our aim is to uncover the relative importance of these error sources. It turns out that the CFO estimation error is much more important than the channel estimation error. This fact leads to useful conclusions concerning the optimal training sequence design for *joint* CFO and channel estimation.

Contents

List of Figures

Notation

Boldface uppercase and boldface lowercase letters denote matrices and column vectors, respectively. Superscripts ^{T}, ^{H} and * denote transpose, conjugate transpose and elementwise conjugation, respectively. $\mathrm{tr}(\cdot)$, $\mathrm{vec}(\cdot)$ and $\mathrm{vech}(\cdot)$ denote the trace, the vectorization and the half-vectorization operator, respectively. $\otimes$ denotes the Kronecker product and $\mathrm{Re}\{\cdot\}$ extracts the real part of a complex number. The eigenvalues of matrix $\mathbf{A}$ are denoted as $\lambda_i(\mathbf{A})$. $\sigma_{\max}(\cdot)$, $\sigma_{\min}(\cdot)$, $\|\cdot\|_2$, $\|\cdot\|_F$, and $k_2(\cdot)$ denote, respectively, the maximum singular value, the minimum singular value, the spectral norm, the Frobenius norm, and the condition number, with respect to the spectral norm, of the matrix argument.

If $\mathbf{A}$ is an $n \times n$ positive semidefinite matrix, then its eigenvalues are ordered such that $\lambda_1(\mathbf{A}) \geq \lambda_2(\mathbf{A}) \geq \cdots \geq \lambda_n(\mathbf{A})$. $\mathbf{A} \geq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is positive semidefinite. a_{ij} denotes the (i,j)-th element of matrix $\mathbf{A}$, and $\mathbf{I}_M$ and $\mathbf{O}_M$ denote the $M \times M$ identity and zero matrix, respectively.

Matrix results

In this section, we present useful matrix results that are used throughout this thesis. For matrices with compatible dimensions [20, pp. 17-19]

$$\mathrm{tr}(\mathbf{ABCD}) = \mathrm{vec}^T(\mathbf{D}^T)(\mathbf{C}^T \otimes \mathbf{A})\mathrm{vec}(\mathbf{B}) \tag{1}$$

$$\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\,\mathrm{vec}(\mathbf{B}). \tag{2}$$

$$\mathbf{AB} \otimes \mathbf{CD} = (\mathbf{A} \otimes \mathbf{C})\,(\mathbf{B} \otimes \mathbf{D}) \tag{3}$$

$$\mathrm{tr}(\mathbf{A} \otimes \mathbf{B}) = \mathrm{tr}(\mathbf{A})\,\mathrm{tr}(\mathbf{B}) \tag{4}$$

and [20, p. 117]

$$\mathbf{K}(\mathbf{A} \otimes \mathbf{B})\mathbf{K}^{\mathbf{H}} = \mathbf{B} \otimes \mathbf{A}. \tag{5}$$

where $\mathbf{K}$ denotes the commutation matrix. If $\mathbf{A}$ and $\mathbf{B}$ are positive semidefinite, then [20, p. 44]

$$\mathrm{tr}(\mathbf{AB}) \leq \mathrm{tr}(\mathbf{A})\,\mathrm{tr}(\mathbf{B}). \tag{6}$$

If $\mathbf{B}$ is positive semidefinite with maximum eigenvalue $\lambda_{\max}(\mathbf{B})$, then [20, p. 44]

$$\mathrm{tr}\left(\mathbf{ABA}^H\right) \leq \lambda_{\max}(\mathbf{B})\,\mathrm{tr}\left(\mathbf{AA}^H\right) \tag{7}$$

For any matrix $\mathbf{A}$ [20, p. 97]

$$\mathrm{vec}(\mathbf{A}) = \mathbf{K}\mathrm{vec}(\mathbf{A}^T). \tag{8}$$

If $\mathbf{A}$ is lower triangular, then [20, p. 99]

$$\mathrm{vec}(\mathbf{A}) = \mathbf{L}^T\mathrm{vech}(\mathbf{A}) \tag{9}$$

and

$$\mathrm{vec}(\mathrm{diag}(\mathrm{diag}(\mathbf{A}))) = \mathbf{L}^T\mathbf{LKL}^T\mathbf{L}\mathrm{vec}(\mathbf{A}) \tag{10}$$

where $\mathbf{L}$ is the elimination matrix [20, ch. 9] and $\mathrm{diag}(\mathrm{diag}(\mathbf{A}))$ is the diagonal matrix whose elements are the diagonal elements of $\mathbf{A}$. Finally, we remind that [27, p. 130]

$$(\mathbf{A} + \Delta\mathbf{A})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\Delta\mathbf{A}\mathbf{A}^{-1} + O(\|\Delta\mathbf{A}\|^2). \tag{11}$$

Chapter 1

Telecommunication Transceivers

In this chapter, we introduce the Single-Input Single-Output (SISO) and the Multiple-Input Multiple-Output (MIMO) channel models and we briefly comment the impairments they induce to a telecommunication system. In the sequel, we present commonly used transceiver structures that combat the unwanted effects caused by the wireless channels.

1.1 Channel models

1.1.1 SISO channels

The transmission over a frequency-selective SISO communication channel is described by the baseband-equivalent signal model

$$y_n = \sum_{l=0}^{L} h_l s_{n-l} + w_n \tag{1.1}$$

where s_n and w_n denote the channel input and additive channel noise, respectively, and the channel impulse response is denoted as $\mathbf{h} \triangleq [h_0 \cdots h_L]^T$.

It is obvious from (1.1) that the received signal suffers from inter-symbol interference (ISI) that has to be eliminated.

1.1.2 MIMO channels

The transmission over a flat fading MIMO communication channel with n_t transmit and n_r receive antennas is described by the baseband-equivalent signal model

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{w} \tag{1.2}$$

where $\mathbf{s}$ is the $n_t \times 1$ channel input vector, $\mathbf{H}$ is the $n_r \times n_t$ channel matrix, and $\mathbf{w}$ is the $n_r \times 1$ additive channel noise.

From (1.2), we observe that the signal received at each antenna suffers from inter-stream interference.

In order to eliminate ISI and inter-stream interference in SISO and MIMO systems the signal has to be processed at the receiver and/or the transmitter. This processing, known as equalization, is performed using properly designed filters.

1.2 Transceivers

The design of equalizing filters can be stated as a simple optimization problem, that of finding the filter parameters that minimize an appropriate cost function.

The most efficient strategy is to consider the joint optimization of the transmit and receive filters, which implies high complexity at both sides of the communication link.

To reduce the complexity at one side of the link, we use the suboptimal solution of the transmit or receive processing, i.e., transmit processing simplifies receivers, while receive processing simplifies transmitters. In any case, the transceiver design depends on the performance requirements and the affordable complexity of the communication scenario under consideration.

Many transmitter and receiver structures mitigating ISI and inter-stream interference have been proposed in the literature, achieving various levels of performance with varying complexity. Linear equalization is a simple technique of low complexity that works well in many cases. In cases where linear processing does not meet the required system performance, non-linear structures (usually, of higher complexity) are used.

In many cases, the computation of the equalizing filters assumes perfect knowledge of system parameters (e.g., the channel state information (CSI), the noise and input second-order statistics, the carrier frequency offset (CFO)) at the transmitter and/or the receiver. But, in real-world systems this assumption is *not* realistic. One way to proceed is to estimate the unknown parameters and use the estimates as if they were the true quantities; this is sometimes called the *mismatched* approach. Another way is to exploit the statistical description of the parameter uncertainties and develop *robust* designs. In both cases, the design of the equalizing filters is based on *inexact* estimates and thus performance degradation is inevitable.

In this thesis, we study the performance degradation of three widely known transceiver structures in terms of the excess mean-square error (EMSE) induced by the parameter uncertainties. More specifically, we estimate the unknown parameters of interest and use the *mismatched* approach to compute the transceiver filters. Perhaps, a more interesting system characterization would result if we used the bit-error rate (BER) as a metric, but, unfortunately, the MSE and BER are not connected by a simple relationship for the transceivers under consideration.

In Chapter 2, we consider the case of transmit processing (precoding or pre-equalization) for the case of a flat fading MIMO channel, and more specifically, we study the performance degradation of the transmit MIMO Wiener filter due to channel and noise second-order statistics uncertainties. We develop second-order EMSE approximations and assuming optimal training we derive simple EMSE approximations in the high SNR cases. Considering the channel estimation errors, we conclude that the EMSE is proportional to the minimum MSE (MMSE). Considering the channel time-variations, we find that the EMSE increases with increasing SNR and for high SNR it reaches an asymptotic value. For the case of noise SOS estimation errors, we show that the EMSE is proportional to the squared noise variance. A comparison of the EMSEs for the cases of estimation errors only shows that the channel estimation errors are much more significant than the noise SOS estimation errors.

In Chapter 3, we consider the non-linear Tomlinson-Harashima (TH) precoder for the flat fading MIMO channel. The specific TH scheme is implemented by a non-linear structure at the transmitter followed by a linear filtering and a modulo operator at the receiver. Again, we compute the associated EMSE induced by the channel estimation errors and the channel time-variations. Considering a packet-based communication scenario where the channel may change (slowly) between successive packets, we conclude that the processing of each packet suffers from errors at both the transmitter and the receiver. We show that the EMSE consists of two components that can be studied separately. The first component is due to the mismatch between the previous channel estimate and the current channel, while the second is due to the mismatch between the current channel and its estimate. We develop a second-order approximation to the EMSE and, then, using optimal training, we focus on the high-SNR regime and derive a simple, informative, and tight (for sufficiently high SNR) EMSE upper bound, which uncovers the basic factors that determine the MIMO-TH performance degradation.

Finally, in Chapter 4, we apply the MMSE linear equalizer to a SISO frequency-selective channel, we quantify the system performance degradation due to channel and CFO estimation errors and we derive useful conclusions for the training sequence design for *joint* channel and CFO estimation. More specifically, the performance degradation is due to the fact that a mismatched MMSE linear equalizer is applied to channel output samples with imperfectly canceled CFO. We uncover that, in many cases of high practical importance, the imperfectly canceled CFO is the main cause of the performance degradation. In these cases, the EMSE is approximately proportional to the CFO estimation error variance, with the proportionality coefficient being independent of the TS, implying that optimal TS design for CFO estimation is also highly relevant for *joint* CFO and channel estimation.

Chapter 2

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

We consider the sensitivity of the transmit MIMO Wiener filter with respect to channel and noise second-order statistics (SOS) uncertainties. Using results from matrix perturbation theory, we derive second-order approximations to the excess mean-square error (EMSE) induced by using the channel or noise SOS estimates as if they were the true quantities. Assuming optimal training and sufficiently high SNR, we develop simple and informative approximations to the EMSE, which indicate that the channel estimation errors are much more significant than the noise SOS estimation errors. Uncertainties due to channel time-variations induce EMSE that increases with increasing SNR and asymptotically tends to a constant value.

2.1 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, [12] and the references therein). In order to keep the mobile units as simple as possible, 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 pre-equalization (or precoding) structures that combat frequency selectivity and/or inter-stream interference and keep the receivers simple, because the only processing required at the receiver is a scalar scaling [12], [13].

The TxWF, which outperforms the two other structures in terms of mean-square error (MSE) and bit-error rate (BER) [12], can be computed if the channel and the input and noise second-order statistics (SOS) are perfectly known at the transmitter. This may happen, for example, in time division duplex (TDD) systems or systems with a feedback information channel. 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 then use the estimates as if they were the true quantities. Estimation errors and/or time-variations introduce uncertainties in the estimated quantities and induce excess MSE (EMSE) leading to TxWF performance degradation.

In this chapter, we consider the sensitivity of the TxWF with respect to channel and noise SOS uncertainties and we develop second-order approximations to the associated EMSEs. While the general expressions are complicated and difficult to interpret, we are able to derive simple and informative EMSE approximations for the high SNR cases. It turns out that the EMSE due to channel estimation errors is proportional to the minimum MSE (MMSE), while the EMSE due to noise SOS estimation errors is proportional to the squared noise variance. On the other hand, the EMSE due to channel time-variations increases for increasing SNR and asymptotically reaches a constant value.

2.2 The Transmit Wiener Filter

2.2.1 The system model

We consider the pre-equalized, baseband-equivalent, discrete-time frequency-flat MIMO system, with n_t transmit and n_r receive antennas (with $n_r \leq n_t$), depicted in Fig. 2.1. This system is described by the expression

$$\hat{\mathbf{s}} = \mathbf{HPs} + \mathbf{w} \tag{2.1}$$

where $\mathbf{s}$ is the $n_r \times 1$ input signal, $\mathbf{P}$ is the $n_t \times n_r$ pre-equalization matrix, $\mathbf{H}$ is the $n_r \times n_t$ channel matrix and $\mathbf{w}$ is the $n_r \times 1$ additive channel noise. The input and noise vectors, $\mathbf{s}$ and $\mathbf{w}$, are assumed to be complex-valued, independent, circular, with covariance matrices $\mathbf{R}_s = \mathbf{I}_{n_r}$ and $\mathbf{R}_w = \sigma_w^2 \mathbf{I}_{n_r}$, respectively; furthermore, the noise is assumed to be Gaussian. This model is particularly suitable for the broadcast scenario, where the users cannot cooperate in order to combat inter-stream interference and, thus, the need for pre-equalization is imperative. In this case, the i-th element of $\mathbf{s}$ is the symbol intended for the

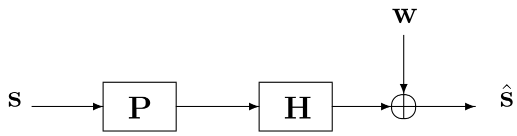

Fig. 2.1. System model

i-th user.

2.2.2 Computation of the TxWF

Our aim is to compute the TxWF $\mathbf{P}$ and the scalar β that minimize the cost function [13]

$$\mathrm{mse}(\mathbf{P},\beta) \triangleq \mathcal{E}\left[\|\mathbf{s}-\beta^{-1}\hat{\mathbf{s}}\|_2^2\right] \tag{2.2}$$

subject to the transmit power constraint

$$\mathcal{E}\left[\|\mathbf{P}\mathbf{s}\|_2^2\right] = \mathrm{E}. \tag{2.3}$$

Function mse$(\cdot)$ can be analytically expressed as

$$\mathrm{mse}(\mathbf{P},\beta) = \mathrm{tr}\left(\mathbf{I}_{n_r}\right) - 2\beta^{-1}\mathrm{Re}\left\{\mathrm{tr}\left(\mathbf{H}\mathbf{P}\right)\right\} + \beta^{-2}\mathrm{tr}\left(\mathbf{H}\mathbf{P}\mathbf{P}^H\mathbf{H}^H\right) + \beta^{-2}\mathrm{tr}\left(\mathbf{R}_w\right). \tag{2.4}$$

The optimal values for this constrained optimization problem are [13]

$$\beta_o = \sqrt{\frac{\mathrm{E}}{\mathrm{tr}\left(\tilde{\mathbf{P}}_o\tilde{\mathbf{P}}_o^H\right)}} \tag{2.5}$$

and $\mathbf{P}_o = \beta_o\tilde{\mathbf{P}}_o$, where

$$\tilde{\mathbf{P}}_o \triangleq \left(\mathbf{H}^H\mathbf{H} + \alpha\mathbf{I}_{n_t}\right)^{-1}\mathbf{H}^H \tag{2.6}$$

and

$$\alpha \triangleq \frac{\mathrm{tr}\left(\mathbf{R}_w\right)}{\mathrm{E}}. \tag{2.7}$$

In [12], quantity α has been defined as inverse SNR.

Using the optimal values $\mathbf{P}_o$ and β_o in (2.4), it can be shown that the MMSE is

$$\begin{aligned}\mathrm{MMSE} &\triangleq \mathrm{mse}(\mathbf{P}_o,\beta_o) = \mathrm{tr}\big(\mathbf{I}_{n_r}\big) - 2\,\mathrm{Re}\left\{\mathrm{tr}\big(\tilde{\mathbf{P}}_o\mathbf{H}\big)\right\} + \mathrm{tr}\big(\mathbf{H}\tilde{\mathbf{P}}_o\tilde{\mathbf{P}}_o^H\mathbf{H}^H\big) + \alpha\,\mathrm{tr}\big(\tilde{\mathbf{P}}_o\tilde{\mathbf{P}}_o^H\big)\\ &\triangleq \mathrm{MSE}(\tilde{\mathbf{P}}_o).\end{aligned} \tag{2.8}$$

2.2.3 Channel and noise SOS uncertainties

In the above development, we assumed that the channel matrix $\mathbf{H}$ and the noise covariance matrix $\mathbf{R}_w$ are perfectly known at the transmitter. Perhaps, the easiest way to obtain estimates of these quantities is through training. In frequency division duplex (FDD) systems, the estimates can be computed at the receiver and communicated to the transmitter via a feedback channel, while in TDD systems they can be computed at the transmitter. In this work, we consider channel estimation at the receiver, i.e., FDD systems. Of course, analogous results hold for TDD systems.

The channel estimate $\hat{\mathbf{H}}$ at the transmitter may suffer from two basic sources of uncertainty, namely, estimation errors and time-variations. We define the channel error or mismatch as

$$\Delta\mathbf{H} \stackrel{\triangle}{=} \hat{\mathbf{H}} - \mathbf{H} \tag{2.9}$$

and we assume that $\mathrm{vec}(\Delta\mathbf{H})$ is complex-valued, circular with

$$\mathbf{R}_{\mathrm{vec}(\Delta\mathbf{H})} \stackrel{\triangle}{=} \mathcal{E}\left[\mathrm{vec}(\Delta\mathbf{H})\mathrm{vec}^H(\Delta\mathbf{H})\right] = \mathbf{\Sigma}. \tag{2.10}$$

Next, we compute $\mathbf{\Sigma}$ for the two cases of interest.

1. *Channel estimation errors:* In this case, we assume that the channel is time-invariant and we estimate it using training. If we denote the $n_t \times N_{\mathrm{tr}}$ training block as $\mathbf{S}_{\mathrm{tr}}$ and the corresponding channel output as $\mathbf{Y}_{\mathrm{tr}}$, then the maximum likelihood (ML) channel estimate is [28, p. 174]

$$\hat{\mathbf{H}} = \mathbf{Y}_{\mathrm{tr}}\mathbf{S}_{\mathrm{tr}}^H\left(\mathbf{S}_{\mathrm{tr}}\mathbf{S}_{\mathrm{tr}}^H\right)^{-1}. \tag{2.11}$$

   Optimal channel estimates are obtained for semi-unitary training matrix $\mathbf{S}_{\mathrm{tr}}$, i.e., $\mathbf{S}_{\mathrm{tr}}\mathbf{S}_{\mathrm{tr}}^H \propto \mathbf{I}_{n_t}$. The corresponding channel estimation error covariance matrix is given by [28, p.175]

$$\mathbf{\Sigma} = \frac{\sigma_w^2}{N_{\mathrm{tr}}}\mathbf{I}_{n_t n_r}. \tag{2.12}$$

2. *Channel time-variations:* In this case, we assume that uncertainties due to channel time-variations dominate those due to channel estimation errors (i.e., we assume that the channel estimate is perfect and we focus on channel time-variations) [1]. We denote with $\mathbf{H} = \mathbf{H}_t$ the true channel at time instant t and with $\hat{\mathbf{H}} = \mathbf{H}_{t-\tau}$ the outdated channel version at the transmitter, where τ is the time needed for the feedback loop.

[1] We introduce the statistical model for the channel time-variations just for *analysis purposes*. Robust precoders exploiting this knowledge (see, for example [6]), are beyond the scope of this paper.

We assume that $\{\mathbf{H}_t\}$ is a stationary matrix random process and, at each time instant t, the elements of $\mathbf{H}_t$ are zero-mean, unit variance i.i.d. Gaussian random variables, yielding $\text{vec}(\mathbf{H}), \text{vec}(\hat{\mathbf{H}}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{n_t n_r})$. The channel coefficients are time-varying according to Jakes' model, with common maximum Doppler frequency f_d. Thus, $\hat{\mathbf{H}}$ and $\mathbf{H}$ can be modeled as jointly Gaussian with cross-correlation [3, p. 93]

$$\mathcal{E}\left[\text{vec}(\mathbf{H})\text{vec}^H(\hat{\mathbf{H}})\right] = \rho_\tau \mathbf{I}_{n_t n_r}$$

where ρ_τ is the normalized correlation coefficient specified by the Jakes model, i.e., $\rho_\tau = J_0(2\pi f_d \tau)$, with $J_0(\cdot)$ the zeroth-order Bessel function of the first kind. In this case, it can be easily proved that

$$\mathbf{\Sigma} = 2\,(1-\rho_\tau)\mathbf{I}_{n_t n_r}. \tag{2.13}$$

We continue with the noise SOS uncertainties. Since we assume that $\mathbf{R}_w = \sigma_w^2 \mathbf{I}_{n_r}$, we define the SOS estimation error as

$$\Delta\mathbf{R}_w \triangleq (\hat{\sigma}_w^2 - \sigma_w^2)\,\mathbf{I}_{n_r}. \tag{2.14}$$

Using training data $\mathbf{S}_{\text{tr}}$, it can be shown that an unbiased noise variance estimate is [24, p. 697]

$$\hat{\sigma}_w^2 = \frac{1}{n_r\,(N_{\text{tr}} - n_t)}\,\text{tr}\left(\mathbf{Y}_{\text{tr}}\mathbf{P}^{\perp}_{\mathbf{S}_{\text{tr}}^H}\mathbf{Y}_{\text{tr}}^H\right) \tag{2.15}$$

where $\mathbf{P}^{\perp}_{\mathbf{S}_{\text{tr}}^H}$ is the orthogonal projector onto the orthogonal complement of the column space of $\mathbf{S}_{\text{tr}}^H$. For more details, the reader is referred to, for example, [28, Sec. 9.4]. Using optimal training, it can be shown that the noise variance estimate (2.15) has variance (we prove it in Appendix 2A)

$$\mathcal{E}\left[\left(\hat{\sigma}_w^2 - \sigma_w^2\right)^2\right] = \frac{\sigma_w^4}{n_r(N_{\text{tr}} - n_t)}. \tag{2.16}$$

2.2.4 EMSE of the TxWF with uncertainties

In this subsection, we develop a second-order approximation to the EMSE induced by channel or noise SOS uncertainties. We denote with $\hat{\tilde{\mathbf{P}}}$ and $\hat{\beta}$ the scaled TxWF and the Wiener scalar computed by using the channel or the noise SOS estimates as if they were the true quantities. The corresponding TxWF is $\hat{\mathbf{P}} \triangleq \hat{\beta}\,\hat{\tilde{\mathbf{P}}}$. The MSE associated with $\hat{\mathbf{P}}$ and $\hat{\beta}$ is

$$\text{mse}(\hat{\mathbf{P}}, \hat{\beta}) = \text{MSE}(\hat{\tilde{\mathbf{P}}}).$$

Using a Taylor expansion of function MSE$(\cdot)$ around point $\tilde{\mathbf{P}}_o$, we obtain

$$\mathrm{MSE}(\hat{\tilde{\mathbf{P}}}) = \mathrm{MSE}(\tilde{\mathbf{P}}_o) + \mathrm{tr}\left(\Delta\tilde{\mathbf{P}}^H \mathrm{MSE}''(\tilde{\mathbf{P}}_o)\Delta\tilde{\mathbf{P}}\right) \tag{2.17}$$

where $\Delta\tilde{\mathbf{P}} \triangleq \hat{\tilde{\mathbf{P}}} - \tilde{\mathbf{P}}_o$ and $\mathrm{MSE}''(\tilde{\mathbf{P}}_o)$ is the second derivative of the function MSE, evaluated at the point $\tilde{\mathbf{P}}_o$. From (2.8), we obtain that [10]

$$\mathrm{MSE}''(\tilde{\mathbf{P}}_o) = \mathbf{H}^H\mathbf{H} + \alpha\mathbf{I}_{n_t}. \tag{2.18}$$

We define the EMSE as

$$\begin{aligned}\mathrm{EMSE}(\hat{\tilde{\mathbf{P}}}) &\triangleq \mathcal{E}\left[\mathrm{MSE}(\hat{\tilde{\mathbf{P}}}) - \mathrm{MSE}(\tilde{\mathbf{P}}_o)\right] = \mathcal{E}\left[\mathrm{tr}\big(\Delta\tilde{\mathbf{P}}^H \mathrm{MSE}''(\tilde{\mathbf{P}}_o)\Delta\tilde{\mathbf{P}}\big)\right] \\ &= \mathcal{E}\left[\mathrm{tr}\left(\Delta\tilde{\mathbf{P}}^H(\mathbf{H}^H\mathbf{H} + \alpha\mathbf{I}_{n_t})\Delta\tilde{\mathbf{P}}\right)\right].\end{aligned} \tag{2.19}$$

2.3 EMSE due to channel uncertainties

In this section, we assume that the transmitter perfectly knows the noise SOS and has obtained a channel estimate $\hat{\mathbf{H}}$, which is used for the computation of the TxWF. In order to compute the EMSE in (2.19), we must develop a first-order approximation to $\Delta\tilde{\mathbf{P}}$ with respect to $\Delta\mathbf{H}$. This is our task in the sequel. If we use in (2.6) the estimate $\hat{\mathbf{H}}$ as if it were the true channel $\mathbf{H}$, then we compute the scaled pre-equalization matrix

$$\hat{\tilde{\mathbf{P}}} = \left(\hat{\mathbf{H}}^H\hat{\mathbf{H}} + \alpha\mathbf{I}_{n_t}\right)^{-1}\hat{\mathbf{H}}^H \tag{2.20}$$

which can be written as

$$\hat{\tilde{\mathbf{P}}} = \Big(\mathbf{H}^H\mathbf{H} + \alpha\mathbf{I}_{n_t} + \underbrace{\mathbf{H}^H\Delta\mathbf{H} + \Delta\mathbf{H}^H\mathbf{H}}_{\mathbf{K}_\Delta} + O(\|\Delta\mathbf{H}\|^2)\Big)^{-1}(\mathbf{H}^H + \Delta\mathbf{H}^H). \tag{2.21}$$

Using the first-order approximation in (11) and definition (2.6), we obtain

$$\hat{\tilde{\mathbf{P}}} = \tilde{\mathbf{P}}_o - (\mathbf{H}^H\mathbf{H} + \alpha\mathbf{I}_{n_t})^{-1}(\mathbf{K}_\Delta\tilde{\mathbf{P}}_o - \Delta\mathbf{H}^H) + O(\|\Delta\mathbf{H}\|^2).$$

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

$$\Delta\tilde{\mathbf{P}} = -\underbrace{(\mathbf{H}^H\mathbf{H} + \alpha\mathbf{I}_{n_t})^{-1}}_{\mathbf{A}}\underbrace{(\mathbf{K}_\Delta\tilde{\mathbf{P}}_o - \Delta\mathbf{H}^H)}_{\boldsymbol{\Delta}} \tag{2.22}$$

and a second-order approximation of the EMSE is given by

$$\begin{aligned}\mathrm{EMSE}(\hat{\tilde{\mathbf{P}}}) &= \mathcal{E}\left[\mathrm{tr}\left(\Delta\tilde{\mathbf{P}}^H(\mathbf{H}^H\mathbf{H} + \alpha\mathbf{I}_{n_t})\Delta\tilde{\mathbf{P}}\right)\right] \overset{(2.22)}{=} \mathcal{E}\left[\mathrm{tr}\left(\boldsymbol{\Delta}^H\mathbf{A}\,\boldsymbol{\Delta}\right)\right] \\ &= \mathcal{E}\left[\mathrm{tr}\left(\mathbf{A}\,\boldsymbol{\Delta}\,\mathbf{I}_{n_r}\,\boldsymbol{\Delta}^H\right)\right] \overset{(1)}{=} \mathcal{E}\left[\mathrm{vec}^H(\boldsymbol{\Delta})\big(\mathbf{I}_{n_r}\otimes\mathbf{A}\big)\mathrm{vec}(\boldsymbol{\Delta})\right] \\ &= \mathrm{tr}\left(\big(\mathbf{I}_{n_r}\otimes\mathbf{A}\big)\,\mathcal{E}\left[\mathrm{vec}(\boldsymbol{\Delta})\mathrm{vec}^H(\boldsymbol{\Delta})\right]\right).\end{aligned} \tag{2.23}$$

From the definitions of $\mathbf{\Delta}$ in (2.22) and $\mathbf{K}_{\Delta}$ in (2.21), we obtain

$$\begin{aligned}\mathrm{vec}(\mathbf{\Delta}) &= \mathrm{vec}(\mathbf{H}^H \Delta\mathbf{H}\tilde{\mathbf{P}}_o) + \mathrm{vec}\left(\Delta\mathbf{H}^H\left(\mathbf{H}\tilde{\mathbf{P}}_o - \mathbf{I}_{n_r}\right)\right)\\ &= \underbrace{(\tilde{\mathbf{P}}_o^T \otimes \mathbf{H}^H)}_{\mathbf{M}_1}\mathrm{vec}(\Delta\mathbf{H}) + \underbrace{((\tilde{\mathbf{P}}_o^T\mathbf{H}^T - \mathbf{I}_{n_r}) \otimes \mathbf{I}_{n_t})}_{\mathbf{M}_2}\mathrm{vec}(\Delta\mathbf{H}^H)\end{aligned} \tag{2.24}$$

where we made use of (2). Using the commutation matrix $\mathbf{K}_{n_t n_r}$ and (8) we obtain

$$\mathrm{vec}(\Delta) = \mathbf{M}_1 \mathrm{vec}(\Delta\mathbf{H}) + \mathbf{M}_2\,\mathbf{K}\mathrm{vec}(\Delta\mathbf{H}^*)$$

where, for notational simplicity, the commutation matrix is denoted as $\mathbf{K}$. Using the circular symmetry of $\Delta\mathbf{H}$ and (2.10), we obtain

$$\mathrm{EMSE}(\hat{\tilde{\mathbf{P}}}) = \mathrm{tr}\left((\mathbf{I}_{n_r} \otimes \mathbf{A})\left(\mathbf{M}_1\,\mathbf{\Sigma}\,\mathbf{M}_1^H + \mathbf{M}_2\,\mathbf{K}\mathbf{\Sigma}^*\mathbf{K}^H\mathbf{M}_2^H\right)\right).$$

Finally, we obtain the expression

$$\mathrm{EMSE}(\hat{\tilde{\mathbf{P}}}) = \mathbf{T}_1 + \mathbf{T}_2 \tag{2.25}$$

where

$$\begin{aligned}\mathbf{T}_1 &\stackrel{\triangle}{=} \mathrm{tr}\left((\mathbf{I}_{n_r} \otimes \mathbf{A})\,\mathbf{M}_1\mathbf{\Sigma}\mathbf{M}_1^H\right) = \mathrm{tr}\left(\mathbf{M}_1^H\,(\mathbf{I}_{n_r} \otimes \mathbf{A})\,\mathbf{M}_1\mathbf{\Sigma}\right)\\ &\stackrel{(2.24)}{=} \mathrm{tr}\left(\left(\tilde{\mathbf{P}}_o^*\tilde{\mathbf{P}}_o^T \otimes \mathbf{H}\mathbf{A}\mathbf{H}^H\right)\mathbf{\Sigma}\right)\end{aligned} \tag{2.26}$$

and

$$\begin{aligned}\mathbf{T}_2 &\stackrel{\triangle}{=} \mathrm{tr}\left((\mathbf{I}_{n_r} \otimes \mathbf{A})\,\mathbf{M}_2\mathbf{K}\mathbf{\Sigma}^*\mathbf{K}^H\mathbf{M}_2^H\right) = \mathrm{tr}\left(\mathbf{K}^H\mathbf{M}_2^H\,(\mathbf{I}_{n_r} \otimes \mathbf{A})\,\mathbf{M}_2\mathbf{K}\mathbf{\Sigma}^*\right)\\ &\stackrel{(2.24)}{=} \mathrm{tr}\left(\mathbf{K}^H\left((\mathbf{H}^*\tilde{\mathbf{P}}_o^* - \mathbf{I}_{n_r})(\tilde{\mathbf{P}}_o^T\mathbf{H}^T - \mathbf{I}_{n_r}) \otimes \mathbf{A}\right)\mathbf{K}\mathbf{\Sigma}^*\right)\\ &= \mathrm{tr}\left(\left(\mathbf{A} \otimes (\mathbf{H}^*\tilde{\mathbf{P}}_o^* - \mathbf{I}_{n_r})(\tilde{\mathbf{P}}_o^T\mathbf{H}^T - \mathbf{I}_{n_r})\right)\mathbf{\Sigma}^*\right)\end{aligned} \tag{2.27}$$

where we made use of expressions (3) and (5).

Until now, we have expressed the EMSE in terms of $\mathbf{\Sigma}$. Expressions (2.25)–(2.27) are admittedly complicated and do not provide significant insight. In the sequel, we assume sufficiently high SNR and we derive simple and informative approximations of the EMSE.

2.3.1 Channel estimation errors and high SNR

In this subsection, we assume that the channel uncertainties are due to estimation errors, implying that $\mathbf{\Sigma} = (\sigma_w^2/N_{\mathrm{tr}})\,\mathbf{I}_{n_t n_r}$. Using the SVD of $\mathbf{H}$, it can be shown that, for $i = 1, \ldots, n_r$ (recall the definition of $\mathbf{A}$ in (2.22))

$$\lambda_i\left(\mathbf{H}\mathbf{A}\mathbf{H}^H\right) = \frac{\lambda_i\left(\mathbf{H}^H\mathbf{H}\right)}{\lambda_i\left(\mathbf{H}^H\mathbf{H}\right) + \alpha}. \tag{2.28}$$

For $\alpha \ll \lambda_{n_r}(\mathbf{H}^H\mathbf{H})$, that is, α much smaller than the smallest nonzero eigenvalue of $\mathbf{H}^H\mathbf{H}$, implying sufficiently high SNR, we obtain

$$\operatorname{tr}\left(\mathbf{HAH}^H\right) = \sum_{i=1}^{n_r} \frac{\lambda_i\left(\mathbf{H}^H\mathbf{H}\right)}{\lambda_i\left(\mathbf{H}^H\mathbf{H}\right)+\alpha} \approx \operatorname{tr}\left(\mathbf{I}_{n_r}\right). \tag{2.29}$$

Another high SNR approximation that will prove useful in the sequel is (the proof is provided in Appendix 2B)

$$\operatorname{tr}\!\left(\tilde{\mathbf{P}}_o\tilde{\mathbf{P}}_o^H\right) \approx \frac{1}{\alpha}\,\mathrm{MMSE}. \tag{2.30}$$

Starting with $\mathbf{T}_1$ in (2.26) and using (4) we obtain

$$\begin{aligned}\mathbf{T}_1 &\overset{(2.12)}{=} \frac{\sigma_w^2}{N_{\mathrm{tr}}}\operatorname{tr}\!\left(\tilde{\mathbf{P}}_o^*\tilde{\mathbf{P}}_o^T \otimes \mathbf{HAH}^H\right) \overset{(2.29)}{\approx} \frac{\sigma_w^2}{N_{\mathrm{tr}}}\operatorname{tr}\left(\tilde{\mathbf{P}}_o^*\tilde{\mathbf{P}}_o^T\right)\operatorname{tr}\left(\mathbf{I}_{n_r}\right)\\ &= \frac{n_r\,\sigma_w^2}{N_{\mathrm{tr}}}\operatorname{tr}\left(\tilde{\mathbf{P}}_o^*\tilde{\mathbf{P}}_o^T\right) \overset{(2.30)}{\approx} \frac{n_r\,\sigma_w^2}{N_{\mathrm{tr}}}\,\frac{1}{\alpha}\,\mathrm{MMSE}.\end{aligned} \tag{2.31}$$

In order to compute an approximation of $\mathbf{T}_2$, we use an expression analogous to (2.28), for $i = 1, \ldots, n_r$

$$\lambda_i\left((\mathbf{H}^*\tilde{\mathbf{P}}_o^* - \mathbf{I}_{n_r})(\tilde{\mathbf{P}}_o^T\mathbf{H}^T - \mathbf{I}_{n_r})\right) = \frac{\alpha^2}{\left(\lambda_i\left(\mathbf{H}^H\mathbf{H}\right)+\alpha\right)^2}. \tag{2.32}$$

For high SNR, the right-hand side of (2.32) goes to zero, yielding

$$\operatorname{tr}\left((\mathbf{H}^*\tilde{\mathbf{P}}_o^* - \mathbf{I}_{n_r})(\tilde{\mathbf{P}}_o^T\mathbf{H}^T - \mathbf{I}_{n_r})\right) \approx 0. \tag{2.33}$$

Thus,

$$\mathbf{T}_2 \overset{(2.12)}{=} \frac{\sigma_w^2}{N_{\mathrm{tr}}}\operatorname{tr}\left(\mathbf{A}\right)\operatorname{tr}\left((\mathbf{H}^*\tilde{\mathbf{P}}_o^* - \mathbf{I}_{n_r})(\tilde{\mathbf{P}}_o^T\mathbf{H}^T - \mathbf{I}_{n_r})\right) \overset{(2.33)}{\approx} 0. \tag{2.34}$$

We conclude that, for sufficiently high SNR, term $\mathbf{T}_2$ is negligible compared with $\mathbf{T}_1$; this statement is in agreement with simulations in Section V. Combining expressions (2.25), (2.31) and (2.34), we obtain

$$\mathrm{EMSE}(\hat{\hat{\mathbf{P}}}) \approx \frac{n_r\,\sigma_w^2}{N_{\mathrm{tr}}}\,\frac{1}{\alpha}\,\mathrm{MMSE} \overset{(2.7)}{=} \frac{\mathrm{E}}{N_{\mathrm{tr}}}\,\mathrm{MMSE}.$$

Thus, for optimal training and sufficiently high SNR

$$\boxed{\mathrm{EMSE}(\hat{\hat{\mathbf{P}}}) \approx \frac{\mathrm{E}}{N_{\mathrm{tr}}}\,\mathrm{MMSE}.} \tag{2.35}$$

We observe that the EMSE is approximately proportional to the MMSE, with the proportionality factor being the ratio of the transmit power, E, to the length of the training block used for channel estimation, N_{tr}. Expression (2.35) 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.

2.3.2 Time-varying channels and high SNR

In this subsection, we assume that the uncertainties due to time-variations dominate those due to estimation errors, yielding $\Sigma = 2\,(1-\rho_\mathcal{T})\,\mathbf{I}_{n_t n_r}$. The only term of the previous analysis that is affected is (2.26), which becomes

$$\mathbf{T}_1 \overset{(2.13)}{=} 2\,(1-\rho_\mathcal{T})\,\mathrm{tr}\Big(\tilde{\mathbf{P}}_o^*\tilde{\mathbf{P}}_o^T \otimes \mathbf{HAH}^H\Big) \overset{(2.29)}{\approx} 2\,(1-\rho_\mathcal{T})\,\mathrm{tr}\left(\tilde{\mathbf{P}}_o^*\tilde{\mathbf{P}}_o^T\right)\mathrm{tr}\left(\mathbf{I}_{n_r}\right). \tag{2.36}$$

giving that

$$\boxed{\mathrm{EMSE}(\hat{\tilde{\mathbf{P}}}) \approx 2\,n_r(1-\rho_\mathcal{T})\mathrm{tr}\left(\tilde{\mathbf{P}}_o^*\tilde{\mathbf{P}}_o^T\right).} \tag{2.37}$$

It is easy to see that

$$\mathrm{tr}\left(\tilde{\mathbf{P}}_o^*\tilde{\mathbf{P}}_o^T\right) = \mathrm{tr}\left(\tilde{\mathbf{P}}_o\tilde{\mathbf{P}}_o^H\right) = \mathrm{tr}\left(\mathbf{HA}^2\mathbf{H}^H\right) = \sum_{i=1}^{n_r}\frac{\lambda_i(\mathbf{HH}^H)}{(\lambda_i(\mathbf{HH}^H+\alpha)^2}$$

is an increasing function of SNR and tends to $\mathrm{tr}\left((\mathbf{HH}^H)^{-1}\right)$ for SNR tending to infinity. Thus, the EMSE increases for increasing SNR and asymptotically attains the value

$$\boxed{\mathrm{EMSE}(\hat{\tilde{\mathbf{P}}}) \approx 2\,n_r(1-\rho_\mathcal{T})\,\mathrm{tr}((\mathbf{HH}^H)^{-1}).} \tag{2.38}$$

Of course, the above approximations are accurate for slow time-variations because fast time-variations introduce large channel uncertainties rendering our asymptotic analysis inaccurate.

2.4 EMSE due to noise SOS uncertainties

In this section, we assume that the channel is perfectly known at the transmitter and the noise SOS estimate $\hat{\mathbf{R}}_w$ is used as if it were the true $\mathbf{R}_w$. Then the scaled precoding matrix becomes

$$\hat{\tilde{\mathbf{P}}} = \left(\mathbf{H}^H\mathbf{H} + \frac{\mathrm{tr}\,(\mathbf{R}_w+\Delta\mathbf{R}_w)}{\mathrm{E}}\,\mathbf{I}_{n_t}\right)^{-1}\mathbf{H}^H. \tag{2.39}$$

Using (11), a first-order approximation to $\Delta\tilde{\mathbf{P}}$, with respect to $\Delta\mathbf{R}_w$, is given by

$$\Delta\tilde{\mathbf{P}} = -\frac{\mathrm{tr}\,(\Delta\mathbf{R}_w)}{\mathrm{E}}\left(\mathbf{H}^H\mathbf{H}+\alpha\mathbf{I}_{n_t}\right)^{-1}\tilde{\mathbf{P}}_o \overset{(2.22)}{=} -\frac{\mathrm{tr}\,(\Delta\mathbf{R}_w)}{\mathrm{E}}\,\mathbf{A}\,\tilde{\mathbf{P}}_o. \tag{2.40}$$

Substituting the above expression in (2.19), we obtain the second-order approximation

$$\begin{aligned}\mathrm{EMSE}(\hat{\tilde{\mathbf{P}}}) &= \mathcal{E}\left[\mathrm{tr}\left(\Delta\tilde{\mathbf{P}}^H(\mathbf{H}^H\mathbf{H}+\alpha\mathbf{I}_{n_t})\Delta\tilde{\mathbf{P}}\right)\right] \\ &= \frac{\mathcal{E}\left[\mathrm{tr}^2\,(\Delta\mathbf{R}_w)\right]}{\mathrm{E}^2}\,\mathrm{tr}\left(\tilde{\mathbf{P}}_o^H\mathbf{A}\tilde{\mathbf{P}}_o\right).\end{aligned} \tag{2.41}$$

2.4.1 Optimal training and high SNR

Considering optimal training and high SNR, recalling that we consider the spatially and temporally white Gaussian noise case and using (2.16), we get

$$\mathcal{E}\left[\mathrm{tr}^2\left(\Delta\mathbf{R}_w\right)\right] = \mathcal{E}\left[n_r^2\left(\hat{\sigma}_w^2 - \sigma_w^2\right)^2\right] = n_r^2\,\mathcal{E}\left[\left(\hat{\sigma}_w^2 - \sigma_w^2\right)^2\right]. \tag{2.42}$$

Using (2.16) and (2.42), (2.41) becomes

$$\mathrm{EMSE}(\hat{\tilde{\mathbf{P}}}) = \frac{n_r^2\,\mathcal{E}\left[\left(\hat{\sigma}_w^2 - \sigma_w^2\right)^2\right]}{\mathrm{E}^2}\,\mathrm{tr}\left(\tilde{P}_o^H\mathbf{A}\tilde{P}_o\right) = \frac{n_r\,\sigma_w^4}{\mathrm{E}^2\left(N_{\mathrm{tr}} - n_t\right)}\,\mathrm{tr}\left(\tilde{\mathbf{P}}_o^H\mathbf{A}\tilde{\mathbf{P}}_o\right). \tag{2.43}$$

Using the definitions of $\tilde{\mathbf{P}}_o$ and $\mathbf{A}$ in (2.6) and (2.22), respectively, we write

$$\mathrm{tr}\left(\tilde{\mathbf{P}}_o^H\mathbf{A}\tilde{\mathbf{P}}_o\right) = \mathrm{tr}\left(\mathbf{H}\left(\mathbf{H}^H\mathbf{H} + \alpha\mathbf{I}_{n_t}\right)^{-3}\mathbf{H}^H\right) = \mathrm{tr}\left(\mathbf{H}\mathbf{A}^3\mathbf{H}^H\right). \tag{2.44}$$

An expression analogous to (2.28), for $i = 1, \ldots, n_r$, is

$$\lambda_i\left(\mathbf{H}\mathbf{A}^3\mathbf{H}^H\right) = \frac{\lambda_i\left(\mathbf{H}^H\mathbf{H}\right)}{\left(\lambda_i\left(\mathbf{H}^H\mathbf{H}\right) + \alpha\right)^3}$$

which, for high SNR, gives

$$\lambda_i\left(\mathbf{H}\mathbf{A}^3\mathbf{H}^H\right) \approx \frac{1}{\lambda_i^2\left(\mathbf{H}^H\mathbf{H}\right)} = \frac{1}{\lambda_i^2\left(\mathbf{H}\mathbf{H}^H\right)}.$$

Thus

$$\mathrm{tr}\left(\tilde{\mathbf{P}}_o^H\mathbf{A}\tilde{\mathbf{P}}_o\right) \approx \sum_{i=1}^{n_r}\frac{1}{\lambda_i^2\left(\mathbf{H}\mathbf{H}^H\right)} = ||\left(\mathbf{H}\mathbf{H}^H\right)^{-1}||_F^2 \tag{2.45}$$

and finally, combining expressions (2.43) and (2.45), we obtain

$$\boxed{\mathrm{EMSE}(\hat{\tilde{\mathbf{P}}}) \approx \frac{n_r\,\sigma_w^4}{\mathrm{E}^2\left(N_{\mathrm{tr}} - n_t\right)}||\left(\mathbf{H}\mathbf{H}^H\right)^{-1}||_F^2.} \tag{2.46}$$

This approximation states that the EMSE is proportional to the squared noise variance, σ_w^4, which decreases very fast for increasing SNR. The proportionality factor is determined by the transmit power, E, the length of the training block, N_{tr}, the number of the transmit and receive antennas n_t and n_r, and the conditioning of the matrix channel $\mathbf{H}$, through the Frobenius norm $||\left(\mathbf{H}\mathbf{H}^H\right)^{-1}||_F^2$. 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**

-0.2646 + 0.1212*j	-0.0456 - 0.2588*j	-0.0081 - 0.7268*j
0.0664 + 0.0179*j	-0.1597 + 0.4986*j	0.0656 - 0.1866*j

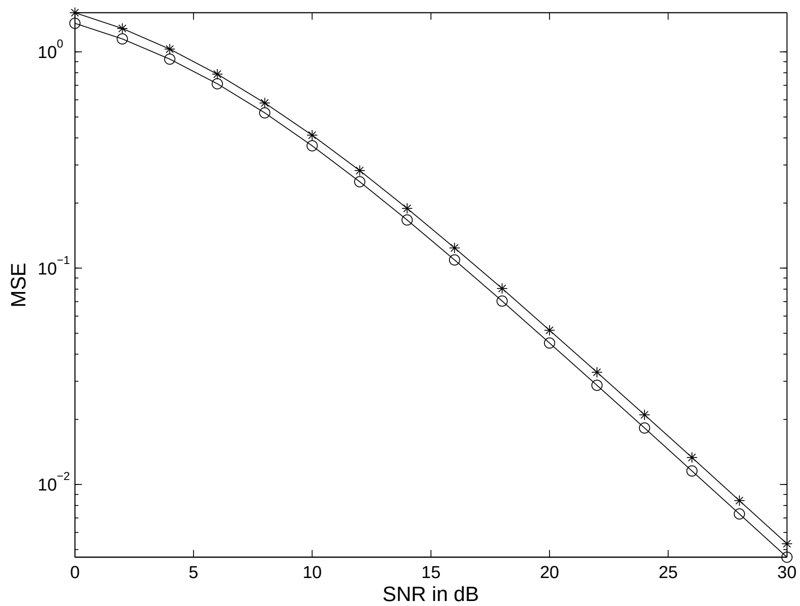

Fig. 2.2. MMSE using the true channel ('-o') and expectation of the MSEs using the channel estimate ('-*').

2.5 Simulation Results

We consider a system with $n_t = 3$ transmit antennas and $n_r = 2$ receive antennas. We consider the channel matrix **H** with elements given in Table I [2]. The noise is spatially and temporally white, circularly symmetric complex Gaussian with variance σ_w^2. We set the transmit power $\mathrm{E} = n_t$. We assume that the training block is composed of $N_{\mathrm{tr}} = 20$ columns.

Simulation 1. *Channel estimation errors.*

In Fig. 2.2, we plot the MMSE (2.8) and the mean of the MSEs computed using the channel estimate (the average is over different realizations of the channel estimation error

[2] Our results hold for *any* channel matrix. Since in our theoretical developments we did not average over the channels, but only over the channel uncertainties, in the simulations we use only one channel realization. We have made analogous observations in extensive simulation studies.

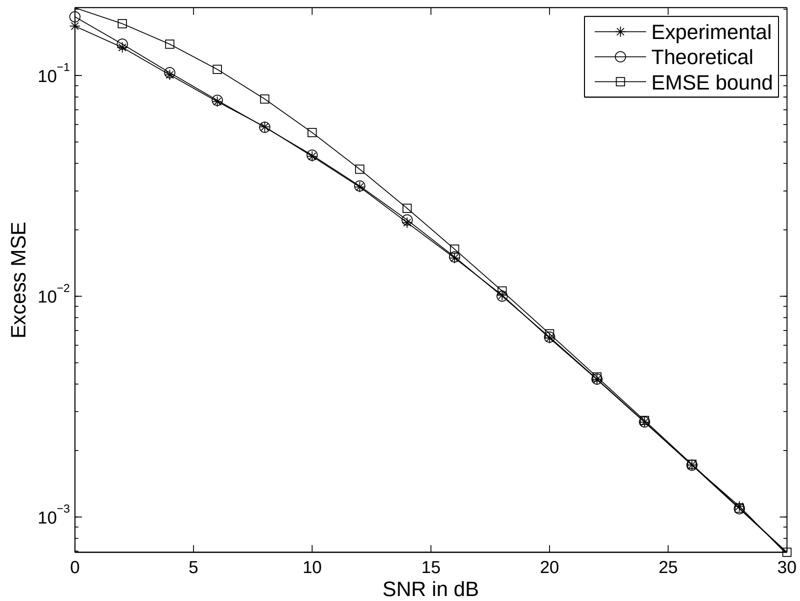

Fig. 2.3. Experimentally computed EMSE, theoretical second-order approximation (2.25), and high SNR approximation (2.35) for the case of channel estimation errors.

$\Delta\mathbf{H}$). We observe that the distance of these two quantities is approximately constant and does not depend on the SNR, verifying expression (2.35).

In Fig. 2.3, we present the experimentally computed EMSE, the theoretical second-order approximation (2.25) and approximation (2.35). We observe that the experimental and theoretical EMSE values practically coincide for SNR higher than 5 dB, while approximation (2.35) is very close to the EMSE, especially at high SNR.

In Fig. 2.4, we plot terms $\mathbf{T}_1$ and $\mathbf{T}_2$ of the theoretical EMSE of (2.25). We observe that, for SNR higher than 7 dB, the contribution of term $\mathbf{T}_2$ to the EMSE is much smaller than the contribution of term $\mathbf{T}_1$, supporting our claim that the EMSE is approximately equal to term $\mathbf{T}_1$ for the high SNR cases.

Simulation 2. *Channel time-variations.*

In Fig. 2.5, we plot the MMSE (2.8) and the mean of the MSEs computed using the outdated channel versions for channel correlation coefficients $\rho_\mathcal{T} = 0.99, 0.9$ (the average is over different realizations of the channel uncertainties due to channel time-variations $\Delta\mathbf{H}$). We observe that the distance of the two curves from the MMSE increases for increasing SNR, and the mean of the MSEs reaches a floor. This happens because the EMSE induced by the channel time-variations increases for increasing SNR and asymptotically attains a

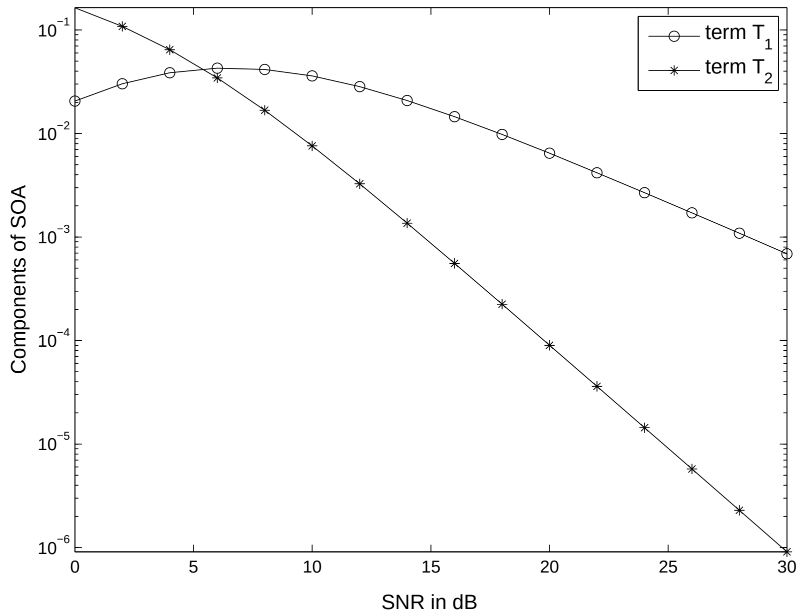

Fig. 2.4. Terms $\mathbf{T}_1$ and $\mathbf{T}_2$ of the EMSE second-order approximation (2.25) for the case of channel estimation errors.

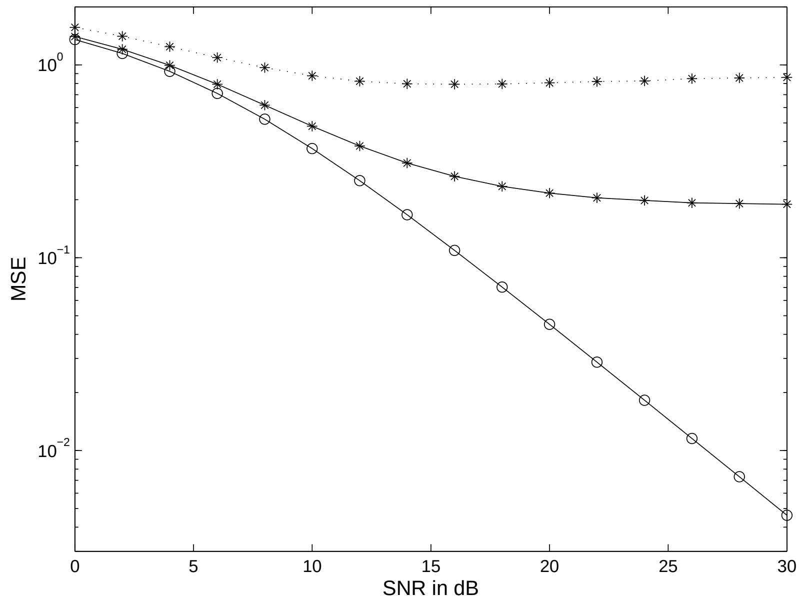

Fig. 2.5. MMSE using the true channel ('-o') and expectation of the MSEs using the channel estimate for $\rho = 0.99$ ('-*') and for $\rho = 0.9$ (dotted line).

limit value.

In Fig. 2.6, we present the theoretical second-order approximation (2.25), the corre-

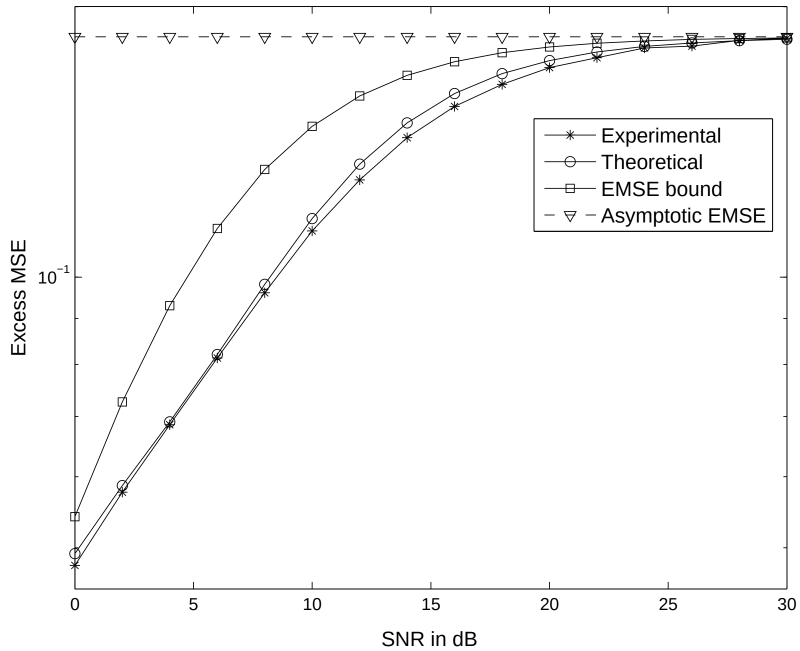

Fig. 2.6. Experimentally computed EMSE, theoretical second-order approximation (2.25), high SNR approximation (2.37) and asymptotic EMSE value (2.38) for channel time-variations ($\rho = 0.99$).

sponding experimentally computed EMSE, the high SNR approximation (2.37) and the asymptotic value (2.38) for channel correlation coefficient equal to $\rho = 0.99$, implying very accurate channel information at the transmitter. We observe that the second-order approximation is very accurate, while (2.37) is a good approximation to the EMSE for SNR higher than 20 dB.

Simulation 3. *Noise estimation errors.*

In Fig. 2.7, we present the theoretical second-order approximation (2.43), the corresponding experimentally computed EMSE and the high SNR approximation (2.46). We observe that the first two quantities practically coincide and approximation (2.46) is very close to the true EMSE for SNR higher than 15 dB.

Comparing the EMSEs for the cases of estimation errors only (see Fig. 2.3 and Fig. 2.7), we observe that the error induced by the channel estimation errors is much more significant than that induced by the noise SOS estimation errors.

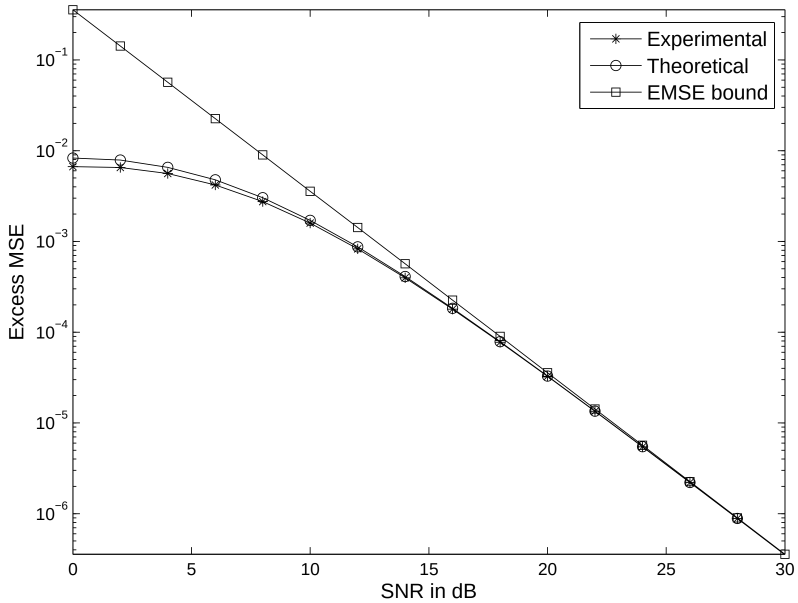

Fig. 2.7. Experimentally computed EMSE, theoretical second-order approximation (2.43), and high SNR approximation (2.46) for the case of noise SOS estimation errors.

2.6 Conclusion

We considered the behavior of the TxWF under channel and noise SOS uncertainties by developing second-order EMSE approximations. We derived simple EMSE approximations in the high SNR cases. Considering the channel estimation errors, we concluded that the EMSE is proportional to the MMSE, with the proportionality factor determined by the transmit power E and the length of the training block N_{tr}. Considering the channel time-variations, we found that the EMSE increases and for high SNR it reaches an asymptotic value. For the case of noise SOS estimation errors, we showed that the EMSE is proportional to the squared noise variance, σ_w^4. A comparison of the EMSEs for the cases of estimation errors only, shows that the error induced by the channel estimate is much more significant than that induced by the noise SOS estimate.

Appendix 2A

Channel and noise variance ML estimates

The ML estimate of the channel gain matrix $\mathbf{H}$ and the noise variance estimate can be derived from training-based estimation [28]. Using the training block $\mathbf{S}_{\mathrm{tr}}$ of dimension $n_t \times N_{\mathrm{tr}}$, the received block of dimension $n_r \times N_{\mathrm{tr}}$ is given by [28]

$$\mathbf{Y}_{\mathrm{tr}} = \mathbf{H}\mathbf{S}_{\mathrm{tr}} + \mathbf{E}_{\mathrm{tr}}$$

where $\mathbf{E}_{\mathrm{tr}}$ is the corresponding $n_r \times N_{\mathrm{tr}}$ 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 $\mathbf{H}$ based on the received training block $\mathbf{Y}_{\mathrm{tr}}$ is given by [28, p. 174]

$$\hat{\mathbf{H}} = \mathbf{Y}_{\mathrm{tr}}\mathbf{S}_{\mathrm{tr}}^H \left(\mathbf{S}_{\mathrm{tr}}\mathbf{S}_{\mathrm{tr}}^H\right)^{-1}.$$

This estimate is unbiased and the covariance matrix of $\mathrm{vec}(\Delta\mathbf{H})$ is given by [28, p. 175]

$$\mathbf{\Sigma} \triangleq \mathcal{E}\left[\mathrm{vec}(\Delta\mathbf{H})\mathrm{vec}^H(\Delta\mathbf{H})\right] = \sigma_w^2 \left(\left(\mathbf{S}_{\mathrm{tr}}\mathbf{S}_{\mathrm{tr}}^H\right)^{-T} \otimes \mathbf{I}_{n_r}\right).$$

As shown in [28, p. 176], the optimal training block $\mathbf{S}_{\mathrm{tr}}$ should satisfy

$$\mathbf{S}_{\mathrm{tr}}\mathbf{S}_{\mathrm{tr}}^H \propto \mathbf{I}_{n_t}.$$

B. ML noise variance estimate

Having estimated the channel matrix $\mathbf{H}$, the ML noise variance estimate is [28, p. 174]

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

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

$$\begin{aligned}\mathcal{E}\left[\mathrm{tr}\left(\mathbf{Y}_{\mathrm{tr}}\mathbf{P}^{\perp}_{\mathbf{S}_{\mathrm{tr}}^H}\mathbf{Y}_{\mathrm{tr}}^H\right)\right] &= \mathcal{E}\left[\mathrm{tr}\left(\left(\mathbf{H}\mathbf{S}_{\mathrm{tr}} + \mathbf{E}_{\mathrm{tr}}\right)\mathbf{P}^{\perp}_{\mathbf{S}_{\mathrm{tr}}^H}\left(\mathbf{S}_{\mathrm{tr}}^H\mathbf{H}^H + \mathbf{E}_{\mathrm{tr}}^H\right)\right)\right] \\ &= \mathcal{E}\left[\mathrm{tr}\left(\mathbf{E}_{\mathrm{tr}}^H\mathbf{E}_{\mathrm{tr}}\mathbf{P}^{\perp}_{\mathbf{S}_{\mathrm{tr}}^H}\right)\right]\end{aligned}$$

giving that

$$\mathcal{E}\left[\mathrm{tr}\left(\mathbf{Y}_{\mathrm{tr}}\mathbf{P}^{\perp}_{\mathbf{S}_{\mathrm{tr}}^H}\mathbf{Y}_{\mathrm{tr}}^H\right)\right] = \left(N_{\mathrm{tr}} - n_t\right) n_r \sigma_w^2.$$

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

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

where

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

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

$$\begin{aligned}
\mathrm{var}\hat{\hat{\sigma}}_w^2 &= \mathcal{E}\left[\left|\sigma_w^2 - \hat{\hat{\sigma}}_w^2\right|^2\right] = \mathcal{E}\left[\left|\sigma_w^2 - \frac{1}{c}\mathrm{tr}\left(\mathbf{P}^{\perp}_{\mathbf{S}^H_{\rm tr}}\mathbf{E}^H_{\rm tr}\mathbf{E}_{\rm tr}\right)\right|^2\right] \\
&= \sigma_w^4 - \frac{2}{c}\,\sigma_w^2 \mathrm{Re}\Bigg\{\underbrace{\mathcal{E}\left[\left|\mathrm{tr}\left(\mathbf{P}^{\perp}_{\mathbf{S}^H_{\rm tr}}\mathbf{E}^H_{\rm tr}\mathbf{E}_{\rm tr}\right)\right|\right]}_{=c\,\sigma_w^2}\Bigg\} + \frac{1}{c^2}\,\mathcal{E}\left[\left|\mathrm{tr}\left(\mathbf{P}^{\perp}_{\mathbf{S}^H_{\rm tr}}\mathbf{E}^H_{\rm tr}\mathbf{E}_{\rm tr}\right)\right|^2\right] \\
&= -\sigma_w^4 + \frac{1}{c^2}\,\underbrace{\mathcal{E}\left[\left|\mathrm{tr}\left(\mathbf{P}^{\perp}_{\mathbf{S}^H_{\rm tr}}\mathbf{E}^H_{\rm tr}\mathbf{E}_{\rm tr}\right)\right|^2\right]}_{b}.
\end{aligned}$$

In order to compute term b, we examine $\mathrm{tr}\left(\mathbf{P}^{\perp}_{\mathbf{S}^H_{\rm tr}}\mathbf{E}^H_{\rm tr}\mathbf{E}_{\rm tr}\right)$

$$\begin{aligned}
\mathrm{tr}\left(\mathbf{P}^{\perp}_{\mathbf{S}^H_{\rm tr}}\mathbf{E}^H_{\rm tr}\mathbf{E}_{\rm tr}\right) &= \mathrm{tr}\left(\mathbf{P}^{\perp}_{\mathbf{S}^H_{\rm tr}}\mathbf{E}^H_{\rm tr}\mathbf{I}_{n_r}\mathbf{E}_{\rm tr}\right) \\
&\overset{(1)}{=} \underbrace{\mathrm{vec}^T\left(\mathbf{E}^T_{\rm tr}\right)}_{\triangleq \mathbf{e}^T}\left(\mathbf{I}_{n_r}\otimes\mathbf{P}^{\perp}_{\mathbf{S}^H_{\rm tr}}\right)\underbrace{\mathrm{vec}(\mathbf{E}^H_{\rm tr})}_{=\mathbf{e}^*} \\
&= \mathbf{e}^T\left(\mathbf{I}_{n_r}\otimes\mathbf{P}^{\perp}_{\mathbf{S}^H_{\rm tr}}\right)\mathbf{e}^*.
\end{aligned}$$

Thus,

$$\begin{aligned}
b &= \mathcal{E}\left[\left|\mathrm{tr}\left(\mathbf{P}^{\perp}_{\mathbf{S}^H_{\rm tr}}\mathbf{E}^H_{\rm tr}\mathbf{E}_{\rm tr}\right)\right|^2\right] = \mathcal{E}\left[\mathrm{tr}\left(\mathbf{P}^{\perp}_{\mathbf{S}^H_{\rm tr}}\mathbf{E}^H_{\rm tr}\mathbf{E}_{\rm tr}\right)\mathrm{tr}^H\left(\mathbf{P}^{\perp}_{\mathbf{S}^H_{\rm tr}}\mathbf{E}^H_{\rm tr}\mathbf{E}_{\rm tr}\right)\right] \\
&= \mathcal{E}\left[\mathbf{e}^T\left(\mathbf{I}_{n_r}\otimes\mathbf{P}^{\perp}_{\mathbf{S}^H_{\rm tr}}\right)\mathbf{e}^*\,\mathbf{e}^T\left(\mathbf{I}_{n_r}\otimes\mathbf{P}^{\perp}_{\mathbf{S}^H_{\rm tr}}\right)\mathbf{e}^*\right] \\
&= \mathcal{E}\left[\mathrm{tr}\left(\left(\mathbf{I}_{n_r}\otimes\mathbf{P}^{\perp}_{\mathbf{S}^H_{\rm tr}}\right)^T\mathbf{e}^*\mathbf{e}^T\left(\mathbf{I}_{n_r}\otimes\mathbf{P}^{\perp}_{\mathbf{S}^H_{\rm tr}}\right)\mathbf{e}^*\mathbf{e}^T\right)\right] \\
&\overset{(1)}{=} \mathcal{E}\left[\mathrm{vec}^T\left(\mathbf{e}\mathbf{e}^H\right)\left(\mathbf{I}_{n_r}\otimes\mathbf{P}^{\perp}_{\mathbf{S}^H_{\rm tr}}\right)^T\otimes\left(\mathbf{I}_{n_r}\otimes\mathbf{P}^{\perp}_{\mathbf{S}^H_{\rm tr}}\right)\mathrm{vec}\left(\mathbf{e}^*\mathbf{e}^T\right)\right] \\
&= \mathcal{E}\left[\mathrm{tr}\left(\left(\mathbf{I}_{n_r}\otimes\mathbf{P}^{\perp}_{\mathbf{S}^H_{\rm tr}}\right)^T\otimes\left(\mathbf{I}_{n_r}\otimes\mathbf{P}^{\perp}_{\mathbf{S}^H_{\rm tr}}\right)\mathrm{vec}\left(\mathbf{e}^*\mathbf{e}^T\right)\mathrm{vec}^T\left(\mathbf{e}\mathbf{e}^H\right)\right)\right] \\
&= \mathrm{tr}\left(\left(\mathbf{I}_{n_r}\otimes\mathbf{P}^{\perp}_{\mathbf{S}^H_{\rm tr}}\right)^T\otimes\left(\mathbf{I}_{n_r}\otimes\mathbf{P}^{\perp}_{\mathbf{S}^H_{\rm tr}}\right)\mathcal{E}\left[\mathrm{vec}\left(\mathbf{e}^*\mathbf{e}^T\right)\mathrm{vec}^T\left(\mathbf{e}\mathbf{e}^H\right)\right]\right).
\end{aligned}$$

Terms $\mathrm{vec}\left(\mathbf{e}^*\mathbf{e}^H\right)$ and $\mathrm{vec}^T\left(\mathbf{e}\mathbf{e}^T\right)$ can be computed analytically, by writing down the exact form of each vector. Thus, having computed these terms, we take expectation, using [15, p. 508]

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

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

$$\mathcal{E}\left[\mathrm{vec}\left(\mathbf{e}^*\mathbf{e}^H\right)\mathrm{vec}^T\left(\mathbf{e}\mathbf{e}^T\right)\right] = \sigma_w^4\mathbf{I}_{n_r N_{\rm tr}} + \sigma_w^4\mathrm{vec}\left(\mathbf{I}_{n_r N_{\rm tr}}\right)\mathrm{vec}^H\left(\mathbf{I}_{n_r N_{\rm tr}}\right).$$

Thus, term b becomes

$$
\begin{aligned}
b &= \mathrm{tr}\left(\left(\mathbf{I}_{n_r}\otimes\mathbf{P}^{\perp}_{\mathbf{S}^H_{\mathrm{tr}}}\right)^T\otimes\left(\mathbf{I}_{n_r}\otimes\mathbf{P}^{\perp}_{\mathbf{S}^H_{\mathrm{tr}}}\right)\left(\sigma_w^4\mathbf{I}_{n_rN_{\mathrm{tr}}}+\sigma_w^4\mathrm{vec}\left(\mathbf{I}_{n_rN_{\mathrm{tr}}}\right)\mathrm{vec}^H\left(\mathbf{I}_{n_rN_{\mathrm{tr}}}\right)\right)\right)\\
&= \sigma_w^4\mathrm{tr}\left(\left(\mathbf{I}_{n_r}\otimes\mathbf{P}^{\perp}_{\mathbf{S}^H_{\mathrm{tr}}}\right)^T\otimes\left(\mathbf{I}_{n_r}\otimes\mathbf{P}^{\perp}_{\mathbf{S}^H_{\mathrm{tr}}}\right)\right)\\
&\quad+\sigma_w^4\mathrm{tr}\left(\mathrm{vec}^H\left(\mathbf{I}_{n_rN_{\mathrm{tr}}}\right)\left(\left(\mathbf{I}_{n_r}\otimes\mathbf{P}^{\perp}_{\mathbf{S}^H_{\mathrm{tr}}}\right)^T\otimes\left(\mathbf{I}_{n_r}\otimes\mathbf{P}^{\perp}_{\mathbf{S}^H_{\mathrm{tr}}}\right)\right)\mathrm{vec}\left(\mathbf{I}_{n_rN_{\mathrm{tr}}}\right)\right)\\
&\overset{(1),(4)}{=} \sigma_w^4 n_r^2\left(N_{\mathrm{tr}}-n_t\right)^2+\sigma_w^4\mathrm{tr}\left(\left(\mathbf{I}_{n_r}\otimes\mathbf{P}^{\perp}_{\mathbf{S}^H_{\mathrm{tr}}}\right)\left(\mathbf{I}_{n_r}\otimes\mathbf{P}^{\perp}_{\mathbf{S}^H_{\mathrm{tr}}}\right)\right)\\
&= \sigma_w^4 n_r^2\left(N_{\mathrm{tr}}-n_t\right)^2+\sigma_w^4\mathrm{tr}\left(\mathbf{I}_{n_r}\otimes\mathbf{P}^{\perp}_{\mathbf{S}^H_{\mathrm{tr}}}\right)\\
&= c^2\sigma_w^4+c\,\sigma_w^4.
\end{aligned}
$$

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

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

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

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

Appendix 2B

A useful approximation

In order to simplify term $\mathrm{tr}\left(\tilde{\mathbf{P}}_o^*\tilde{\mathbf{P}}_o^T\right)$ in the high SNR cases, i.e., $\alpha \ll \lambda_{n_r}(\mathbf{H}^H\mathbf{H})$, we notice that

$$
\mathrm{tr}\left(\tilde{\mathbf{P}}_o^*\tilde{\mathbf{P}}_o^T\right)=\mathrm{tr}\left(\tilde{\mathbf{P}}_o\tilde{\mathbf{P}}_o^H\right).
$$

Using the definitions of MMSE and $\tilde{\mathbf{P}}_o$ in (2.8) and (2.6), respectively, we get

$$
\begin{aligned}
\mathrm{MMSE} &= \mathrm{tr}\left(\mathbf{I}_{n_r}\right)-2\mathrm{tr}\left(\tilde{\mathbf{P}}_o\mathbf{H}\right)+\mathrm{tr}\left(\tilde{\mathbf{P}}_o\tilde{\mathbf{P}}_o^H\mathbf{H}^H\mathbf{H}\right)+\alpha\,\mathrm{tr}\left(\tilde{\mathbf{P}}_o\tilde{\mathbf{P}}_o^H\right)\\
&= \mathrm{tr}\left(\mathbf{I}_{n_r}\right)-2\mathrm{tr}\left(\mathbf{A}\mathbf{H}^H\mathbf{H}\right)+\mathrm{tr}\left(\mathbf{A}\mathbf{H}^H\mathbf{H}\mathbf{A}\mathbf{H}^H\mathbf{H}\right)+\alpha\,\mathrm{tr}\left(\tilde{\mathbf{P}}_o\tilde{\mathbf{P}}_o^H\right).
\end{aligned}
$$

Using approximations analogous to (2.29), we can write for the high SNR cases

$$
\mathrm{MMSE}\approx\mathrm{tr}\left(\mathbf{I}_{n_r}\right)-2\mathrm{tr}\left(\mathbf{I}_{n_r}\right)+\mathrm{tr}\left(\mathbf{I}_{n_r}\right)+\alpha\,\mathrm{tr}\left(\tilde{\mathbf{P}}_o\tilde{\mathbf{P}}_o^H\right)=\alpha\,\mathrm{tr}\left(\tilde{\mathbf{P}}_o\tilde{\mathbf{P}}_o^H\right).
$$

Finally, we get

$$
\mathrm{tr}\left(\tilde{\mathbf{P}}_o\tilde{\mathbf{P}}_o^H\right)\approx\frac{1}{\alpha}\,\mathrm{MMSE}. \tag{2.47}
$$

Chapter 3

On the sensitivity of the MIMO Tomlinson-Harashima precoder with respect to channel uncertainties

The multiple-input multiple-output Tomlinson-Harashima (MIMO-TH) precoder is a well-known structure that mitigates interstream interference in flat fading MIMO systems. The MIMO-TH filters are designed by assuming perfect channel state information (CSI) at both the transmitter and the receiver. However, in practice, channel estimates are available instead of the true channels. In this chapter, we assess the MIMO-TH performance degradation in the cases where the channel estimates are used as if they were the true channels. More specifically, we develop second-order and high-SNR approximations to the excess mean-square error induced by channel uncertainties, uncovering the factors that determine the MIMO-TH performance degradation in practice. Numerical experiments are in agreement with our theoretical developments.

3.1 Introduction

Interstream interference is a problem commonly encountered in MIMO communication systems. Many receiver structures mitigating interstream interference have been proposed in the literature, achieving various levels of performance with varying complexity. Prominent among them is the MIMO decision feedback equalizer (DFE). This non-linear receiver works efficiently but may suffer from error propagation. This disadvantage can be overcome by moving the feedback loop of the DFE to the transmitter, resulting in the so-called Tomlinson-Harashima precoder. In this chapter, we consider the TH precoder proposed in

Appendix E of [8]. The design of the TH precoder assumes perfect channel state information (CSI) at both the transmitter and the receiver; see, for example, [8], [14], [22], [29] and [30]. However, since CSI uncertainties *always* exist in real-world systems, due to, e.g., channel estimation errors, this assumption is *not* realistic. One way to proceed is to use the channel estimate as if it were the true channel; this is sometimes called the *mismatched* or *naive* approach. Another way is to exploit the statistical description of the channel uncertainties and develop *robust* designs; see, for example, [6], [23] and [25]. However, in all cases, the design of the MIMO-TH filters is based on *inexact* channel estimates and thus performance degradation is inevitable.

We consider a packet-based communication scenario where the channel may change (slowly) between successive packets. During each packet, the receiver estimates the channel and feeds its estimate back to the transmitter. This estimate is used for the design of the TH precoding filter that will be applied to the next packet. Thus, the TH precoding filter suffers from channel estimation errors (that occur at the receiver) and usually also suffers from mismatch due to channel time-variations, because the next packet may pass through a (slightly) different channel. Upon arrival of the packet, the receiver estimates the current channel (which is fed back to the transmitter) and proceeds to equalization and detection. Thus, the processing of each packet suffers from errors at both the transmitter and the receiver. Obviously, these errors degrade the MIMO-TH performance. We quantify this degradation by assessing the associated excess mean-square error. We show that the EMSE consists of two components that can be studied separately. The first component is due to the mismatch between the previous channel estimate and the current channel, while the second is due to the mismatch between the current channel and its estimate. We develop a second-order approximation to the EMSE which, in our experiments, is very accurate for SNR higher than 5 dB. However, this approximation is quite complicated and thus difficult to interpret. We focus on the high-SNR regime and derive a simple, informative, and tight (for sufficiently high SNR) EMSE upper bound, which uncovers the basic factors that determine the MIMO-TH performance degradation.

3.1.1 Notation

During our study, we shall develop first- and second-order approximations, with respect to channel uncertainties, as well as high-SNR approximations. In order to distinguish among these cases, we shall use the symbols $\simeq$, $\approx$, and $\approxeq$, respectively.

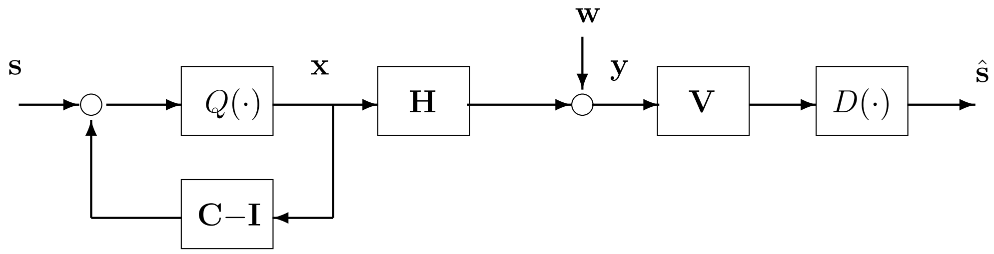

Fig. 3.1. System model.

3.2 The MIMO-TH Precoder

3.2.1 The system model

We consider the baseband-equivalent discrete-time frequency-flat MIMO system depicted in Fig. 3.1, with n_t transmit and n_r receive antennas (with $n_r \geq n_t$). The input-output relation of the channel is

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} \tag{3.1}$$

where $\mathbf{x}$ is the $n_t \times 1$ channel input vector, $\mathbf{H}$ is the $n_r \times n_t$ channel matrix, and $\mathbf{w}$ is the $n_r \times 1$ additive channel noise. The channel input symbols x_k, $k = 1, \ldots, n_t$, are successively generated from the data symbols s_k, $k = 1, \ldots, n_t$, as shown in Fig. 3.1, where the feedback loop consists of the feedback matrix $\mathbf{C}$ and the modulo operator $Q_M(\cdot)$. If $\mathbf{s}$ is a vector with independent identically distributed (i.i.d.) elements s_k (drawn from an M-QAM constellation), then it can be shown that $\mathbf{x}$ consists of uncorrelated random variables, with covariance matrix $\mathbf{R}_\mathbf{x} = \sigma_x^2 \mathbf{I}_{n_t}$, where $\sigma_x^2 \simeq 2M^2/12$ [8, p. 462]. The noise vector, $\mathbf{w}$, is assumed to be complex-valued circular Gaussian with covariance matrix $\mathbf{R}_\mathbf{w} = \sigma_w^2 \mathbf{I}_{n_r}$.

3.2.2 Optimal MMSE MIMO-TH

In this subsection, we briefly present the computation of the MMSE MIMO-TH filters, following the approach of the Appendix E of [8]. The error signal before the receiver's modulo operator (see Fig. 3.1) is

$$\mathbf{e} = \mathbf{V}\mathbf{y} - \mathbf{C}\mathbf{x} \tag{3.2}$$

and the mean-squared error is defined as

$$\text{mse}(\mathbf{C}, \mathbf{V}) \stackrel{\triangle}{=} \mathcal{E}\left[\|\mathbf{e}\|_2^2\right]. \tag{3.3}$$

The function mse(·) can be expressed as

$$\begin{aligned}\mathrm{mse}(\mathbf{C},\mathbf{V}) = \mathrm{tr}\big(\mathbf{V}(\sigma_x^2\mathbf{H}\mathbf{H}^H + \sigma_w^2\mathbf{I}_{n_r})\mathbf{V}^H\big) \\ - 2\,\sigma_x^2\mathrm{Re}\,\{\mathrm{tr}(\mathbf{V}\mathbf{H}\mathbf{C}^H)\} + \sigma_x^2\mathrm{tr}(\mathbf{C}\mathbf{C}^H).\end{aligned} \tag{3.4}$$

For any $\mathbf{C}$, minimization of $\mathrm{mse}(\mathbf{C},\mathbf{V})$ with respect to $\mathbf{V}$, yields

$$\mathbf{V} = \mathbf{C}\mathbf{H}^H(\mathbf{H}\mathbf{H}^H + \zeta\mathbf{I}_{n_r})^{-1}$$

where $\zeta \triangleq \sigma_w^2/\sigma_x^2$. By substituting this value to $\mathrm{mse}(\mathbf{C},\mathbf{V})$, we obtain

$$\mathrm{MSE}(\mathbf{C}) \triangleq \sigma_w^2\mathrm{tr}\big(\mathbf{C}(\mathbf{H}^H\mathbf{H} + \zeta\mathbf{I}_{n_t})^{-1}\mathbf{C}^H\big). \tag{3.5}$$

Minimization of $\mathrm{MSE}(\mathbf{C})$ with respect to $\mathbf{C}$, subject to the constraint that $\mathbf{C}$ be a lower triangular matrix with diagonal elements equal to 1, gives [1], [8]

$$\mathbf{C}_o = \mathbf{G}\mathbf{R} \tag{3.6}$$

where $\mathbf{R}$ is the lower triangular matrix satisfying the modified Cholesky factorization

$$\mathbf{H}^H\mathbf{H} + \zeta\mathbf{I}_{n_t} = \mathbf{R}^H\mathbf{R} \tag{3.7}$$

and

$$\mathbf{G} = \mathrm{diag}\left(r_{11}^{-1}, \ldots, r_{n_t n_t}^{-1}\right). \tag{3.8}$$

Using $\mathbf{C}_o$, we compute the optimal $\mathbf{V}$ as

$$\mathbf{V}_o = \mathbf{C}_o\mathbf{H}^H(\mathbf{H}\mathbf{H}^H + \zeta\mathbf{I}_{n_r})^{-1}. \tag{3.9}$$

Substituting $\mathbf{C}_o$ and $\mathbf{V}_o$ in (3.4), we obtain

$$\begin{aligned}\mathrm{MMSE} &\triangleq \mathrm{mse}(\mathbf{C}_o,\mathbf{V}_o) = \mathrm{MSE}(\mathbf{C}_o) \\ &= \sigma_w^2\,\mathrm{tr}\big(\mathbf{C}_o(\mathbf{H}^H\mathbf{H} + \zeta\mathbf{I}_{n_t})^{-1}\mathbf{C}_o^H\big).\end{aligned} \tag{3.10}$$

Using (3.6), we derive the alternative expression for the MMSE

$$\begin{aligned}\mathrm{MMSE} &= \sigma_w^2\,\mathrm{tr}(\mathbf{G}\mathbf{G}^H) \\ &= \sigma_w^2\,\mathrm{tr}\left(\mathrm{diag}\left(|r_{11}|^{-2}, \ldots, |r_{n_t n_t}|^{-2}\right)\right).\end{aligned} \tag{3.11}$$

3.2.3 Channel uncertainties

After the description of the ideal case, where we assumed that the channel $\mathbf{H}$ is perfectly known at the receiver and the transmitter, we proceed to a realistic scenario where both the transmitter and the receiver possess channel estimates. More specifically, we consider a frequency division duplex system and focus on the transmission of packet i. During the transmission of packet $(i-1)$, the receiver estimates the true channel, $\mathbf{H}_{i-1}$, as $\bar{\mathbf{H}}_{i-1}$. This estimate is communicated to the transmitter through a feedback channel and is used for precoding packet i. The true channel during the transmission of packet i, $\mathbf{H}_i$, is estimated at the receiver as $\bar{\mathbf{H}}_i$. Thus, in general, the channel estimate used at the transmitter for precoding packet i, $\bar{\mathbf{H}}_{i-1}$, suffers from *both* estimation errors and errors due to channel time-variations (other potential error sources are quantization errors and feedback channel errors - the following analysis can easily incorporate quantization errors, while the same does not happen for the feedback channel errors). On the other hand, the channel estimate at the receiver for packet i, $\bar{\mathbf{H}}_i$, suffers *only* from estimation errors.

In order to assess the associated performance degradation, we adopt the following statistical models for the channel inaccuracies.

1. *Channel estimation errors:* During each packet, we use training and estimate the channel using the maximum likelihood (ML) method, i.e., we assume that the channel is *constant* but *unknown*. The $n_t \times N_{\text{tr}}$ training block for packet i, $\mathbf{S}_i$, is multiplexed with the precoded information vectors (for example, it may be at the start of the packet) but is *not* precoded (we note that $\mathbf{S}_i$ may be the same for all i). If $\mathbf{Y}_i$ denotes the channel output corresponding to $\mathbf{S}_i$, then the ML estimate of $\mathbf{H}_i$ is [28, p. 174]
$$\bar{\mathbf{H}}_i = \mathbf{Y}_i \mathbf{S}_i^H \left(\mathbf{S}_i \mathbf{S}_i^H\right)^{-1}. \tag{3.12}$$
The channel estimation error is defined as
$$\Delta \mathbf{H}_{\text{est},i} \stackrel{\triangle}{=} \bar{\mathbf{H}}_i - \mathbf{H}_i. \tag{3.13}$$
Optimal channel estimates are obtained for semi-unitary training matrices, i.e., $\mathbf{S}_i \mathbf{S}_i^H = \sigma_x^2 N_{\text{tr}} \mathbf{I}_{n_t}$, and the optimal channel estimation error covariance matrix is [28, p. 175]
$$\boldsymbol{\Sigma}_{\text{est}} \stackrel{\triangle}{=} \mathcal{E}\left[\text{vec}(\Delta \mathbf{H}_{\text{est},i}) \text{vec}^H(\Delta \mathbf{H}_{\text{est},i})\right] = \frac{\sigma_w^2}{\sigma_x^2 N_{\text{tr}}} \mathbf{I}_{n_t n_r}. \tag{3.14}$$
We note that channel estimation errors associated with different packets are independent due to the assumed noise independence.

2. *Channel time-variations:* We adopt a commonly used statistical model describing the time evolution of the channel (the model is used only for analysis purposes and is not exploited during channel estimation). We denote with τ the time difference between two successive packets. We assume that $\{\mathbf{H}_i\}$ is a stationary matrix random process where, for all i, the elements of $\mathbf{H}_i$ are unit variance i.i.d. circular Gaussian random variables, i.e.,

$$\mathrm{vec}(\mathbf{H}_{i-1}), \mathrm{vec}(\mathbf{H}_i) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{n_t n_r}).$$

We assume that the channel coefficients are time-varying according to Jakes' model, with common maximum Doppler frequency f_d. Thus, $\mathbf{H}_{i-1}$ and $\mathbf{H}_i$ can be modeled as jointly Gaussian with cross-correlation [3, p. 93]

$$\mathcal{E}\left[\mathrm{vec}(\mathbf{H}_i)\mathrm{vec}^H(\mathbf{H}_{i-1})\right] = \rho_\tau \mathbf{I}_{n_t n_r} \tag{3.15}$$

where ρ_τ is the normalized correlation coefficient specified by the Jakes model, i.e., $\rho_\tau = J_0(2\pi f_d \tau)$, with $J_0(\cdot)$ the zeroth-order Bessel function of the first kind. If we define the channel error due to time-variations as

$$\Delta\mathbf{H}_{\mathrm{tv},i} \overset{\triangle}{=} \mathbf{H}_i - \mathbf{H}_{i-1} \tag{3.16}$$

then the associated error covariance matrix is independent of i and is given by

$$\mathbf{\Sigma}_{\mathrm{tv}} \overset{\triangle}{=} \mathcal{E}\left[\mathrm{vec}(\Delta\mathbf{H}_{\mathrm{tv},i})\mathrm{vec}^H(\Delta\mathbf{H}_{\mathrm{tv},i})\right] = 2\,(1-\rho_\tau)\mathbf{I}_{n_t n_r}. \tag{3.17}$$

Finally, we note that it is natural to assume that the errors due to channel time-variations are independent of the channel estimation errors because they are originating from independent phenomena, i.e., the first from the random channel evolution in time and the second from the additive channel noise.

In the sequel, for notational convenience, we neglect index i. We denote with $\mathbf{H}$ the true channel, with $\hat{\mathbf{H}}$ the channel estimate at the transmitter, and with $\tilde{\mathbf{H}}$ the channel estimate at the receiver. We define the mismatch at the transmitter and the receiver as

$$\Delta\mathbf{H}_{\mathrm{Tx}} \overset{\triangle}{=} \hat{\mathbf{H}} - \mathbf{H}, \quad \Delta\mathbf{H}_{\mathrm{Rx}} \overset{\triangle}{=} \tilde{\mathbf{H}} - \mathbf{H}. \tag{3.18}$$

It can be easily shown that $\Delta\mathbf{H}_{\mathrm{Tx}}$ and $\Delta\mathbf{H}_{\mathrm{Rx}}$ are zero mean with covariance matrices

$$\mathcal{E}\left[\mathrm{vec}(\Delta\mathbf{H}_{\mathrm{Tx}})\mathrm{vec}^H(\Delta\mathbf{H}_{\mathrm{Tx}})\right] = \mathbf{\Sigma}_{\mathrm{est}} + \mathbf{\Sigma}_{\mathrm{tv}} \tag{3.19}$$

and

$$\mathcal{E}\left[\mathrm{vec}(\Delta\mathbf{H}_{\mathrm{Rx}})\mathrm{vec}^H(\Delta\mathbf{H}_{\mathrm{Rx}})\right] = \mathbf{\Sigma}_{\mathrm{est}} \tag{3.20}$$

respectively. Furthermore, $\Delta\mathbf{H}_{\text{Rx}}$ and $\Delta\mathbf{H}_{\text{Tx}}$ are independent.

We close this subsection by mentioning that $\zeta = \sigma_w^2/\sigma_x^2$ is also required for the computation of the filters at both the transmitter and the receiver. We assume that $\sigma_x^2 = 2M^2/12$ is known at both sides and σ_w^2 is estimated at the receiver (for more details we refer to [28, Sec. 9.4]); then, the estimate is sent to the transmitter through a feedback channel. It turns out that the variance of the noise variance estimation error is $O(\sigma_w^4)$ and thus, for sufficiently high SNR, the error in ζ is negligible compared with the channel estimation error. Thus, we assume that ζ is perfectly known.

3.2.4 MIMO-TH: the mismatched approach

In this subsection, we follow the mismatched approach and compute the MIMO-TH filters using the channel estimates $\hat{\mathbf{H}}$ and $\tilde{\mathbf{H}}$ as if they were the true channel $\mathbf{H}$. The transmitter, based on (3.6), computes and uses

$$\hat{\mathbf{C}} = \hat{\mathbf{G}}\hat{\mathbf{R}} \tag{3.21}$$

where $\hat{\mathbf{R}}^H\hat{\mathbf{R}} = \hat{\mathbf{H}}^H\hat{\mathbf{H}} + \zeta\mathbf{I}_{n_t}$ and $\hat{\mathbf{G}} = \text{diag}(\hat{r}_{11}^{-1}, \ldots, \hat{r}_{n_t n_t}^{-1})$. We note that since the receiver knows $\hat{\mathbf{H}}$, it can compute and use $\hat{\mathbf{C}}$.

Given that the transmitter uses $\hat{\mathbf{C}}$, the input estimation error becomes

$$\hat{\mathbf{e}} = \mathbf{V}\hat{\mathbf{y}} - \hat{\mathbf{C}}\hat{\mathbf{x}} \tag{3.22}$$

where $\hat{\mathbf{x}}$ is the channel input produced by the feedback filter $\hat{\mathbf{C}}$ and $\hat{\mathbf{y}} = \mathbf{H}\hat{\mathbf{x}} + \mathbf{w}$. In order to compute the "optimal" filter at the receiver, we follow steps analogous to those of subsection 3.2.2. Then, it can be shown that the filter that minimizes $\mathcal{E}\left[\|\hat{\mathbf{e}}\|_2^2\right]$ is

$$\hat{\mathbf{V}} = \hat{\mathbf{C}}\mathbf{H}^H(\mathbf{H}\mathbf{H}^H + \zeta\mathbf{I}_{n_r})^{-1}. \tag{3.23}$$

The best the receiver can do is to use its current channel estimate $\tilde{\mathbf{H}}$ as if it were $\mathbf{H}$ and compute[1]

$$\hat{\hat{\mathbf{V}}} = \hat{\mathbf{C}}\tilde{\mathbf{H}}^H(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H + \zeta\mathbf{I}_{n_r})^{-1}. \tag{3.24}$$

Using (3.21) and (3.24) in (3.4), we obtain that the MSE achieved by the mismatched approach is

$$\begin{aligned} \text{mse}(\hat{\mathbf{C}}, \hat{\hat{\mathbf{V}}}) = {} & \sigma_x^2 \text{tr}\big(\hat{\hat{\mathbf{V}}}(\mathbf{H}\mathbf{H}^H + \zeta\mathbf{I}_{n_r})\hat{\hat{\mathbf{V}}}^{\mathbf{H}}\big) \\ & - 2\,\sigma_x^2 \text{Re}\left\{\text{tr}(\hat{\hat{\mathbf{V}}}\mathbf{H}\hat{\mathbf{C}}^H)\right\} + \sigma_x^2 \text{tr}(\hat{\mathbf{C}}\hat{\mathbf{C}}^H). \end{aligned} \tag{3.25}$$

[1] It can be proved that if the receiver uses $\hat{\mathbf{H}}$ instead of $\tilde{\mathbf{H}}$, then the performance degrades dramatically. The proof can be made available by the authors upon request.

The EMSE is defined as

$$\text{EMSE}(\hat{\mathbf{C}}, \hat{\tilde{\mathbf{V}}}) \triangleq \mathcal{E}\left[\text{mse}\big(\hat{\mathbf{C}}, \hat{\tilde{\mathbf{V}}}\big) - \text{mse}\big(\mathbf{C}_o, \mathbf{V}_o\big)\right] \tag{3.26}$$

where the expectation is with respect to the channel uncertainties. Our main task in the sequel is to quantify $\text{EMSE}(\hat{\mathbf{C}}, \hat{\tilde{\mathbf{V}}})$.

3.3 EMSE - Second-order analysis

In this section, we develop a second-order approximation to $\text{EMSE}(\hat{\mathbf{C}}, \hat{\tilde{\mathbf{V}}})$, with respect to channel uncertainties.

We start by considering two unrealistic and, thus, seemingly, useless cases. Their usefulness will become evident shortly.

1. *Channel uncertainties* only *at the transmitter:* We assume that the transmitter possesses the channel estimate $\hat{\mathbf{H}}$ while the receiver has perfect CSI. Thus, the transmitter and the receiver use filters $\hat{\mathbf{C}}$ and $\hat{\mathbf{V}}$, defined in (3.21) and (3.23), respectively. Substituting these values into (3.4), we obtain that the associated MSE is

$$\begin{aligned} \text{MSE}(\hat{\mathbf{C}}) &\triangleq \text{mse}(\hat{\mathbf{C}}, \hat{\mathbf{V}}) \\ &= \sigma_w^2 \text{tr}\left(\hat{\mathbf{C}}(\mathbf{H}^H\mathbf{H} + \zeta \mathbf{I}_{n_t})^{-1}\hat{\mathbf{C}}^H\right). \end{aligned} \tag{3.27}$$

The corresponding EMSE is defined as

$$\text{EMSE}(\hat{\mathbf{C}}) \triangleq \mathcal{E}\left[\text{MSE}\big(\hat{\mathbf{C}}\big) - \text{MSE}\big(\mathbf{C}_o\big)\right]. \tag{3.28}$$

2. *Channel uncertainties* only *at the receiver:* We assume that the transmitter has perfect CSI and the receiver possesses the channel estimate $\tilde{\mathbf{H}}$. Thus, the transmitter uses $\mathbf{C}_o$ defined in (3.6), while the receive filter, denoted as $\tilde{\mathbf{V}}$, is computed using the optimal transmit filter $\mathbf{C}_o$ and the channel estimate $\tilde{\mathbf{H}}$, as

$$\tilde{\mathbf{V}} = \mathbf{C}_o\tilde{\mathbf{H}}^H(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H + \zeta\mathbf{I}_{n_r})^{-1}. \tag{3.29}$$

Substituting (3.6) and (3.29) into (3.4), we obtain that the associated MSE is

$$\begin{aligned} \text{MSE}_o(\tilde{\mathbf{V}}) &\triangleq \text{mse}(\mathbf{C}_o, \tilde{\mathbf{V}}) \\ &= \sigma_x^2 \text{tr}\big(\tilde{\mathbf{V}}(\mathbf{H}\mathbf{H}^H + \zeta\mathbf{I}_{n_r})\tilde{\mathbf{V}}^H\big) \\ &\quad - 2\sigma_x^2 \text{Re}\left\{\text{tr}(\tilde{\mathbf{V}}\mathbf{H}\mathbf{C}_o^H)\right\} + \sigma_x^2\text{tr}(\mathbf{C}_o\mathbf{C}_o^H). \end{aligned} \tag{3.30}$$

The corresponding EMSE is defined as

$$\mathrm{EMSE}(\tilde{\mathbf{V}}) \stackrel{\triangle}{=} \mathcal{E}\left[\mathrm{MSE}_o\big(\tilde{\mathbf{V}}\big) - \mathrm{MSE}_o\big(\mathbf{V}_o\big)\right]. \tag{3.31}$$

The next result shows that $\mathrm{EMSE}(\hat{\mathbf{C}}, \hat{\tilde{\mathbf{V}}})$ can be decomposed into two terms that correspond to these unrealistic cases.

Proposition 1. The EMSE induced by channel inaccuracies at both the transmitter and the receiver can be approximated as

$$\mathrm{EMSE}(\hat{\mathbf{C}}, \hat{\tilde{\mathbf{V}}}) \approx \mathrm{EMSE}(\hat{\mathbf{C}}) + \mathrm{EMSE}(\tilde{\mathbf{V}}). \tag{3.32}$$

Proof: The proof is provided in Appendix 3A and is based on the fact that the channel errors $\Delta\mathbf{H}_{\mathrm{Tx}}$ and $\Delta\mathbf{H}_{\mathrm{Rx}}$ are zero-mean and independent. □

In the sequel, we develop second-order approximations to $\mathrm{EMSE}(\hat{\mathbf{C}})$ and $\mathrm{EMSE}(\tilde{\mathbf{V}})$.

3.3.1 Channel uncertainties *only* at the transmitter

Using a Taylor expansion of the function $\mathrm{MSE}(\mathbf{C})$ in (3.27) around $\mathbf{C}_o$, we obtain

$$\mathrm{MSE}(\hat{\mathbf{C}}) = \mathrm{MSE}(\mathbf{C}_o) + \mathrm{tr}\left(\Delta\mathbf{C}\,\mathrm{MSE}''(\mathbf{C}_o)\,\Delta\mathbf{C}^H\right) \tag{3.33}$$

where $\Delta\mathbf{C} \stackrel{\triangle}{=} \hat{\mathbf{C}} - \mathbf{C}_o$ and $\mathrm{MSE}''(\mathbf{C}_o)$ is the second derivative of $\mathrm{MSE}(\mathbf{C})$ evaluated at $\mathbf{C}_o$[2]. It can be shown that [10]

$$\mathrm{MSE}''(\mathbf{C}_o) = \sigma_w^2\,(\mathbf{H}^H\mathbf{H} + \zeta\mathbf{I}_{n_t})^{-1}. \tag{3.34}$$

Using (3.28), (3.33), and (3.34), we obtain

$$\begin{aligned}\mathrm{EMSE}(\hat{\mathbf{C}}) &= \mathcal{E}\left[\mathrm{tr}\left(\Delta\mathbf{C}\,\mathrm{MSE}''(\mathbf{C}_o)\Delta\mathbf{C}^H\right)\right] \\ &= \sigma_w^2\,\mathcal{E}\left[\mathrm{tr}\left(\Delta\mathbf{C}\mathbf{A}^{-1}\Delta\mathbf{C}^H\right)\right]\end{aligned} \tag{3.35}$$

where

$$\mathbf{A} \stackrel{\triangle}{=} \mathbf{H}^H\mathbf{H} + \zeta\mathbf{I}_{n_t}. \tag{3.36}$$

The following lemma gives a second-order approximation to $\mathrm{EMSE}(\hat{\mathbf{C}})$.

Lemma 1. A second-order approximation to $\mathrm{EMSE}(\hat{\mathbf{C}})$ is given by

$$\mathrm{EMSE}(\hat{\mathbf{C}}) \approx \sum_{i=1}^{3} \mathbf{B}_i \tag{3.37}$$

[2]The first derivative of $\mathrm{MSE}(\mathbf{C})$ at $\mathbf{C}_o$ vanishes because $\mathbf{C}_o$ is the minimizer of $\mathrm{MSE}(\mathbf{C})$.

where terms $\mathbf{B}_i$ are defined as

$$\mathbf{B}_1 \triangleq \left(\alpha + \frac{\sigma_w^2}{\sigma_x^2 N_{\text{tr}}}\right) \sigma_w^2 \text{tr}\left(\left(\mathbf{I}_{n_t} \otimes \mathbf{G}^2\right) \mathbf{L}^T\mathbf{L}\left(\mathbf{R}^{-T}\mathbf{R}^{-*} \otimes \mathbf{R}^{-H}\mathbf{H}^H\mathbf{H}\mathbf{R}^{-1}\right)\mathbf{L}^T\mathbf{L}\right) \quad (3.38)$$

$$\mathbf{B}_2 \triangleq -2\left(\alpha + \frac{\sigma_w^2}{\sigma_x^2 N_{\text{tr}}}\right) \sigma_w^2 \text{tr}\left((\mathbf{I}_{n_t} \otimes \mathbf{G}^2)\mathbf{P}(\mathbf{R}^{-T}\mathbf{R}^{-*} \otimes \mathbf{R}^{-H}\mathbf{H}^H\mathbf{H}\mathbf{R}^{-1})\mathbf{P}\right) \quad (3.39)$$

$$\mathbf{B}_3 \triangleq \left(\alpha + \frac{\sigma_w^2}{\sigma_x^2 N_{\text{tr}}}\right) \sigma_w^2 \text{tr}\left((\mathbf{I}_{n_t} \otimes \mathbf{G}^2)\mathbf{L}^T\mathbf{L}(\mathbf{R}^{-T}\mathbf{H}^T\mathbf{H}^*\mathbf{R}^{-*} \otimes \mathbf{R}^{-H}\mathbf{R}^{-1})\mathbf{L}^T\mathbf{L}\right). \quad (3.40)$$

In these expressions, $\mathbf{L}$ is the elimination matrix and $\mathbf{P} \triangleq \mathbf{L}^T\mathbf{L}\mathbf{K}\mathbf{L}^T\mathbf{L}$, where $\mathbf{K}$ is the commutation matrix. The scalar α is defined as $\alpha \triangleq 2(1-\rho_\tau)$ (see (3.17)).

Proof: The proof is provided in Appendix 3B. □

3.3.2 Channel uncertainties *only* at the receiver

Using a Taylor expansion of the function $\text{MSE}_o(\mathbf{V})$ in (3.30) around $\mathbf{V}_o$, we obtain

$$\text{MSE}_o(\tilde{\mathbf{V}}) = \text{MSE}_o(\mathbf{V}_o) + \text{tr}\left(\Delta\mathbf{V}\text{MSE}_o''(\mathbf{V}_o)\Delta\mathbf{V}^H\right) \quad (3.41)$$

where $\Delta\mathbf{V} \triangleq \tilde{\mathbf{V}} - \mathbf{V}_o$, and $\text{MSE}_o''(\mathbf{V}_o)$ is the second derivative of $\text{MSE}_o(\mathbf{V})$ evaluated at $\mathbf{V}_o$[3]. It can be shown that [10]

$$\text{MSE}_o''(\mathbf{V}_o) = \sigma_x^2(\mathbf{H}\mathbf{H}^H + \zeta\mathbf{I}_{n_r}). \quad (3.42)$$

Using (3.31), (3.30), and (3.42), we obtain

$$\begin{aligned}\text{EMSE}(\tilde{\mathbf{V}}) &= \mathcal{E}\left[\text{tr}\left(\Delta\mathbf{V}\,\text{MSE}_o''(\mathbf{V}_o)\Delta\mathbf{V}^H\right)\right] \\ &= \sigma_x^2\,\mathcal{E}\left[\text{tr}\left(\Delta\mathbf{V}\,\mathbf{B}\Delta\mathbf{V}^H\right)\right]\end{aligned} \quad (3.43)$$

where

$$\mathbf{B} \triangleq \mathbf{H}\mathbf{H}^H + \zeta\mathbf{I}_{n_r}. \quad (3.44)$$

The following lemma gives a second-order approximation to $\text{EMSE}(\tilde{\mathbf{V}})$.

Lemma 2. A second-order approximation to the $\text{EMSE}(\tilde{\mathbf{V}})$ is given by

$$\text{EMSE}(\tilde{\mathbf{V}}) \approx \mathbf{T}_1 + \mathbf{T}_2 \quad (3.45)$$

where

$$\mathbf{T}_1 \triangleq \frac{\sigma_w^2}{N_{\text{tr}}}\,\text{tr}\left(\mathbf{H}^T\mathbf{B}^{-T}\mathbf{H}^*\right)\text{tr}\left(\mathbf{V}_o\mathbf{V}_o^H\right) \quad (3.46)$$

[3]The first derivative of $\text{MSE}_o(\mathbf{V})$ at $\mathbf{V}_o$ is zero because $\mathbf{V}_o$ minimizes the function $\text{MSE}_o(\mathbf{V})$.

and

$$\mathbf{T}_2 \triangleq \frac{\sigma_w^2}{N_{\text{tr}}} \operatorname{tr}\left((\mathbf{C}_o - \mathbf{V}_o\mathbf{H})^H(\mathbf{C}_o - \mathbf{V}_o\mathbf{H})\right) \operatorname{tr}\left(\mathbf{B}^{-T}\right). \tag{3.47}$$

Proof: The proof is provided in Appendix 3C. □

Substituting (3.37) and (3.45) into (3.32), we obtain a second-order approximation to the EMSE induced by channel uncertainties at both the transmitter and the receiver. Admittedly, this approximation is complicated and difficult to interpret. In the sequel, we shall develop simple and insightful high-SNR expressions.

3.4 EMSE - High-SNR approximations

In this section, we focus on the high-SNR regime and we derive a simple upper bound to EMSE($\hat{\mathbf{C}}$) and a simple approximation to EMSE($\tilde{\mathbf{V}}$). Putting these expressions together, we obtain a simple high-SNR EMSE upper bound for the mismatched MIMO-TH precoder. Finally, we average over the channel statistics and obtain a simple high-SNR upper bound for the expected value of the EMSE to MMSE ratio.

High-SNR regime means "small" σ_w^2. Our results will be derived either by ignoring $O(\sigma_w^2)$ terms compared with $O(1)$ terms or by ignoring $O(\sigma_w^4)$ terms compared with $O(\sigma_w^2)$ terms. We proceed by presenting some high-SNR approximations that will be useful in the sequel.

Using the definition of matrix $\mathbf{R}$ in (3.7), it can be shown that for high SNR

$$\mathbf{R}^{-H}\mathbf{H}^H\mathbf{H}\mathbf{R}^{-1} \approxeq \mathbf{I}_{n_t} \tag{3.48}$$

and

$$\operatorname{tr}(\mathbf{R}^{-T}\mathbf{R}^{-*}) \approxeq \operatorname{tr}((\mathbf{H}^H\mathbf{H})^{-1}). \tag{3.49}$$

Furthermore (the proof is provided in Appendix 3D)

$$\operatorname{tr}\left(\mathbf{V}_o\mathbf{V}_o^H\right) \approxeq \frac{1}{\sigma_w^2}\,\text{MMSE}. \tag{3.50}$$

Using the matrix inversion lemma [9], it can be shown that

$$\mathbf{H}^H(\mathbf{H}\mathbf{H}^H + \zeta\mathbf{I}_{n_r})^{-1} = (\mathbf{H}^H\mathbf{H} + \zeta\mathbf{I}_{n_t})^{-1}\mathbf{H}^H. \tag{3.51}$$

Then, using (3.44), (3.51) and the high-SNR assumption, we get

$$\begin{aligned}\operatorname{tr}\left(\mathbf{H}^T\mathbf{B}^{-T}\mathbf{H}^*\right) &= \operatorname{tr}\left(\mathbf{H}^T(\mathbf{H}^*\mathbf{H}^T + \zeta\mathbf{I}_{n_r})^{-1}\mathbf{H}^*\right) \\ &= \operatorname{tr}\left((\mathbf{H}^T\mathbf{H}^* + \zeta\mathbf{I}_{n_t})^{-1}\mathbf{H}^T\mathbf{H}^*\right) \approxeq \operatorname{tr}\left(\mathbf{I}_{n_t}\right).\end{aligned} \tag{3.52}$$

Finally, using (3.9) and (3.51), we can write matrix $\mathbf{C}_o - \mathbf{V}_o\mathbf{H}$ as

$$\begin{aligned}\mathbf{C}_o - \mathbf{V}_o\mathbf{H} &= \mathbf{C}_o\left(\mathbf{I}_{n_t} - \mathbf{H}^H(\mathbf{H}\mathbf{H}^H + \zeta\mathbf{I}_{n_r})^{-1}\mathbf{H}\right) \\ &= \mathbf{C}_o\left(\mathbf{I}_{n_t} - (\mathbf{H}^H\mathbf{H} + \zeta\mathbf{I}_{n_t})^{-1}\mathbf{H}^H\mathbf{H}\right)\end{aligned} \tag{3.53}$$

and for high SNR

$$\mathbf{C}_o - \mathbf{V}_o\mathbf{H} \approxeq \mathbf{O}_{n_t}. \tag{3.54}$$

3.4.1 High SNR - channel uncertainties *only* at the transmitter

Lemma 3: In the high-SNR regime, the following approximate inequality holds

$$\text{EMSE}(\hat{\mathbf{C}}) \lessapprox 2\,(1-\rho_\tau)\,(n_t-1)\text{tr}\left((\mathbf{H}^H\mathbf{H})^{-1}\right)\ \text{MMSE}. \tag{3.55}$$

Proof: Using (3.48) in (3.38) and ignoring the term that involves σ_w^4, we obtain

$$\begin{aligned}\mathbf{B}_1 &\approxeq \alpha\,\sigma_w^2\text{tr}\left(\left(\mathbf{I}_{n_t}\otimes\mathbf{G}^2\right)\mathbf{L}^T\mathbf{L}\left(\mathbf{R}^{-T}\mathbf{R}^{-*}\otimes\mathbf{I}_{n_t}\right)\mathbf{L}^T\mathbf{L}\right) \\ &\overset{(a)}{\leq} \alpha\,\sigma_w^2\,\text{tr}\big((\mathbf{I}_{n_t}\otimes\mathbf{G}^2)(\mathbf{R}^{-T}\mathbf{R}^{-*}\otimes\mathbf{I}_{n_t})\big) \\ &\overset{(3)}{=} \alpha\,\sigma_w^2\,\text{tr}(\mathbf{R}^{-T}\mathbf{R}^{-*}\otimes\mathbf{G}^2) \\ &\overset{(4)}{=} \alpha\,\sigma_w^2\,\text{tr}(\mathbf{R}^{-T}\mathbf{R}^{-*})\,\text{tr}(\mathbf{G}^2)\end{aligned} \tag{3.56}$$

where at point (a) we used the structure of the elimination and the commutation matrices and the fact that matrices $\mathbf{G}^2$ and $\mathbf{R}^{-T}\mathbf{R}^{-*}$ have positive diagonal elements.

Using (3.48) in (3.39) and ignoring the term that involves σ_w^4, we obtain

$$\begin{aligned}\mathbf{B}_2 &\approxeq -2\alpha\,\sigma_w^2\text{tr}\left((\mathbf{I}_{n_t}\otimes\mathbf{G}^2)\mathbf{P}(\mathbf{R}^{-T}\mathbf{R}^{-*}\otimes\mathbf{I}_{n_t})\mathbf{P}\right) \\ &= -2\,\alpha\,\sigma_w^2\,\text{tr}\left(\mathbf{G}^2\mathbf{R}^{-T}\mathbf{R}^{-*}\right).\end{aligned} \tag{3.57}$$

The proof of the last equality is provided in Appendix 3E for the $n_r \times 2$ case (the generalization is easy).

Finally, using (3.48) in (3.40) and ignoring the term involving σ_w^4, we obtain

$$\begin{aligned}\mathbf{B}_3 &\approxeq \alpha\,\sigma_w^2\text{tr}\left((\mathbf{I}_{n_t}\otimes\mathbf{G}^2)\mathbf{L}^T\mathbf{L}(\mathbf{I}_{n_t}\otimes\mathbf{R}^{-H}\mathbf{R}^{-1})\mathbf{L}^T\mathbf{L}\right) \\ &\overset{(b)}{\leq} \alpha\,\sigma_w^2\,\text{tr}\big((\mathbf{I}_{n_t}\otimes\mathbf{G}^2)(\mathbf{I}_{n_t}\otimes\mathbf{R}^{-H}\mathbf{R}^{-1})\big) \\ &\overset{(3)}{=} \alpha\,\sigma_w^2\,\text{tr}(\mathbf{I}_{n_t}\otimes\mathbf{G}^2\mathbf{R}^{-H}\mathbf{R}^{-1}) \\ &= n_t\,\alpha\,\sigma_w^2\,\text{tr}(\mathbf{G}^2\mathbf{R}^{-H}\mathbf{R}^{-1}) \\ &= n_t\,\alpha\,\sigma_w^2\,\text{tr}(\mathbf{G}^2\mathbf{R}^{-T}\mathbf{R}^{-*})\end{aligned} \tag{3.58}$$

where at point (b) we used the structure of the elimination and the commutation matrices and the fact that matrices $\mathbf{G}^2$ and $\mathbf{R}^{-H}\mathbf{R}^{-1}$ have positive diagonal elements.

Combining expressions (3.37) and (3.56)–(3.58), we obtain

$$\begin{aligned}\mathrm{EMSE}(\hat{\mathbf{C}}) &\lessapprox \alpha\,\sigma_w^2\,\mathrm{tr}(\mathbf{R}^{-T}\mathbf{R}^{-*})\,\mathrm{tr}(\mathbf{G}^2)\\ &\quad + (n_t-2)\,\alpha\,\sigma_w^2\,\mathrm{tr}(\mathbf{G}^2\mathbf{R}^{-T}\mathbf{R}^{-*})\\ &\overset{(6)}{\leq} \alpha\,\sigma_w^2\,(n_t-1)\,\mathrm{tr}(\mathbf{R}^{-T}\mathbf{R}^{-*})\,\mathrm{tr}(\mathbf{G}^2).\end{aligned} \tag{3.59}$$

Using (3.59), (3.11) and (3.49), we conclude with the following bound

$$\mathrm{EMSE}(\hat{\mathbf{C}}) \lessapprox \alpha\,(n_t-1)\mathrm{tr}((\mathbf{H}^H\mathbf{H})^{-1})\,\mathrm{MMSE}. \tag{3.60}$$

Finally, recalling the definition of α as $\alpha := 2\,(1-\rho_\mathcal{T})$, we obtain (3.55) to prove Lemma 3.
□

3.4.2 High SNR - channel uncertainties *only* at the receiver

Lemma 4: In the high-SNR regime, the following approximation holds

$$\mathrm{EMSE}(\tilde{\mathbf{V}}) \approxeq \frac{n_t}{N_{\mathrm{tr}}}\,\mathrm{MMSE}. \tag{3.61}$$

Proof: Starting with $\mathbf{T}_1$ in (3.46) and using (3.50) and (3.52), we obtain

$$\begin{aligned}\mathbf{T}_1 &\approxeq \frac{\sigma_w^2}{N_{\mathrm{tr}}}\,\mathrm{tr}\left(\mathbf{V}_o\mathbf{V}_o^H\right)\mathrm{tr}\left(\mathbf{I}_{n_t}\right)\\ &\approxeq \frac{n_t}{N_{\mathrm{tr}}}\,\mathrm{MMSE}.\end{aligned} \tag{3.62}$$

Using (3.54) in (3.47) we get

$$\mathbf{T}_2 \approxeq 0. \tag{3.63}$$

We conclude that, for sufficiently high SNR, term $\mathbf{T}_2$ is negligible compared with $\mathbf{T}_1$. Combining expressions (3.45), (3.62) and (3.63), we obtain (3.61) to prove Lemma 4. □

3.4.3 High SNR - channel uncertainties at both the transmitter and the receiver

Proposition 2: The high-SNR MIMO-TH EMSE induced by channel uncertainties at both the transmitter and the receiver is upper bounded as

$$\begin{aligned}\mathrm{EMSE}(\hat{\mathbf{C}},\hat{\tilde{\mathbf{V}}}) &\lessapprox 2\,(1-\rho_\mathcal{T})\,(n_t-1)\mathrm{tr}((\mathbf{H}^H\mathbf{H})^{-1})\,\mathrm{MMSE}\\ &\quad + \frac{n_t}{N_{\mathrm{tr}}}\,\mathrm{MMSE}.\end{aligned} \tag{3.64}$$

Proof: The proof requires only the substitution of (3.55) and (3.61) into (3.32). □

We observe that the EMSE is upper bounded by an expression proportional to the MMSE. The proportionality factor is determined by the system parameters n_t and N_{tr}, the channel correlation coefficient $\rho_\mathcal{T}$ and the conditioning of the channel matrix through $\text{tr}((\mathbf{H}^H\mathbf{H})^{-1})$.

In the simulations section, we shall observe that this high-SNR bound is in many cases tight because the EMSE due to the channel inaccuracies only at the receiver dominates the EMSE due to channel inaccuracies only at the transmitter.

3.4.4 High SNR - averaging over the channels

In this subsection, we compute the average, over the channels, of the EMSE to MMSE ratio.

Proposition 3: Taking expectation with respect to the channels in (3.64), we obtain the following bound for the average EMSE to MMSE ratio, for $n_r > n_t$,

$$\mathcal{E}\left[\frac{\text{EMSE}(\hat{\mathbf{C}}, \hat{\hat{\mathbf{V}}})}{\text{MMSE}}\right] \lessapprox 2\left(1-\rho_\mathcal{T}\right)\frac{n_t(n_t-1)}{n_r-n_t} + \frac{n_t}{N_{\text{tr}}}. \tag{3.65}$$

Proof: Bound (3.64) can be written as

$$\frac{\text{EMSE}(\hat{\mathbf{C}}, \hat{\hat{\mathbf{V}}})}{\text{MMSE}} \lessapprox 2\left(1-\rho_\mathcal{T}\right)(n_t-1)\text{tr}((\mathbf{H}^H\mathbf{H})^{-1}) + \frac{n_t}{N_{\text{tr}}}. \tag{3.66}$$

If we take expectation with respect to the channel, we get

$$\mathcal{E}\left[\frac{\text{EMSE}(\hat{\mathbf{C}}, \hat{\hat{\mathbf{V}}})}{\text{MMSE}}\right] \lessapprox 2\left(1-\rho_\mathcal{T}\right)(n_t-1)\mathcal{E}\left[\text{tr}((\mathbf{H}^H\mathbf{H})^{-1})\right] + \frac{n_t}{N_{\text{tr}}}. \tag{3.67}$$

It can be shown that if the elements of $\mathbf{H}$ are zero-mean, unit variance i.i.d. circular complex Gaussian random variables and $n_r > n_t$, then [19]

$$\mathcal{E}\left[\text{tr}((\mathbf{H}^H\mathbf{H})^{-1})\right] = \frac{n_t}{n_r-n_t}. \tag{3.68}$$

Substituting (3.68) in (3.67), we prove (3.65). □

We observe that the average EMSE to MMSE ratio is upper bounded by an expression which depends on the system parameters n_t, n_r and N_{tr}, and the channel correlation coefficient $\rho_\mathcal{T}$.

Table II

Elements of channel matrix **H**

0.2877 + 0.2097*j	1.5537 - 1.0653*j	0.3450 - 0.5177*j	-0.8714 - 0.2760*j
-0.2433 - 0.5179*j	-0.8435 - 0.2245*j	0.1448 + 1.0146*j	-0.2163 - 0.5226*j
-1.0170 - 0.5243*j	0.5028 - 0.4757*j	-1.1749 + 0.5322*j	1.0848 + 0.5731*j
-0.7636 - 0.6970*j	-0.6756 + 0.3384*j	0.4666 - 0.0437*j	0.2691 + 1.0737*j
0.4901 + 0.0910*j	0.0804 + 0.0937*j	-0.5181 - 0.7709*j	-0.4012 - 0.0189*j
-0.5236 + 0.9038*j	0.2247 + 0.2181*j	0.6343 - 0.8332*j	-0.5841 + 0.1868*j

3.5 Simulation Results

In the first part of our experiments, we illustrate Propositions 1 and 2 using a *specific* channel realization, by taking averages over the channel uncertainties. More specifically, we consider a system with $n_t = 4$ transmit antennas and $n_r = 6$ receive antennas and channel matrix $\mathbf{H}$ with elements given in Table II[4]. The noise is spatially and temporally white, circularly symmetric complex Gaussian with variance σ_w^2. The input symbols are i.i.d., drawn from a 4-QAM constellation. We assume that the training block consists of $N_{\text{tr}} = 10$ columns. We set the channel correlation coefficient equal to $\rho_\mathcal{T} = 0.99$. We define the SNR as the ratio of the total receive power to the total noise power

$$\text{SNR} \triangleq \frac{\mathcal{E}_{\mathbf{x}}\left[\text{tr}(\mathbf{H}\mathbf{x}\mathbf{x}^H\mathbf{H}^H)\right]}{\mathcal{E}_{\mathbf{w}}\left[\text{tr}(\mathbf{w}\mathbf{w}^H)\right]} = \frac{\sigma_x^2\|\mathbf{H}\|_F^2}{n_r\sigma_w^2}. \tag{3.69}$$

In Fig. 3.2, we plot the MMSE (3.10), the average of the MSEs for the case of channel inaccuracies only at the transmitter (the average is over $\Delta\mathbf{H}_{\text{Tx}}$), the average of the MSEs for the case of channel inaccuracies only at the receiver (the average is over $\Delta\mathbf{H}_{\text{Rx}}$), and finally the average of the MSEs for inaccuracies at both the transmitter and the receiver. We observe that the EMSE component due to $\Delta\mathbf{H}_{\text{Rx}}$ is significantly larger than that due to $\Delta\mathbf{H}_{\text{Tx}}$. This observation is in agreement with our theoretical results because the high-SNR approximations (3.55) and (3.61) indicate that both EMSEs are proportional to the MMSE, with the proportionality factor in (3.61) being larger than the one in (3.55), as long as the channel matrix, $\mathbf{H}$, is well conditioned and the channel correlation coefficient, $\rho_\mathcal{T}$, is relatively large. An explanation of this phenomenon might be the fact that in the

[4]Analogous results have been obtained in extended simulations with other channel realizations.

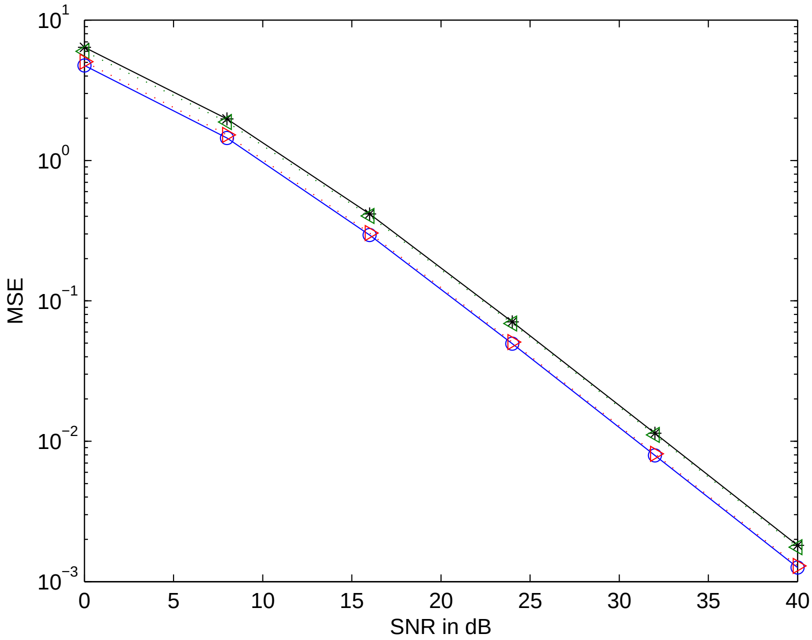

Fig. 3.2. MMSE using the true channel ('-o'), expectation of the MSEs for channel inaccuracies only at transmitter ('▷'), expectation of the MSEs for channel inaccuracies only at the receiver ('◁') and expectation of the MSEs for channel inaccuracies at both the transmitter and the receiver ('-*').

first case the receiver is optimized by taking into account the channel uncertainties at the transmitter while something analogous does not happen in the latter case.

In Fig. 3.3, we present the experimentally computed EMSE, the theoretical second-order approximation as the sum of (3.37) and (3.45), and the EMSE bound in (3.64). We observe that the experimental and theoretical EMSE values practically coincide for SNR higher than 5 dB. Also, the EMSE bound is very close to the true EMSE for SNR higher than 15 dB.

In the second part of our experiments, we take averages over the channel matrices by assuming that the elements of $\mathbf{H}$ are i.i.d. $\mathcal{CN}(0,1)$. The SNR in this case is defined as

$$\mathrm{SNR} \triangleq \frac{\mathcal{E}_{\mathbf{x},\mathbf{H}}\left[\mathrm{tr}(\mathbf{H}\mathbf{x}\mathbf{x}^H\mathbf{H}^H)\right]}{\mathcal{E}_{\mathbf{w}}\left[\mathrm{tr}(\mathbf{w}\mathbf{w}^H)\right]} = \frac{\sigma_x^2\, n_t}{\sigma_w^2}. \tag{3.70}$$

In Fig. 3.4, we plot the experimentally computed EMSE and the theoretical second-order approximation, i.e., the sum of (3.37) and (3.45), averaged over different channel realizations, for the parameters defined above. We observe that the two curves coincide for SNR higher than 7 dB, meaning that our analysis holds for this case too, although it is difficult to give a simple expression for the theoretical second-order approximation.

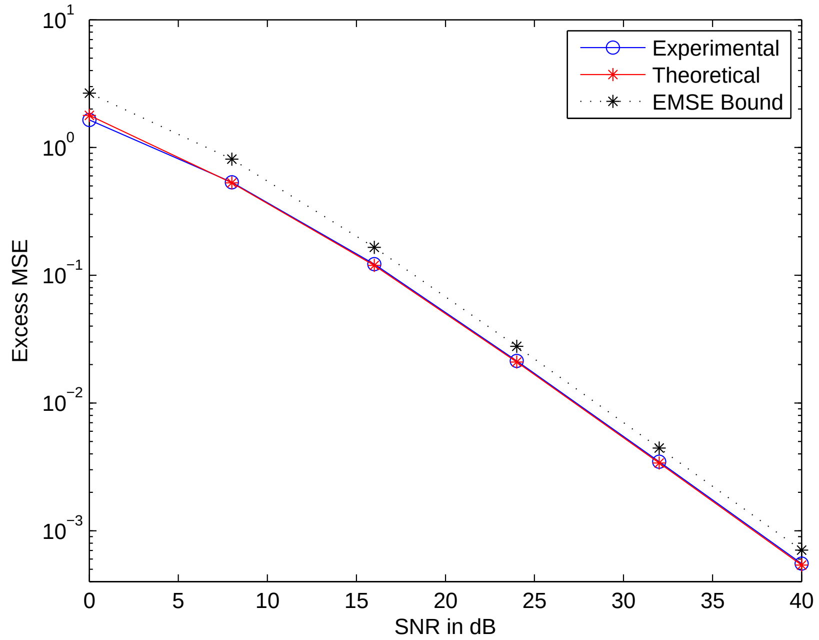

Fig. 3.3. Experimentally computed EMSE, theoretical second-order approximation (sum of (3.37) and (3.45)), and EMSE bound in (3.64).

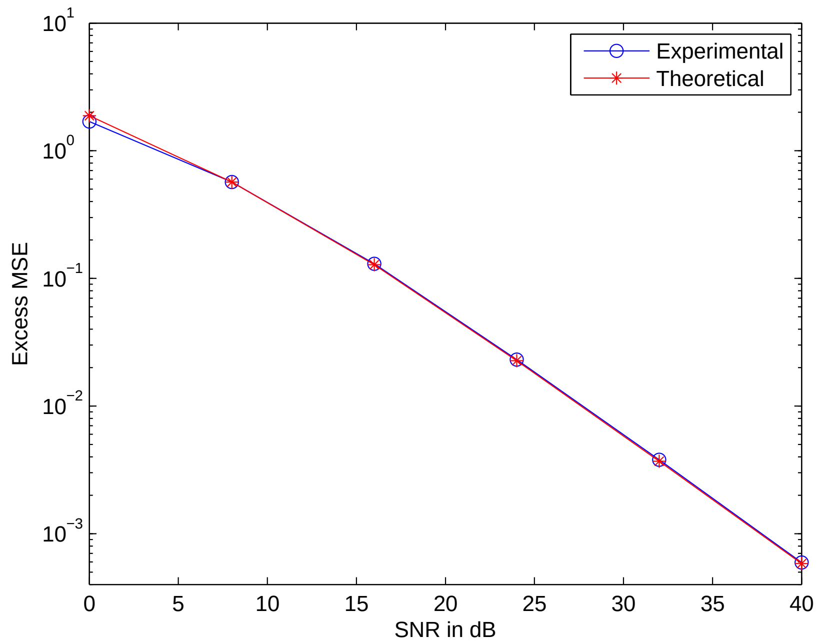

Fig. 3.4. Experimentally computed EMSE and theoretical second-order approximation (sum of (3.37) and (3.45)) averaged over different channel realizations.

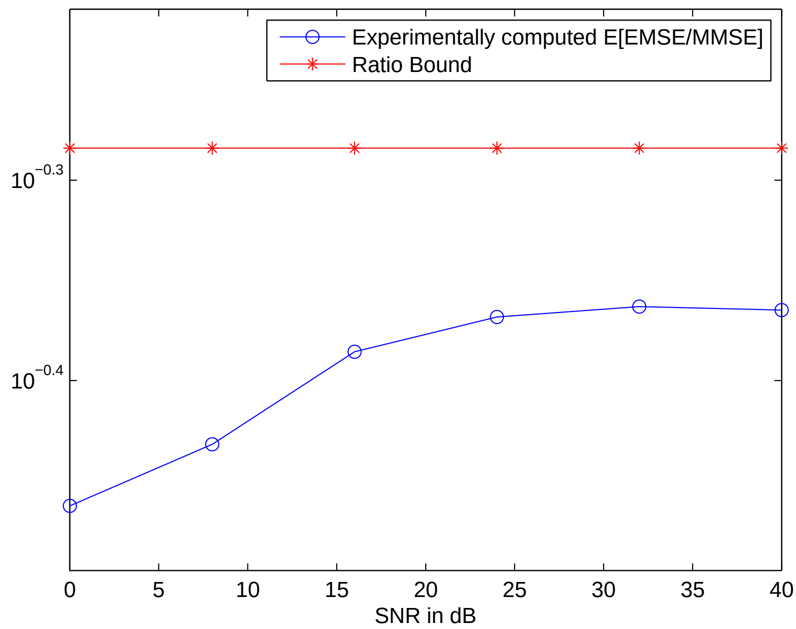

Fig. 3.5. Experimentally computed averaged ratio $\dfrac{\text{EMSE}}{\text{MMSE}}$ and the corresponding bound in (3.65).

In Fig. 3.5, we plot the experimental average ratio $\frac{\text{EMSE}}{\text{MMSE}}$ and the simple bound in (3.65). We observe that the bound in (3.65) is very close to the true average EMSE to MMSE ratio, which attains a constant value for sufficiently high SNR.

3.6 Conclusion

We considered the sensitivity of the mismatched MIMO-TH with respect to channel estimation errors and channel time-variations. We developed a second-order EMSE approximation which, unfortunately, was difficult to interpret. We focused on the high-SNR regime and derived a simple and informative EMSE upper bound that uncovers the factors that determine the sensitivity of the MIMO-TH precoder with respect to channel uncertainties at both the transmitter and the receiver. Numerical experiments were in agreement with our theoretical analysis.

Appendix 3A

Proof of Proposition 1: The aim is to compute the EMSE assuming channel inaccuracies at both the transmitter and the receiver. The matrix filters used in this case are given by (3.21) and (3.24).

We have already defined $\Delta\mathbf{C}$ and $\Delta\mathbf{V}$, as $\Delta\mathbf{C} \triangleq \hat{\mathbf{C}} - \mathbf{C}_o$ and $\Delta\mathbf{V} \triangleq \tilde{\mathbf{V}} - \mathbf{V}_o$, respectively. We have also mentioned (and prove in Appendices II, III) that $\Delta\mathbf{C}$ depends only on $\Delta\mathbf{H}_{\mathrm{Tx}}$, while $\Delta\mathbf{V}$ depends only on $\Delta\mathbf{H}_{\mathrm{Rx}}$. We recall that $\Delta\mathbf{H}_{\mathrm{Tx}}$ and $\Delta\mathbf{H}_{\mathrm{Rx}}$ are independent.

In order to compute the EMSE defined in (3.26), we define $\Delta\check{\mathbf{V}} \triangleq \hat{\tilde{\mathbf{V}}} - \mathbf{V}_o$ and use (3.24) and (3.29). Then

$$\begin{aligned}\Delta\check{\mathbf{V}} &\triangleq \hat{\tilde{\mathbf{V}}} - \mathbf{V}_o \\ &= (\mathbf{C}_o + \Delta\mathbf{C})\tilde{\mathbf{H}}^H(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H + \zeta\mathbf{I}_{n_r})^{-1} - \mathbf{V}_o \\ &= \tilde{\mathbf{V}} - \mathbf{V}_o + \Delta\mathbf{C}\tilde{\mathbf{H}}^H(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H + \zeta\mathbf{I}_{n_r})^{-1} \\ &= \Delta\mathbf{V} + \Delta\mathbf{C}\,\underbrace{\tilde{\mathbf{H}}^H(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H + \zeta\mathbf{I}_{n_r})^{-1}}_{\mathbf{X}}.\end{aligned} \tag{3.71}$$

We observe that $\Delta\check{\mathbf{V}}$ depends on both $\Delta\mathbf{H}_{\mathrm{Tx}}$ and $\Delta\mathbf{H}_{\mathrm{Rx}}$, through $\Delta\mathbf{C}$ and $\Delta\mathbf{V}$, respectively. If we write term $\mathbf{X}$ using (11) and then keep only the first-order terms, we get

$$\begin{aligned}\mathbf{X} = \mathbf{H}^H\mathbf{B}^{-1} - \mathbf{H}^H\mathbf{B}^{-1}(\underbrace{\mathbf{H}\Delta\mathbf{H}_{\mathrm{Rx}}^H + \Delta\mathbf{H}_{\mathrm{Rx}}\mathbf{H}^H}_{\Delta\mathbf{B}})\,\mathbf{B}^{-1} \\ + \Delta\mathbf{H}_{\mathrm{Rx}}^H\,\mathbf{B}^{-1} + O(\|\Delta\mathbf{H}_{\mathrm{Rx}}\|^2) \\ \simeq \mathbf{H}^H\mathbf{B}^{-1} - \mathbf{H}^H\mathbf{B}^{-1}\Delta\mathbf{B}\,\mathbf{B}^{-1} + \Delta\mathbf{H}_{\mathrm{Rx}}^H\,\mathbf{B}^{-1}\end{aligned} \tag{3.72}$$

where matrix $\mathbf{B}$ is defined in (3.44). Combining (3.71) and (3.72), we get

$$\Delta\check{\mathbf{V}} \simeq \Delta\mathbf{V} + \Delta\mathbf{C}\mathbf{H}^H\mathbf{B}^{-1}. \tag{3.73}$$

Next, we return to the EMSE definition in (3.26). We first substitute $\hat{\mathbf{C}}$ and $\hat{\tilde{\mathbf{V}}}$ with $\hat{\mathbf{C}} = \mathbf{C}_o + \Delta\mathbf{C}$ and $\hat{\tilde{\mathbf{V}}} = \Delta\check{\mathbf{V}} + \mathbf{V}_o$, respectively, in (3.25). Then, using the definition (3.26) and keeping only the second-order terms, we get

$$\begin{aligned}\mathrm{EMSE}(\hat{\mathbf{C}}, \hat{\tilde{\mathbf{V}}}) = \underbrace{\sigma_x^2\mathcal{E}\left[\mathrm{tr}(\Delta\check{\mathbf{V}}\mathbf{B}\Delta\check{\mathbf{V}}^H)\right]}_{t_1} - \underbrace{\sigma_x^2\mathcal{E}\left[\mathrm{tr}(\Delta\check{\mathbf{V}}\mathbf{H}\Delta\mathbf{C}^H)\right]}_{t_2} \\ - \underbrace{\sigma_x^2\mathcal{E}\left[\mathrm{tr}(\Delta\mathbf{C}\mathbf{H}^H\Delta\check{\mathbf{V}})\right]}_{t_3} + \sigma_x^2\mathcal{E}\left[\mathrm{tr}(\Delta\mathbf{C}\Delta\mathbf{C}^H)\right].\end{aligned} \tag{3.74}$$

Using (3.73), and the fact that $\Delta\mathbf{H}_{\mathrm{Tx}}$ and $\Delta\mathbf{H}_{\mathrm{Rx}}$ are zero mean and independent (which implies independence between $\Delta\mathbf{C}$ and $\Delta\mathbf{V}$), terms t_i become

$$\begin{aligned} t_1 &\approx \sigma_x^2\,\mathcal{E}\left[\mathrm{tr}(\Delta\mathbf{V}\mathbf{B}\Delta\mathbf{V}^H)\right] + \sigma_x^2\,\mathcal{E}\left[\mathrm{tr}(\Delta\mathbf{C}\mathbf{H}^H\mathbf{B}^{-1}\mathbf{H}\Delta\mathbf{C}^H)\right] \\ &\overset{(3.43)}{=} \mathrm{EMSE}(\tilde{\mathbf{V}}) + \sigma_x^2\,\mathcal{E}\left[\mathrm{tr}(\Delta\mathbf{C}\mathbf{H}^H\mathbf{B}^{-1}\mathbf{H}\Delta\mathbf{C}^H)\right] \end{aligned} \tag{3.75}$$

and

$$t_2 = t_3 \approx \sigma_x^2\,\mathcal{E}\left[\mathrm{tr}(\Delta\mathbf{C}\mathbf{H}^H\mathbf{B}^{-1}\mathbf{H}\Delta\mathbf{C}^H)\right]. \tag{3.76}$$

Finally, we combine (3.74)–(3.76) and use (3.51) and (3.35) to get

$$\mathrm{EMSE}(\hat{\mathbf{C}}, \hat{\tilde{\mathbf{V}}}) \approx \mathrm{EMSE}(\hat{\mathbf{C}}) + \mathrm{EMSE}(\tilde{\mathbf{V}}).$$

Appendix 3B

Proof of Lemma 1: The aim is to develop a second-order approximation to $\mathrm{EMSE}(\hat{\mathbf{C}})$. Towards this purpose, we must develop a first-order approximation to $\Delta\mathbf{C}$ with respect to $\Delta\mathbf{H}_{\mathrm{Tx}}$. Using (3.21) and defining $\Delta\mathbf{G} \triangleq \hat{\mathbf{G}} - \mathbf{G}$ and $\Delta\mathbf{R} \triangleq \hat{\mathbf{R}} - \mathbf{R}$, we obtain

$$\hat{\mathbf{C}} = (\mathbf{G} + \Delta\mathbf{G})(\mathbf{R} + \Delta\mathbf{R}) = \mathbf{C} + \mathbf{G}\Delta\mathbf{R} + \Delta\mathbf{G}\mathbf{R} + \Delta\mathbf{G}\Delta\mathbf{R}.$$

Thus, a first-order approximation to $\Delta\mathbf{C}$, with respect to $\Delta\mathbf{R}$ and $\Delta\mathbf{G}$, is

$$\Delta\mathbf{C} \simeq \mathbf{G}\Delta\mathbf{R} + \Delta\mathbf{G}\mathbf{R}. \tag{3.77}$$

Next, we derive first-order approximations to $\Delta\mathbf{R}$ and $\Delta\mathbf{G}$, with respect to $\Delta\mathbf{H}_{\mathrm{Tx}}$. We start with $\Delta\mathbf{R}$. We remind that $\mathbf{R}^H\mathbf{R} = \mathbf{A}$ and

$$\hat{\mathbf{R}}^H\hat{\mathbf{R}} = \hat{\mathbf{H}}^H\hat{\mathbf{H}} + \zeta\mathbf{I}_{n_t} \simeq \mathbf{A} + \underbrace{\left(\mathbf{H}^H\Delta\mathbf{H}_{\mathrm{Tx}} + \Delta\mathbf{H}_{\mathrm{Tx}}^H\mathbf{H}\right)}_{\Delta\mathbf{A}}.$$

Using a result for the Cholesky factorization of a perturbed positive definite matrix [27], we obtain

$$\hat{\mathbf{R}} \triangleq \mathbf{R} + \Delta\mathbf{R} \simeq \mathbf{R} + \mathbf{F}_L\mathbf{R}$$

where $\mathbf{F}_L$ is the lower triangular part of matrix $\mathbf{F} \triangleq \mathbf{R}^{-H}\Delta\mathbf{A}\,\mathbf{R}^{-1}$, with diagonal elements equal to half the diagonal elements of $\mathbf{F}$. Thus,

$$\Delta\mathbf{R} \simeq \mathbf{F}_L\mathbf{R}. \tag{3.78}$$

For term $\Delta\mathbf{G}$, we have

$$\begin{aligned}\Delta\mathbf{G} &\triangleq \hat{\mathbf{G}} - \mathbf{G}\\ &= \operatorname{diag}\left(\frac{1}{\hat{r}_{11}}, \cdots, \frac{1}{\hat{r}_{n_tn_t}}\right) - \operatorname{diag}\left(\frac{1}{r_{11}}, \cdots, \frac{1}{r_{n_tn_t}}\right)\\ &\overset{(a)}{\simeq} \operatorname{diag}\left(\frac{1}{r_{11}} - \frac{\Delta r_{11}}{r_{11}^2}, \cdots, \frac{1}{r_{n_tn_t}} - \frac{\Delta r_{n_tn_t}}{r_{n_tn_t}^2}\right)\\ &- \operatorname{diag}\left(\frac{1}{r_{11}}, \cdots, \frac{1}{r_{n_tn_t}}\right)\end{aligned}$$

where at point (a) we used the first-order approximation

$$\frac{1}{\hat{r}_{ii}} \simeq \frac{1}{r_{ii}} - \frac{\Delta r_{ii}}{r_{ii}^2}.$$

Thus,

$$\Delta\mathbf{G} \simeq -\operatorname{diag}\left(\frac{\Delta r_{11}}{r_{11}^2}, \ldots, \frac{\Delta r_{n_tn_t}}{r_{n_tn_t}^2}\right). \tag{3.79}$$

Finally, using (3.78), (3.79) and the definition of matrix $\mathbf{G}$ in (3.8), we get

$$\begin{aligned}\Delta\mathbf{G} &\simeq -\operatorname{diag}\left(\operatorname{diag}\left(\mathbf{F}_L\right)\right) \operatorname{diag}\left(\operatorname{diag}\left(\mathbf{R}\right)\right)^{-1}\\ &= -\frac{1}{2} \operatorname{diag}\left(\operatorname{diag}\left(\mathbf{F}\right)\right) \mathbf{G}.\end{aligned} \tag{3.80}$$

Up to this point, we have expressed terms $\Delta\mathbf{R}$ and $\Delta\mathbf{G}$ as functions of the matrix $\mathbf{F}$, which, in turn, is a linear function of $\Delta\mathbf{H}_{\mathrm{Tx}}$. Next, we return to (3.35) and using (1) we write the EMSE as

$$\operatorname{EMSE}(\hat{\mathbf{C}}) = \sigma_w^2 \operatorname{tr}\left((\mathbf{A}^{-T} \otimes \mathbf{I}_{n_t})\mathcal{E}[\operatorname{vec}(\Delta\mathbf{C})\operatorname{vec}^H(\Delta\mathbf{C})]\right). \tag{3.81}$$

Using (3.77), (3.78) and (3.80), and defining $\mathbf{D}_{\mathbf{F}_L} \triangleq \operatorname{diag}(\operatorname{diag}(\mathbf{F}_L))$, we can express term $\operatorname{vec}(\Delta\mathbf{C})$ as

$$\begin{aligned}\operatorname{vec}(\Delta\mathbf{C}) &\simeq \operatorname{vec}(\mathbf{G}\Delta\mathbf{R}) + \operatorname{vec}(\Delta\mathbf{G}\mathbf{R})\\ &\simeq \operatorname{vec}(\mathbf{G}\mathbf{F}_L\mathbf{R}) + \operatorname{vec}(-\mathbf{D}_{\mathbf{F}_L}\mathbf{G}\mathbf{R})\\ &\overset{(2)}{=} (\mathbf{R}^T \otimes \mathbf{G})\operatorname{vec}(\mathbf{F}_L) - (\mathbf{R}^T\mathbf{G}^T \otimes \mathbf{I}_{n_t})\operatorname{vec}(\mathbf{D}_{\mathbf{F}_L})\\ &\overset{(*)}{=} (\mathbf{R}^T \otimes \mathbf{G})\operatorname{vec}(\mathbf{F}_L) - (\mathbf{R}^T\mathbf{G}^T \otimes \mathbf{I}_{n_t})\operatorname{vec}(\mathbf{D}_{\mathbf{F}_L})\\ &\quad \pm (\mathbf{R}^T \otimes \mathbf{G})\operatorname{vec}(\mathbf{D}_{\mathbf{F}_L})\\ &= (\mathbf{R}^T \otimes \mathbf{G})(\operatorname{vec}(\mathbf{F}_L) + \operatorname{vec}(\mathbf{D}_{\mathbf{F}_L}))\\ &\quad - ((\mathbf{R}^T\mathbf{G}^T \otimes \mathbf{I}_{n_t}) + (\mathbf{R}^T \otimes \mathbf{G}))\operatorname{vec}(\mathbf{D}_{\mathbf{F}_L})\\ &= (\mathbf{R}^T \otimes \mathbf{G})\operatorname{vec}(\mathbf{F}_l)\\ &\quad - ((\mathbf{R}^T\mathbf{G}^T \otimes \mathbf{I}_{n_t}) + (\mathbf{R}^T \otimes \mathbf{G}))\operatorname{vec}(\mathbf{D}_{\mathbf{F}_L})\end{aligned} \tag{3.82}$$

where $\mathbf{F}_l$ is the lower triangular part of $\mathbf{F}$. At point (*) we add and subtract the same term in order to simplify our calculations.

Next, we express $\text{vec}(\mathbf{F}_l)$ and $\text{vec}(\mathbf{D}_{\mathbf{F}_L})$ in terms of $\text{vec}(\Delta\mathbf{H}_{\text{Tx}})$. Using (2), (3) and (9), we can write

$$\begin{aligned}
\text{vec}(\mathbf{F}_l) &= \mathbf{L}^T\text{vech}(\mathbf{F}_l) = \mathbf{L}^T\text{vech}(\mathbf{F}) \\
&= \mathbf{L}^T\text{vech}(\mathbf{R}^{-H}\Delta\mathbf{A}\mathbf{R}^{-1}) \\
&\overset{(9)}{=} \mathbf{L}^T\mathbf{L}(\mathbf{R}^{-T}\otimes\mathbf{R}^{-H})\text{vec}(\Delta\mathbf{A}) \\
&= \mathbf{L}^T\mathbf{L}(\mathbf{R}^{-T}\otimes\mathbf{R}^{-H})\text{vec}(\mathbf{H}^H\Delta\mathbf{H}_{\text{Tx}} + \Delta\mathbf{H}_{\text{Tx}}^H\mathbf{H}) \\
&\overset{(2)}{=} \mathbf{L}^T\mathbf{L}(\mathbf{R}^{-T}\otimes\mathbf{R}^{-H})(\mathbf{I}_{n_t}\otimes\mathbf{H}^H)\text{vec}(\Delta\mathbf{H}_{\text{Tx}}) \\
&\quad + \mathbf{L}^T\mathbf{L}(\mathbf{R}^{-T}\otimes\mathbf{R}^{-H})(\mathbf{H}^T\otimes\mathbf{I}_{n_t})\text{vec}(\Delta\mathbf{H}_{\text{Tx}}^H) \\
&\overset{(3)}{=} \mathbf{L}^T\mathbf{L}(\mathbf{R}^{-T}\otimes\mathbf{R}^{-H}\mathbf{H}^H)\text{vec}(\Delta\mathbf{H}_{\text{Tx}}) \\
&\quad + \mathbf{L}^T\mathbf{L}(\mathbf{R}^{-T}\mathbf{H}^T\otimes\mathbf{R}^{-H})\text{vec}(\Delta\mathbf{H}_{\text{Tx}}^H).
\end{aligned} \tag{3.83}$$

We continue with $\text{vec}(\mathbf{D}_{\mathbf{F}_L})$. Using (2), (3), (10) and defining $\mathbf{P} \triangleq \mathbf{L}^T\mathbf{L}\mathbf{K}\mathbf{L}^T\mathbf{L}$, we obtain

$$\begin{aligned}
\text{vec}(\mathbf{D}_{\mathbf{F}_L}) &\overset{(10)}{=} \frac{1}{2}\mathbf{L}^T\mathbf{L}\mathbf{K}\mathbf{L}^T\mathbf{L}\text{vec}(\mathbf{F}) = \frac{1}{2}\mathbf{P}\,\text{vec}(\mathbf{R}^{-H}\Delta\mathbf{A}\mathbf{R}^{-1}) \\
&\overset{(2)}{=} \frac{1}{2}\mathbf{P}\,(\mathbf{R}^{-T}\otimes\mathbf{R}^{-H})\text{vec}(\mathbf{H}^H\Delta\mathbf{H}_{\text{Tx}} + \Delta\mathbf{H}_{\text{Tx}}^H\mathbf{H}) \\
&= \frac{1}{2}\mathbf{P}\,(\mathbf{R}^{-T}\otimes\mathbf{R}^{-H})(\mathbf{I}_{n_t}\otimes\mathbf{H}^H)\text{vec}(\Delta\mathbf{H}_{\text{Tx}}) \\
&\quad + \frac{1}{2}\mathbf{P}\,(\mathbf{R}^{-T}\otimes\mathbf{R}^{-H})(\mathbf{H}^T\otimes\mathbf{I}_{n_t})\text{vec}(\Delta\mathbf{H}_{\text{Tx}}^H) \\
&\overset{(3)}{=} \frac{1}{2}\mathbf{P}\,(\mathbf{R}^{-T}\otimes\mathbf{R}^{-H}\mathbf{H}^H)\text{vec}(\Delta\mathbf{H}_{\text{Tx}}) \\
&\quad + \frac{1}{2}\mathbf{P}\,(\mathbf{R}^{-T}\mathbf{H}^T\otimes\mathbf{R}^{-H})\text{vec}(\Delta\mathbf{H}_{\text{Tx}}^H).
\end{aligned} \tag{3.84}$$

We return to (3.82), and using (3.83), (3.84) and (8), after some calculations, we obtain

$$\begin{aligned}
\text{vec}(\Delta\mathbf{C}) &\simeq \mathbf{M}_1\text{vec}(\Delta\mathbf{H}_{\text{Tx}}) + \mathbf{M}_2\text{vec}(\Delta\mathbf{H}_{\text{Tx}}^H) \\
&= \mathbf{M}_1\text{vec}(\Delta\mathbf{H}_{\text{Tx}}) + \mathbf{M}_2\mathbf{K}\text{vec}(\Delta\mathbf{H}_{\text{Tx}}^*)
\end{aligned} \tag{3.85}$$

where

$$\begin{aligned}
\mathbf{M}_1 &\triangleq (\mathbf{R}^T\otimes\mathbf{G})\mathbf{L}^T\mathbf{L}(\mathbf{R}^{-T}\otimes\mathbf{R}^{-H}\mathbf{H}^H) \\
&\quad - \frac{1}{2}\big((\mathbf{R}^T\mathbf{G}^T\otimes\mathbf{I}_{n_t}) + (\mathbf{R}^T\otimes\mathbf{G})\big)\mathbf{P}(\mathbf{R}^{-T}\otimes\mathbf{R}^{-H}\mathbf{H}^H)
\end{aligned} \tag{3.86}$$

and

$$\begin{aligned}
\mathbf{M}_2 &\triangleq (\mathbf{R}^T\otimes\mathbf{G})\mathbf{L}^T\mathbf{L}(\mathbf{R}^{-T}\mathbf{H}^T\otimes\mathbf{R}^{-H}) \\
&\quad - \frac{1}{2}\big((\mathbf{R}^T\mathbf{G}^T\otimes\mathbf{I}_{n_t}) + (\mathbf{R}^T\otimes\mathbf{G})\big)\mathbf{P}(\mathbf{R}^{-T}\mathbf{H}^T\otimes\mathbf{R}^{-H}).
\end{aligned} \tag{3.87}$$

Using the circular symmetry of $\Delta\mathbf{H}_{\text{Tx}}$, (3.19), (3.17), (3.14) and (3.81), we obtain

$$\begin{aligned}\text{EMSE}(\hat{\mathbf{C}}) &\approx \sigma_w^2 \operatorname{tr}\left(\left(\mathbf{A}^{-T} \otimes \mathbf{I}_{n_t}\right) \mathbf{M}_1(\boldsymbol{\Sigma}_{\text{est}} + \boldsymbol{\Sigma}_{\text{tv}})\mathbf{M}_1^H\right) \\ &\quad + \sigma_w^2 \operatorname{tr}\left(\left(\mathbf{A}^{-T} \otimes \mathbf{I}_{n_t}\right) \mathbf{M}_2\mathbf{K}(\boldsymbol{\Sigma}_{\text{est}} + \boldsymbol{\Sigma}_{\text{tv}})^*\mathbf{K}^H\mathbf{M}_2^H\right) \\ &= \left(\alpha + \frac{\sigma_w^2}{\sigma_x^2 N_{\text{tr}}}\right) \sigma_w^2 \operatorname{tr}\left(\left(\mathbf{A}^{-T} \otimes \mathbf{I}_{n_t}\right) \mathbf{M}_1\mathbf{M}_1^H\right) \\ &\quad + \left(\alpha + \frac{\sigma_w^2}{\sigma_x^2 N_{\text{tr}}}\right) \sigma_w^2 \operatorname{tr}\left(\left(\mathbf{A}^{-T} \otimes \mathbf{I}_{n_t}\right) \mathbf{M}_2\mathbf{M}_2^H\right)\end{aligned} \tag{3.88}$$

where we also defined the scalar α, as $\alpha \triangleq 2\,(1 - \rho_\tau)$, and used that $\mathbf{K}\mathbf{K}^H = \mathbf{I}$. Using (3.86) and (3.87) and after some calculations, it can be shown that the second-order EMSE approximation (3.88) can be expressed as

$$\text{EMSE}(\hat{\mathbf{C}}) \approx \sum_{i=1}^{3} \mathbf{B}_i.$$

where terms $\mathbf{B}_i$ are given in (3.38)–(3.40). During the calculations, we also used that, from the definition of matrices $\mathbf{R}$ and $\mathbf{A}$ in (3.7) and (3.36), respectively, we get

$$\mathbf{R}^*\mathbf{A}^{-T}\mathbf{R}^T = \mathbf{I}_{n_t}. \tag{3.89}$$

Appendix 3C

Proof of Lemma 2: The aim is to develop a second-order approximation to $\text{EMSE}(\tilde{\mathbf{V}})$. In order to compute the EMSE defined in (3.43), we must develop a first-order approximation to $\Delta\mathbf{V}$ with respect to $\Delta\mathbf{H}_{\text{Rx}}$, which is defined as $\Delta\mathbf{H}_{\text{Rx}} \triangleq \tilde{\mathbf{H}} - \mathbf{H}$. We can write $\tilde{\mathbf{V}}$ from (3.29) as

$$\tilde{\mathbf{V}} = \mathbf{C}_o(\mathbf{H}^H + \Delta\mathbf{H}_{\text{Rx}}^H)\left(\mathbf{H}\mathbf{H}^H + \zeta\mathbf{I}_{n_r} + \underbrace{\mathbf{H}\Delta\mathbf{H}_{\text{Rx}}^H + \Delta\mathbf{H}_{\text{Rx}}\mathbf{H}^H}_{\Delta\mathbf{B}} + O(\|\Delta\mathbf{H}_{\text{Rx}}\|^2)\right)^{-1}. \tag{3.90}$$

Using (11) and the definition of $\mathbf{V}_o$ in (3.9), we obtain

$$\tilde{\mathbf{V}} \simeq \mathbf{V}_o - \left(\mathbf{V}_o\Delta\mathbf{B} - \mathbf{C}_o\Delta\mathbf{H}_{\text{Rx}}^H\right)\mathbf{B}^{-1}.$$

Thus, a first-order approximation to $\Delta\mathbf{V}$ is

$$\Delta\mathbf{V} \simeq \underbrace{-\left(\mathbf{V}_o\Delta\mathbf{B} - \mathbf{C}_o\Delta\mathbf{H}_{\text{Rx}}^H\right)\mathbf{B}^{-1}}_{\mathbf{K}_\Delta} \tag{3.91}$$

and a second-order approximation of the EMSE is given by

$$\begin{aligned}
\text{EMSE}(\tilde{\mathbf{V}}) &= \sigma_x^2\,\mathcal{E}\left[\text{tr}(\Delta\mathbf{V}\mathbf{B}\Delta\mathbf{V}^H)\right] \\
&\overset{(3.91)}{\approx} \sigma_x^2\,\mathcal{E}\left[\text{tr}(\mathbf{K}_\Delta\mathbf{B}^{-1}\mathbf{K}_\Delta^H)\right] \\
&= \sigma_x^2\,\mathcal{E}\left[\text{tr}(\mathbf{I}_{n_t}\mathbf{K}_\Delta\mathbf{B}^{-1}\mathbf{K}_\Delta^H)\right] \\
&\overset{(1)}{=} \sigma_x^2\,\mathcal{E}\left[\text{vec}^H(\mathbf{K}_\Delta)\big(\mathbf{B}^{-T}\otimes\mathbf{I}_{n_t}\big)\text{vec}(\mathbf{K}_\Delta)\right] \\
&= \sigma_x^2\,\text{tr}\left(\big(\mathbf{B}^{-T}\otimes\mathbf{I}_{n_t}\big)\;\mathcal{E}\left[\text{vec}(\mathbf{K}_\Delta)\text{vec}^H(\mathbf{K}_\Delta)\right]\right).
\end{aligned} \tag{3.92}$$

From the definitions of $\mathbf{K}_\Delta$ in (3.91), $\Delta\mathbf{B}$ in (3.90), and (2), we obtain

$$\begin{aligned}
\text{vec}(\mathbf{K}_\Delta) &= -\text{vec}\left(\mathbf{V}_o\Delta\mathbf{H}_{\text{Rx}}\mathbf{H}^H\right) + \text{vec}\left((\mathbf{C}_o-\mathbf{V}_o\mathbf{H})\Delta\mathbf{H}_{\text{Rx}}^H\right) \\
&= \underbrace{-(\mathbf{H}^*\otimes\mathbf{V}_o)}_{\mathbf{U}_1}\text{vec}(\Delta\mathbf{H}_{\text{Rx}}) \\
&\quad + \underbrace{\big(\mathbf{I}_{n_r}\otimes(\mathbf{C}_o-\mathbf{V}_o\mathbf{H})\big)}_{\mathbf{U}_2}\text{vec}(\Delta\mathbf{H}_{\text{Rx}}^H).
\end{aligned} \tag{3.93}$$

Using (8), we get

$$\text{vec}(\mathbf{K}_\Delta) = \mathbf{U}_1\text{vec}(\Delta\mathbf{H}_{\text{Rx}}) + \mathbf{U}_2\mathbf{K}\text{vec}(\Delta\mathbf{H}_{\text{Rx}}^*).$$

Using the circular symmetry of $\Delta\mathbf{H}_{\text{Rx}}$ and (3.20), we obtain

$$\begin{aligned}
\text{EMSE}(\tilde{\mathbf{V}}) &\approx \sigma_x^2\,\text{tr}\left((\mathbf{B}^{-T}\otimes\mathbf{I}_{n_t})\mathbf{U}_1\,\mathbf{\Sigma}_{\text{est}}\,\mathbf{U}_1^H\right) \\
&\quad + \sigma_x^2\,\text{tr}\left((\mathbf{B}^{-T}\otimes\mathbf{I}_{n_t})\mathbf{U}_2\mathbf{K}\mathbf{\Sigma}_{\text{est}}^*\mathbf{K}^H\mathbf{U}_2^H\right).
\end{aligned}$$

Finally, using (3.14) we obtain the expression

$$\text{EMSE}(\tilde{\mathbf{V}}) \approx \mathbf{T}_1 + \mathbf{T}_2$$

where

$$\begin{aligned}
\mathbf{T}_1 &\overset{\triangle}{=} \sigma_x^2\,\text{tr}\left((\mathbf{B}^{-T}\otimes\mathbf{I}_{n_t})\mathbf{U}_1\,\mathbf{\Sigma}_{\text{est}}\,\mathbf{U}_1^H\right) \\
&= \frac{\sigma_w^2}{N_{\text{tr}}}\,\text{tr}\left(\mathbf{U}_1^H(\mathbf{B}^{-T}\otimes\mathbf{I}_{n_t})\mathbf{U}_1\right) \\
&\overset{(3.93)}{=} \frac{\sigma_w^2}{N_{\text{tr}}}\,\text{tr}\left(\mathbf{H}^T\mathbf{B}^{-T}\mathbf{H}^*\otimes\mathbf{V}_o\mathbf{V}_o^H\right) \\
&\overset{(4)}{=} \frac{\sigma_w^2}{N_{\text{tr}}}\,\text{tr}\left(\mathbf{H}^T\mathbf{B}^{-T}\mathbf{H}^*\right)\text{tr}\left(\mathbf{V}_o\mathbf{V}_o^H\right)
\end{aligned}$$

and

$$\begin{aligned}\mathbf{T}_2 &\triangleq \sigma_x^2 \operatorname{tr}\left((\mathbf{B}^{-T} \otimes \mathbf{I}_{n_t})\mathbf{U}_2\, \mathbf{K}\boldsymbol{\Sigma}_{\text{est}}^* \mathbf{K}^H\, \mathbf{U}_2^H\right) \\ &= \frac{\sigma_w^2}{N_{\text{tr}}} \operatorname{tr}\left(\mathbf{K}^H \mathbf{U}_2^H (\mathbf{B}^{-T} \otimes \mathbf{I}_{n_t})\mathbf{U}_2\, \mathbf{K}\right) \\ &\overset{(3.93)}{=} \frac{\sigma_w^2}{N_{\text{tr}}} \operatorname{tr}\left(\mathbf{K}^H \left(\mathbf{B}^{-T} \otimes (\mathbf{C}_o - \mathbf{V}_o\mathbf{H})^H (\mathbf{C}_o - \mathbf{V}_o\mathbf{H})\right) \mathbf{K}\right) \\ &\overset{(5)}{=} \frac{\sigma_w^2}{N_{\text{tr}}} \operatorname{tr}\left(((\mathbf{C}_o - \mathbf{V}_o\mathbf{H})^H (\mathbf{C}_o - \mathbf{V}_o\mathbf{H}) \otimes \mathbf{B}^{-T})\right) \\ &\overset{(4)}{=} \frac{\sigma_w^2}{N_{\text{tr}}} \operatorname{tr}\left((\mathbf{C}_o - \mathbf{V}_o\mathbf{H})^H (\mathbf{C}_o - \mathbf{V}_o\mathbf{H})\right) \operatorname{tr}\left(\mathbf{B}^{-T}\right).\end{aligned}$$

Appendix 3D

In this Appendix, we simplify term $\operatorname{tr}(\mathbf{V}_o\mathbf{V}_o^H)$ in the high-SNR regime (i.e., $\zeta \to 0$). Using (3.6), (3.7), (3.9), and (3.51), we write matrix $\mathbf{V}_o$ as

$$\mathbf{V}_o = \mathbf{G}\mathbf{R}^{-H}\mathbf{H}^H.$$

Then, using (3.48), we get

$$\begin{aligned}\operatorname{tr}\left(\mathbf{V}_o\mathbf{V}_o^H\right) &= \operatorname{tr}\left(\mathbf{G}\mathbf{R}^{-H}\mathbf{H}^H\mathbf{H}\mathbf{R}^{-1}\mathbf{G}^H\right) \\ &\approxeq \operatorname{tr}\left(\mathbf{G}\mathbf{G}^H\right).\end{aligned}$$

Finally, using (3.11), we get

$$\operatorname{tr}\left(\mathbf{V}_o\mathbf{V}_o^H\right) \approxeq \frac{1}{\sigma_w^2}\,\text{MMSE}.$$

Appendix 3E

In this Appendix, we prove the second equality in (3.57) for the $n_r \times 2$ case (i.e., $n_t = 2$). The aim is to simplify the trace term of the first line of (3.57)

$$\operatorname{tr}\left((\mathbf{I}_{n_t} \otimes \mathbf{G}^2)\mathbf{P}(\mathbf{R}^{-T}\mathbf{R}^{-*} \otimes \mathbf{I}_{n_t})\mathbf{P}\right). \tag{3.94}$$

For notational simplicity, we define matrices $\mathbf{Q}$ and $\mathbf{Z}$, as $\mathbf{Q} \triangleq \mathbf{R}^{-T}\mathbf{R}^{-*}$ and $\mathbf{Z} \triangleq \mathbf{G}^2$.

We first write the matrices inside the trace operator of (3.94). For the $n_r \times 2$ case,

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and

$$\mathbf{I}_2 \otimes \mathbf{Z} = \begin{bmatrix} \mathbf{Z} & \mathbf{O}_2 \\ \mathbf{O}_2 & \mathbf{Z} \end{bmatrix}.$$

For the other Kronecker product we get

$$\mathbf{Q} \otimes \mathbf{I}_2 = \begin{bmatrix} q_{11} & 0 & q_{12} & 0 \\ 0 & q_{11} & 0 & q_{12} \\ q_{21} & 0 & q_{22} & 0 \\ 0 & q_{21} & 0 & q_{22} \end{bmatrix}.$$

Then, the product of the matrices inside the trace operator is

$$(\mathbf{I}_{n_t} \otimes \mathbf{G}^2)\mathbf{P}(\mathbf{R}^{-T}\mathbf{R}^{-*} \otimes \mathbf{I}_{n_t})\mathbf{P} = \begin{bmatrix} z_{11}q_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & z_{22}q_{22} \end{bmatrix}$$

and is obvious that

$$\begin{aligned} \operatorname{tr}\left((\mathbf{I}_{n_t} \otimes \mathbf{G}^2)\mathbf{P}(\mathbf{R}^{-T}\mathbf{R}^{-*} \otimes \mathbf{I}_{n_t})\mathbf{P}\right) &= \operatorname{tr}(\mathbf{ZQ}) \\ &= \operatorname{tr}\left(\mathbf{G}^2\mathbf{R}^{-T}\mathbf{R}^{-*}\right). \end{aligned} \tag{3.95}$$

Using an analogous procedure, it can be shown that result (3.95) holds for the general $n_r \times n_t$ case.

Chapter 4

On the training sequence design for joint channel and CFO estimation in frequency-selective single-carrier systems with MMSE linear equalizers

We consider optimal training sequence design for joint channel and carrier frequency offset (CFO) estimation in frequency-selective single-carrier systems with an MMSE linear equalizer. Performance degradation is due to the fact that a *mismatched* MMSE linear equalizer is applied to channel output samples with *imperfectly canceled* CFO. Our aim in this work is to uncover the relative importance of these error sources. Toward this end, we develop asymptotic expressions for the excess mean square error induced by the channel and CFO estimation errors. We show that, in many cases of high practical importance, the excess mean square error is approximately proportional to the CFO estimation error variance, with the proportionality factor being independent of the training sequence. Thus, in these cases, performance degradation is mainly caused by the imperfectly canceled CFO and optimal training sequence design for CFO estimation is also highly relevant for *joint* channel and CFO estimation.

4.1 Introduction

The receiver of a communication system has to deal with synchronization issues like frame synchronization, carrier frequency offset (CFO) estimation and correction, symbol-timing recovery, and channel estimation and equalization. Data-aided (DA) techniques perform these tasks using training sequences (TSs). DA techniques decrease the symbol rate but are commonly used in practice due to their overall good performance. A critical step toward successful application of DA techniques in practice is optimal TS design.

An important problem that usually arises in packet-based wireless systems is that of the *joint* frequency-selective channel and CFO estimation. Optimal TS design for this problem has been considered in [28], where the optimized cost function was the worst-case, in terms of channel realization, *asymptotic* Cramér-Rao bound (CRB). However, all diagonal elements of the CRB were assigned equal weight, which might *not* be optimal since "*... presumably channel estimation errors will have a different impact, e.g., on bit-error rate, than frequency estimation errors*" [28].

It seems that the *unequal weighting* problem cannot be resolved unless one considers specific receiver structures. Toward this end, Ciblat *et al.* computed the second-order statistics (power spectrum) of the training sequence that minimizes, under certain assumptions, the mean square error achieved by the mismatched MMSE linear equalizer [4]. However, in our opinion, the cost function appearing in [4, eq. (26)] does not clearly uncover the relative importance of the channel and CFO estimation errors to the performance degradation of the MMSE linear equalizer.

Performance degradation is due to the fact that a *mismatched* MMSE linear equalizer is applied to channel output samples with *imperfectly canceled* CFO. Our aim in this chapter is to uncover the relative importance of these error sources. We assume that training consists of a block of $N_{\rm tr}$ consecutive training symbols and compute asymptotic expressions for the excess MSE induced by the channel and CFO estimation errors. Under the *small ideal MMSE* assumption, we derive a simple and informative EMSE approximation which reveals that

- the placement of the TS at the middle of the transmitted data packet is a good practice;
- the resulting EMSE is approximately proportional to the CFO estimation error variance, with the proportionality factor being *independent* of the TS; thus, optimal TS design for CFO estimation is also highly relevant for *joint* channel and CFO estimation.

4.2 The channel model

We consider the linear baseband-equivalent discrete-time frequency-selective channel described by the input-output relation

$$z_n = \sum_{l=0}^{L} h_l a_{n-l} + w_n \tag{4.1}$$

where a_n and w_n denote the channel input and additive channel noise, respectively. The input samples are i.i.d., zero mean, circular, with variance σ_a^2, and the noise samples are i.i.d., zero-mean, circularly symmetric complex Gaussian, with variance σ_w^2. The channel impulse response is denoted as $\mathbf{h} \triangleq [h_0 \cdots h_L]^T$. By stacking $(M+1)$ consecutive output samples, we construct the vector $\mathbf{z}_{n:n-M} \triangleq [z_n \ \cdots \ z_{n-M}]^T$ which can be expressed as

$$\mathbf{z}_{n:n-M} = \mathbf{H}\mathbf{a}_{n:n-M-L} + \mathbf{w}_{n:n-M} \tag{4.2}$$

where the definitions of $\mathbf{a}_{n:n-M-L}$ and $\mathbf{w}_{n:n-M}$ are obvious and $\mathbf{H}$ is the $(M+1)\times(M+L+1)$ Toeplitz filtering matrix defined as

$$\mathbf{H} \triangleq \begin{bmatrix} h_0 & \cdots & h_L & & \\ & \ddots & & \ddots & \\ & & h_0 & \cdots & h_L \end{bmatrix}. \tag{4.3}$$

If angular CFO ω is present, then the channel output is given by

$$r_n = \mathrm{e}^{j\omega n} \sum_{l=0}^{L} h_l a_{n-l} + w_n. \tag{4.4}$$

If we stack $(M+1)$ consecutive output samples, we construct the vector $\mathbf{r}_{n:n-M} \triangleq [r_n \ \cdots \ r_{n-M}]^T$ which can be expressed as

$$\mathbf{r}_{n:n-M} = \mathbf{\Gamma}_{n:n-M}(\omega)\mathbf{H}\mathbf{a}_{n:n-L-M} + \mathbf{w}_{n:n-M} \tag{4.5}$$

where $\mathbf{\Gamma}_{n:n-M}(\omega) \triangleq \mathrm{diag}(\mathrm{e}^{j\omega n}, \ldots, \mathrm{e}^{j\omega(n-M)})$.

4.3 The MMSE linear equalizer

4.3.1 Channel and CFO estimation

We assume that the N_{tr} input samples $\mathbf{a}_{\mathrm{tr}} \triangleq [a_{n_1} \ \cdots \ a_{n_2}]^T$, with $N_{\mathrm{tr}} \triangleq n_2 - n_1 + 1$, are known at the receiver and used for training purposes. Collecting the channel output samples that depend only on the training symbols, we obtain

$$\mathbf{y} \triangleq \mathbf{r}_{n_2:n_1+L} = \mathbf{\Gamma}_{n_2:n_1+L}(\omega)\mathbf{A}\mathbf{h} + \mathbf{w}_{n_2:n_1+L} \tag{4.6}$$

where $\mathbf{A}$ is the $(N_{\rm tr} - L) \times (L+1)$ Hankel matrix constructed by the training symbols as

$$\mathbf{A} \triangleq \begin{bmatrix} a_{n_2} & \cdots & a_{n_2-L} \\ \vdots & \ddots & \vdots \\ a_{n_1+L} & \cdots & a_{n_1} \end{bmatrix}.$$

The joint ML CFO and channel estimates are given by [22]

$$\hat{\omega} = \underset{\tilde{\omega}}{\rm argmax}\{\mathbf{y}^H \mathbf{\Gamma}_{n_2:n_1+L}(\tilde{\omega})\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{\Gamma}^H_{n_2:n_1+L}(\tilde{\omega})\mathbf{y}\} \tag{4.7}$$

and

$$\hat{\mathbf{h}} = (\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{\Gamma}^H_{n_2:n_1+L}(\hat{\omega})\mathbf{y}. \tag{4.8}$$

We define the CFO and channel estimation errors as $\Delta\omega \triangleq \hat{\omega} - \omega$ and $\Delta\mathbf{h} \triangleq \hat{\mathbf{h}} - \mathbf{h}$, respectively. We assume that $N_{\rm tr}$ is sufficiently large so that the above ML estimators are (approximately) unbiased and efficient.[1] Thus, $\Delta\omega$ and $\Delta\mathbf{h}$ are (approximately) zero mean with second-order statistics (approximately) equal to those indicated by the *finite sample* CRBs. More specifically, if we define $\mathbf{K} \triangleq {\rm diag}(n_2, \ldots, n_1 + L)$ and $\mathbf{P}^{\perp}_{\mathcal{R}(\mathbf{A})}$ the orthogonal projector onto the orthogonal complement of the column space of $\mathbf{A}$, then [28]

$$\sigma^2_{\Delta\omega} \triangleq \mathcal{E}\left[(\Delta\omega)^2\right] = \frac{1}{2}\sigma^2_w\left[{\rm tr}\left(\mathbf{h}^H\mathbf{A}^H\mathbf{K}\mathbf{P}^{\perp}_{\mathcal{R}(\mathbf{A})}\mathbf{K}\mathbf{A}\mathbf{h}\right)\right]^{-1} \tag{4.9}$$

$$\mathbf{C} \triangleq \mathcal{E}\left[\Delta\mathbf{h}\Delta\mathbf{h}^H\right] = \sigma^2_w(\mathbf{A}^H\mathbf{A})^{-1} + \sigma^2_{\Delta\omega}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{K}\mathbf{A}\mathbf{h}\mathbf{h}^H\mathbf{A}^H\mathbf{K}\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1} \tag{4.10}$$

$$\mathbf{C}_t \triangleq \mathcal{E}[\Delta\mathbf{h}\Delta\mathbf{h}^T] = -\sigma^2_{\Delta\omega}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{K}\mathbf{A}\mathbf{h}\mathbf{h}^T\mathbf{A}^T\mathbf{K}\mathbf{A}^*(\mathbf{A}^H\mathbf{A})^{-T} \tag{4.11}$$

and

$$\mathcal{E}[\Delta\omega\Delta\mathbf{h}] = j\,\sigma^2_{\Delta\omega}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{K}\mathbf{A}\mathbf{h}. \tag{4.12}$$

If we define $D_i \triangleq {\rm diag}(0, 1, \ldots, i)$, then it can be easily shown that

$$\sigma^2_{\Delta\omega} = \frac{1}{2}\sigma^2_w\left[{\rm tr}\left(\mathbf{h}^H\mathbf{A}^H\mathbf{D}_{N_{\rm tr}-L-1}\mathbf{P}^{\perp}_{\mathcal{R}(\mathbf{A})}\mathbf{D}_{N_{\rm tr}-L-1}\mathbf{A}\mathbf{h}\right)\right]^{-1} \tag{4.13}$$

which implies that the CFO estimation error variance is *independent* of the training positions. However, the same does *not* apply to the channel estimation error second-order statistics. This occurs because the channel estimate $\hat{\mathbf{h}}$ in (4.8) is in fact a least-squares estimate based on the CFO-corrected channel output samples. Its accuracy is determined by the propagated CFO estimation error that exists in $\Gamma_{n_2:n_1+L}(\hat{\omega})$.

[1] In the Simulations section, we shall see that our estimates meet the Cramér-Rao bounds for small values of $N_{\rm tr}$.

In order to achieve more accurate channel estimation, we use a slightly different model. More specifically, we write (4.6) as

$$\mathbf{y} = \mathbf{\Gamma}_{\frac{N_{\mathrm{tr}}-L}{2}-1:-\frac{N_{\mathrm{tr}}-L}{2}}(\omega)\mathbf{A}\mathbf{h}' + \mathbf{w}_{n_2:n_1+L} \tag{4.14}$$

where $\mathbf{h}' \triangleq \mathrm{e}^{j\omega\xi}\mathbf{h}$, with $\xi \triangleq n_1 + \frac{N_{\mathrm{tr}}+L}{2}$, i.e., ξ is the middle position of $\mathbf{y}$. Then, the ML estimate of $\mathbf{h}'$ is

$$\hat{\mathbf{h}}' = (\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{\Gamma}^H_{\frac{N_{\mathrm{tr}}-L}{2}-1:-\frac{N_{\mathrm{tr}}-L}{2}}(\hat{\omega})\,\mathbf{y}. \tag{4.15}$$

If we define $\Delta\mathbf{h}' \triangleq \hat{\mathbf{h}}' - \mathbf{h}'$ and $\mathbf{K}' \triangleq \mathrm{diag}\left(\frac{N_{\mathrm{tr}}-L}{2} - 1, \ldots, -\frac{N_{\mathrm{tr}}-L}{2}\right)$, then the *finite sample* CRBs indicate that

$$\mathbf{C}' \triangleq \mathcal{E}\left[\Delta\mathbf{h}'\Delta\mathbf{h}'^H\right] = \sigma_w^2(\mathbf{A}^H\mathbf{A})^{-1} + \sigma^2_{\Delta\omega}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{K}'\mathbf{A}\mathbf{h}'\mathbf{h}'^H\mathbf{A}^H\mathbf{K}'\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1} \tag{4.16}$$

$$\mathbf{C}'_t \triangleq \mathcal{E}[\Delta\mathbf{h}'\Delta\mathbf{h}'^T] = -\sigma^2_{\Delta\omega}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{K}'\mathbf{A}\mathbf{h}'\mathbf{h}'^T\mathbf{A}^T\mathbf{K}'\mathbf{A}^*(\mathbf{A}^H\mathbf{A})^{-T} \tag{4.17}$$

and

$$\mathcal{E}[\Delta\omega\Delta\mathbf{h}'] = j\sigma^2_{\Delta\omega}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{K}'\mathbf{A}\mathbf{h}'. \tag{4.18}$$

In the sequel, we assume that the true channel is $\mathbf{h}'$ and, consequently, the second-order statistics of $\Delta\omega$ and $\Delta\mathbf{h}'$ are (approximately) equal to those appearing in (4.9), (4.16)–(4.18). Finally, we assume that the input and noise variances, σ_a^2 and σ_w^2, are known at the receiver. In fact, we assume that $\sigma_a^2 = 1$, while the noise variance estimation error is negligible compared with the channel and CFO estimation error.

4.3.2 The ideal MMSE linear equalizer

In this subsection, we assume perfect channel and CFO knowledge and compute the ideal MMSE linear equalizer for the channel $\mathbf{h}'$. An order-M delay-d linear equalizer is defined as $\mathbf{f} \triangleq [f_0 \;\cdots\; f_M]^T$. Its output at time instant n, $\tilde{a}_{n-d}$, is an estimate of the delayed channel input, a_{n-d}, and is given by

$$\tilde{a}_{n-d} = \mathbf{f}^H\mathbf{z}_{n:n-M}. \tag{4.19}$$

The input symbol estimation error is defined as

$$e_n \triangleq \tilde{a}_{n-d} - a_{n-d} = \mathbf{f}^H\mathbf{z}_{n:n-M} - \mathbf{e}_d^H\mathbf{a}_{n:n-M-L}$$

where $\mathbf{e}_d$ is the $(M+L+1)\times 1$ vector with 1 at the $(d+1)$-st position and zeros elsewhere. The mean square input symbol estimation error can be expressed as

$$\mathrm{MSE}(\mathbf{f}) \triangleq \mathcal{E}_{a,w}\left[|e_n|^2\right] = \mathbf{f}^H\left(\mathbf{H}'\mathbf{H}'^H + \sigma_w^2\mathbf{I}_{M+1}\right)\mathbf{f} - 2\mathrm{Re}\{\mathbf{f}^H\mathbf{H}'\mathbf{e}_d\} + 1. \tag{4.20}$$

The order-M delay-d MMSE linear equalizer is given by [24, Section 2.7.3]

$$\mathbf{f} = \left(\mathbf{H}'\mathbf{H}'^H + \sigma_w^2 \mathbf{I}_{M+1}\right)^{-1} \mathbf{H}'\mathbf{e}_d = \mathbf{R}_z^{-1}\mathbf{H}'\mathbf{e}_d \tag{4.21}$$

where $\mathbf{R}_z \triangleq \mathcal{E}_{a,w}\left[\mathbf{z}_{n:n-M}\mathbf{z}_{n:n-M}^H\right] = \mathbf{H}'\mathbf{H}'^H + \sigma_w^2\mathbf{I}_{M+1}$.

4.3.3 Mismatched MMSE linear equalizer

If we do not know the channel and the CFO, we may estimate them and use the estimates as if they were the true quantities. Adopting the channel model presented in (4.14), the channel output is expressed as

$$r'_n = \mathrm{e}^{j\omega(n-\xi)} \sum_{l=0}^{L} h'_l a_{n-l} + w_n. \tag{4.22}$$

After the computation of the CFO estimate $\hat{\omega}$, we proceed to CFO correction and obtain

$$s'_n = \mathrm{e}^{-j\hat{\omega}(n-\xi)} r'_n = \mathrm{e}^{j(\omega-\hat{\omega})(n-\xi)} \sum_{l=0}^{L} h'_l a_{n-l} + \mathrm{e}^{-j\hat{\omega}(n-\xi)} w_n. \tag{4.23}$$

If we stack $(M+1)$ consecutive CFO-corrected output samples, we construct the vector $\mathbf{s}'_{n:n-M} \triangleq [s'_n \ \cdots \ s'_{n-M}]^T$, which can be expressed as

$$\mathbf{s}'_{n:n-M} = \mathrm{e}^{j\Delta\omega\xi}\boldsymbol{\Gamma}_{n:n-M}(-\Delta\omega)\mathbf{H}'\mathbf{a}_{n:n-L-M} + \mathrm{e}^{j\hat{\omega}\xi}\boldsymbol{\Gamma}_{n:n-M}(-\hat{\omega})\mathbf{w}_{n:n-M}. \tag{4.24}$$

If we use in (4.21) the channel estimate $\hat{\mathbf{h}}'$ as if it were the true channel, we compute the mismatched MMSE equalizer

$$\hat{\mathbf{f}} = \left(\hat{\mathbf{H}}'\hat{\mathbf{H}}'^H + \sigma_w^2\mathbf{I}_{M+1}\right)^{-1} \hat{\mathbf{H}}'\mathbf{e}_d. \tag{4.25}$$

The input symbol estimation error at the output of the mismatched equalizer at the time instant n is

$$\hat{e}_n = \hat{\mathbf{f}}^H \mathbf{s}'_{n:n-M} - \mathbf{e}_d^H \mathbf{a}_{n:n-L-M} \tag{4.26}$$

and the mean square estimation error of a_{n-d} is

$$\begin{aligned} \mathrm{MSE}_n(\hat{\mathbf{f}}, \hat{\omega}) \triangleq \mathcal{E}_{a,w}\left[|\,\hat{e}_n|^2\right] = \hat{\mathbf{f}}^H \left(\boldsymbol{\Gamma}_{n:n-M}(-\Delta\omega)\mathbf{H}'\mathbf{H}'^H\boldsymbol{\Gamma}_{n:n-M}^H(-\Delta\omega) + \sigma_w^2\mathbf{I}_{M+1}\right)\hat{\mathbf{f}} \\ - 2\,\mathrm{Re}\{\mathrm{e}^{j\Delta\omega\xi}\,\hat{\mathbf{f}}^H\boldsymbol{\Gamma}_{n:n-M}(-\Delta\omega)\mathbf{H}'\mathbf{e}_d\} + 1. \end{aligned} \tag{4.27}$$

We observe that the mean square estimation error is time-dependent and, thus, not suitable for training design.

4.3.4 Excess MSE

The excess MSE at the time instant n is defined as

$$\text{EMSE}_n(\hat{\mathbf{f}}, \hat{\omega}) \stackrel{\triangle}{=} \mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}[\text{MSE}_n(\hat{\mathbf{f}}, \hat{\omega})] - \text{MSE}(\mathbf{f}). \tag{4.28}$$

Using slightly different notation, it has been proved in [18, eq. (22) and (27)] that the mismatched equalizer $\hat{\mathbf{f}}$ can be expressed as

$$\hat{\mathbf{f}} = \mathbf{f} - \mathbf{R}_z^{-1}\left(\mathbf{R}^*\Delta\mathbf{h}' + \mathbf{G}\Delta\mathbf{h}'^*\right) + \mathcal{O}\left(\|\Delta\mathbf{h}'\|^2\right) \tag{4.29}$$

where

1. $\mathbf{R}$ is the $(M+1)\times(L+1)$ Hankel matrix constructed by

$$\mathbf{r} \stackrel{\triangle}{=} \mathbf{c} - \mathbf{e}_d \tag{4.30}$$

   with $\mathbf{c}$ being the combined (channel-equalizer) impulse response, i.e., $\mathbf{c} \stackrel{\triangle}{=} \mathbf{H}'^T\mathbf{f}^*$;

2. $\mathbf{G} \stackrel{\triangle}{=} \mathbf{H}'\mathbf{F}^T$, where $\mathbf{F}$ is the $(L+1)\times(L+M+1)$ Toeplitz filtering matrix constructed by $\mathbf{f}$.

The equalizer mismatch is defined as $\Delta\mathbf{f} \stackrel{\triangle}{=} \hat{\mathbf{f}} - \mathbf{f}$.

The following proposition provides an asymptotic EMSE approximation.

Proposition 1. *The EMSE induced by the channel and CFO estimation errors at time instant n, for $n \in \{d+1,\ldots,n_1+d-1\} \cup \{n_2+d+1,\ldots,N+d\}$,[2] can be expressed as*

$$\text{EMSE}_n(\hat{\mathbf{f}}, \hat{\omega}) = \mathbf{T}_1 + \mathbf{T}_2(n) + \mathbf{T}_3(n) + \mathcal{O}\left(\frac{n^2\,\sigma_w^3}{R^{7/2}}\right) + \mathcal{O}\left(\frac{M^2\,\sigma_w^2}{R^3}\right) \tag{4.31}$$

where

$$\mathbf{T}_1 \stackrel{\triangle}{=} \text{tr}\left(\mathbf{R}_z^{-1}\left(\mathbf{R}^*\mathbf{C}'\mathbf{R}^T + \mathbf{G}\mathbf{C}'^*\mathbf{G}^H + \mathbf{G}\mathbf{C}_t'^*\mathbf{R}^T + \mathbf{R}^*\mathbf{C}_t'\mathbf{G}^H\right)\right) \tag{4.32a}$$

$$\mathbf{T}_2(n) \stackrel{\triangle}{=} \sigma_{\Delta\omega}^2\text{Re}\{\mathbf{f}^H\mathbf{D}_{n:n-M}'^2\mathbf{H}'\mathbf{e}_d\} \tag{4.32b}$$

$$\begin{aligned}\mathbf{T}_3(n) \stackrel{\triangle}{=} 2\sigma_{\Delta\omega}^2\text{Re}\big\{&\mathbf{h}'^H\mathbf{A}^H\mathbf{K}'\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{R}^T\mathbf{R}_z^{-1}\mathbf{D}_{n:n-M}'\mathbf{H}'\mathbf{e}_d \\ &- \mathbf{h}'^T\mathbf{A}^T\mathbf{K}'\mathbf{A}^*(\mathbf{A}^H\mathbf{A})^{-T}\mathbf{G}^H\mathbf{R}_z^{-1}\mathbf{D}_{n:n-M}'\mathbf{H}'\mathbf{e}_d\big\}\end{aligned} \tag{4.32c}$$

where $\mathbf{D}_{n:n-M}' \stackrel{\triangle}{=} \text{diag}((n-\xi),\ldots,(n-M-\xi))$ and $R \stackrel{\triangle}{=} N_{\text{tr}} - L$.

Proof: The proof is provided in Appendix 4A. □

[2] We do not compute the EMSE for the training symbols a_n, $n = n_1,\ldots,n_2$.

Remark 1: Term $\mathbf{T}_1$ involves only the channel estimation error second-order statistics. In fact, it is the EMSE that would result if the mismatched equalizer were applied to perfectly CFO-corrected channel output samples [18, eq. (28)]. On the other hand, $\mathbf{T}_2(n)$ involves only the CFO estimation error variance, while $\mathbf{T}_3(n)$ involves both the channel and CFO estimation errors. □

4.4 Small ideal MMSE assumption and average EMSE

Since the time-dependent EMSE expression of (4.31) is not suitable for training design, we make the assumption of "small ideal MMSE" and adopt as cost function the average EMSE. Then, we derive a very simple and informative average EMSE approximation.

4.4.1 "Small ideal MMSE" assumption

We assume that the *ideal* MMSE is sufficiently small, i.e., the equalizer length is sufficiently large, the SNR is sufficiently high and the delay is chosen carefully. This assumption defines a scenario of very high practical importance because it refers to the cases where the use of the MMSE linear equalizer seems most suitable. If this assumption does not hold, then a more complicated (and, probably, more computationally demanding) equalizer structure, e.g., a nonlinear equalizer, seems necessary. Under this assumption, the vector $\mathbf{r}$, defined in (4.30), becomes "small." More specifically, it has been proved in [18, eq. (29)], that $\|\mathbf{r}\|_2^2 \leq \text{MMSE}$, which implies that $\|\mathbf{r}\|_2 = \mathcal{O}\left(\sqrt{\text{MMSE}}\right)$. Thus, terms that involve matrix $\mathbf{R}$, which is constructed by vector $\mathbf{r}$, are "small" in comparison with terms that involve matrix $\mathbf{G}$. Consequently, terms $\mathbf{T}_1$ and $\mathbf{T}_3(n)$ of (4.32a) and (4.32c), respectively, can be approximated as

$$\mathbf{T}_1 \approx \text{tr}\left(\mathbf{R}_z^{-1}\mathbf{G}\mathbf{C}'^{*}\mathbf{G}^H\right) \tag{4.33}$$

and

$$\mathbf{T}_3(n) \approx -2\sigma_{\Delta\omega}^2 \,\text{Re}\left\{\mathbf{h}'^T\mathbf{A}^T\mathbf{K}'\mathbf{A}^*(\mathbf{A}^H\mathbf{A})^{-T}\mathbf{G}^H\mathbf{R}_z^{-1}\mathbf{D}'_{n:n-M}\mathbf{H}'\mathbf{e}_d\right\}. \tag{4.34}$$

4.4.2 Average EMSE

A cost function that is more relevant for TS design is the EMSE time-average, across the time instances that correspond to the unknown transmitted data, defined as [4]

$$\begin{aligned}\mathrm{EMSE}(\hat{\mathbf{f}},\hat{\omega}) &\triangleq \frac{1}{n_1-1}\sum_{n=d+1}^{n_1+d-1}\mathrm{EMSE}_n(\hat{\mathbf{f}},\hat{\omega}) + \frac{1}{N-n_2}\sum_{n=n_2+d+1}^{N+d}\mathrm{EMSE}_n(\hat{\mathbf{f}},\hat{\omega}) \\ &= \mathbf{T}_1 + \frac{1}{n_1-1}\sum_{n=d+1}^{n_1+d-1}\left(\mathbf{T}_2(n)+\mathbf{T}_3(n)\right) + \frac{1}{N-n_2}\sum_{n=n_2+d+1}^{N+d}\left(\mathbf{T}_2(n)+\mathbf{T}_3(n)\right).\end{aligned} \tag{4.35}$$

If we write matrix $\mathbf{D}'_{n:n-M}$ as

$$\mathbf{D}'_{n:n-M} = (n-\xi)\,\mathbf{I}_{M+1} - \mathbf{D}_M \tag{4.36}$$

then terms $\mathbf{T}_2(n)$ of (4.32b) and $\mathbf{T}_3(n)$ of (4.34) can be expressed as

$$\mathbf{T}_2(n) = \sigma^2_{\Delta\omega}\left[(n-\xi)^2\,\mathrm{Re}\left\{\mathbf{f}^H\mathbf{H}'\mathbf{e}_d\right\} - 2(n-\xi)\,\mathrm{Re}\left\{\mathbf{f}^H\mathbf{D}_M\mathbf{H}'\mathbf{e}_d\right\} + \mathrm{Re}\left\{\mathbf{f}^H\mathbf{D}_M^2\mathbf{H}'\mathbf{e}_d\right\}\right] \tag{4.37}$$

$$\mathbf{T}_3(n) \approx -2\sigma^2_{\Delta\omega}\,\mathrm{Re}\left\{\mathbf{h}'^T\mathbf{A}^T\mathbf{K}'\mathbf{A}^*(\mathbf{A}^H\mathbf{A})^{-T}\mathbf{G}^H\mathbf{R}_z^{-1}\left((n-\xi)\,\mathbf{I}_{M+1}-\mathbf{D}_M\right)\mathbf{H}'\mathbf{e}_d\right\}. \tag{4.38}$$

If we define

$$c_1 \triangleq \frac{1}{n_1-1}\sum_{n=d+1}^{n_1+d-1} n^2 + \frac{1}{N-n_2}\sum_{n=n_2+d+1}^{N+d} n^2 \tag{4.39}$$

and

$$c_2 \triangleq \frac{1}{n_1-1}\sum_{n=d+1}^{n_1+d-1} n + \frac{1}{N-n_2}\sum_{n=n_2+d+1}^{N+d} n \tag{4.40}$$

then it is easy to show that

$$\begin{aligned}\mathbf{T}_2 &\triangleq \frac{1}{n_1-1}\sum_{n=d+1}^{n_1+d-1}\mathbf{T}_2(n) + \frac{1}{N-n_2}\sum_{n=n_2+d+1}^{N+d}\mathbf{T}_2(n) \\ &= \sigma^2_{\Delta\omega}\Big[\underbrace{\left(c_1-2\,c_2\xi+2\xi^2\right)\mathrm{Re}\{\mathbf{f}^H\mathbf{H}'\mathbf{e}_d\}}_{t_{21}}\underbrace{-2(c_2-2\xi)\mathrm{Re}\{\mathbf{f}^H\mathbf{D}_M\mathbf{H}'\mathbf{e}_d\}}_{t_{22}}\underbrace{+2\mathrm{Re}\{\mathbf{f}^H\mathbf{D}_M^2\mathbf{H}'\mathbf{e}_d\}}_{t_{23}}\Big]\end{aligned} \tag{4.41}$$

and

$$\begin{aligned}\mathbf{T}_3 &\triangleq \frac{1}{n_1-1}\sum_{n=d+1}^{n_1+d-1}\mathbf{T}_3(n) + \frac{1}{N-n_2}\sum_{n=n_2+d+1}^{N+d}\mathbf{T}_3(n) \\ &\approx -2\sigma^2_{\Delta\omega}\,\mathrm{Re}\Big\{\mathbf{h}'^T\mathbf{A}^T\mathbf{K}'\mathbf{A}^*(\mathbf{A}^H\mathbf{A})^{-T}\mathbf{G}^H\mathbf{R}_z^{-1}\Big(\underbrace{(c_2-2\,\xi)}_{t_{31}}\,\mathbf{I}_{M+1}-2\,\mathbf{D}_M\Big)\mathbf{H}'\mathbf{e}_d\Big\}.\end{aligned} \tag{4.42}$$

Obviously, both $\mathbf{T}_2$ and $\mathbf{T}_3$ depend on ξ. It turns out that there does *not* exist a *unique channel independent* ξ that is optimal, i.e., attains minimum EMSE. If we put $\xi = \frac{c_2}{2}$,[3] then term t_{21} is minimized[4] and terms t_{22} and t_{31} vanish. In the sequel, we adopt this simple choice (however, we do not claim optimality, in general). Then, if we define

$$\begin{aligned} c &\triangleq \left(c_1 - \frac{c_2^2}{2}\right) \\ &= \frac{1}{6}\Big(N(N+N_{\text{tr}}) + N_{\text{tr}}^2 + 4d(d-L-1) + L(L+2)\Big) = \mathcal{O}\left(N^2\right). \end{aligned} \tag{4.43}$$

it can be shown that

$$\mathbf{T}_2 = \sigma_{\Delta\omega}^2 \left[\, c \operatorname{Re}\{\mathbf{f}^H \mathbf{H}' \mathbf{e}_d\} + 2 \operatorname{Re}\{\mathbf{f}^H \mathbf{D}_M^2 \mathbf{H}' \mathbf{e}_d\}\right] \tag{4.44}$$

and

$$\mathbf{T}_3 \approx 4\,\sigma_{\Delta\omega}^2 \operatorname{Re}\left\{\mathbf{h}'^T \mathbf{A}^T \mathbf{K}' \mathbf{A}^* (\mathbf{A}^H \mathbf{A})^{-T} \mathbf{G}^H \mathbf{R}_z^{-1} \mathbf{D}_M \mathbf{H}' \mathbf{e}_d\right\}. \tag{4.45}$$

Thus, the EMSE is approximately equal to the sum of the three terms in (4.33), (4.44) and (4.45). In the next subsection, we study these terms in detail.

4.4.3 Comparison of terms $\mathbf{T}_1$, $\mathbf{T}_2$ and $\mathbf{T}_3$

Term $\mathbf{T}_1$

From (4.33) and the definition of $\mathbf{G}$, we obtain

$$\mathbf{T}_1 \approx \operatorname{tr}\left(\mathbf{C}'^* \mathbf{G}^H \mathbf{R}_z^{-1} \mathbf{G}\right) = \operatorname{tr}\left(\mathbf{C}'^* \mathbf{F}^* \mathbf{H}'^H \mathbf{R}_z^{-1} \mathbf{H}' \mathbf{F}^T\right). \tag{4.46}$$

Using the facts (i) $\mathbf{A} \geq \mathbf{B}$ implies that $\mathbf{A}^{-1} \leq \mathbf{B}^{-1}$ [11, p. 471] and (ii) $\mathbf{P}_{\mathbf{H}'^H} \leq \mathbf{I}_{L+M+1}$, where $\mathbf{P}_{\mathbf{H}'^H}$ is the projector onto the column space of $\mathbf{H}'^H$, we obtain

$$\mathbf{H}'^H \left(\mathbf{H}'\mathbf{H}'^H + \sigma_w^2 \mathbf{I}\right)^{-1} \mathbf{H}' \leq \mathbf{H}'^H \left(\mathbf{H}'\mathbf{H}'^H\right)^{-1} \mathbf{H}' = \mathbf{P}_{\mathbf{H}'^H} \leq \mathbf{I}_{M+L+1}. \tag{4.47}$$

Using (4.47) and the property in (7) we obtain (recall the definition of $\mathbf{R}_z$)

$$\begin{aligned} \mathbf{T}_1 &\approx \operatorname{tr}\left(\mathbf{C}'^* \mathbf{F}^* \mathbf{H}'^H \mathbf{R}_z^{-1} \mathbf{H}' \mathbf{F}^T\right) \\ &\lessapprox \operatorname{tr}\left(\mathbf{C}'^* \mathbf{F}^* \mathbf{F}^T\right) \leq \lambda_{\max}\left(\mathbf{C}'\right) \operatorname{tr}\left(\mathbf{F}^* \mathbf{F}^T\right) = \lambda_{\max}\left(\mathbf{C}'\right) \|\mathbf{F}\|_F^2 \\ &= \lambda_{\max}\left(\mathbf{C}'\right)(L+1)\|\mathbf{f}\|_2^2. \end{aligned} \tag{4.48}$$

[3] This implies that the training block is placed close to the middle of the packet.
[4] We shall see that t_{21} is the most significant EMSE term.

Using asymptotic expansions, it is proved in Appendix 4B that, if $\mathbf{A}^H\mathbf{A}$ is invertible (and not very ill-conditioned), then the first term of $\mathbf{C}'$ in (4.16) is much larger than the second. In this case,

$$\lambda_{\max}\left(\mathbf{C}'\right) \approx \frac{\sigma_w^2}{\lambda_{\min}(\mathbf{A}^H\mathbf{A})} \tag{4.49}$$

and

$$\mathbf{T}_1 \lessapprox \frac{(L+1)\|\mathbf{f}\|_2^2\,\sigma_w^2}{\lambda_{\min}(\mathbf{A}^H\mathbf{A})}. \tag{4.50}$$

Term $\mathbf{T}_2$

Term $\mathrm{Re}\{\mathbf{f}^H\mathbf{H}'\mathbf{e}_d\}$ is the $(d+1)$-st coefficient of the combined (channel-equalizer) impulse response. Using the definition of $\mathbf{f}$ in (4.21) and expression (4.47), it can be shown that $\mathrm{Re}\{\mathbf{f}^H\mathbf{H}'\mathbf{e}_d\}$ is always smaller than 1, and, under our small MMSE assumption, it is very close to 1. Thus, $t_{21} \simeq c$. On the other hand, using the definition of $\mathbf{f}$ in (4.21), the sub-multiplicative property of the matrix norms, and the singular value decomposition (SVD) of $\mathbf{H}'$, it can be shown that $t_{23} = 2\mathrm{Re}\{\mathbf{f}^H\mathbf{D}_M^2\mathbf{H}'\mathbf{e}_d\} \leq 2M^2k_2(\mathbf{H}')$. If N is sufficiently large and $\mathbf{H}'$ is not very ill-conditioned, then $t_{21} \gg t_{23}$ and

$$\mathbf{T}_2 \simeq c\,\sigma_{\Delta\omega}^2. \tag{4.51}$$

Term $\mathbf{T}_3$

Using the SVD of $\mathbf{A}$, it can be easily shown that (we prove it in Appendix 4C)

$$\|\mathbf{A}^T\mathbf{K}'\mathbf{A}^*(\mathbf{A}^H\mathbf{A})^{-T}\|_2 \leq \frac{\sigma_{\max}(\mathbf{A})}{\sigma_{\min}(\mathbf{A})}\,\|\mathbf{K}'\|_2 = \frac{R}{2}\,k_2(\mathbf{A}). \tag{4.52}$$

In the same manner, it can be shown that (recall that $\mathbf{G} = \mathbf{H}'\mathbf{F}^T$) (we prove it in Appendix 4C)

$$\|\mathbf{G}^H\mathbf{R}_z^{-1}\mathbf{D}_M\mathbf{H}'\mathbf{e}_d\|_2 \leq \|\mathbf{F}\|_2\frac{\sigma_{\max}(\mathbf{H}')}{\sigma_{\min}(\mathbf{H}')}M = M\,\|\mathbf{F}\|_2 k_2(\mathbf{H}'). \tag{4.53}$$

Thus,

$$\mathbf{T}_3 \leq \left(2\,MR\,\|\mathbf{F}\|_2 k_2(\mathbf{A})k_2(\mathbf{H}')\right)\,\sigma_{\Delta\omega}^2. \tag{4.54}$$

Comparison of $\mathbf{T}_2$ and $\mathbf{T}_3$

If N is sufficiently large and $\mathbf{A}$ and $\mathbf{H}'$ are not very ill-conditioned, then, from (4.51) and (4.54), we conclude that $\mathbf{T}_2 \gg \mathbf{T}_3$.

Comparison of $\mathbf{T}_1$ and $\mathbf{T}_2$

Using [2, eq. (10)], we can derive the following asymptotic expression

$$\sigma^2_{\Delta\omega} \approx \frac{6\,\sigma_w^2}{R^2\,\mathbf{h}^H\mathbf{A}^H\mathbf{A}\mathbf{h}}. \tag{4.55}$$

Thus

$$\mathbf{T}_2 \approx \frac{6\,c\,\sigma_w^2}{R^2\mathbf{h}^H\mathbf{A}^H\mathbf{A}\mathbf{h}}. \tag{4.56}$$

Using (4.50) and (4.56), we derive the following approximate bound

$$\begin{aligned}\frac{\mathbf{T}_1}{\mathbf{T}_2} &\lessapprox \frac{(L+1)R^2\,\|\mathbf{f}\|_2^2\,\mathbf{h}^H\mathbf{A}^H\mathbf{A}\mathbf{h}}{6\,c\,\lambda_{\min}(\mathbf{A}^H\mathbf{A})} \le \frac{(L+1)R^2\|\mathbf{f}\|_2^2\lambda_{\max}(\mathbf{A}^H\mathbf{A})}{6\,c\,\lambda_{\min}(\mathbf{A}^H\mathbf{A})} \\ &= k_2(\mathbf{A}^H\mathbf{A})(L+1)\|\mathbf{f}\|_2^2\,\alpha\end{aligned} \tag{4.57}$$

where $\alpha \triangleq \frac{R^2}{6\,c} = \mathcal{O}\left(\frac{R^2}{N^2}\right)$. We note that bound (4.57) becomes approximate equality if $\mathbf{A}^H\mathbf{A} = R\mathbf{I}_{L+1}$, i.e., the training sequence is orthogonal.

Thus, if α is sufficiently small, i.e., R is sufficiently small with respect to N (recall that $R = N_{\text{tr}} - L$), and $\mathbf{A}$ is not very ill-conditioned, then term $\mathbf{T}_1$ is much smaller than $\mathbf{T}_2$ and an approximate expression for the EMSE is given by

$$\boxed{\text{EMSE}(\hat{\mathbf{f}}, \hat{\omega}) \simeq c\,\sigma^2_{\Delta\omega}.} \tag{4.58}$$

That is, the EMSE is approximately proportional to the CFO estimation error variance, with the proportionality factor being *independent* of the training sequence. Consequently, training sequences that are optimal for CFO estimation seem also very good candidates for *joint* channel and CFO estimation. Optimal training sequence design for CFO estimation has been extensively studied; see, for example, [2], [16] and [21]. This topic is beyond the scope of this paper.

Remark 2: If we consider perfect channel estimation and errors due to imperfect CFO cancellation, we can easily show that the resulting EMSE is approximately equal to term $\mathbf{T}_2$ (for sufficiently large N). Thus, expression (4.58) implies that, under the small ideal MMSE assumption, the performance degradation caused by the imperfectly canceled CFO is much more significant than that caused by the mismatched equalizer. □

4.5 Simulation Results

In this section, we check our theoretical results with simulations. We present results for channel order $L = 3$ (the channel coefficients are given in Table III), equalizer order $M = 8$,

Table III

Channel Impulse Response **h**

0.0010-0.0311*j	-0.0066+0.0825*j	-0.9451+0.3051*j	-0.0144-0.0757*j

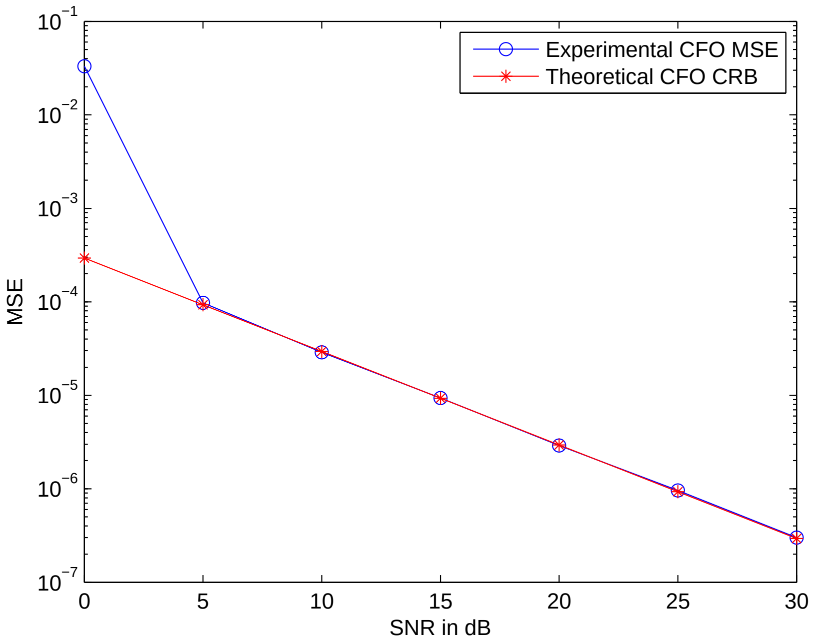

Fig. 4.1. Experimental MSE of the ML CFO estimator and the CFO CRB for the channel realization in Table III.

delay $d = 6$, packet length $N = 300$ and TS length $N_{\text{tr}} = 30$ ($\alpha = 0.0073$). The data symbols are i.i.d., BPSK symbols. The training symbols, which are also i.i.d. BPSK, have been placed close to the middle of the transmitted packet, i.e., $\xi = \frac{c_2}{2}$. The results presented in the sequel have been derived by using the binary sequence that corresponds to the hexadecimal number $198153E6$ (we have observed, through exhaustive search, that this sequence has "good" performance for both CFO and joint channel and CFO estimation).

In Figures 4.1 and 4.2, we compare the experimental MSE of the ML CFO and channel estimators (in (4.7) and (4.15), respectively) with the corresponding quantities derived by the CRBs in (4.9) and (4.16). We observe that the experimental and theoretical results practically coincide for SNR higher than 5 dB. Thus, the assumption that the *finite sample* CRBs are met is valid in this case.

In Fig. 4.3, we plot the experimentally computed EMSE and the EMSE theoretical

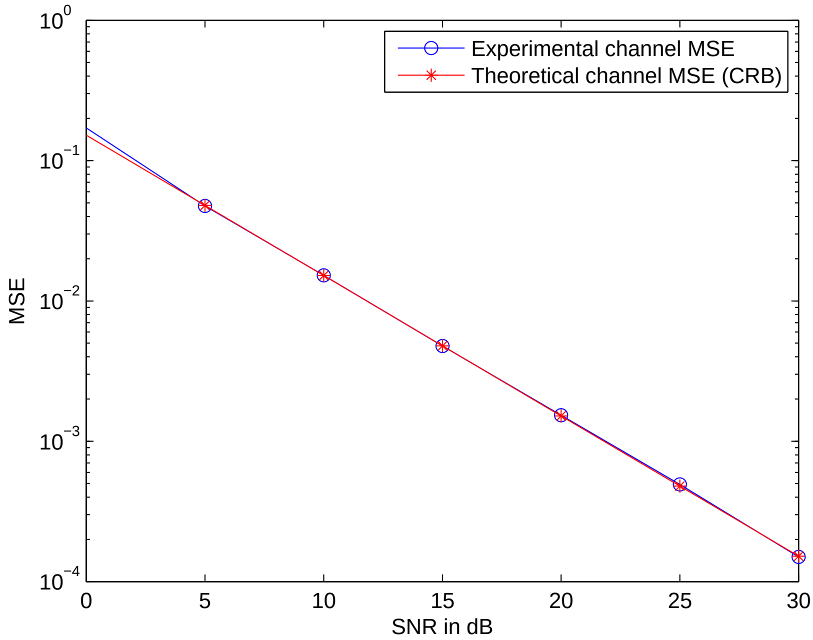

Fig. 4.2. Experimental MSE of the ML channel estimator and theoretical channel MSE (CRB) for the channel realization in Table III.

approximation in (4.31) versus the time instances n, for SNR equal to 30 dB (as mentioned above, we do not compute the EMSE for the known training symbols). We observe that the expression (4.31) is a very good approximation to the true time-dependent EMSE. Moreover, we observe that the EMSE increases as we move away from the training symbol positions.

In Fig. 4.4, we compare the experimentally computed time-average EMSE and the time-average of the EMSE theoretical approximation in (4.31). We observe that our theoretical result practically coincides with the true EMSE for SNR higher than 15 dB.

In Fig. 4.5, we present the time-average of the theoretical terms $\mathbf{T}_1$ of (4.32a) and $\mathbf{T}_3$ of (4.32c) and their corresponding high SNR approximations (4.33) and (4.34), which ignore terms that involve matrix $\mathbf{R}$. Since the approximations practically coincide with the time-averages for sufficiently high SNR, we conclude that terms that involve matrix $\mathbf{R}$ are indeed negligible compared to terms that involve matrix $\mathbf{G}$.

In Fig. 4.6, we present the time-averages of the three EMSE terms $\mathbf{T}_1$, $\mathbf{T}_2$ and $\mathbf{T}_3$ in (4.33), (4.44) and (4.45), respectively, and their sum, i.e., the approximate EMSE. We observe that $\mathbf{T}_2$ is very close to the approximate EMSE, while terms $\mathbf{T}_1$ and $\mathbf{T}_3$ are much smaller.

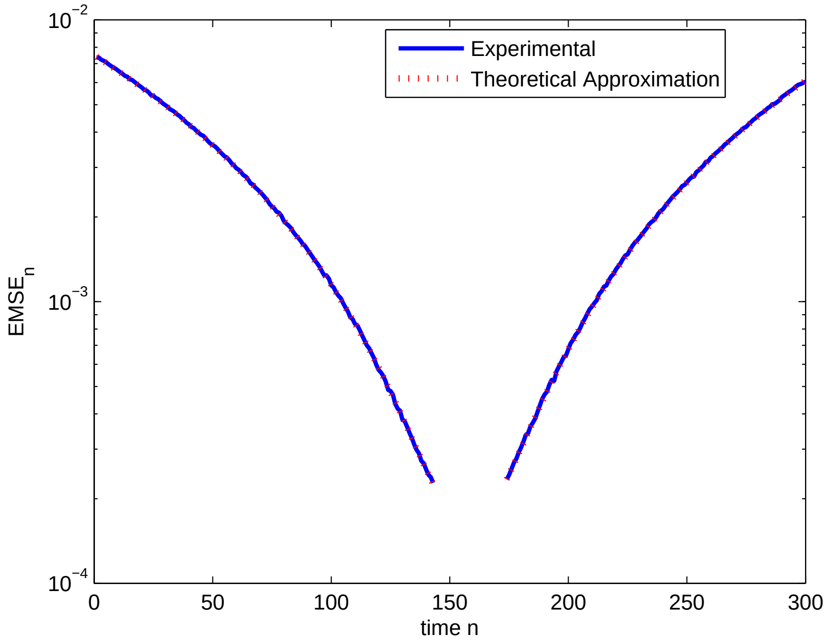

Fig. 4.3. Experimentally computed EMSE and the EMSE theoretical approximation in (4.31) versus n.

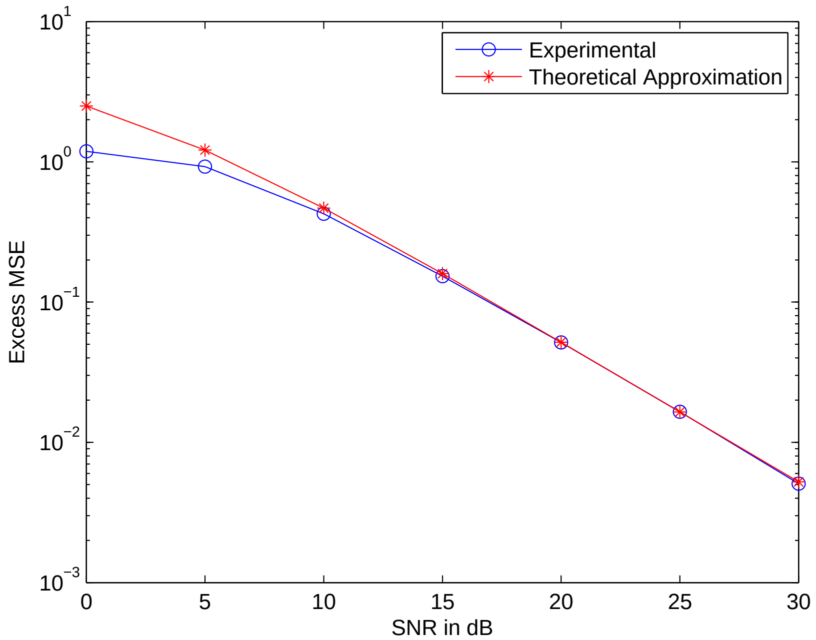

Fig. 4.4. Experimentally computed EMSE and the EMSE theoretical approximation in (4.31).

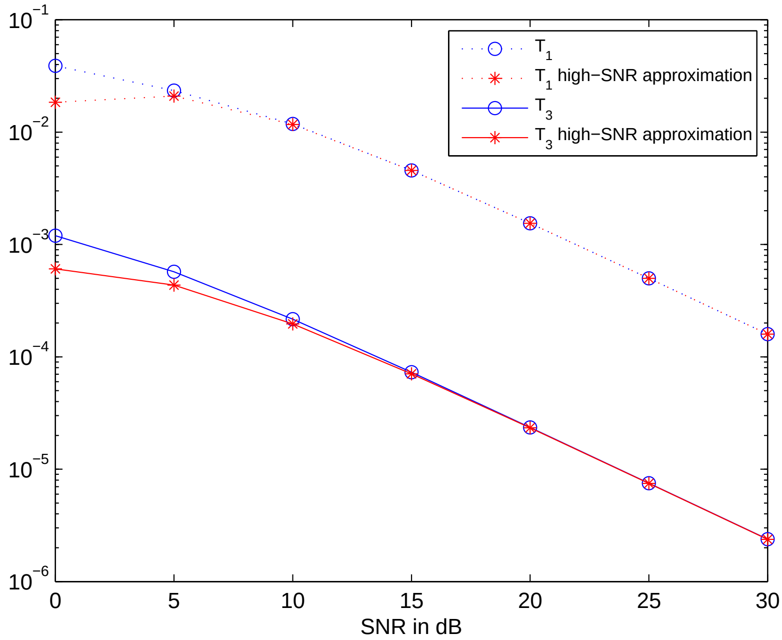

Fig. 4.5. Terms $\mathbf{T}_1$ and $\mathbf{T}_3$ and their high SNR approximations (i.e., terms involving $\mathbf{R}$ are neglected).

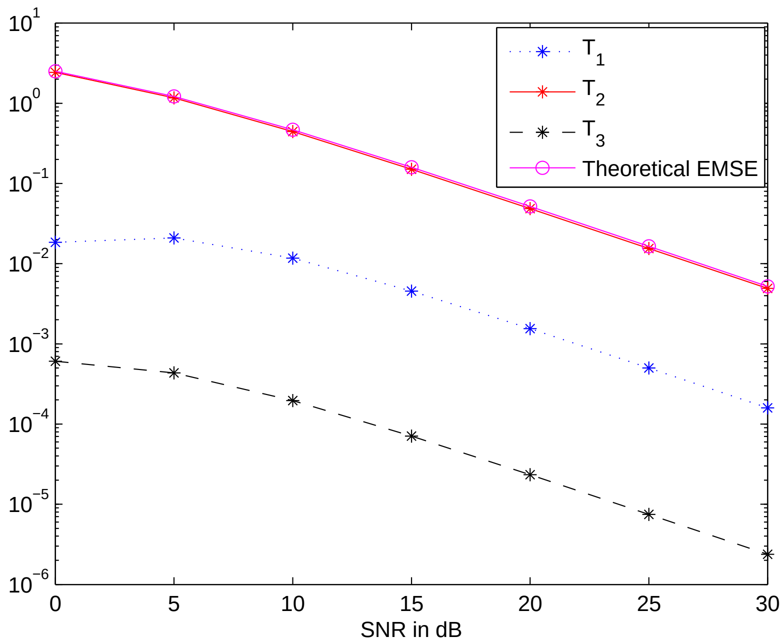

Fig. 4.6. Final expressions for terms $\mathbf{T}_1$, $\mathbf{T}_2$ and $\mathbf{T}_3$ in (4.33), (4.44) and (4.45) and their sum.

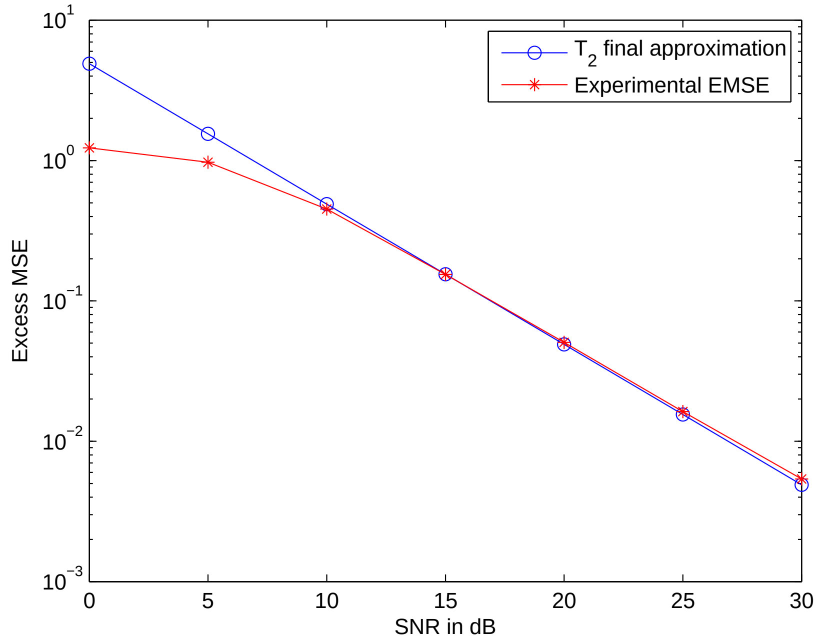

Fig. 4.7. Final EMSE theoretical expression in (4.58).

Finally, in Fig. 4.7 we plot the experimental EMSE and the simple EMSE approximation (4.58). We observe that the very simple and informative expression of (4.58) is indeed a very good EMSE approximation.

We also present results for $\alpha = 0.0261$, i.e., we use the same channel and the same TS, but we reduce the packet length to $N = 150$. A bigger value for parameter α means that the ratio of $\mathbf{T}_1$ to $\mathbf{T}_2$ increases, and, thus, the final approximation in (4.58) becomes less accurate. More specifically, in Fig. 4.8, we plot the two significant EMSE terms, $\mathbf{T}_1$ and $\mathbf{T}_2$, for $\alpha = 0.0073$ and $\alpha = 0.0261$. We note that term $\mathbf{T}_1$ *remains the same* in both cases, since it depends only on the first term of $\mathbf{C}'$ in (4.16) (see (4.33)). On the other hand, when we reduce N, term $\mathbf{T}_2$ decreases, since the CFO estimation error variance remains the same and $\mathcal{C}$ decreases (see (4.51)). Thus, the final approximation in (4.58) is less accurate for $\alpha = 0.0261$ than for $\alpha = 0.0073$.

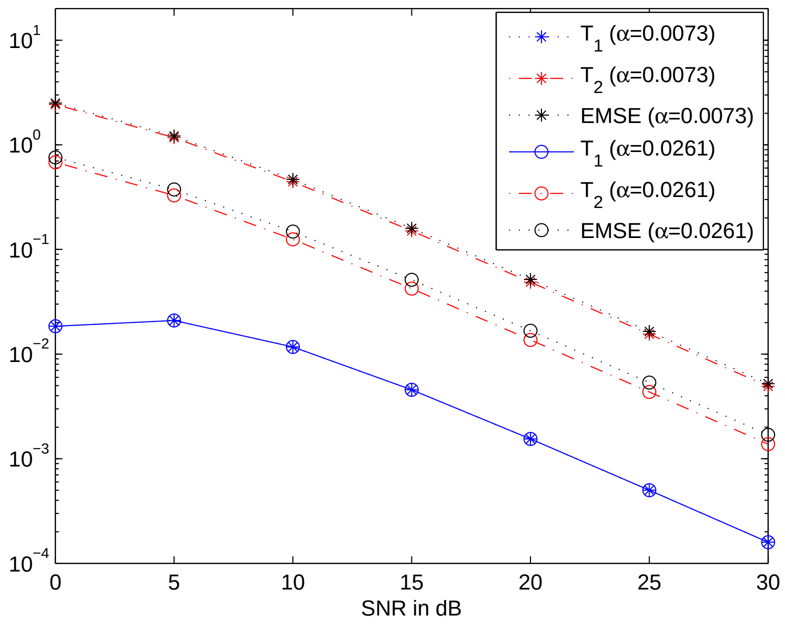

Fig. 4.8. Terms $\mathbf{T}_1$ and $\mathbf{T}_2$ and the EMSE for $\alpha = 0.0073$ and $\alpha = 0.0261$.

4.6 Conclusion

We considered optimal TS design for joint channel and CFO estimation in frequency-selective single-carrier systems with an MMSE linear equalizer. Performance degradation is due to the fact that a mismatched MMSE linear equalizer is applied to channel output samples with imperfectly canceled CFO. We uncovered that, in many cases of high practical importance, the imperfectly canceled CFO is the main cause of the performance degradation. In these cases, the EMSE is approximately proportional to the CFO estimation error variance, with the proportionality coefficient being independent of the TS, implying that optimal TS design for CFO estimation is also highly relevant for *joint* CFO and channel estimation.

Appendix 4A

Proof of Proposition 1: If we use the definition of the equalizer mismatch, $\Delta\mathbf{f} \triangleq \hat{\mathbf{f}} - \mathbf{f}$, in (4.27), we get

$$\begin{aligned}\mathrm{MSE}_n(\hat{\mathbf{f}}, \hat{\omega}) = &\underbrace{\mathbf{f}^H \left(\boldsymbol{\Gamma}_{n:n-M}(-\Delta\omega)\mathbf{H}'\mathbf{H}'^H\boldsymbol{\Gamma}^H_{n:n-M}(-\Delta\omega) + \sigma_w^2\mathbf{I}_{M+1}\right)\mathbf{f}}_{t_1} \\ &\underbrace{+\Delta\mathbf{f}^H \left(\boldsymbol{\Gamma}_{n:n-M}(-\Delta\omega)\mathbf{H}'\mathbf{H}'^H\boldsymbol{\Gamma}^H_{n:n-M}(-\Delta\omega) + \sigma_w^2\mathbf{I}_{M+1}\right)\Delta\mathbf{f}}_{t_2} \\ &\underbrace{+2\,\mathrm{Re}\{\mathbf{f}^H \left(\boldsymbol{\Gamma}_{n:n-M}(-\Delta\omega)\mathbf{H}'\mathbf{H}'^H\boldsymbol{\Gamma}^H_{n:n-M}(-\Delta\omega) + \sigma_w^2\mathbf{I}_{M+1}\right)\Delta\mathbf{f}\}}_{t_3} \\ &\underbrace{-2\,\mathrm{Re}\{\mathrm{e}^{j\Delta\omega\xi}\mathbf{f}^H\boldsymbol{\Gamma}_{n:n-M}(-\Delta\omega)\mathbf{H}'\mathbf{e}_d\}}_{t_4}\underbrace{-2\,\mathrm{Re}\{\mathrm{e}^{j\Delta\omega\xi}\Delta\mathbf{f}^H\boldsymbol{\Gamma}_{n:n-M}(-\Delta\omega)\mathbf{H}'\mathbf{e}_d\}}_{t_5}+1.\end{aligned} \tag{4.59}$$

We define matrix $\boldsymbol{\Gamma}'_{n:n-M}(-\Delta\omega) \triangleq \mathrm{e}^{j\Delta\omega\xi}\boldsymbol{\Gamma}_{n:n-M}(-\Delta\omega) = \mathrm{diag}(\mathrm{e}^{-j\Delta\omega(n-\xi)}, \ldots, \mathrm{e}^{-j\Delta\omega(n-M-\xi)})$. Using the expression $\exp(x) = 1 + x + \frac{x^2}{2} + O(x^3)$, we obtain

$$\boldsymbol{\Gamma}_{n:n-M}(-\Delta\omega) = \mathbf{I}_{M+1} - j\Delta\omega\mathbf{D}_{n:n-M} - \frac{1}{2}\Delta\omega^2\mathbf{D}^2_{n:n-M} + \mathcal{O}_p\left(\frac{n^3\sigma_w^3}{R^{9/2}}\right) \tag{4.60}$$

$$\boldsymbol{\Gamma}'_{n:n-M}(-\Delta\omega) = \mathbf{I}_{M+1} - j\Delta\omega\mathbf{D}'_{n:n-M} - \frac{1}{2}\Delta\omega^2\mathbf{D}'^2_{n:n-M} + \mathcal{O}_p\left(\frac{n^3\sigma_w^3}{R^{9/2}}\right) \tag{4.61}$$

where $\mathbf{D}_{n:n-M} \triangleq \mathrm{diag}(n, \ldots, n-M)$, $\mathbf{D}'_{n:n-M} \triangleq \mathrm{diag}(n-\xi, \ldots, n-M-\xi)$ and $R \triangleq N_{\mathrm{tr}}-L$.

We will write analytically the five terms defined in (4.59), using (4.60) and (4.61). We will also take the expected value of each term with respect to $\Delta\mathbf{h}'$ and $\Delta\omega$.

Term t_1

Using (4.60) we obtain

$$\begin{aligned}\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}[t_1] &= \mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[\mathbf{f}^H\Big(\big(\overset{(1)}{\mathbf{I}}_{M+1} - j\Delta\omega\overset{(2)}{\mathbf{D}}_{n:n-M} - \overset{(3)}{\frac{1}{2}\Delta\omega^2\mathbf{D}^2_{n:n-M}}\big)\mathbf{H}'\mathbf{H}'^H \\ &\quad \times \big(\overset{(1)}{\mathbf{I}}_{M+1} + j\Delta\omega\overset{(2)}{\mathbf{D}}_{n:n-M} - \overset{(3)}{\frac{1}{2}\Delta\omega^2\mathbf{D}^2_{n:n-M}}\big) + \sigma_w^2\mathbf{I}_{M+1}\Big)\mathbf{f}\Big] \\ &= \mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[\mathbf{f}^H\big(\overset{(11)}{\mathbf{H}'\mathbf{H}'^H} + \sigma_w^2\mathbf{I}_{M+1}\big)\mathbf{f} + \overset{(12)}{j\Delta\omega\mathbf{f}^H\mathbf{H}'\mathbf{H}'^H\mathbf{D}_{n:n-M}\mathbf{f}} \\ &\quad - \overset{(13)}{\frac{1}{2}\Delta\omega^2\mathbf{f}^H\mathbf{H}'\mathbf{H}'^H\mathbf{D}^2_{n:n-M}\mathbf{f}} - \overset{(21)}{j\Delta\omega\mathbf{f}^H\mathbf{D}_{n:n-M}\mathbf{H}'\mathbf{H}'^H\mathbf{f}} \\ &\quad + \overset{(22)}{\Delta\omega^2\mathbf{f}^H\mathbf{D}_{n:n-M}\mathbf{H}'\mathbf{H}'^H\mathbf{D}_{n:n-M}\mathbf{f}} + \overset{(23)}{j\frac{1}{2}\Delta\omega^3\mathbf{f}^H\mathbf{D}_{n:n-M}\mathbf{H}'\mathbf{H}'^H\mathbf{D}^2_{n:n-M}\mathbf{f}} \\ &\quad - \overset{(31)}{\frac{1}{2}\Delta\omega^2\mathbf{f}^H\mathbf{D}^2_{n:n-M}\mathbf{H}'\mathbf{H}'^H\mathbf{f}} - \overset{(32)}{j\frac{1}{2}\Delta\omega^3\mathbf{f}^H\mathbf{D}^2_{n:n-M}\mathbf{H}'\mathbf{H}'^H\mathbf{D}_{n:n-M}\mathbf{f}} \\ &\quad + \overset{(33)}{\frac{1}{4}\Delta\omega^4\mathbf{f}^H\mathbf{D}^2_{n:n-M}\mathbf{H}'\mathbf{H}'^H\mathbf{D}^2_{n:n-M}\mathbf{f}}\Big].\end{aligned} \tag{4.62}$$

Using that $\mathcal{E}_{\Delta\omega}[\Delta\omega] = 0$, $\mathcal{E}_{\Delta\omega}[\Delta\omega^3] = 0$ (the ML estimator is practically unbiased and Gaussian) and that the term proportional to $\Delta\omega^4$ is practically equal to zero, we obtain

$$\begin{aligned}\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}[t_1] &= \mathbf{f}^H\mathbf{R}_z\mathbf{f} + \sigma^2_{\Delta\omega}\mathbf{f}^H\mathbf{D}_{n:n-M}\mathbf{H}'\mathbf{H}'^H\mathbf{D}_{n:n-M}\mathbf{f} \\ &\quad - \frac{1}{2}\sigma^2_{\Delta\omega}\mathbf{f}^H\mathbf{H}'\mathbf{H}'^H\mathbf{D}^2_{n:n-M}\mathbf{f} - \frac{1}{2}\sigma^2_{\Delta\omega}\mathbf{f}^H\mathbf{D}^2_{n:n-M}\mathbf{H}'\mathbf{H}'^H\mathbf{f}.\end{aligned} \tag{4.63}$$

We write the last three terms of (4.63) using that $\mathbf{D}_{n:n-M} = n\mathbf{I}_{M+1} - \mathbf{D}_M$ as

1.

$$\begin{aligned}\sigma^2_{\Delta\omega}\mathbf{f}^H\mathbf{D}_{n:n-M}\mathbf{H}'\mathbf{H}'^H\mathbf{D}_{n:n-M}\mathbf{f} &= n^2\,\sigma^2_{\Delta\omega}\mathbf{f}^H\mathbf{H}'\mathbf{H}'^H\mathbf{f} + \sigma^2_{\Delta\omega}\mathbf{f}^H\mathbf{D}_M\mathbf{H}'\mathbf{H}'^H\mathbf{D}_M\mathbf{f} \\ &\quad - n\,\sigma^2_{\Delta\omega}\mathbf{f}^H\mathbf{D}_M\mathbf{H}'\mathbf{H}'^H\mathbf{f} - n\,\sigma^2_{\Delta\omega}\mathbf{f}^H\mathbf{H}'\mathbf{H}'^H\mathbf{D}_M\mathbf{f}.\end{aligned} \tag{4.64}$$

2.

$$\begin{aligned}-\frac{1}{2}\,\sigma^2_{\Delta\omega}\mathbf{f}^H\mathbf{H}'\mathbf{H}'^H\mathbf{D}^2_{n:n-M}\mathbf{f} &= -\frac{1}{2}\,n^2\,\sigma^2_{\Delta\omega}\mathbf{f}^H\mathbf{H}'\mathbf{H}'^H\mathbf{f} + n\,\sigma^2_{\Delta\omega}\mathbf{f}^H\mathbf{H}'\mathbf{H}'^H\mathbf{D}_M\mathbf{f} \\ &\quad - \frac{1}{2}\sigma^2_{\Delta\omega}\mathbf{f}^H\mathbf{H}'\mathbf{H}'^H\mathbf{D}^2_M\mathbf{f}.\end{aligned} \tag{4.65}$$

3.

$$\begin{aligned}-\frac{1}{2}\,\sigma^2_{\Delta\omega}\mathbf{f}^H\mathbf{D}^2_{n:n-M}\mathbf{H}'\mathbf{H}'^H\mathbf{f} &= -\frac{1}{2}\,n^2\,\sigma^2_{\Delta\omega}\mathbf{f}^H\mathbf{H}'\mathbf{H}'^H\mathbf{f} + n\,\sigma^2_{\Delta\omega}\mathbf{f}^H\mathbf{D}_M\mathbf{H}'\mathbf{H}'^H\mathbf{f} \\ &\quad - \frac{1}{2}\sigma^2_{\Delta\omega}\mathbf{f}^H\mathbf{D}^2_M\mathbf{H}'\mathbf{H}'^H\mathbf{f}.\end{aligned} \tag{4.66}$$

Using (4.64)-(4.66) in (4.63) we obtain

$$\begin{aligned}\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}[t_1] &= \mathbf{f}^H\mathbf{R}_z\mathbf{f} + \sigma^2_{\Delta\omega}\mathbf{f}^H\mathbf{D}_M\mathbf{H}'\mathbf{H}'^H\mathbf{D}_M\mathbf{f} \\ &\quad - \frac{1}{2}\sigma^2_{\Delta\omega}\mathbf{f}^H\mathbf{H}'\mathbf{H}'^H\mathbf{D}^2_M\mathbf{f} - \frac{1}{2}\sigma^2_{\Delta\omega}\mathbf{f}^H\mathbf{D}^2_M\mathbf{H}'\mathbf{H}'^H\mathbf{f} \\ &= \mathbf{f}^H\mathbf{R}_z\mathbf{f} + \mathcal{O}\left(\frac{M^2\,\sigma^2_w}{R^3}\right).\end{aligned} \tag{4.67}$$

Term t_2

Using (4.60) we obtain

$$\begin{aligned}\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}[t_2] &= \mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[\Delta\mathbf{f}^H\Big(\big(\overset{(1)}{\mathbf{I}}_{M+1} - j\Delta\omega\overset{(2)}{\mathbf{D}}_{n:n-M} - \overset{(3)}{\frac{1}{2}\Delta\omega^2\mathbf{D}^2_{n:n-M}}\big)\mathbf{H}'\mathbf{H}'^H \\ &\qquad \times\big(\overset{(1)}{\mathbf{I}}_{M+1} + j\Delta\omega\overset{(2)}{\mathbf{D}}_{n:n-M} - \overset{(3)}{\frac{1}{2}\Delta\omega^2\mathbf{D}^2_{n:n-M}}\big) + \sigma^2_w\mathbf{I}_{M+1}\Big)\Delta\mathbf{f}\Big] \\ &= \mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[\overset{(11)}{\Delta\mathbf{f}^H(\mathbf{H}'\mathbf{H}'^H + \sigma^2_w\mathbf{I}_{M+1})\Delta\mathbf{f}} + \overset{(12)}{j\Delta\omega\Delta\mathbf{f}^H\mathbf{H}'\mathbf{H}'^H\mathbf{D}_{n:n-M}\Delta\mathbf{f}} \\ &\quad - \overset{(13)}{\frac{1}{2}\Delta\omega^2\Delta\mathbf{f}^H\mathbf{H}'\mathbf{H}'^H\mathbf{D}^2_{n:n-M}\Delta\mathbf{f}} - \overset{(21)}{j\Delta\omega\Delta\mathbf{f}^H\mathbf{D}_{n:n-M}\mathbf{H}'\mathbf{H}'^H\Delta\mathbf{f}} \\ &\quad + \overset{(22)}{\Delta\omega^2\Delta\mathbf{f}^H\mathbf{D}_{n:n-M}\mathbf{H}'\mathbf{H}'^H\mathbf{D}_{n:n-M}\Delta\mathbf{f}} + \overset{(23)}{j\frac{1}{2}\Delta\omega^3\Delta\mathbf{f}^H\mathbf{D}_{n:n-M}\mathbf{H}'\mathbf{H}'^H\mathbf{D}^2_{n:n-M}\Delta\mathbf{f}} \\ &\quad - \overset{(31)}{\frac{1}{2}\Delta\omega^2\Delta\mathbf{f}^H\mathbf{D}^2_{n:n-M}\mathbf{H}'\mathbf{H}'^H\Delta\mathbf{f}} - \overset{(32)}{j\frac{1}{2}\Delta\omega^3\Delta\mathbf{f}^H\mathbf{D}^2_{n:n-M}\mathbf{H}'\mathbf{H}'^H\mathbf{D}_{n:n-M}\Delta\mathbf{f}} \\ &\quad + \overset{(33)}{\frac{1}{4}\Delta\omega^4\Delta\mathbf{f}^H\mathbf{D}^2_{n:n-M}\mathbf{H}'\mathbf{H}'^H\mathbf{D}^2_{n:n-M}\Delta\mathbf{f}}\Big].\end{aligned} \tag{4.68}$$

We keep second and third-order error terms (*we keep third-order terms just to see the lower-order term we neglect for* t_2). Then

$$\begin{aligned}\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}[t_2] &= \mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[\Delta\mathbf{f}^H\mathbf{R}_z\Delta\mathbf{f} + j\Delta\omega\Delta\mathbf{f}^H\mathbf{H}'\mathbf{H}'^H\mathbf{D}_{n:n-M}\Delta\mathbf{f} \\ &\quad - j\Delta\omega\Delta\mathbf{f}^H\mathbf{D}_{n:n-M}\mathbf{H}'\mathbf{H}'^H\Delta\mathbf{f}\Big].\end{aligned} \tag{4.69}$$

We write the last two terms of (4.69) using that $\mathbf{D}_{n:n-M} = n\mathbf{I}_{M+1} - \mathbf{D}_M$ as

1.

$$\begin{aligned}\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[j\Delta\omega\Delta\mathbf{f}^H\mathbf{H}'\mathbf{H}'^H\mathbf{D}_{n:n-M}\Delta\mathbf{f}\Big] &= \mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[j\,n\,\Delta\omega\Delta\mathbf{f}^H\mathbf{H}'\mathbf{H}'^H\Delta\mathbf{f}\Big] \\ &\quad - \mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[j\Delta\omega\Delta\mathbf{f}^H\mathbf{H}'\mathbf{H}'^H\mathbf{D}_M\Delta\mathbf{f}\Big].\end{aligned} \tag{4.70}$$

2.

$$\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[-j\Delta\omega\Delta\mathbf{f}^H\mathbf{D}_{n:n-M}\mathbf{H}'\mathbf{H}'^H\Delta\mathbf{f}\Big] = \mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[-j\,n\,\Delta\omega\Delta\mathbf{f}^H\mathbf{H}'\mathbf{H}'^H\Delta\mathbf{f}\Big] + \mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[j\Delta\omega\Delta\mathbf{f}^H\mathbf{D}_M\mathbf{H}'\mathbf{H}'^H\Delta\mathbf{f}\Big]. \quad (4.71)$$

Using (4.70)-(4.71) in (4.69) we obtain

$$\begin{aligned}\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}[t_2] &= \mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[\Delta\mathbf{f}^H\mathbf{R}_z\Delta\mathbf{f}\Big] + \mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[-j\Delta\omega\Delta\mathbf{f}^H\mathbf{H}'\mathbf{H}'^H\mathbf{D}_M\Delta\mathbf{f}\Big] \\ &\quad + \mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[j\Delta\omega\Delta\mathbf{f}^H\mathbf{D}_M\mathbf{H}'\mathbf{H}'^H\Delta\mathbf{f}\Big] \\ &= \mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[\Delta\mathbf{f}^H\mathbf{R}_z\Delta\mathbf{f}\Big] + 2\,\mathrm{Re}\Big\{\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[j\Delta\omega\Delta\mathbf{f}^H\mathbf{D}_M\mathbf{H}'\mathbf{H}'^H\Delta\mathbf{f}\Big]\Big\} \\ &= \mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[\Delta\mathbf{f}^H\mathbf{R}_z\Delta\mathbf{f}\Big] + \mathcal{O}\left(\frac{M\,\sigma_w^3}{R^{5/2}}\right).\end{aligned} \quad (4.72)$$

Term t_3

Using (4.60) we obtain

$$\begin{aligned}\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}[t_3] &= 2\,\mathrm{Re}\Big\{\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[\mathbf{f}^H\Big(\big(\overset{(1)}{\mathbf{I}_{M+1}} - j\Delta\omega\overset{(2)}{\mathbf{D}_{n:n-M}} - \overset{(3)}{\tfrac{1}{2}\Delta\omega^2\mathbf{D}^2_{n:n-M}}\big)\mathbf{H}'\mathbf{H}'^H \\ &\quad \times \big(\overset{(1)}{\mathbf{I}_{M+1}} + j\Delta\omega\overset{(2)}{\mathbf{D}_{n:n-M}} - \overset{(3)}{\tfrac{1}{2}\Delta\omega^2\mathbf{D}^2_{n:n-M}}\big) + \sigma_w^2\mathbf{I}_{M+1}\Big)\Delta\mathbf{f}\Big]\Big\} \\ &= 2\,\mathrm{Re}\Big\{\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[\overset{(11)}{\mathbf{f}^H\big(\mathbf{H}'\mathbf{H}'^H + \sigma_w^2\mathbf{I}_{M+1}\big)\Delta\mathbf{f}} + \overset{(12)}{j\Delta\omega\mathbf{f}^H\mathbf{H}'\mathbf{H}'^H\mathbf{D}_{n:n-M}\Delta\mathbf{f}} \\ &\quad - \overset{(13)}{\tfrac{1}{2}\Delta\omega^2\mathbf{f}^H\mathbf{H}'\mathbf{H}'^H\mathbf{D}^2_{n:n-M}\Delta\mathbf{f}} - \overset{(21)}{j\Delta\omega\mathbf{f}^H\mathbf{D}_{n:n-M}\mathbf{H}'\mathbf{H}'^H\Delta\mathbf{f}} \\ &\quad + \overset{(22)}{\Delta\omega^2\mathbf{f}^H\mathbf{D}_{n:n-M}\mathbf{H}'\mathbf{H}'^H\mathbf{D}_{n:n-M}\Delta\mathbf{f}} + \overset{(23)}{j\tfrac{1}{2}\Delta\omega^3\mathbf{f}^H\mathbf{D}_{n:n-M}\mathbf{H}'\mathbf{H}'^H\mathbf{D}^2_{n:n-M}\Delta\mathbf{f}} \\ &\quad - \overset{(31)}{\tfrac{1}{2}\Delta\omega^2\mathbf{f}^H\mathbf{D}^2_{n:n-M}\mathbf{H}'\mathbf{H}'^H\Delta\mathbf{f}} - \overset{(32)}{j\tfrac{1}{2}\Delta\omega^3\mathbf{f}^H\mathbf{D}^2_{n:n-M}\mathbf{H}'\mathbf{H}'^H\mathbf{D}_{n:n-M}\Delta\mathbf{f}} \\ &\quad + \overset{(33)}{\tfrac{1}{4}\Delta\omega^4\mathbf{f}^H\mathbf{D}^2_{n:n-M}\mathbf{H}'\mathbf{H}'^H\mathbf{D}^2_{n:n-M}\Delta\mathbf{f}}\Big]\Big\}.\end{aligned} \quad (4.73)$$

We keep first and second-order error terms (*we keep second-order terms just to see the lower-order term we neglect for t_3*). We will write the second-order terms of (4.73) in detail using that $\mathbf{D}_{n:n-M} = n\mathbf{I}_{M+1} - \mathbf{D}_M$

1.

$$2\,\mathrm{Re}\Big\{\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[j\Delta\omega\mathbf{f}^H\mathbf{H}'\mathbf{H}'^H\mathbf{D}_{n:n-M}\Delta\mathbf{f}\Big]\Big\} = 2\,\mathrm{Re}\Big\{\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[j\,n\,\Delta\omega\mathbf{f}^H\mathbf{H}'\mathbf{H}'^H\Delta\mathbf{f}\Big]\Big\} - 2\,\mathrm{Re}\Big\{\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[j\Delta\omega\mathbf{f}^H\mathbf{H}'\mathbf{H}'^H\mathbf{D}_M\Delta\mathbf{f}\Big]\Big\}. \tag{4.74}$$

2.

$$2\,\mathrm{Re}\Big\{\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[-j\Delta\omega\mathbf{f}^H\mathbf{D}_{n:n-M}\mathbf{H}'\mathbf{H}'^H\Delta\mathbf{f}\Big]\Big\} = -2\,\mathrm{Re}\Big\{\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[j\,n\,\Delta\omega\mathbf{f}^H\mathbf{H}'\mathbf{H}'^H\Delta\mathbf{f}\Big]\Big\} + 2\,\mathrm{Re}\Big\{\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[j\Delta\omega\mathbf{f}^H\mathbf{D}_M\mathbf{H}'\mathbf{H}'^H\Delta\mathbf{f}\Big]\Big\}. \tag{4.75}$$

Thus, using (4.74)-(4.75) in (4.73), and by ignoring terms of higher order, we obtain

$$\begin{aligned}\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}[t_3] &= 2\,\mathrm{Re}\Big\{\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[\mathbf{f}^H\mathbf{R}_z\Delta\mathbf{f}\Big]\Big\} - 2\,\mathrm{Re}\Big\{\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[j\Delta\omega\mathbf{f}^H\mathbf{H}'\mathbf{H}'^H\mathbf{D}_M\Delta\mathbf{f}\Big]\Big\}\\ &\quad + 2\,\mathrm{Re}\Big\{\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[j\Delta\omega\mathbf{f}^H\mathbf{D}_M\mathbf{H}'\mathbf{H}'^H\Delta\mathbf{f}\Big]\Big\}\\ &= 2\,\mathrm{Re}\Big\{\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[\mathbf{e}_d^H\mathbf{H}'^H\Delta\mathbf{f}\Big]\Big\} - 2\,\mathrm{Re}\Big\{\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[j\Delta\omega\mathbf{f}^H\mathbf{H}'\mathbf{H}'^H\mathbf{D}_M\Delta\mathbf{f}\Big]\Big\}\\ &\quad + 2\,\mathrm{Re}\Big\{\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[j\Delta\omega\mathbf{f}^H\mathbf{D}_M\mathbf{H}'\mathbf{H}'^H\Delta\mathbf{f}\Big]\Big\}\\ &= 2\,\mathrm{Re}\Big\{\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[\mathbf{e}_d^H\mathbf{H}'^H\Delta\mathbf{f}\Big]\Big\} + \mathcal{O}\left(\frac{M\,\sigma_w^2}{R^3}\right)\end{aligned} \tag{4.76}$$

where we have used that $\mathbf{f} = \mathbf{R}_z^{-1}\mathbf{H}'\mathbf{e}_d$.

Term t_4

Using (4.61) we obtain

$$\begin{aligned}\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}[t_4] &\approx -2\,\mathrm{Re}\Big\{\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[\mathrm{e}^{j\Delta\omega\xi}\mathbf{f}^H\mathbf{\Gamma}_{n:n-M}(-\Delta\omega)\mathbf{H}'\mathbf{e}_d\Big]\Big\}\\ &= -2\,\mathrm{Re}\Big\{\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[\mathbf{f}^H\mathbf{\Gamma}'_{n:n-M}(-\Delta\omega)\mathbf{H}'\mathbf{e}_d\Big]\Big\}\\ &= -2\,\mathrm{Re}\Big\{\mathbf{f}^H\mathbf{H}'\mathbf{e}_d\Big\} + 2\,\mathrm{Re}\Big\{\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[j\Delta\omega\mathbf{f}^H\mathbf{D}'_{n:n-M}\mathbf{H}'\mathbf{e}_d\Big]\Big\}\\ &\quad + \mathrm{Re}\Big\{\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[\Delta\omega^2\mathbf{f}^H\mathbf{D}'^2_{n:n-M}\mathbf{H}'\mathbf{e}_d\Big]\Big\}.\end{aligned} \tag{4.77}$$

Using that $\mathcal{E}_{\Delta\omega}[\Delta\omega] = 0$, we get

$$\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}[t_4] \approx -2\,\mathrm{Re}\Big\{\mathbf{f}^H\mathbf{H}'\mathbf{e}_d\Big\} + \mathrm{Re}\Big\{\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[\Delta\omega^2\mathbf{f}^H\mathbf{D}'^2_{n:n-M}\mathbf{H}'\mathbf{e}_d\Big]\Big\}. \tag{4.78}$$

Term t_5

Using (4.61) we obtain

$$\begin{aligned}\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}[t_5] &\approx -2\,\mathrm{Re}\Big\{\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[\mathrm{e}^{j\Delta\omega\xi}\Delta\mathbf{f}^H\mathbf{\Gamma}_{n:n-M}(-\Delta\omega)\mathbf{H}'\mathbf{e}_d\Big]\Big\}\\ &= -2\,\mathrm{Re}\Big\{\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[\Delta\mathbf{f}^H\mathbf{\Gamma}'_{n:n-M}(-\Delta\omega)\mathbf{H}'\mathbf{e}_d\Big]\Big\}\\ &= -2\,\mathrm{Re}\Big\{\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[\Delta\mathbf{f}^H\mathbf{H}'\mathbf{e}_d\Big]\Big\} + 2\,\mathrm{Re}\Big\{\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[j\Delta\omega\Delta\mathbf{f}^H\mathbf{D}'_{n:n-M}\mathbf{H}'\mathbf{e}_d\Big]\Big\}\\ &\quad + \mathrm{Re}\Big\{\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[\Delta\omega^2\Delta\mathbf{f}^H\mathbf{D}'^2_{n:n-M}\mathbf{H}'\mathbf{e}_d\Big]\Big\}\\ &= -2\,\mathrm{Re}\Big\{\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[\Delta\mathbf{f}^H\mathbf{H}'\mathbf{e}_d\Big]\Big\} + \mathcal{O}\left(\frac{n^2\,\sigma_w^3}{R^{7/2}}\right).\end{aligned} \tag{4.79}$$

From the definition of the EMSE and using (4.67), (4.72), (4.76), (4.78) and (4.79) we get

$$\begin{aligned}\mathrm{EMSE}_n(\hat{\mathbf{f}},\hat{\omega}) = \mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[&\Delta\mathbf{f}^H\mathbf{R}_z\Delta\mathbf{f} + \Delta\omega^2\mathrm{Re}\{\mathbf{f}^H\mathbf{D}'^2_{n:n-M}\mathbf{H}'\mathbf{e}_d\}\\ &+ 2\mathrm{Re}\{j\Delta\omega\Delta\mathbf{f}^H\mathbf{D}'_{n:n-M}\mathbf{H}'\mathbf{e}_d\}\Big] + \mathcal{O}\left(\frac{n^2\,\sigma_w^3}{R^{7/2}}\right) + \mathcal{O}\left(\frac{M^2\,\sigma_w^2}{R^3}\right).\end{aligned} \tag{4.80}$$

The first term of (4.80) is computed as

$$\begin{aligned}\mathbf{T}_1 &\overset{\triangle}{=} \mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\big[\Delta\mathbf{f}^H\mathbf{R}_z\Delta\mathbf{f}\big]\\ &\overset{(4.29)}{=} \mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\Big[\big(\Delta\mathbf{h}'^H\mathbf{R}_o^T + \Delta\mathbf{h}'^T\mathbf{G}_o^H\big)\mathbf{R}_z^{-1}\big(\mathbf{R}_o^*\Delta\mathbf{h}' + \mathbf{G}_o\Delta\mathbf{h}'^*\big)\Big]\\ &\overset{(4.16),(4.17)}{=} \mathrm{tr}\left(\mathbf{R}_z^{-1}\left(\mathbf{R}_o^*\mathbf{C}'\mathbf{R}_o^T + \mathbf{G}_o\mathbf{C}'^*\mathbf{G}_o^H + \mathbf{G}_o\mathbf{C}'^*_t\mathbf{R}_o^T + \mathbf{R}_o^*\mathbf{C}'_t\mathbf{G}_o^H\right)\right).\end{aligned} \tag{4.81}$$

We continue with the second term of (4.80), which is time-dependent

$$\mathbf{T}_2(n) \overset{\triangle}{=} \mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\big[\Delta\omega^2\mathrm{Re}\{\mathbf{f}^H\mathbf{D}'^2_{n:n-M}\mathbf{H}'\mathbf{e}_d\}\big] = \sigma^2_{\Delta\omega}\mathrm{Re}\{\mathbf{f}^H\mathbf{D}'^2_{n:n-M}\mathbf{H}'\mathbf{e}_d\}. \tag{4.82}$$

The third term of (4.80) is also time-dependent, and is computed as

$$\begin{aligned}\mathbf{T}_3(n) &\overset{\triangle}{=} \mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\big[2\mathrm{Re}\{j\Delta\omega\Delta\mathbf{f}^H\mathbf{D}'_{n:n-M}\mathbf{H}'\mathbf{e}_d\}\big] = 2\mathrm{Re}\big\{\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\big[j\Delta\omega\Delta\mathbf{f}^H\mathbf{D}'_{n:n-M}\mathbf{H}'\mathbf{e}_d\big]\big\}\\ &\overset{(4.29)}{=} 2\mathrm{Re}\big\{\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\big[j\Delta\omega\left(\Delta\mathbf{h}'^H\mathbf{R}_o^T + \Delta\mathbf{h}'^T\mathbf{G}_o^H\right)\mathbf{R}_z^{-1}\mathbf{D}'_{n:n-M}\mathbf{H}'\mathbf{e}_d\big]\big\}\\ &= 2\mathrm{Re}\big\{\mathcal{E}_{\Delta\mathbf{h}',\Delta\omega}\big[j\Delta\omega\Delta\mathbf{h}'^H\mathbf{R}_o^T\mathbf{R}_z^{-1}\mathbf{D}'_{n:n-M}\mathbf{H}'\mathbf{e}_d + j\Delta\omega\Delta\mathbf{h}'^T\mathbf{G}_o^H\mathbf{R}_z^{-1}\mathbf{D}'_{n:n-M}\mathbf{H}'\mathbf{e}_d\big]\big\}\\ &\overset{(4.18)}{=} 2\sigma^2_{\Delta\omega}\,\mathrm{Re}\big\{\mathbf{h}'^H\mathbf{A}^H\mathbf{K}'\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{R}_o^T\mathbf{R}_z^{-1}\mathbf{D}'_{n:n-M}\mathbf{H}'\mathbf{e}_d\\ &\qquad - \mathbf{h}'^T\mathbf{A}^T\mathbf{K}'\mathbf{A}^*(\mathbf{A}^H\mathbf{A})^{-T}\mathbf{G}_o^H\mathbf{R}_z^{-1}\mathbf{D}'_{n:n-M}\mathbf{H}'\mathbf{e}_d\big\}.\end{aligned} \tag{4.83}$$

Appendix 4B

If the covariance matrix of the training sequence is $\mathcal{R}$, then [2, Appendix A]

$$\frac{1}{R}\mathbf{A}^H\mathbf{A} = \mathcal{R} + \mathcal{O}\left(\frac{1}{R}\right) \tag{4.84}$$

and

$$\frac{1}{\frac{R^2}{2}}\mathbf{A}^H\mathbf{D}_{R-1}\mathbf{A} = \mathcal{R} + \mathcal{O}\left(\frac{1}{R}\right). \tag{4.85}$$

If $\mathcal{R}$ is invertible, then, using the first-order expansion (11) we obtain

$$\left(\mathbf{A}^H\mathbf{A}\right)^{-1} = \frac{1}{R}\mathcal{R}^{-1} + \mathcal{O}\left(\frac{1}{R^2}\right)\mathcal{R}^{-2}. \tag{4.86}$$

Furthermore,

$$\begin{aligned}\mathbf{A}^H\mathbf{K}'\mathbf{A} &= \mathbf{A}^H\left(-\mathbf{D}_{R-1} + \left(\frac{R}{2} - 1\right)\mathbf{I}_R\right)\mathbf{A} \\ &= -\mathbf{A}^H\mathbf{D}_{R-1}\mathbf{A} + \left(\frac{R}{2} - 1\right)\mathbf{A}^H\mathbf{A} \\ &= -\frac{R^2}{2}\mathcal{R} - \mathcal{O}(R) + \left(\frac{R}{2} - 1\right)(R\mathcal{R} + \mathcal{O}(1)) \\ &= \mathcal{O}(R).\end{aligned} \tag{4.87}$$

Then

$$\begin{aligned}\mathbf{A}^H\mathbf{K}'\mathbf{A}\left(\mathbf{A}^H\mathbf{A}\right)^{-1} &= \mathcal{O}(R)\left(\frac{1}{R}\mathcal{R}^{-1} + \mathcal{O}\left(\frac{1}{R^2}\right)\mathcal{R}^{-2}\right) \\ &= \mathcal{O}(1)\mathcal{R}^{-1} + \mathcal{O}\left(\frac{1}{R}\right)\mathcal{R}^{-2} \\ &= \mathcal{O}(1).\end{aligned} \tag{4.88}$$

Using (4.88) and (4.55) it is easy to see that the second term of $\mathbf{C}'$ in (4.16) is $\mathcal{O}\left(\frac{\sigma_w^2}{R^3}\right)$, while, using (4.86) we obtain that the first term of $\mathbf{C}'$ is $\mathcal{O}\left(\frac{\sigma_w^2}{R}\right)$. Thus, if the covariance of the training sequence is invertible (and not very ill-conditioned), then the first term of $\mathbf{C}'$ in (4.16) is much larger than the second.

Appendix 4C

Using the singular value decomposition (SVD), matrix $\mathbf{A}$ can be written as

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H \tag{4.89}$$

where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices with dimensions $(N_{\text{tr}}-L)\times(N_{\text{tr}}-L)$ and $(L+1)\times(L+1)$ respectively, while $\boldsymbol{\Sigma}$ is the $(N_{\text{tr}}-L)\times(L+1)$ matrix with the singular values of $\mathbf{A}$ in its diagonal, and all the off the diagonal elements equal to zero. Using (4.89) we obtain

$$\left(\mathbf{A}^H\mathbf{A}\right)^{-T} = \mathbf{V}^*\boldsymbol{\Sigma}_1\mathbf{V}^T \tag{4.90}$$

where $\boldsymbol{\Sigma}_1 = \left(\boldsymbol{\Sigma}^H\boldsymbol{\Sigma}\right)^{-1} = \text{diag}\left(\sigma_1^{-2}(\mathbf{A}),\ldots,\sigma_{L+1}^{-2}(\mathbf{A})\right)$. Thus, in order to prove (4.52) we use the submultiplicative property of the matrix norms and expressions (4.89) and (4.90) to get

$$\begin{aligned}
\|\mathbf{A}^T\mathbf{K}'\mathbf{A}^*(\mathbf{A}^H\mathbf{A})^{-T}\|_2 &= \|\mathbf{V}^*\boldsymbol{\Sigma}^T\mathbf{U}^T\mathbf{K}'\mathbf{U}^*\boldsymbol{\Sigma}^*\boldsymbol{\Sigma}_1\mathbf{V}^T\|_2 \\
&\leq \|\mathbf{V}^*\|_2\|\boldsymbol{\Sigma}^T\|_2\|\mathbf{U}^T\|_2\|\mathbf{K}'\|_2\|\mathbf{U}^*\|_2\|\boldsymbol{\Sigma}^*\boldsymbol{\Sigma}_1\|_2\|\mathbf{V}^T\|_2 \\
&= \sigma_{\max}(\mathbf{A})\|\mathbf{K}'\|_2\frac{1}{\sigma_{\min}(\mathbf{A})} = \frac{\sigma_{\max}(\mathbf{A})}{\sigma_{\min}(\mathbf{A})}\|\mathbf{K}'\|_2 = \frac{R}{2}\,k_2(\mathbf{A})
\end{aligned} \tag{4.91}$$

where for the last equality we used the definition of matrix $\mathbf{K}'$ and the condition number with respect to the spectral norm.

In the same manner we prove the inequality in (4.53). We first write the SVD of matrix $\mathbf{H}'$ as

$$\mathbf{H}' = \mathbf{U}_1\boldsymbol{\Lambda}\mathbf{V}_1^H \tag{4.92}$$

where $\mathbf{U}_1$ and $\mathbf{V}_1$ are unitary matrices with dimensions $(M+1)\times(M+1)$ and $(M+L+1)\times(M+L+1)$ respectively, while $\boldsymbol{\Lambda}$ is the $(M+1)\times(M+L+1)$ matrix with the singular values of $\mathbf{H}'$ in its diagonal, and all the off the diagonal elements equal to zero. Using (4.92) and the definition of $\mathbf{R}_z$ we obtain

$$\mathbf{R}_z^{-1} = \mathbf{U}_1\boldsymbol{\Lambda}_1\mathbf{U}_1^H \tag{4.93}$$

where $\boldsymbol{\Lambda}_1 = \text{diag}\left(\frac{1}{\sigma_1^2(\mathbf{H}')+\sigma_w^2},\ldots,\frac{1}{\sigma_{M+1}^2(\mathbf{H}')+\sigma_w^2}\right)$.

Using (4.92), (4.93) and that $\mathbf{G} = \mathbf{H}'\mathbf{F}^T$ we obtain

$$\begin{aligned}
\|\mathbf{G}^H\mathbf{R}_z^{-1}\mathbf{D}_M\mathbf{H}'\mathbf{e}_d\|_2 &= \|\mathbf{F}^*\mathbf{V}_1\boldsymbol{\Lambda}^H\boldsymbol{\Lambda}_1\mathbf{U}_1^H\mathbf{D}_M\mathbf{U}_1\boldsymbol{\Lambda}\mathbf{V}_1^H\mathbf{e}_d\|_2 \\
&\overset{(*)}{\leq} \|\mathbf{F}\|_2\frac{1}{\sigma_{\min}(\mathbf{H}')}M\sigma_{\max}(\mathbf{H}') \\
&= \|\mathbf{F}\|_2\frac{\sigma_{\max}(\mathbf{H}')}{\sigma_{\min}(\mathbf{H}')}M = M\,\|\mathbf{F}\|_2 k_2(\mathbf{H}')
\end{aligned} \tag{4.94}$$

where at point (*) we have used that the singular values of $\mathbf{\Lambda}^H\mathbf{\Lambda}_1$ are equal to $\sigma_i(\mathbf{\Lambda}^H\mathbf{\Lambda}_1) = \frac{\sigma_i(\mathbf{H}')}{\sigma_i^2(\mathbf{H}') + \sigma_w^2}$, $i = 1, \ldots, M+1$, and $\sigma_i(\mathbf{\Lambda}^H\mathbf{\Lambda}_1) = \frac{\sigma_i(\mathbf{H}')}{\sigma_i^2(\mathbf{H}') + \sigma_w^2} \leq \frac{\sigma_i(\mathbf{H}')}{\sigma_i^2(\mathbf{H}')} = \frac{1}{\sigma_i(\mathbf{H}')}$. Thus, $\|\mathbf{\Lambda}^H\mathbf{\Lambda}_1\|_2 = \max(\sigma_i(\mathbf{\Lambda}^H\mathbf{\Lambda}_1)) \leq \frac{1}{\sigma_{\min}(\mathbf{H}')}$.

Bibliography

[1] N. Al-Dhahir and A. H. Sayed, "The finite-length multi-input multi-output MMSE-DFE," *IEEE Trans. Signal Process.*, vol. 48, no. 10. pp. 2921-2936, Oct. 2000.

[2] O. Besson and P. Stoica, "Training sequence selection for frequency offset estimation in frequency-selective channels," *Digital Signal Processing*, vol. 13, no. 1, pp. 106-127, Jan. 2003.

[3] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj and H. V. Poor. *MIMO Wireless Communications.* Cambridge University Press, 2007.

[4] P. Ciblat, P. Bianchi and M. Ghogho, "Training Sequence Optimization for Joint Channel and Frequency Offset Estimation," *IEEE Trans. Signal Proc.*, vol. 56, no. 8, part 1, pp. 3424-3436, Aug. 2008.

[5] F. A. Dietrich, P. Breun, and W. Utschick, "Robust Tomlinson-Harashima Precoding for the Wireless Broadcast Channel, " *IEEE Trans. Signal Process.*, vol. 55, no. 2, pp. 631-644, Feb. 2007.

[6] F. A. Dietrich, R. Hunger, M. Joham and W. Utschick, "Robust transmit Wiener filter for time division duplex systems," *Proc. 3rd IEEE Int. Symp. on Signal Processing and Information Technology*, pp. 415-418, Darmstadt, Dec. 2003.

[7] A. A. D' Amico and M. Morelli, "Joint TX-RX MMSE Design for MIMO Multicarrier Systems with Tomlinson-Harashima Pre-coding," *IEEE Trans. Wireless Communications*, vol. 7, no. 8, pp. 3118-3127, Aug. 2008.

[8] R. F.H. Fischer. *Precoding and Signal Shaping for Digital Communications.* John Wiley, 2002.

[9] G. Golub and C. Van Loan. *Matrix Computations.* The John Hopkins University Press, Baltimore, MD, 1996.

[10] A. Hjørungnes, D. Gesbert, "Complex-Valued Matrix Differentiation: Techniques and Key Results," *IEEE Trans. on Communications*, Vol. 55, no. 6, pp. 2740 - 2746, June 2007.

[11] R. A Horn and C. R. Johnson. *Matrix analysis.* Cambridge University Press, 1985.

[12] 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.

[13] 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.

[14] M. Joham, D. Schmidt, J. Brehmer, W. Utschick, "Finite-length MMSE Tomlinson-Harashima precoding for frequency selective vector channels," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 3073-3088, June 2007.

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

[16] Y-D Kim, J. K. Lim, C. Suh, Y.H. Lee, "Designing training sequences for carrier frequency estimation in frequency-selective channels," *IEEE Trans. Vehicular Technology*, vol. 55, no. 1, pp. 151-157, Jan. 2006.

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

[18] A. P. Liavas and D. Tsipouridou, "On the performance of the mismatched MMSE and the LS linear equalizers," *IEEE Trans. Signal Proc.*, vol. 55, no. 7, part 1, pp. 3302-3311, July 2007.

[19] A. Lozano, A. M. Tulino, and S. Verdu, "Multiple-antenna capacity in the low-power regime," *IEEE Trans. Inf. Theory*, Vol. 49, no. 10, pp. 2527 - 2544, Oct. 2003.

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

[21] H. Minn, X. Fu, and V. Bhargava, "Optimal Periodic Training signal for frequency offset estimation in frequency-selective fading channels," *IEEE Trans. on Comm.* , vol. 54, pp. 1081-1096, Jun. 2006.

[22] M. Morelli and U. Mengali, "Carrier frequency estimation for transmissions over selective channels," *IEEE Trans. on Comm.*, vol. 48, no. 9, pp. 1580-1589, Sep. 2000.

[23] M. Payaro, A. Pascual-Iserte, A.I. Perez-Neira, M.A. Lagunas, "Robust Design of Spatial Tomlinson-Harashima Precoding in the Presence of Errors in the CSI," *IEEE Trans. Wireless Communications*, vol. 6, no. 7, pp. 2396-2401, July 2007.

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

[25] M. Boutros Shenouda and T. N. Davidson, "Tomlinson-Harashima precoding for broadcast channels with uncertainty," *IEEE J. Select. Areas Commun.*, vol. 25, no. 7, pp. 1380-1389, Sept. 2007.

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

[27] G. W. Stewart, On the perturbation of LU, Cholesky, and QR factorizations, SIAM Journal on Matrix Analysis and Applications, v.14 n.4, p.1141-1145, Oct. 1993.

[28] P. Stoica and O. Besson, "Training sequence design for frequency offset and frequency-selective channel estimation," *IEEE Trans. on Comm.*, vol. 51, no. 11, pp. 1910-1917, Nov. 2003.

[29] C. Windpassinger, R. F.H. Fischer, T. Vencel, and J. B. Huber, "Precoding in Multi-antenna and Multiuser Communications," *IEEE Trans. Wireless Communications*, vol. 3, no. 4, pp. 1305-1316, July 2004.

[30] Y. Zhu, K.B. Letaief, "Frequency domain equalization with tomlinson-harashima precoding for single carrier broadband MIMO systems," *IEEE Trans. Wireless Communications*, vol. 6, no. 12, pp. 4420-4431, Dec. 2007.