On the Robustness of the Finite-Length MMSE-DFE With Respect to Channel and Second-Order Statistics Estimation Errors

Athanasios P. Liavas, *Member, IEEE*

***Abstract*—The finite-length minimum mean square error decision-feedback equalizer (MMSE-DFE) is an efficient structure mitigating intersymbol interference (ISI) introduced by practically all communication channels at high-enough symbol rates. The filters constituting the MMSE-DFE, as well as related performance measures, can be computed by assuming perfect knowledge of the channel impulse response and the input and noise second-order statistics (SOS). In practice, we estimate the unknown quantities, and thus, inevitable estimation errors arise. In this paper, we model the estimation errors as small perturbations, and we derive a second-order approximation to the excess MSE. Furthermore, we derive second-order approximations to the mean excess MSE in terms of the parameter estimation error covariance matrices and simple and informative bounds, revealing the factors that govern the behavior of MMSE-DFE under mismatch. Simulations confirm that the derived second-order approximations provide accurate estimates of the MMSE-DFE performance degradation due to mismatch.**

***Index Terms*—Finite-length MMSE-DFE, performance analysis under mismatch, perturbation analysis.**

I. Introduction

IT is well-known that intersymbol interference (ISI) severely impedes reliable high-speed digital communication over bandlimited channels. Many linear and nonlinear structures have been proposed to mitigate ISI. These structures differ greatly in the assumptions they make, their computational complexity, and their performance [1]. The finite-length MMSE-DFE has proved to be an efficient structure toward ISI mitigation in packet-based communication systems [2]. It is determined by two optimal filters, namely, the feedforward and the feedback filter. These filters, as well as related performance measures, can be computed by assuming perfect knowledge of the channel impulse response and the input and additive channel noise SOS [2]. In practice, we estimate the unknown quantities, and thus, inevitable estimation errors arise, resulting in channel and SOS mismatch. Consequently, the analysis of the robustness of the finite-length MMSE-DFE with respect to mismatch is of great importance.

This problem was first considered in [3], where the authors developed closed-form expressions for the perturbed MMSE-DFE filters and the corresponding performance measures. In this work, we present a detailed second-order perturbation analysis that explicitly reveals the factors that govern the performance of the MMSE-DFE under mismatch.

The rest of the paper is organized as follows. In Section II, we assume perfect knowledge of the channel impulse response and the input and noise SOS, and we recall known results concerning the finite-length MMSE-DFE. In Section III, we model the channel impulse response and SOS estimation errors as small perturbations, and we derive a second-order approximation to the excess MSE. In Section IV, we assume that we estimate only one quantity at a time (with the others being perfectly known), and we develop expressions for the second-order approximation to the mean excess MSE in terms of the

a) channel;
b) noise SOS;
c) input SOS;

estimation error covariance matrices. Using some approximations, in cases a) and c), we develop simple and informative bounds, revealing the factors that govern the behavior of the MMSE-DFE under mismatch. We show that the size of the elements of the feedforward filter and the residual impulse response (to be defined later) determine the magnification of the estimation errors governing the behavior of the MMSE-DFE under mismatch. In Section V, simulation studies validate the usefulness of the derived second-order expressions.

II. Finite-Length MMSE-DFE

In this section, for the convenience of the reader and in order to fix notation, we recapitulate known results concerning the finite-length MMSE-DFE [2]. We assume that the channel impulse response and the input and noise SOS are perfectly known.

A. Channel Model

Let us consider a baseband discrete-time fractionally spaced noisy communication channel modeled by the νth-order one-input/p-output linear time-invariant system depicted in Fig. 1. Its input–output relation is given by the convolution

$$\mathbf{y}_n = \sum_{i=0}^{\nu} \mathbf{h}_i x_{n-i} + \mathbf{n}_n$$

where x_n denotes the input sequence, and the p-dimensional vectors $\mathbf{y}_n$, $\mathbf{n}_n$ and $\mathbf{h}_i$ denote, respectively, the terms of the

Manuscript received January 28, 2002; revised June 20, 2002. This work was supported in part by the ΠΕΝΕΔ-99Ε Δ83 "ISODIA" program of the Greek Secretariat for Research and Technology. The associate editor coordinating the review of this paper and approving it for publication was Dr. Alexei Gorokhov.

The author is with the Department of Mathematics, University of the Aegean, Samos, Greece (e-mail: liavas@aegean.gr).

Digital Object Identifier 10.1109/TSP.2002.804083.

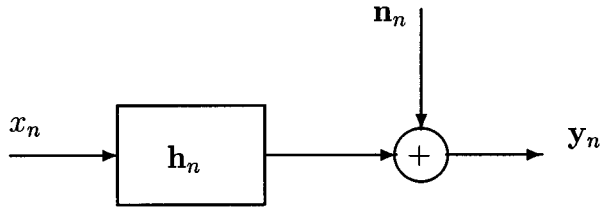

Fig. 1. Channel model.

output, noise, and channel finite impulse response (FIR) sequences. The impulse response terms $\mathbf{h}_i$ are vectors composed of the samples of the continuous-time impulse response modeling the combined effect of the transmit filter, the physical channel, and the receive filter [2]. By grouping the impulse response terms, we construct the impulse response vector $\mathcal{H}_\nu \triangleq \left[\mathbf{h}_0^T \ \cdots \ \mathbf{h}_\nu^T\right]^T$, where superscript ^{T} denotes transpose.

By stacking N_f successive output samples, we construct the data vector

$$\mathbf{y}_{n:n-N_f+1} \triangleq \left[\mathbf{y}_n^T \ \cdots \ \mathbf{y}_{n-N_f+1}^T\right]^T$$

which can be expressed as

$$\mathbf{y}_{n:n-N_f+1} = \mathbf{H}\mathbf{x}_{n:n-N_f-\nu+1} + \mathbf{n}_{n:n-N_f+1}$$

where the $pN_f \times (\nu + N_f)$ filtering matrix $\mathbf{H}$ is defined as

$$\mathbf{H} \triangleq \begin{bmatrix} \mathbf{h}_0 & \cdots & \cdots & \mathbf{h}_\nu & & \\ & \ddots & & & \ddots & \\ & & \mathbf{h}_0 & \cdots & \cdots & \mathbf{h}_\nu \end{bmatrix}$$

and the definitions of $\mathbf{x}_{n:n-N_f-\nu+1}$ and $\mathbf{n}_{n:n-N_f+1}$ are obvious.

B. Finite-Length MMSE-DFE

Our aim is to recover (a delayed version of) the input sequence x_n by passing the noisy output data $\mathbf{y}_n$ through an equalizer structure. To this end, we employ the finite-length DFE depicted in Fig. 2. The DFE is determined by the following parameter vectors.

1) $\mathbf{w} \triangleq \left[\mathbf{w}_0^T \ \cdots \ \mathbf{w}_{N_f-1}^T\right]^T$, which denotes the p-input/1-output length-N_f feedforward filter.
2) $\mathbf{b} \triangleq [1 \ b_1 \ \cdots \ b_{N_b}]^T$, which determines the single-input/single-output length-N_b strictly causal feedback filter. The settings of the feedback filter are $\{-b_1, \ldots, -b_{N_b}\}$.

Assuming that *past decisions are correct* and considering the delay Δ, the error between the desired output $x_{n-\Delta}$ and the input to the decision device $\mathcal{D}$ is given by

$$\begin{aligned} e_n &\triangleq x_{n-\Delta} - \left(\sum_{i=0}^{N_f-1} \mathbf{w}_i^T \mathbf{y}_{n-i} - \sum_{i=1}^{N_b} b_i x_{n-\Delta-i}\right) \\ &= \mathbf{b}^T \mathbf{x}_{n-\Delta:n-\Delta-N_b} - \mathbf{w}^T \mathbf{y}_{n:n-N_f+1} \\ &= \widetilde{\mathbf{b}}^T \mathbf{x}_{n:n-N_f-\nu+1} - \mathbf{w}^T \mathbf{y}_{n:n-N_f+1} \end{aligned} \quad (1)$$

where we defined $\widetilde{\mathbf{b}} \triangleq \left[\mathbf{0}_{1\times\Delta} \mathbf{b}^T \mathbf{0}_{1\times s}\right]^T$, with $\mathbf{0}_{i\times j}$ denoting the $i \times j$ zero matrix, and $s \triangleq N_f + \nu - \Delta - N_b - 1$. In order to simplify notation, we will omit the subscripts from $\mathbf{x}_{n:n-N_f-\nu+1}$, $\mathbf{y}_{n:n-N_f+1}$ and $\mathbf{n}_{n:n-N_f+1}$.

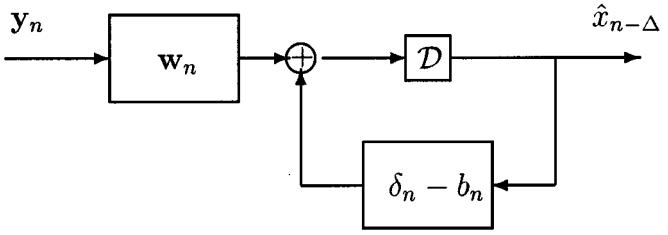

Fig. 2. Finite-length DFE.

The MMSE-DFE settings are computed by minimizing the mean square error (**MSE**), which can be expressed as

$$\begin{aligned} \mathbf{MSE} &\triangleq \mathcal{E}\left[e_n^2\right] = \mathcal{E}\left[\left(\widetilde{\mathbf{b}}^T\mathbf{x} - \mathbf{w}^T\mathbf{y}\right)\left(\mathbf{x}^T\widetilde{\mathbf{b}} - \mathbf{y}^T\mathbf{w}\right)\right] \\ &= \widetilde{\mathbf{b}}^T\mathbf{R}_{xx}\widetilde{\mathbf{b}} - \widetilde{\mathbf{b}}^T\mathbf{R}_{xy}\mathbf{w} - \mathbf{w}^T\mathbf{R}_{yx}\widetilde{\mathbf{b}} + \mathbf{w}^T\mathbf{R}_{yy}\mathbf{w} \end{aligned} \quad (2)$$

where

$$\mathbf{R}_{xx} \triangleq \mathcal{E}\left[\mathbf{x}\mathbf{x}^T\right], \ \mathbf{R}_{xy} \triangleq \mathcal{E}\left[\mathbf{x}\mathbf{y}^T\right] = \mathbf{R}_{xx}\mathbf{H}^T = \mathbf{R}_{yx}^T \quad (3)$$

$$\mathbf{R}_{yy} \triangleq \mathcal{E}\left[\mathbf{y}\mathbf{y}^T\right] = \mathbf{H}\mathbf{R}_{xx}\mathbf{H}^T + \mathbf{R}_{nn}, \ \mathbf{R}_{nn} \triangleq \mathcal{E}\left[\mathbf{n}\mathbf{n}^T\right]. \quad (4)$$

At the optimal settings, the error e_n is uncorrelated with the data vector $\mathbf{y}$, i.e., $\mathcal{E}\left[e_n\mathbf{y}^T\right] = \mathbf{0}_{1\times pN_f}$, yielding [2]

$$\mathbf{R}_{yx}\widetilde{\mathbf{b}} = \mathbf{R}_{yy}\mathbf{w} \Longrightarrow \mathbf{w} = \mathbf{R}_{yy}^{-1}\mathbf{R}_{yx}\widetilde{\mathbf{b}}.$$

Substituting the above expression for $\mathbf{w}$ into (2), we obtain

$$\mathbf{MSE} = \widetilde{\mathbf{b}}^T\mathbf{R}\widetilde{\mathbf{b}}$$

where

$$\mathbf{R} \triangleq \mathbf{R}_{xx} - \mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\mathbf{R}_{yx}. \quad (5)$$

If we define

$$\mathbf{R}_\Delta \triangleq \left[\mathbf{0}_{(N_b+1)\times\Delta}\mathbf{I}_{N_b+1}\mathbf{0}_{(N_b+1)\times s}\right]\mathbf{R}\begin{bmatrix}\mathbf{0}_{\Delta\times(N_b+1)} \\ \mathbf{I}_{N_b+1} \\ \mathbf{0}_{s\times(N_b+1)}\end{bmatrix}$$

where $\mathbf{I}_i$ denotes the $i \times i$ identity matrix, then the MSE is expressed as

$$\mathbf{MSE} = \mathbf{b}^T\mathbf{R}_\Delta\mathbf{b}$$

and it can be shown [2] that it is minimized for

$$\mathbf{b}_o = \frac{\mathbf{R}_\Delta^{-1}\mathbf{e}_0}{\mathbf{e}_0^T\mathbf{R}_\Delta^{-1}\mathbf{e}_0}, \quad \mathbf{w}_o = \mathbf{R}_{yy}^{-1}\mathbf{R}_{yx}\widetilde{\mathbf{b}}_o \quad (6)$$

where $\mathbf{e}_0$ is the vector with 1 at its first position and zeros elsewhere. The corresponding minimum MSE $\mathcal{M}$ is derived by putting in (2) $\widetilde{\mathbf{b}} = \widetilde{\mathbf{b}}_o$ and $\mathbf{w} = \mathbf{w}_o$. Alternative expressions are given by

$$\mathcal{M} = \widetilde{\mathbf{b}}_o^T\mathbf{R}\widetilde{\mathbf{b}}_o = \mathbf{b}_o^T\mathbf{R}_\Delta\mathbf{b}_o = \frac{1}{\mathbf{e}_0^T\mathbf{R}_\Delta^{-1}\mathbf{e}_0}. \quad (7)$$

III. MMSE-DFE: Performance Analysis Under Mismatch

In the previous section, we assumed exact knowledge of the channel impulse response $\mathcal{H}_\nu$ and the input and noise SOS $\mathbf{R}_{xx}$ and $\mathbf{R}_{nn}$. In practice, we estimate the unknown quantities and, thus, inevitable estimation errors arise, resulting in channel and SOS mismatch. In this section, we provide a second-order approximation to the excess MSE induced by channel and SOS mismatch.

A. Framework

Let us assume that an estimation procedure has furnished the estimates $\{\widehat{\mathbf{h}}_i\}_{i=0}^{\hat{\nu}}$, leading to the impulse response vector estimate $\widehat{\mathcal{H}}_{\hat{\nu}} \triangleq \left[\widehat{\mathbf{h}}_0^T \ \cdots \ \widehat{\mathbf{h}}_{\hat{\nu}}^T\right]^T$. We consider the case $\hat{\nu} < \nu$, that is, the estimated channel order $\hat{\nu}$ is less than the true channel order ν. The case $\hat{\nu} \geq \nu$ is analogous.

For the purposes of analysis, we have to compare the true and estimated impulse responses. At first sight, this might seem difficult because the vectors collecting their terms, that is, $\mathcal{H}_\nu$ and $\widehat{\mathcal{H}}_{\hat{\nu}}$ have different lengths. In order to overcome this difficulty, we take into account the fact that, due to pulse shaping and dispersive effects in the transmission medium, $\mathcal{H}_\nu$ is usually composed of tails of small leading and trailing terms, and we augment the estimated impulse response vector $\widehat{\mathcal{H}}_{\hat{\nu}}$ with leading and trailing zeros, obtaining the vector

$$\widehat{\mathcal{H}}_\nu^{m_1} \triangleq \left[\mathbf{0}_{1\times pm_1} \ \widehat{\mathcal{H}}_{\hat{\nu}}^T \ \mathbf{0}_{1\times p(\nu-\hat{\nu}-m_1)}\right]^T$$

whose length equals the length of $\mathcal{H}_\nu$. Then, we define

$$m_1^* \triangleq \arg\min_{m_1} \left\| \mathcal{H}_\nu - \widehat{\mathcal{H}}_\nu^{m_1} \right\|_2$$

where $\|\cdot\|_2$ denotes, depending on the argument, the matrix or vector 2-norm. That is, pm_1^* is the number of leading zeros we must insert in front of $\widehat{\mathcal{H}}_{\hat{\nu}}$ so that the augmented impulse response vector estimate becomes closest to $\mathcal{H}_\nu$. In the sequel, we will work with the augmented impulse response vector estimate $\widehat{\mathcal{H}}_\nu \triangleq \widehat{\mathcal{H}}_\nu^{m_1^*}$. We note that working with $\widehat{\mathcal{H}}_\nu$ instead of $\widehat{\mathcal{H}}_{\hat{\nu}}$ simply amounts to insertion of an extra delay of m_1^* time units.

We consider our channel estimate as being good if $\mathcal{H}_\nu$ and $\widehat{\mathcal{H}}_\nu$ are close to each other. In terms of the associated filtering matrices, we express this condition as

$$\Delta\mathbf{H} \triangleq \widehat{\mathbf{H}} - \mathbf{H}, \quad \|\Delta\mathbf{H}\|_2 \ll \|\mathbf{H}\|_2.$$

The estimation errors in the input and noise SOS can be expressed in an analogous manner. Thus, let us assume that we have estimated $\widehat{\mathbf{R}}_{xx}$ and $\widehat{\mathbf{R}}_{nn}$ such that

$$\begin{aligned}\Delta\mathbf{R}_{xx} &\triangleq \widehat{\mathbf{R}}_{xx} - \mathbf{R}_{xx}, \quad \|\Delta\mathbf{R}_{xx}\|_2 \ll \|\mathbf{R}_{xx}\|_2\\ \Delta\mathbf{R}_{nn} &\triangleq \widehat{\mathbf{R}}_{nn} - \mathbf{R}_{nn}, \quad \|\Delta\mathbf{R}_{nn}\|_2 \ll \|\mathbf{R}_{nn}\|_2.\end{aligned}$$

B. MMSE-DFE: Perturbation Analysis

In the sequel, we derive analytic expressions describing the behavior of the MMSE-DFE in the presence of estimation errors. We start by noting that under mismatch, efforts toward computation of $\mathbf{R}_{xy}$, $\mathbf{R}_{yx}$, and $\mathbf{R}_{yy}$ lead to[1]

$$\widehat{\mathbf{R}}_{xy} \triangleq \widehat{\mathbf{R}}_{xx}\widehat{\mathbf{H}}^T = \widehat{\mathbf{R}}_{yx}^T, \quad \widehat{\mathbf{R}}_{yy} \triangleq \widehat{\mathbf{H}}\widehat{\mathbf{R}}_{xx}\widehat{\mathbf{H}}^T + \widehat{\mathbf{R}}_{nn}.$$

The corresponding first-order perturbations $\Delta\mathbf{R}_{xy}$, $\Delta\mathbf{R}_{yx}$, and $\Delta\mathbf{R}_{yy}$ are computed as

$$\begin{aligned}\widehat{\mathbf{R}}_{xy} =& \widehat{\mathbf{R}}_{xx}\widehat{\mathbf{H}}^T = (\mathbf{R}_{xx} + \Delta\mathbf{R}_{xx})(\mathbf{H} + \Delta\mathbf{H})^T\\ \cong& \mathbf{R}_{xy} + \underbrace{\left(\Delta\mathbf{R}_{xx}\mathbf{H}^T + \mathbf{R}_{xx}\Delta\mathbf{H}^T\right)}_{\Delta\mathbf{R}_{xy}}\end{aligned} \tag{8}$$

where $\cong$ denotes first-order approximation. It is easy to see that $\Delta\mathbf{R}_{yx} = \Delta\mathbf{R}_{xy}^T$ and

$$\Delta\mathbf{R}_{yy} = \Delta\mathbf{H}\mathbf{R}_{xx}\mathbf{H}^T + \mathbf{H}\Delta\mathbf{R}_{xx}\mathbf{H}^T + \mathbf{H}\mathbf{R}_{xx}\Delta\mathbf{H}^T + \Delta\mathbf{R}_{nn}. \tag{9}$$

Then, efforts toward computing $\mathbf{R}$ and $\mathbf{R}_\Delta$ yield

$$\begin{aligned}\widehat{\mathbf{R}} \triangleq& \widehat{\mathbf{R}}_{xx} - \widehat{\mathbf{R}}_{xy}\widehat{\mathbf{R}}_{yy}^{-1}\widehat{\mathbf{R}}_{yx}\\ \widehat{\mathbf{R}}_\Delta \triangleq& \left[\mathbf{0}_{(N_b+1)\times\Delta} \ \mathbf{I}_{N_b+1}\mathbf{0}_{(N_b+1)\times s}\right]\widehat{\mathbf{R}}\begin{bmatrix}\mathbf{0}_{\Delta\times(N_b+1)}\\ \mathbf{I}_{N_b+1}\\ \mathbf{0}_{s\times(N_b+1)}\end{bmatrix}.\end{aligned}$$

In order to derive the first-order perturbation $\Delta\mathbf{R}$, we use the well-known first-order expansion [4, p. 131]

$$(\mathbf{A} + \Delta\mathbf{A})^{-1} \cong \mathbf{A}^{-1} - \mathbf{A}^{-1}\Delta\mathbf{A}\mathbf{A}^{-1} \tag{10}$$

and we obtain

$$\begin{aligned}\Delta\mathbf{R} = \Delta\mathbf{R}_{xx} - \mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\Delta\mathbf{R}_{yx} - \Delta\mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\mathbf{R}_{yx}\\ + \mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\Delta\mathbf{R}_{yy}\mathbf{R}_{yy}^{-1}\mathbf{R}_{yx}.\end{aligned} \tag{11}$$

The perturbation $\Delta\mathbf{R}_\Delta$ can be easily derived from the definitions of $\widehat{\mathbf{R}}_\Delta$ and $\Delta\mathbf{R}$.

The resulting "optimal" filters are given by

$$\widehat{\mathbf{b}}_o = \frac{\widehat{\mathbf{R}}_\Delta^{-1}\mathbf{e}_0}{\mathbf{e}_0^T\widehat{\mathbf{R}}_\Delta^{-1}\mathbf{e}_0}, \quad \widehat{\mathbf{w}}_o = \widehat{\mathbf{R}}_{yy}^{-1}\widehat{\mathbf{R}}_{yx}\widetilde{\widehat{\mathbf{b}}}_o \tag{12}$$

where $\widetilde{\widehat{\mathbf{b}}}_o$ is the appropriately zero-padded version of $\widehat{\mathbf{b}}_o$ [recall the definition of $\mathbf{b}$ in terms of $\mathbf{b}$ after (1)].

Assuming correct past decisions, the corresponding MSE can be expressed as

$$\begin{aligned}\widehat{\mathcal{M}} \triangleq& \mathcal{E}\left[\left(\widetilde{\widehat{\mathbf{b}}}_o^T\mathbf{x} - \widehat{\mathbf{w}}_o^T\mathbf{y}\right)\left(\mathbf{x}^T\widetilde{\widehat{\mathbf{b}}}_o - \mathbf{y}^T\widehat{\mathbf{w}}_o\right)\right]\\ =& \widetilde{\widehat{\mathbf{b}}}_o^T\mathbf{R}_{xx}\widetilde{\widehat{\mathbf{b}}}_o - \widetilde{\widehat{\mathbf{b}}}_o^T\mathbf{R}_{xy}\widehat{\mathbf{w}}_o - \widehat{\mathbf{w}}_o^T\mathbf{R}_{yx}\widetilde{\widehat{\mathbf{b}}}_o + \widehat{\mathbf{w}}_o^T\mathbf{R}_{yy}\widehat{\mathbf{w}}_o.\end{aligned} \tag{13}$$

By inspection of (2) and (13), we deduce that $\widehat{\mathcal{M}}$ is the value of the constrained quadratic function $\mathbf{MSE}$ at the point $\left[\widetilde{\widehat{\mathbf{b}}}_o^T \ \widehat{\mathbf{w}}_o^T\right]^T$, which is "close" to the optimal point $\left[\widetilde{\mathbf{b}}_o^T\mathbf{w}_o^T\right]^T$. If we denote with $\Delta\widetilde{\mathbf{b}}_o$ and $\Delta\mathbf{w}_o$ the first-order perturbations in

[1]Another way to estimate $\mathbf{R}_{yy}$ is through the sample autocovariance of the channel output data. The way adopted in the text offers lower computational complexity.

quantities $\widetilde{\mathbf{b}}_o$ and $\mathbf{w}_o$, respectively, we obtain the second-order approximation

$$\widehat{\mathcal{M}} \simeq \mathcal{M} + \text{first-order error terms} + \left[\Delta\widetilde{\mathbf{b}}_o^T \; \Delta\mathbf{w}_o^T\right] \begin{bmatrix} \mathbf{R}_{xx} & -\mathbf{R}_{xy} \\ -\mathbf{R}_{yx} & \mathbf{R}_{yy} \end{bmatrix} \begin{bmatrix} \Delta\widetilde{\mathbf{b}}_o \\ \Delta\mathbf{w}_o \end{bmatrix} \quad (14)$$

where $\simeq$ denotes second-order approximation. The summand "first-order error terms" is identically zero due to the optimality of the point $\left[\widetilde{\mathbf{b}}_o^T \; \mathbf{w}_o^T\right]^T$. Thus, the excess MSE is approximated by the second-order error terms.

We note that an expression analogous to (14) has appeared in [3]. The difference is that the true errors were in the place of the first-order approximations $\Delta\widetilde{\mathbf{b}}_o$ and $\Delta\mathbf{w}_o$, and a known channel order assumption was implicitly used. However, (14), as well as its analog in [3], are not very informative because they do not explicitly reveal the factors that govern the size of the excess MSE. This is our subject in the sequel.

At first, we derive first-order approximations to the perturbations on $\mathbf{b}_o$ and $\mathbf{w}_o$. Then, these expressions will be used for the derivation of the second-order approximation to the excess MSE.

Result 1: Let $\Delta\mathbf{b}_o$ and $\Delta\mathbf{w}_o$ be the first-order perturbations in quantities $\mathbf{b}_o$ and $\mathbf{w}_o$, respectively. Then

$$\Delta\mathbf{b}_o = \left(\mathbf{b}_o^T \Delta\mathbf{R}_\Delta \mathbf{b}_o\right) \mathbf{R}_\Delta^{-1}\mathbf{e}_0 - \mathbf{R}_\Delta^{-1}\Delta\mathbf{R}_\Delta\mathbf{b}_o \quad (15)$$

$$\Delta\mathbf{w}_o = \mathbf{R}_{yy}^{-1}\left(\mathbf{R}_{yx}\Delta\widetilde{\mathbf{b}}_o + \Delta\mathbf{R}_{yx}\widetilde{\mathbf{b}}_o - \Delta\mathbf{R}_{yy}\mathbf{w}_o\right) \quad (16)$$

where $\Delta\widetilde{\mathbf{b}}_o$ is the appropriately zero-padded version of $\Delta\mathbf{b}_o$.

Proof: The proof can be constructed by performing calculations in (12) using (10) and the first-order approximation

$$\frac{1}{a+\Delta a} \cong \frac{1}{a} - \frac{\Delta a}{a^2}$$

and ignoring higher order error terms. □

We now proceed to the derivation of the second-order approximation to the excess MSE.

Result 2: By denoting with SOT the second-order error terms in (14), we obtain

$$\text{SOT} = \left(\widetilde{\mathbf{b}}_o^T\Delta\mathbf{R}_{xy} - \mathbf{w}_o^T\Delta\mathbf{R}_{yy}\right)\mathbf{R}_{yy}^{-1} \times \left(\Delta\mathbf{R}_{yx}\widetilde{\mathbf{b}}_o - \Delta\mathbf{R}_{yy}\mathbf{w}_o\right) + \Delta\widetilde{\mathbf{b}}_o^T\mathbf{R}\Delta\widetilde{\mathbf{b}}_o. \quad (17)$$

Proof: The proof can be constructed by performing tedious but straightforward calculations after substitution of (16) for $\Delta\mathbf{w}_o$ in the second-order error terms of (14). □

Using (15) and the relation

$$\Delta\widetilde{\mathbf{b}}_o^T\mathbf{R}\Delta\widetilde{\mathbf{b}}_o = \Delta\mathbf{b}_o^T\mathbf{R}_\Delta\Delta\mathbf{b}_o$$

we can derive, after some straightforward calculations, an alternative expression for the second term of (17), as

$$\Delta\widetilde{\mathbf{b}}_o^T\mathbf{R}\Delta\widetilde{\mathbf{b}}_o = \mathbf{b}_o^T\Delta\mathbf{R}_\Delta\mathbf{R}_\Delta^{-1}\Delta\mathbf{R}_\Delta\mathbf{b}_o - \frac{\left(\mathbf{b}_o^T\Delta\mathbf{R}_\Delta\mathbf{b}_o\right)^2}{\mathcal{M}}. \quad (18)$$

Expression (17) provides an approximation to the excess MSE when the channel and the noise and input SOS estimates contain errors. A more insightful analysis results by considering the corresponding expression when errors are introduced by one quantity only (the other two quantities are assumed to be perfectly known). This is our subject in the sequel.

IV. Mean Excess MSE

A. Channel Estimation Errors

In this subsection, we assume that the input and noise SOS are perfectly known, and we derive an analytic expression for the mean excess MSE in terms of the channel estimation error covariance matrix

$$\mathbf{R}_{\Delta\mathcal{H}_\nu} \triangleq \mathcal{E}\left[\Delta\mathcal{H}_\nu \, \Delta\mathcal{H}_\nu^T\right]$$

where $\Delta\mathcal{H}_\nu \triangleq \widehat{\mathcal{H}}_\nu - \mathcal{H}_\nu$. To that end, we will find a matrix $\mathcal{S}$ such that

$$\text{SOT} = \Delta\mathcal{H}_\nu^T\mathcal{S}\Delta\mathcal{H}_\nu = \text{Tr}\left(\Delta\mathcal{H}_\nu^T\mathcal{S}\Delta\mathcal{H}_\nu\right)$$

where Tr denotes the matrix trace. Then, using the identity $\text{Tr}(AB) = \text{Tr}(BA)$, we will obtain

$$\mathcal{E}[\text{SOT}] = \mathcal{E}\left[\text{Tr}\left(\Delta\mathcal{H}_\nu^T\mathcal{S}\Delta\mathcal{H}_\nu\right)\right] = \text{Tr}\left(\mathcal{S}\mathbf{R}_{\Delta\mathcal{H}_\nu}\right)$$

relating the mean excess MSE with the channel estimation error covariance matrix.

We simplify somewhat the results by making the common assumption that $\mathbf{R}_{xx} = \mathbf{I}_{N_f+\nu}$. Under the above assumptions, many simplifications arise. For example, from (3) and (5), we obtain

$$\mathbf{R}_{xy} = \mathbf{H}^T = \mathbf{R}_{yx}^T, \quad \mathbf{R} = \mathbf{I}_{N_f+\nu} - \mathbf{H}^T\mathbf{R}_{yy}^{-1}\mathbf{H} \quad (19)$$

and from (8) and (11), we obtain

$$\Delta\mathbf{R}_{xy} = \Delta\mathbf{H}^T = \Delta\mathbf{R}_{yx}^T, \; \Delta\mathbf{R}_{yy} = \mathbf{H}\Delta\mathbf{H}^T + \Delta\mathbf{H}\mathbf{H}^T \quad (20)$$

$$\Delta\mathbf{R} = -\mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\Delta\mathbf{H}\mathbf{R} - \mathbf{R}\Delta\mathbf{H}^T\mathbf{R}_{yy}^{-1}\mathbf{R}_{yx}. \quad (21)$$

In the sequel, we shall make use of the following relations and definitions.

1) The product $\mathbf{w}_o^T\mathbf{H}$ expresses the convolution of the multichannel $\mathcal{H}_\nu$ and the feedforward filter $\mathbf{w}_o$ and, due to the commutativity property of the convolution, is equal to $\mathcal{H}_\nu^T\mathcal{W}_o$, where $\mathcal{W}_o$ is the $p(\nu+1)\times(N_f+\nu)$ filtering matrix constructed by the block vector $\mathbf{w}_o$. Thus

$$\mathbf{w}_o^T\mathbf{H} = \mathcal{H}_\nu^T\mathcal{W}_o. \quad (22)$$

2) We define the combined (channel-feedforward filter) impulse response $\mathbf{f}_o \triangleq \mathbf{H}^T\mathbf{w}_o$ and the residual impulse response

$$\mathbf{g}_o \triangleq \widetilde{\mathbf{b}}_o - \mathbf{f}_o \quad (23)$$

with elements $g_{o,i}$ for $i = 1,\ldots,N_f+\nu$. It is easy to show that under the previous assumptions, $\mathbf{g}_o = \mathbf{R}\widetilde{\mathbf{b}}_o$. We note that if the ideal MMSE-DFE performs well, then the elements of the residual impulse response $\mathbf{g}_o$ are small (in particular, $g_{o,\Delta} = \mathcal{M}$).

3) Simple calculations show that

$$\mathbf{g}_o^T\mathbf{H}^T = \mathcal{H}_\nu^T\mathcal{G}_o \tag{24}$$

where $\mathcal{G}_o$ is the block Hankel matrix defined as

$$\mathcal{G}_o \triangleq \begin{bmatrix} g_{o,1}\mathbf{I}_p & g_{o,2}\mathbf{I}_p & \cdots & g_{o,N_f}\mathbf{I}_p \\ g_{o,2}\mathbf{I}_p & g_{o,3}\mathbf{I}_p & \cdots & g_{o,N_f+1}\mathbf{I}_p \\ \vdots & \vdots & & \vdots \\ g_{o,\nu+1}\mathbf{I}_p & g_{o,\nu+2}\mathbf{I}_p & \cdots & g_{o,N_f+\nu}\mathbf{I}_p \end{bmatrix}.$$

Now, we consider each term of SOT separately.

1) Term SOT1: Let us consider the first term of SOT in (17), which is denoted SOT1. Using (20), (22), (23), and (24), we obtain

$$\begin{aligned}\widetilde{\mathbf{b}}_o^T\Delta\mathbf{R}_{xy} - \mathbf{w}_o^T\Delta\mathbf{R}_{yy} &= \widetilde{\mathbf{b}}_o^T\Delta\mathbf{H}^T - \mathbf{w}_o^T\left(\mathbf{H}\Delta\mathbf{H}^T + \Delta\mathbf{H}\mathbf{H}^T\right) \\ &= \widetilde{\mathbf{b}}_o^T\Delta\mathbf{H}^T - \mathbf{f}_o^T\Delta\mathbf{H}^T - \mathbf{w}_o^T\Delta\mathbf{H}\mathbf{H}^T \\ &= \mathbf{g}_o^T\Delta\mathbf{H}^T - \mathbf{w}_o^T\Delta\mathbf{H}\mathbf{H}^T \\ &= \Delta\mathcal{H}_\nu^T\mathcal{G}_o - \Delta\mathcal{H}_\nu^T\mathcal{W}_o\mathbf{H}^T.\end{aligned}$$

Thus, the term SOT1 can be expressed as

$$\text{SOT1} = \Delta\mathcal{H}_\nu^T\underbrace{\left(\mathcal{G}_o - \mathcal{W}_o\mathbf{H}^T\right)}_{\mathcal{A}_1}\mathbf{R}_{yy}^{-1}\left(\mathcal{G}_o^T - \mathbf{H}\mathcal{W}_o^T\right)\Delta\mathcal{H}_\nu. \tag{25}$$

In the sequel, we consider the second term of SOT, which is denoted SOT2 and is expanded in two terms in (18).

2) Term SOT2$_1$: The first term on the right-hand side of (18), which is denoted SOT2$_1$, can be expressed as

$$\text{SOT2}_1 = \widetilde{\mathbf{b}}_o^T\Delta\mathbf{R}\mathbf{M}\Delta\mathbf{R}\widetilde{\mathbf{b}}_o$$

with

$$\mathbf{M} \triangleq \begin{bmatrix} \mathbf{0}_{(N_b+1)\times\Delta} \\ \mathbf{I}_{N_b+1} \\ \mathbf{0}_{(N_b+1)\times s} \end{bmatrix}\mathbf{R}_\Delta^{-1}\left[\mathbf{0}_{(N_b+1)\times\Delta}\ \mathbf{I}_{N_b+1}\ \mathbf{0}_{(N_b+1)\times s}\right].$$

Using (6) and (21), we obtain

$$\begin{aligned}\text{SOT2}_1 = &\left(\mathbf{w}_o^T\Delta\mathbf{H}\mathbf{R} + \mathbf{g}_o^T\Delta\mathbf{H}^T\mathbf{R}_{yy}^{-1}\mathbf{R}_{yx}\right)\mathbf{M} \\ &\times\left(\mathbf{R}\Delta\mathbf{H}^T\mathbf{w}_o + \mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\Delta\mathbf{H}\mathbf{g}_o\right)\end{aligned}$$

which, using (22) and (24), gives

$$\begin{aligned}\text{SOT2}_1 = \Delta\mathcal{H}_\nu^T&\underbrace{\left(\mathcal{W}_o\mathbf{R} + \mathcal{G}_o\mathbf{R}_{yy}^{-1}\mathbf{R}_{yx}\right)}_{\mathcal{A}_2}\mathbf{M} \\ &\times\left(\mathbf{R}\mathcal{W}_o^T + \mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\mathcal{G}_o^T\right)\Delta\mathcal{H}_\nu.\end{aligned} \tag{26}$$

3) Term SOT2$_2$: Considering the numerator of the second term of SOT2 in (18), we obtain, from (6) and (21)

$$\begin{aligned}\widetilde{\mathbf{b}}_o^T\Delta\mathbf{R}\widetilde{\mathbf{b}}_o &= -\widetilde{\mathbf{b}}_o^T\mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\Delta\mathbf{H}\mathbf{R}\widetilde{\mathbf{b}}_o - \widetilde{\mathbf{b}}_o^T\mathbf{R}\Delta\mathbf{H}^T\mathbf{R}_{yy}^{-1}\mathbf{R}_{yx}\widetilde{\mathbf{b}}_o \\ &= -\mathbf{w}_o^T\Delta\mathbf{H}\mathbf{g}_o - \mathbf{g}_o^T\Delta\mathbf{H}^T\mathbf{w}_o \\ &= -\Delta\mathcal{H}_\nu^T\mathcal{W}_o\mathbf{g}_o - \mathbf{g}_o^T\mathcal{W}_o^T\Delta\mathcal{H}_\nu.\end{aligned}$$

The square of this term is given by

$$\Delta\mathcal{H}_\nu^T\mathcal{A}_3\Delta\mathcal{H}_\nu$$

with

$$\mathcal{A}_3 \triangleq 4\mathcal{W}_o\mathbf{g}_o\mathbf{g}_o^T\mathcal{W}_o^T.$$

Thus, the second term of SOT2, which is denoted SOT2$_2$, becomes

$$\text{SOT2}_2 = \frac{1}{\mathcal{M}}\Delta\mathcal{H}_\nu^T\mathcal{A}_3\Delta\mathcal{H}_\nu. \tag{27}$$

Using (25)–(27), we derive the second-order approximation for the mean excess MSE as

$$\boxed{\mathcal{E}\left[\text{SOT}\right] = \mathcal{E}\left[\text{Tr}\left(\Delta\mathcal{H}_\nu^T\mathcal{S}\Delta\mathcal{H}_\nu\right)\right] = \text{Tr}\left(\mathcal{S}\mathbf{R}_{\Delta\mathcal{H}_\nu}\right)} \tag{28}$$

where

$$\boxed{\mathcal{S} \triangleq \mathcal{A}_1\mathbf{R}_{yy}^{-1}\mathcal{A}_1^T + \mathcal{A}_2\mathbf{M}\mathcal{A}_2^T - \tfrac{\mathcal{A}_3}{\mathcal{M}}.} \tag{29}$$

Expression (28) gives a relation between the second-order approximation to the mean excess MSE and the channel estimation error covariance matrix. Admittedly, (29) is complicated, and in order to get the necessary physical insight, we resort to simplifications and/or approximations.

Assuming that the elements of the residual impulse response $g_{o,i}$ are small, we may neglect the terms involving $\mathcal{G}_o$ in SOT1 and SOT2$_1$ (this assumption is more likely to be satisfied in the medium and high SNR cases).

If we assume further that the additive channel noise is zero-mean white with variance σ^2, we obtain from (25)

$$\begin{aligned}\mathcal{A}_1\mathbf{R}_{yy}^{-1}\mathcal{A}_1^T &\approx \mathcal{W}_o\mathbf{H}^T\mathbf{R}_{yy}^{-1}\mathbf{H}\mathcal{W}_o^T \\ &= \mathcal{W}_o\mathbf{H}^T\left(\mathbf{H}\mathbf{H}^T + \sigma^2\mathbf{I}_{pN_f}\right)^{-1}\mathbf{H}\mathcal{W}_o^T.\end{aligned} \tag{30}$$

It is an easy exercise to prove that (see [4, p. 138])

$$\lambda_i\left(\mathbf{H}^T\mathbf{R}_{yy}^{-1}\mathbf{H}\right) = \frac{\sigma_i^2(\mathbf{H})}{\sigma_i^2(\mathbf{H}) + \sigma^2} \le 1 \tag{31}$$

where $\lambda_i(\cdot)$ and $\sigma_i(\cdot)$, denote, respectively, the ith eigenvalue and the ith singular value of the matrix argument.

Considering the term SOT2$_1$ in (26), we obtain

$$\mathcal{A}_2\mathbf{M}\mathcal{A}_2^T \approx \mathcal{W}_o(\mathbf{R}\,\mathbf{M}\,\mathbf{R})\mathcal{W}_o^T. \tag{32}$$

Using (19) and (31), we obtain

$$\lambda_{\max}(\mathbf{R}) = 1 - \frac{\sigma_{\min}^2(\mathbf{H})}{\sigma_{\min}^2(\mathbf{H}) + \sigma^2}.$$

In Appendix A, we show that the matrix $\mathbf{R} - \mathbf{R}\,\mathbf{M}\,\mathbf{R}$ is semi-positive definite, i.e.,

$$\mathbf{R} - \mathbf{R}\,\mathbf{M}\,\mathbf{R} \ge 0 \tag{33}$$

which gives

$$\lambda_{\max}\left(\mathbf{R}\,\mathbf{M}\,\mathbf{R}\right) \le \lambda_{\max}(\mathbf{R}) \le 1. \tag{34}$$

Ignoring the term SOT2$_2$, we obtain, from (29), (30), and (32)

$$\mathcal{S} \approx \mathcal{W}_o\underbrace{\left(\mathbf{H}^T\mathbf{R}_{yy}^{-1}\mathbf{H} + \mathbf{R}\,\mathbf{M}\,\mathbf{R}\right)}_{\mathcal{X}}\mathcal{W}_o^T. \tag{35}$$

Using the inequalities $|\mathrm{Tr}(A^TB)| \leq ||A||_F||B||_F$ and $||AB||_F \leq ||A||_F||B||_2$, where $||\cdot||_F$ denotes the Frobenious norm of the matrix argument and the fact that $||\mathbf{A}||_2 = \lambda_{\max}(\mathbf{A})$ for a semi-positive definite matrix $\mathbf{A}$, we obtain from (28), (31), (34), and (35)

$$\begin{aligned}\mathcal{E}[\mathrm{SOT}] \leq & ||\mathcal{S}||_F||\mathbf{R}_{\Delta\mathcal{H}_\nu}||_F \\ \leq & ||\mathcal{W}_o||_F||\mathcal{X}||_2||\mathcal{W}_o^T||_2||\mathbf{R}_{\Delta\mathcal{H}_\nu}||_F \\ \leq & 2||\mathcal{W}_o||_F||\mathcal{W}_o||_2||\mathbf{R}_{\Delta\mathcal{H}_\nu}||_F \\ \leq & 2\,||\mathcal{W}_o||_F^2\,||\mathbf{R}_{\Delta\mathcal{H}_\nu}||_F \\ = & 2\,(\nu+1)\,||\mathbf{w}_o||_2^2\,||\mathbf{R}_{\Delta\mathcal{H}_\nu}||_F. \end{aligned} \tag{36}$$

Our aim was not to derive the best possible bound but a bound with a simple interpretation. Expression (36) is informative because it provides a simple bound for the mean excess MSE in terms of the channel estimation error covariance matrix and the feedforward filter of the ideal MMSE-DFE $\mathbf{w}_o$. If $\mathbf{w}_o$ has large elements, then the channel estimation errors may be significantly magnified, resulting in large excess MSE. If, on the other hand, its elements are not large, then we do not have significant error magnification. Of course, if the channel estimation errors are large, then the excess MSE may be large, irrespective of the size of the elements of $\mathbf{w}_o$.

B. Noise SOS Estimation Errors

When the channel and the input SOS are perfectly known and the only inaccuracies are due to the noise SOS estimation errors $\Delta\mathbf{R}_{nn}$, it can be shown that

$$\begin{aligned}\Delta\mathbf{R}_{xy} = & \mathbf{0} = \Delta\mathbf{R}_{yx}^T, \quad \Delta\mathbf{R}_{yy} = \Delta\mathbf{R}_{nn} \\ \Delta\mathbf{R} = & \mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\Delta\mathbf{R}_{nn}\mathbf{R}_{yy}^{-1}\mathbf{R}_{yx}\end{aligned}$$

giving

$$\begin{aligned}\mathrm{SOT1} = & \mathbf{w}_o^T\Delta\mathbf{R}_{nn}\,\mathbf{R}_{yy}^{-1}\Delta\mathbf{R}_{nn}\,\mathbf{w}_o \\ \mathrm{SOT2}_1 = & \mathbf{w}_o^T\Delta\mathbf{R}_{nn}\,\mathbf{R}_{yy}^{-1}\mathbf{R}_{yx}\,\mathbf{M}\,\mathbf{R}_{xy}\,\mathbf{R}_{yy}^{-1}\Delta\mathbf{R}_{nn}\,\mathbf{w}_o \\ \mathrm{SOT2}_2 = & \frac{1}{\mathcal{M}}\,\left(\mathbf{w}_o^T\Delta\mathbf{R}_{nn}\,\mathbf{w}_o\right)^2.\end{aligned}$$

Using the relation

$$\Delta\mathbf{R}_{nn}\mathbf{w}_o = \mathrm{Vec}\,(\Delta\mathbf{R}_{nn}\mathbf{w}_o) = \left(\mathbf{w}_o^T \otimes \mathbf{I}_{pN_f}\right)\mathrm{Vec}\,(\Delta\mathbf{R}_{nn})$$

where $\mathrm{Vec}(\cdot)$ denotes the vectorization operator and $\otimes$ denotes the Kronecker product, the previous terms become

$$\begin{aligned}\mathrm{SOT1} = & \mathrm{Vec}^T\,(\Delta\mathbf{R}_{nn})\,\mathcal{B}_1\,\mathrm{Vec}\,(\Delta\mathbf{R}_{nn}) \\ \mathrm{SOT2}_1 = & \mathrm{Vec}^T\,(\Delta\mathbf{R}_{nn})\,\mathcal{B}_2\,\mathrm{Vec}\,(\Delta\mathbf{R}_{nn}) \\ \mathrm{SOT2}_2 = & \mathrm{Vec}^T\,(\Delta\mathbf{R}_{nn})\,\mathcal{B}_3\,\mathrm{Vec}\,(\Delta\mathbf{R}_{nn})\end{aligned}$$

where

$$\begin{aligned}\mathcal{B}_1 \triangleq & \left(\mathbf{w}_o^T \otimes \mathbf{I}_{pN_f}\right)^T\mathbf{R}_{yy}^{-1}\left(\mathbf{w}_o^T \otimes \mathbf{I}_{pN_f}\right) \\ \mathcal{B}_2 \triangleq & \left(\mathbf{w}_o^T \otimes \mathbf{I}_{pN_f}\right)^T\mathbf{R}_{yy}^{-1}\mathbf{R}_{yx}\,\mathbf{M}\,\mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\left(\mathbf{w}_o^T \otimes \mathbf{I}_{pN_f}\right) \\ \mathcal{B}_3 \triangleq & \frac{1}{\mathcal{M}}\,\left(\mathbf{w}_o^T \otimes \mathbf{w}_o^T\right)^T\left(\mathbf{w}_o^T \otimes \mathbf{w}_o^T\right).\end{aligned}$$

If we define

$$\mathrm{Cov}\,(\mathrm{Vec}\,(\Delta\mathbf{R}_{nn})) \triangleq \mathcal{E}\left[\mathrm{Vec}\,(\Delta\mathbf{R}_{nn})\,\mathrm{Vec}^T\,(\Delta\mathbf{R}_{nn})\right]$$

then

$$\boxed{\mathcal{E}\,[\mathrm{SOT}] = \mathrm{Tr}\,[(\mathcal{B}_1 + \mathcal{B}_2 - \mathcal{B}_3)\,\mathrm{Cov}\,(\mathrm{Vec}\,(\Delta\mathbf{R}_{nn}))]\,.} \tag{37}$$

A simple and informative bound like (36) does not appear at hand. Direct application of the matrix product trace inequality gives

$$\mathcal{E}\,[\mathrm{SOT}] \leq ||\mathcal{B}_1 + \mathcal{B}_2 - \mathcal{B}_3||_F\,||\mathrm{Cov}\,(\mathrm{Vec}\,(\Delta\mathbf{R}_{nn}))\,||_F.$$

The magnification of the noise SOS errors depends on the size of $\mathbf{w}_o$, $\mathbf{R}_{yy}^{-1}$, and $\mathbf{M}$. A more informative and useful bound remains to be found.

C. Input SOS Estimation Errors

When the channel and the noise SOS are perfectly known and the sole inaccuracies are due to input SOS estimation errors $\Delta\mathbf{R}_{xx}$, it can be shown that

$$\begin{aligned}\Delta\mathbf{R}_{xy} = & \Delta\mathbf{R}_{xx}\,\mathbf{H}^T = \Delta\mathbf{R}_{yx}^T, \quad \Delta\mathbf{R}_{yy} = \mathbf{H}\Delta\mathbf{R}_{xx}\mathbf{H}^T \\ \Delta\mathbf{R} = & \mathbf{R}\,\Delta\mathbf{R}_{xx}\,\mathbf{R}.\end{aligned}$$

These simplifications result in

$$\begin{aligned}\mathrm{SOT1} = & \mathbf{g}_o^T\Delta\mathbf{R}_{xx}\mathbf{H}^T\mathbf{R}_{yy}^{-1}\mathbf{H}\Delta\mathbf{R}_{xx}\mathbf{g}_o \\ \mathrm{SOT2}_1 = & \mathbf{g}_o^T\Delta\mathbf{R}_{xx}\mathbf{R}\,\mathbf{M}\,\mathbf{R}\,\Delta\mathbf{R}_{xx}\,\mathbf{g}_o \\ \mathrm{SOT2}_2 = & \frac{1}{\mathcal{M}}\left(\mathbf{g}_o^T\Delta\mathbf{R}_{xx}\,\mathbf{g}_o\right)^2.\end{aligned}$$

Thus, we obtain

$$\begin{aligned}\mathrm{SOT1} = & \mathrm{Vec}^T\,(\Delta\mathbf{R}_{xx})\,\mathcal{C}_1\,\mathrm{Vec}\,(\Delta\mathbf{R}_{xx}) \\ \mathrm{SOT2}_1 = & \mathrm{Vec}^T\,(\Delta\mathbf{R}_{xx})\,\mathcal{C}_2\,\mathrm{Vec}\,(\Delta\mathbf{R}_{xx}) \\ \mathrm{SOT2}_2 = & \mathrm{Vec}^T\,(\Delta\mathbf{R}_{xx})\,\mathcal{C}_3\,\mathrm{Vec}\,(\Delta\mathbf{R}_{xx})\end{aligned}$$

where

$$\begin{aligned}\mathcal{C}_1 \triangleq & \left(\mathbf{g}_o^T \otimes \mathbf{I}_{N_f+\nu}\right)^T\mathbf{H}^T\mathbf{R}_{yy}^{-1}\mathbf{H}\left(\mathbf{g}_o^T \otimes \mathbf{I}_{N_f+\nu}\right) \\ \mathcal{C}_2 \triangleq & \left(\mathbf{g}_o^T \otimes \mathbf{I}_{N_f+\nu}\right)^T\mathbf{R}\,\mathbf{M}\,\mathbf{R}\left(\mathbf{g}_o^T \otimes \mathbf{I}_{N_f+\nu}\right) \\ \mathcal{C}_3 \triangleq & \frac{1}{\mathcal{M}}\left(\mathbf{g}_o^T \otimes \mathbf{g}_o^T\right)^T\left(\mathbf{g}_o^T \otimes \mathbf{g}_o^T\right).\end{aligned}$$

If we define

$$\mathrm{Cov}\,(\mathrm{Vec}\,(\Delta\mathbf{R}_{xx})) \triangleq \mathcal{E}\left[\mathrm{Vec}\,(\Delta\mathbf{R}_{xx})\,\mathrm{Vec}^T\,(\Delta\mathbf{R}_{xx})\right]$$

then

$$\boxed{\mathcal{E}[\mathrm{SOT}] = \mathrm{Tr}\,[(\mathcal{C}_1 + \mathcal{C}_2 - \mathcal{C}_3)\,\mathrm{Cov}\,(\mathrm{Vec}\,(\Delta\mathbf{R}_{xx}))]\,.} \tag{38}$$

Ignoring the term $\mathcal{C}_3$, using (31) and (34), and following the same chain of inequalities as in (36), we obtain

$$\begin{aligned}\mathcal{E}[\mathrm{SOT}] \leq & 2\left\|\mathbf{g}_o^T \otimes \mathbf{I}_{N_f+\nu}\right\|_F^2\,||\mathrm{Cov}\,(\mathrm{Vec}\,(\Delta\mathbf{R}_{xx}))||_F \\ = & 2\,(N_f+\nu)\,||\mathbf{g}_o||_2^2\,||\mathrm{Cov}\,(\mathrm{Vec}\,(\Delta\mathbf{R}_{xx}))||_F\,.\end{aligned} \tag{39}$$

This expression implies that in the cases where the terms of the residual impulse response $\mathbf{g}_o$ are small, the MMSE-DFE is robust w.r.t. input SOS estimation errors. In [6], the authors had observed that the MMSE-DFE is surprisingly insensitive w.r.t.

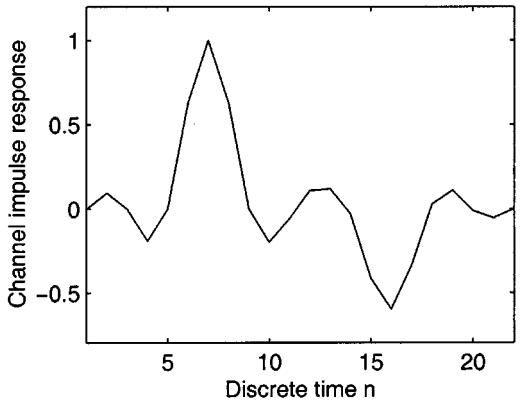

Fig. 3. Channel impulse response.

input SOS variations. The analysis of this subsection may serve as an explanation of this phenomenon.

V. Simulations

In our simulations, we use the communication channel whose impulse response is depicted in Fig. 3. It models a multipath scenario resulting in severe ISI and is derived by oversampling, by a factor of 2, the continuous-time channel impulse response

$$h(t) = p(t) - 0.6\,p(t - 4.45\,T)$$

where $p(t)$ is the (truncated) pulse with a raised-cosine spectrum and roll-off factor $\beta = 0.22$. The truncation interval of $p(t)$ is $[-3T, 3T]$, where T denotes the symbol period, and the sampling instants are the integer multiples of $T/2$. The vector impulse response $\{\mathbf{h}_i\}_{i=0}^{10}$ of the corresponding one-input/two-output system is constructed by grouping together the even and the odd terms of this oversampled impulse response.

The input is a BPSK signal, taking, with equal probability, the values ± 1, yielding $\mathbf{R}_{xx} = \mathbf{I}_{N_f+\nu}$. At the multichannel output, we add temporally and spatially white Gaussian noise with variance σ_n^2. Hence, $\mathbf{R}_{nn} = \sigma_n^2 \mathbf{I}_{2N_f}$. We define the SNR as

$$\text{SNR} \triangleq 10 \log_{10} \frac{\mathcal{E}\left[\|\mathbf{w}_n\|_2^2\right]}{\mathcal{E}\left[\|\mathbf{n}_n\|_2^2\right]}$$

where $\mathbf{w}_n$ is the noiseless output at time n, which is defined as $\mathbf{w}_n \triangleq \sum_{i=0}^{\nu} \mathbf{h}_i x_{n-i}$.

In the sequel, we consider the performance of the MMSE-DFE with filter lengths $N_f = 8$ and $N_b = 6$ for the delay $\Delta = 5$.

1) Channel Estimation Errors: At first, we assume that the input and additive white Gaussian channel noise SOS are perfectly known. In addition, we assume that the channel order is perfectly known. The channel is estimated by the application of the maximum-likelihood (least-squares) method to the training sequence consisting of $N_{\text{tr}} = 26$ consecutive training symbols (a packet consists of $N_p = 148$ data symbols) [5, Sect. 15.2]. The channel estimate is used for the computation of the DFE filters. In Fig. 4, we plot the MMSE $\mathcal{M}$ (dotted line), the mean of the theoretical excess MSE $\mathcal{E}[\widehat{\mathcal{M}} - \mathcal{M}]$ (solid line), $\mathcal{E}$ [SOT] (dashed line), and the bound of the third line of (36). Quantity $\mathcal{E}[\widehat{\mathcal{M}} - \mathcal{M}]$ is experimentally computed over 500 independent input and additive noise realizations (the training sequence remains the same over these realizations) as

$$\mathcal{E}[\widehat{\mathcal{M}} - \mathcal{M}] = \frac{1}{500} \sum_{i=1}^{500} \left(\widehat{\mathcal{M}}_i - \mathcal{M}\right)$$

where $\widehat{\mathcal{M}}_i - \mathcal{M}$ is the excess MSE of the ith realization. The channel estimation error covariance matrix $\mathbf{R}_{\Delta\mathcal{H}_\nu}$, which is used for the computation of $\mathcal{E}$ [SOT], can be computed by extending results of [5, Sect. 15.2] to the one-input/two-output channel setting. We observe the following.

1) For SNR lower than 10 dB, $\mathcal{E}$ [SOT] slightly overestimates $\mathcal{E}[\widehat{\mathcal{M}} - \mathcal{M}]$, whereas for higher SNRs, the two quantities practically coincide. This has been observed in many simulations with various channel shapes, validating the usefulness of the second-order approximation SOT to the excess MSE.
2) For all SNRs, the excess MSE is larger than the MMSE. The same has been observed in simulations with severe ISI channels, whereas for less severe channels (i.e., channels with small delay spread), the excess MSE is usually smaller than the MMSE.
3) The bound follows, in general, the changes of the mean excess MSE but is conservative.

In Fig. 5, we depict the case where $\hat{\nu} = 8$, that is, the assumed channel order is smaller than the true channel order (recall that

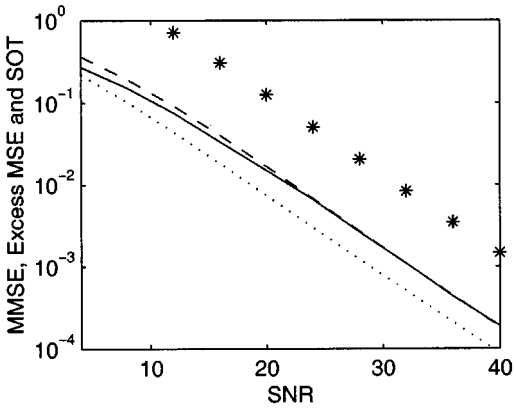

Fig. 4. Channel estimation errors: MMSE $\mathcal{M}$ (dotted line), mean excess MSE $\mathcal{E}[\widehat{\mathcal{M}} - \mathcal{M}]$ (solid line), $\mathcal{E}$[SOT] (dashed line), and bound (36). ($^*-$) versus SNR.

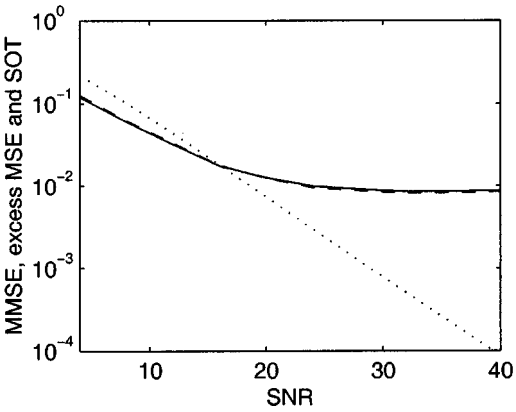

Fig. 5. Channel estimation errors (undermodeling). MMSE $\mathcal{M}$ (dotted line), mean excess MSE $\mathcal{E}[\widehat{\mathcal{M}} - \mathcal{M}]$ (solid line), and $\mathcal{E}$[SOT] (dashed line) versus SNR.

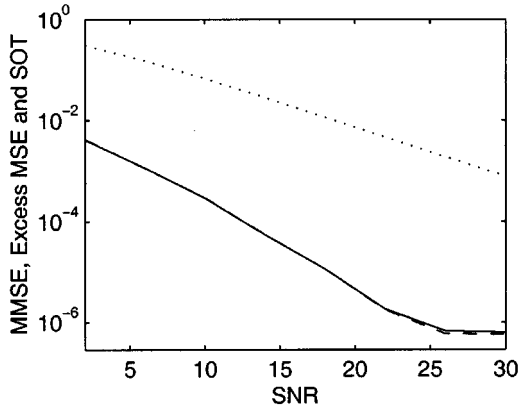

Fig. 6. Noise SOS estimation errors. MMSE $\mathcal{M}$ (dotted line), mean excess MSE $\mathcal{E}[\widehat{\mathcal{M}}-\mathcal{M}]$ (solid line), and $\mathcal{E}[\mathrm{SOT}]$ (dashed line) versus SNR.

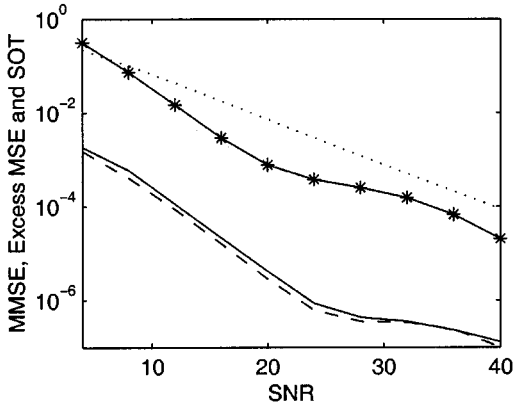

Fig. 7. Input SOS estimation errors. MMSE $\mathcal{M}$ (dotted line), mean excess MSE $\mathcal{E}[\widehat{\mathcal{M}}-\mathcal{M}]$ (solid line), $\mathcal{E}[\mathrm{SOT}]$ (dashed line), and bound (39) ($^*-$) versus SNR.

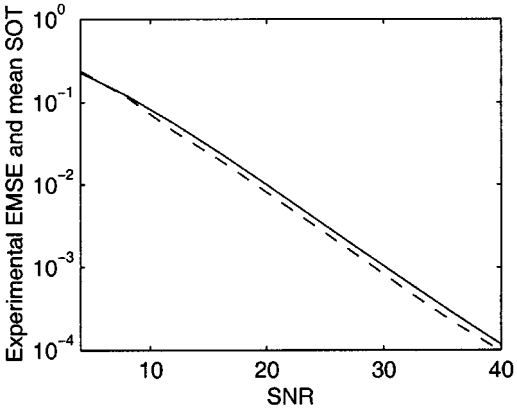

Fig. 8. Experimentally computed excess MSE (using the actual past decisions) (dashed line) and $\mathcal{E}[\mathrm{SOT}]$ (solid line) versus SNR.

$\nu = 10$). More specifically, our least-squares channel estimation procedure ignores the last two terms of the true channel impulse response. With the dotted line, we plot the MMSE-DFE performance assuming perfect channel knowledge, whereas with the dashed and solid lines, we plot, respectively, the mean excess MSE and its second-order approximation $\mathcal{E}[\mathrm{SOT}]$ (the channel estimation error covariance matrix has been computed experimentally). We observe that $\mathcal{E}[\mathrm{SOT}]$ provides a very accurate estimate for the mean excess MSE. Furthermore, we observe that for low SNR, (slight) undermodeling does not lead to dramatic performance degradation of the MMSE-DFE, whereas for high SNR, there is an unavoidable error floor, which depends on the size of unmodeled part.

2) Noise SOS Estimation Errors: Next, we assume that the channel impulse response and input SOS are perfectly known, and we compute the excess MSE introduced by the estimation of the noise variance (we assume that the additive channel noise is white Gaussian, as is the case). Using the training input samples and the channel impulse response, we compute *correctly* the corresponding $(N_{\mathrm{tr}}-\nu)$ 2-D noise samples $\mathbf{n}_n$, for $n = n_1+1,\ldots,n_1+N_{\mathrm{tr}}-\nu$, where n_1 depends on the position of the training sequence in the input data packet. Then, we estimate the noise variance as

$$\widehat{\sigma}^2 = \frac{1}{2(N_{\mathrm{tr}}-\nu)} \sum_{n=n_1+1}^{n_1+N_{\mathrm{tr}}-\nu} \mathbf{n}_n^T \mathbf{n}_n.$$

It can be shown that $\widehat{\sigma}^2$ is an unbiased estimate of σ^2 with variance

$$\mathcal{E}\left[\left(\widehat{\sigma}^2-\sigma^2\right)^2\right] = \frac{\sigma^4}{N_{\mathrm{tr}}-\nu}.$$

Thus, we derive

$$\mathrm{Cov}\left(\mathrm{Vec}\left(\Delta\mathbf{R}_{nn}\right)\right) = \frac{\sigma^4}{N_{\mathrm{tr}}-\nu}\,\mathrm{Vec}\left(\mathbf{I}_{pN_f}\right)\,\mathrm{Vec}^T\left(\mathbf{I}_{pN_f}\right).$$

In Fig. 6, we plot the MMSE $\mathcal{M}$ (dotted line), the mean theoretical excess MSE $\mathcal{E}[\widehat{\mathcal{M}}-\mathcal{M}]$ (solid line), and $\mathcal{E}[\mathrm{SOT}]$ (dashed line). We observe that the second-order approximation provides very accurate estimates of the excess MSE for all SNRs and that the excess MSE due to noise variance estimation errors is much lower than the MMSE.

3) Input SOS Estimation Errors: Finally, we consider the influence of the input SOS estimation errors. To this end, we assume that the channel impulse response and the noise SOS are perfectly known and that the error is made in the input variance. Thus, for illustration purposes, we assume that the input is white with "estimated" variance $1+\delta\sigma_x^2$, where $\delta\sigma_x^2$ is a zero-mean random variable with variance $\mathrm{var}(\delta\sigma_x^2)$ (we just want to check the accuracy of (38) without having in mind a particular estimator of the input variance). In Fig. 7, we plot the MMSE $\mathcal{M}$ (dotted line), the mean theoretical excess MSE $\mathcal{E}[\widehat{\mathcal{M}}-\mathcal{M}]$ (solid line), $\mathcal{E}[\mathrm{SOT}]$ (dashed line), and bound (39), resulting from $\mathrm{var}(\delta\sigma_x^2) = 0.04$. We observe that the second-order approximation provides very accurate estimates of the excess MSE for all SNRs and that the excess MSE due to noise variance estimation errors is much lower than the MMSE. Bound (39) is pessimistic. Furthermore, we observe that the mean excess MSE decreases for increasing the SNR despite the fact that $\mathrm{var}(\delta\sigma_x^2)$ remains constant over all SNRs. This may be explained by the fact that the terms of the residual impulse response $\mathbf{g}_o$, which govern the excess MSE in this case, decrease for increased SNR.

We emphasize that quantities $\mathcal{M}$, $\widehat{\mathcal{M}}$ and S.O.T. have been computed by assuming that previous decisions were correct. In Fig. 8, we plot the experimentally computed mean excess MSE (using the *actual* past decisions) and $\mathcal{E}[\mathrm{SOT}]$. We observe that $\mathcal{E}[\mathrm{SOT}]$ is a very accurate measure of the actual excess MSE.

VI. Conclusion

We considered the behavior of the finite-length MMSE-DFE in cases of channel and SOS estimation errors. Assuming that the estimation errors are small, we derived a second-order approximation to the excess MSE. We also derived second-order approximations to the mean excess MSE in terms of the parameter estimation error covariance matrices and simple and informative bounds for the excess MSE. The sensitivity w.r.t. channel estimation errors is mainly determined by the size of the elements of the feedforward filter, whereas the sensitivity w.r.t. input SOS errors is governed by the size of the terms of the residual impulse response. Simulations were in agreement with our theoretical results and showed that the second-order approximation to the excess MSE is an accurate measure of the actual excess MSE.

Appendix

Let the symmetric positive definite matrix A be partitioned as

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}$$

where A_{11}, A_{22}, and A_{33} are square, and $A_{21} = A_{12}^T$, $A_{31} = A_{13}^T$, and $A_{32} = A_{23}^T$, and let B be the matrix whose nonzero part corresponds to A_{22} and equals A_{22}^{-1}, i.e.,

$$B = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & A_{22}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

The difference $A - A\,B\,A$ is

$$\begin{bmatrix} A_{11} - A_{12}A_{22}^{-1}A_{21} & \mathbf{0} & A_{13} - A_{12}A_{22}^{-1}A_{23} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ A_{31} - A_{32}A_{22}^{-1}A_{21} & \mathbf{0} & A_{33} - A_{32}A_{22}^{-1}A_{23} \end{bmatrix}.$$

Let us define

$$C \triangleq \begin{bmatrix} A_{11} - A_{12}A_{22}^{-1}A_{21} & A_{13} - A_{12}A_{22}^{-1}A_{23} \\ A_{31} - A_{32}A_{22}^{-1}A_{21} & A_{33} - A_{32}A_{22}^{-1}A_{23} \end{bmatrix}.$$

Since $A > 0$, we have that for any nonzero vectors x_1, x_2, and x_3 of appropriate dimensions

$$[\,x_1^T \quad x_2^T \quad x_3^T\,] \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} > 0.$$

In particular, for $x_2 = -A_{22}^{-1}A_{21}\,x_1 - A_{22}^{-1}A_{23}\,x_3$, we obtain that

$$[\,x_1^T \quad x_3^T\,]\,C \begin{bmatrix} x_1 \\ x_3 \end{bmatrix} > 0$$

proving that $A - A\,B\,A$ is semi-positive definite. Putting $\mathbf{R}$ in the place of A and $\mathbf{M}$ in the place of B, we obtain (33).

Acknowledgment

The author is very grateful to an anonymous reviewer for the very insightful review that significantly helped to improve the quality of the paper.

References

[1] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.

[2] N. Al-Dhahir and J. M. Cioffi, "MMSE decision-feedback equalizers: Finite-length results," *IEEE Trans. Inform. Theory*, vol. 41, pp. 961–975, July 1995.

[3] ——, "Mismatched finite-complexity MMSE desicion feedback equalizers," *IEEE Trans. Signal Processing*, vol. 45, pp. 935–944, Apr. 1997.

[4] G. W. Stewart and J.-G. Sun, *Matrix Perturbation Theory*. New York: Academic, 1990.

[5] H. Meyr, M. Moeneclaey, and S. Fechtel, *Digital Communication Receivers: Synchronization, Channel Estimation and Signal Processing*. New York: Wiley, 1998.

[6] J. Cioffi *et al.*, "MMSE decision-feedback equalizers and coding-part II: Coding results," *IEEE Trans. Commun.*, vol. 43, pp. 2582–2594, Oct. 1995.

Athanasios P. Liavas (M'94) was born in Pyrgos, Greece, in 1966. He received the diploma and the Ph.D. degrees in computer engineering from the University of Patras, Patras, Greece, in 1989 and 1993, respectively.

From 1996 to 1998, he was a research fellow at the Insitut National des Télécommunications, Evry, France, under the framework of the Training and Mobility of Researchers (TMR) program of the European Commission. In 2001, he joined the Department of Mathematics, University of the Aegean, Samos, Greece, as an assistant professor. His research interests include signal processing for communications and biomedical signal processing.

Dr. Liavas is a member of the Technical Chamber of Greece.