Katholieke Universiteit Leuven

Departement Elektrotechniek

ESAT-SISTA/TR 2004-02

On the relation between time-domain equalizers and per-tone equalizers for DMT-based systems [1]

Koen Vanbleu, Geert Ysebaert, Gert Cuypers and Marc Moonen [2]

Accepted for publication in IEEE Transactions on Signal Processing
September 2004

[1] This report is available by anonymous ftp from *ftp.esat.kuleuven.ac.be* in the directory *pub/sista/vanbleu/reports/04-02.pdf*

[2] K.U.Leuven, Dept. of Electrical Engineering (ESAT), Research group SISTA, Kasteelpark Arenberg 10, 3001 Leuven-Heverlee, Belgium, Tel. 32/16/32 18 41, Fax 32/16/32 19 70, WWW: *http://www.esat.kuleuven.ac.be/sista*. E-mail: *koen.vanbleu@esat.kuleuven.ac.be*. This research work was carried out at the ESAT laboratory of the K.U. Leuven, in the frame of IUAP P5/22 and P5/11, GOA-MEFISTO-666, Research Project FWO nr.G.0196.02 and was partially sponsored by Alcatel-Bell. The scientific responsibility is assumed by its authors.

On Time-Domain And Frequency-Domain MMSE-Based TEQ Design For DMT Transmission

Koen Vanbleu*, *Student Member, IEEE,* Geert Ysebaert, *Student Member, IEEE,* Gert Cuypers, *Member, IEEE,* Marc Moonen, *Member, IEEE*

***Abstract*— We reconsider the MMSE time-domain equalizer (TEQ), bitrate maximizing TEQ (BM-TEQ) and per-tone equalizer design (PTEQ) for DMT transmission and cast them in a common least-squares (LS) based framework. The MMSE-TEQ design criterion can be formulated as a constrained linear least-squares (CLLS) criterion that minimizes a time-domain (TD) error energy. From this CLLS-based TD-MMSE-TEQ criterion, we derive two new least-squares (LS) based frequency-domain (FD) MMSE-TEQ design criteria: a CLLS-based FD-MMSE-TEQ criterion and a so-called separable nonlinear LS (SNLLS) based FD-MMSE-TEQ design. Finally, the original BM-TEQ design is shown to be equivalent to a so-called iteratively-reweighted (IR) version of the SNLLS-based FD-MMSE-TEQ design. This LS-based framework then results in the following contributions. The new, IR-SNLLS-based BM-TEQ design criterion gives rise to an elegant, iterative, fast converging, Gauss-Newton-based design algorithm that exploits the separability property. The intermediate FD-MMSE-TEQ designs establish a link between the TD-MMSE-TEQ on one hand and the BM-TEQ and the PTEQ on the other hand. Moreover, the considered LS-based equalizer designs — although at first sight very different in nature — exhibit remarkable correspondence when turned into equivalent generalized eigenvalue problems.**

***Index Terms*— Discrete multitone, digital subscriber lines, time-domain equalization, channel shortening, separable nonlinear least-squares, constrained linear least-squares, iteratively reweighted least-squares**

I. Introduction

Discrete multitone (DMT) modulation and orthogonal frequency division multiplexing (OFDM) are all-digital multicarrier modulation schemes. DMT modulation is adopted as the transmission format for asymmetric digital subscriber lines (ADSL) [1] and very high bit rate digital subscriber lines (VDSL) [2]; OFDM is adopted for wireless local area applications, e.g., IEEE 802.11 (Wi-Fi) [3] and HiperLAN/2 [4].

DMT/OFDM schemes divide the available bandwidth into parallel subchannels or tones. The incoming bitstream is split into parallel streams of symbols that are used to QAM-modulate the different tones. The modulation is done by means of an inverse discrete Fourier transform (IDFT). Before transmission of a DMT/OFDM symbol, a cyclic prefix of ν samples is added. If the channel impulse response length is smaller than or equal to $\nu + 1$, demodulation can be implemented by means of a DFT, followed by a (complex-valued) 1-tap frequency domain equalizer (FEQ) per tone to compensate for channel amplitude and phase effects.

Practical ADSL channel impulse responses can be very long, hence a large prefix would be required. However, a long prefix introduces a large overhead, resulting in a reduced bitrate. An existing solution is to insert a (real-valued) T-tap time domain equalizer (TEQ) before demodulation that shortens the channel impulse response to $\nu+1$ samples. Many TEQ design algorithms have been developed. We will focus on so-called *minimum mean-square error (MMSE) TEQ designs* [5], [6] with a unit energy constraint (UEC) [7], [8], [9] and the recently proposed *bitrate maximizing TEQ design* (BM-TEQ) [10]. Whereas the MMSE-TEQ minimizes a time-domain error energy (i.e., the error between the TEQ output and the output of an FIR filter with a so-called target impulse response (TIR) of length $\nu + 1$), the BM-TEQ design takes the multicarrier aspect of DMT modulation into account and maximizes the aggregated number of bits transmitted over all subchannels. Other TEQ design methods of interest can be found in [11], which was the first paper to propose and formulate bitrate maximization as a criterion for TEQ design, in [12], which improved upon [11] and suggested the low-complexity min-ISI TEQ design, and in [13], [14], which present blind adaptive TEQ design methods based on the maximum shortening-SNR TEQ of [15]. In [16], an attractive alternative equalization scheme is proposed that always performs at least as well as — and usually better than — a TEQ based receiver while keeping complexity during data transmission at the same level: a complex-valued MMSE equalizer is then used for each tone separately, hence the term "*per-tone equalization*" (PTEQ).

It is noted that, at first sight, the considered equalizer designs (MMSE-TEQ, BM-TEQ and PTEQ) are very different in nature: the MMSE-TEQ is the solution of a constrained, linear MMSE criterion in the TEQ and (auxiliary) TIR coefficients; the PTEQ design is rather based on an unconstrained, linear MMSE criterion that only depends on the PTEQ coefficients; the BM-TEQ criterion is a nonlinear function of the TEQ coefficients and asks for a nonlinear optimization routine.

In this paper, we reconsider these equalizer designs and cast them in a *common least-squares (LS) based framework*.

This research work was carried out at the ESAT laboratory of the Katholieke Universiteit Leuven, in the frame of the Belgian Programme on Interuniversity Attraction Poles, initiated by the Belgian Federal Science Policy Office IUAP P5/22 ('Dynamical Systems and Control: Computation, Identification and Modelling') and P5/11 ('Mobile multimedia communication systems and networks'), the Concerted Research Action GOA-MEFISTO-666 (Mathematical Engineering for Information and Communication Systems Technology) of the Flemish Government, Research Project FWO nr.G.0196.02 ('Design of efficient communication techniques for wireless time-dispersive multi-user MIMO systems') and IWT Project 030054 SOLIDT ('Solutions for xDSL interoperability, deployment and new technologies'), and was partially sponsored by Alcatel-Bell. The scientific responsibility is assumed by its authors.

The classical MMSE-TEQ design criterion can be formulated as a constrained linear least-squares (CLLS) based MMSE criterion that minimizes a TD error energy. Starting from this CLLS-based TD-MMSE-TEQ design, we derive two new intermediate MMSE-TEQ designs that minimize a sum of frequency-domain (FD) or subchannel error energies (i.e., after DFT demodulation), rather than a TD error energy: the first FD-MMSE-TEQ design is based on a CLLS criterion, while the other is based on a — related — separable nonlinear LS (SNLLS) criterion. Finally, the BM-TEQ design criterion, originally presented in [10], is shown to be equivalent to a so-called iteratively-reweighted (IR) version of the SNLLS-based FD-MMSE-TEQ design criterion, referred to as an *IR-SNLLS-based BM-TEQ design criterion*. This common LS-based framework then gives rise to the following contributions.

- As for solving the original BM-TEQ design criterion, it was suggested in [10] to resort to classical, oftentimes complex, non-linear optimization algorithms. Thanks to its LS-based formulation, the equivalent IR-SNLLS-based BM-TEQ design criterion can be optimized with an elegant, fast converging, *iterative Gauss-Newton (GN) based algorithm* that exploits the separability property of the design criterion, which results from the specific TEQ-FEQ-based receiver structure. It will be shown that the GN algorithm's complexity and convergence speed allow to design a BM-TEQ at a reasonable computational cost. Moreover, the GN algorithm ends up in local optima with a similar bitrate as provided by the original BM-TEQ design criterion of [10], rendering it a worthwhile approach.
- The two, new, intermediate FD-MMSE-TEQ designs allow to establish a *link* between the MMSE-TEQ on one hand and the BM-TEQ as well as the PTEQ on the other hand.
- The close relationship between the LS-based formulations of the TD-MMSE-TEQ, FD-MMSE-TEQ, BM-TEQ and PTEQ design criteria is made explicit by turning the LS-based formulations into *generalized eigenvalue (GEV) problems* $\mathbf{B}\mathbf{w} = \lambda\mathbf{A}\mathbf{w}$ where, loosely speaking, $\mathbf{A}$ is an autocorrelation metric of the receive signal y_l and $\mathbf{B}$ depends on a crosscorrelation metric between transmit (TX) and receive (RX) signal x_l and y_l.

The LS-based framework is presented in **Section II**. Each of the considered TEQ designs is treated in a subsection: first, the CLLS-based TD-MMSE-TEQ design criterion is revisited; modifications are then introduced to obtain the CLLS-based and the SNLLS-based FD-MMSE-TEQ design criteria; finally, the IR-SNLLS-based BM-TEQ design criterion is established. The iterative GN-based algorithm for optimizing the IR-SNLLS-based BM-TEQ design criterion is developed in **Section III**; its computational cost is also discussed. **Section IV** shows that the PTEQ design is a special case of the presented FD-MMSE-TEQ designs, when considering a single tone. In this section, the relation between all LS-based formulations and their GEV problem description is also summarized and tabulated. **Section V** compares the performance of the considered equalizer designs and illustrates the fast convergence of the iterative GN-based algorithm for optimizing the IR-SNLLS-based BM-TEQ design criterion.

Notation

Below, we introduce the notation that is used in the paper.

- $\mathcal{S}_a$ is the set of N_a active tones, e.g., tones 38 to 256 for downstream ADSL; n is a tone index; N is the (I)DFT size; $\mathcal{F}_N$ is a unitary DFT matrix of size N; the (m,n)-th entry is given by $\frac{1}{\sqrt{N}}e^{-j2\pi mn/N}$, $m = 0,\ldots,N-1$, $n = 0,\ldots,N-1$; $\mathcal{F}_{\mathcal{S}_a}$ is an $N_a \times N$ submatrix of $\mathcal{F}_N$ with only the N_a rows of the active tones $\mathcal{S}_a$; the n-th DFT row is $\mathcal{F}_n$.
- A tilde over a variable distinguishes frequency-domain symbols from time-domain symbols, e.g. the $N_a \times 1$ TX symbol vector at time k, $\tilde{\mathbf{x}}_k$; vectors are typeset in bold lowercase while matrices are in bold uppercase; the subscript N is used to denote frequency-domain vectors or matrices that take the N DFT bins (rather than only the active tones in $\mathcal{S}_a$) into account (e.g., the $N \times 1$ TX symbol vector, $\tilde{\mathbf{x}}_{k,N}$, see below).
- $\mathbf{w}$ is the time-domain equalizer (TEQ, T taps) and $\mathbf{w}_n$ is the per-tone equalizer (PTEQ) for tone n.
- $\tilde{\mathbf{d}}$ is the $N_a \times 1$ vector of FEQ coefficients; the corresponding $N \times 1$ vector (padded with do-not-care entries for non-active tones) is $\tilde{\mathbf{d}}_N$; $\tilde{d}_n$ is the FEQ for tone n. Sometimes, we use a joint TEQ-FEQ parameter vector, either with all FEQs $\tilde{\mathbf{d}}$ or with only the FEQ $\tilde{d}_n$, denoted as $\boldsymbol{\theta} = \begin{bmatrix} \mathbf{w}^H & \tilde{\mathbf{d}}^H \end{bmatrix}^H$ and $\boldsymbol{\theta}_n = \begin{bmatrix} \mathbf{w}^H & \tilde{d}_n^* \end{bmatrix}^H$, respectively.
- The DMT symbol index is k. The k-th (frequency-domain) $N_a \times 1$ DMT symbol vector that is fed to the modulating IDFT is $\tilde{\mathbf{x}}_k$; the corresponding $N \times 1$ symbol vector (padded with zeros for non-active tones) is $\tilde{\mathbf{x}}_{k,N}$; the symbol on tone n is $\tilde{x}_{k,n}$. The cyclic prefix length is ν.
- $\mathcal{E}\{\cdot\}$ is the expectation operator; $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^*$ denote the transpose, Hermitian and complex conjugate operator, respectively; $\mathrm{diag}(\mathbf{a})$ is a diagonal matrix with the entries of $\mathbf{a}$ on the diagonal. $\mathbf{a} \odot \mathbf{b}$ is the elementwise multiplication of 2 vectors of equal length. $\mathbf{I}_M$ is the $M \times M$ identity matrix.
- A subscript with the number of data points, e.g., L samples or K DMT symbols, is used to distinguish between a (deterministic) correlation *estimate*, e.g., $\boldsymbol{\Sigma}^2_{L,\mathbf{y}} = \frac{1}{L}\sum_{l=1}^{L}\mathbf{y}_l\mathbf{y}_l^T$ or $\boldsymbol{\sigma}_{K,n,\tilde{x}\tilde{\mathbf{y}}} = \frac{1}{K}\sum_{k=1}^{K}\tilde{x}_{k,n}^*\tilde{\mathbf{y}}_{k,n}$, and the (stochastic) *true* correlation $\boldsymbol{\Sigma}^2_{\mathbf{y}} = \mathcal{E}\left\{\mathbf{y}_l\mathbf{y}_l^T\right\}$ or $\boldsymbol{\sigma}_{n,\tilde{x}\tilde{\mathbf{y}}} = \mathcal{E}\left\{\tilde{x}_{k,n}^*\tilde{\mathbf{y}}_{k,n}\right\}$. Under the assumption that the involved signals are wide-sense stationary (WSS), it holds that $\lim_{L\to\infty}\boldsymbol{\Sigma}^2_{L,\mathbf{y}} = \boldsymbol{\Sigma}^2_{\mathbf{y}}$ and $\lim_{K\to\infty}\boldsymbol{\sigma}_{K,n,\tilde{x}\tilde{\mathbf{y}}} = \boldsymbol{\sigma}_{n,\tilde{x}\tilde{\mathbf{y}}}$.

II. LS-BASED FRAMEWORK FOR TIME-DOMAIN AND FREQUENCY-DOMAIN MMSE-TEQ DESIGNS

In this section, an LS-based framework for TEQ design is established. Several LS-based TEQ design criteria are presented and discussed. Starting from a CLLS-based formulation of the TD-MMSE-TEQ design criterion, two new intermediate

FD-MMSE-TEQ design criteria are derived. These establish a link between the CLLS-based TD-MMSE-TEQ design criterion on one hand and the IR-SNLLS-based formulation of the BM-TEQ design criterion on the other hand. This framework then allows to further develop and highlight our main contributions in Section III and Section IV.

A. CLLS-based TD-MMSE-TEQ design

The idea of MMSE-based channel shortening was first presented in [5]. The idea was later applied to DMT-based systems and referred to as MMSE-TEQ design in [6]. An MMSE-TEQ minimizes the *time-domain* (TD) MSE between the output of the T-tap TEQ, $y_{l,\mathbf{w}} = \mathbf{y}_l^T\mathbf{w}$, with l the sample index[1] and $\mathbf{y}_l = \begin{bmatrix} y_l & \cdots & y_{l-T+1} \end{bmatrix}^T$ a vector of RX samples[2], and the output $\mathbf{x}_l^T\mathbf{b}$ of a virtual FIR channel, the so-called TIR $\mathbf{b}$ of length $\nu+1$, which is fed with a vector of TX samples $\mathbf{x}_l = \begin{bmatrix} x_l & \cdots & x_{l-\nu} \end{bmatrix}^T$:

$$\min_{\mathbf{w},\mathbf{b}} \mathcal{E}\left\{|e_l|^2\right\} = \min_{\mathbf{w},\mathbf{b}} \mathcal{E}\left\{\left|\mathbf{y}_l^T\mathbf{w} - \mathbf{x}_l^T\mathbf{b}\right|^2\right\} \tag{1}$$

It follows from **Figure 1** that this TD-MMSE-TEQ criterion

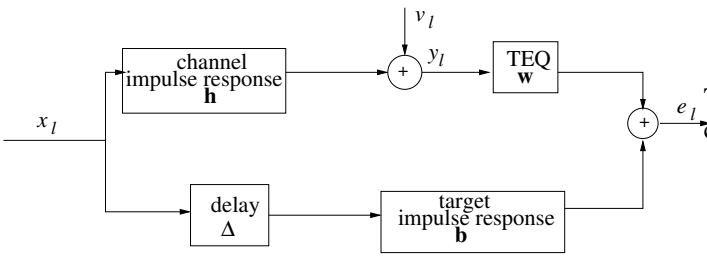


Fig. 1. Classical TD-MMSE-TEQ design.

(1) forces the joint channel-TEQ impulse response to have a main energy window of length $\nu+1$. To avoid the trivial solution $\mathbf{w}=\mathbf{0}$, $\mathbf{b}=\mathbf{0}$, a nontriviality constraint is added. Several constraints have been presented in [7]: a unit tap constraint $\mathbf{e}_i^T\mathbf{w}=1$ or $\mathbf{e}_i^T\mathbf{b}=1$ (with $\mathbf{e}_i$ the canonical vector with a 1 on position i), a unit norm constraint[3] $\|\mathbf{w}\|^2 = \mathbf{w}^T\mathbf{w} = 1$ or $\|\mathbf{b}\|^2 = \mathbf{b}^T\mathbf{b} = 1$ and a weighted unit norm constraint $\mathbf{w}^T\mathbf{G}\mathbf{w} = 1$ or $\mathbf{b}^T\mathbf{G}\mathbf{b} = 1$ (with $\mathbf{G}$ a positive definite matrix). Recently, particular choices of $\mathbf{G}$ have been considered. So-called unit energy constraints (UEC) on the TEQ output, $\mathcal{E}\left\{\left|\mathbf{y}_l^T\mathbf{w}\right|^2\right\} = 1$, or the TIR output, $\mathcal{E}\left\{\left|\mathbf{x}_l^T\mathbf{b}\right|^2\right\} = 1$, have been suggested in [8] and lead to:

$$\mathbf{w}^T\boldsymbol{\Sigma}_{\mathbf{y}}^2\mathbf{w} = 1 \text{ or } \mathbf{b}^T\boldsymbol{\Sigma}_{\mathbf{x}}^2\mathbf{b} = 1 \tag{2}$$

where $\boldsymbol{\Sigma}_{\mathbf{y}}^2 = \mathcal{E}\left\{\mathbf{y}_l\mathbf{y}_l^T\right\}$ and $\boldsymbol{\Sigma}_{\mathbf{x}}^2 = \mathcal{E}\left\{\mathbf{x}_l\mathbf{x}_l^T\right\}$ are autocorrelation matrices. It has been shown in [8] that both UECs lead to the same TEQ solution (up to a scaling). In the sequel, we will focus on the TD-MMSE-TEQ with a UEC to derive new FD-MMSE-TEQ design criteria, which allow to establish a relationship with the BM-TEQ and PTEQ.

Under the assumption that the involved signals are WSS and for $L \to \infty$, the stochastic linear MMSE criterion (1) with UEC constraint on $\mathbf{w}$ (2) is equivalent to the following deterministic **constrained linear least-squares (CLLS) based TD-MMSE-TEQ** criterion:

$$\min_{\mathbf{w},\mathbf{b}} \frac{1}{L}\sum_{l=1}^{L}\left|\mathbf{y}_l^T\mathbf{w} - \mathbf{x}_l^T\mathbf{b}\right|^2 \text{ s.t. } \mathbf{w}^T\boldsymbol{\Sigma}_{L,\mathbf{y}}^2\mathbf{w} = 1 \tag{3}$$

with L the total number of available data samples and $\boldsymbol{\Sigma}_{L,\mathbf{y}}^2 = \frac{1}{L}\sum_{l=1}^{L}\mathbf{y}_l\mathbf{y}_l^T$ an estimate of $\boldsymbol{\Sigma}_{\mathbf{y}}^2$. The UEC $\mathbf{b}^T\boldsymbol{\Sigma}_{L,\mathbf{x}}^2\mathbf{b} = 1$ with $\boldsymbol{\Sigma}_{L,\mathbf{x}}^2 = \frac{1}{L}\sum_{l=1}^{L}\mathbf{x}_l\mathbf{x}_l^T$ gives again the same TEQ solution (up to a scaling) [8]. Due to the so-called orthogonality condition, the solution for the unconstrained vector $\mathbf{b}$ depends on $\mathbf{w}$ through [9]:

$$\mathbf{b} = \left(\boldsymbol{\Sigma}_{L,\mathbf{x}}^2\right)^{-1}\boldsymbol{\Sigma}_{L,\mathbf{xy}}\mathbf{w} \tag{4}$$

with $\boldsymbol{\Sigma}_{L,\mathbf{xy}} = \frac{1}{L}\sum_{l=1}^{L}\mathbf{x}_l\mathbf{y}_l^T$. Combining (4) with (3) gives:

$$\min_{\mathbf{w}} \mathbf{w}^T\left[\boldsymbol{\Sigma}_{L,\mathbf{y}}^2 - \boldsymbol{\Sigma}_{L,\mathbf{xy}}^T\left(\boldsymbol{\Sigma}_{L,\mathbf{x}}^2\right)^{-1}\boldsymbol{\Sigma}_{L,\mathbf{xy}}\right]\mathbf{w} \quad \text{s.t. } \mathbf{w}^T\boldsymbol{\Sigma}_{L,\mathbf{y}}^2\mathbf{w} = 1 \tag{5}$$

This optimization criterion is equivalent to the following criteria:

$$\min_{\mathbf{w}} \frac{\mathbf{w}^T\left[\boldsymbol{\Sigma}_{L,\mathbf{y}}^2 - \boldsymbol{\Sigma}_{L,\mathbf{xy}}^T\left(\boldsymbol{\Sigma}_{L,\mathbf{x}}^2\right)^{-1}\boldsymbol{\Sigma}_{L,\mathbf{xy}}\right]\mathbf{w}}{\mathbf{w}^T\boldsymbol{\Sigma}_{L,\mathbf{y}}^2\mathbf{w}} = \mathbf{1} - \max_{\mathbf{w}} \frac{\mathbf{w}^{\mathbf{T}}\overbrace{\left[\boldsymbol{\Sigma}_{L,\mathbf{xy}}^T\left(\boldsymbol{\Sigma}_{L,\mathbf{x}}^2\right)^{-1}\boldsymbol{\Sigma}_{L,\mathbf{xy}}\right]}^{\mathbf{B}}\mathbf{w}}{\mathbf{w}^{\mathbf{T}}\underbrace{\boldsymbol{\Sigma}_{L,\mathbf{y}}^2}_{\mathbf{A}}\mathbf{w}} \tag{6}$$

Hence, solving the CLLS-based TD-MMSE-TEQ design criterion (3) reduces to the maximization of the generalized Rayleigh quotient $\frac{\mathbf{w}^T\mathbf{B}\mathbf{w}}{\mathbf{w}^T\mathbf{A}\mathbf{w}}$ in (6). The solution to (6) is seen to be the dominant generalized eigenvector of the GEV problem [17]

$$\mathbf{B}\mathbf{w} = \lambda\mathbf{A}\mathbf{w} \tag{7}$$

with the matrix pair

$$(\mathbf{B},\mathbf{A}) = \left(\boldsymbol{\Sigma}_{L,\mathbf{xy}}^T\left(\boldsymbol{\Sigma}_{L,\mathbf{x}}^2\right)^{-1}\boldsymbol{\Sigma}_{L,\mathbf{xy}}, \boldsymbol{\Sigma}_{L,\mathbf{y}}^2\right) \tag{8}$$

Note: The authors of [15] were the first to put forth a GEV formulation in the context of TEQ design: they presented a TEQ design based on a so-called maximum shortening SNR (MSSNR). Recently, an alternative GEV-based formulation of the MMSE-TEQ design was presented in [9]: based on (4), the MMSE-TEQ criterion with UNC on $\mathbf{b}$ is turned into an MMSE-TEQ criterion with weighted UNC on $\mathbf{w}$, $\mathbf{w}^T\mathbf{G}\mathbf{w} = 1$, which resembles (5), be it with a different constraint weighting matrix $\mathbf{G}$. Remarkably, the MSSNR-TEQ design turns out to be a special case of this MMSE-TEQ design where the noise is disregarded in the optimization [9].

[1]This early TEQ design ignores the DMT block transmission nature: it is sample-based, hence the use of a sample index l rather than the DMT symbol index k.

[2]The received signal y_l and vector $\mathbf{y}_l$ depend on the synchronization delay Δ, a measure for the channel group delay, which we do not mention explicitly here, for the sake of conciseness.

[3]In [7], the term "unit energy constraint" was used for this case.

B. CLLS-based FD-MMSE-TEQ design

As the TD-MMSE-TEQ design in (1) is *sample*-based and minimizes a *TD MSE*, the DMT *block* transmission structure (including the cyclic prefix and the *FD symbols of interest*) is not easily accounted for: oftentimes, the signals x_l and y_l are assumed stationary, i.e., $\mathbf{\Sigma}^2_{\mathbf{x}}$ is approximated by an identity matrix (even when not all tones are active) and $\mathbf{\Sigma}^2_{\mathbf{y}}$ is assumed Toeplitz. Also, there is no clear relation between minimizing the TD MSE $\mathcal{E}\left\{|e_l|^2\right\}$ in (1) and bitrate optimization. In this section and in the next section, we suggest two new, alternative MMSE-TEQ designs that do take into account the DMT block transmission structure and minimize a FD instead of TD MSE. The steps, explained below, that lead to these FD-MMSE-TEQ design criteria intuitively suggest that they are better related than the TD-MMSE-TEQ design criterion to the multicarrier, hence frequency domain nature of DMT equalization. Especially, the separable nonlinear LS (SNLLS) based FD-MMSE-TEQ criterion, developed in Section II-C, closely approaches the BM-TEQ performance (see Section V) and the design criterion is closely related to the so-called iteratively-reweighted SNLLS-based BM-TEQ design criterion, developed in Section II-D.

We propose to first rewrite the TD-MMSE-TEQ design criterion (1) on a *per-DMT-symbol* basis:

$$\min_{\mathbf{w},\mathbf{b}} \underbrace{\mathcal{E}\left\{\|\mathbf{Y}_k\mathbf{w}-\mathbf{X}_k\mathbf{b}\|^2\right\}}_{\mathcal{E}\left\{\|\mathbf{e}_k\|^2\right\}} \text{ s.t. } \mathbf{w}^T\mathbf{\Sigma}^2_{\mathbf{Y}}\mathbf{w}=1 \tag{9}$$

where $\mathbf{\Sigma}^2_{\mathbf{Y}} = \mathcal{E}\left\{\mathbf{Y}_k^T\mathbf{Y}_k\right\}$ is an autocorrelation matrix and where the Toeplitz matrices $\mathbf{Y}_k$ (size $N\times T$) and $\mathbf{X}_k$ (size $N\times(\nu+1)$ and columnwise circulant; $\mathbf{X}_k$ incorporates the cyclic prefix in the upper right triangle with side ν) are defined as:

$$\mathbf{Y}_k = \begin{bmatrix} y_{k,0} & y_{k,-1} & \cdots & y_{k,-T+1} \\ y_{k,1} & y_{k,2} & \ddots & y_{k,-T+2} \\ \vdots & \ddots & \ddots & \vdots \\ y_{k,N-1} & y_{k,N-2} & \cdots & y_{k,N-T} \end{bmatrix} \tag{10}$$

$$\mathbf{X}_k = \begin{bmatrix} x_{k,0} & x_{k,N-1} & \cdots & x_{k,N-\nu} \\ \vdots & \ddots & \ddots & \vdots \\ x_{k,\nu} & \ddots & x_{k,0} & x_{k,N-1} \\ x_{k,\nu+1} & \ddots & x_{k,1} & x_{k,0} \\ \vdots & \ddots & \ddots & \vdots \\ x_{k,N-1} & \cdots & x_{k,N-\nu} & x_{k,N-\nu-1} \end{bmatrix}$$

where $y_{k,i}$ and $x_{k,i}$ denote the TD sample with index i of the k-th RX and TX DMT symbol, respectively, i.e., $y_{k,i} = y_{(k-1)(N+\nu)+\nu+i}$, $i=-T+1,\ldots,N-1$ and $x_{k,i} = x_{(k-1)(N+\nu)+\nu+i}$, $i=0,\ldots,N-1$ (note that the subscript $(k-1)(N+\nu)+\nu+i$ is an absolute time index). The first term $\mathbf{Y}_k\mathbf{w}$ in (9) is the convolution of the k-th RX DMT symbol with the TEQ and hence is the $N\times 1$ TEQ output vector that is fed to the RX DFT. The second term $\mathbf{X}_k\mathbf{b}$ in (9) is the convolution of the k-th TX DMT symbol and the TIR $\mathbf{b}$.

In a second step, $\mathbf{X}_k$ is extended with $N-\nu-1$ columns to a circulant matrix $\mathbf{X}_{k,\mathrm{C}}$ of size $N\times N$, so that

$$\mathbf{X}_k\mathbf{b} = \mathbf{X}_{k,\mathrm{C}}\begin{bmatrix}\mathbf{b}\\ \mathbf{0}_{(N-\nu-1)\times 1}\end{bmatrix} \tag{11}$$

with (12) on the facing page A DFT-based decomposition of the circulant matrix $\mathbf{X}_{k,\mathrm{C}} = \mathcal{F}_N^H\tilde{\mathbf{X}}_{k,N,\mathrm{D}}\mathcal{F}_N$ with $\tilde{\mathbf{X}}_{k,N,\mathrm{D}} = \mathrm{diag}(\tilde{\mathbf{x}}_{k,N})$ is then plugged in:

$$\begin{aligned}\mathbf{X}_k\mathbf{b} &= \mathbf{X}_{k,\mathrm{C}}\begin{bmatrix}\mathbf{b}\\ \mathbf{0}\end{bmatrix} \\ &= \mathcal{F}_N^H\tilde{\mathbf{X}}_{k,N,\mathrm{D}}\underbrace{\mathcal{F}_N\begin{bmatrix}\mathbf{b}\\ \mathbf{0}\end{bmatrix}}_{\tilde{\mathbf{b}}_N} = \mathcal{F}_N^H\tilde{\mathbf{X}}_{k,N,\mathrm{D}}\tilde{\mathbf{b}}_N\end{aligned} \tag{13}$$

where $\tilde{\mathbf{b}}_N$ is the N-point DFT of the TIR $\mathbf{b}$ padded with zeros. Thirdly, the cost function and constraint (9) are transformed to the frequency domain:

$$\begin{aligned}\min_{\mathbf{w},\mathbf{b}}\mathcal{E}\left\{\|\mathbf{e}_k\|^2\right\} &= \min_{\mathbf{w},\mathbf{b}}\mathcal{E}\left\{\|\mathcal{F}_N\mathbf{e}_k\|^2\right\} \\ &= \min_{\mathbf{w},\mathbf{b}}\mathcal{E}\left\{\left\|\tilde{\mathbf{Y}}_{k,N}\mathbf{w}-\tilde{\mathbf{X}}_{k,N,\mathrm{D}}\tilde{\mathbf{b}}_N\right\|^2\right\}\end{aligned} \tag{14}$$

$$\text{s.t. } \mathbf{w}^T\mathbf{\Sigma}^2_{N,\tilde{\mathbf{Y}}}\mathbf{w}=1,\ \tilde{\mathbf{b}}_N = \mathcal{F}\begin{bmatrix}\mathbf{b}\\ \mathbf{0}\end{bmatrix} \text{ and real } \mathbf{w} \text{ and } \mathbf{b} \tag{15}$$

The first term of the error vector $\mathcal{F}_N\mathbf{e}_k$ in (14) corresponds to the RX DFT output:

$$\underbrace{\mathcal{F}_N(\mathbf{Y}_k\mathbf{w})}_{\{1\}} = \underbrace{(\mathcal{F}_N\mathbf{Y}_k)\mathbf{w}}_{\{2\}} = \tilde{\mathbf{Y}}_{k,N}\mathbf{w} = \tilde{\mathbf{y}}_{k,N,\mathbf{w}} \tag{16}$$

which can either be computed as $\{1\}$ the DFT of the TEQ output $\mathbf{Y}_k\mathbf{w}$ or $\{2\}$ as a linear combination $\mathbf{w}$ of the N-point sliding DFT of the k-th RX DMT symbol, $\tilde{\mathbf{Y}}_{k,N} = \mathcal{F}_N\mathbf{Y}_k$ (see [10] for details). The autocorrelation matrix in (15) is then given by $\mathbf{\Sigma}^2_{N,\tilde{\mathbf{Y}}} = \mathcal{E}\left\{\tilde{\mathbf{Y}}^H_{k,N}\tilde{\mathbf{Y}}_{k,N}\right\} = \mathcal{E}\left\{\mathbf{Y}_k^T\mathbf{Y}_k\right\} = \mathbf{\Sigma}^2_{\mathbf{Y}}$. The additional, second constraint in (15) comes from the definition of $\tilde{\mathbf{b}}_N$ in (13) and stems from the original TD-MMSE-TEQ design that imposes channel shortening by means of a TIR $\mathbf{b}$ of length $\nu+1$. Criterion (9) and (14-15) are thus fully equivalent.

Finally, by selecting the appropriate elements of $\mathcal{F}_N\mathbf{e}_k$ in (14) and modifying the constraints (15) accordingly, one can restrict the criterion to the N_a active tones of $\mathcal{S}_a$:

$$\min_{\mathbf{w},\mathbf{b}}\underbrace{\mathcal{E}\left\{\|\mathcal{F}_{\mathcal{S}_a}\mathbf{e}_k\|^2\right\}}_{\mathcal{E}\left\{\|\tilde{\mathbf{e}}_k\|^2\right\}} = \min_{\mathbf{w},\mathbf{b}}\mathcal{E}\left\{\left\|\tilde{\mathbf{Y}}_k\mathbf{w}-\tilde{\mathbf{X}}_{k,\mathrm{D}}\tilde{\mathbf{b}}\right\|^2\right\} \tag{17}$$

$$\text{s.t. } \mathbf{w}^T\mathbf{\Sigma}^2_{\tilde{\mathbf{Y}}}\mathbf{w}=1,\ \tilde{\mathbf{b}} = \mathcal{F}_{\mathcal{S}_a}\begin{bmatrix}\mathbf{b}\\ \mathbf{0}\end{bmatrix} \text{ and real } \mathbf{w} \text{ and } \mathbf{b} \tag{18}$$

where $\tilde{\mathbf{e}}_k = \mathcal{F}_{\mathcal{S}_a}\mathbf{e}_k$, $\tilde{\mathbf{Y}}_k = \mathcal{F}_{\mathcal{S}_a}\mathbf{Y}_k$, $\tilde{\mathbf{X}}_{k,\mathrm{D}} = \mathrm{diag}(\tilde{\mathbf{x}}_k)$ and $\mathbf{\Sigma}^2_{\tilde{\mathbf{Y}}} = \mathcal{E}\left\{\tilde{\mathbf{Y}}^H_k\tilde{\mathbf{Y}}_k\right\}$. The (modified) second constraint, $\tilde{\mathbf{b}} =$

$$\mathbf{X}_{k,\mathrm{C}} = \left[\begin{array}{cccc|ccc} x_{k,0} & x_{k,N-1} & \cdots & x_{k,N-\nu} & x_{k,N-\nu-1} & \cdots & x_{k,1} \\ x_{k,1} & x_{k,0} & \ddots & x_{k,N-\nu+1} & x_{k,N-\nu} & & x_{k,2} \\ \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{k,N-1} & \cdots & & x_{k,N-\nu-1} & x_{k,N-\nu-2} & \cdots & x_{k,0} \end{array}\right] \tag{12}$$

$\mathcal{F}_{\mathcal{S}_a}\begin{bmatrix}\mathbf{b}\\ \mathbf{0}\end{bmatrix}$, still follows from imposing channel shortening by means of a TIR $\mathbf{b}$ of length $\nu+1$: from (13), it is seen that

$$\begin{aligned}\mathcal{F}_{\mathcal{S}_a}\mathbf{X}_k\mathbf{b} &= \underbrace{\mathcal{F}_{\mathcal{S}_a}\mathcal{F}_N^H}_{\mathbf{S}_{N_a}}\underbrace{\mathrm{diag}(\tilde{\mathbf{x}}_{k,N})}_{\tilde{\mathbf{X}}_{k,N,\mathrm{D}}}\underbrace{\mathcal{F}_N\begin{bmatrix}\mathbf{b}\\ \mathbf{0}\end{bmatrix}}_{\tilde{\mathbf{b}}_N} \\ &= \underbrace{\mathrm{diag}(\tilde{\mathbf{x}}_k)}_{\tilde{\mathbf{X}}_{k,\mathrm{D}}}\underbrace{\mathbf{S}_{N_a}\mathcal{F}_N\begin{bmatrix}\mathbf{b}\\ \mathbf{0}\end{bmatrix}}_{\tilde{\mathbf{b}}} = \tilde{\mathbf{X}}_{k,\mathrm{D}}\underbrace{\mathcal{F}_{\mathcal{S}_a}\begin{bmatrix}\mathbf{b}\\ \mathbf{0}\end{bmatrix}}_{\tilde{\mathbf{b}}}\end{aligned} \tag{19}$$

where the $N_a\times N$ selection matrix $\mathbf{S}_{N_a}=\mathcal{F}_{\mathcal{S}_a}\mathcal{F}_N^H$ is obtained from the $N\times N$ identity matrix $\mathbf{I}_N$ by selecting the N_a rows corresponding to the tones $\mathcal{S}_a$. This TEQ design criterion (17-18) has been previously presented in [18]. If we drop the constraints on $\tilde{\mathbf{b}}$ in (18), i.e., the constraints of $\mathbf{b}$ being real and having length $\nu+1$, and solve instead

$$\min_{\mathbf{w},\tilde{\mathbf{b}}}\mathcal{E}\left\{\left\|\tilde{\mathbf{Y}}_k\mathbf{w}-\tilde{\mathbf{X}}_{k,\mathrm{D}}\tilde{\mathbf{b}}\right\|^2\right\}\ \text{s.t. }\mathbf{w}^T\boldsymbol{\Sigma}^2_{\tilde{\mathbf{Y}}}\mathbf{w}=1\text{ and real }\mathbf{w} \tag{20}$$

we obtain an *FD-MMSE-TEQ design criterion* in the (typically) real-valued TEQ $\mathbf{w}$ and the complex-valued vector $\tilde{\mathbf{b}}$ instead of $\mathbf{b}$, that minimizes a sum of FD error energies at the DFT output of the active tones $\mathcal{S}_a$. By dropping the constraints on $\mathbf{b}$, the search space of $\tilde{\mathbf{b}}$ is enlarged, so that the FD error energy in (20) is smaller than in (17-18). The impact of removing the constraints on $\tilde{\mathbf{b}}$ is studied in the simulations of Section V.

The optimum solution for the unconstrained $\tilde{\mathbf{b}}$, which follows from the so-called orthogonality principle applied to (20), is a vector with the inverses of the unbiased MMSE (uMMSE) FEQs $\tilde{\mathbf{d}}^{\mathrm{uMMSE}}$ (also referred to as zero-forcing FEQs), which are in fact the optimal choice of FEQs for a given $\mathbf{w}$ [10], [19]:

$$\tilde{d}_n^{\mathrm{uMMSE}}=\frac{\sigma^2_{n,\tilde{x}}}{\boldsymbol{\sigma}_{n,\tilde{x}\tilde{\mathbf{y}}}\mathbf{w}}=\frac{1}{\tilde{b}_n} \tag{21}$$

The numerator $\sigma^2_{n,\tilde{x}}=\mathcal{E}\left\{|\tilde{x}_{k,n}|^2\right\}$ is the variance of $\tilde{x}_{k,n}$; the denominator is the crosscorrelation $\mathcal{E}\left\{\tilde{x}^*_{k,n}\tilde{y}_{k,n,\mathbf{w}}\right\}$ between the RX DFT output $\tilde{y}_{k,n,\mathbf{w}}$ and the TX symbol $\tilde{x}_{k,n}$, both on tone n. It follows from (16) for a particular tone n, i.e., $\tilde{y}_{k,n,\mathbf{w}}=\mathcal{F}_n\tilde{\mathbf{Y}}_k\mathbf{w}=\tilde{\mathbf{y}}_{k,n}\mathbf{w}$, that this crosscorrelation is equal to $\boldsymbol{\sigma}_{n,\tilde{x}\tilde{\mathbf{y}}}\mathbf{w}$, with $\boldsymbol{\sigma}_{n,\tilde{x}\tilde{\mathbf{y}}}=\mathcal{E}\left\{\tilde{x}^*_{k,n}\tilde{\mathbf{y}}_{k,n}\right\}$ the $1\times T$ crosscorrelation vector of $\tilde{x}_{k,n}$ and the n-th sliding DFT output $\tilde{\mathbf{y}}_{k,n}=\mathcal{F}_n\mathbf{Y}_k$ (see [10] for details). Solving (20) then minimizes the sum of error energies between the DFT outputs and the scaled desired symbols $\frac{\tilde{x}_{k,n}}{\tilde{d}_n^{\mathrm{uMMSE}}}$.

Following the derivation in [8], a design criterion that leads to the same TEQ as (20) (up to a scaling) is obtained with the alternative UEC $\tilde{\mathbf{b}}^H\boldsymbol{\Sigma}^2_{\tilde{\mathbf{x}},\mathrm{D}}\tilde{\mathbf{b}}=1$ where $\boldsymbol{\Sigma}^2_{\tilde{\mathbf{x}},\mathrm{D}}=\mathcal{E}\left\{\tilde{\mathbf{X}}^*_{k,\mathrm{D}}\tilde{\mathbf{X}}_{k,\mathrm{D}}\right\}$ is a diagonal matrix with the variance of the TX symbols $\tilde{\mathbf{x}}_k$ on the diagonal. In case of equipowered TX symbols $\tilde{x}_{k,n}$, this constraint reduces to a unit norm constraint $\tilde{\mathbf{b}}^H\tilde{\mathbf{b}}=1$.

Under the assumption of WSS signals and for $K\to\infty$, a deterministic **CLLS-based FD-MMSE-TEQ design criterion**, equivalent with the stochastic criterion (20), is given by:

$$\min_{\mathbf{w},\tilde{\mathbf{b}}}\frac{1}{K}\sum_{k=1}^{K}\left\|\tilde{\mathbf{Y}}_k\mathbf{w}-\tilde{\mathbf{X}}_{k,\mathrm{D}}\tilde{\mathbf{b}}\right\|^2 \quad \text{s.t. }\mathbf{w}^T\boldsymbol{\Sigma}^2_{K,\tilde{\mathbf{Y}}}\mathbf{w}=1\text{ and with real }\mathbf{w} \tag{22}$$

where K is the number of available DMT symbols and with $\boldsymbol{\Sigma}^2_{K,\tilde{\mathbf{Y}}}=\frac{1}{K}\sum_{k=1}^{K}\tilde{\mathbf{Y}}_k^H\tilde{\mathbf{Y}}_k$. Note that $\boldsymbol{\Sigma}^2_{K,\tilde{\mathbf{Y}}}=\sum_{n\in\mathcal{S}_a}\boldsymbol{\Sigma}^2_{K,n,\tilde{\mathbf{y}}}$ with $\boldsymbol{\Sigma}^2_{K,n,\tilde{\mathbf{y}}}=\frac{1}{K}\sum_{k=1}^{K}\tilde{\mathbf{y}}_{k,n}^H\tilde{\mathbf{y}}_{k,n}$ the autocorrelation matrix of the n-th sliding DFT output $\tilde{\mathbf{y}}_{k,n}$. Solving (22) for the unconstrained $\tilde{\mathbf{b}}$ (again using the orthogonality principle) gives an LS estimate (for a given $\mathbf{w}$) of the inverses of the uMMSE FEQs (21):

$$\tilde{\mathbf{b}}_K=\left(\boldsymbol{\Sigma}^2_{K,\tilde{\mathbf{x}},\mathrm{D}}\right)^{-1}\boldsymbol{\Sigma}_{K,\tilde{\mathbf{x}}\tilde{\mathbf{Y}}}\mathbf{w}\text{ with entries }\tilde{b}_{K,n}=\frac{\boldsymbol{\sigma}_{K,n,\tilde{x}\tilde{\mathbf{y}}}\mathbf{w}}{\sigma^2_{K,n,\tilde{x}}} \tag{23}$$

where $\boldsymbol{\Sigma}^2_{K,\tilde{\mathbf{x}},\mathrm{D}}=\frac{1}{K}\sum_{k=1}^{K}\tilde{\mathbf{X}}^*_{k,\mathrm{D}}\tilde{\mathbf{X}}_{k,\mathrm{D}}$ is diagonal with diagonal elements $\sigma^2_{K,n,\tilde{x}}=\frac{1}{K}\sum_{k=1}^{K}|\tilde{x}_{k,n}|^2$, and where the rows of the $N_a\times T$ crosscorrelation matrix estimate $\boldsymbol{\Sigma}_{K,\tilde{\mathbf{x}}\tilde{\mathbf{Y}}}=\frac{1}{K}\sum_{k=1}^{K}\tilde{\mathbf{X}}^*_{k,\mathrm{D}}\tilde{\mathbf{Y}}_k$ are the crosscorrelation vectors $\boldsymbol{\sigma}_{K,n,\tilde{x}\tilde{\mathbf{y}}}=\frac{1}{K}\sum_{k=1}^{K}\tilde{x}^*_{k,n}\tilde{\mathbf{y}}_{k,n}$. The CLLS-based FD-MMSE-TEQ design criterion (22) is very similar to the original CLLS-based TD-MMSE-TEQ (3). Hence, combining (23) and (22) results in an optimization criterion that is very similar to (5), namely:

$$\min_{\mathbf{w}}\mathbf{w}^T\left[\boldsymbol{\Sigma}^2_{K,\tilde{\mathbf{Y}}}-\boldsymbol{\Sigma}^H_{K,\tilde{\mathbf{x}}\tilde{\mathbf{Y}}}\left(\boldsymbol{\Sigma}^2_{K,\tilde{\mathbf{x}},\mathrm{D}}\right)^{-1}\boldsymbol{\Sigma}_{K,\tilde{\mathbf{x}}\tilde{\mathbf{Y}}}\right]\mathbf{w} \quad \text{s.t. }\mathbf{w}^T\boldsymbol{\Sigma}^2_{K,\tilde{\mathbf{Y}}}\mathbf{w}=1\text{ and with real }\mathbf{w} \tag{24}$$

This optimization criterion (24) is equivalent to:

$$\min_{\mathbf{w}}\frac{\mathbf{w}^T\left[\boldsymbol{\Sigma}^2_{K,\tilde{\mathbf{Y}}}-\boldsymbol{\Sigma}^H_{K,\tilde{\mathbf{x}}\tilde{\mathbf{Y}}}\left(\boldsymbol{\Sigma}^2_{K,\tilde{\mathbf{x}},\mathrm{D}}\right)^{-1}\boldsymbol{\Sigma}_{K,\tilde{\mathbf{x}}\tilde{\mathbf{Y}}}\right]\mathbf{w}}{\mathbf{w}^T\boldsymbol{\Sigma}^2_{K,\tilde{\mathbf{Y}}}\mathbf{w}} \tag{25}$$

$$=1-\max_{\mathbf{w}}\frac{\mathbf{w}^T\left[\boldsymbol{\Sigma}^H_{K,\tilde{\mathbf{x}}\tilde{\mathbf{Y}}}\left(\boldsymbol{\Sigma}^2_{K,\tilde{\mathbf{x}},\mathrm{D}}\right)^{-1}\boldsymbol{\Sigma}_{K,\tilde{\mathbf{x}}\tilde{\mathbf{Y}}}\right]\mathbf{w}}{\mathbf{w}^T\boldsymbol{\Sigma}^2_{K,\tilde{\mathbf{Y}}}\mathbf{w}}\text{ with real }\mathbf{w} \tag{26}$$

$$=1-\max_{\mathbf{w}}\frac{\mathbf{w}^T\left[\Re\left\{\boldsymbol{\Sigma}^H_{K,\tilde{\mathbf{x}}\tilde{\mathbf{Y}}}\left(\boldsymbol{\Sigma}^2_{K,\tilde{\mathbf{x}},\mathrm{D}}\right)^{-1}\boldsymbol{\Sigma}_{K,\tilde{\mathbf{x}}\tilde{\mathbf{Y}}}\right\}\right]\mathbf{w}}{\mathbf{w}^T\Re\left\{\boldsymbol{\Sigma}^2_{K,\tilde{\mathbf{Y}}}\right\}\mathbf{w}} \tag{27}$$

The constraint in (25) of a real-valued TEQ is imposed in (27) by inserting the real operators $\Re\{\cdot\}$. The equivalence of (25) and (27) then follows from the Hermitian property of the matrices $\boldsymbol{\Sigma}^2_{K,\tilde{\mathbf{Y}}}$ and $\boldsymbol{\Sigma}^H_{K,\tilde{\mathbf{x}}\tilde{\mathbf{Y}}}\left(\boldsymbol{\Sigma}^2_{K,\tilde{\mathbf{x}},\mathrm{D}}\right)^{-1}\boldsymbol{\Sigma}_{K,\tilde{\mathbf{x}}\tilde{\mathbf{Y}}}$, which

ensures that the Rayleigh quotient in (25) is real. Criterion (27) reduces to a GEV problem (7), which is closely related to (8), namely with

$$\begin{aligned}(\mathbf{B},\mathbf{A}) &= \left(\Re\left\{\boldsymbol{\Sigma}^H_{K,\tilde{\mathbf{x}}\tilde{\mathbf{Y}}}\left(\boldsymbol{\Sigma}^2_{K,\tilde{\mathbf{x}},\mathrm{D}}\right)^{-1}\boldsymbol{\Sigma}_{K,\tilde{\mathbf{x}}\tilde{\mathbf{Y}}}\right\},\Re\left\{\boldsymbol{\Sigma}^2_{K,\tilde{\mathbf{Y}}}\right\}\right)\\ &= \left(\Re\left\{\sum_{n\in\mathcal{S}_a}\sigma^{-2}_{K,n,\tilde{x}}\boldsymbol{\sigma}^H_{K,n,\tilde{x}\tilde{\mathbf{y}}}\boldsymbol{\sigma}_{K,n,\tilde{x}\tilde{\mathbf{y}}}\right\},\right. \qquad (28)\\ &\quad \left.\Re\left\{\sum_{n\in\mathcal{S}_a}\boldsymbol{\Sigma}^2_{K,n,\tilde{\mathbf{y}}}\right\}\right)\end{aligned}$$

Simulation results with the CLLS-based TD-MMSE-TEQ (3) and CLLS-based FD-MMSE-TEQ (22) are included in Section V and compared with the CLLS-based FD-MMSE-TEQ design with constraints on $\tilde{\mathbf{b}}$ (17-18) of [18].

C. SNLLS-based FD-MMSE-TEQ design

Instead of minimizing the sum of *DFT output* energies, an alternative FD criterion is obtained by minimizing the sum of *FEQ output* energies:

$$\begin{aligned}&\min_{\mathbf{w},\tilde{\mathbf{d}}}\mathcal{E}\left\{\|\tilde{\mathbf{e}}_{k,\boldsymbol{\theta}}\|^2\right\} \qquad (29)\\ &=\min_{\mathbf{w},\tilde{\mathbf{d}}}\mathcal{E}\left\{\left\|\mathrm{diag}(\tilde{\mathbf{d}})\tilde{\mathbf{Y}}_k\mathbf{w}-\tilde{\mathbf{x}}_k\right\|^2\right\}\ \text{with real } \mathbf{w}\end{aligned}$$

where $\tilde{\mathbf{e}}_{k,\boldsymbol{\theta}}$ is the FEQ output error vector, which depends on both the real-valued TEQ $\mathbf{w}$ and complex-valued FEQ parameters $\tilde{\mathbf{d}}$, i.e., $\boldsymbol{\theta}=\begin{bmatrix}\mathbf{w}^H & \tilde{\mathbf{d}}^H\end{bmatrix}^H$; its first term $\mathrm{diag}(\tilde{\mathbf{d}})\tilde{\mathbf{Y}}_k\mathbf{w}$ is a vector with the FEQ outputs. The UEC constraint of (20) has been dropped as the criterion (29) does not have a trivial solution anymore. Under the assumption of WSS signals and for $K\rightarrow\infty$, this criterion corresponds to a separable nonlinear LS (SNLLS) criterion [20], [21]:

$$\min_{\boldsymbol{\theta}}\frac{1}{K}\sum_{k=1}^{K}\underbrace{\left\|\mathrm{diag}(\tilde{\mathbf{d}})\tilde{\mathbf{Y}}_k\mathbf{w}-\tilde{\mathbf{x}}_k\right\|^2}_{\|\tilde{\mathbf{e}}_{k,\boldsymbol{\theta}}\|^2}\ \text{with real } \mathbf{w} \qquad (30)$$

which we refer to as an **SNLLS-based FD-MMSE-TEQ** design criterion: although the error $\tilde{\mathbf{e}}_{k,\boldsymbol{\theta}}$ is nonlinear in $\boldsymbol{\theta}$, the TEQ $\mathbf{w}$ and FEQs $\tilde{\mathbf{d}}$ appear linearly, hence the separability property. It follows from this separability property that solving (29) as a linear MMSE problem in $\tilde{\mathbf{d}}$, while keeping $\mathbf{w}$ fixed, results in the biased MMSE (bMMSE) FEQs for the given $\mathbf{w}$ [10], [19]:

$$\tilde{d}_n^{\mathrm{MMSE}}=\frac{\mathbf{w}^T\boldsymbol{\sigma}^H_{n,\tilde{x}\tilde{\mathbf{y}}}}{\mathbf{w}^T\boldsymbol{\Sigma}^2_{n,\tilde{\mathbf{y}}}\mathbf{w}} \qquad (31)$$

where the numerator is equal to the complex conjugate of the denominator of the uMMSE FEQ (21) and where the denominator is the autocorrelation of the DFT output, i.e., $\mathcal{E}\left\{|\tilde{y}_{k,n,\mathbf{w}}|^2\right\}=\mathbf{w}^T\boldsymbol{\Sigma}^2_{n,\tilde{\mathbf{y}}}\mathbf{w}$ with $\boldsymbol{\Sigma}^2_{n,\tilde{\mathbf{y}}}$ defined earlier.

The CLLS-based FD-MMSE-TEQ design (22) and the SNLLS-based FD-MMSE-TEQ design (30) establish a link between the TD-MMSE-TEQ design (3) and a new IR-SNLLS-based formulation of the BM-TEQ design, which is developed in the next section. Moreover, in Section IV-A, it will become clear that both the CLLS-based and the SNLLS-based FD-MMSE-TEQ designs are equivalent to the PTEQ design when considering only a single tone. It will be shown in Section II-D that the SNLLS-based FD-MMSE-TEQ design criterion, as well as the IR-SNLLS-based BM-TEQ design criterion, can be optimized iteratively with a sequence of GEV problems (7). An alternative, fast converging GN-based design algorithm that exploits the separability property in both criteria is presented in Section III. The simulations of Section V show that the SNLLS-based FD-MMSE-TEQ (30) consistently outperforms the CLLS-based TD-MMSE-TEQ design (3) and closely approaches the IR-SNLLS-based BM-TEQ performance.

D. IR-SNLLS-based BM-TEQ design

The bitrate maximizing TEQ (BM-TEQ) is the solution to the following constrained nonlinear optimization problem in the joint TEQ-FEQ parameter vector $\boldsymbol{\theta}=\begin{bmatrix}\mathbf{w}^H & \tilde{\mathbf{d}}^H\end{bmatrix}^H$:

$$\max_{\boldsymbol{\theta}}\sum_{n\in\mathcal{S}_a}\log_2\left(1+\frac{\mathrm{SNR}_{n,\boldsymbol{\theta}_n}}{\Gamma_n}\right) \qquad (32)$$

$$\text{with}\quad \mathrm{SNR}_{n,\boldsymbol{\theta}_n}=\frac{\sigma^2_{n,\tilde{x}}}{\mathcal{E}\left\{|\tilde{e}_{k,n,\boldsymbol{\theta}_n}|^2\right\}}=\frac{\sigma^2_{n,\tilde{x}}}{\mathcal{E}\left\{|\tilde{d}_n\tilde{\mathbf{y}}_{k,n}\mathbf{w}-\tilde{x}_{k,n}|^2\right\}}$$

$$\text{subject to a real } \mathbf{w} \text{ and } \tilde{d}_n=\frac{\sigma^2_{n,\tilde{x}}}{\boldsymbol{\sigma}_{n,\tilde{x}\tilde{\mathbf{y}}}\mathbf{w}},\ \forall n\in\mathcal{S}_a \qquad (33)$$

i.e., maximizing the number of bits per DMT symbol (given a certain SNR gap Γ_n between SNR_n and the SNR required to achieve Shannon capacity, typically assumed to be independent of the equalizer [22]), over the joint TEQ-FEQ parameters $\boldsymbol{\theta}$, subject to the use of uMMSE FEQs (33) (see also (21)), which render the subchannel SNR model in (32) exact [10], [23]. In [10], it has been shown that this optimization problem (32-33) is equivalent to minimizing the BM-TEQ cost function:

$$\mathbf{w}=\arg\min_{\mathbf{w}}\sum_{n\in\mathcal{S}_a}\log\left(\frac{\mathbf{w}^T\mathbf{B}_n\mathbf{w}}{\mathbf{w}^T\mathbf{A}_n\mathbf{w}}\right) \qquad (34)$$

where the tone-dependent matrices $\mathbf{A}_n$ and $\mathbf{B}_n$ are independent of $\mathbf{w}$ and depend on the second-order statistics $\sigma^2_{n,\tilde{x}}$, $\boldsymbol{\Sigma}^2_{n,\tilde{\mathbf{y}}}=\mathcal{E}\left\{\tilde{\mathbf{y}}^H_{k,n}\tilde{\mathbf{y}}_{k,n}\right\}$ and $\boldsymbol{\sigma}_{n,\tilde{x}\tilde{\mathbf{y}}}=\mathcal{E}\left\{\tilde{x}^*_{k,n}\tilde{\mathbf{y}}_{k,n}\right\}$:

$$\begin{aligned}\mathbf{A}_n &= \Gamma_n\sigma^2_{n,\tilde{x}}\boldsymbol{\Sigma}^2_{n,\tilde{\mathbf{y}}}+(1-\Gamma_n)\boldsymbol{\sigma}^H_{n,\tilde{x}\tilde{\mathbf{y}}}\boldsymbol{\sigma}_{n,\tilde{x}\tilde{\mathbf{y}}} \qquad (35)\\ \mathbf{B}_n &= \Gamma_n\left(\sigma^2_{n,\tilde{x}}\boldsymbol{\Sigma}^2_{n,\tilde{\mathbf{y}}}-\boldsymbol{\sigma}^H_{n,\tilde{x}\tilde{\mathbf{y}}}\boldsymbol{\sigma}_{n,\tilde{x}\tilde{\mathbf{y}}}\right) \qquad (36)\end{aligned}$$

In the appendix, we show, based on (32-33), that, under the assumption of WSS signals and for $K\rightarrow\infty$, this original BM-TEQ criterion (34) is equivalent to the following **iteratively reweighted (IR) SNLLS-based bitrate maximizing TEQ** design criterion (explained below and to be compared to (30)) [24]:

$$\min_{\boldsymbol{\theta}}\frac{1}{K}\sum_{k=1}^{K}\underbrace{\left\|\mathrm{diag}\left(\sqrt{\check{\boldsymbol{\gamma}}_{K,\mathbf{w}_{\mathrm{prev}}}}\right)\left(\mathrm{diag}(\tilde{\mathbf{d}})\tilde{\mathbf{Y}}_k\mathbf{w}-\tilde{\mathbf{x}}_k\right)\right\|^2}_{\left\|\mathrm{diag}\left(\sqrt{\check{\boldsymbol{\gamma}}_{K,\mathbf{w}_{\mathrm{prev}}}}\right)\tilde{\mathbf{e}}_{k,\boldsymbol{\theta}}\right\|^2}\ \text{with real } \mathbf{w} \qquad (37)$$

This is clearly a weighted LS criterion where the entry of the $N_a\times 1$ weight vector $\check{\boldsymbol{\gamma}}_{K,\mathbf{w}_{\mathrm{prev}}}$ that corresponds to tone n is

found to equal:

$$\check{\gamma}_{K,n,\mathbf{w}} = \frac{\left(\mathrm{SNR}_{K,n,\mathbf{w}}+1\right)^2}{\sigma^2_{K,n,\tilde{x}}\left(\mathrm{SNR}_{K,n,\mathbf{w}}+\Gamma_n\right)} \quad (38)$$

$$\mathrm{SNR}_{K,n,\mathbf{w}} = \frac{1}{\rho^{-2}_{K,n,\mathbf{w}}-1} = \frac{\sigma^2_{K,n,\tilde{x}}}{\frac{1}{K}\sum_{k=1}^{K}\left|\tilde{e}_{k,n,\boldsymbol{\theta}_n}\right|^2} \quad (39)$$

$$\rho^2_{K,n,\mathbf{w}} = \frac{\left|\boldsymbol{\sigma}_{K,n,\tilde{x}\tilde{\mathbf{y}}}\mathbf{w}\right|^2}{\sigma^2_{K,n,\tilde{x}}\left(\mathbf{w}^T\boldsymbol{\Sigma}^2_{K,n,\tilde{\mathbf{y}}}\mathbf{w}\right)} \quad (40)$$

The error vector $\tilde{\mathbf{e}}_{k,\boldsymbol{\theta}} = \mathrm{diag}(\tilde{\mathbf{d}})\tilde{\mathbf{Y}}_k\mathbf{w} - \tilde{\mathbf{x}}_k$ is the same error vector as in the SNLLS-based FD-MMSE-TEQ design criterion (30), hence (37) is also an SNLLS-based criterion. According to [20], [21], SNLLS criteria such as the SNLLS-based FD-MMSE-TEQ design criterion (30) and the IR-SNLLS-based BM-TEQ design criterion (37) are solved iteratively, e.g., by alternately updating the parameters $\mathbf{w}$ and $\tilde{\mathbf{d}}$. The SNLLS-based FD-MMSE-TEQ design criterion (30) is indeed an unweighted version of, hence closely related to, the IR-SNLLS-based BM-TEQ design criterion (37). The algorithms for solving the IR-SNLLS-based BM-TEQ design criterion (37), which are given below, then also apply to the SNLLS-based FD-MMSE-TEQ design criterion (30) by setting the weights $\check{\boldsymbol{\gamma}}_{K,\mathbf{w}_{\mathrm{prev}}}$ equal to 1.

The cost criterion (37) is a so-called *iteratively reweighted* (IR) LS criterion. An IR-LS criterion forms a representation of a non-linear cost criterion, which can be written as a weighted LS criterion where the weights $\check{\boldsymbol{\gamma}}_{K,\mathbf{w}_{\mathrm{prev}}}$ depend on the LS errors $\mathbf{e}_{k,\boldsymbol{\theta}}$, $k = 1,\ldots,K$ (here: via the subchannel SNRs (39)), hence on the optimization parameters $\boldsymbol{\theta}$. IR-LS criteria are typically solved as an iterative sequence of weighted LS criteria (here: an SNLLS criterion) with fixed weights in each iteration, computed with the parameter estimates from the previous iteration, $\boldsymbol{\theta}_{\mathrm{prev}}$. According to [24], convergence occurs provided that the weights are bounded and non-increasing in the (absolute value) of the LS errors. For a non-convex cost function, the IR-LS algorithm leads to a local optimum. *For $K \to \infty$ and under the assumption of wide-sense stationarity of the involved signals, the iterative sequence of (deterministic) LS criteria (37) is equivalent to the original (stochastic) BM-TEQ criterion (34).*

Classical, oftentimes prohibitively complex, non-linear iterative batch optimization algorithms were proposed in [10] for optimizing the original non-linear BM-TEQ design criterion (34). The IR-SNLLS-based formulation of the BM-TEQ design criterion allows for elegant and tractable alternatives, as explained below. The i-th iteration step of a possible iterative algorithm for optimizing the IR-SNLLS-based BM-TEQ criterion consists of the computation of

1) an estimate of the weights, $\check{\boldsymbol{\gamma}}_{K,\mathbf{w}_{\mathrm{prev}}}$, based on (38-40), evaluated in the TEQ estimate $\mathbf{w}_{\mathrm{prev}} = \mathbf{w}_{i-1}$ of the previous iteration;
2) an LS estimate of the bMMSE FEQs in (31), $\tilde{\mathbf{d}}^{\mathrm{MMSE}}_{K,i}$, which are the solution of (37) for a fixed $\mathbf{w}_{i-1}$:

$$\tilde{d}^{\mathrm{MMSE}}_{K,n,i} = \frac{\mathbf{w}^T_{i-1}\boldsymbol{\sigma}^H_{K,n,\tilde{x}\tilde{\mathbf{y}}}}{\mathbf{w}^T_{i-1}\boldsymbol{\Sigma}^2_{K,n,\tilde{\mathbf{y}}}\mathbf{w}_{i-1}} \quad (41)$$

3) a new BM-TEQ estimate $\mathbf{w}_i$, which is the LS solution of (37) for a fixed $\tilde{\mathbf{d}}_{K,i} = \tilde{\mathbf{d}}^{\mathrm{MMSE}}_{K,i}$:

$$\mathbf{w}_i = \Re\left\{\sum_{n\in\mathcal{S}_a}\check{\gamma}_{K,n,\mathbf{w}_{i-1}}\left|\tilde{d}_{K,n,i}\right|^2\boldsymbol{\Sigma}^2_{K,n,\tilde{\mathbf{y}}}\right\}^{-1} \times \Re\left\{\sum_{n\in\mathcal{S}_a}\check{\gamma}_{K,n,\mathbf{w}_{i-1}}\tilde{d}^*_{K,n,i}\boldsymbol{\sigma}^H_{K,n,\tilde{x}\tilde{\mathbf{y}}}\right\} \quad (42)$$

For a complex-valued TEQ, the $\Re$-operators must be omitted. The iterations for solving the SNLLS-based FD-MMSE-TEQ design criterion (30) do not include the first step, i.e., the weights $\check{\gamma}_{K,n,\mathbf{w}_{\mathrm{prev}}}$ are always equal to 1.

The TEQ (42) is the solution (up to a scaling) of a GEV problem (7) with matrix pair:

$$(\mathbf{B},\mathbf{A}) = \left(\Re\left\{\sum_{n\in\mathcal{S}_a}\check{\gamma}_{K,n,\mathbf{w}_{\mathrm{prev}}}\tilde{d}^*_{K,n,i}\boldsymbol{\sigma}^H_{K,n,\tilde{x}\tilde{\mathbf{y}}}\right\}\times \Re\left\{\sum_{n\in\mathcal{S}_a}\check{\gamma}_{K,n,\mathbf{w}_{\mathrm{prev}}}\tilde{d}_{K,n,i}\boldsymbol{\sigma}_{K,n,\tilde{x}\tilde{\mathbf{y}}}\right\}, \Re\left\{\sum_{n\in\mathcal{S}_a}\check{\gamma}_{K,n,\mathbf{w}_{\mathrm{prev}}}\left|\tilde{d}_{K,n,i}\right|^2\boldsymbol{\Sigma}^2_{K,n,\tilde{\mathbf{y}}}\right\}\right) \quad (43)$$

As $\mathbf{A}$ is a positive-definite matrix and $\mathbf{B}$ has rank 1, the corresponding GEV problem has only 1 strictly positive GEV; the other GEVs are zero; the dominant generalized eigenvector has then indeed the closed form solution (42).

III. An iterative Gauss-Newton based algorithm for BM-TEQ design

The above presented three-step algorithm that alternately updates $\mathbf{w}$ and $\tilde{\mathbf{d}}$ converges only linearly. It is argued in [20], [21] that the convergence speed of an iterative algorithm that optimizes an SNLLS problem can be increased dramatically and becomes quadratic by exploiting the separability property in an appropriate way. Step (3), which solves for $\mathbf{w}$ while keeping $\tilde{\mathbf{d}}$ fixed, can be better replaced by a much faster converging updating step, based on a Gauss-Newton (GN) update of the joint parameter vector $\boldsymbol{\theta}$:

$$\boldsymbol{\theta}_i \leftarrow \boldsymbol{\theta}_{i-1} - \alpha_i\left(\nabla^2_{\boldsymbol{\theta}_{i-1}}J\right)^{-1}\nabla_{\boldsymbol{\theta}_{i-1}}J \quad (44)$$

where $\boldsymbol{\theta}_{i-1} = \begin{bmatrix}\mathbf{w}^H_{i-1} & \tilde{\mathbf{d}}^H_{K,i}\end{bmatrix}^H$ with $\mathbf{w}_{i-1} = \mathbf{w}_{\mathrm{prev}}$ obtained in the GN update of the previous iteration $i-1$ and $\tilde{\mathbf{d}}_{K,i}$ obtained by evaluating (41) (see step 2 of the current iteration i above), where $\nabla_{\boldsymbol{\theta}_{i-1}}J$ and $\nabla^2_{\boldsymbol{\theta}_{i-1}}J$ are the gradient and a positive-definite Hessian approximation, respectively, of the IR-SNLLS-based BM-TEQ design criterion J in (37), evaluated in $\boldsymbol{\theta}_{i-1}$, and where α_i is a step size, obtained through a line search algorithm [20]. The gradient $\nabla_{\boldsymbol{\theta}_{i-1}}J$ and approximate Hessian $\nabla^2_{\boldsymbol{\theta}_{i-1}}J$ in (44) for a *complex-valued*[4]

[4]The expressions for a real-valued TEQ are similar, but more cumbersome, because $\boldsymbol{\theta}$ then contains real-valued TEQ coefficients and complex-valued FEQ coefficients.

TEQ are given by:

$$\nabla_{\boldsymbol{\theta}} J = \begin{bmatrix} \nabla_{\mathbf{w}} J \\ \nabla_{\tilde{\mathbf{d}}} J \end{bmatrix} = \frac{1}{K} \sum_{k=1}^{K} \check{\mathbf{Y}}_{k,\boldsymbol{\theta}}^H \mathrm{diag}\left(\check{\boldsymbol{\gamma}}_{K,\mathbf{w}_{\mathrm{prev}}}\right) \tilde{\mathbf{e}}_{k,\boldsymbol{\theta}} \quad (45)$$

$$\nabla_{\boldsymbol{\theta}}^2 J = \frac{1}{K} \sum_{k=1}^{K} \check{\mathbf{Y}}_{k,\boldsymbol{\theta}}^H \mathrm{diag}\left(\check{\boldsymbol{\gamma}}_{K,\mathbf{w}_{\mathrm{prev}}}\right) \check{\mathbf{Y}}_{k,\boldsymbol{\theta}} \quad (46)$$

with $\check{\mathbf{Y}}_{k,\boldsymbol{\theta}} = \begin{bmatrix} \mathrm{diag}\left(\tilde{\mathbf{d}}\right) \tilde{\mathbf{Y}}_k & \mathrm{diag}(\tilde{\mathbf{Y}}_k \mathbf{w}) \end{bmatrix}$ and where $\nabla_{\mathbf{w}} J$ and $\nabla_{\tilde{\mathbf{d}}} J$ are $T \times 1$ and $N_a \times 1$ subvectors of $\nabla_{\boldsymbol{\theta}} J$, respectively. Based on (46), we define the following block-based description for the approximate Hessian:

$$\nabla_{\boldsymbol{\theta}}^2 J = \begin{bmatrix} \mathbf{E}_{\tilde{\mathbf{d}}} & \mathbf{F}_{\boldsymbol{\theta}} \\ \mathbf{F}_{\boldsymbol{\theta}}^H & \mathbf{G}_{\mathbf{w},\mathrm{D}} \end{bmatrix} \quad (47)$$

with a diagonal submatrix $\mathbf{G}_{\mathbf{w},\mathrm{D}}$ and

$$\begin{aligned} \mathbf{E}_{\tilde{\mathbf{d}}} &= \frac{1}{K} \sum_{k=1}^{K} \tilde{\mathbf{Y}}_k^H \mathrm{diag}\left(\check{\boldsymbol{\gamma}}_{K,\mathbf{w}_{\mathrm{prev}}} \odot \left|\tilde{\mathbf{d}}\right|^2\right) \tilde{\mathbf{Y}}_k \quad (48) \\ &= \sum_{n \in \mathcal{S}_a} \check{\gamma}_{K,n,\mathbf{w}_{\mathrm{prev}}} \left|\tilde{d}_n\right|^2 \boldsymbol{\Sigma}_{K,n,\tilde{\mathbf{y}}}^2 \end{aligned}$$

$$\begin{aligned} \mathbf{F}_{\boldsymbol{\theta}}^H &= \frac{1}{K} \sum_{k=1}^{K} \mathrm{diag}\left(\tilde{\mathbf{Y}}_k \mathbf{w}\right)^* \mathrm{diag}\left(\check{\boldsymbol{\gamma}}_{K,\mathbf{w}_{\mathrm{prev}}} \odot \tilde{\mathbf{d}}\right) \tilde{\mathbf{Y}}_k \quad (49) \\ &= \begin{bmatrix} \vdots \\ \left[\check{\gamma}_{K,n,\mathbf{w}_{\mathrm{prev}}} \tilde{d}_n \left(\mathbf{w}^H \boldsymbol{\Sigma}_{K,n,\tilde{\mathbf{y}}}^2\right)\right] \\ \vdots \end{bmatrix} \Bigg\updownarrow N_a \text{ rows} \end{aligned}$$

$$\begin{aligned} \mathbf{G}_{\mathbf{w},\mathrm{D}} &= \frac{1}{K} \sum_{k=1}^{K} \mathrm{diag}\left(\check{\boldsymbol{\gamma}}_{K,\mathbf{w}_{\mathrm{prev}}} \odot \left|\tilde{\mathbf{Y}}_k \mathbf{w}\right|^2\right) \quad (50) \\ &= \mathrm{diag}\left(\begin{bmatrix} \cdots & \check{\gamma}_{K,n,\mathbf{w}_{\mathrm{prev}}} \left(\mathbf{w}^H \Sigma_{K,n,\tilde{\mathbf{y}}}^2 \mathbf{w}\right) & \cdots \end{bmatrix}\right) \end{aligned}$$

The inverse of $\left(\nabla_{\boldsymbol{\theta}}^2 J\right)^{-1}$ is then given by [25]:

$$\begin{bmatrix} \mathbf{Q}_{\boldsymbol{\theta}}^{-1} & -\mathbf{Q}_{\boldsymbol{\theta}}^{-1} \mathbf{F}_{\boldsymbol{\theta}} \mathbf{G}_{\mathbf{w},\mathrm{D}}^{-1} \\ -\mathbf{G}_{\mathbf{w},\mathrm{D}}^{-1} \mathbf{F}_{\boldsymbol{\theta}}^H \mathbf{Q}_{\boldsymbol{\theta}}^{-1} & \mathbf{G}_{\mathbf{w},\mathrm{D}}^{-1} \left(\mathbf{I}_{N_a} + \mathbf{F}_{\boldsymbol{\theta}}^H \mathbf{Q}_{\boldsymbol{\theta}}^{-1} \mathbf{F}_{\boldsymbol{\theta}} \mathbf{G}_{\mathbf{w},\mathrm{D}}^{-1}\right) \end{bmatrix} \quad (51)$$

with $\mathbf{Q}_{\boldsymbol{\theta}}$ the so-called Schur complement of $\mathbf{G}_{\mathbf{w},\mathrm{D}}$ [17]:

$$\mathbf{Q}_{\boldsymbol{\theta}} = \mathbf{E}_{\tilde{\mathbf{d}}} - \mathbf{F}_{\boldsymbol{\theta}} \mathbf{G}_{\mathbf{w},\mathrm{D}}^{-1} \mathbf{F}_{\boldsymbol{\theta}}^H \quad (52)$$

Although considering (44), we in fact only use the GN updating of $\mathbf{w}$, as $\tilde{\mathbf{d}}$ is updated in each iteration using (41), i.e., by exploiting the separability property of the SNLLS-based criterion. Moreover, it follows from (41) that $\nabla_{\tilde{\mathbf{d}}_{i-1}} J$ is an all-zero vector in each iteration. For a complex valued TEQ, step 3 in each iteration is then given by:

$$\mathbf{w}_i \leftarrow \mathbf{w}_{i-1} - \alpha_i \mathbf{Q}_{\boldsymbol{\theta}_{i-1}}^{-1} \left(\nabla_{\mathbf{w}_{i-1}} J\right) \quad (53)$$

with

$$\nabla_{\mathbf{w}_{i-1}} J = \sum_{n \in \mathcal{S}_a} \check{\gamma}_{K,n,\mathbf{w}_{\mathrm{prev}}} \tilde{d}_{K,n,i}^* \left(\tilde{d}_{K,n,i} \boldsymbol{\Sigma}_{K,n,\tilde{\mathbf{y}}}^2 \mathbf{w} - \boldsymbol{\sigma}_{K,n,\tilde{x}\tilde{\mathbf{y}}}^H\right) \quad (54)$$

One can show that the updating, in the case of a real-valued TEQ, becomes:

$$\mathbf{w}_i \leftarrow \mathbf{w}_{i-1} - \alpha_i \left(\Re\left\{\mathbf{Q}_{\boldsymbol{\theta}_{i-1}}\right\}\right)^{-1} \Re\left\{\nabla_{\mathbf{w}_{i-1}} J\right\} \quad (55)$$

After each iteration, we normalize the TEQ to remove the scalar parameter ambiguity of (37), namely that if $\boldsymbol{\theta} = \begin{bmatrix} \mathbf{w}^H & \tilde{\mathbf{d}}^H \end{bmatrix}^H$ is a solution of (37), then $\boldsymbol{\theta} = \begin{bmatrix} \eta \mathbf{w}^H & \eta^{-1} \tilde{\mathbf{d}}^H \end{bmatrix}^H$ is also a solution.

The non-convexity of the cost criterion causes local optima, which could correspond to unacceptably low bitrates. Although there is no guarantee that poorly performing local optima will not be encountered in practice, all simulations with the above described algorithm for a set of test channels, noise environments with and without RFI, synchronization delays, equalizer lengths and random initializations $\mathbf{w}_0$ converged to a local optimum with an acceptable bitrate.

The iterative GN algorithm can also be used for optimizing the SNLLS-based FD-MMSE-TEQ design criterion (30) by setting the weights $\check{\gamma}_{K,n,\mathbf{w}_{\mathrm{prev}}}$ equal to 1.

As suggested in [10], the second-order statistics $\boldsymbol{\Sigma}_{K,n,\tilde{\mathbf{y}}}^2$ and $\boldsymbol{\sigma}_{K,n,\tilde{x}\tilde{\mathbf{y}}}$, required for the above outline iterative GN algorithm can be computed, based on a channel and noise estimate or recursively, based on (part of) the training signal with cyclic prefix that is transmitted during connection set-up. Below, an estimate of the dominant computational cost terms (in number of real multiplications per iteration[5]) for the case of a real-valued TEQ is given. The following steps in the algorithm are the most expensive:

- the computation of the complex-valued vectors $\boldsymbol{\Sigma}_{K,n,\tilde{\mathbf{y}}}^2 \mathbf{w}_{i-1}$, $n \in \mathcal{S}_a$, required for the weights $\check{\boldsymbol{\gamma}}_{K,\mathbf{w}_{i-1}}$ as well as for $\mathbf{F}_{\boldsymbol{\theta}_{i-1}}$, $\mathbf{G}_{\mathbf{w}_{i-1},\mathrm{D}}$ and $\nabla_{\mathbf{w}_{i-1}} J$, costs around $\mathcal{O}(2N_a T^2)$ real multiplications;
- the computation of the real part of $\mathbf{E}_{\tilde{\mathbf{d}}_{i-1}}$ in (48), which is required for $\Re\{\mathbf{Q}_{\boldsymbol{\theta}_{i-1}}\}$ in (55) and which is a symmetric matrix (hence roughly half of the elements are to be computed), costs around $\mathcal{O}(N_a T^2/2)$ real multiplications;
- the computation of the real part of $\mathbf{F}_{\boldsymbol{\theta}} \mathbf{G}_{\mathbf{w},\mathrm{D}}^{-1} \mathbf{F}_{\boldsymbol{\theta}}^H$, which is also a symmetric matrix required for $\Re\{\mathbf{Q}_{\boldsymbol{\theta}_{i-1}}\}$, costs around$\mathcal{O}(N_a T^2)$ real multiplications;
- the matrix inversion $\Re\{\mathbf{Q}_{\boldsymbol{\theta}_{i-1}}\}^{-1}$ costs $\mathcal{O}(T^3/6)$ real multiplications.

All other steps (e.g., the computation of the scalars $\mathbf{w}_{i-1}^T \boldsymbol{\sigma}_{K,n,\tilde{x}\tilde{\mathbf{y}}}^H$ and $\mathbf{w}_{i-1}^T (\Sigma_{K,n,\tilde{\mathbf{y}}}^2 \mathbf{w}_{i-1})$, $n \in \mathcal{S}_a$, the weights $\check{\boldsymbol{\gamma}}_{K,\mathbf{w}_{i-1}}$ and the MMSE FEQs (41)) have a complexity of $\mathcal{O}(N_a T)$ or less, hence the overall computational cost per iteration is dominated by $\mathcal{O}(\frac{7}{2} N_a T^2) + \mathcal{O}(\frac{1}{6} T^3)$ real multiplications. From the simulations in Section V, we learn that roughly 30 iterations suffice for convergence. For typical values of N_a and T, e.g., $N_a = 219$ and $T = 32$, one iteration requires 8×10^5 multiplications, so roughly 24×10^6 multiplications are needed until convergence. For comparison, the blind, adaptive channel shortening algorithm presented in [14], which is a stochastic gradient algorithm (hence does not need

[5]A multiplication of two complex-valued numbers is counted as 4 real multiplications and a multiplication of a real-valued and a complex-valued number counts as 2 real multiplications. The number of required additions and multiplications is comparable.

to store second-order statistics), requires only around 25600 multiplications per update (for a channel impulse response length around 200 samples), but convergence is much slower and takes more than 40000 updates.

The connection set-up phase of an ADSL modem typically lasts for several seconds; e.g., the training signal with cyclic prefix is around 16000 DMT symbols or 4 seconds long [1]. Assuming that 1 second is used for the recursive estimation of the second-order statistics $\mathbf{\Sigma}^2_{K,n,\tilde{\mathbf{y}}}$ and $\boldsymbol{\sigma}_{K,n,\tilde{x}\tilde{\mathbf{y}}}$ (computational cost: $\mathcal{O}(N_a T)$ per DMT symbol, see also [23]) and 3 seconds are used for the iterative GN algorithm, the latter requires a reasonable cost of around 5.2×10^6 real multiplications per second. Although in general the large computational cost and the presence of local optima may render a nonlinear optimization algorithm prohibitive for a real-time implementation, the above described iterative GN algorithm for IR-SNLLS-based BM-TEQ design turns out to be a worthwhile approach. Alternative recursive/adaptive approaches are pursued in [23].

IV. Relations between LS-based equalizer design criteria

In this section, the common LS-based framework of Section II is further established. It is shown in Section IV-A that the CLLS-based and SNLLS-based FD-MMSE-TEQ criteria reduce to the PTEQ design criterion when considering a single tone. In Section IV-B, the close relationship between the LS-based formulations of the TD-MMSE-TEQ, FD-MMSE-TEQ, BM-TEQ and PTEQ design criteria is made explicit by tabulating the corresponding generalized eigenvalue (GEV) problems.

A. Relation between FD-MMSE-TEQ designs and PTEQ design

In [16], an alternative DMT equalization scheme was presented, based on the key observation (16) that the equalization and the DFT can be swapped, i.e., that the equalization can be moved behind the demodulating DFT ($\{2\}$ in (16)). This allows to equalize each tone n separately by means of a complex-valued equalizer $\mathbf{w}_n$ that linearly combines the entries of the n-th sliding DFT output $\tilde{\mathbf{y}}_{k,n}$, hence the term *per-tone equalization* (PTEQ). The bitrate performance with a PTEQ is at least as good as — and usually much better than — the bitrate achieved with a TEQ, while the computational complexity during data transmission remains roughly the same. The PTEQ $\mathbf{w}_n$ for a particular tone n is the solution to the following linear **LS-based MMSE-PTEQ criterion**:

$$\min_{\mathbf{w}_n} \frac{1}{K} \sum_{k=1}^{K} \left| \tilde{\mathbf{y}}_{k,n} \mathbf{w}_n - \tilde{x}_{k,n} \right|^2 \tag{56}$$

where the *complex-valued* PTEQ $\mathbf{w}_n$ combines the task of the FEQ and TEQ. The solution to (56) is given by

$$\mathbf{w}_n = \left(\mathbf{\Sigma}^2_{K,n,\tilde{\mathbf{y}}} \right)^{-1} \boldsymbol{\sigma}^H_{K,n,\tilde{x}\tilde{\mathbf{y}}} \tag{57}$$

where $\mathbf{\Sigma}^2_{K,n,\tilde{\mathbf{y}}}$ and $\boldsymbol{\sigma}_{K,n,\tilde{x}\tilde{\mathbf{y}}}$ have been defined in the previous section.

The PTEQ design criterion (56) is closely related to the CLLS-based and SNLLS-based FD-MMSE-TEQ design criteria of the previous section when only a single tone is considered. Criteria (22) and (30) then reduce to

$$\min_{\mathbf{w},\tilde{b}_n} \frac{1}{K} \sum_{k=1}^{K} |\tilde{e}_{k,n}|^2 \;=\; \min_{\mathbf{w},\tilde{b}_n} \frac{1}{K} \sum_{k=1}^{K} \left| \tilde{\mathbf{y}}_{k,n} \mathbf{w} - \tilde{x}_{k,n} \tilde{b}_n \right|^2 \tag{58}$$
$$\text{s.t. } \mathbf{w}^T \mathbf{\Sigma}^2_{K,n,\tilde{\mathbf{y}}} \mathbf{w} = 1 \text{ and real } \mathbf{w}$$

and

$$\min_{\mathbf{w},\tilde{d}_n} \frac{1}{K} \sum_{k=1}^{K} |\tilde{e}_{k,n,\boldsymbol{\theta}_n}|^2 \;=\; \min_{\mathbf{w},\tilde{d}_n} \frac{1}{K} \sum_{k=1}^{K} \left| \tilde{d}_n \tilde{\mathbf{y}}_{k,n} \mathbf{w} - \tilde{x}_{k,n} \right|^2 \tag{59}$$
$$\text{with real } \mathbf{w}$$

By eliminating the linearly appearing LS-based bMMSE FEQ, which follows from (31):

$$\tilde{d}_n = \frac{\mathbf{w}^T \boldsymbol{\sigma}^H_{K,n,\tilde{x}\tilde{\mathbf{y}}}}{\mathbf{w}^T \mathbf{\Sigma}_{K,n,\tilde{\mathbf{y}}} \mathbf{w}} \tag{60}$$

the SNLLS-based criterion (59) is reduced to a nonlinear criterion in $\mathbf{w}$:

$$\min_{\mathbf{w}} \frac{1}{K} \sum_{k=1}^{K} \left| \frac{\mathbf{w}^T \boldsymbol{\sigma}^H_{K,n,\tilde{x}\tilde{\mathbf{y}}}}{\mathbf{w}^T \mathbf{\Sigma}^2_{K,n,\tilde{\mathbf{y}}} \mathbf{w}} \tilde{\mathbf{y}}_{k,n} \mathbf{w} - \tilde{x}_{k,n} \right|^2 \tag{61}$$

$$= \min_{\mathbf{w}} \sigma^2_{K,n,\tilde{\mathbf{x}}} - \frac{\mathbf{w}^T \boldsymbol{\sigma}^H_{K,n,\tilde{x}\tilde{\mathbf{y}}} \boldsymbol{\sigma}_{K,n,\tilde{x}\tilde{\mathbf{y}}} \mathbf{w}}{\mathbf{w}^T \mathbf{\Sigma}^2_{n,\tilde{\mathbf{y}}} \mathbf{w}} \text{ with real } \mathbf{w}$$

$$= \sigma^2_{K,n,\tilde{x}} \left(1 - \max_{\mathbf{w}} \frac{\sigma^{-2}_{K,n,\tilde{\mathbf{x}}} \left(\mathbf{w}^T \Re \left\{ \boldsymbol{\sigma}^H_{K,n,\tilde{x}\tilde{\mathbf{y}}} \boldsymbol{\sigma}_{K,n,\tilde{x}\tilde{\mathbf{y}}} \right\} \mathbf{w} \right)}{\mathbf{w}^T \Re \left\{ \mathbf{\Sigma}^2_{K,n,\tilde{\mathbf{y}}} \right\} \mathbf{w}} \right) \tag{62}$$

The constraint in (61) of a real-valued TEQ is imposed in (62) by inserting the real operators $\Re\{\cdot\}$. The equivalence of (61) and (62) then follows from the Hermitian property of the matrices $\boldsymbol{\sigma}^H_{K,n,\tilde{x}\tilde{\mathbf{y}}} \boldsymbol{\sigma}_{K,n,\tilde{x}\tilde{\mathbf{y}}}$ and $\mathbf{\Sigma}^2_{K,n,\tilde{\mathbf{y}}}$, which ensures that the Rayleigh quotient in (61) is real. Criterion (62) is equivalent to a GEV problem (7) with

$$(\mathbf{B}, \mathbf{A}) = \left(\sigma^{-2}_{K,n,\tilde{\mathbf{x}}} \Re \left\{ \boldsymbol{\sigma}^H_{K,n,\tilde{x}\tilde{\mathbf{y}}} \boldsymbol{\sigma}_{K,n,\tilde{x}\tilde{\mathbf{y}}} \right\}, \Re \left\{ \mathbf{\Sigma}^2_{K,n,\tilde{\mathbf{y}}} \right\} \right) \tag{63}$$

where the $\Re$-operators ensure a real-valued TEQ $\mathbf{w}$. For a complex-valued $\mathbf{w}$, the $\Re$-operators should be dropped and the matrix $\mathbf{B}$ is rank 1 (for a real-valued TEQ, its rank is 2). Then, there is only one strictly positive GEV (the other GEVs are zero), and the corresponding dominant generalized eigenvector $\mathbf{w}$ indeed has the closed form PTEQ solution (57) (up to a scaling), meaning that the complex-valued solution for (59) is the PTEQ (57) (up to a scaling). It also follows straightforwardly from (28) that the CLLS-based FD-MMSE-TEQ design criterion (22) for a single tone n is equivalent to the same GEV problem (63), hence the CLLS-based criterion (58) and the SNLLS-based criterion (59) have the same solution $\mathbf{w}$. The real-valued or complex-valued PTEQ must be multiplied with an estimate of the uMMSE or bMMSE FEQ, (21) or (31). Note that it follows from (57) and (60) that the bMMSE FEQ estimate for a complex-valued PTEQ is equal to one, meaning that a (complex-valued) PTEQ effectively incorporates a bMMSE FEQ estimate.

	$\mathbf{B}$	$\mathbf{A}$
CLLS-based TD-MMSE-TEQ	$\mathbf{\Sigma}_{L,\mathbf{xy}}^T \left(\mathbf{\Sigma}_{L,\mathbf{x}}^2\right)^{-1} \mathbf{\Sigma}_{L,\mathbf{xy}}$	$\mathbf{\Sigma}_{L,\mathbf{y}}^2$
CLLS-based FD-MMSE-TEQ	$\Re\left\{\mathbf{\Sigma}_{K,\tilde{\mathbf{x}}\tilde{\mathbf{Y}}}^H \left(\mathbf{\Sigma}_{K,\tilde{\mathbf{x}},\mathrm{D}}^2\right)^{-1} \mathbf{\Sigma}_{K,\tilde{\mathbf{x}}\tilde{\mathbf{Y}}}\right\} = \sum_{n\in\mathcal{S}_a} \sigma_{K,n,\tilde{\mathbf{x}}}^{-2} \Re\left\{\boldsymbol{\sigma}_{K,n,\tilde{x}\tilde{\mathbf{y}}}^H \boldsymbol{\sigma}_{K,n,\tilde{x}\tilde{\mathbf{y}}}\right\}$	$\Re\left\{\mathbf{\Sigma}_{K,\tilde{\mathbf{Y}}}^2\right\} = \sum_{n\in\mathcal{S}_a} \Re\left\{\mathbf{\Sigma}_{K,n,\tilde{\mathbf{y}}}^2\right\}$
SNLLS-based FD-MMSE-TEQ	$\Re\left\{\sum_{n\in\mathcal{S}_a} \tilde{d}_{K,n,i}^* \boldsymbol{\sigma}_{K,n,\tilde{x}\tilde{\mathbf{y}}}^H\right\} \times \Re\left\{\sum_{n\in\mathcal{S}_a} \tilde{d}_{K,n,i} \boldsymbol{\sigma}_{K,n,\tilde{x}\tilde{\mathbf{y}}}\right\}$	$\Re\left\{\sum_{n\in\mathcal{S}_a} \left\lvert\tilde{d}_{K,n,i}\right\rvert^2 \mathbf{\Sigma}_{K,n,\tilde{\mathbf{y}}}^2\right\}$
IR-SNLLS-based BM-TEQ	$\Re\left\{\sum_{n\in\mathcal{S}_a} \tilde{\gamma}_{K,n,\mathbf{w}_{\mathrm{prev}}} \tilde{d}_{K,n,i}^* \boldsymbol{\sigma}_{K,n,\tilde{x}\tilde{\mathbf{y}}}^H\right\} \times \Re\left\{\sum_{n\in\mathcal{S}_a} \tilde{\gamma}_{K,n,\mathbf{w}_{\mathrm{prev}}} \tilde{d}_{K,n,i} \boldsymbol{\sigma}_{K,n,\tilde{x}\tilde{\mathbf{y}}}\right\}$	$\Re\left\{\sum_{n\in\mathcal{S}_a} \tilde{\gamma}_{K,n,\mathbf{w}_{\mathrm{prev}}} \left\lvert\tilde{d}_{K,n,i}\right\rvert^2 \mathbf{\Sigma}_{K,n,\tilde{\mathbf{y}}}^2\right\}$
PTEQ	$\sigma_{K,n,\tilde{\mathbf{x}}}^{-2} \Re\left\{\boldsymbol{\sigma}_{K,n,\tilde{x}\tilde{\mathbf{y}}}^H \boldsymbol{\sigma}_{K,n,\tilde{x}\tilde{\mathbf{y}}}\right\}$	$\Re\left\{\mathbf{\Sigma}_{K,n,\tilde{\mathbf{y}}}^2\right\}$

TABLE I

REAL-VALUED TEQ AND PTEQ DESIGNS FORMULATED AS A GEV PROBLEM $\mathbf{Bw} = \lambda\mathbf{Aw}$. COMPLEX TEQS OR PTEQS ARE OBTAINED BY OMITTING THE $\Re$-OPERATORS.

B. Relation between the GEV problem formulations

Throughout the text, each LS-based criterion has been shown to be equivalent to a GEV problem (7), with the SNLLS-based criteria giving rise to an iterative sequence of GEV problems. **Table I** summarizes the encountered matrix pairs $(\mathbf{B}, \mathbf{A})$ (for real-valued TEQ and PTEQ designs) and shows that the $\mathbf{A}$ matrices are closely related autocorrelation matrices of the RX signal y_l, while the $\mathbf{B}$ matrices are closely related, possibly low-rank, matrices determined by a crosscorrelation metric between the RX and TX signal y_l and x_l. Complex-valued TEQs or PTEQs are obtained by omitting the $\Re$-operators in Table I.

V. SIMULATIONS

Figure 2 shows bitrate performance plots for the considered equalizer designs with 32 taps (both real-valued and complex-valued TEQs and PTEQs are considered). The bitrate is depicted for 8 downstream CSA loops (tones 38 to 256) with strong front-end filtering to separate up- and downstream transmission (see [10], [23] for details). All simulations use the same synchronization delay Δ, which is determined by the first sample index of the channel impulse response window of $\nu + 1$ samples with maximum energy. As this choice ensures that most of the desired energy is already within this window, this choice typically results in good performance. A delay optimization can further enhance the bitrate; the delay sensitivity of the equalizer designs is investigated below. The bitrate (32) is measured by averaging over 1024 DMT symbols. An SNR gap Γ_n of 12.8 dB is used, based on a target bit error rate of 10^{-7}, a coding gain of 3 dB and a noise margin of 6 dB [22]. The noise in **Figure 2a** is a superposition of AWG noise at -140dBm/Hz, residual echo and near-end crosstalk from 24 ADSL disturbers. In **Figure 2b**, severe RFI (7 RFIs with carrier frequencies 540, 650, 680, 760, 790, 840 and 1080kHz; the first two RFIs have a power of -30dBm, the remaining five have a power of -50dBm) is added. RFI, especially ingress from AM radio stations, can be an important interferer in ADSL. It has been shown in [10], [23] that the BM-TEQ and the PTEQ can effectively mitigate RFI.

The last bar in both plots shows the theoretical maximum achievable bitrate, i.e., the bitrate that is achieved with perfect equalization [12]; in the case of RFI, it is assumed that the theoretical maximum achievable bitrate is obtained when all tones are free of RFI, except the two neighbouring tones of an RFI. It is clear from Figure 2b that in this RFI case, the IR-SNLLS-based BM-TEQ and PTEQ (bars 7 to 10) outperform the suboptimal TEQ designs (bars 1 to 6) and approach the theoretical maximum bitrate (bar 11). The sub-optimal TEQ designs do not guarantee that a complex-valued equalizer performs better than a real-valued equalizer. The SNLLS-based FD-MMSE-TEQ (bars 5 and 6) consistently outperforms the CLLS-based FD-MMSE-TEQ design (bars 3 and 4) and closely approaches the IR-SNLLS-based BM-TEQ performance (bars 7 en 8); apparently, it makes more sense to minimize the sum of FEQ output energies than the sum of DFT output energies. The impact of dropping the constraints on $\tilde{\mathbf{b}}$ in (18) on the performance of the CLLS-based FD-MMSE-TEQ design (22) appears significant: with these constraints on $\tilde{\mathbf{b}}$, the CLLS-based TD-MMSE-TEQ (3) (first bar) and FD-MMSE-TEQ (17-18) (second bar) achieve similar performance on most loops; without these constraints (bar 3 and 4), the CLLS-based FD-MMSE-TEQ (22) design fails. The SNLLS-based FD-MMSE-TEQ design and IR-SNLLS-based BM-TEQ design have been computed using the iterative GN design algorithm suggested in Section III, initialized with $\mathbf{w}_0 = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T$. Note that, despite the non-convexity of the SNLLS-based FD-MMSE-TEQ and IR-SNLLS-based BM-TEQ design criteria, the iterative GN algorithm ends up in a well-performing local optimum. Simulations with other initializations resulted in very similar performance.

Figure 2 assumes that all equalizer designs are based on the same synchronization delay Δ, which is determined based on the highest energy window of the channel impulse response. A delay optimization can further enhance the performance. Therefore, we depict in **Figure 3** the bitrate performance of several real-valued TEQ designs as a function of Δ for the CSA4 loop (without RFI). As is expected, the performance of

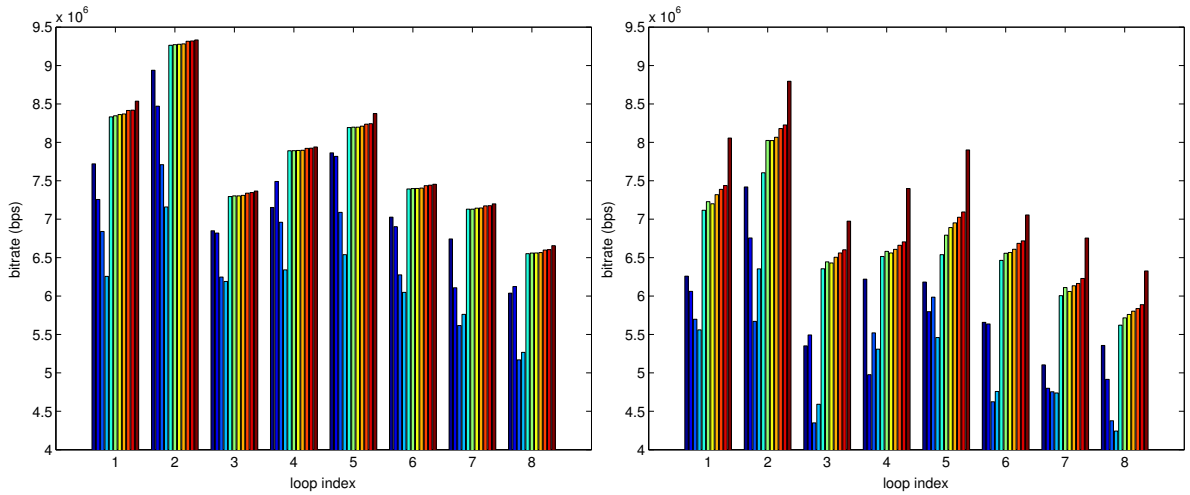


Fig. 2. Bitrate performance of the considered TEQ and PTEQ designs for 8 CSA loops. From left to right: (real-valued) CLLS-based TD-MMSE-TEQ, real-valued CLLS-based FD-MMSE-TEQ with constraints on $\tilde{\mathbf{b}}$ (17-18), real-valued and complex-valued CLLS-based FD-MMSE-TEQ (22), real-valued and complex-valued SNLLS-based FD-MMSE-TEQ, real-valued and complex-valued IR-SNLLS-based BM-TEQ, real-valued and complex-valued PTEQ, theoretical maximum achievable bitrate. (a) Without RFI. (b) With RFI.

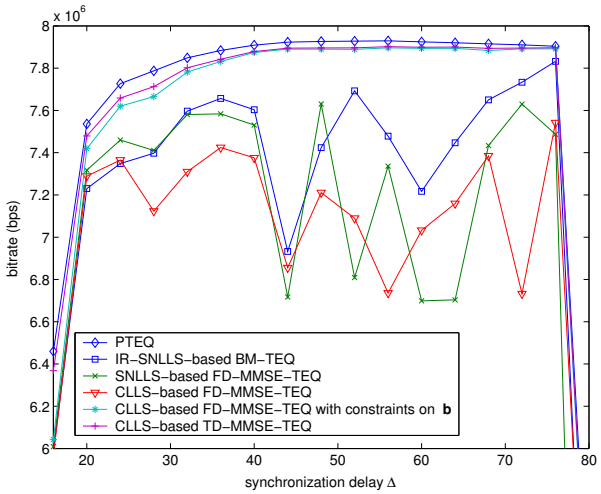


Fig. 3. Bitrate performance of several real-valued TEQ designs and the PTEQ design as a function of the synchronization delay Δ for the CSA4 loop (without RFI). The highest energy window of the channel impulse response starts at $\Delta = 46$.

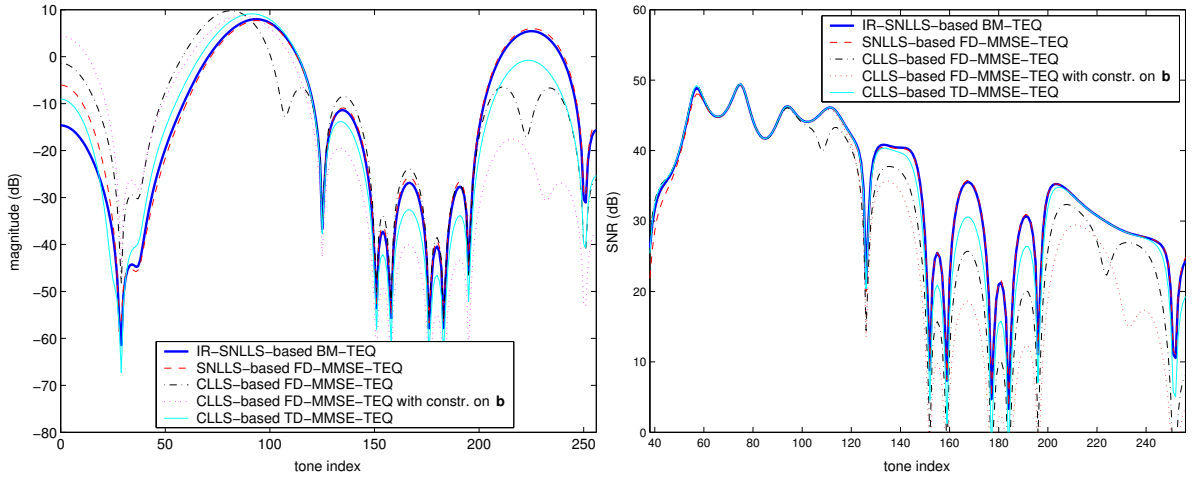
the bitrate maximizing approaches (BM-TEQ and PTEQ) is smooth as a function of Δ and close to optimal performance is obtained for $\Delta = 46$, which is the first sample of the highest energy window of the channel impulse response. The other TEQ designs have a non-smooth behaviour as a function of Δ.

Figure 4 shows the amplitude of the TEQ frequency response as well as the corresponding SNR per tone for the (real-valued) CLLS-based TD-MMSE-TEQ design, the CLLS-based FD-MMSE-TEQ designs (with and without constraints on $\tilde{\mathbf{b}}$), the SNLLS-based FD-MMSE-TEQ design and the IR-SNLLS-based BM-TEQ design for the CSA4 loop in the case of RFI and $T = 32$. All TEQ designs put clear notches on tones that coincide with the RFIs. The spectra and the SNR per tone of the IR-SNLLS-based BM-TEQ and SNLLS-based FD-MMSE-TEQ, which achieve similar bitrates, coincide over the entire downstream band; the TD-MMSE-TEQ comes quite close, except between the tones 150 and 195 where RFI dominates. The CLLS-based FD-MMSE-TEQ designs, which achieve a lower bitrate, have a considerably different spectrum, especially above tone 125: apparently, the RFIs are not sufficiently notched out so that a lower SNR per tone is obtained above tone 125.

Figure 5 illustrates the fast convergence of the iterative GN-based algorithm for designing the SNLLS-based FD-MMSE-TEQ and the IR-SNLLS-based BM-TEQ. The CSA7 loop with RFI and a complex TEQ are given as an example. Two different initializations are depicted: initialization from scratch with $\mathbf{w}_0 = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T$ and initialization with the CLLS-based TD-MMSE-TEQ design. Initialization from scratch requires only 30 iterations, while convergence with a CLLS-based TD-MMSE-TEQ requires less than 10 iterations for both TEQ designs.

VI. Conclusions

In this paper, we have reconsidered the TD-MMSE-TEQ, BM-TEQ and PTEQ design for DMT equalization. Although these designs are at first sight completely different in nature, we have shown that they can all be formulated either as a CLLS-based or an SNLLS-based design. This has lead to an LS-based TEQ design framework: an IR-SNLLS-based BM-TEQ design criterion, which is equivalent to the original BM-TEQ design criterion, has been presented. New CLLS-based and SNLLS-based FD-MMSE-TEQ designs allowed to link



(a) TEQ frequency response

(b) SNR per tone

Fig. 4. TEQ frequency responses and SNR per tone for the CSA4 loop in the case of RFI ($T = 32$).

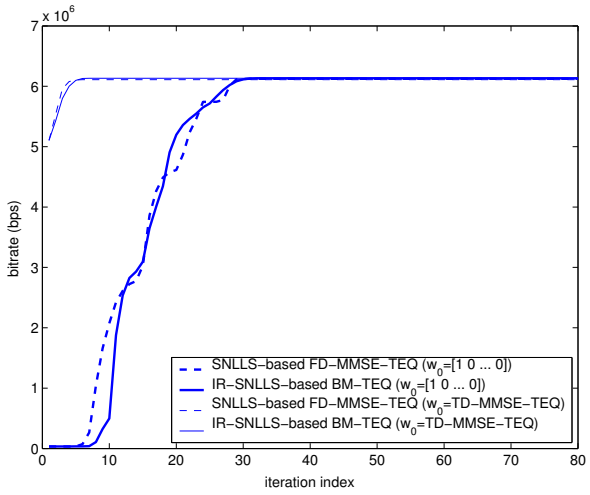


Fig. 5. Convergence of the iterative Gauss-Newton-based algorithm for the SNLLS-based FD-MMSE-TEQ design and the IR-SNLLS-based BM-TEQ design with 2 different initializations (from scratch, i.e., $\mathbf{w}_0 = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T$ and from a CLLS-based TD-MMSE-TEQ design). CSA7 loop with RFI; $T = 32$.

the TD-MMSE-TEQ design to the BM-TEQ and PTEQ design. Moreover, the IR-SNLLS-based formulation of the BM-TEQ design has lead to a fast-converging iterative GN-based BM-TEQ design algorithm with reasonable computational cost. Finally, all considered equalizer designs have been shown to be closely related, when turning their LS-based formulations into GEV-based descriptions.

ACKNOWLEDGEMENT

The authors wish to thank the anonymous reviewers for their valuable comments and suggestions.

REFERENCES

[1] International Telecommunications Union (ITU), "Recommendation G.992.1. Asymmetric digital subscriber line (ADSL) transceivers," July 1999.

[2] The European Telecomm. Standards Inst. (ETSI), "Transmission and multiplexing (TM); Access transmission systems on metallic access cables; Very high speed digital subscriber line (VDSL). TS 101 270 v1.3.1." July 2003.

[3] T. I. of Electrical and E. Engineers, "Wireless LAN medium access control (MAC) and physical layer (PHY) specifications. IEEE std. 802.11a," 1999.

[4] The European Telecomm. Standards Inst. (ETSI), "Broadband radio access networks (BRAN); HIPERLAN type 2; Physical (PHY) layer. TS 101 475 v1.3.1." Dec. 2001.

[5] D. D. Falconer and F. R. Magee, "Adaptive channel memory truncation for maximum likelihood sequence estimation," *Bell Sys. Tech. Journal*, pp. 1541–1562, Nov. 1973.

[6] J. S. Chow and J. M. Cioffi, "A cost-effective maximum likelihood receiver for multicarrier systems," in *Proc. IEEE Int. Conf. Communications (ICC)*, vol. 2, June 1992, pp. 948–952.

[7] N. Al-Dhahir and J. M. Cioffi, "Efficiently computed reduced-parameter input-aided MMSE equalizers for ML detection: A unified approach," *IEEE Trans. Information Theory*, vol. 42, no. 3, pp. 903–915, May 1996.

[8] G. Ysebaert, K. Van Acker, M. Moonen, and B. De Moor, "Constraints in channel shortening equalizer design for DMT-based systems," *Signal Processing*, vol. 83, no. 3, pp. 641–648, Mar. 2003.

[9] D. Daly, C. Heneghan, and A. D. Fagan, "Minimum mean-squared error impulse response shortening for discrete multitone transceivers," *IEEE Trans. Signal Processing*, vol. 52, no. 1, pp. 301–306, Jan. 2004.

[10] K. Vanbleu, G. Ysebaert, G. Cuypers, M. Moonen, and K. Van Acker, "Bitrate maximizing time-domain equalizer design for DMT-based systems," *IEEE Trans. Communications*, vol. 52, no. 6, pp. 871–876, June 2004.

[11] N. Al-Dhahir and J. M. Cioffi, "Optimum finite-length equalization for multicarrier transceivers," *IEEE Trans. Communications*, vol. 44, no. 1, pp. 56–64, Jan. 1996.

[12] G. Arslan, B. L. Evans, and S. Kiaei, "Equalization for discrete multitone transceivers to maximize bit rate," *IEEE Trans. Signal Processing*, vol. 49, no. 12, pp. 3123–3135, Dec. 2001.

[13] R. K. Martin, J. Balakrishnan, W. A. Sethares, and C. R. Johnson, Jr., "A blind, adaptive TEQ for multicarrier systems," *IEEE Signal Processing Lett.*, vol. 9, no. 11, pp. 341–343, Nov. 2002.

[14] J. Balakrishnan, R. K. Martin, and C. R. Johnson, Jr., "Blind, adaptive channel shortening by sum-squared auto-correlation minimization (SAM)," *IEEE Trans. Signal Processing*, vol. 51, no. 12, pp. 3086–3093, Dec. 2003.

[15] P. J. W. Melsa, R. C. Younce, and C. E. Rohrs, "Impulse response shortening for discrete multitone transceivers," *IEEE Trans. Communications*, vol. 44, no. 12, pp. 1662–1672, Dec. 1996.

[16] K. Van Acker, G. Leus, M. Moonen, O. van de Wiel, and T. Pollet, "Per-tone equalization for DMT-based systems," *IEEE Trans. Communications*, vol. 49, no. 1, pp. 109–119, Jan. 2001.

[17] G. H. Golub and C. F. Van Loan, *Matrix Computations - Third Edition*. The John Hopkins University Press, 1996.

[18] K. Van Acker, G. Leus, M. Moonen, S. Claes, and O. van de Wiel, "An improved optimization algorithm for time domain equalizer design in ADSL modems," in *Proc. Intern. Workshop on Copper Wire Access Systems "Bridging the Last Copper Drop" (CWAS)*, Budapest, Hungary, 1997.

[19] T. Starr, J. M. Cioffi, and P. J. Silvermann, *Understanding Digital Subscriber Line Technology*. Englewood Cliffs, NJ: Prentice Hall, 1999.

[20] A. Bjorck, *Least-Squares Methods*. Elsevier, 1987.

[21] G. H. Golub and V. Pereyra, "Separable nonlinear least squares: the variable projection method and its applications," *Institute of Physics, Inverse Problems*, vol. 19, no. 2, pp. R1–R26, Apr. 2003.

[22] J. M. Cioffi, "A multicarrier primer," ANSI Doc., T1E1.4 Tech. Subcommittee, no. 91-157, 1991.

[23] K. Vanbleu, G. Ysebaert, G. Cuypers, and M. Moonen, "Adaptive bitrate maximizing time-domain equalization design for DMT-based systems," ESAT, K.U.Leuven, Belgium, Tech. Rep. ESAT-SISTA/TR 2003-168. *Submitted to IEEE Trans. Signal Processing*. Available at ftp://ftp.esat.kuleuven.ac.be/pub/sista/vanbleu/reports/03-168.pdf, May 2004.

[24] R. H. Byrd and D. A. Payne, "Convergence of the iteratively reweighted least squares algorithm for robust regression," The Johns Hopkins University, Baltimore, MD, Tech. Rep. 313, June 1979.

[25] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Prentice Hall, 2000.

APPENDIX

In this appendix, we derive the **IR-SNLLS-based bitrate maximizing FD-MMSE-TEQ design criterion** for a complex-valued TEQ, based on the constrained BM-TEQ optimization problem (32-33), which is repeated here for convenience:

$$\max_{\boldsymbol{\theta}} \sum_{n\in\mathcal{S}_a} \log_2\left(1+\frac{\mathrm{SNR}_{n,\boldsymbol{\theta}_n}}{\Gamma_n}\right) \tag{64}$$

$$\text{with } \mathrm{SNR}_{n,\boldsymbol{\theta}_n} = \frac{\sigma^2_{n,\tilde{x}}}{\mathcal{E}\left\{|\tilde{e}_{k,n,\boldsymbol{\theta}_n}|^2\right\}} = \frac{\sigma^2_{n,\tilde{x}}}{\mathcal{E}\left\{\left|\tilde{d}_n^{\mathrm{uMMSE}}\tilde{\mathbf{y}}_{k,n}\mathbf{w}-\tilde{x}_{k,n}\right|^2\right\}}$$

$$\text{subject to } \tilde{d}_n^{\mathrm{uMMSE}} = \frac{\sigma^2_{n,\tilde{x}}}{\mathcal{E}\left\{\tilde{y}^*_{k,n,\mathbf{w}}\tilde{x}_{k,n}\right\}} = \frac{\sigma^2_{n,\tilde{x}}}{\boldsymbol{\sigma}_{n,\tilde{x}\tilde{\mathbf{y}}}\mathbf{w}},\ \forall n\in\mathcal{S}_a \tag{65}$$

A constrained optimization problem is typically restated as the minimization of a Lagrangian cost function:

$$\begin{aligned} J_{\mathrm{L}}(\boldsymbol{\theta},\boldsymbol{\lambda}) &= \sum_{n\in\mathcal{S}_a}\left[\log\left(\left(1+\frac{\mathrm{SNR}_{n,\boldsymbol{\theta}_n}}{\Gamma_n}\right)^{-1}\right)\right. \\ &\quad \left.-\Re\left\{\lambda_n^*\left(\tilde{d}_n^{\mathrm{uMMSE}}\mathcal{E}\left\{\tilde{x}^*_{k,n}\tilde{y}_{k,n,\mathbf{w}}\right\}-\sigma^2_{n,\tilde{x}}\right)\right\}\right] \end{aligned} \tag{66}$$

where N_a Lagrange parameters λ_n have been introduced and where the real operator $\Re\{\cdot\}$ ensures a real constraint term. Setting the gradient of J_{L} w.r.t. $\tilde{d}_n = \tilde{d}_n^{\mathrm{uMMSE}}$, denoted as $\nabla_{\tilde{d}_n} J_{\mathrm{L}} = \left(\frac{\partial}{\partial \tilde{d}_n^*} J_{\mathrm{L}}\right)$[6], equal to zero:

$$\begin{aligned} \nabla_{\tilde{d}_n} J_{\mathrm{L}} &= \underbrace{\frac{\mathrm{SNR}^2_{n,\boldsymbol{\theta}_n}}{\sigma^2_{n,\tilde{x}}(\mathrm{SNR}_{n,\boldsymbol{\theta}_n}+\Gamma_n)}}_{\gamma_{n,\boldsymbol{\theta}_n}} \mathcal{E}\left\{\tilde{y}^*_{k,n,\mathbf{w}}\tilde{e}_{k,n,\boldsymbol{\theta}_n}\right\} \\ &\quad -\lambda_n\mathcal{E}\left\{\tilde{y}^*_{k,n,\mathbf{w}}\tilde{x}_{k,n}\right\}=0 \end{aligned} \tag{67}$$

[6]The partial derivative of $\tilde{d}_n$ and its complex conjugate w.r.t. itself are defined as $\frac{\partial}{\partial \tilde{d}_n}\tilde{d}_n = 1$ and $\frac{\partial}{\partial \tilde{d}_n}\tilde{d}_n^* = 0$, respectively.

and using the uMMSE FEQ constraint (65), equation (67) results in an expression for λ_n:

$$\begin{aligned} \lambda_n &= \gamma_{n,\boldsymbol{\theta}_n}\frac{\mathcal{E}\left\{\tilde{y}^*_{k,n,\mathbf{w}}\tilde{e}_{k,n,\boldsymbol{\theta}_n}\right\}}{\mathcal{E}\left\{\tilde{y}^*_{k,n,\mathbf{w}}\tilde{x}_{k,n}\right\}} \\ &= \gamma_{n,\boldsymbol{\theta}_n}\frac{\tilde{d}_n\mathcal{E}\left\{|\tilde{y}_{k,n,\mathbf{w}}|^2\right\}-\mathcal{E}\left\{\tilde{y}^*_{k,n,\mathbf{w}}\tilde{x}_{k,n}\right\}}{\mathcal{E}\left\{\tilde{y}^*_{k,n,\mathbf{w}}\tilde{x}_{k,n}\right\}} \\ &= \gamma_{n,\boldsymbol{\theta}_n}\left(\frac{\sigma^2_{n,\tilde{x}}\mathcal{E}\left\{|\tilde{y}_{k,n,\mathbf{w}}|^2\right\}}{\left|\mathcal{E}\left\{\tilde{y}^*_{k,n,\mathbf{w}}\tilde{x}_{k,n}\right\}\right|^2}-1\right) \\ &= \gamma_{n,\boldsymbol{\theta}_n}\left(\rho^{-2}_{n,\mathbf{w}}-1\right)=\frac{\gamma_{n,\boldsymbol{\theta}_n}}{\mathrm{SNR}_{n,\mathbf{w}}} \end{aligned} \tag{68}$$

The last two equations follow from the definition of the square normalized crosscorrelation between the n-th DFT output $\tilde{y}_{k,n,\mathbf{w}}$ and the n-th TX symbol $\tilde{x}_{k,n}$ [10]:

$$\rho^2_{n,\mathbf{w}} = \frac{\left|\mathcal{E}\left\{\tilde{y}^*_{k,n,\mathbf{w}}\tilde{x}_{k,n}\right\}\right|^2}{\sigma^2_{n,\tilde{x}}\mathcal{E}\left\{|\tilde{y}_{k,n,\mathbf{w}}|^2\right\}} = \frac{\tilde{d}_n^{\mathrm{bMMSE}}}{\tilde{d}_n^{\mathrm{uMMSE}}} \tag{69}$$

and the subchannel SNR, which was shown in [10] to be a function of $\rho^2_{n,\mathbf{w}}$:

$$\mathrm{SNR}_{n,\mathbf{w}} = \frac{1}{\rho^{-2}_{n,\mathbf{w}}-1} \tag{70}$$

The gradient of J_{L} w.r.t. a complex-valued TEQ $\mathbf{w}$ is given by:

$$\begin{aligned} \nabla_{\mathbf{w}} J_{\mathrm{L}} &= \sum_{n\in\mathcal{S}_a}\gamma_{n,\boldsymbol{\theta}_n}\tilde{d}_n^{\mathrm{uMMSE}*}\mathcal{E}\left\{\tilde{\mathbf{y}}^H_{k,n}\tilde{e}_{k,n,\boldsymbol{\theta}_n}\right\} \\ &\quad -\lambda_n\tilde{d}_n^{\mathrm{uMMSE}*}\mathcal{E}\left\{\tilde{\mathbf{y}}^H_{k,n}\tilde{x}_{k,n}\right\} \end{aligned} \tag{71}$$

In case of a real-valued TEQ, $2\Re\{\nabla_{\mathbf{w}} J_{\mathrm{L}}\}$ should be used. Combining the gradient w.r.t. $\mathbf{w}$ (71) and $\tilde{\mathbf{d}}$ (67) and filling in the Lagrange parameter λ_n (68), we obtain, in matrix notation:

$$\begin{aligned} \nabla_{\boldsymbol{\theta}} J_L &= \mathcal{E}\left\{\begin{bmatrix}\tilde{\mathbf{Y}}^H_k\mathrm{diag}\left(\tilde{\mathbf{d}}^{\mathrm{uMMSE}}\right)^*\\ \mathrm{diag}(\tilde{\mathbf{Y}}_k\mathbf{w})^*\end{bmatrix}\mathrm{diag}\left(\boldsymbol{\gamma}_{\mathbf{w}}\right)\right. \\ &\quad \left.\left[\mathrm{diag}\left(\tilde{\mathbf{d}}^{\mathrm{uMMSE}}\right)\tilde{\mathbf{Y}}_k\mathbf{w}-\mathrm{diag}(\boldsymbol{\rho}^{-2}_{\mathbf{w}})\tilde{\mathbf{x}}_k\right]\right\} \end{aligned} \tag{72}$$

The entries of the $N_a\times 1$ vector $\boldsymbol{\gamma}_{\mathbf{w}}$ are $\gamma_{n,\mathbf{w}} = \gamma_{n,\boldsymbol{\theta}_n}$, defined in (67), and are a function of the subchannel SNR (70). The vector $\boldsymbol{\rho}^{-2}_{\mathbf{w}}$ is short-hand notation for a vector whose entries are the inverses of $\rho^2_{n,\mathbf{w}}$. Making use of (69) and (70), the gradient (72) can be expressed in terms of bMMSE FEQs instead of uMMSE FEQs:

$$\begin{aligned} \nabla_{\boldsymbol{\theta}} J_L &= \mathcal{E}\left\{\begin{bmatrix}\tilde{\mathbf{Y}}^H_k\mathrm{diag}\left(\tilde{\mathbf{d}}^{\mathrm{bMMSE}}\right)^*\\ \mathrm{diag}(\tilde{\mathbf{Y}}_k\mathbf{w})^*\end{bmatrix}\mathrm{diag}\left(\check{\boldsymbol{\gamma}}_{\mathbf{w}}\right)\right. \\ &\quad \left.\left[\mathrm{diag}\left(\tilde{\mathbf{d}}^{\mathrm{bMMSE}}\right)\tilde{\mathbf{Y}}_k\mathbf{w}-\tilde{\mathbf{x}}_k\right]\right\} \end{aligned} \tag{73}$$

$$= \mathcal{E}\left\{\check{\mathbf{Y}}^H_{k,\boldsymbol{\theta}}\mathrm{diag}\left(\check{\boldsymbol{\gamma}}_{\mathbf{w}}\right)\tilde{\mathbf{e}}_{k,\boldsymbol{\theta}}\right\} \tag{74}$$

with $\check{\gamma}_{n,\mathbf{w}} = \frac{(\mathrm{SNR}_{n,\mathbf{w}}+1)^2}{\sigma^2_{n,\tilde{x}}(\mathrm{SNR}_{n,\mathbf{w}}+\Gamma_n)}$. Replacing the expectation operator in (74) and (69) with a time-average over K DMT

symbols results in a generalized version (with weighting $\boldsymbol{\gamma}_{\check{K},\mathbf{w}}$ and multiple-error vector $\tilde{\mathbf{e}}_{k,\boldsymbol{\theta}}$) of the orthogonality condition that is typically encountered when solving LS criteria:

$$\frac{1}{K}\sum_{k=1}^{K}\check{\mathbf{Y}}_{k,\boldsymbol{\theta}}^{H}\mathrm{diag}\left(\check{\boldsymbol{\gamma}}_{K,\mathbf{w}}\right)\tilde{\mathbf{e}}_{k,\boldsymbol{\theta}}=\mathbf{0} \tag{75}$$

More specifically, it corresponds to the orthogonality condition, encountered in so-called iteratively-reweighted (IR) LS criteria. After replacing the weights $\check{\boldsymbol{\gamma}}_{K,\mathbf{w}}$ with $\check{\boldsymbol{\gamma}}_{K,\mathbf{w}_{\mathrm{prev}}}$, i.e., with weights $\check{\boldsymbol{\gamma}}_{K,\mathbf{w}_{\mathrm{prev}}}$ based on estimates $\mathrm{SNR}_{K,n,\mathbf{w}_{\mathrm{prev}}}$ and $\rho^2_{K,n,\mathbf{w}_{\mathrm{prev}}}$ (over K symbols) that are evaluated in the parameter estimate $\mathbf{w}_{\mathrm{prev}}$ of the previous iteration, this orthogonality condition gives rise to the following so-called iteratively-reweighted (IR) LS-based BM-TEQ design criterion criterion:

$$\min_{\boldsymbol{\theta}}\frac{1}{K}\sum_{k=1}^{K}\left\|\mathrm{diag}\left(\sqrt{\check{\boldsymbol{\gamma}}_{K,\mathbf{w}_{\mathrm{prev}}}}\right)\tilde{\mathbf{e}}_{k,\boldsymbol{\theta}}\right\|^2 \tag{76}$$

This criterion can be shown to be also valid in case of real-valued TEQs.