On the Performance of the Mismatched MMSE and the LS Linear Equalizers

Athanasios P. Liavas, *Member, IEEE*, and Despoina Tsipouridou

***Abstract*—We consider two widely referenced trained finite-length linear equalizers, namely, the mismatched minimum mean square error (MMSE) equalizer and the least-squares (LS) equalizer. Using matrix perturbation theory, we express both of them as perturbations of the ideal MMSE equalizer and we derive insightful analytical expressions for their excess mean square error. We observe that, in general, the mismatched MMSE equalizer performs (much) better than the LS equalizer. We attribute this phenomenon to the fact that the LS equalizer *implicitly* estimates the input second-order statistics, while the mismatched MMSE equalizer uses perfect knowledge. Thus, assuming that the input second-order statistics are known at the receiver, which is usually the case, the use of the mismatched MMSE equalizer is preferable, in general.**

***Index Terms*—Intersymbol interference, least-squares (LS) equalization, minimum mean square error (MMSE) equalization, performance analysis.**

I. Introduction

LINEAR equalization is a well-known receiver technique for combatting intersymbol interference (ISI). Linear equalizers are usually computed by minimization of either the mean square error (MSE) cost function or the least-squares (LS) error cost function.

If we know the channel impulse response and the input and noise first- and second-order statistics, then we can compute the minimum MSE (MMSE) equalizer [1, Sec. 2.7.3]. An assumption that will be used throughout this work is that the input is i.i.d., with zero-mean and unit variance and the noise is white Gaussian (thus, the receiver can exploit this information). Practically always, the channel impulse response and the noise variance are unknown at the receiver. A common approach toward the design of the MMSE equalizer, in the cases where the channel impulse response and the noise second-order statistics are unknown at the receiver, is to estimate them using training data and, then, use the estimates as if they were the true quantities. We denote the resulting equalizer *mismatched MMSE* equalizer.

On the other hand, in the direct LS equalization approach, we use the training data and directly compute at the receiver the LS optimal equalizer without the intermediate computation of the channel impulse response and the noise second-order statistics (see, for example, [2, Sec. 13.2.1] for the real-valued case). We note that the LS equalizer cannot exploit the knowledge about the input second-order statistics that is available at the receiver.

Despite the fact that these approaches are widely referenced, there does not exist, to our knowledge, any study concerning their relative performance. In this paper, we compare their performance by adopting as performance measure the MSE. Using matrix perturbation theory, we express both the mismatched MMSE and the LS equalizers as perturbations of the ideal MMSE equalizer, that is, the equalizer that minimizes the MSE by assuming exact knowledge of the true quantities. Then, we derive second-order approximations to the excess MSE associated with the mismatched MMSE and the LS equalizers (computation of the excess MSE for many adaptive algorithms appears in [1, Ch. 7] and the references therein). The analytic expressions provide significant insight into the behavior of the above mentioned equalizers. In general, the mismatched MMSE equalizer performs (much) better than the LS equalizer. The main reason for this superiority is the fact that the LS equalizer *implicitly* estimates the input second-order statistics, while the mismatched MMSE equalizer uses perfect knowledge.

A. Notation and Matrix Results

Superscripts ^{T}, *, and ^{H} denote, respectively, transpose, componentwise conjugate, and conjugate transpose. $\mathbf{E}[\cdot]$ denotes expectation, $\mathbf{I}_M$ denotes the $M \times M$ identity matrix, $\mathbf{0}_{L\times M}$ denotes the $L \times M$ zero matrix, $\mathrm{tr}(\mathbf{A})$ and $\|\mathbf{A}\|_F$ denote, respectively, the trace and the Frobenius norm of matrix $\mathbf{A}$, and $\|\mathbf{x}\|_2$ denotes the Euclidean norm of vector $\mathbf{x}$. $\mathbf{A}\otimes\mathbf{B}$ denotes the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$ and $\mathrm{vec}(\cdot)$ denotes the vectorization operator. We remind that for matrices with compatible dimensions [3, p. 19]

$$(\mathbf{A}\otimes\mathbf{C})(\mathbf{B}\otimes\mathbf{D}) = \mathbf{AB}\otimes\mathbf{CD} \tag{1}$$

and [3, p. 17]

$$\mathrm{tr}(\mathbf{ABCD}) = \mathrm{vec}^T\left(\mathbf{D}^T\right)\left(\mathbf{C}^T\otimes\mathbf{A}\right)\mathrm{vec}(\mathbf{B}). \tag{2}$$

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

$$(\mathbf{A}+\Delta\mathbf{A})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\Delta\mathbf{A}\mathbf{A}^{-1}. \tag{3}$$

The rest of the paper is structured as follows. In Section II, we present the ISI channel model. In Sections III and IV, respectively, we consider the mismatched MMSE and the LS equalizers and we develop second-order approximations to their excess MSE. These approximations provide significant insight into the behavior of the mismatched MMSE and the LS equalizer and lead to a comparison of their performance in

Manuscript received April 8, 2006; revised October 27, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Yuan-Pei Lin.

The authors are with the Department of Electronic and Computer Engineering, Technical University of Crete, 73100 Chania, Greece (e-mail: liavas@telecom.tuc.gr; despoina@telecom.tuc.gr).

Digital Object Identifier 10.1109/TSP.2007.894392

1053-587X/$25.00 © 2007 IEEE

Section V. In Section VI, we illustrate our theoretical results with simulations and we conclude the paper in Section VII.

II. Channel Model

We consider the discrete-time baseband-equivalent noisy communication channel described by the convolution

$$y_n = \sum_{k=0}^{M} h_k s_{n-k} + w_n \tag{4}$$

where s_n, y_n and w_n denote, respectively, the channel input, output and noise samples at time instant n. The channel impulse response is denoted $\mathbf{h} \triangleq [h_0 \dots h_M]^T$. Communication is based on length-N data packets $\{s_1, \dots, s_N\}$. By stacking $L+1$ consecutive output samples, we obtain $\mathbf{y}_{n:n-L} \triangleq [y_n \quad \dots \quad y_{n-L}]^T$[1], which can be expressed as

$$\mathbf{y}_{n:n-L} = \mathbf{H}\mathbf{s}_{n:n-L-M} + \mathbf{w}_{n:n-L} \tag{5}$$

where $\mathbf{H}$ is the $(L+1)\times(L+M+1)$ filtering matrix defined as

$$\mathbf{H} \triangleq \begin{bmatrix} h_0 & \cdots & h_M & & \\ & \ddots & & \ddots & \\ & & h_0 & \cdots & h_M \end{bmatrix}. \tag{6}$$

The noise samples are assumed to be i.i.d., zero-mean, circularly symmetric complex Gaussian with variance σ_w^2. The input samples are assumed to be i.i.d., zero-mean, complex-valued circular, with variance 1.

III. Finite-Length MMSE Linear Equalizer

A. Ideal Finite-Length MMSE Linear Equalizer

An order-L delay-d linear equalizer is determined by vector $\mathbf{f} \triangleq [f_0 \quad \dots \quad f_L]^T$. Its output at time instant n, $\tilde{s}_{n-d}$, is an estimate of the (delayed) channel input, s_{n-d}, and is given by

$$\tilde{s}_{n-d} = \sum_{i=0}^{L} f_i^* y_{n-i} = \mathbf{f}^H \mathbf{y}_{n:n-L}. \tag{7}$$

The symbol estimation error at time instant n, e_n, is expressed as

$$e_n \triangleq \tilde{s}_{n-d} - s_{n-d} = \mathbf{f}^H \mathbf{y}_{n:n-L} - \mathbf{e}_d^H \mathbf{s}_{n:n-L-M}$$

where $\mathbf{e}_d$ is the $(L+M+1)\times 1$ vector with 1 at the $(d+1)$st position and zeros elsewhere.

The mean-square symbol estimation error can be expressed as a function of $\mathbf{f}$ as follows:

$$\begin{aligned}\mathbf{MSE}(\mathbf{f}) &\triangleq \mathbf{E}\left[|e_n|^2\right] \\ &= \mathbf{f}^H \mathbf{R}_\mathbf{y} \mathbf{f} - \mathbf{f}^H \mathbf{H} \mathbf{e}_d - \mathbf{e}_d^H \mathbf{H}^H \mathbf{f} + 1\end{aligned} \tag{8}$$

where

$$\mathbf{R}_\mathbf{y} \triangleq \mathbf{E}\left[\mathbf{y}_{n:n-L}\mathbf{y}_{n:n-L}^H\right] = \mathbf{H}\mathbf{H}^H + \sigma_w^2 \mathbf{I}_{L+1}.$$

[1]If $\mathbf{a}$ is a vector and $i > j$, we define $\mathbf{a}_{i:j} \triangleq [a_i \; a_{i-1} \; \dots \; a_{j+1} \; a_j]^T$.

The order-L delay-d ideal MMSE linear equalizer, $\mathbf{f}_o$, is given by the expression [1, Sec. 2.7.3]

$$\mathbf{f}_o = \mathbf{R}_\mathbf{y}^{-1} \mathbf{H} \mathbf{e}_d. \tag{9}$$

If we know the channel impulse response and the input and noise second-order statistics, then we may proceed to the design of the MMSE equalizer. Usually, the channel $\mathbf{h}$ and the noise variance σ_w^2 are unknown at the receiver (recall that we have set the variance of the i.i.d. input equal to 1). A common approach towards the design of the MMSE equalizer is to estimate $\mathbf{h}$ and σ_w^2 using training data and then use the estimates as if they were the true quantities.

In the sequel, we shall assume that the $N_{\rm tr}$ input samples $\{s_{n_1}, \dots, s_{n_2}\}$, with $N_{\rm tr} \triangleq n_2 - n_1 + 1$, are known at the receiver and used for training purposes.

B. Channel and Noise Variance Estimation

In this subsection, we assume that the channel $\mathbf{h}$ is constant but unknown and we estimate it using the maximum likelihood (ML) method. Collecting the channel output samples that depend *only* on the training samples, we obtain

$$\mathbf{y}_{n_2:n_1+M} = \mathbf{S}_1 \mathbf{h} + \mathbf{w}_{n_2:n_1+M}$$

where

$$\mathbf{S}_1 \triangleq \begin{bmatrix} s_{n_2} & \cdots & s_{n_2-M} \\ \vdots & \ddots & \vdots \\ s_{n_1+M} & \cdots & s_{n_1} \end{bmatrix}.$$

The ML channel estimate, which, in this case, coincides with the LS estimate, is given by [1, p. 697]

$$\begin{aligned}\hat{\mathbf{h}} &= \left(\mathbf{S}_1^H \mathbf{S}_1\right)^{-1} \mathbf{S}_1^H \mathbf{y}_{n_2:n_1+M} \\ &= \mathbf{h} + \left(\mathbf{S}_1^H \mathbf{S}_1\right)^{-1} \mathbf{S}_1^H \mathbf{w}_{n_2:n_1+M}.\end{aligned}$$

The channel estimation error, $\Delta\mathbf{h} \triangleq \hat{\mathbf{h}} - \mathbf{h}$, is of first-order with respect to the noise, that is, its elements are linear combinations of the noise samples $\{w_{n_1+M}, \dots, w_{n_2}\}$. It can be easily shown that $\Delta\mathbf{h}$ is zero-mean, circular, with covariance matrix

$$\mathbf{R}_{\Delta\mathbf{h}} \triangleq \mathbf{E}\left[\Delta\mathbf{h}\Delta\mathbf{h}^H\right] = \sigma_w^2 \left(\mathbf{S}_1^H \mathbf{S}_1\right)^{-1}. \tag{10}$$

In the above expression, we regarded the training sequence as deterministic. If we consider it as stochastic, then we must take expectation with respect to the training samples as well (in Appendix, we provide guidelines for this computation). $\Delta\mathbf{h}$ is independent of s_n and w_n for $n \notin \{n_1, \dots, n_2\}$. Thus, we consider $\Delta\mathbf{h}$ as a random vector independent of the random input and noise.

An unbiased estimate of the noise variance is given by [1, p. 697]

$$\hat{\sigma}_w^2 = \frac{1}{N_{\rm tr}+1}(\mathbf{y} - \mathbf{S}_1\hat{\mathbf{h}})^H(\mathbf{y} - \mathbf{S}_1\hat{\mathbf{h}}). \tag{11}$$

The noise variance estimation error, $\Delta\sigma_w^2 \triangleq \hat{\sigma}_w^2 - \sigma_w^2$, is of second-order with respect to the noise, that is, it is linear combination of terms that are products of independent noise samples (in fact, it can be shown that the variance of $\hat{\sigma}_w^2$ is $(\sigma_w^4/(N_{\rm tr}+1)))$. Consequently, it is negligible in comparison with $\Delta\mathbf{h}$, for sufficiently high SNR. Thus, in the sequel, we shall assume that $\hat{\sigma}_w^2 = \sigma_w^2$.

Minimization of $\mathrm{tr}(\mathbf{R}_{\Delta\mathbf{h}})$ is achieved by selecting the training sequence such that [5, pp. 787–788]

$$\mathbf{S}_1^H\mathbf{S}_1 = (N_{\rm tr} - M)\mathbf{I}_{M+1}. \tag{12}$$

Training sequences that closely satisfy this constraint can be constructed by periodic extension, with period P, of the so-called CAZAC (constant amplitude zero autocorrelation) sequences, with examples given in [5, Table 15-3]. It can be easily checked, e.g., by a computer search, that if $P > (M+1)$ and $N_{\rm tr} = kP + M$, with k being a positive integer, then the corresponding training sequences *perfectly* satisfy constraint (12). In this case

$$\mathbf{R}_{\Delta\mathbf{h}} = \frac{\sigma_w^2}{N_{\rm tr} - M}\mathbf{I}_{M+1}. \tag{13}$$

C. Computation of the Excess MSE of the Mismatched MMSE Equalizer

The analysis of this subsection resembles that of Section IV-A of [6]. In order to preserve the readability of the paper and to introduce the definitions of the involved quantities, we briefly present the whole analysis.

If we use in (9) the channel estimate $\hat{\mathbf{h}}$ as if it were the true channel, we compute the mismatched MMSE equalizer

$$\hat{\mathbf{f}} = \hat{\mathbf{R}}_{\mathbf{y}}^{-1}\hat{\mathbf{H}}\mathbf{e}_d = \left(\hat{\mathbf{H}}\hat{\mathbf{H}}^H + \sigma_w^2\mathbf{I}_{L+1}\right)^{-1}\hat{\mathbf{H}}\mathbf{e}_d \tag{14}$$

where $\hat{\mathbf{H}}$ is the $(L+1)\times(L+M+1)$ filtering matrix constructed by $\hat{\mathbf{h}}$. For later use, we define $\Delta\mathbf{H} \triangleq \hat{\mathbf{H}} - \mathbf{H}$. Applying $\hat{\mathbf{f}}$ for input estimation, we obtain the estimation error

$$\hat{e}_n = \hat{\mathbf{f}}^H\mathbf{y}_{n:n-L} - \mathbf{e}_d^H\mathbf{s}_{n:n-L-M}. \tag{15}$$

Taking expectation, with respect to the input and the noise, we obtain

$$\begin{aligned}\mathbf{E}\left[|\hat{e}_n|^2\right] &= \hat{\mathbf{f}}^H\mathbf{R}_{\mathbf{y}}\hat{\mathbf{f}} - \hat{\mathbf{f}}^H\mathbf{H}\mathbf{e}_d - \mathbf{e}_d^H\mathbf{H}^H\hat{\mathbf{f}} + 1\\ &= \mathbf{MSE}(\hat{\mathbf{f}}).\end{aligned} \tag{16}$$

Taylor expansion of function $\mathbf{MSE}(\cdot)$ around the point $\mathbf{f}_o$ leads to

$$\mathbf{MSE}(\hat{\mathbf{f}}) = \mathbf{MSE}(\mathbf{f}_o) + \Delta\mathbf{f}^H\mathbf{R}_{\mathbf{y}}\Delta\mathbf{f} \tag{17}$$

where $\Delta\mathbf{f} \triangleq \hat{\mathbf{f}} - \mathbf{f}_o$. In the above expansion, we used the fact that the gradient of $\mathbf{MSE}(\cdot)$ at point $\mathbf{f}_o$ is zero, due to the optimality of $\mathbf{f}_o$.

Taking expectation, with respect to the channel estimation errors, we obtain the excess MSE (EMSE)

$$\begin{aligned}\mathbf{EMSE}(\hat{\mathbf{f}}) &\triangleq \mathbf{E}\left[\mathbf{MSE}(\hat{\mathbf{f}}) - \mathbf{MSE}(\mathbf{f}_o)\right]\\ &= \mathbf{E}\left[\Delta\mathbf{f}^H\mathbf{R}_{\mathbf{y}}\Delta\mathbf{f}\right].\end{aligned} \tag{18}$$

In the following, we derive an analytic expression for the EMSE achieved by the mismatched MMSE equalizer in terms of the channel estimation error covariance matrix, $\mathbf{R}_{\Delta\mathbf{h}}$.

We start by providing a first-order approximation to $\Delta\mathbf{f}$. Ignoring products of error terms inside the parenthesis in (14), we obtain

$$\hat{\mathbf{f}} = (\mathbf{R}_{\mathbf{y}} + \boldsymbol{\Delta}_1)^{-1}(\mathbf{H}\mathbf{e}_d + \boldsymbol{\delta}_1)$$

where

$$\boldsymbol{\Delta}_1 \triangleq \Delta\mathbf{H}\mathbf{H}^H + \mathbf{H}\Delta\mathbf{H}^H \tag{19}$$

$$\boldsymbol{\delta}_1 \triangleq \Delta\mathbf{H}\mathbf{e}_d. \tag{20}$$

Using approximation (3), we obtain

$$\hat{\mathbf{f}} = \left(\mathbf{R}_{\mathbf{y}}^{-1} - \mathbf{R}_{\mathbf{y}}^{-1}\boldsymbol{\Delta}_1\mathbf{R}_{\mathbf{y}}^{-1}\right)(\mathbf{H}\mathbf{e}_d + \boldsymbol{\delta}_1). \tag{21}$$

Ignoring products of error terms in (21), we obtain the first-order approximation

$$\hat{\mathbf{f}} = \mathbf{f}_o - \mathbf{R}_{\mathbf{y}}^{-1}(\boldsymbol{\Delta}_1\mathbf{f}_o - \boldsymbol{\delta}_1) \tag{22}$$

which leads to

$$\Delta\mathbf{f} = -\mathbf{R}_{\mathbf{y}}^{-1}(\boldsymbol{\Delta}_1\mathbf{f}_o - \boldsymbol{\delta}_1). \tag{23}$$

Using (18), we obtain the second-order approximation

$$\begin{aligned}\mathbf{EMSE}(\hat{\mathbf{f}}) &= \mathbf{E}[(\boldsymbol{\Delta}_1\mathbf{f}_o - \boldsymbol{\delta}_1)^H\mathbf{R}_{\mathbf{y}}^{-1}(\boldsymbol{\Delta}_1\mathbf{f}_o - \boldsymbol{\delta}_1)]\\ &= \mathrm{tr}\left(\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{E}\left[(\boldsymbol{\Delta}_1\mathbf{f}_o - \boldsymbol{\delta}_1)(\boldsymbol{\Delta}_1\mathbf{f}_o - \boldsymbol{\delta}_1)^H\right]\right).\end{aligned} \tag{24}$$

In order to compute the above expectation, we must express

$$\boldsymbol{\Delta}_1\mathbf{f}_o - \boldsymbol{\delta}_1 = \Delta\mathbf{H}\mathbf{H}^H\mathbf{f}_o + \mathbf{H}\Delta\mathbf{H}^H\mathbf{f}_o - \Delta\mathbf{H}\mathbf{e}_d$$

in terms of $\Delta\mathbf{h}$.

1) We start with $\Delta\mathbf{H}\mathbf{H}^H\mathbf{f}_o - \Delta\mathbf{H}\mathbf{e}_d$. If we define the combined (channel-MMSE equalizer) impulse response

$$\mathbf{c}_o \triangleq \mathbf{H}^T\mathbf{f}_o^* \tag{25}$$

and

$$\mathbf{r}_o \triangleq \mathbf{c}_o - \mathbf{e}_d = [r_0 \quad \ldots \quad r_{L+M+1}]^T \tag{26}$$

then it follows easily that

$$\begin{aligned}\Delta\mathbf{H}\mathbf{H}^H\mathbf{f}_o - \Delta\mathbf{H}\mathbf{e}_d &= \Delta\mathbf{H}(\mathbf{c}_o^* - \mathbf{e}_d)\\ &= \Delta\mathbf{H}\mathbf{r}_o^* = \mathbf{R}_o^*\Delta\mathbf{h}\end{aligned}$$

where $\mathbf{R}_o$ is the $(L+1)\times(M+1)$ Hankel matrix

$$\mathbf{R}_o \triangleq \begin{bmatrix} r_0 & r_1 & \cdots & r_M \\ r_1 & r_2 & \cdots & r_{M+1} \\ \vdots & \vdots & \iddots & \cdots \\ r_L & r_{L+1} & \cdots & r_{L+M+1} \end{bmatrix}.$$

2) Due to the commutativity property of the convolution, term $\mathbf{H}\Delta\mathbf{H}^H\mathbf{f}_o$ can be expressed as

$$\mathbf{H}\Delta\mathbf{H}^H\mathbf{f}_o = \mathbf{H}\mathbf{F}_o^T\Delta\mathbf{h}^* = \mathbf{G}_o\Delta\mathbf{h}^*.$$

where $\mathbf{F}_o$ is the $(M+1)\times(L+M+1)$ filtering matrix constructed by vector $\mathbf{f}_o$.

Thus, we obtain that

$$\boldsymbol{\Delta}_1\mathbf{f}_o - \boldsymbol{\delta}_1 = \mathbf{R}_o^*\Delta\mathbf{h} + \mathbf{G}_o\Delta\mathbf{h}^* \tag{27}$$

which, using (24), gives (note that the cross-terms vanish due to the circular symmetry of $\Delta\mathbf{h}$)

$$\boxed{\mathbf{EMSE}(\hat{\mathbf{f}}) = \operatorname{tr}\left(\mathbf{R}_{\mathbf{y}}^{-1}\left(\mathbf{R}_o^*\mathbf{R}_{\Delta\mathbf{h}}\mathbf{R}_o^T + \mathbf{G}_o\mathbf{R}_{\Delta\mathbf{h}}^*\mathbf{G}_o^H\right)\right).} \tag{28}$$

D. Simplifications in the High-SNR Cases

It can be easily shown that

$$\mathbf{MMSE} \triangleq \mathbf{MSE}(\mathbf{f}_o) = \|\mathbf{r}_o\|_2^2 + \sigma_w^2\|\mathbf{f}_o\|_2^2. \tag{29}$$

Thus, for sufficiently high SNR, mid-range delays and sufficiently large equalizer length (implying small MMSE), the first term of the sum in (28), which is constructed using $\mathbf{r}_o$, is of the order of the MMSE and thus negligible compared to the second (this claim is confirmed in the simulations section). In this case, we approximate the EMSE in (28) as

$$\mathbf{EMSE}(\hat{\mathbf{f}}) \approx \operatorname{tr}\left(\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{G}_o\mathbf{R}_{\Delta\mathbf{h}}^*\mathbf{G}_o^H\right). \tag{30}$$

The most insightful case arises when $\mathbf{R}_{\Delta\mathbf{h}} = (\sigma_w^2/(N_{\text{tr}}-M))\mathbf{I}_{M+1}$, that is, when we perform optimal channel estimation [recall (13)]. A first-order approximation to $\mathbf{R}_{\mathbf{y}}^{-1}$, with respect to σ_w^2, is

$$\mathbf{R}_{\mathbf{y}}^{-1} = (\mathbf{H}\mathbf{H}^H)^{-1} - \sigma_w^2(\mathbf{H}\mathbf{H}^H)^{-2}.$$

Reminding that $\mathbf{G}_o = \mathbf{H}\mathbf{F}_o^T$, expression (30) becomes

$$\begin{aligned}\mathbf{EMSE}(\hat{\mathbf{f}}) &\approx \frac{\sigma_w^2}{N_{\text{tr}}-M}\operatorname{tr}\left(\mathbf{G}_o^H\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{G}_o\right) \\ &\approx \frac{\sigma_w^2}{N_{\text{tr}}-M}\operatorname{tr}\left(\mathbf{F}_o^*\mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1}\mathbf{H}\mathbf{F}_o^T\right) \\ &\overset{(a)}{=} \frac{\sigma_w^2}{N_{\text{tr}}-M}\operatorname{tr}\left(\mathbf{F}_o^*\mathbf{P}_{\mathbf{H}^H}\mathbf{F}_o^T\right) \\ &\overset{(b)}{\leq} \frac{\sigma_w^2}{N_{\text{tr}}-M}\operatorname{tr}\left(\mathbf{F}_o^*\mathbf{F}_o^T\right) \\ &= \frac{\sigma_w^2}{N_{\text{tr}}-M}\|\mathbf{F}_o\|_F^2 \\ &\overset{(c)}{=} \frac{(M+1)\sigma_w^2}{N_{\text{tr}}-M}\|\mathbf{f}_o\|_2^2\end{aligned}$$

where at point (a) we defined the projector onto the column space of $\mathbf{H}^H$, $\mathbf{P}_{\mathbf{H}^H}$, at point (b) we used the positive semi-definiteness of $\mathbf{I}_{L+M+1} - \mathbf{P}_{\mathbf{H}^H}$ and at point (c) we used the structure of $\mathbf{F}_o$.

Thus, we have that[2]

$$\boxed{\mathbf{EMSE}(\hat{\mathbf{f}}) \lesssim \frac{(M+1)\|\mathbf{f}_o\|_2^2}{N_{\text{tr}}-M}\sigma_w^2.} \tag{31}$$

In extensive simulation studies, we have observed that this bound is a very good approximation to the (experimentally computed) true EMSE, for SNR higher than 5 dB. This approximation states that the EMSE of the mismatched MMSE equalizer is proportional to the channel noise variance σ_w^2 (note that, since $\sigma_s^2 = 1$, σ_w^2 equals the inverse SNR), with the proportionality coefficient determined by the parameters N_{tr} and M and the 2-norm of the ideal MMSE equalizer $\mathbf{f}_o$.

IV. TRAINED LS LINEAR EQUALIZER

The order-L delay-d trained LS linear equalizer solves the minimization problem (for the real-valued case, see, for example, [2, p. 249])

$$\mathbf{f}_{\text{LS}} = \arg\min_{\mathbf{f}} \|\mathbf{Y}\mathbf{f}^* - \mathbf{s}_{n_2:n_1}\|_2^2$$

where $\mathbf{Y}$ is the $N_{\text{tr}}\times(L+1)$ Hankel matrix defined as

$$\mathbf{Y} \triangleq \begin{bmatrix} y_{n_2+d} & \cdots & y_{n_2+d-L} \\ \vdots & \iddots & \vdots \\ y_{n_1+d} & \cdots & y_{n_1+d-L} \end{bmatrix}.$$

Assuming that matrix $\mathbf{Y}^T\mathbf{Y}^*$ is invertible, we obtain

$$\mathbf{f}_{\text{LS}} = \left(\mathbf{Y}^T\mathbf{Y}^*\right)^{-1}\mathbf{Y}^T\mathbf{s}_{n_2:n_1}^*. \tag{32}$$

Using the channel input-output relation (4), we can express matrix $\mathbf{Y}^T$ as

$$\mathbf{Y}^T = \mathbf{H}\mathbf{S}_2 + \mathbf{W}^T$$

where $\mathbf{S}_2$ is the $(L+M+1)\times N_{\text{tr}}$ Hankel matrix defined as

$$\mathbf{S}_2 \triangleq \begin{bmatrix} s_{n_2+d} & s_{n_2+d-1} & \cdots & s_{n_1+d} \\ \vdots & \vdots & \iddots & \vdots \\ s_{n_2+d-L-M} & s_{n_2+d-1-L-M} & \cdots & s_{n_1+d-L-M} \end{bmatrix}$$

and $\mathbf{W}$ contains noise samples and is of the same form as $\mathbf{Y}$. Thus, we obtain

$$\mathbf{Y}^T\mathbf{Y}^* = \mathbf{H}\mathbf{S}_2\mathbf{S}_2^H\mathbf{H}^H + \mathbf{W}^T\mathbf{W}^* + \mathbf{H}\mathbf{S}_2\mathbf{W}^* + \mathbf{W}^T\mathbf{S}_2^H\mathbf{H}^H \tag{33}$$

and

$$\mathbf{Y}^T\mathbf{s}_{n_2:n_1}^* = \mathbf{Y}^T\mathbf{S}_2^H\mathbf{e}_d = \mathbf{H}\mathbf{S}_2\mathbf{S}_2^H\mathbf{e}_d + \mathbf{W}^T\mathbf{S}_2^H\mathbf{e}_d. \tag{34}$$

[2]Using (29), we can derive the simple and informative bound

$$\mathbf{EMSE}(\hat{\mathbf{f}}) < \frac{M+1}{N_{\text{tr}}-M}\mathbf{MMSE}.$$

However, since MMSE reaches a floor for high enough SNR, this bound also exhibits a floor and, thus, is much less useful than the one appearing in (31), especially at high SNR.

In the following, we express random quantities[3] $\mathbf{S}_2\mathbf{S}_2^H$, $\mathbf{W}^T\mathbf{W}^*$ and $\mathbf{S}_2\mathbf{W}^*$ as perturbations of their mean values. At first, we note that

$$\mathbf{E}\left[\mathbf{S}_2\mathbf{S}_2^H\right] = N_{\text{tr}}\mathbf{I}_{L+M+1}$$
$$\mathbf{E}\left[\mathbf{W}^T\mathbf{W}^*\right] = N_{\text{tr}}\sigma_w^2\mathbf{I}_{L+1}$$

and

$$\mathbf{E}\left[\mathbf{S}_2\mathbf{W}^*\right] = \mathbf{0}_{(L+M+1)\times(L+1)}.$$

Then, we make the substitutions

$$\mathbf{S}_2\mathbf{S}_2^H = N_{\text{tr}}(\mathbf{I}_{L+M+1} + \mathbf{\Delta_s}) \quad (35)$$
$$\mathbf{W}^T\mathbf{W}^* = N_{\text{tr}}(\sigma_w^2\mathbf{I}_{L+1} + \mathbf{\Delta_w}) \quad (36)$$
$$\mathbf{S}_2\mathbf{W}^* = N_{\text{tr}}\mathbf{\Delta_{sw}} \quad (37)$$

where $N_{\text{tr}}\mathbf{\Delta_s}$, $N_{\text{tr}}\mathbf{\Delta_w}$, and $N_{\text{tr}}\mathbf{\Delta_{sw}}$ are the corresponding perturbations. Term $\mathbf{\Delta_{sw}}$ is of first-order with respect to the noise, while term $\Delta_{\mathbf{w}}$ is of second-order with respect to the noise and thus is negligible in comparison with $\mathbf{\Delta_{sw}}$ at sufficiently high SNR. Thus, in the sequel, we shall assume that $\mathbf{\Delta_w} = \mathbf{0}$.

By substituting the values of (35)–(37) in (33) and (34), we obtain

$$\mathbf{Y}^T\mathbf{Y}^* = N_{\text{tr}}\left(\mathbf{R_y} + \mathbf{\Delta}_2\right)$$

where

$$\mathbf{\Delta}_2 \triangleq \mathbf{H}\mathbf{\Delta_s}\mathbf{H}^H + \mathbf{H}\mathbf{\Delta_{sw}} + \mathbf{\Delta_{sw}^H}\mathbf{H}^H$$

and

$$\mathbf{Y}^T\mathbf{s}^*_{n_2:n_1} = N_{\text{tr}}\left(\mathbf{He}_d + \boldsymbol{\delta}_2\right)$$

where

$$\boldsymbol{\delta}_2 \triangleq \mathbf{H}\mathbf{\Delta_s}\mathbf{e}_d + \mathbf{\Delta_{sw}^H}\mathbf{e}_d.$$

If we substitute the above values in (32) and use approximation (3), we obtain

$$\mathbf{f}_{\text{LS}} = \left(\mathbf{R_y}^{-1} - \mathbf{R_y}^{-1}\mathbf{\Delta}_2\mathbf{R_y}^{-1}\right)\left(\mathbf{He}_d + \boldsymbol{\delta}_2\right). \quad (38)$$

By ignoring products of error terms in (38) and using (9), we obtain the first-order approximation

$$\mathbf{f}_{\text{LS}} = \mathbf{f}_o - \mathbf{R_y}^{-1}\left(\mathbf{\Delta}_2\mathbf{f}_o - \boldsymbol{\delta}_2\right). \quad (39)$$

We observe that the direct LS equalizer can be expressed as a perturbation of the MMSE equalizer, similarly to (22), leading to an expression for the EMSE analogous to (24). More specifically

$$\mathbf{EMSE}(\mathbf{f}_{\text{LS}}) = \text{tr}\left(\mathbf{R_y}^{-1}\mathbf{E}\left[(\Delta_2\mathbf{f}_o - \boldsymbol{\delta}_2)(\Delta_2\mathbf{f}_o - \boldsymbol{\delta}_2)^H\right]\right). \quad (40)$$

A. Computation of the EMSE of the LS Equalizer

In this subsection, we provide an analytic expression for the EMSE induced by using the LS equalizer instead of the ideal MMSE equalizer. At first, we express analytically term $\mathbf{\Delta}_2\mathbf{f}_o - \boldsymbol{\delta}_2$ as follows:

$$\mathbf{\Delta}_2\mathbf{f}_o - \boldsymbol{\delta}_2 = \underbrace{\mathbf{H}\mathbf{\Delta_s}\left(\mathbf{c}_o^* - \mathbf{e}_d\right)}_{\mathbf{t}_1} + \underbrace{\mathbf{H}\mathbf{\Delta_{sw}}\mathbf{f}_o}_{\mathbf{t}_2} + \underbrace{\mathbf{\Delta_{sw}^H}\left(\mathbf{c}_o^* - \mathbf{e}_d\right)}_{\mathbf{t}_3}. \quad (41)$$

Using the fact that the input and noise sequences are mutually independent, circular and zero-mean, it can be shown that the cross-covariance terms $\mathbf{E}\left[\mathbf{t}_i\mathbf{t}_j^H\right]$, for i, j =1, 2, 3 and $i \neq j$, vanish identically. From (40) and (41), we obtain

$$\mathbf{EMSE}\left(\mathbf{f}_{\text{LS}}\right) = \sum_{i=1}^{3}\mathbf{E}\left[\text{tr}\left(\mathbf{R_y}^{-1}\mathbf{t}_i\mathbf{t}_i^H\right)\right]. \quad (42)$$

In the sequel, we shall consider the three terms of the above sum.

1) Using (26) and (2), we obtain

$$\begin{aligned}\mathcal{T}_1 &\triangleq \mathbf{E}\left[\text{tr}\left(\mathbf{R_y}^{-1}\mathbf{t}_1\mathbf{t}_1^H\right)\right]\\ &= \mathbf{E}\left[\text{tr}\left(\mathbf{R_y}^{-1}\mathbf{H}\mathbf{\Delta_s}\mathbf{r}_o^*\mathbf{r}_o^T\mathbf{\Delta_s^H}\mathbf{H}^H\right)\right]\\ &= \mathbf{E}\left[\text{tr}\left(\underbrace{\mathbf{H}^H\mathbf{R_y}^{-1}\mathbf{H}}_{A}\underbrace{\mathbf{\Delta_s}}_{B}\underbrace{\mathbf{r}_o^*\mathbf{r}_o^T}_{C}\underbrace{\mathbf{\Delta_s^H}}_{D}\right)\right]\\ &= \mathbf{E}\left[\text{vec}^H(\mathbf{\Delta_s})\left(\mathbf{r}_o\mathbf{r}_o^H \otimes \mathbf{H}^H\mathbf{R_y}^{-1}\mathbf{H}\right)\text{vec}(\mathbf{\Delta_s})\right]\\ &= \text{tr}\left(\left(\mathbf{r}_o\mathbf{r}_o^H \otimes \mathbf{H}^H\mathbf{R_y}^{-1}\mathbf{H}\right)\mathbf{E}\left[\text{vec}(\mathbf{\Delta_s})\text{vec}^H(\mathbf{\Delta_s})\right]\right).\end{aligned} \quad (43)$$

The expression for the covariance matrix of $\text{vec}(\mathbf{\Delta_s})$, denoted $\mathbf{R}_{\text{vec}(\mathbf{\Delta_s})}$, is provided in the Appendix.

Remark 1: If $\mathbf{S}_2\mathbf{S}_2^H = N_{\text{tr}}\mathbf{I}_{L+M+1}$, then term $\mathbf{\Delta_s}$ and, thus, $\mathcal{T}_1$ identically *vanish*. This can be achieved if we construct sequence $\{s_{n_1+d-L-M}, \ldots, s_{n_2+d}\}$ by periodic extension, with period P, of the CAZAC sequences of [5, Table 15-3], with $P > L + M + 1$ and $N_{\text{tr}} = kP + L + M$, with k being a positive integer. However, in this case, the number of the samples that do not carry information but are used for "training purposes" is $N_{\text{tr}} + L + M$. This should be taken into account for a fair comparison between the mismatched MMSE and the LS equalizer.

Remark 2: If $\mathbf{S}_2$ is considered as deterministic, then we do not have to take expectation and $\mathcal{T}_1$ depends, through $\mathbf{\Delta_s}$, on the particular realization of $\mathbf{S}_2$.

Remark 3: If the number of training samples is small, then $\Delta_{\mathbf{s}}$ may be large [see (53)], and thus the first-order approximation in (39), with respect to $\mathbf{\Delta_s}$, is not very accurate. However, higher-order approximations seem very complicated and do not offer significant further insight into the behavior of the LS equalizer.

Remark 4: Term $\mathcal{T}_1$ depends on $\mathbf{r}_o$, whose 2-norm squared is approximately equal to the MMSE for sufficiently high SNR, and $\mathbf{\Delta_s}$, which depends on the number of training samples but is independent of the SNR. Thus, $\mathcal{T}_1$ will reach a *floor* and remain almost constant for high enough SNR. This fact makes the significant difference between the mismatched MMSE and the LS equalizer, in the cases where $\mathcal{T}_1$ does not identically vanish.

[3] In order to keep the analysis as general as possible, we consider matrices $\mathbf{S}_2$ as random. However, there exist cases of particular interest, for example, when $\mathbf{S}_2\mathbf{S}_2^H = N_{\text{tr}}\mathbf{I}_{L+M+1}$, where it is preferable to consider $\mathbf{S}_2$ as deterministic.

2) The second term of the right-hand side of (42) is expressed as

$$\mathcal{T}_2 \triangleq \mathbf{E}\left[\mathrm{tr}\left(\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{t}_2\mathbf{t}_2^H\right)\right] = \mathbf{E}\left[\mathrm{tr}\left(\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{H}\Delta_{\mathbf{sw}}\mathbf{f}_o\left(\mathbf{H}\Delta_{\mathbf{sw}}\mathbf{f}_o\right)^H\right)\right]. \quad (44)$$

We recall that $\Delta_{\mathbf{sw}} = (1/N_{\mathrm{tr}})\mathbf{S}_2\mathbf{W}^*$. It can be easily shown that

$$\mathbf{W}^*\mathbf{f}_o = \mathtt{F}_o\mathbf{w}^*_{n_2+d:n_1+d-L}$$

where $\mathtt{F}_o$ is the $N_{\mathrm{tr}} \times (N_{\mathrm{tr}} + L)$ filtering matrix constructed by $\mathbf{f}_o$. Furthermore

$$\mathrm{vec}(\mathbf{S}_2) = \mathcal{M}\mathbf{s}_{n_2+d:n_1+d-L-M}$$

where $\mathcal{M}$ is the $N_{\mathrm{tr}}(L+M+1) \times (N_{\mathrm{tr}} + L + M)$ matrix

$$\mathcal{M} \triangleq \begin{bmatrix} \mathbf{I}_{L+M+1} & & \\ \mathbf{0} & \mathbf{I}_{L+M+1} & \\ \vdots & \ddots & \\ \underbrace{\mathbf{0}\ \ldots\ \mathbf{0}}_{N_{\mathrm{tr}}-1} & & \mathbf{I}_{L+M+1} \end{bmatrix}$$

and $\mathbf{0}$ denotes the $(L+M+1) \times 1$ zero vector. In the sequel, we shall drop the long subscripts and denote the corresponding vectors as $\mathbf{s}$ and $\mathbf{w}$. Thus, $\mathcal{T}_2$ is given by the expression

$$\begin{aligned}\mathcal{T}_2 &= \frac{1}{N_{\mathrm{tr}}^2}\mathbf{E}_{\mathbf{s}}\left[\mathrm{tr}\left(\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{H}\mathbf{S}_2\mathtt{F}_o\mathbf{E}_{\mathbf{w}}[\mathbf{w}^*\mathbf{w}^T]\mathtt{F}_o^H\mathbf{S}_2^H\mathbf{H}^H\right)\right] \\ &= \frac{\sigma_w^2}{N_{\mathrm{tr}}^2}\mathbf{E}_{\mathbf{s}}\left[\mathrm{tr}\left(\underbrace{\mathbf{H}^H\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{H}}_{A}\underbrace{\mathbf{S}_2}_{B}\underbrace{\mathtt{F}_o\mathtt{F}_o^H}_{C}\underbrace{\mathbf{S}_2^H}_{D}\right)\right] \\ &= \frac{\sigma_w^2}{N_{\mathrm{tr}}^2}\mathbf{E}_{\mathbf{s}}\left[\mathrm{tr}\left((\mathtt{F}_o^*\mathtt{F}_o^T \otimes \mathbf{H}^H\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{H})\mathrm{vec}(\mathbf{S}_2)\mathrm{vec}^H(\mathbf{S}_2)\right)\right] \\ &= \frac{\sigma_w^2}{N_{\mathrm{tr}}^2}\mathrm{tr}\left(\left(\mathtt{F}_o^*\mathtt{F}_o^T \otimes \mathbf{H}^H\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{H}\right)\mathcal{M}\mathcal{M}^H\right). \quad (45)\end{aligned}$$

Remark 5: If $\mathbf{S}_2$ is regarded as deterministic, then we do not have to take expectation with respect to the training samples. In this case, $\mathcal{T}_2$ depends on the specific realization of $\mathbf{S}_2$ and equals the expression (without the expectation) in the second or third line of (45).

3) The third term of the right-hand side of (42) is expressed as

$$\mathcal{T}_3 \triangleq \mathbf{E}\left[\mathrm{tr}\left(\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{t}_3\mathbf{t}_3^H\right)\right] = \mathbf{E}\left[\mathrm{tr}\left(\mathbf{R}_{\mathbf{y}}^{-1}\boldsymbol{\Delta}_{\mathbf{sw}}^H\mathbf{r}_o^*\left(\boldsymbol{\Delta}_{\mathbf{sw}}^H\mathbf{r}_o^*\right)^H\right)\right]. \quad (46)$$

It can be shown that

$$\mathbf{S}_2^H\mathbf{r}_o^* = \mathtt{R}_o^*\mathbf{s}^*$$

where $\mathtt{R}_o$ is the $N_{\mathrm{tr}} \times (N_{\mathrm{tr}} + L + M)$ filtering matrix constructed by $\mathbf{r}_o$ and

$$\mathrm{vec}(\mathbf{W}^T) = \mathcal{A}\mathbf{w}$$

where $\mathcal{A}$ is the $N_{\mathrm{tr}}(L+1) \times (N_{\mathrm{tr}} + L)$ matrix with the same form as $\mathcal{M}$ but with $\mathbf{I}_{L+1}$ replacing $\mathbf{I}_{L+M+1}$ and, of course, the analogous changes to the zero vectors. Thus, $\mathcal{T}_3$ is given by the expression

$$\begin{aligned}\mathcal{T}_3 &= \frac{1}{N_{\mathrm{tr}}^2}\mathbf{E}_{\mathbf{s},\mathbf{w}}\left[\mathrm{tr}\left(\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{W}^T\mathbf{S}_2^H\mathbf{r}_o^*\mathbf{r}_o^T\mathbf{S}_2\mathbf{W}^*\right)\right] \\ &= \frac{1}{N_{\mathrm{tr}}^2}\mathbf{E}_{\mathbf{w}}\left[\mathrm{tr}\left(\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{W}^T\mathtt{R}_o^*\mathbf{E}_{\mathbf{s}}[\mathbf{s}^*\mathbf{s}^T]\mathtt{R}_o^T\mathbf{W}^*\right)\right] \\ &= \frac{1}{N_{\mathrm{tr}}^2}\mathbf{E}_{\mathbf{w}}\left[\mathrm{tr}\left((\mathtt{R}_o\mathtt{R}_o^H \otimes \mathbf{R}_{\mathbf{y}}^{-1})\mathrm{vec}\left(\mathbf{W}^T\right)\mathrm{vec}^H\left(\mathbf{W}^T\right)\right)\right] \\ &= \frac{\sigma_w^2}{N_{\mathrm{tr}}^2}\mathrm{tr}\left((\mathtt{R}_o\mathtt{R}_o^H \otimes \mathbf{R}_{\mathbf{y}}^{-1})\mathcal{A}\mathcal{A}^H\right). \quad (47)\end{aligned}$$

Remark 6: If $\mathbf{S}_2$ is regarded as deterministic, then $\mathcal{T}_3$ depends on the specific realization of $\mathbf{S}_2$. Using analogous techniques, it can be shown that in this case

$$\mathcal{T}_3 = \frac{\sigma_w^2}{N_{\mathrm{tr}}^2}\mathrm{tr}\left(\mathbf{R}_{\mathbf{y}}^{-1}\left(\mathbf{r}_o^H\mathbf{S}_2^* \otimes \mathbf{I}_{L+1}\right)\mathcal{A}\mathcal{A}^H\left(\mathbf{S}_2^T\mathbf{r}_o \otimes \mathbf{I}_{L+1}\right)\right). \quad (48)$$

B. Simplifications in the High-SNR Cases

In the high-SNR cases, for the mid-range delays and sufficiently large equalizer lengths, $\mathcal{T}_3$ will be negligible compared with $\mathcal{T}_2$, since it is constructed using $\mathbf{r}_o$. An approximation to $\mathcal{T}_2$ is derived as follows:

$$\begin{aligned}\mathcal{T}_2 &\approx \frac{\sigma_w^2}{N_{\mathrm{tr}}^2}\mathrm{tr}\left(\mathcal{M}^H\left(\mathtt{F}_o^*\mathtt{F}_o^T \otimes \mathbf{P}_{\mathbf{H}^H}\right)\mathcal{M}\right) \\ &\overset{(d)}{\leq} \frac{\sigma_w^2}{N_{\mathrm{tr}}^2}\mathrm{tr}\left(\mathcal{M}^H\left(\mathtt{F}_o^*\mathtt{F}_o^T \otimes \mathbf{I}_{L+M+1}\right)\mathcal{M}\right) \\ &\overset{(e)}{=} \frac{\sigma_w^2}{N_{\mathrm{tr}}^2}\mathrm{tr}\left(\underbrace{\mathcal{M}^H\left(\mathtt{F}_o^* \otimes \mathbf{I}_{L+M+1}\right)}_{\mathcal{B}}\left(\mathtt{F}_o^T \otimes \mathbf{I}_{L+M+1}\right)\mathcal{M}\right) \\ &= \frac{\sigma_w^2}{N_{\mathrm{tr}}^2}\|\mathcal{B}\|_F^2 \quad (49)\end{aligned}$$

where at point (d) we used the fact that $\mathbf{P}_{\mathbf{H}^H} \leq \mathbf{I}_{L+M+1}$ and at point (e) we used (1).

Due to the structure of $\mathcal{M}$, it holds that

$$\|\mathcal{B}\|_F^2 = \|\mathtt{F}_o^* \otimes \mathbf{I}_{L+M+1}\|_F^2 = (L+M+1)N_{\mathrm{tr}}\|\mathbf{f}_o\|_2^2.$$

Thus we derive the bound

$$\boxed{\mathcal{T}_2 \lesssim \frac{(L+M+1)\|\mathbf{f}_o\|_2^2}{N_{\mathrm{tr}}}\sigma_w^2}. \quad (50)$$

In extensive simulation studies, we have observed that this bound is a very good approximation to $\mathcal{T}_2$ and, thus, to the EMSE of the LS equalizer, when $\boldsymbol{\Delta}_{\mathbf{s}}$ and, thus, $\mathcal{T}_1$, is zero. $\mathcal{T}_2$ is proportional to the noise variance (or inverse SNR), σ_w^2, just like the EMSE of the mismatched MMSE equalizer, with the only difference being in the associated constants.

V. Comparison of the Mismatched MMSE and the LS Linear Equalizers

In this section, we use the derived results and compare the mismatched MMSE and the LS equalizers. At first, we make the comparison for two extreme but important cases.

1) *Case 1:* $\mathbf{S}_1^H\mathbf{S}_1 = (N_{\rm tr} - M)\mathbf{I}_{M+1}$. This is the ideal case for the mismatched MMSE equalizer, yielding $\mathbf{R}_{\Delta\mathbf{h}} = \left(\sigma_w^2/(N_{\rm tr} - M)\right)\mathbf{I}_{M+1}$. For sufficiently high SNR, we obtain

$$\frac{\mathbf{EMSE}(\mathbf{f}_{\rm LS})}{\mathbf{EMSE}(\hat{\mathbf{f}})} \approx \frac{\mathcal{T}_1 + \mathcal{T}_2}{\mathbf{EMSE}(\hat{\mathbf{f}})}. \quad (51)$$

Using bounds (31) and (50) as approximate equalities, we obtain

$$\frac{\mathcal{T}_2}{\mathbf{EMSE}(\hat{\mathbf{f}})} \approx \frac{(L+M+1)(N_{\rm tr} - M)}{(M+1)N_{\rm tr}} \quad (52)$$

which is larger than 1 for

$$N_{\rm tr} > \frac{M(L+M+1)}{L}$$

and converges to $(L+M+1)/(M+1)$ for $N_{\rm tr}$ tending to infinity. However, ratio (51) is much larger, due to the contribution of $\mathcal{T}_1$, which exhibits a floor depending on the number of the training samples and the equalizer length. Consequently, *the ratio of the EMSEs in* (51) *tends to infinity as the SNR tends to infinity.* In this case, the mismatched MMSE equalizer (significantly) outperforms the LS equalizer.

2) *Case 2:* $\mathbf{S}_2\mathbf{S}_2^H = N_{\rm tr}\mathbf{I}_{L+M+1}$. This is the ideal case for the LS equalizer, because $\mathbf{\Delta_s}$ and, thus, $\mathcal{T}_1$, identically vanish. At high enough SNR, we have that $\mathbf{EMSE}(\mathbf{f}_{\rm LS}) \approx \mathcal{T}_2$. For the values of M and L we use in the simulations ($M = 4$ and $L \leq 10$), we have uncovered one CAZAC sequence leading to zero $\mathbf{\Delta_s}$. It is the eighth sequence appearing in [5, Table 15.3]. The EMSE for the mismatched MMSE is given by (30), with $\mathbf{R}_{\Delta\mathbf{h}} = \sigma_w^2\left(\mathbf{S}_1\mathbf{S}_1^H\right)^{-1}$, with $\mathbf{S}_1$ being the Hankel matrix constructed by the training samples $\{s_{n_1+d-L-M}, \ldots, s_{n_2+d}\}$. In the simulations section, we shall see that the mismatched MMSE equalizer outperforms the LS equalizer, in this case, as well.

Finally, if the training samples are random, then the mismatched MMSE (significantly) outperforms the LS equalizer because the EMSE of the mismatched MMSE is proportional to the noise variance (recall that in this case $\mathbf{R}_{\Delta\mathbf{h}} = \sigma_w^2\mathbf{E}\left[(\mathbf{S}_1^H\mathbf{S}_1)^{-1}\right]$), while the EMSE of the LS equalizer (and, more specifically, term $\mathcal{T}_1$) reaches a floor and remains constant for high enough SNR.

VI. Simulations

In this section, we check our theoretical results with simulations. We present the results for a randomly generated channel having impulse response $\mathbf{h}$ given in Table I, with $M = 4$, equalizer order $L = 10$ and delay $d = 6$. Analogous results have been observed in extensive simulations with other parameter values.

TABLE I
Channel Impulse Response

$h(0)$	$-0.6981 + 0.2199i$
$h(1)$	$+0.123 + 0.4803i$
$h(2)$	$-0.2785 + 0.0841i$
$h(3)$	$-0.218 - 0.2012i$
$h(4)$	$-0.0581 - 0.206i$

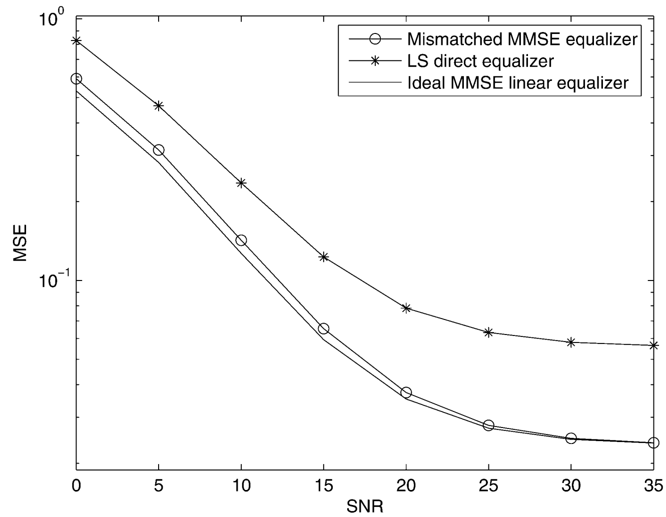

Fig. 1. MSE achieved by the ideal MMSE, the mismatched MMSE and the LS equalizers.

1) Simulation 1: Ideal input for the mismatched MMSE equalizer.

The training sequence is constructed by periodic extension of the seventh sequence of [5, Table 15-3] and contains $N_{\rm tr} = 28$ training samples.

In Fig. 1, we present the MSE achieved by the MMSE, the mismatched MMSE and the LS equalizers. At first, we observe that all equalizers exhibit an MSE floor. This was to be expected because a finite-length linear equalizer cannot equalize perfectly an FIR channel, even in the noiseless case. Furthermore, we observe that the mismatched MMSE significantly outperforms the LS equalizer in terms of MSE. More specifically:

1) the MSE of the mismatched MMSE equalizer tends to the MMSE for increasing SNR, because, as we see in (31), the EMSE of the mismatched MMSE equalizer is proportional to the noise variance, which goes to zero as the SNR goes to infinity;
2) the MSE of the LS equalizer does *not* tend to the MMSE for increasing SNR, because the EMSE of the LS equalizer does *not* tend to zero for increasing SNR, due to the fact that term $\mathcal{T}_1$ is almost constant for high enough SNR. The value of $\mathcal{T}_1$ depends on the size of $\mathbf{r}_o$, which reaches a floor for sufficiently high SNR, and $\mathbf{\Delta_s}$, which depends on the number of training samples.

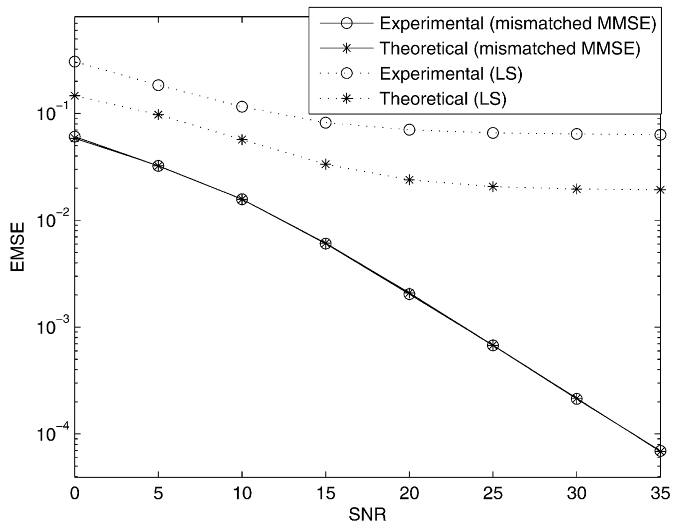

Fig. 2. Theoretical (second-order approximation) and experimentally computed EMSE for the mismatched and the LS equalizers.

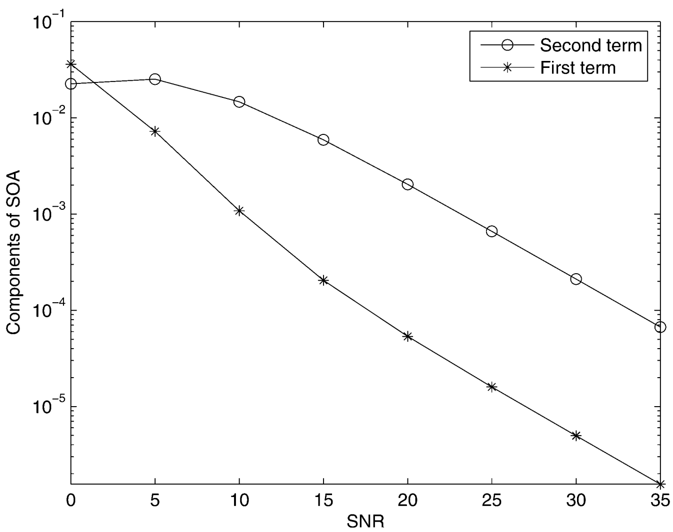

Fig. 3. First and second components of the second-order approximation EMSE (28) for the mismatched MMSE equalizer.

In Fig. 2, we present the theoretical second-order approximations (SOAs) (28) and (42) and the associated experimentally computed EMSEs for the mismatched MMSE and the LS equalizer. We observe that the theoretical EMSE for the mismatched MMSE equalizer practically coincides with the experimental EMSE. On the other hand, the theoretical expression is less accurate for the LS equalizer. This happens because, for small number of training samples, the first-order approximation in (39), with respect to $\mathbf{\Delta_s}$, is not very accurate. We have observed that this discrepancy disappears for sufficiently large number of training samples (e.g., $N_{\rm tr} \approx 100$).

In Fig. 3, we plot the first and second components of the theoretical EMSE of (28). We observe that, for SNR higher than 5 dB, the contribution of the term that involves matrix $\mathbf{R}_o$ is much smaller (about two orders of magnitude) than the contribution of the term involving $\mathbf{G}_o$, confirming our claim appearing before (30).

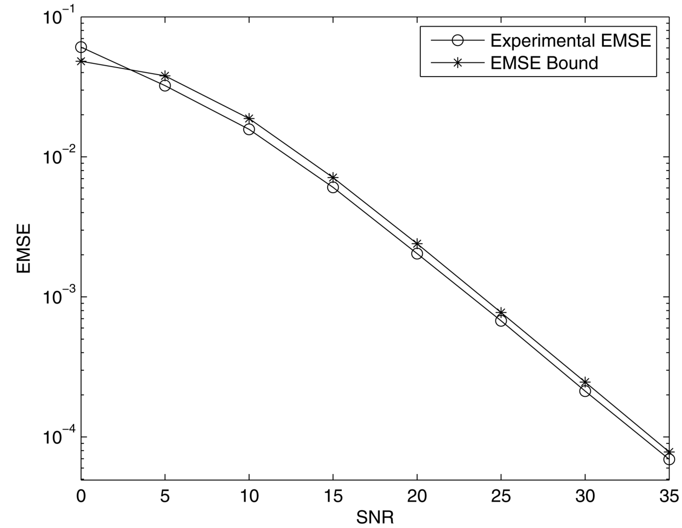

Fig. 4. Experimentally computed EMSE and bound (31) for the mismatched MMSE equalizer.

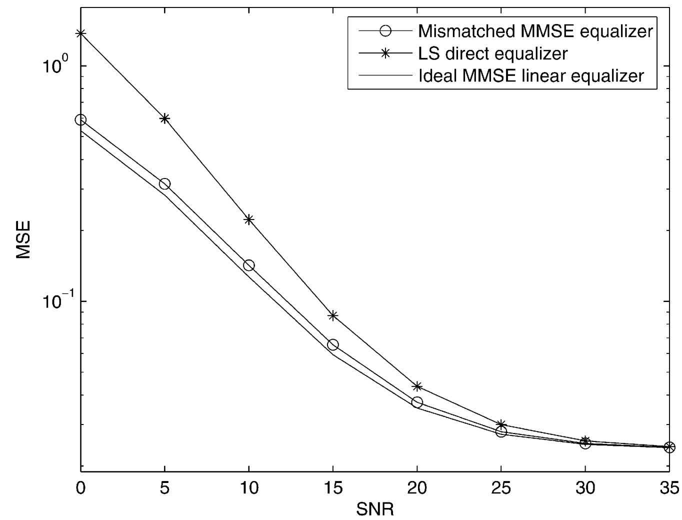

Fig. 5. MSE achieved by the ideal MMSE, the mismatched MMSE and the LS equalizers.

In Fig. 4, we plot the experimentally computed EMSE for the mismatched MMSE equalizer and bound (31), confirming that bound (31) is a very good approximation to the EMSE of the mismatched MMSE equalizer for SNR higher than 5 dB.

2) Simulation 2: Ideal input for the LS equalizer.

In order to have $\mathbf{\Delta_s}$ identically zero, we use training sequence $\{s_{n_1+d-L-M}, \ldots, s_{n_2+d}\}$, with 30 samples, constructed by the periodic extension of the eighth sequence of [5, Table 15-3].

In Fig. 5, we plot the MSE achieved by the MMSE, the mismatched MMSE and the LS equalizers. We observe that the mismatched MMSE equalizer outperforms the LS equalizer in this case as well. For high enough SNR, all quantities practically coincide. This happens because, in this case, the EMSEs for both the mismatched and the LS equalizers are proportional to the noise variance (recall that $\mathcal{T}_1$ vanishes identically) and go to zero as the SNR goes to infinity.

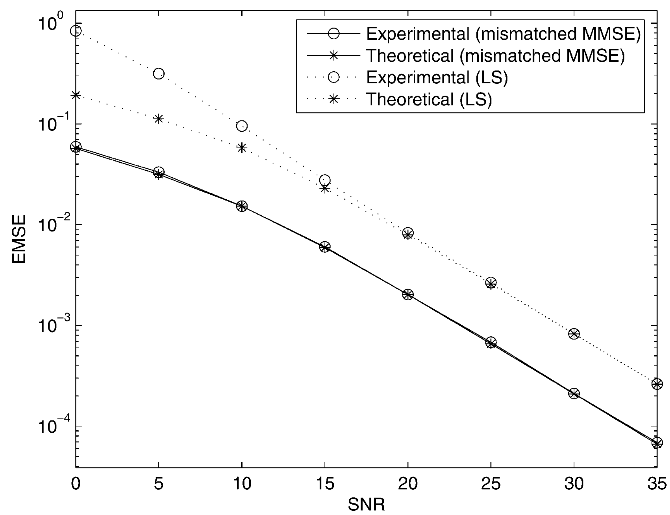

Fig. 6. Theoretical (second-order approximation) and experimentally computed EMSE for the mismatched and the LS equalizers.

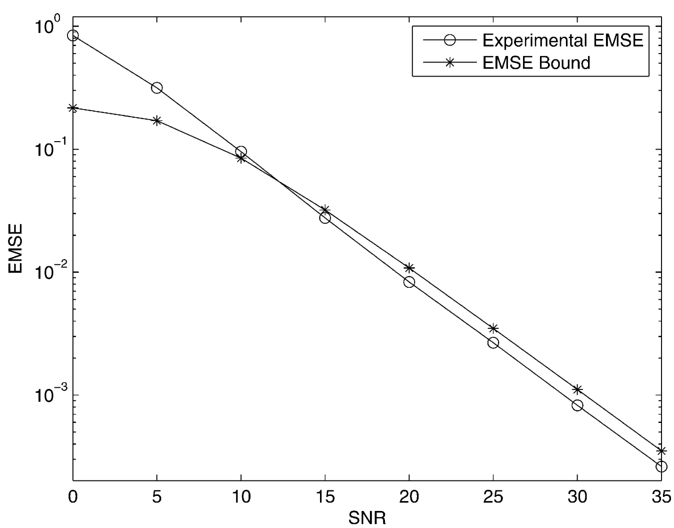

Fig. 7. Experimentally computed EMSE and bound (50) for the LS equalizer.

In Fig. 6, we check the accuracy of the theoretical expressions for the EMSE. We observe that our expressions are very accurate over a very wide range of SNRs. Furthermore, we observe that the EMSE of the mismatched MMSE is smaller than that of the LS equalizer and that their ratio tends to a limit for increasing the SNR (this becomes obvious from the fact that the curves representing the EMSEs become parallel for high enough SNR).

Finally, in Fig. 7, we observe that bound (50) is a very good approximation of the EMSE of the LS equalizer for SNR higher than 10 dB.

The case with random training samples resembles the one presented in Simulation 1 and thus we do not present graphical simulation results.

VII. Conclusion

We considered two widely referenced trained finite-length linear equalizers, namely, the mismatched MMSE and the LS equalizer. Using matrix perturbation theory, we expressed both of them as perturbations of the ideal MMSE equalizer and we derived insightful analytical expressions for their excess MSE. We observed that, in general, the mismatched MMSE equalizer performs (much) better than the LS equalizer. This happens because the LS equalizer implicitly estimates the input second-order statistics, while the mismatched MMSE equalizer uses perfect knowledge. Thus, when the input second-order statistics are known at the receiver, which is usually the case, the use of the mismatched MMSE equalizer is preferable, in general. In this work, we assumed that the order of the channel impulse response, M, is known. An interesting topic might be the comparison of the mismatched MMSE and the LS equalizers in the cases where the channel order is underestimated.

Appendix

A. Expression for $\mathbf{R}_{\text{vec}(\mathbf{\Delta_s})}$*:* We remind that $\mathbf{\Delta_s} = (1/N_{\text{tr}})\mathbf{S}_2\mathbf{S}_2^H - \mathbf{I}_{L+M+1}$. We denote the (i,j)-th element of $\mathbf{\Delta_s}$ as $\mathbf{\Delta_s}(i,j)$. If $|s_n| = 1$, then $\mathbf{\Delta_s}(i,i) = 0$. For $j \neq i$, we have

$$\mathbf{\Delta_s}(i,j) = \frac{1}{N_{\text{tr}}} \sum_{l=n_1+d+1}^{n_2+d+1} s_{l-i}s_{l-j}^* = \mathbf{\Delta_s}^*(j,i)$$

where the last equality is due to the Hermitian structure of $\mathbf{\Delta_s}$.

The autocorrelation of $\mathbf{\Delta_s}(i,j)$ and $\mathbf{\Delta_s}(x,y)$ is given by

$$\mathbf{E}\left[\mathbf{\Delta_s}(i,j)\mathbf{\Delta_s}^*(x,y)\right] = \begin{cases} \frac{N_{\text{tr}}-|i-x|}{N_{\text{tr}}^2}, & \text{if } j-i=y-x \neq 0 \\ 0, & \text{otherwise.} \end{cases}$$

The element $\mathbf{\Delta_s}(i,j)$ lies at the $((j-1)(L+M+1)+i)$-th position of $\text{vec}(\mathbf{\Delta_s})$. Introducing indices $k = (j-1)(L+M+1)+i$ and $l = (y-1)(L+M+1)+x$, for $i,j,x,y = 1,\ldots,L+M+1$, we obtain

$$\mathbf{R}_{\text{vec}(\mathbf{\Delta_s})}(k,l) = \begin{cases} \frac{N_{\text{tr}}-|x-i|}{N_{\text{tr}}^2}, & \text{if } j-i=y-x \neq 0 \\ 0, & \text{otherwise.} \end{cases} \quad (53)$$

B. Channel Estimation Error Covariance Matrix in the Case of Random Training Samples: In the sequel, we provide guidelines for the computation of the channel estimation error covariance matrix, $\mathbf{R}_{\Delta\mathbf{h}}$, when the training samples are independent zero-mean unit-variance random variables. If $\mathbf{S}_1^H\mathbf{S}_1 \neq (N_{\text{tr}} - M)\mathbf{I}_{M+1}$, then we can express $\mathbf{S}_1^H\mathbf{S}_1$ as the perturbation

$$\mathbf{S}_1^H\mathbf{S}_1 = (N_{\text{tr}} - M)\mathbf{I}_{M+1} + \Delta_{\mathbf{s}_1}$$

with $\mathbf{E}[\Delta_{\mathbf{s}_1}] = \mathbf{0}$. If we define $\alpha \triangleq N_{\text{tr}} - M$, then $\mathbf{E}\left[\left(\mathbf{S}_1^H\mathbf{S}_1\right)^{-1}\right]$ can be approximated to second order, with respect to $\mathbf{\Delta}_{\mathbf{s}_1}$, as follows:

$$\begin{aligned}\mathbf{E}\left[\left(\mathbf{S}_1^H\mathbf{S}_1\right)^{-1}\right] &= \mathbf{E}\left[\alpha^{-1}\mathbf{I}_{M+1} - \alpha^{-2}\Delta_{\mathbf{s}_1} + \alpha^{-3}\Delta_{\mathbf{s}_1}^2\right] \\ &= \alpha^{-1}\mathbf{I}_{M+1} + \alpha^{-3}\mathbf{E}\left[\Delta_{\mathbf{s}_1}^2\right].\end{aligned}$$

Then, using techniques analogous to those of Section A of this Appendix, we can compute the above expectations.

REFERENCES

[1] A. Sayed, *Fundamentals of Adaptive Filtering*. New York: Wiley, 2003.
[2] C. R. Johnson, Jr and W. A. Sethares, *Telecommunication Breakdown*. Englewood Cliffs, NJ: Prentice-Hall, 2004.
[3] H. Luetkepohl, *Handbook of Matrices*. New York: Wiley, 1996.
[4] G. Stewart and J.-G. Sun, *Matrix Perturbation Theory*. New York: Academic, 1990.
[5] H. Meyr, M. Moeneclaey, and S. A. Fechtel, *Digital Communication Receivers*. New York: Wiley, 1998.
[6] A. P. Liavas, "On the robustness of the finite-length MMSE-DFE with respect to channel and second-order statistics estimation errors," *IEEE Trans. Signal Process.*, vol. 50, no. 11, pp. 2866–2874, Nov. 2002.

Athanasios P. Liavas (M'89) was born in Pyrgos, Greece, in 1966. He received the Diploma and Ph.D. degrees in computer engineering from the University of Patras, Patras, Greece, in 1989 and 1993, respectively.

He is currently an Associate Professor at the Department of Electronic and Computer Engineering, Technical University of Crete, Chania, Greece. His recent research interests lie in the areas of signal processing for communications and information theory.

Dr. Liavas is a member of the Technical Chamber of Greece and is currently serving as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING.

Despoina Tsipouridou was born in Drama, Greece, in 1980. She received the Diploma degree in electrical and computer engineering from the Aristotle University of Thessaloniki, Thessaloniki, Greece, and the M.Sc. degree in electronic and computer engineering from the Technical University of Crete, Greece, in 2004 and 2006, respectively. She is currently working toward the Ph.D. degree in the Telecommunications Division, Department of Electronic and Computer Engineering, Technical University of Crete, Chania, Greece.

Her research interests are in the area of signal processing for communications.