Robust Design of Adaptive Equalizers

Markus Rupp, *Senior Member, IEEE,*

Abstract

Although equalizers promise to improve the signal to noise energy ratio, zero forcing equalizers are derived classically in a deterministic setting minimizing intersymbol interference, while minimum mean square error equalizer solutions are derived in a stochastic context based on quadratic Wiener cost functions. In this contribution we show that it is possible –and in our opinion even simpler– to derive the classical results in a purely deterministic setup, interpreting both equalizer types as least squares solutions. This in turn allows the introduction of a simple linear reference model for equalizers, which supports the exact derivation of a family of iterative and recursive algorithms with robust behavior. The framework applies equally to multi-user transmissions and multiple-input multiple-output channels. A major contribution is that due to the reference approach the adaptive equalizer problem can equivalently be treated as an adaptive system identification problem for which very precise statements are possible with respect to convergence, robustness and l_2-stability. Robust adaptive equalizers are much more desirable as they guarantee a much stronger form of stability than conventional in the mean square sense convergence. Even some blind channel estimation schemes can now be included in the form of recursive algorithms and treated under this general framework.

Index Terms

linear equalizers, least mean squares, minimum mean square error, zero forcing, reference modeling, blind channel estimation, recursive adaptive filters, iterative adaptive filters, convergence, robustness, l_2-stability

I. Introduction

Modern digital receivers in wireless and cable-based systems are unimaginable without equalizers in some form. The earliest mentions of digital equalizers were probably made in 1961 by E.Gibson [1, 2] and in 1964 by M.Rappeport [3]; they became popular due to the milestone work by R.W.Lucky [4, 5] in 1965 and in 1966, published in the Bell System Technical Journal, who also coined the expression "Zero Forcing" (ZF). In 1967 M.E.Austin introduced the Decision Feedback Equalizer (DFE) concept [6] that was soon to be thoroughly

This work has been funded by the NFN SISE project S10611-N13 (National Research Network Signal and Information Processing in Science and Engineering).

The author is with the Institute of Telecommunications at the Vienna University of Technology.

analyzed [7, 8]. A.Gersho proposed the Minimum Mean Square Error (MMSE) equalizer in 1969 [9] and D.M.Brady introduced Fractionally Spaced Equalizers (FSE) in 1970 [10], which inspired a large amount of significant work [11–13]. In 1972 G.Forney included the Maximum Likelihood principle [14]. In the 80s overview papers by S.U.H.Qureshi were popular [15, 16] and the major driving forces in research became blind equalization methods, introduced by Y.Sato in 1975 [17] and D.N.Godard [18]. A first overview paper on blind methods was by A.Benveniste and M.Goursat [19], although the breakthrough in this field came later in 1991 by L.Tang, G.Xu, and T.Kailath [20, 21]. Further papers of significance are by J.M.Cioffi et al. [22, 23] as well as J.Treichler [24]. Extensions for Multiple-Input Multiple-Output (MIMO) transmissions start with A.Duel-Hallen [25] in 1992, followed by many others, of which we cite just a few selected papers [26–30]. More recent overview papers are provided in [31–35].

In all the above literature, linear equalizers were derived either on the deterministic ZF approach for equalizers of ZF type or on the stochastic MMSE approach for equalizers of the MMSE type. We present a new formulation of the equalizer problem based on a weighted Least Squares (LS) approach. This deterministic view is very much in line with Lucky's original formulation, leaving out all signal and noise properties (up to the noise variance) but at the same time offers new insight in the equalizer solutions, as they share common LS orthogonality properties. This novel LS interpretation allows a very general formulation to cover a multitude of equalizer problems, including different channel models, multiple antennas as well as multiple users.

In practice the equalizer problem is not yet solved once the solution is known, as it typically involves a matrix inversion, a mathematical operation that is highly complex and challenging in low-cost fixed-point devices. Adaptive algorithms are thus commonly employed to approximate the results. Such adaptive algorithms for equalization purposes come in two flavors, an iterative (also off-line or batch process) approach as well as a recursive approach (also on-line or data-driven process) that readjusts its estimates on each new data element that is being observed. Both approaches have their benefits and drawbacks. If channel estimation has been performed in a previous step (for various reasons), then the iterative algorithm based on the channel's impulse response may be most effective. On the other hand, it is not required to compute first the channel's impulse response if only the equalizer solution is of interest. In particular in time-variant scenario's one may not have the chance to continuously estimate the channel and compute equalizer solutions iteratively and therefore, a recursive solution that is able to track changes, may be the only hope for good results.

However, such adaptive algorithms require a deep understanding of their properties as selecting their only free parameter, the step-size, turns out to be crucial. While for decades adaptive filter designers were highly satisfied when they could prove convergence in the mean-square sense, more and more situations now become known, in which this approach has proved to be insufficient, since despite the convergence in the mean square sense, worst case sequences exist that cause the algorithm to diverge. This observation started with Feintuch's adaptive IIR algorithm [36, 37] and the class of adaptive filters with a linear filter in the error path [38, 39] but has recently arisen in other adaptive filters [40, 41] as well as in adaptive equalizers [42]. A robust algorithm design [43, 44], on the other hand, is much more suited to solving the equalization problem as it can guarantee the adaptive algorithm will not diverge in every case. In this contribution we therefore show how to design robust, adaptive filters for linear equalizers.

A. Notation and Organization

Table I provides an overview of the most common variables and their dimensions. The paper is organized as follows. In Section II common transmission models in wireless systems are introduced along with their optimal linear equalizer solutions. In Section III we propose a least squares linear reference model for ZF and a weighted LS model for MMSE that allows the interpretation of equalizers as a standard system identification problem. In Section IV we discuss iterative forms (i.e., algorithms that contain all required data and iterate to find an optimal solution) and in Section V recursive forms (i.e., algorithms that take on new data for each adaptation step) of adaptive equalizers, as well as blind channel estimators for equalization, some of them well known in the literature, others entirely new. Due to the novel reference approach, we are able to design them as robust algorithms and provide convergence conditions. In Section VI we present selected Matlab experiments to validate our findings. Finally, Section VII concludes the paper.

II. Transmission Models and their Optimal Linear Equalizers

Throughout this paper we assume that the individual transmit signal elements s_k have unit energy, $\mathrm{E}[|s_k|^2] = 1$, and the noise variance at the receiver is given by $\mathrm{E}[|v_k|^2] = N_o$. We are considering several similar but distinct transmission schemes:

1) **Single-User transmission over frequency selective SISO channels**

The following Single-User (SU) transmission describes frequency selective (also called time dispersive) Single-

Variable	Dimension	Comment
$\mathbf{s}_k$	$\mathbb{C}^{SN_TM\times 1}$	transmit symbol vector at time instant k
$\mathbf{s}_{k,m}$	$\mathbb{C}^{SN_T\times 1}$	transmit symbol vector of user m
$\mathbf{s}_{k,t,m}$	$\mathbb{C}^{S\times 1}$	transmit symbol vector from antenna t of user m
$\mathbf{r}_k$	$\mathbb{C}^{RN_R\times 1}$	receive symbol vector at time instant k
$\mathbf{H}$	$\mathbb{C}^{RN_R\times SN_TM}$	compound channel matrix
$\mathbf{H}_m$	$\mathbb{C}^{RN_R\times SN_T}$	compound channel matrix of user m
$\mathbf{H}_{t,m}$	$\mathbb{C}^{RN_R\times S}$	compound channel matrix from transmit antenna t of user m
$\mathbf{e}_\tau$	$\mathbb{R}^{S\times 1}$	unit vector, "1" at delay lag τ
$\mathbf{e}_t$	$\mathbb{R}^{N_T\times 1}$	unit vector, "1" at transmit antenna t
$\mathbf{e}_{t,m}$	$\mathbb{R}^{SN_T\times 1}$	unit vector, "1" at transmit antenna t of user m
$\mathbf{e}_{\tau,t,m}$	$\mathbb{R}^{S\times 1}$	unit vector, "1" at position τ, t, m
$\mathbf{f}^{\text{ZF}}_{\tau,t,m}$	$\mathbb{C}^{RN_R\times 1}$	ZF LS solution
$\mathbf{f}^{\text{MMSE}}_{\tau,t,m}$	$\mathbb{C}^{RN_R\times 1}$	MMSE LS solution
$\hat{\mathbf{f}}_k$	$\mathbb{C}^{RN_R\times 1}$	equalizer estimate
$\tilde{\mathbf{f}}_k$	$\mathbb{C}^{RN_R\times 1}$	equalizer parameter estimation error
$\mathbf{v}^{\text{ZF}}_{\tau,t,m}$	$\mathbb{C}^{SN_TM\times 1}$	ZF modelling error
$\mathbf{v}^{\text{MMSE}}_{\tau,t,m}$	$\mathbb{C}^{SN_TM\times 1}$	MMSE modelling error
$\tilde{v}^{\text{ZF}}_{k,t,m}$	$\mathbb{C}$	ZF compound noise
$\tilde{v}^{\text{MMSE}}_{k,t,m}$	$\mathbb{C}$	MMSE compound noise

TABLE I

OVERVIEW OF MOST COMMON VARIABLES.

Input Single-Output (SISO) transmission scenarios:

$$\mathbf{r}_k = \mathbf{H}\mathbf{s}_k + \mathbf{v}_k. \tag{1}$$

Here, the vector $\mathbf{s}_k = [s_k, s_{k-1}, ..., s_{k-S+1}]^T$ consists of the current and $S-1$ past symbols according to the span $L < S$ of the channel $\mathbf{H}$, which is typically of Toeplitz form as shown in (2). The received vector is defined as $\mathbf{r}_k = [r_k, r_{k-1}, ..., r_{k-R+1}]^T$. Let the transmission be disturbed by additive noise $\mathbf{v}_k$, being of the same dimension as $\mathbf{r}_k$.

$$\begin{bmatrix} r_k \\ r_{k-1} \\ \vdots \\ r_{k-R+1} \end{bmatrix} = \begin{bmatrix} h_0 & h_1 & \dots & h_{L-1} & & & \\ & \ddots & \ddots & & \ddots & & \\ & & h_0 & h_1 & & \dots & h_{L-1} \end{bmatrix} \begin{bmatrix} s_k \\ s_{k-1} \\ \vdots \\ s_{k-S+1} \end{bmatrix} + \begin{bmatrix} v_k \\ v_{k-1} \\ \vdots \\ v_{k-R+1} \end{bmatrix}. \tag{2}$$

Note that for a Toeplitz form channel $\mathbf{H} \in \mathbb{C}^{R\times S}$ we have $R \leq S$. A linear equalizer applies an FIR filter $\mathbf{f}$

on the received signal $\mathbf{r}_k$ so that $\mathbf{f}^H\mathbf{r}_k$ is an estimate of $s_{k-\tau} = \mathbf{e}_\tau^T\mathbf{s}_k$ for some delayed version of s_k. A unit vector $\mathbf{e}_\tau = [0, ..., 0, 1, 0, ..0]^T$ with a single one at position $\tau \in \{0, 1, ..., S-1\}$ facilitates the description.

2) **Single-User transmission over frequency flat MIMO channels**

The transmissions follow the same form as described in (1), although with a different meaning as we transmit over N_T antennas and receive by N_R. Such Multiple-Input Multiple-Output systems are generally referred to as MIMO systems. The transmit vector $\mathbf{s}_k = [s_{1,k}, s_{2,k}, ..., s_{N_T,k}]^T$ is of dimension $1 \times N_T$, the channel matrix $\mathbf{H} \in \mathbb{C}^{N_R \times N_T}$, and thus the receive vector and the additive noise vector are of dimension $1 \times N_R$. Here, N_T is the number of transmit antennas, whilst N_R denotes the number of receive antennas. As in the previous case, we assume each entry of the transmit vector $\mathbf{s}_k$ to have unit power. Unlike the earlier situation, however, we have to distinguish $N_R > N_T$ (underdetermined LS solution) and $N_R < N_T$ (overdetermined LS solution). For $N_R = N_T$ both solutions coincide. In order to detect the various entries of the transmit vector $\mathbf{s}_k$, we again employ a unit vector $\mathbf{e}_t$: $\mathbf{e}_t^T\mathbf{s}_k = s_{t,k}$. Note however that in contrast to the previous channel model, a set of N_T different vectors $\mathbf{e}_t; t = 1, 2, .., N_T$ will be employed in order to select all N_T transmitted symbols while in the frequency selective SISO case a single vector $\mathbf{e}_\tau$ is sufficient. Early work on linear MIMO equalization can be found in [29, 35]. Note that precoding matrices are often applied in particular in modern cellular systems such as HSDPA and LTE. In this case the concatenation of the precoding matrix and the wireless channel has to be considered as a new compound channel. Such precoding matrices can also have an impact on the dimension of the transmit vector $\mathbf{s}_k$ as in many cases fewer symbols than rank($\mathbf{H}$) are transmitted at each time instant k. A particular form of this is given when the precoding matrix shrinks to a vector, in which case we talk about beamforming where only one symbol stream is transmitted.

3) **Multi-User transmission over flat frequency MIMO channels**

In a Multi-User (MU) scenario the received vector captures all M incoming user signals

$$\mathbf{r}_k = \sum_{m=1}^{M} \mathbf{H}_m\mathbf{s}_{k,m} + \mathbf{v}_k = \mathbf{H}\mathbf{s}_k + \mathbf{v}_k. \tag{3}$$

The channel matrices $\mathbf{H}_m; m = 1, 2, .., M$ hereby denote matrices of dimension $N_R \times N_T$, describing flat frequency MIMO transmissions. As before a unit vector is required to select the desired transmit symbol but now we also have to select the user whose symbols we want to detect, thus a set of vectors $\mathbf{e}_{t,m}; t = 1, 2, .., N_T$ is required for each user $m = 1, 2, .., M$: $s_{k,t,m} = \mathbf{e}_{t,m}^T\mathbf{s}_k$, where $\mathbf{s}_k$ denotes a concatenated vector over all M transmit vectors $\mathbf{s}_{k,m}$ and the compound channel matrix $\mathbf{H} = [\mathbf{H}_1, \mathbf{H}_2, ..., \mathbf{H}_M]$.

4) **Multi-User transmission over arbitrary channels**

The previous concepts can be combined to model more complicated scenarios. Consider for example MU transmissions over frequency selective MIMO channels as the most elaborate case. To simplify matters we assume that all corresponding SISO channels require a memory of L taps and that all transmitters are comprised of N_T transmit antennas. A large compound channel matrix $\mathbf{H}$ can now be defined as

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{1,1,1} & \mathbf{H}_{2,1,1} & \dots & \mathbf{H}_{N_T,1,1} & \mathbf{H}_{1,1,2} & \dots & \mathbf{H}_{N_T,1,M} \\ \mathbf{H}_{1,2,1} & \mathbf{H}_{2,2,1} & \dots & \mathbf{H}_{N_T,2,1} & \mathbf{H}_{1,2,2} & \dots & \mathbf{H}_{N_T,2,M} \\ & \vdots & & & & & \\ \mathbf{H}_{1,N_R,1} & \mathbf{H}_{2,N_R,1} & \dots & \mathbf{H}_{N_T,N_R,1} & \mathbf{H}_{1,N_R,2} & \dots & \mathbf{H}_{N_T,N_R,M} \end{bmatrix}, \tag{4}$$

where the indexes in $\mathbf{H}_{t,r,m}$ indicate the frequency selective SISO channel (of Toeplitz structure) of user m from transmit antenna t to receive antenna r. Consequently, the compound channel matrix $\mathbf{H} \in \mathbb{C}^{RN_R \times SN_T M}$. The transmission model for a single receive antenna r simply reads

$$\mathbf{r}_{k,r} = \sum_{t=1}^{N_T} \sum_{m=1}^{M} \mathbf{H}_{t,r,m} \mathbf{s}_{k,t,m} + \mathbf{v}_k, \tag{5}$$

$\mathbf{s}_k$ being a concatenated transmit vector, comprised of all (SISO) transmit vectors $\mathbf{s}_{k,t,m}$ for each transmit antenna t and user m at time instant k. Concatenating all N_R receive vectors in a single vector $\mathbf{r}_k$, we obtain the general transmission model $\mathbf{r}_k = \mathbf{H}\mathbf{s}_k + \mathbf{v}_k$, the channel being the compound matrix as defined above. As before we can envision a unit vector $\mathbf{e}_{\tau,t,m}; \tau \in \{0, 1, ..., S-1\}$ that selects the transmit symbol at delay lag τ, originating from transmit antenna $t; t = 1, 2, .., N_T$ of user $m; m = 1, 2, .., M$: $s_{k-\tau,t,m} = \mathbf{e}_{\tau,t,m}^T \mathbf{s}_k$.

Although an SIR or SINR formulation is very intuitive for designing transmission systems, equalizers are usually not derived in the context of maximizing SINR but in minimizing a mean square error (referred to as Minimum MSE=MMSE) instead. The relation between MMSE equalizers and maximizing SINR is well known, see for example [45, 46] and is being used successfully in recent research [47, 48]. We briefly re-derive it here in a consistent form to show relations to SIR as well as to modern MU equalizers.

A. *Maximizing SIR and SINR*

To understand the vast amount of research and information available on this subject, one has to ask the question "What is the purpose of an equalizer?". While Lucky's original work focused on the SU scenario, attempting a minimax approach to combat Inter-Symbol-Interference (ISI), today we typically view the equalizer in terms of Signal-to-Interference Ratio (SIR) or Signal-to-Interference and Noise Ratio (SINR). If a signal, say s_k, is transmitted through a frequency selective channel, a mixture of ISI, additive noise and signals from other users, Multi-User Interference (MUI), is received. If signals are transmitted by multiple antennas, then additional so-called spatial ISI (SP-ISI) is introduced. The ratio of the received signal power P_s and all disturbance terms before an equalizer (indicated by the index 'be') is easily described as:

$$\mathrm{SINR}_{\mathrm{be}} = \frac{P_s}{P_{\mathrm{ISI}} + P_{\mathrm{SP-ISI}} + P_{\mathrm{MUI}} + N_o}. \tag{6}$$

The task of the equalizer is to improve the situation, i.e., to increase this ratio. A linear filter applied to the observed signal can for example result in an increased $P_s' > P_s$, utilizing useful parts of P_{ISI} and $P_{\mathrm{SP-ISI}}$, while the remaining $P_{\mathrm{ISI}}' < P_{\mathrm{ISI}}$ and/or $P_{\mathrm{SP-ISI}}' < P_{\mathrm{SP-ISI}}$ and/or $P_{\mathrm{MUI}}' < P_{\mathrm{MUI}}$ is decreased. Unfortunately, the noise power N_0 as well as its power spectral density is in general also changed when an equalizer filter is applied. At best it can be considered possible to achieve the post-equalization SINR (the index 'ae' denotes after equalization):

$$\mathrm{SINR}_{\mathrm{ae}} = \max_{\text{Equalizer strategy}} \mathrm{SINR}_{\mathrm{be}} = \frac{P_s'}{P_{\mathrm{ISI}}' + P_{\mathrm{SP-ISI}}' + P_{\mathrm{MUI}}' + N_o'} \le \frac{P_s + P_{\mathrm{ISI}} + P_{\mathrm{SP-ISI}}}{N_o}, \tag{7}$$

where the equalizer manages to treat the ISI and SP-ISI as useful signal whilst at the same time eliminating the MUI (for example by successive interference cancellation). The ratio of $\mathrm{SINR}_{\mathrm{be}}$ to the eventually achieved $\mathrm{SINR}_{\mathrm{ae}}$ is considered as the equalizer gain. The purpose of this paper is to develop a unified view on the SINR and SNR relation to the MMSE and ZF equalizer, which permits the introduction of a simple linear reference model as well as its application in an adaptive system identification framework. In correspondence with the four transmission scenarios described above, there also exist four related linear equalizers that will be discussed below.

1) **Single-User equalizer on frequency selective SISO channels**

We start with the simplest case: a single user scenario. Now consider the original equalizer problem of maximizing the signal to interference ratio in the absence of noise in (6). By applying a linear filter $\mathbf{f}$ we obtain one part that is proportional to the signal of interest, $\mathbf{e}_\tau^T \mathbf{s}_k = s_{k-\tau}$ and a remaining ISI part:

$$\max_{\mathbf{f}} \mathrm{SIR} = \max_{\mathbf{f}} \frac{P_{\mathrm{S}}}{P_{\mathrm{I}}} = \max_{\mathbf{f}} \frac{\left\|\mathbf{e}_\tau^T \mathbf{H}^H \mathbf{f}\right\|^2}{\mathbf{f}^H \mathbf{H}\mathbf{H}^H \mathbf{f} - \left\|\mathbf{e}_\tau^T \mathbf{H}^H \mathbf{f}\right\|_2^2} = \max_{\mathbf{f}} \frac{\mathbf{f}^H \mathbf{H} \mathbf{e}_\tau \mathbf{e}_\tau^T \mathbf{H}^H \mathbf{f}}{\mathbf{f}^H \mathbf{H}\left(\mathbf{I} - \mathbf{e}_\tau \mathbf{e}_\tau^T\right)\mathbf{H}^H \mathbf{f}} \tag{8}$$

or, alternatively if the noise is also included to maximize the signal to interference-plus-noise ratio

$$\max_{\mathbf{f}} \mathrm{SINR} = \max_{\mathbf{f}} \frac{P_{\mathrm{S}}}{P_{\mathrm{I}} + N_0} = \max_{\mathbf{f}} \frac{\mathbf{f}^H \mathbf{H} \mathbf{e}_\tau \mathbf{e}_\tau^T \mathbf{H}^H \mathbf{f}}{\mathbf{f}^H \left[\mathbf{H}\left(\mathbf{I} - \mathbf{e}_\tau \mathbf{e}_\tau^T\right)\mathbf{H}^H + N_0 \mathbf{I}\right]\mathbf{f}}. \tag{9}$$

Here, P_S and P_{I} denote the signal and interference power (ISI only in the case of for SISO), respectively, while $\mathbf{I}$ denotes the identity matrix. The introduction of our unit vector $\mathbf{e}_\tau$ now takes on a new interpretation. Not only does it select the symbol we wish to detect from the (concatenated) transmit vector $\mathbf{e}_\tau^T \mathbf{s}_k = s_{k-\tau}$, but it also selects the sub-channels $\mathbf{H}\mathbf{e}_\tau$ in $\mathbf{H}$ that carry relevant information for the equalizer. Note that both terms, SIR as well as SINR, can uniquely be mapped into the following problem

$$\max_{\mathbf{f}} \frac{P_{\mathrm{S}}}{P_{\mathrm{S}}+P_{\mathrm{I}}} = \max_{\mathbf{f}} \frac{\mathrm{SIR}}{\mathrm{SIR}+1} = \gamma^{\mathrm{SIR}} \leq 1 \tag{10}$$

and correspondingly

$$\max_{\mathbf{f}} \frac{P_{\mathrm{S}}}{P_{\mathrm{S}}+P_{\mathrm{I}}+N_o} = \max_{\mathbf{f}} \frac{\mathrm{SINR}}{\mathrm{SINR}+1} = \gamma^{\mathrm{SINR}} \leq 1. \tag{11}$$

As both provide a monotone mapping from SIR to γ^{SIR} or SINR to γ^{SINR}, we can equivalently maximize these new terms and obtain (here shown on the example of SINR):

$$\max_{\mathbf{f}} \gamma^{\mathrm{SINR}} = \max_{\mathbf{f}} \frac{\mathbf{f}^H \mathbf{H}\mathbf{e}_\tau \mathbf{e}_\tau^T \mathbf{H}^H \mathbf{f}}{\mathbf{f}^H \left[\mathbf{H}\mathbf{H}^H + N_o \mathbf{I}\right] \mathbf{f}} \tag{12}$$

the solution of which is given by the eigenvector to the largest eigenvalue of the numerator term $\mathbf{H}\mathbf{e}_\tau \mathbf{e}_\tau^T \mathbf{H}^H$:

$$\mathbf{f}_\tau^{\max \mathrm{SINR}} = \alpha(\mathbf{H}\mathbf{H}^H + N_o \mathbf{I})^{-1}\mathbf{H}\mathbf{e}_\tau = \alpha \mathbf{f}_\tau^{\mathrm{MMSE}}, \tag{13}$$

which is obviously proportional to the MMSE solution. More precisely: any vector $\mathbf{f}^{\max \mathrm{SINR}}$ from (13) maximizes the SINR but only one solution of this is also the MMSE solution, thus a better way to put it is that *the MMSE solution also maximizes SINR* (but not the opposite).

Correspondingly, maximizing γ^{SIR} we find:

$$\max_{\mathbf{f}} \gamma^{\mathrm{SIR}} = \max_{\mathbf{f}} \frac{\mathbf{f}^H \mathbf{H}\mathbf{e}_\tau \mathbf{e}_\tau^T \mathbf{H}^H \mathbf{f}}{\mathbf{f}^H \mathbf{H}\mathbf{H}^H \mathbf{f}} \tag{14}$$

with the well-known (overdetermined) ZF solution (see (27) below)

$$\mathbf{f}_\tau^{\max \mathrm{SIR}} = \alpha(\mathbf{H}\mathbf{H}^H)^{-1}\mathbf{H}\mathbf{e}_\tau = \alpha \mathbf{f}_\tau^{\mathrm{ZF,o}}. \tag{15}$$

Analog to the MMSE case, we can now state that any vector $\mathbf{f}^{\max \mathrm{SIR}}$ maximizes the SIR but only one solution of this is identical to the ZF solution.

As the solution is known explicitly we find $\gamma_{\max,\tau}^{\mathrm{SIR}} = \mathbf{e}_\tau^T \mathbf{H}^H [\mathbf{H}\mathbf{H}^H]^{-1}\mathbf{H}\mathbf{e}_\tau \leq 1$ and thus the right means to select the optimal value of τ in order to maximize

$$\tau_{\mathrm{opt}} = \arg\max_\tau \gamma_{\max,\tau}^{\mathrm{SIR}}.$$

Similarly, for the MMSE solution we maximize

$$\gamma_{\max,\tau}^{\mathrm{SINR}} = \mathbf{e}_\tau^T \mathbf{H}^H [\mathbf{H}\mathbf{H}^H + N_o \mathbf{I}]^{-1}\mathbf{H}\mathbf{e}_\tau. \tag{16}$$

2) **Single-User equalizer on frequency flat MIMO channels**

The ZF and MMSE equalizer for this transmission scenario can be derived straightforwardly following the previous concept. We thus briefly present the results. For the overdetermined case ($N_T > N_R$) the ZF solution for the t-th entry of transmit vector $\mathbf{s}_k$ is obtained for

$$\mathbf{f}_t^{\max \text{SIR,o}} = \alpha(\mathbf{H}\mathbf{H}^H)^{-1}\mathbf{H}\mathbf{e}_t = \alpha \mathbf{f}_t^{\text{ZF,o}}; t = 1, 2, .., N_T. \tag{17}$$

while the underdetermined case ($N_R > N_T$) reads

$$\mathbf{f}_t^{\max \text{SIR,u}} = \alpha\mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{e}_t = \alpha \mathbf{f}_t^{\text{ZF,u}}; t = 1, 2, .., N_T. \tag{18}$$

In the following we indicate underdetermined solutions by an additional index 'u' and overdetermined solutions by 'o'. We omit the index if the context makes it clear what is meant or if they coincide, as in the case of the MMSE equalizer.

The MMSE equalizer solution for this transmission mode is given by

$$\mathbf{f}_t^{\max \text{SINR}} = \alpha(\mathbf{H}\mathbf{H}^H + N_o\mathbf{I})^{-1}\mathbf{H}\mathbf{e}_t = \alpha \mathbf{f}_t^{\text{MMSE}}; t = 1, 2, .., N_T. \tag{19}$$

Formally the solutions appear in the same form as before but note that in the SISO channel, τ could be selected to maximize SINR, while now we have no such liberty and have to accept the SINR that comes out for each value of t.

3) **Multi-User equalizer on frequency flat MIMO channels**

We now show that this method can easily be extended to an MU scenario, where we have M users transmitting simultaneously, as explained in (3). The linear equalizer that maximizes the SINR of the signal from transmit antenna t of user m is defined by

$$\max_{\mathbf{f}} \gamma_{t,m}^{\text{SINR}} = \max_{\mathbf{f}} \frac{\mathbf{f}^H\mathbf{H}_m\mathbf{e}_{t,m}\mathbf{e}_{t,m}^T\mathbf{H}_m^H\mathbf{f}}{\mathbf{f}^H\left[\sum_{m=1}^M \mathbf{H}_m\mathbf{H}_m^H + N_o\mathbf{I}\right]\mathbf{f}} \tag{20}$$

with the solution

$$\mathbf{f}_{t,m}^{\max \text{SINR}} = \alpha\left(\sum_{m=1}^M \mathbf{H}_m\mathbf{H}_m^H + N_o\mathbf{I}\right)^{-1}\mathbf{H}_m\mathbf{e}_{t,m} = \alpha \mathbf{f}_{t,m}^{\text{MU}-\text{MMSE}}, \tag{21}$$

which is proportional to the solution of the MMSE MU equalizer. Knowing the individual channels $\mathbf{H}_m; m = 1, ..., M$ involved, it is thus straightforward to find the linear solution to optimize the SINR in an MU scenario. The solution is typically overdetermined as long as $N_R < MN_T$. The corresponding ZF solution, maximizing SIR, is obtained by omitting N_o. Correspondingly, an MU setup that maximizes SIR can be computed simply by setting N_o to zero in the above solution. Note further that similar concepts occur for equalizers in CDMA systems, where due to the signatures of the various users a rich interference is introduced by the design [49][Chapter 6.2].

4) **Multi-User equalizer (interference aware equalizer) on arbitrary channels**

The linear equalizer that maximizes the SINR of the signal at time lag τ from transmit antenna t of user m, received over N_R antennas is defined by

$$\max_{\mathbf{f}} \gamma_{\tau,t,m}^{\text{SINR}} = \max_{\mathbf{f}} \frac{\mathbf{f}^H \mathbf{H}_{t,m} \mathbf{e}_{\tau,t,m} \mathbf{e}_{\tau,t,m}^T \mathbf{H}_{t,m}^H \mathbf{f}}{\mathbf{f}^H \left[\sum_{t'=1}^{N_T} \sum_{m=1}^{M} \mathbf{H}_{t',m} \mathbf{H}_{t',m}^H + N_o \mathbf{I} \right] \mathbf{f}} \tag{22}$$

with the compound matrix

$$\mathbf{H}_{t,m} = \left[\mathbf{H}_{t,1,m}, \mathbf{H}_{t,2,m}, ..., \mathbf{H}_{t,N_R,m} \right]^T. \tag{23}$$

We find the solution

$$\mathbf{f}_{\tau,t,m}^{\max \text{SINR}} = \alpha \left(\sum_{t'=1}^{N_T} \sum_{m=1}^{M} \mathbf{H}_{t',m} \mathbf{H}_{t',m}^H + N_o \mathbf{I} \right)^{-1} \mathbf{H}_{t,m} \mathbf{e}_{\tau,t,m} = \alpha \mathbf{f}_{\tau,t,m}^{\text{MU-MMSE}}. \tag{24}$$

In the case of the ZF solution we have to be careful. Depending on the actual numbers of antennas and users, both situations can appear: if $SN_T M > RN_R$ we have the overdetermined case, i.e., we simply omit $N_o \mathbf{I}$ in (24), while for $SN_T M < RN_R$ the underdetermined solution has to be taken:

$$\mathbf{f}_{\tau,t,m}^{\max \text{SIR,u}} = \alpha \mathbf{H}_{t,m} \left(\sum_{t'=1}^{N_T} \sum_{m=1}^{M} \mathbf{H}_{t',m}^H \mathbf{H}_{t',m} \right)^{-1} \mathbf{e}_{\tau,t,m} = \alpha \mathbf{f}_{\tau,t,m}^{\text{MU-ZF,u}}. \tag{25}$$

For the MMSE solution both are again identical. Such equalizers have also been referred to as interference aware equalizers in the context of UMTS-HSDPA transmissions [50–52], as they also take other users into account.

For all such equalizer solutions (MMSE and ZF) corresponding views exist, interpreting the equalizer as the linear solution that maximizes SINR or SIR. As long as we have a general transmission model $\mathbf{r}_k = \mathbf{H}\mathbf{s}_k + \mathbf{v}_k$, we always find that the MMSE solution also maximizes SINR, or alternatively, the ZF solution maximizes SIR.

III. A REFERENCE MODEL FOR EQUALIZATION

While classical literature views the equalizer problem as minimizing a mean square error, we show in the following section that this is in fact not required and a purely deterministic approach based on a least squares modeling is possible. This approach in turn leads to the novel interpretation of the adaptive equalizer problem in terms of a classic system identification problem. For such problems, however, a much stronger l_2-stability and robustness has been derived in the past to ensure convergence of the adaptive algorithms under worst case conditions. In order to apply such robust techniques, we first have to show the equivalent system identification approach for equalizers. We start with the ZF equalizer and then continue with its MMSE counterpart.

A. ZF Equalizer

A solution to the ZF equalizer problem is equivalently given by the following LS formulation

$$\mathbf{f}_{\tau,t,m}^{\mathrm{ZF}} = \arg\min_{\mathbf{f}} \left\| \mathbf{H}^H \mathbf{f} - \mathbf{e}_{\tau,t,m} \right\|_2^2 = \arg\min_{\mathbf{f}} \left\| \mathbf{H}^H \left[\mathbf{f} - (\mathbf{H}\mathbf{H}^H)^{-1} \mathbf{H} \mathbf{e}_{\tau,t,m} \right] \right\|_2^2 \tag{26}$$

with $\mathbf{e}_{\tau,t,m}$ indicating a unit vector with a single one entry at position τ, for transmit antenna t of user m, thus $\mathbf{e}_{\tau,t,m}^T \mathbf{s}_k = s_{k-\tau,t,m}$, the transmit signal at antenna t of user m that will be decoded at delay lag τ. Note that this form of derivation requires no signal or noise information, focusing instead only on properties of linear time-invariant systems of finite length (FIR); it thus entirely ignores the presence of noise. This is identical to Lucky's original formulations [4], where system properties were the focus and the particular case of $N_T = 1, M = 1$ was considered.

If $RN_R < SN_T M$ (for example in Lucky's SISO frequency selective scenario, we have $R < S$) the solution to this problem is obviously given by

$$\mathbf{f}_{\tau,t,m}^{\mathrm{ZF,o}} = (\mathbf{H}\mathbf{H}^H)^{-1} \mathbf{H} \mathbf{e}_{\tau,t,m}, \tag{27}$$

commonly known as the ZF solution. Note that this is a so-called overdetermined LS solution as we have more equations than entries in $\mathbf{f}_{\tau,l,m}^{\mathrm{ZF}}$. When $RN_R > SN_T M$ an alternative so-called underdetermined LS solution exists, as long as $\mathrm{rank}(\mathbf{H}) = SN_T M$:

$$\mathbf{f}_{\tau,l,m}^{\mathrm{ZF,u}} = \mathbf{H}(\mathbf{H}^H \mathbf{H})^{-1} \mathbf{e}_{\tau,t,m} \tag{28}$$

and requires independent consideration as will be provided further on in this section.

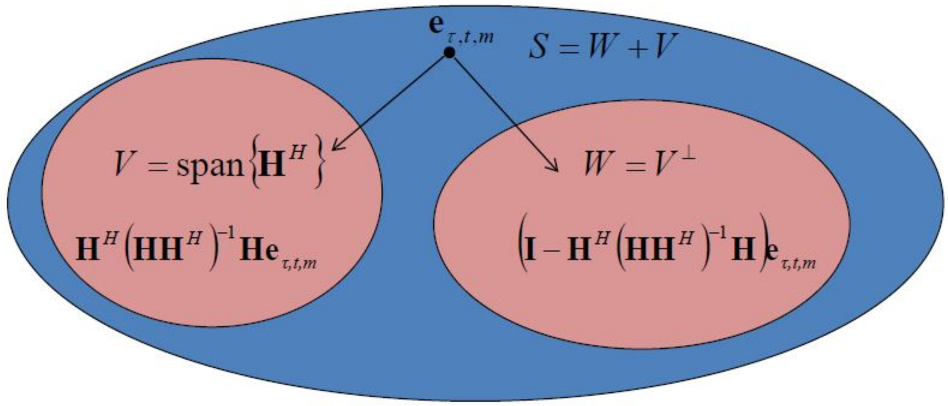
Let us first consider the overdetermined case of Eqn. (27). As ISI does not vanish for finite length vectors, we propose the following reference model for ZF equalizers

$$\mathbf{e}_{\tau,t,m} = \mathbf{H}^H \mathbf{f}_{\tau,t,m}^{\mathrm{ZF,o}} + \mathbf{v}_{\tau,t,m}^{\mathrm{ZF,o}} \tag{29}$$

with the modeling error vector

$$\mathbf{v}_{\tau,t,m}^{\mathrm{ZF,o}} = \left(\mathbf{I} - \mathbf{H}^H (\mathbf{H}\mathbf{H}^H)^{-1} \mathbf{H} \right) \mathbf{e}_{\tau,t,m}. \tag{30}$$

The term $\mathbf{v}_{\tau,t,m}^{\mathrm{ZF,o}}$ models ISI, SP-ISI and MUI. The larger the equalizer length RN_R, the smaller the ISI, e.g., measured in $\|\mathbf{v}_{\tau,t,m}^{\mathrm{ZF,o}}\|_2^2$. The cursor position τ also influences the result.



$$\begin{aligned}\mathbf{e}_{\tau,t,m} &= \mathbf{H}^H\left(\mathbf{H}\mathbf{H}^H\right)^{-1}\mathbf{H}\mathbf{e}_{\tau,t,m} + \left(\mathbf{I} - \mathbf{H}^H\left(\mathbf{H}\mathbf{H}^H\right)^{-1}\mathbf{H}\right)\mathbf{e}_{\tau,t,m} \\ &= \mathbf{H}^H\mathbf{f}^{ZF,o}_{\tau,t,m} \qquad\qquad + \qquad \mathbf{v}^{ZF,o}_{\tau,t,m}\end{aligned}$$

Fig. 1. *The overdetermined ZF equalizer can be interpreted as a linear combination from two complementary subspaces, the equalizer in the range of* $\mathbf{H}^H$ *and the modeling error from its orthogonal complement.*

Due to its projection properties we find that the outcome of the reference model lies in the range of $\mathbf{H}^H$ with an additive term $\mathbf{v}^{\mathrm{ZF,o}}_{\tau,t,m}$ from its orthogonal complement:

$$\mathbf{H}\mathbf{v}^{\mathrm{ZF,o}}_{\tau,t,m} = \mathbf{0}, \tag{31}$$

$$\mathbf{f}^{\mathrm{ZF,o}\,H}_{\tau,t,m}\mathbf{H}\mathbf{v}^{\mathrm{ZF,o}}_{\tau,t,m} = 0, \tag{32}$$

$$\mathbf{v}^{\mathrm{ZF,o}\,H}_{\tau,t,m}\mathbf{v}^{\mathrm{ZF,o}}_{\tau,t,m} = \mathbf{v}^{\mathrm{ZF,o}\,H}_{\tau,t,m}\mathbf{e}_{\tau,t,m} = \mathbf{e}^T_{\tau,t,m}\left(\mathbf{I} - \mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1}\mathbf{H}\right)\mathbf{e}_{\tau,t,m}. \tag{33}$$

This last term can be interpreted as the energy of the modeling error but equally describes the remaining ISI power $P'_{\mathrm{ISI}} + P'_{\mathrm{SP-ISI}} + P'_{\mathrm{MUI}}$. Figure 1 illustrates the ZF equalizer reference model.

What changes for the underdetermined solution, i.e., if $RN_R > SN_TM$? Such situations may occur if we simply have two ($N_R = 2$) sensors (antennas) for observations on an SU SIMO channel, resulting in a stacked matrix $\mathbf{H}^H = [\mathbf{H}_1^H, \mathbf{H}_2^H]$. In this case we find that

$$\mathbf{e}_{\tau,t,m} = \mathbf{H}^H\mathbf{f}^{\mathrm{ZF,u}}_{\tau,t,m}, \tag{34}$$

i.e., the equalizer removes all ISI, without introducing any additional modeling noise, thus $\mathbf{v}^{\mathrm{ZF,u}}_{\tau,t,m} = \mathbf{0}$. Such situations typically occur in SU situations ($M = 1$), either in frequency selective SIMO transmissions when the number of observations $R > S$ or in MIMO transmissions if $N_R > N_T$. In such cases the remaining ISI becomes zero.

How does a received signal look after such ZF-equalization? We apply $\mathbf{f}_{\tau,t,m}^{\mathrm{ZF,o}}$ on the observation vector and obtain

$$\mathbf{f}_{\tau,t,m}^{\mathrm{ZF,o}\,H}\mathbf{r}_k \;\;=\;\; s_{k-\tau,t,m} - \mathbf{v}_{\tau,t,m}^{\mathrm{ZF,o}\,H}\mathbf{s}_k + \mathbf{f}^{\mathrm{ZF,o}\,H}\mathbf{v}_k = s_{k-\tau,t,m} + \tilde{v}_{k,t,m}^{\mathrm{ZF,o}}. \tag{35}$$

Let us now consider the underdetermined case, i.e., $RN_R > SN_TM$. We then find that the ISI components become zero, i.e., $\mathbf{v}_{\tau,t,m}^{\mathrm{ZF,u}} = \mathbf{0}$. The outcome of the ZF equalizer now reads

$$\mathbf{f}_{\tau,t,m}^{\mathrm{ZF,u}\,H}\mathbf{r}_k \;\;=\;\; s_{k-\tau,t,m} + \mathbf{f}^{\mathrm{ZF,u}\,H}\mathbf{v}_k. \tag{36}$$

While in the overdetermined case some ISI remains, the underdetermined case only displays the additive noise term. Both relations serve as ZF reference model, as we apply them to adaptive algorithms further ahead. Note that both ISI and additive noise are often treated as a compound noise $\tilde{v}_{k,t,m}^{\mathrm{ZF}}$ as indicated in (35). Summarizing the result of this section provides the following statement:

Theorem 3.1: The LS solution $\mathbf{f}_{\tau,t,m}^{\mathrm{ZF}}$ from (27) for the overdetermined and (28) for the underdetermined case, respectively, applied to the observation vector $\mathbf{r}_k$, defines a linear reference model in which the desired output signal is $s_{k-\tau,t,m}$, originating from a transmitted signal over antenna t of user m, and corrupted by additive compound noise $\tilde{v}_{k,t,m}^{\mathrm{ZF}}$. The compound noise is defined by $\mathbf{f}_{\tau,t,m}^{\mathrm{ZF}\,H}\mathbf{v}_k$ and has an additional component in the overdetermined case, i.e., the modeling noise $\mathbf{v}_{\tau,t,m}^{\mathrm{ZF,o}\,H}\mathbf{s}_k$, defined by the modeling error vector $\mathbf{v}_{\tau,t,m}^{\mathrm{ZF,o}}$ in (30).

B. MMSE Equalizer

MMSE solutions are typically derived [31] on the basis of signal and noise statistics, e.g., by

$$\mathbf{f}_{\tau,t,m}^{\mathrm{MMSE}} = \arg\min_{\mathbf{f}} \mathrm{E}\left[|\mathbf{f}^H\mathbf{r}_k - s_{k-\tau,t,m}|^2\right]. \tag{37}$$

However, the linear MMSE solution can alternatively be defined by

$$\begin{aligned}\mathbf{f}_{\tau,t,m}^{\mathrm{MMSE}} &= \arg\min_{\mathbf{f}}\left(\left\|\mathbf{H}^H\mathbf{f} - \mathbf{e}_{\tau,t,m}\right\|_2^2 + N_o\left\|\mathbf{f}\right\|_2^2\right)\\ &= \arg\min_{\mathbf{f}}\left\|\left(\mathbf{H}\mathbf{H}^H + N_o\mathbf{I}\right)^{\frac{1}{2}}\left[\mathbf{f} - (\mathbf{H}\mathbf{H}^H + N_o\mathbf{I})^{-1}\mathbf{H}\mathbf{e}_{\tau,t,m}\right]\right\|_2^2 + \mathrm{MMSE}\end{aligned} \tag{38}$$

with an additional term according to the additive noise variance N_o. We consider here white noise; alternative forms with colored noise, as originating for example from fractionally spaced equalizers, are straightforward; one only has to replace $N_o\mathbf{I}$ with $\mathbf{R}_{\mathrm{vv}}$, the autocorrelation matrix of the noise.

This formulation (38) now reveals that the MMSE problem equivalently can be written as a weighted LS problem with the weighting matrix $\left(\mathbf{H}\mathbf{H}^H + N_o\mathbf{I}\right)^{\frac{1}{2}}$. Note that the last term of (38)

$$\text{MMSE} = \mathbf{e}_{\tau,t,m}^T \left[\mathbf{I} - \mathbf{H}^H(\mathbf{H}\mathbf{H}^H + N_o\mathbf{I})^{-1}\mathbf{H}\right] \mathbf{e}_{\tau,t,m} \tag{39}$$

defines the minimum mean square error. As the term is independent of $\mathbf{f}$, it can thus be dropped when minimizing (38). The well-known MMSE solution is now obviously

$$\mathbf{f}_{\tau,t,m}^{\text{MMSE}} = (\mathbf{H}\mathbf{H}^H + N_o\mathbf{I})^{-1}\mathbf{H}\mathbf{e}_{\tau,t,m}. \tag{40}$$

Similarly to the ZF equalizer, an overdetermined solution for $RN_R < SN_TM$ also exists here:

$$\mathbf{f}_{\tau,t,m}^{\text{MMSE,o}} = \mathbf{H}(\mathbf{H}^H\mathbf{H} + N_o\mathbf{I})^{-1}\mathbf{e}_{\tau,t,m}. \tag{41}$$

Under white noise both solutions are in fact identical $\mathbf{f}_{\tau,t,m}^{\text{MMSE,o}} = \mathbf{f}_{\tau,t,m}^{\text{MMSE}}$, which is very different to the ZF equalizer (for the proof see Appendix A). The same procedure can be applied to DFE or FSE structures, the noise terms, however, are then partially zero or correlated and care needs to be taken when treating such cases. The equivalence (13) to the maximum SINR solution, however, is preserved. Correspondingly to the reference model for ZF equalizers in (29), we can now also define a reference model for MMSE equalizers

$$\mathbf{e}_{\tau,t,m} = \mathbf{H}^H\mathbf{f}_{\tau,t,m}^{\text{MMSE}} + \mathbf{v}_{\tau,t,m}^{\text{MMSE}} \tag{42}$$

with the modeling error

$$\mathbf{v}_{\tau,t,m}^{\text{MMSE}} = \left(\mathbf{I} - \mathbf{H}^H(\mathbf{H}\mathbf{H}^H + N_o\mathbf{I})^{-1}\mathbf{H}\right)\mathbf{e}_{\tau,t,m}. \tag{43}$$

Note however, that unlike in the case of the ZF solution the modeling error is not orthogonal to the MMSE solution, i.e., $\mathbf{v}_{\tau,t,m}^{\text{MMSE}\,H}\mathbf{H}^H\mathbf{f}_{\tau,t,m}^{\text{MMSE}} \neq 0$. MMSE equalizers are typically designed on the basis of observations rather than system parameters. Multiplying the signal vector with $\mathbf{e}_{\tau,t,m}$ we obtain

$$\mathbf{e}_{\tau,t,m}^T\mathbf{s}_k = s_{k-\tau,t,m} = \mathbf{f}_{\tau,t,m}^{\text{MMSE}\,H}\mathbf{H}\mathbf{s}_k + \mathbf{v}_{\tau,t,m}^{\text{MMSE}\,H}\mathbf{s}_k. \tag{44}$$

How does a received signal look after such MMSE-equalization? We apply $\mathbf{f}_{\tau,t,m}^{\text{MMSE}}$ on the observation vector and obtain

$$\mathbf{f}_{\tau,t,m}^{\text{MMSE}\,H}\mathbf{r}_k = s_{k-\tau,t,m} - \mathbf{v}_{\tau,t,m}^{\text{MMSE}\,H}\mathbf{s}_k + \mathbf{f}_{\tau,t,m}^{\text{MMSE}\,H}\mathbf{v}_k = s_{k-\tau,t,m} + \tilde{v}_{k,t,m}^{\text{MMSE}}. \tag{45}$$

From classic equalizer theory it is well known that the remaining ISI energy of the ZF equalizer is smaller than that of the MMSE part. This can be shown by comparing the energies of $\mathbf{v}_{\tau,t,m}^{\text{ZF}}$ and $\mathbf{v}_{\tau,t,m}^{\text{MMSE}}$ (see Appendix B)

$$\|\mathbf{v}_{\tau,t,m}^{\text{ZF}}\|_2^2 \leq \|\mathbf{v}_{\tau,t,m}^{\text{MMSE}}\|_2^2 \tag{46}$$

while the remaining ISI and noise energy is smaller for MMSE:

$$\|\mathbf{v}_{\tau,t,m}^{\mathrm{ZF}}\|_2^2 + N_o\|\mathbf{f}_{\tau,t,m}^{\mathrm{ZF}}\|_2^2 \geq \|\mathbf{v}_{\tau,t,m}^{\mathrm{MMSE}}\|_2^2 + N_o\|\mathbf{f}_{\tau,t,m}^{\mathrm{MMSE}}\|_2^2 \tag{47}$$

Let us summarize the most important findings of this section in the following statement:

Theorem 3.2: The weighted LS solution $\mathbf{f}_{\tau,t,m}^{\mathrm{MMSE}}$ from (40), applied to the observation vector $\mathbf{r}_k$, defines a linear reference model in which the desired output signal is $s_{k-\tau,t,m}$, originating from a transmitted signal over antenna t of user m, corrupted by additive compound noise $\tilde{v}_{k,t,m}^{\mathrm{MMSE}}$. The compound noise is defined by $\mathbf{f}^{\mathrm{MMSE}\,H}\mathbf{v}_k$ as well as by the modeling noise $\mathbf{v}_{k,t,m}^{\mathrm{MMSE}\,H}\mathbf{s}_k$, defined by the modeling error vector $\mathbf{v}_{k,t,m}^{\mathrm{MMSE}}$ in (43).

In conclusion, the adaptive equalizer problem has thus taken on the form of an identification problem as depicted in Figure 2. The linear system with impulse response $\mathbf{f}_{\tau,t,m}^{\mathrm{ref}}$ is estimated as $\hat{\mathbf{f}}_{\tau,t,m}$ by an adaptive equalizer algorithm. Here, 'ref' stands for either MMSE or ZF. The outcome of the reference system is disturbed by the compound noise $\tilde{v}_{k,t,m}^{\mathrm{ref}}$ (see (45)) and constructs a noisy reference symbol $s_{k-\tau,t,m}$. The adaptive filter with its output $\hat{s}_{k-\tau,t,m}+\hat{v}_{k,t,m}^{\mathrm{ref}}$ tries to resemble $s_{k-\tau,t,m}+\tilde{v}_{k,t,m}^{\mathrm{ref}}$. The distorted error signal $\tilde{e}_{k,t,m}$ is applied to the adaptive filter in order to adjust the equalizer solution, as will be explained in the following section.

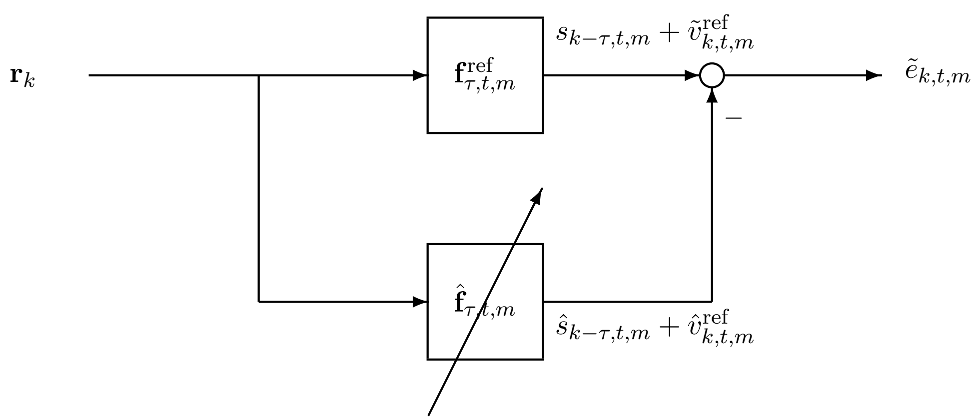


Fig. 2. *Adaptive equalizer as a system identification problem.*

IV. Iterative Algorithms

Equalizer solutions requiring matrix inverses are highly complex and numerically challenging, in particular when the matrix size is 50 or over [53]. In this section we apply the reference model description obtained in the previous section to iterative algorithms, as they offer the possibility of implementing the equalizer in a robust manner, being particularly suited to fixed-point implementations and furthermore avoid the need for an explicit matrix inversion. An iterative algorithm, as referred to here, is one that possesses all data and attempts to achieve an optimal solution. In the literature such algorithms are also referred to at times as off-line or batch algorithms since they require no new data during their operation. In this contribution we show convergence conditions for numerous known and novel algorithms but do not deal with the question of when to stop the iterations.

The design processes for iterative algorithms follow a common path. A channel $\mathbf{H}$ is applied in order to obtain the desired equalizer estimate. Due to our ZF and MMSE reference models in (29) and (42), respectively, we can define an observation error $\mathbf{e} - \mathbf{X}^H \hat{\mathbf{f}}_{l-1} = \mathbf{X}^H \mathbf{f}^{\mathrm{ref}} + \bar{\mathbf{v}} - \mathbf{X}^H \hat{\mathbf{f}}_{l-1}$. Thus all that is needed is a unit vector $\mathbf{e}$, the matrix $\mathbf{X}$ (which in the simplest case of ZF is the channel matrix $\mathbf{H}$) and our equalizer estimate. In the following, we omit the indexes for the various solutions in order to simplify the description. An iterative gradient algorithm takes on the following update, starting at a given value $\hat{\mathbf{f}}_0$:

$$\begin{aligned} \hat{\mathbf{f}}_l &= \hat{\mathbf{f}}_{l-1} + \mu \mathbf{X}\mathbf{B}\left(\mathbf{e} - \mathbf{X}^H \hat{\mathbf{f}}_{l-1}\right) \\ &= \hat{\mathbf{f}}_{l-1} + \mu \mathbf{X}\mathbf{B}\left(\mathbf{X}^H \left[\mathbf{f}^{\mathrm{ref}} - \hat{\mathbf{f}}_{l-1}\right] + \bar{\mathbf{v}}\right) \qquad ; l = 1, 2, ..., \end{aligned} \tag{48}$$

where the matrix $\mathbf{B}$ is considered to be positive definite and Hermitian. Given the set $\{\mathbf{X}, \mathbf{B}, \mathbf{e}\}$, the algorithm iterates until it approaches the desired solution $\mathbf{f}^{\mathrm{ref}}$.

Theorem 4.1: The iterative-algorithm updating Equation (48) converges in the parameter error vector to zero, i.e., $\lim_{k\to\infty} \hat{\mathbf{f}}_k = \mathbf{f}^{\mathrm{ref}}$ if both the following conditions hold:

1) The absolute value of all eigenvalues of $\mathbf{I} - \mu \mathbf{X}\mathbf{B}\mathbf{X}^H$ is smaller than one, or correspondingly

$$0 < \mu < \frac{2}{\max \lambda\{\mathbf{X}\mathbf{B}\mathbf{X}^H\}}. \tag{49}$$

2) The filtered noise $\mathbf{X}\mathbf{B}\bar{\mathbf{v}} = \mathbf{0}$.

Proof: Let us consider the update equation (48) in terms of the parameter error vector $\tilde{\mathbf{f}}_l = \mathbf{f}^{\mathrm{ref}} - \hat{\mathbf{f}}_l$:

$$\begin{aligned} \tilde{\mathbf{f}}_l &= \tilde{\mathbf{f}}_{l-1} - \mu \mathbf{X}\mathbf{B}\left(\mathbf{X}^H \tilde{\mathbf{f}}_{l-1} + \bar{\mathbf{v}}\right) \qquad ; l = 1, 2, ..., \\ &= \left(\mathbf{I} - \mu \mathbf{X}\mathbf{B}\mathbf{X}^H\right) \tilde{\mathbf{f}}_{l-1} - \mu \mathbf{X}\mathbf{B}\bar{\mathbf{v}}. \end{aligned} \tag{50}$$

Thus, as long as the step-size μ guarantees that all eigenvalues of $\mathbf{I} - \mu \mathbf{X}\mathbf{B}\mathbf{X}^H$ are smaller than one in magnitude, and the noise satisfies $\mathbf{X}\mathbf{B}\bar{\mathbf{v}} = \mathbf{0}$, the parameter error vector $\tilde{\mathbf{f}}_l$ will converge to zero.

We show in the following examples that both conditions are satisfied for all ZF and MMSE type iterative equalizer algorithms.

Iterative ZF algorithm on arbitrary channels:

Starting with an initial value $\mathbf{f}_0$ (which can be the zero vector), we arrive at the ZF iterative algorithm for $\mathbf{X} = \mathbf{H}$:

$$\hat{\mathbf{f}}_l = \hat{\mathbf{f}}_{l-1} + \mu \mathbf{H}(\mathbf{e}_{\tau,t,m} - \mathbf{H}^H \hat{\mathbf{f}}_{l-1}) \qquad ; l = 1, 2, ... \tag{51}$$

With the reference model in (29) we can introduce the parameter error vector $\tilde{\mathbf{f}}_l = \mathbf{f}^{\mathrm{ZF}}_{\tau,t,m} - \hat{\mathbf{f}}_l$ and obtain

$$\tilde{\mathbf{f}}_l = (\mathbf{I} - \mu \mathbf{H}\mathbf{H}^H)\tilde{\mathbf{f}}_{l-1} - \mu \mathbf{H}\mathbf{v}^{\mathrm{ZF}}_{\tau,t,m}. \tag{52}$$

We recognize that the noise condition is satisfied, as property $\mathbf{H}\mathbf{v}^{\mathrm{ZF}}_{\tau,t,m} = \mathbf{0}$ for $\bar{\mathbf{v}} = \mathbf{v}^{\mathrm{ZF}}_{\tau,t,m}$ from (31) holds independent if we have an over- or underdetermined equalizer. Convergence conditions for the step-size μ are now also readily derived, being dependent on the largest eigenvalue of $\mathbf{H}\mathbf{H}^H$, i.e.,

$$0 < \mu < \frac{2}{\max \lambda\{\mathbf{H}\mathbf{H}^H\}}. \tag{53}$$

As computing the largest eigenvalue may be a computationally expensive task, simpler bounds are of interest, even though they may be conservative.

1) A classic conservative bound is given by

$$0 < \mu < \frac{2}{\mathrm{trace}(\mathbf{H}\mathbf{H}^H)} \tag{54}$$

and can be computed with low complexity once the matrix $\mathbf{H}$ is known.

2) For an SU in a frequency selective SISO channel, the channel $\mathbf{H}$ is defined by a single Toeplitz matrix, the largest eigenvalue of which can also be bounded by $\max_\Omega |H(e^{j\Omega})|$, with $H(e^{j\Omega})$ denoting the Fourier transform of the channel impulse response. The corresponding condition for the step-size reads now

$$0 < \mu < \frac{2}{\max_\Omega |H(e^{j\Omega})|^2}. \tag{55}$$

Such a step-size may be more conservative than the condition (53) but is also more practical to find. In the simulation examples presented in Section VI the bound so obtained is very close to the theoretical value (53). Note that for a SIMO case, i.e., $N_T = 1$ but N_R receive antennas, a similar bound can be found:

$$0 < \mu < \frac{2}{\max_\Omega \sum_{r=1}^{N_R} |H_r(e^{j\Omega})|^2}, \tag{56}$$

where H_r denotes the frequency response for each of the N_R sub-channels.

Fast convergent iterative ZF Algorithm:

As the convergence of the previous equalizer algorithm is dependent on the channel matrix $\mathbf{H}$, the algorithm exhibits much slower convergence for some channels than for others, even for optimal step-sizes. The analysis of the algorithm shows that the optimal matrix $\mathbf{B}$ that ensures fastest convergence is given by $\mathbf{B} = \left[\mathbf{H}\mathbf{H}^H\right]^{-1}$, which is exactly the inverse whose computation we are attempting to avoid with the iterative approach. If, however, some a-priori knowledge is present on the channel class (e.g., Pedestrian B or Vehicular A), then we can precompute the mean value over an ensemble of channels from a specific class, for example $E\left[[\mathbf{H}\mathbf{H}^H]^{-1}\right] = \overline{\mathbf{R}_{\mathrm{HH}}^{-1}}$. In this case the algorithm updates read:

$$\hat{\mathbf{f}}_l = \hat{\mathbf{f}}_{l-1} + \mu\overline{\mathbf{R}_{\mathrm{HH}}^{-1}}\mathbf{H}(\mathbf{e}_{\tau,t,m} - \mathbf{H}^H\hat{\mathbf{f}}_{l-1}) \qquad ; l = 1, 2, ... \tag{57}$$

when $RN_R < SN_TM$ and

$$\hat{\mathbf{f}}_l = \hat{\mathbf{f}}_{l-1} + \mu\mathbf{H}\overline{\mathbf{R}_{\mathrm{HH}}^{-1}}(\mathbf{e}_{\tau,t,m} - \mathbf{H}^H\hat{\mathbf{f}}_{l-1}) \qquad ; l = 1, 2, ... \tag{58}$$

when $RN_R > SN_TM$ with the corresponding step-size bound in (49). An example of how this can be applied in provided in the first example of Section 6.

Iterative MMSE algorithm on arbitrary channels:

Starting with our MMSE reference model (45) we consider the following update equation

$$\hat{\mathbf{f}}_k = \hat{\mathbf{f}}_{k-1} + \mu\mathbf{X}(\mathbf{H}\mathbf{e}_{\tau,t,m} - (\mathbf{H}\mathbf{H}^H + N_o\mathbf{I})\hat{\mathbf{f}}_{k-1}) \tag{59}$$

for which $\mathbf{f}^{\mathrm{ref}} = \mathbf{f}_{\tau,t,m}^{\mathrm{MMSE}}$ is the desired solution. Recalling (42), i.e., $\mathbf{e}_{\tau,t,m} = \mathbf{H}^H\mathbf{f}_{\tau,t,m}^{\mathrm{MMSE}} + \mathbf{v}_{\tau,t,m}^{\mathrm{MMSE}}$ we identify the noise as $\bar{\mathbf{v}} = \mathbf{v}_{\tau,t,m}^{\mathrm{MMSE}}$ and we find that the error is composed of

$$\mathbf{H}\mathbf{e}_{\tau,t,m} - (\mathbf{H}\mathbf{H}^H + N_o\mathbf{I})\hat{\mathbf{f}}_{k-1} \quad = \quad (\mathbf{H}\mathbf{H}^H + N_o\mathbf{I})[\mathbf{f}_{\tau,t,m}^{\mathrm{MMSE}} - \hat{\mathbf{f}}_{k-1}] - N_o\mathbf{f}_{\tau,t,m}^{\mathrm{MMSE}} + \mathbf{H}\mathbf{v}_{\tau,t,m}^{\mathrm{MMSE}}. \tag{60}$$

Recalling with (43) that $\mathbf{v}_{\tau,t,m}^{\mathrm{MMSE}} = \left(\mathbf{I} - \mathbf{H}^H(\mathbf{H}\mathbf{H}^H + N_o\mathbf{I})^{-1}\mathbf{H}\right)\mathbf{e}_{\tau,t,m}$ we find that $-N_o\mathbf{f}_{\tau,t,m}^{\mathrm{MMSE}} + \mathbf{H}\mathbf{v}_{\tau,t,m}^{\mathrm{MMSE}} = \mathbf{0}$. We can thus reformulate (59) into

$$\tilde{\mathbf{f}}_k = \tilde{\mathbf{f}}_{k-1} - \mu\mathbf{X}(\mathbf{H}\mathbf{H}^H + N_o\mathbf{I})\tilde{\mathbf{f}}_{k-1}. \tag{61}$$

If we select $\mathbf{X} = \mathbf{I}$, we can identify $\mathbf{B} = \mathbf{H}\mathbf{H}^H + N_o\mathbf{I}$ and we find as a condition for convergence that

$$0 < \mu < \frac{2}{\max\lambda\{\mathbf{H}\mathbf{H}^H + N_o\mathbf{I}\}}. \tag{62}$$

Alternatively, a speed-up algorithm is possible with $\mathbf{X} = (\mathbf{R}_{\mathrm{HH}} + N_o\mathbf{I})^{-1}$ as an approximation or even better with $\mathbf{X} = (\mathbf{H}\mathbf{H}^H + N_o\mathbf{I})(\mathbf{R}_{\mathrm{HH}} + N_o\mathbf{I})^{-2}$.

V. Recursive adaptive filter algorithms

In contrast to iterative algorithms, recursive algorithms try to improve their estimation in the presence of new data that has not been previously applied. Such algorithms are sometimes referred to as on-line algorithms as they require new data for each adaptation step. Due to this property they typically behave adaptively as well. In addition to their learning behavior, recursive algorithms also have a tracking behavior that allows them to follow channel changes, a feature that makes them very interesting for many applications. Furthermore, there is typically no need for the channel matrix $\mathbf{H}$ to be known for such algorithms, offering a solution of very low complexity.

As previously in the case of iterative algorithms we start with the construction of an error signal. By applying the observations $\mathbf{X}_k$ (which can be observation vectors $\mathbf{r}_k$ or composite matrices of several of them) on the linear equalizer $\mathbf{f}^{\mathrm{ref}}$ we obtain the desired data signal with additive noise, thus $\mathbf{X}_k\mathbf{f}^{\mathrm{ref}} = \mathbf{s}_{k-\tau} - \bar{\mathbf{v}}_k$. The noise term is defined by the corresponding reference model (35), (36), or (45).

A recursive algorithm thus continuously takes on new data values. Its update equations take on the form

$$\begin{aligned}\hat{\mathbf{f}}_k &= \hat{\mathbf{f}}_{k-1} + \mu\mathbf{Y}_k\left(\mathbf{s}_{k-\tau} - \mathbf{X}_k^H\hat{\mathbf{f}}_{k-1}\right)\\ &= \hat{\mathbf{f}}_{k-1} + \mu\mathbf{Y}_k\left(\mathbf{X}_k^H\left[\mathbf{f}^{\mathrm{ref}} - \hat{\mathbf{f}}_{k-1}\right] + \bar{\mathbf{v}}_k\right) \qquad ; k = 1, 2, ...,\end{aligned} \tag{63}$$

A. Robust Adaptive Filter Algorithms

The analysis strategy is thus to reformulate the updates (63) in terms of the parameter error vector $\tilde{\mathbf{f}}_{k-1} = \mathbf{f}^{\mathrm{ref}} - \hat{\mathbf{f}}_{k-1}$ and find conditions on the step-size μ and the modeling noise sequence $\bar{\mathbf{v}}_k$ to guarantee l_2-stability of the recursive form. We refer to results from the standard literature [43, 44], repeated here for convenience of the reader, only bringing the algorithms into the required formulation. Note that l_2-stability is a much stronger form than the mean square sense convergence that is mostly associated to adaptive filters. This form of stability includes the desired property that there should be no data or noise sequences (or initial values) to cause divergence of the algorithm. Algorithms that behave in this way are thus called robust. In practice the differences may not be so obvious, as sequences that cause divergence may not often occur. On the other hand, a systematic search for them can reveal that a "trusted" algorithm is in fact by no means not robust [40].

Theorem 5.1: Consider Update (63) in which both $\mathbf{X}_k = \mathbf{r}_k$ and $\mathbf{Y}_k = \mathbf{B}\mathbf{r}_k$ are vectors with a constant positive definite Hermitian matrix $\mathbf{B}$ and noise $\bar{\mathbf{v}}_k = \tilde{v}_k\mathbf{r}_k$.

1) The algorithm is called robust and l_2-stable for step-sizes $0 < \mu < \max_k \frac{1}{\mathbf{r}_k^H\mathbf{B}\mathbf{r}_k}$ (alternatively for a time

variant step-size $0 < \mu_k < \frac{2}{\mathbf{r}_k^H \mathbf{B} \mathbf{r}_k}$). Consequently there is no sequence (noise or data) that causes the algorithm to diverge [43, 44].

2) Furthermore, if the noise $\tilde{v}_k$ is l_2 bounded, i.e., $\sum \mu_k |\tilde{v}_k|^2 < S_v < \infty$, the undistorted a priori estimation error $\sqrt{\mu_k} e_{a,k} = \sqrt{\mu_k} \mathbf{r}_k^H \tilde{\mathbf{f}}_{k-1}$ becomes zero [43, 44].

3) If additionally the regression vector $\mathbf{r}_k$ is of persistent excitation, the parameter error vector also becomes zero $\lim_{k\to\infty} \tilde{\mathbf{f}}_{k-1} = \mathbf{0}$ [43, 44].

4) If the compound noise $\tilde{v}_k$ is not l_2 bounded but has bounded variance, for a normalized step-size $\alpha = \mu_k \mathbf{r}_k^H \mathbf{B} \mathbf{r}_k$, we obtain convergence in the mean square sense for $0 < \alpha < 2$ and the steady-state misadjustment of the adaptive filter amounts to $\mathcal{M} = \frac{\alpha \sigma_{\tilde{v}}^2}{2-\alpha}$ [42].

There are even extensions towards non symmetric-matrices $\mathbf{B}$ [42, 54], which are not considered here. At first sight these conditions may appear very similar to those mentioned previously for iterative algorithms; note, however, that they are distinctively different as they involve observed data rather than estimated channels.

Recursive MMSE and ZF algorithm on arbitrary channels:

The first truly recursive algorithm we present here is the well-known LMS algorithm for equalization (see for example Rappaport [32] for the SU SISO case). We identify here $\mathbf{B} = \mathbf{I}$. Starting with an initial value $\hat{\mathbf{f}}_0$

$$\begin{aligned} \hat{\mathbf{f}}_k &= \hat{\mathbf{f}}_{k-1} + \mu_k \mathbf{r}_k [s_{k-\tau,t,m}^* - \mathbf{r}_k^H \hat{\mathbf{f}}_{k-1}] \qquad ; k = 1, 2, ... \\ &= \hat{\mathbf{f}}_{k-1} + \mu_k \mathbf{r}_k [\mathbf{r}_k^H (\mathbf{f}_{\tau,t,m}^{\text{MMSE}} - \hat{\mathbf{f}}_{k-1}) + \tilde{v}_{k,t,m}^{\text{MMSE}}]. \end{aligned} \tag{64}$$

The desired l_2-stability is achieved as long as the step-size condition is satisfied. According to [43, 44] for a constant step-size this is equivalent to

$$0 < \mu < \min_k \frac{2}{\|\mathbf{r}_k\|_2^2}. \tag{65}$$

In this case it makes sense to apply a time-variant step-size μ_k with step-size condition

$$0 < \mu_k < \frac{2}{\|\mathbf{r}_k\|_2^2}, \tag{66}$$

which can easily be satisfied for example by a normalized step-size. Due to the linear MMSE reference model we can now relate the adaptive equalizer algorithm to the well-known l_2-stability of adaptive filters [43, 44]; for further details see [42].

According to (45) the noise is $\tilde{v}_{k,t,m} = -\mathbf{v}_{\tau,t,m}^{\text{MMSE}\,H} \mathbf{s}_k + \mathbf{f}_{\tau,t,m}^{\text{MMSE}\,H} \mathbf{v}_k$. Thus even in the absence of additive noise v_k, the compound noise sequence will not be bounded. We can thus not expect the recursive, adaptive MMSE equalizer (without sophisticated step-size control) to converge to zero. If additive noise $\mathbf{v}_k$ is absent, only the underdetermined ZF solution can converge to its corresponding reference. In this case MMSE and ZF solutions are identical.

Otherwise if additive noise remains unbounded but is of finite variance, a steady-state mismatch larger than zero remains. Following Property 4 of Theorem 5.1, with a normalized step-size $\mu_k = \alpha/\|\mathbf{r}_k\|_2^2$, we obtain a misadjustment of

$$\mathcal{M}^{\mathrm{MMSE}} = \frac{\alpha}{2-\alpha}\left(\|\mathbf{v}_{\tau,t,m}^{\mathrm{MMSE}}\|_2^2 + N_o\|\mathbf{f}_{\tau,t,m}^{\mathrm{MMSE}}\|_2^2\right). \tag{67}$$

In [42] simulation results are presented that validate such relations. Note that we can follow the same arguments as before, by simply replacing the reference model with the ZF model. We then obtain the same stability conditions (66) just with a different misadjustment:

$$\mathcal{M}^{\mathrm{ZF}} = \frac{\alpha}{2-\alpha}\left(\|\mathbf{v}_{\tau,t,m}^{\mathrm{ZF}}\|_2^2 + N_o\|\mathbf{f}_{\tau,t,m}^{\mathrm{ZF}}\|_2^2\right). \tag{68}$$

We can thus not truly argue that the recursive algorithm leads to the MMSE or the ZF solution as there is a considerable misadjustment for both algorithms. However, the distance to the two solutions in the mean square sense is smaller for MMSE than for ZF (see Appendix B). As in the case of previous algorithms, a speed-up variant with $\mathbf{B} = (\mathbf{R}_{\mathrm{HH}} + N_o\mathbf{I})^{-1}$ is possible.

Alternative recursive ZF algorithm on arbitrary channels:

Consider the following potential update for a recursive ZF algorithm (starting with an initial estimate $\hat{\mathbf{f}}_0$) in which we replace the regression vector $\mathbf{r}_k$ of the previous recursive MMSE algorithm with $\mathbf{H}\mathbf{s}_k$:

$$\begin{aligned}
\hat{\mathbf{f}}_k &= \hat{\mathbf{f}}_{k-1} + \mu_k\mathbf{H}\mathbf{s}_k[s_{k-\tau,t,m}^* - \mathbf{r}_k^H\hat{\mathbf{f}}_{k-1}] \qquad ; k = 1, 2, ... \\
&= \hat{\mathbf{f}}_{k-1} + \mu_k\mathbf{H}\mathbf{s}_k[\mathbf{r}_k^H(\mathbf{f}_{\tau,t,m}^{\mathrm{ZF}} - \hat{\mathbf{f}}_{k-1}) + \tilde{v}_{k,t,m}^{\mathrm{ZF}}] \\
&= \hat{\mathbf{f}}_{k-1} + \mu_k\mathbf{H}\mathbf{s}_k[\mathbf{s}_k^H\mathbf{H}^H\tilde{\mathbf{f}}_{k-1} + \mathbf{v}_k^H\tilde{\mathbf{f}}_{k-1} + \tilde{v}_{k,t,m}^{\mathrm{ZF}}]
\end{aligned} \tag{69}$$

The last line reveals that additional noise occurs, proportional to the parameter error vector. Algorithms of such structure have been analyzed in [40, 42] and it has been shown that they exhibit convergence in the mean-square sense with sufficiently small step-size, but also that they are not robust since unbounded noise sequences (unbounded energy but with bounded variance) can be found that cause the algorithm to diverge even for very small step-sizes. We thus conclude that in this set-up of recursive algorithms, there is no known robust form that truly achieves a ZF (or MMSE) solution under unbounded noise.

B. *Adaptive equalizer algorithms in block form*

The situation is more difficult if $\mathbf{X}_k$ and $\mathbf{Y}_k$ are truly matrices. In this case the theoretical results from the literature are not directly applicable and thus need to be extended first. Such scenarios occur if the update is not performed at every time instant but blocked, say after every T instances. Applied to the equalizer problem we can state the following:

Theorem 5.2: Consider the recursive block algorithm with update (63) for which $\mathbf{Y}_k = \mathbf{X}_k\mathbf{B}_k$ with a positive definite Hermitian matrix $\mathbf{B}_k$.

1) The algorithm is called robust and l_2-stable if the step-size μ can be selected so that $0 \leq \mu \leq \frac{1}{\lambda_{\max}\left(\mathbf{B}_k^{H/2}\tilde{\mathbf{X}}_k^H\tilde{\mathbf{X}}_k\mathbf{B}_k^{1/2}\right)}$.
2) Furthermore, if the noise sequence $\sqrt{\mu_k}\|\mathbf{B}_k^{H/2}\bar{\mathbf{v}}_k\|$ is bounded, i.e., $\sum_{k=1}^{K}\mu_k\|\mathbf{B}_k^{H/2}\bar{\mathbf{v}}_k\|_2 < S_v < \infty$, then the undistorted a priori error vector $\mathbf{e}_{a,k} = \mathbf{X}_k^H\tilde{\mathbf{f}}_{k-1}$ becomes zero.
3) If additionally a persistent excitation condition is satisfied for $\sqrt{\mu_k}\mathbf{X}_k$, we also can conclude that the parameter error vector $\tilde{\mathbf{f}}_k$ converges to zero.

Proof: See Appendix C.

Iterative/recursive block MMSE algorithm: Consider a set of receiver vectors $\mathbf{r}_k$ obtained over a time period T. Stacking the corresponding receiver vectors $\mathbf{r}_k$ in a matrix $\mathbf{R}_k = [\mathbf{r}_k, \mathbf{r}_{k-1}, ..., \mathbf{r}_{k-T+1}]$ and correspondingly the transmitted symbols $\mathbf{S}_k = [\mathbf{s}_k, \mathbf{s}_{k-1}, ..., \mathbf{s}_{k-T+1}]$, we can reformulate the previous transmission model in block form to $\mathbf{R}_k = \mathbf{H}\mathbf{S}_k + \mathbf{V}_k$ in which we also stacked the noise vectors accordingly into $\mathbf{V}_k$. We only have to assume that the channel $\mathbf{H}$ remains constant during the period T. Applying (45) in matrix form, we obtain the reference model

$$\mathbf{f}_{\tau,t,m}^{\mathrm{MMSE}\,H}\mathbf{R}_k = \mathbf{s}_{k-\tau,t,m} - \mathbf{v}_{\tau,t,m}^{\mathrm{MMSE}\,H}\mathbf{S}_k + \mathbf{f}_{\tau,t,m}^{\mathrm{MMSE}\,H}\mathbf{V}_k = \mathbf{s}_{k-\tau,t,m} + \tilde{\mathbf{v}}_{k,t,m}^{\mathrm{MMSE}}. \tag{70}$$

The signal to decode $\mathbf{s}_{k-\tau,t,m} \in \mathbb{C}^{T\times 1}$ is now in vector form, as is the compound noise $\tilde{\mathbf{v}}_{k,t,m}^{\mathrm{MMSE}} \in \mathbb{C}^{T\times 1}$. For each time instant k it is possible to first derive an *iterative* algorithm, starting with the initial value $\hat{\mathbf{f}}_0$:

$$\hat{\mathbf{f}}_l = \hat{\mathbf{f}}_{l-1} + \mu_k\mathbf{R}_k\left(\mathbf{s}_{k-\tau,t,m}^* - \mathbf{R}_k^H\hat{\mathbf{f}}_{l-1}\right) \qquad ;l=1,2.... \tag{71}$$

With the reference model (70), the parameter error vector $\tilde{\mathbf{f}}_l = \mathbf{f} - \hat{\mathbf{f}}_l$ at time instant k is given by

$$\tilde{\mathbf{f}}_l = \left(\mathbf{I} - \mu_k\mathbf{R}_k\mathbf{R}_k^H\right)\tilde{\mathbf{f}}_{l-1} + \mu_k\mathbf{R}_k\tilde{\mathbf{v}}_{k,t,m}^{\mathrm{MMSE}} \qquad ;l=1,2,.... \tag{72}$$

Convergence conditions can now be derived on the basis of the largest eigenvalue of matrix $\mathbf{R}_k\mathbf{R}_k^H$: $0 < \mu_k < \frac{2}{\max\lambda\{\mathbf{R}_k\mathbf{R}_k^H\}}$. Note, however, that due to the noise components in $\mathbf{R}_k$ a final error remains: $\tilde{\mathbf{f}}_\infty = [\mathbf{R}_k\mathbf{R}_k^H]^{-1}\mathbf{R}_k\tilde{\mathbf{v}}_{k,t,m}^{\mathrm{MMSE}} \neq \mathbf{0}$. If we envision an ensemble of random data and noise sets $\mathbf{S}_k$ and $\mathbf{V}_k$, convergence can only be found in the mean sense, i.e., $\lim_{k\to\infty}\mathrm{E}[\tilde{\mathbf{f}}_l] = \mathbf{0}$.

The entire algorithm remains an iterative algorithm and not a recursive method as long as k remains fixed. As soon as k changes, the recursive aspect enters the algorithm. Note, however, that the convergence condition then becomes time-variant as for each time instant k another $\mathbf{R}_k\mathbf{R}_k^H$ appears and thus the step-size needs to be adapted. Based on the previous analysis, convergence in the mean sense remains. Also note that such algorithms are quite popular, often appearing under the name of data-reuse adaptive filters [55–60].

Recursive single-user block MMSE algorithm: Rather than applying an iterative method (or mixed iterative/recursive method), we can reformulate the algorithm in terms of a purely recursive procedure in which new data are being processed at every time instant k at a block basis of period T:

$$\hat{\mathbf{f}}_k = \hat{\mathbf{f}}_{k-1} + \mu_k \mathbf{R}_k \left(\mathbf{s}^*_{k-\tau,t,m} - \mathbf{R}_k^H \hat{\mathbf{f}}_{k-1} \right) \qquad ; k = 1, 2.... \tag{73}$$

Note that the algorithm now tries to improve the solution at every time instant k under varying conditions $\{\mathbf{R}_k, \mathbf{s}_{k-\tau,t,m}\}$. To guarantee l_2-stability, the step-size μ_k needs to be changed over time:

$$0 < \mu_k < \frac{1}{\max \lambda\{\mathbf{R}_k \mathbf{R}_k^H\}}. \tag{74}$$

More conservative bounds over trace($\mathbf{R}_k\mathbf{R}_k^H$), which can be derived with less complexity, are possible. Following Theorem 5.2 we identify $\mathbf{Y}_k = \mathbf{X}_k = \mathbf{R}_k$ and we can conclude robustness and l_2-stabilty. Similar to the fast convergent ZF algorithm, a fast convergent MMSE algorithm is also possible. However, it requires pre-computation of the inverse of $\mathbf{R}_k\mathbf{R}_k^H$ at every time instant k, which is prohibitive from a complexity point of view. Alternatively, if the channel class is a priori known, a fixed mean value $\mathbf{B} = E\left[[\mathbf{R}_k\mathbf{R}_k^H]^{-1}\right]$ can be employed, i.e., $\mathbf{Y}_k = \mathbf{R}_k\mathbf{B}$ rather than $\mathbf{R}_k$ as a gradient term for speeding up the convergence.

C. Recursive blind MMSE channel estimation

All iterative equalizer methods discussed so far require the knowledge of channel matrix $\mathbf{H}$. The reference model also facilitates deriving robust algorithms for channel estimation. While these are well known algorithms for system identification once a training sequence is employed, it is of greater interest to derive algorithms when such training sequences are entirely missing. In such situations we call this a blind algorithm. Let us first consider an iterative algorithm that employs two SISO channels, say $\mathbf{H}_1$ and $\mathbf{H}_2$, the cost function for which is given by

$$\begin{aligned}
\mathbf{h}^{\mathrm{B}} &= \arg \min_{\mathbf{h}=\begin{bmatrix}\mathbf{h}_2\\ \mathbf{h}_1\end{bmatrix}} \left\|\mathbf{H}_1^H\mathbf{h}_2 - \mathbf{H}_2^H\mathbf{h}_1\right\|_2^2 \\
&= \arg\min_{\mathbf{h}} \left\|\left[\mathbf{H}_1^H, -\mathbf{H}_2^H\right]\mathbf{h}\right\|_2^2 \\
&= \arg\min_{\mathbf{h}} \left\|\mathbf{H}^H\mathbf{h}\right\|_2^2. \qquad (75)
\end{aligned}$$

If $\mathbf{h}_1$ is proportional to the impulse response in channel Toeplitz matrix $\mathbf{H}_1$ and $\mathbf{h}_2$ proportional to $\mathbf{H}_2$, then we find $\mathbf{H}_1^H\mathbf{h}_2 = \mathbf{H}_2^H\mathbf{h}_1$ and so the cost function becomes zero and thus minimal. If we can guarantee $R > S$ we can derive an iterative algorithm, starting at initial value $\hat{\mathbf{h}}_0$ as:

$$\hat{\mathbf{h}}_l = \hat{\mathbf{h}}_{l-1} - \mu\mathbf{H}(\mathbf{H}^H\hat{\mathbf{h}}_{l-1}) \qquad ; l = 1, 2, ..., \qquad (76)$$

with the compound channel $\mathbf{H}$ and channel estimate $\hat{\mathbf{h}}_{l-1}$ for the stacked channel $\mathbf{h}^{\mathrm{B}}$ as defined above. Convergence is guaranteed for $0 < \mu < \frac{2}{\max\lambda\{\mathbf{H}\mathbf{H}^H\}}$. One can also view the blind algorithm above as a standard iterative algorithm solving an LS problem for which the reference is zero, thus explaining the minus sign in (76). Utilizing the compound channel $\mathbf{H}$, step-size bounds for convergence of the algorithm can be derived as before. As it stands, the algorithm does not yet make any sense, as the channel would need to be known in order to blindly estimate it.

A true blind algorithm does not know the channel matrices and has to operate on the received vectors, i.e., $\mathbf{r}_k^T = [\mathbf{r}_k^{(1)T}, -\mathbf{r}_k^{(2)T}]$, of which T vectors are stacked into the corresponding matrix $\mathbf{R}_k^H = [\mathbf{R}_k^{(1)H}, -\mathbf{R}_k^{(2)H}]$. We can thus formulate a recursive algorithm for blind channel estimation, starting with initial value $\hat{\mathbf{h}}_0$:

$$\begin{aligned}
\hat{\mathbf{h}}_k &= \hat{\mathbf{h}}_{k-1} - \mu_k\mathbf{R}_k\mathbf{R}_k^H\hat{\mathbf{h}}_{k-1} \qquad ; k = 1, 2.... \qquad (77)\\
\tilde{\mathbf{h}}_k &= \tilde{\mathbf{h}}_{k-1} + \mu_k\mathbf{R}_k(\mathbf{R}_k^H\mathbf{h}^{\mathrm{B}} - \mathbf{R}_k^H\tilde{\mathbf{h}}_{k-1}) \\
&= \tilde{\mathbf{h}}_{k-1} - \mu_k\mathbf{R}_k(\mathbf{R}_k^H\tilde{\mathbf{h}}_{k-1} - \mathbf{V}_k^H\mathbf{h}^{\mathrm{B}}), \qquad (78)
\end{aligned}$$

where we used $\mathbf{V}_k^H = [\mathbf{V}_k^{(1)H}, -\mathbf{V}_k^{(2)H}]$. Stability is guaranteed for step-sizes that satisfy (74). Naturally, for convergence some other conditions need to be satisfied: persistent excitation of s_k, channels that are co-prime and bounded noise [61]. The classic blind channel estimation can thus be straightforwardly formulated as a robust, recursive algorithm as an alternative to the highly-complex SVD methods originally proposed in [20, 21], adaptive variants in [62]. Robust blind equalizers have been treated in [63].

VI. Experimental Results

In order to corroborate our theoretical findings, we present selected Matlab examples, in which we elaborate previous experimental settings, provided in [42], based on a set of seven channel impulse responses of finite length:

$$\begin{aligned}
h_k^{(1)} &= \frac{1}{1-0.9q^{-1}}[\delta_k] = 0.9^k; k=0,1,2..M-1 && (79)\\
h_k^{(2)} &= \frac{1}{1-0.8q^{-1}}[\delta_k] = 0.8^k; k=0,1,2..M-1 && (80)\\
h_k^{(3)} &= \frac{1}{1-q^{-1}+0.5q^{-2}}[\delta_k]; k=0,1,..,M-1 && (81)\\
h_k^{(4)} &= \frac{1}{1-1.4q^{-1}+0.6q^{-2}}[\delta_k]; k=0,1,..,M-1 && (82)\\
h_k^{(5)} &= \delta_k + 0.6\delta_{k-2} && (83)\\
h_k^{(6)} &= \delta_k + \delta_{k-2} && (84)\\
h_k^{(7)} &= \delta_k + \delta_{k-1}, && (85)
\end{aligned}$$

where we used the Kronecker delta δ_k to describe a unit impulse and the delay operator $q^{-1}[x_k] = x_{k-1}$. In [42] the MMSE equalizer algorithm (64) was demonstrated to run on channel $h^{(4)}$. An excellent agreement with the steady-state values (67) was found.

Iterative ZF equalizer: In our first example we select the length of the channel to be $M = 50$ for which even the first four impulse responses have decayed considerably. If we run an iterative receiver (also of 50 taps) according to (51), the result for $h^{(3)}$ is depicted on the left-hand side of Figure 3, with $\mathbf{f}_0$ denoting the ZF solution and $\hat{\mathbf{f}}_l$ denoting its estimate. Based on the convergence condition (53), it is possible to compute the exact step-size bound (0.255), given the channel matrix $\mathbf{H}$. Also shown in the figure are the conservative bound (54), which is the smallest step-size (0.017) in the figure, resulting in the slowest convergence speed and (55), which is just a fraction smaller (0.25 vs. 0.255) than the step-size bound. The average inverse autocorrelation $\overline{\mathbf{R}_{\mathrm{HH}}^{-1}}$ is computed over all seven channels (79)-(85), and applied in the algorithm's updates. This results in a considerable speed-up in the iterations as proposed in (57) and is depicted on the right-hand side of Figure 3.

Non-robust alternative recursive ZF algorithm: In the second example we consider the recursive algorithm in (69). We again apply channel $h^{(3)}$ for $M = 50$ taps and compare the outcome with the ZF solution (also of 50 taps). Additive noise was $N_0 = 0.01$ and QPSK for data transmission over the channel. Figure 4 displays the obtained relative system mismatch for various normalized step-sizes α, i.e., $\mu = \alpha\frac{1}{\|\mathbf{H}\mathbf{s}_k\|_2^2}$. On the left we display

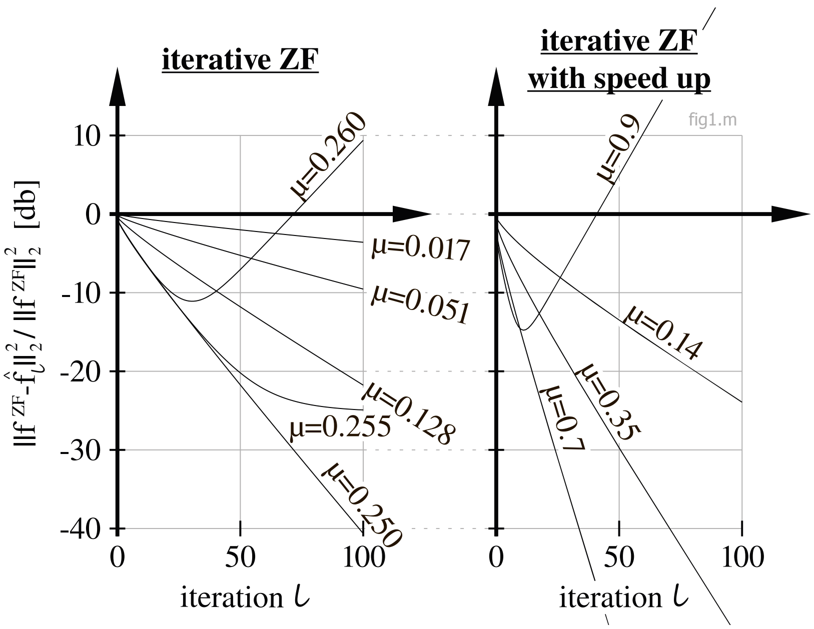


Fig. 3. *Left: Convergence behavior of the iterative ZF algorithm (51), Right: Iterative ZF algorithm with speed-up variant (57).*

the behavior for random data sequences. As expected for normalized step-sizes $0 < \alpha < 2$ we observe convergence in the mean square sense. On the right of the figure we show the behavior when searching for the worst case data sequence. We observe that even for very small step-sizes, the algorithm diverges and thus behaves non-robustly as predicted.

Blind channel estimation and equalizer: In the final example we simulated a blind channel estimation according to (77). We selected two channels, channel $h_k^{(3)}$ and $h_k^{(4)}$ from the list above, but shortened them to five elements ($M = 5$). We ran the experiment with additive noise $N_o = 0.01$ and averaged over 50 runs. The left-hand side of Figure 5 depicts the results when using only the instanteneous regression vector $\mathbf{r}_k$ while the right-hand side of the figure applies a block mode with $T = 5$. As a figure of merit we compute the closeness of the two estimated channels to the true channels by the correlation coefficient:

$$2|r_{12}|^2 = \frac{|\mathbf{h}_1^H \hat{\mathbf{h}}_2|^2}{\|\mathbf{h}_1\|_2^2 \|\hat{\mathbf{h}}_2^2\|_2} + \frac{|\mathbf{h}_2^H \hat{\mathbf{h}}_1|^2}{\|\mathbf{h}_2\|_2^2 \|\hat{\mathbf{h}}_1\|_2^2}. \tag{86}$$

In Figure 5 the distance of the correlation coefficient to one, i.e., $1 - |r_{1,2}|^2$ is depicted. In both cases the step-sizes are normalized to the conservative bounds, i.e., $\frac{\mu}{\|\mathbf{r}_k\|_2^2}$ and $\frac{\mu}{\mathrm{tr}(\mathbf{R}_k \mathbf{R}_k^H)}$, respectively.

All Matlab examples can be downloaded from https://www.nt.tuwien.ac.at/downloads/featured-downloads.

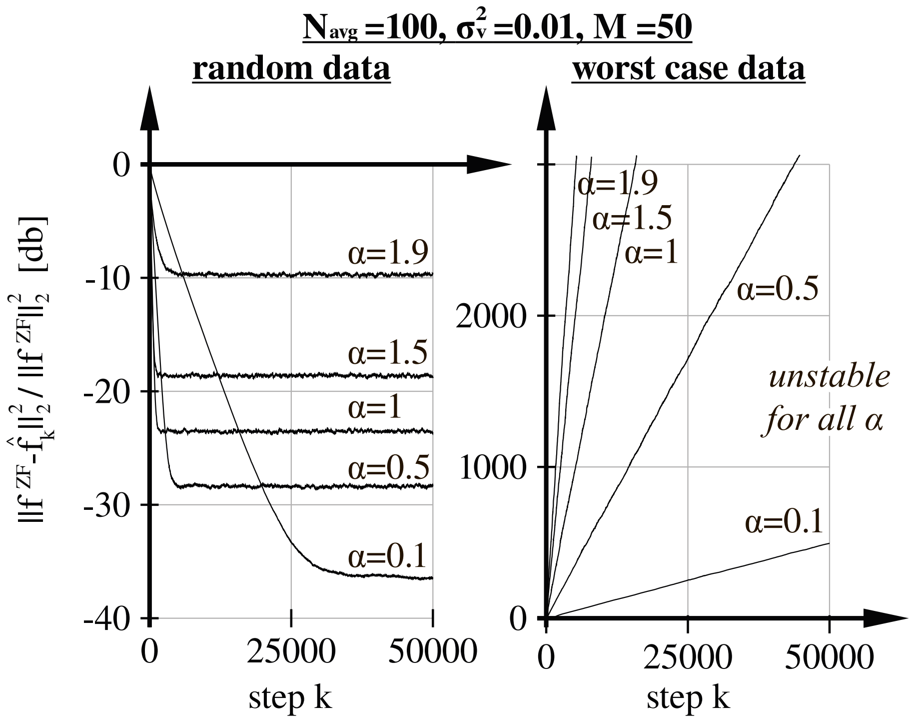


Fig. 4. *Left: Convergence behavior of the recursive ZF algorithm (69) under random data; Right: Convergence behavior under worst case data sequences.*

VII. Conclusion

We have started with well-known SIR and SINR formulations for linear equalizers and shown their connection to equivalent ZF and MMSE formulations. Due to an LS approach it is now possible to derive the classical equalizer types with an alternative formulation, an LS formulation for ZF and a weighted LS formulation for MMSE equalizers. This in turn resulted in a linear reference model for both. Based on such a linear reference model, it is possible to derive iterative as well as recursive forms of equalizers that are robust. Conditions for their robustness were presented, and in particular ranges for their only free parameter, the step-size, were presented to guarantee robust learning. The reader may be disappointed not to encounter the classical algorithm from Lucky [4, 5]. As this recursive algorithm is not l_2-stable, it does not fit into our robustness framework and has been treated elsewhere in a different context [42].

Acknowledgments

The author wishes to thank the anonymous reviewers for their helpful comments, Sebastian Caban for preparing high quality figures and Nick Shaw for his language corrections.

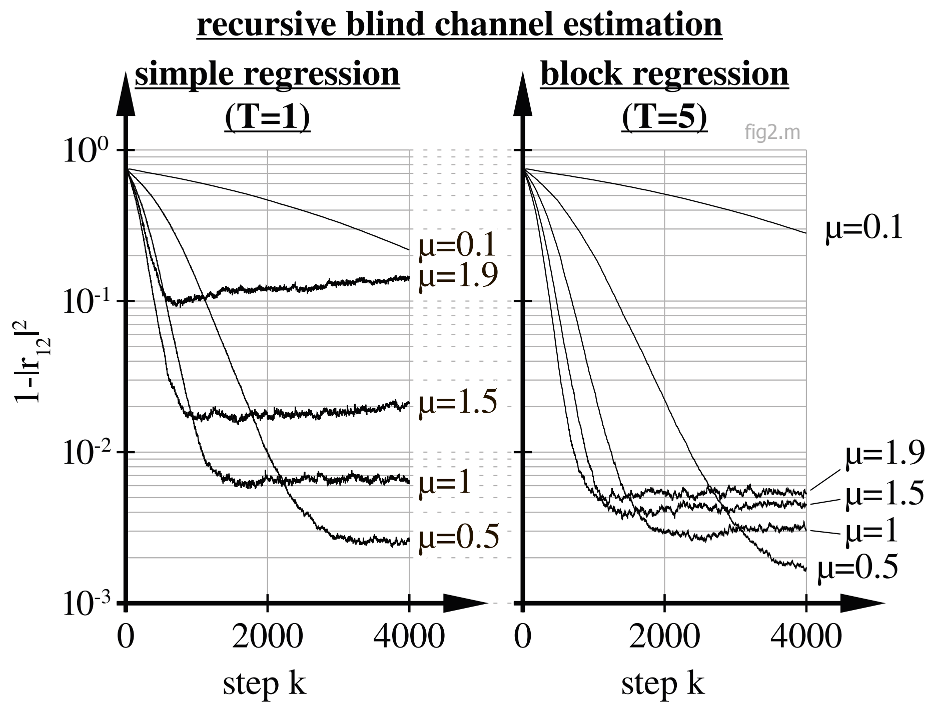


Fig. 5. *Left: Convergence behavior of recursive channel estimator (77) with instantaneous observations ($T = 1$); Right: Convergence behavior of channel estimator in block mode ($T = 5$).*

Appendix

A. Appendix A

We want to show that both MMSE solutions

$$\mathbf{f}_{\tau,t,m}^{\mathrm{MMSE}} = (\mathbf{H}\mathbf{H}^H + N_o\mathbf{I})^{-1}\mathbf{H}\mathbf{e}_{\tau,t,m} = \mathbf{H}(\mathbf{H}^H\mathbf{H} + N_o\mathbf{I})^{-1}\mathbf{e}_{\tau,t,m} \tag{87}$$

are equivalent. Although the reader can find the result also in the literature [64], we repeat it here for convenience. We first apply the matrix inversion lemma [34]

$$[\mathbf{A} + \mathbf{C}\mathbf{C}^H]^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{C}[\mathbf{I} + \mathbf{C}^H\mathbf{A}^{-1}\mathbf{C}]^{-1}\mathbf{C}^H\mathbf{A}^{-1} \tag{88}$$

and multiply from the right by $\mathbf{C}$:

$$\begin{aligned} [\mathbf{A} + \mathbf{C}\mathbf{C}^H]^{-1}\mathbf{C} &= \mathbf{A}^{-1}\mathbf{C} - \mathbf{A}^{-1}\mathbf{C}[\mathbf{I} + \mathbf{C}^H\mathbf{A}^{-1}\mathbf{C}]^{-1}\mathbf{C}^H\mathbf{A}^{-1}\mathbf{C} \\ &= \mathbf{A}^{-1}\mathbf{C}\left[\mathbf{I} - [\mathbf{I} + \mathbf{C}^H\mathbf{A}^{-1}\mathbf{C}]^{-1}\mathbf{C}^H\mathbf{A}^{-1}\mathbf{C}\right] \\ &= \mathbf{A}^{-1}\mathbf{C}[\mathbf{I} + \mathbf{C}^H\mathbf{A}^{-1}\mathbf{C}]^{-1}. \end{aligned} \tag{89}$$

We now substitute $\mathbf{A} = N_o\mathbf{I}$ and $\mathbf{C} = \mathbf{H}$ and immediately obtain

$$(\mathbf{H}\mathbf{H}^H + N_o\mathbf{I})^{-1}\mathbf{H} = \mathbf{H}(\mathbf{H}^H\mathbf{H} + N_o\mathbf{I})^{-1}, \tag{90}$$

which in turn delivers the desired result.

B. Appendix B

We find

$$
\begin{aligned}
\|\mathbf{v}_{\tau,t,m}^{\text{ZF,u}}\|_2^2 &= 0, && (91)\\
\|\mathbf{v}_{\tau,t,m}^{\text{ZF,o}}\|_2^2 &= \mathbf{e}_{\tau,t,m}^T[\mathbf{I}-\mathbf{H}^H[\mathbf{H}\mathbf{H}^H]^{-1}\mathbf{H}]\mathbf{e}_{\tau,t,m}, && (92)\\
\|\mathbf{f}_{\tau,t,m}^{\text{ZF,u}}\|_2^2 &= \mathbf{e}_{\tau,t,m}^T[\mathbf{H}^H\mathbf{H}]^{-1}\mathbf{e}_{\tau,t,m}, && (93)\\
\|\mathbf{f}_{\tau,t,m}^{\text{ZF,o}}\|_2^2 &= \mathbf{e}_{\tau,t,m}^T\mathbf{H}^H[\mathbf{H}\mathbf{H}^H]^{-2}\mathbf{H}\mathbf{e}_{\tau,t,m}, && (94)\\
\|\mathbf{v}_{\tau,t,m}^{\text{MMSE}}\|_2^2 &= \mathbf{e}_{\tau,t,m}^T[\mathbf{I}-\mathbf{H}^H\underbrace{[\mathbf{H}\mathbf{H}^H+N_o\mathbf{I}]^{-1}}_{\overline{\mathbf{Q}}}\mathbf{H}]^2\mathbf{e}_{\tau,t,m} && (95)\\
\|\mathbf{f}_{\tau,t,m}^{\text{MMSE}}\|_2^2 &= \mathbf{e}_{\tau,t,m}^T\mathbf{H}^H\overline{\mathbf{Q}}^2\mathbf{H}\mathbf{e}_{\tau,t,m} = \mathbf{e}_{\tau,t,m}^T\underbrace{(\mathbf{H}^H\mathbf{H}+N_o\mathbf{I})^{-1}}_{\mathbf{Q}}\mathbf{H}^H\mathbf{H}(\mathbf{H}^H\mathbf{H}+N_o\mathbf{I})^{-1}\mathbf{e}_{\tau,t,m} && (96)
\end{aligned}
$$

resulting in

$$
\begin{aligned}
\|\mathbf{v}_{\tau,t,m}^{\text{ZF,u}}\|_2^2+N_o\|\mathbf{f}_{\tau,t,m}^{\text{ZF,u}}\|_2^2 &= N_o\mathbf{e}_{\tau,t,m}^T[\mathbf{H}^H\mathbf{H}]^{-1}\mathbf{e}_{\tau,t,m}, && (97)\\
\|\mathbf{v}_{\tau,t,m}^{\text{ZF,o}}\|_2^2+N_o\|\mathbf{f}_{\tau,t,m}^{\text{ZF,o}}\|_2^2 &= \mathbf{e}_{\tau,t,m}^T[\mathbf{I}-\mathbf{H}^H[\mathbf{H}\mathbf{H}^H]^{-1}\overline{\mathbf{Q}}^{-1}[\mathbf{H}\mathbf{H}^H]^{-1}\mathbf{H}]\mathbf{e}_{\tau,t,m}, && (98)\\
\|\mathbf{v}_{\tau,t,m}^{\text{MMSE}}\|_2^2+N_o\|\mathbf{f}_{\tau,t,m}^{\text{MMSE}}\|_2^2 &= \mathbf{e}_{\tau,t,m}^T[\mathbf{I}-\mathbf{H}^H\overline{\mathbf{Q}}\mathbf{H}]\mathbf{e}_{\tau,t,m} = \mathbf{e}_{\tau,t,m}^T[\mathbf{I}-\mathbf{H}^H\mathbf{H}\mathbf{Q}]\mathbf{e}_{\tau,t,m}, && (99)
\end{aligned}
$$

where we made use of the fact that $\mathbf{H}\mathbf{Q}=\overline{\mathbf{Q}}\mathbf{H}$.

We consider the difference of

$$
\begin{aligned}
\|\mathbf{v}_{\tau,t,m}^{\text{ZF,o}}\|_2^2 &- \|\mathbf{v}_{\tau,t,m}^{\text{MMSE}}\|_2^2 = \\
&= \mathbf{e}_{\tau,t,m}^T[\mathbf{I}-\mathbf{H}^H[\mathbf{H}\mathbf{H}^H]^{-1}\mathbf{H}-[\mathbf{I}-\mathbf{H}^H\overline{\mathbf{Q}}\mathbf{H}]^2]\mathbf{e}_{\tau,t,m} && (100)\\
&= -\mathbf{e}_{\tau,t,m}^T\mathbf{H}^H\left([\mathbf{H}\mathbf{H}^H]^{-1}+\overline{\mathbf{Q}}+N_o\overline{\mathbf{Q}}^2\right)\mathbf{H}\mathbf{e}_{\tau,t,m}.
\end{aligned}
$$

In order for the difference $\|\mathbf{v}_{\tau,t,m}^{\text{ZF}}\|_2^2-\|\mathbf{v}_{\tau,t,m}^{\text{MMSE}}\|_2^2<0$ we have to have $[\mathbf{H}\mathbf{H}^H]^{-1}+\overline{\mathbf{Q}}+N_o\overline{\mathbf{Q}}^2>0$[1], which is obviously true. As for the underdetermined case $\|\mathbf{v}_{\tau,t,m}^{\text{ZF,u}}\|_2^2=0$, the difference is also negative.

We then consider the difference of

$$
\begin{aligned}
&\|\mathbf{v}_{\tau,t,m}^{\text{ZF,u}}\|_2^2+N_o\|\mathbf{f}_{\tau,t,m}^{\text{ZF,u}}\|_2^2-\left(\|\mathbf{v}_{\tau,t,m}^{\text{MMSE}}\|_2^2+N_o\|\mathbf{f}_{\tau,t,m}^{\text{MMSE}}\|_2^2\right)\\
&= N_o\mathbf{e}_{\tau,t,m}[(\mathbf{H}^H\mathbf{H})^{-1}-\mathbf{Q}]\mathbf{e}_{\tau,t,m}. && (101)
\end{aligned}
$$

[1]The notation >0 for a matrix means that the matrix is positive definite.

As $[(\mathbf{H}^H\mathbf{H})^{-1} - \mathbf{Q}] > 0$, we can conclude that the compound noise from the underdetermined ZF solution is larger than or equal to that of the MMSE solution.

Finally we consider the difference of

$$\begin{aligned} & \|\mathbf{v}_{\tau,t,m}^{\text{ZF,o}}\|_2^2 + N_o\|\mathbf{f}_{\tau,t,m}^{\text{ZF,o}}\|_2^2 - \left(\|\mathbf{v}_{\tau,t,m}^{\text{MMSE}}\|_2^2 + N_o\|\mathbf{f}_{\tau,t,m}^{\text{MMSE}}\|_2^2\right) \\ = & \; -\mathbf{e}_{\tau,t,m}\mathbf{H}^H[(\mathbf{H}^H\mathbf{H})^{-1}\overline{\mathbf{Q}}^{-1}(\mathbf{H}^H\mathbf{H})^{-1} - \overline{\mathbf{Q}}]\mathbf{H}\mathbf{e}_{\tau,t,m}. \end{aligned} \tag{102}$$

With the relation $(\mathbf{H}^H\mathbf{H})^{-1}\overline{\mathbf{Q}}^{-1}(\mathbf{H}^H\mathbf{H})^{-1} \geq (\mathbf{H}^H\mathbf{H})^{-1} > (\mathbf{H}^H\mathbf{H} + N_o\mathbf{I})^{-1} = \overline{\mathbf{Q}}$ we can conclude that the overdetermined ZF solution also results in an MSE larger than that of the MMSE solution.

C. Appendix C

As we assume $\mathbf{B}_k$ to be positive symmetric, we can split it into its roots: $\mathbf{B}_k = \mathbf{B}_k^{1/2}\mathbf{B}_k^{H/2}$. In this case we can substitute $\mathbf{X}_k\mathbf{B}_k^{1/2} = \tilde{\mathbf{X}}_k$. Starting with the initial parameter error vector $\tilde{\mathbf{f}}_0$ we find that

$$\begin{aligned} \tilde{\mathbf{f}}_k &= \tilde{\mathbf{f}}_{k-1} - \mu_k\mathbf{X}_k\mathbf{B}\left(\mathbf{X}_k^H\tilde{\mathbf{f}}_{k-1} + \bar{\mathbf{v}}_k\right) \qquad ; k = 1, 2, ..., \\ &= \tilde{\mathbf{f}}_{k-1} - \mu_k\tilde{\mathbf{X}}_k\left(\tilde{\mathbf{X}}_k^H\tilde{\mathbf{f}}_{k-1} + \mathbf{B}_k^{H/2}\bar{\mathbf{v}}_k\right). \end{aligned} \tag{103}$$

We define the undistorted a prior error vector $\mathbf{e}_{a,k} = \tilde{\mathbf{X}}_k^H\tilde{\mathbf{f}}_{k-1}$ and obtain the local relation

$$||\tilde{\mathbf{f}}_k||^2 + \mu_k||\mathbf{e}_{a,k}||^2 \leq ||\tilde{\mathbf{f}}_{k-1}||^2 + \mu_k||\mathbf{B}_k^{H/2}\bar{\mathbf{v}}_k||^2 \tag{104}$$

as long as $0 \leq \mu \leq \frac{1}{\lambda_{\max}\left(\mathbf{B}_k^{H/2}\tilde{\mathbf{X}}_k^H\tilde{\mathbf{X}}_k\mathbf{B}_k^{1/2}\right)}$. Summing up the terms, we find the global relation over a time horizon from 0 to K:

$$||\tilde{\mathbf{f}}_{K+1}||^2 + \sum_{k=1}^{K}\mu_k||\mathbf{e}_{a,k}||^2 \leq ||\tilde{\mathbf{f}}_0||^2 + \sum_{k=1}^{K}\mu_k||\mathbf{B}_k^{H/2}\bar{\mathbf{v}}_k||^2, \tag{105}$$

which already guarantees robustness and thus l_2-stability. If additionally the noise is bounded, i.e., $\sum_{k=1}^{K}\mu_k||\mathbf{B}_k^{H/2}\bar{\mathbf{v}}_k||^2 < S_v < \infty$, we can conclude convergence of $\sqrt{\mu_k}||\mathbf{e}_{a,k}|| \to 0$ as it must be a Cauchy sequence. Since $\mathbf{e}_{a,k} = \tilde{\mathbf{X}}_k^H\tilde{\mathbf{f}}_{k-1}$, we require a persistent excitation condition for $\sqrt{\mu_k}\tilde{\mathbf{X}}_k$ in order to conclude that the parameter error vector $\tilde{\mathbf{f}}_{k-1} \to \mathbf{0}$ for $\sqrt{\mu_k}||\mathbf{e}_{a,k}|| \to 0$.

REFERENCES

[1] E.D.Gibson and W.G.Burke, Jr., "Automatic equalization of standard voice telephone lines for digital data transmission," ACF Industries, Riverdale, Md., ACFE Rept. 88-0275-101, Aug. 1961.

[2] E.D.Gibson, "Automatic Equalization Using Transversal Equalizers," IEEE Transactions on Communication Technology, vol. 13, no. 3, pp. 380, Sep. 1965.

[3] M.Rappeport, "Automatic Equalization of Data Transmission Facility Distortion Using Transversal Equalizers," IEEE Transactions on Communication Technology, vol. 12, no. 3, pp. 65–73, Sep. 1964.

[4] R.W.Lucky, “Automatic equalization for digital communication,” in *Bell System Technical Journal,* vol. 44, pp. 547–588, April 1965.

[5] R.W.Lucky, “Techniques for adaptive equalization of digital communication systems,” in *Bell System Technical Journal,* vol. 45, pp. 255–286, Feb. 1966.

[6] M.E.Austin, “Decision feedback equalization for digital communication over dispersive channels,” Technical report 437, MIT Lincoln Lab, USA, Aug. 1967.

[7] D.George, R.Bowen, and J.Storey, “An Adaptive Decision Feedback Equalizer,” IEEE Transactions on Communication Technology, vol. 19, no. 3, pp. 281–293, June 1971.

[8] J.Salz, “Optimum mean square decision feedback equalization,” in *Bell System Technical Journal,* vol. 52, pp. 1341–1373, Oct. 1973.

[9] A.Gersho, “Adaptive equalization in highly dispersive channels for data transmission,” in *Bell System Technical Journal,* vol. 48, pp. 55–70, 1969.

[10] D.E.Brady, “An adaptive coherent diversity receiver for data transmission through dispersive media,” in Conference Record of ICC’70, pp.21.35-21.40, 1970.

[11] G.Ungerboeck, “Fractional tap-spacing equalizer and consequences for clock recovery in data modems,” IEEE Transactions on Communications, vol. COM-24, pp. 856–864, Aug. 1976.

[12] S.U.H.Qureshi and G.D.Fomey, Jr., “Performance and properties of a T/2 equalizer,” in Conference Record of National Telecommunication Conference, Los Angeles, CA, Dec. 1977, pp. 11.1.1–11.1.9.

[13] R.D.Gitlin and S.B.Weinstein, “Fractionally spaced equalization: An improved digital transversal equalizer,” Bell System Technical Journal, vol. 60, pp. 275-296, Feb. 1981.

[14] G.Forney, “Maximum Likelihood sequence estimation of digital sequences in the presence of intersymbol interference,” in *IEEE Transactions on Information Theory,* vol. 18, no. 3, pp. 363–378, May 1972.

[15] S.U.H.Qureshi, “Adaptive equalization,” IEEE Communications Magazine, vol. 20, no. 2, pp. 9–16 , March 1982.

[16] S.U.H.Qureshi, “Adaptive equalization,” Proceedings of the IEEE, vol. 73, no. 9, pp. 1349–1387, Sep. 1985.

[17] Y.Sato, “A method of self-recovering equalization for multilevel amplitude modulation systems,” IEEE Transactions on Communications, vol. COM-23, pp. 679–682, June 1975.

[18] D.N.Godard, “Self recovering equalization and carrier tracking in two-dimensional data communication systems,” IEEE Transactions on Communications, vol. COM-28, pp. 1867–1875, Nov. 1980.

[19] A.Benveniste and M.Goursat, “Blind Equalizers,” IEEE Transactions on Communications, vol. 32, no. 8, pp. 871–883, Aug. 1984.

[20] L.Tong, G.Xu, and T.Kailath, “A new approach to blind identification and equalization of multipath channels,” Conference Record of the 25th. Asilomar Conference of Signals, Systems and Computers, Monterey, USA, Nov. 1991.

[21] L.Tong and S.Perreau, “Multichannel blind channel estimation: From subspace to maximum likelihood methods,” Proceedings IEEE, vol. 86, pp.1951–1968, Oct.1998.

[22] J.M.Cioffi, G.Dudevoir, M.Eyuboglu, and G.D.Forney, Jr, “MMSE decision feedback equalization and coding-Part I, ” in *IEEE Transactions on Communications,* pp. 2582–2594, vol. 43, no. 10, Oct. 1995.

[23] N.Al-Dhahir and J.M.Cioffi, “MMSE Decision Feedback Equalizers: Finite Length Results, ” in *IEEE Transactions on Information Theory,* vol. 41, no. 4, pp. 961–975, July 1995.

[24] J.R.Treichler, I.Fijalkow, and C.R.Johnson, Jr., “Fractionally spaced equalizer. How long should they really be?,” *IEEE Signal Processing Magazine*, vol. 13, pp. 65–81, May 1996.

[25] A.Duel-Hallen, “Equalizers for multiple input/output channels and PAM systems with cyclostationary input sequences,” IEEE Journal

on Selected Areas in Communications, vol. 10, no. 3, pp. 630–639, April 1992.

[26] J.Yang and S.Roy, "Joint Transmitter and Receiver Optimization for Multiple-Input-Multiple-Output Systems with Decision Feedback," IEEE Transactions on Information Theory, pp. 1334–1347, Sep. 1994.

[27] I.Ghauri and D.Slock, "Linear Receiver for the CDMA Downlink Exploiting Orthogonality of Spreading Sequences," 32rd. Asilomar Conference, pp.650–655, vol.1, Nov. 98.

[28] M.Rupp, M.Guillaud, and S.Das, "On MIMO Decoding Algorithms for UMTS," Conference Record of the Thirty-Fifth Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, pp. 975–979, vol. 2, Nov. 2001.

[29] B.A.Bjerke and J.G.Proakis, "Equalization and Decoding for Multiple-Input Multiple-Output Wireless Channels," EURASIP Journal on Applied Signal Processing, no. 3, pp. 249–266, 2002.

[30] L.Mailaender, "Linear MIMO equalization for CDMA downlink signals with code reuse," IEEE Transactions on Wireless Communications, vol. 4, no. 5, pp. 2423–2434, Sept. 2005.

[31] E.A.Lee, D.G.Messerschmitt, Digital Communications, Kluver Academic Publisher, 1994.

[32] T.Rappaport, *Wireless Communications*, Prentice Hall, 1996.

[33] J.Proakis, Digital Communications, McGraw-Hill, 2000.

[34] S.Haykin, *Adaptive Filter Theory*, fourth edition, Prentice Hall, 2002.

[35] M.Rupp and A.Burg, *Chapter: Algorithms for Equalization in Wireless Applications* in Adaptive Signal Processing: Application to Real-World Problems, Springer, 2003.

[36] P.L.Feintuch, "An adaptive recursive LMS filter," Proc. IEEE, vol. 64, no. 11, pp. 1622-1624, Nov. 1976.

[37] C.R.Johnson, Jr. and M.G.Larimore, "Comments on and additions to "An adaptive recursive LMS filter"," Proc. IEEE, vol. 65, no.9, pp. 1399-1402, Sep. 1977.

[38] M.Rupp and A.H.Sayed, "On the Stability and Convergence of Feintuch's Algorithm for Adaptive IIR Filtering"; International Conference on Acoustics, Speech and Signal Processing, Detroit, MI, USA, May 1995, vol. 2, pp. 1388 - 1391.

[39] M.Rupp and A.H.Sayed, "A time-domain feedback analysis of filtered-error adaptive gradient algorithms," IEEE Transactions on Signal Processing, vol. 44, no. 6, pp. 1428-1439, June 1996.

[40] R.Dallinger and M.Rupp, "A Strict Stability Limit for Adaptive Gradient Type Algorithms," Proc. of Asilomar Conference on Signals, Systems, and Computers, 1-4. Nov. 2009.

[41] M.Rupp, "Pseudo Affine Projection Algorithms Revisited: Robustness and Stability Analysis," IEEE Transactions on Signal Procing, vol. 59, no. 5, pp. 2017 - 2023, May 2011.

[42] M.Rupp, "Convergence Properties of Adaptive Equalizer Algorithms," IEEE Trans. on Signal Processing, vol. 59, no. 6, pp. 2562 - 2574, June 2011.

[43] A.H.Sayed and M.Rupp, "Error-energy bounds for adaptive gradient algorithms," IEEE Transactions on Signal Processing, vol. 44, no. 8, pp. 1982-1989, Aug. 1996.

[44] A.H.Sayed and M.Rupp, *Chapter: Robustness Issues in Adaptive Filters* in The DSP Handbook, CRC Press, 1998.

[45] M.Tsatsanis and Z.Xu,"Performance Analysis of Minimum Variance CDMA Receivers," IEEE Transactions on Signal Processing, vol. 46, no. 11, Nov. 1998

[46] R.A.Monzingo and T.W.Miller, Introduction of Adaptive Arrays. New York: Wiley, 1980.

[47] D.A.Schmidt, C.Shi, R.A.Berry, M.L.Honig, and W.Utschick, "Minimum Mean Squared Error Interference Alignment," 43th Asilomar Conference of Signals, Systems and Computers, Monterey, USA, 2009.

[48] M.Wrulich, PhD thesis at TU Vienna: “System-Level Modeling and Optimization of MIMO HSDPA Networks”, http://publik.tuwien.ac.at/files/PubDat_181165.pdf

[49] S.Verdu, Multiuser Detection. Cambridge University Press, 1998.

[50] M.Wrulich, C.Mehlführer, and M.Rupp, “Interference aware MMSE equalization for MIMO TxAA,” in Proc. IEEE 3rd International Symposium on Communications, Control and Signal Processing (ISCCSP), 2008, pp. 1585–1589.

[51] C.Mehlführer, M.Wrulich, and M.Rupp, “Intra-cell interference aware equalization for TxAA HSDPA,” in Proc. IEEE 3rd International Symposium on Wireless Pervasive Computing, 2008, pp. 406–409.

[52] M.Wrulich, C.Mehlführer, and M.Rupp, “Managing the Interference Structure of MIMO HSDPA: A Multi-User Interference Aware MMSE Receiver with Moderate Complexity,” IEEE Trans. on Wireless Communications, vol.9, no.4, pp. 1472 - 1482, April 2010.

[53] C.Dumard, F.Kaltenberger, K.Freudenthaler, ”Low-Cost Approximate LMMSE Equalizer Based on Krylov Subspace Methods for HSDPA,” IEEE Transactions on Wireless Communications, vol.6, no.5, pp. 1610-1614, May 2007.

[54] M.Rupp, “On gradient type adaptive filters with non-symmetric matrix step-sizes,” in Proc. of ICASSP, Prague, Czech Republic, May 2011.

[55] S.Shaffer and C.S.Williams, “Comparison of LMS, alpha LMS, and Data Reusing LMS Algorithms,” Proc. of Asilomar Conference, Nov. 1983, pp. 260-264.

[56] S.Roy and J.J.Shynk, “Analysis of the data-reusing LMS algorithm,” Proc. of the 32nd Midwest Symposium on Circuits and Systems, vol. 2, Aug. 1989, pp. 1127-1130.

[57] M.L.R.de Campos, P.S.R.Diniz, and J.A.Apolinario Jr., “On normalized data-reusing and affine-projections algorithms,” Proc. of the 6th IEEE International Conference on Electronics, Circuits and Systems, vol. 2, Sep. 1999, pp. 843-846.

[58] P.S.R.Diniz and S.Werner, “Set-membership binormalized data-reusing LMS algorithms,” IEEE Transactions on Signal Processing, vol. 51, no. 1, pp. 124-134, Jan. 2003.

[59] H.-C.Shin, W.-J.Song, and A.H.Sayed, “Mean-square performance of data-reusing adaptive algorithms,” IEEE Signal Processing Letters, vol. 12, no. 12, Dec. 2005, pp. 851-854.

[60] H.-C.Chen and O.T.-C.Chen, “Convergence performance analyses of fast data-reusing normalized least mean squared algorithm,” IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences, vol. E90-A, no.1, pp. 249-256, Jan. 2007.

[61] T.Kailath, Linear Systems, Prentice Hall, NJ, 1980.

[62] Y.Huang and J.Benesty, “Adaptive multi-channel least mean square and Newton algorithms for blind channel identification,” Signal Processing, vol. 82, pp. 1127-1138, Aug. 2002.

[63] M.Rupp and A.H.Sayed, “On the convergence of blind adaptive equalizers for constant modulus signals,” IEEE Transactions on Communications, vol. 48, no. 5, pp. 795-803, May 2000.

[64] S.M.Kay, Fundamentals of Statistical Signal Processing: Estimation Theory. Prentice Hall, 1993.

Markus Rupp received his Dipl.-Ing. degree in 1988 at the University of Saarbrücken, Germany and his Dr.-Ing. degree in 1993 at the Technische Universität Darmstadt, Germany, where he worked with Eberhardt Hänsler on designing new algorithms for acoustical and electrical echo compensation. From November 1993 until July 1995, he had a postdoctoral position at the University of Santa Barbara, California with Sanjit Mitra where he worked with Ali H. Sayed on a robustness description of adaptive filters with impact on neural networks and active noise control. From October 1995 until August 2001 he was a member of Technical Staff in the Wireless Technology Research Department of Bell-Labs at Crawford Hill, NJ, where he worked on various topics related to adaptive equalization and rapid implementation for IS-136, 802.11 and UMTS. Since October

2001 he is a full professor for Digital Signal Processing in Mobile Communications at the Vienna University of Technology where he founded the Christian-Doppler Laboratory for Design Methodology of Signal Processing Algorithms in 2002 at the Institute for Communications and RF Engineering (now institute of telecommunications). He served as Dean from 2005-2007. He was associate editor of IEEE Transactions on Signal Processing from 2002-2005, is currently associate editor of JASP EURASIP Journal of Advances in Signal Processing, and JES EURASIP Journal on Embedded Systems. He is elected AdCom member of EURASIP since 2004 and serving as president of EURASIP from 2009-2010. He authored and co-authored more than 400 papers and patents on adaptive filtering, wireless communications, and rapid prototyping, as well as automatic design methods.