Convergence Properties of Adaptive Equalizer Algorithms

Markus Rupp, *Senior Member, IEEE*

***Abstract*—In this paper, we provide a thorough stability analysis of two well known adaptive algorithms for equalization based on a novel least squares reference model that allows to treat the equalizer problem equivalently as system identification problem. While not surprising the adaptive minimum mean-square error (MMSE) equalizer algorithm behaves l_2–stable for a wide range of step-sizes, the even older zero-forcing (ZF) algorithm however behaves very differently. We prove that the ZF algorithm generally does not belong to the class of robust algorithms but can be convergent in the mean square sense. We furthermore provide conditions on the upper step-size bound to guarantee such mean squares convergence. We specifically show how noise variance of added channel noise and the channel impulse response influences this bound. Simulation examples validate our findings.**

***Index Terms*—Adaptive gradient type filters, error bounds, l_2-stability, mean-square-convergence, mismatch, robustness, zero forcing.**

I. Introduction

Modern digital receivers in wireless and cable-based systems are not considerable without equalizers in some form. The first mentioning of digital equalizers was by Lucky [1], [2] in 1965 and 1966 at the Bell System Technical Journal, who also coined the expression "zero forcing" (ZF). Correspondingly, the MMSE formulation was provided by Gersho [3]. Further milestone papers are by Forney [4], Cioffi *et al.* [5], Al-Dhahir *et al.* [6] as well as Treichler *et al.* [7]. Good overviews in adaptive equalization are provided in [8]–[11] and [12].

Consider the following transmission over a time dispersive (frequency selective) channel model:

$$\mathbf{r}_k = \mathbf{H}\mathbf{s}_k + \mathbf{v}_k. \quad (1)$$

Manuscript received July 15, 2010; revised October 09, 2010 and December 22, 2010; accepted February 18, 2011. Date of publication March 03, 2011; date of current version May 18, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Isao Yamada. This work has been funded by the NFN SISE project S10609 (National Research Network Signal and Information Processing in Science and Engineering). Section IV of the paper appears as a conference paper entitled "On Gradient Type Adaptive Filters with Non-Symmetric Matrix Step-Sizes" in the *Proceedings of the International Conference on Acoustics, Speech and Signal Processing* (ICASSP), Prague, Czech Republic, May 22–27, 2011.

The author is with the Vienna University of Technology, Institute of Telecommunications, 1040 Vienna, Austria (e-mail: mrupp@nt.tuwien.ac.at).

This paper has supplementary downloadable multimedia material available at http://ieeexplore.ieee.org provided by the authors. This includes all Matlab code, which shows the simulation examples. This material is 11 kB in size.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2011.2121905

TABLE I
Overview of Most Common Variables

Variable	Dimension	Name
$\mathbf{s}_k$	$\mathbb{C}^{S\times 1}$	transmit symbol vector
$\mathbf{r}_k$	$\mathbb{C}^{R\times 1}$	receive symbol vector
$\mathbf{H}$	$\mathbb{C}^{R\times S}$	channel matrix
$\mathbf{e}_\tau$	$\mathbb{R}^{S\times 1}$	unit vector, "1" at position τ
$\mathbf{f}_{\text{ZF}}$	$\mathbb{C}^{R\times 1}$	ZF solution
$\mathbf{f}_{\text{MMSE}}$	$\mathbb{C}^{R\times 1}$	MMSE solution
$\hat{\mathbf{f}}_k$	$\mathbb{C}^{R\times 1}$	equalizer estimate
$\tilde{\mathbf{f}}_k$	$\mathbb{C}^{R\times 1}$	equalizer estimation error
$\mathbf{v}_{\text{ZF}}$	$\mathbb{C}^{S\times 1}$	ZF modelling noise
$\mathbf{v}_{\text{MMSE}}$	$\mathbb{C}^{S\times 1}$	MMSE modelling noise
$\bar{v}_{\text{ZF},k}$	$\mathbb{C}$	ZF compound noise
$\bar{v}_{\text{MMSE},k}$	$\mathbb{C}$	MMSE compound noise
$\tilde{w}_{\text{ZF},k}$	$\mathbb{C}$	ZF algorithmic compound noise

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

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

Throughout this paper, we will assume that transmit signals s_k have unit energy that is $\mathbb{E}[|s_k|^2] = 1$, and the noise variance is given by $\mathbb{E}[|v_k|^2] = N_o$ without loss of generality. Note that for a Toeplitz form channel $\mathbf{H}$ we have $R \leq S$ which refers to a Single-Input Single-Output (SISO) case. This contribution only focuses on the SISO model in (1); extensions towards multiantenna (MIMO) and/or multiuser (MU) scenarios will be treated elsewhere. A linear equalizer applies an FIR filter $\mathbf{f}$ on received signal $\mathbf{r}_k$ so that $\hat{s}_{k-\tau} = \mathbf{f}^H\mathbf{r}_k$ is an estimate of $s_{k-\tau}$ for some delayed version of s_k. The optimal selection of τ will not be treated here.

Table I provides an overview of the most common variables and their dimensions. In Section II we introduce a reference model for linear equalizers based on a deterministic least squares approach. We thus avoid the commonly used MSE approach. Section III formulates the problem in terms of adaptive equalizers. While the MMSE adaptive equalizer is straightforward to analyze, Lucky's classical adaptive ZF equalizer turns out to be more difficult. Here, we have to solve two problems that is adaptive filters with arbitrary matrices as step-sizes and parameter error terms appearing in the additive noise term which is addressed in Sections IV and V, respectively. With this new knowledge we return to the original problem and finally analyze the adaptive ZF equalizer algorithm in Section VI. Simulation results support our findings. A conclusion is provided in Section VII.

II. A Reference Model for Equalization

A. ZF Equalizer

A solution to the ZF equalizer problem is equivalently given by the following least-squares formulation:

$$\begin{aligned}\mathbf{f}_{\mathrm{ZF}} &= \arg\min_{\mathbf{f}} \left\|\mathbf{H}^H\mathbf{f}-\mathbf{e}_\tau\right\|_2^2 \\ &= \arg\min_{\mathbf{f}} \left\|\mathbf{H}^H\left[\mathbf{f}-(\mathbf{H}\mathbf{H}^H)^{-1}\mathbf{H}\mathbf{e}_\tau\right]\right\|_2^2 \end{aligned} \quad (3)$$

with $\mathbf{e}_\tau$ indicating a unit vector with a single one entry at position τ, thus $\mathbf{e}_\tau^T\mathbf{s}_k = s_{k-\tau}$. The resulting LS solution is called the ZF solution $\mathbf{f}_{\mathrm{ZF}} \in \mathbb{C}^{R\times 1}$. Note that this form of derivation does not require signal or noise information but focuses only on properties of linear time-invariant systems of finite length (FIR); it thus ignores the presence of noise entirely. This is identical to the original formulations by Lucky [1], [2] where system properties were the focus.

In a SISO scenario, we have $R < S$ and the solution to this problem is obviously given by

$$\mathbf{f}_{\mathrm{ZF}} = (\mathbf{H}\mathbf{H}^H)^{-1}\mathbf{H}\mathbf{e}_\tau. \quad (4)$$

As the ZF solution leads to ISI for finite length vectors, we propose the following reference model for ZF equalizers

$$\mathbf{e}_\tau = \mathbf{H}^H\mathbf{f}_{\mathrm{ZF}} + \mathbf{v}_{\mathrm{ZF}} \quad (5)$$

with modeling noise $\mathbf{v}_{\mathrm{ZF}} \in \mathbb{C}^{S\times 1}$

$$\mathbf{v}_{\mathrm{ZF}} = \left(\mathbf{I}-\mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1}\mathbf{H}\right)\mathbf{e}_\tau. \quad (6)$$

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

$$\mathbf{H}\mathbf{v}_{\mathrm{ZF}} = \mathbf{0} \quad (7)$$

$$\mathbf{f}_{\mathrm{ZF}}^H\mathbf{H}\mathbf{v}_{\mathrm{ZF}} = 0 \quad (8)$$

$$\mathbf{v}_{\mathrm{ZF}}^H\mathbf{v}_{\mathrm{ZF}} = \mathbf{v}_{\mathrm{ZF}}^H\mathbf{e}_\tau = \mathbf{e}_\tau^T\left(\mathbf{I}-\mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1}\mathbf{H}\right)\mathbf{e}_\tau. \quad (9)$$

The last term can be interpreted as the energy of the modeling error but equally describes the remaining ISI power.

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

$$\begin{aligned}\mathbf{f}_{\mathrm{ZF}}^H\mathbf{r}_k &= s_{k-\tau} - \mathbf{v}_{\mathrm{ZF}}^H\mathbf{s}_k + \mathbf{f}_{\mathrm{ZF}}^H\mathbf{v}_k \quad (10)\\ &= s_{k-\tau} + \bar{v}_{\mathrm{ZF},k}. \quad (11)\end{aligned}$$

Such relation serves as SISO ZF reference model as we will apply it to adaptive algorithms further ahead. Note that often ISI as well as additive noise is treated equivalently as a compound noise $\bar{v}_{\mathrm{ZF},k}$ as indicated in (11).

B. MMSE Equalizer

Correspondingly, the well-known MMSE solution $\mathbf{f}_{\mathrm{MMSE}} \in \mathbb{C}^{R\times 1}$ can be obtained including noise variance N_o from a spectrally white noise

$$\mathbf{f}_{\mathrm{MMSE}} = (\mathbf{H}\mathbf{H}^H + N_o\mathbf{I})^{-1}\mathbf{H}\mathbf{e}_\tau \quad (12)$$

as a solution of the weighted LS problem

$$\begin{aligned}\mathbf{f}_{\mathrm{MMSE}} &= \arg\min_{\mathbf{f}}\left(\left\|\mathbf{H}^H\mathbf{f}-\mathbf{e}_\tau\right\|_2^2 + N_o\|\mathbf{f}\|_2^2\right)\\ &= \arg\min_{\mathbf{f}}\left\|(\mathbf{H}\mathbf{H}^H+N_o\mathbf{I})^{\frac{1}{2}}\times\left[\mathbf{f}-(\mathbf{H}\mathbf{H}^H+N_o\mathbf{I})^{-1}\mathbf{H}\mathbf{e}_\tau\right]\right\|_2^2\\ &\quad + \mathrm{MMSE}.\end{aligned} \quad (13)$$

Correspondingly to the reference model for ZF equalizers in (5), we can now also define a reference model for MMSE equalizers

$$\mathbf{e}_\tau = \mathbf{H}^H\mathbf{f}_{\mathrm{MMSE}} + \mathbf{v}_{\mathrm{MMSE}} \quad (14)$$

with modeling noise $\mathbf{v}_{\mathrm{MMSE}} \in \mathbb{C}^{S\times 1}$

$$\mathbf{v}_{\mathrm{MMSE}} = \left(\mathbf{I}-\mathbf{H}^H(\mathbf{H}\mathbf{H}^H+N_o\mathbf{I})^{-1}\mathbf{H}\right)\mathbf{e}_\tau. \quad (15)$$

Note, however, that different to the ZF solution the modeling error is not orthogonal to the MMSE solution, that is $\mathbf{v}_{\mathrm{MMSE}}^H\mathbf{H}^H\mathbf{f}_{\mathrm{MMSE}} \neq 0$. Multiplying the signal vector with $\mathbf{e}_\tau$ we obtain

$$\mathbf{e}_\tau^T\mathbf{s}_k = s_{k-\tau} = \mathbf{f}_{\mathrm{MMSE}}^H\mathbf{H}\mathbf{s}_k + \mathbf{v}_{\mathrm{MMSE}}^H\mathbf{s}_k. \quad (16)$$

How does a received signal look after such MMSE-equalization? We apply $\mathbf{f}_{\mathrm{MMSE}}$ on the observation vector and obtain analogously to (11):

$$\mathbf{f}_{\mathrm{MMSE}}^H\mathbf{r}_k = s_{k-\tau} - \mathbf{v}_{\mathrm{MMSE}}^H\mathbf{s}_k + \mathbf{f}_{\mathrm{MMSE}}^H\mathbf{v}_k = s_{k-\tau} + \bar{v}_{\mathrm{MMSE},k}. \quad (17)$$

Note that compound noise $\bar{v}_{\mathrm{MMSE},k}$ is now different when compared to the ZF solution in (11).

III. Adaptive Equalizers

A. Recursive MMSE Algorithm

We start with the classical adaptive MMSE equalizer as it is much easier to analyze than its ZF counterpart. Such algorithm is also known under the name least-mean-square (LMS) algorithm for equalization (see, for example, Rappaport [8]). Recursive algorithms try to improve their estimation in the presence of new data that has not been applied before. Due to this property they typically behave adaptively as well. This allows the reinterpretation of the algorithm in terms of a recursive procedure in which new data are being processed at every time instant k. Starting with some initial value $\hat{\mathbf{f}}_0$, the equalizer estimate $\hat{\mathbf{f}}_k \in \mathbb{C}^{R\times 1}$ reads

$$\hat{\mathbf{f}}_k = \hat{\mathbf{f}}_{k-1} + \mu_k \mathbf{r}_k (s^*_{k-\tau} - \mathbf{r}_k^H \hat{\mathbf{f}}_{k-1}) \quad ; k = 1, 2, \ldots. \tag{18}$$

In order to perform a stability analysis we introduce parameter error vector $\tilde{\mathbf{f}}_k = \mathbf{f}_{\mathrm{MMSE}} - \hat{\mathbf{f}}_k$ as well as the reference model (17) and obtain

$$\begin{aligned}\tilde{\mathbf{f}}_k &= \tilde{\mathbf{f}}_{k-1} - \mu_k \mathbf{r}_k (\mathbf{r}_k^H \tilde{\mathbf{f}}_{k-1} - \bar{v}^*_{\mathrm{MMSE},k}) \\ &= \left(\mathbf{I} - \mu_k \mathbf{r}_k \mathbf{r}_k^H\right) \tilde{\mathbf{f}}_{k-1} + \mu_k \mathbf{r}_k \bar{v}^*_{\mathrm{MMSE},k}. \end{aligned} \tag{19}$$

The description of the MMSE equalizer is thus identical to a classical system identification problem. As we know the behavior of a standard LMS algorithm for such circumstances we can deduce immediately the results from there for our MMSE equalizer problem. While results concerning the classical convergence in the mean square sense are already known [3], new results are possible concerning the adaptive filter misadjustment and relative system mismatch due to the knowledge of compound noise term $\bar{v}_{\mathrm{MMSE},k}$. We will show this further ahead in Section VI with some simulation results.

Entirely new is the convergence of these adaptive equalizers in the l_2–sense. According to [13], [14], and [15, Ch. 17], the LMS algorithm can be described in terms of robustness, showing l_2–stability. Different to the classic approaches in which the driving signals are stochastic processes, a robustness description does not require any statistic for the signals or noise. In fact, the l_2–stability is guaranteed for any driving signal and noise sequence. With this new interpretation of an adaptive MMSE equalizer as system identification problem, we can thus directly adapt the results from literature and state the following for the adaptive MMSE equalizer algorithm.

Theorem 3.1: The adaptive MMSE equalizer with Update (18) is l_2-stable from its input uncertainties $\{\tilde{\mathbf{f}}_0, \{\sqrt{\mu_k}\bar{v}_{\mathrm{MMSE},k}\}_{1<k<N}\}$ to its output errors $\{\sqrt{\mu_k}\mathbf{r}_k^H\tilde{\mathbf{f}}_{k-1}\}_{1<k<N}$ for a constant step-size if

$$0 < \mu < \min_k \frac{2}{\|\mathbf{r}_k\|_2^2} \tag{20}$$

and for a time-variant step-size μ_k if

$$0 < \mu_k < \frac{2}{\|\mathbf{r}_k\|_2^2}. \tag{21}$$

We like to note that the derivation of the theorem can be found in [13] in which the small gain theorem was applied. The bounds are thus conservative and not necessarily tight. A recent discussion on this is presented in [16]. Further note that the statement in Theorem 3.1 relates to the convergence of the undistorted *a priori* error $\mathbf{r}_k^H\tilde{\mathbf{f}}_{k-1}$. If the noise energy is bounded, so is the *a priori* error energy and thus for $N \to \infty$ the *a priori* error needs to converge to zero (Cauchy series). In order to conclude that parameter error $\tilde{\mathbf{f}}_k$ converges to zero, a further persistent excitation argument on the observed data $\mathbf{r}_k$ is required [15], [17].

B. Recursive ZF Algorithm

The original ZF algorithm [2] is given by the following update equation, starting with some initial value $\hat{\mathbf{f}}_0$ to estimate $\hat{\mathbf{f}}_k \in \mathbb{C}^{R\times 1}$

$$\hat{\mathbf{f}}_k = \hat{\mathbf{f}}_{k-1} + 2\mu_k \mathbf{P}\mathbf{s}_k (s^*_{k-\tau} - \mathbf{r}_k^H \hat{\mathbf{f}}_{k-1}) \quad ; k = 1, 2, \ldots. \tag{22}$$

Matrix $\mathbf{P} = [\mathbf{I}, \mathbf{0}] \in \mathbb{R}^{R\times S}$ is required to shorten the long vector $\mathbf{s}_k$ in (1) from S to its length R. We have deliberately selected a step-size $2\mu_k$ as will become clear later.

Applying the same method as before, introducing parameter error vector $\tilde{\mathbf{f}}_k = \mathbf{f}_{\mathrm{ZF}} - \hat{\mathbf{f}}_k$ and the ZF reference model (11), we now obtain

$$\begin{aligned}\tilde{\mathbf{f}}_k &= \tilde{\mathbf{f}}_{k-1} - 2\mu_k \mathbf{P}\mathbf{s}_k (\mathbf{r}_k^H \tilde{\mathbf{f}}_{k-1} + \mathbf{s}_k^H \mathbf{v}_{\mathrm{ZF}} - \mathbf{v}_k^H \mathbf{f}_{\mathrm{ZF}}) \\ &= \tilde{\mathbf{f}}_{k-1} - 2\mu_k \mathbf{P}\mathbf{s}_k (\mathbf{s}_k^H \mathbf{H}^H \tilde{\mathbf{f}}_{k-1} + \bar{w}_{\mathrm{ZF},k}). \end{aligned} \tag{23}$$

Here, we abbreviated

$$\bar{w}_{\mathrm{ZF},k} = \mathbf{v}_k^H \tilde{\mathbf{f}}_{k-1} + \mathbf{s}_k^H \mathbf{v}_{\mathrm{ZF}} - \mathbf{v}_k^H \mathbf{f}_{\mathrm{ZF}} \tag{24}$$

$$= \mathbf{v}_k^H \tilde{\mathbf{f}}_{k-1} - \bar{v}^*_{\mathrm{ZF},k} \tag{25}$$

an algorithmic compound noise that contains an additional component when compared of the corresponding ZF value $\bar{v}_{\mathrm{ZF},k}$ in (11), depending on the parameter error vector itself. We split excitation vector $\mathbf{s}_k$ into two parts:

$$\mathbf{s}_k^T = [\mathbf{s}_{u,k}^T, \mathbf{s}_{l,k}^T] = [\mathbf{s}_k^T \mathbf{P}^T, \mathbf{s}_k^T \mathbf{Q}^T] \tag{26}$$

$$\mathbf{Q} = [\mathbf{0}, \mathbf{I}] \in \mathbb{R}^{R\times S} \tag{27}$$

$$\mathbf{H} = \begin{bmatrix} \underbrace{\begin{matrix} h_0 & h_1 & \ldots & h_{R-1} \\ & \ddots & \ddots & \\ & & h_0 & h_1 \\ & & & h_0 \end{matrix}}_{\mathbf{H}_u} & \underbrace{\begin{matrix} h_R & \ldots & h_{L-1} & & & \\ \ddots & \ddots & & \ddots & & \\ h_2 & \ldots & h_R & \ldots & h_{L-1} & \\ h_1 & \ldots & h_{R-1} & h_R & \ldots & h_{L-1} \end{matrix}}_{\mathbf{H}_l} \end{bmatrix} \tag{28}$$

as well as the Toeplitz matrix [compare with (2)] shown in (28) at the bottom of the previous page for $R < L$.

$$\mathbf{H}_u = \mathbf{H}\mathbf{P}^T \tag{29}$$
$$\mathbf{H}_l = \mathbf{H}\mathbf{Q}^T. \tag{30}$$

We can thus also split the inner vector product

$$\mathbf{s}_k^H \mathbf{H}^H \tilde{\mathbf{f}}_{k-1} = \mathbf{s}_{u,k}^H \mathbf{H}_u^H \tilde{\mathbf{f}}_{k-1} + \mathbf{s}_{l,k}^H \mathbf{H}_l^H \tilde{\mathbf{f}}_{k-1}. \tag{31}$$

By this splitting operation, we obtain an upper triangular matrix $\mathbf{H}_u$ as shown in (28). We assume that the cursor index is selected in such a way that the main diagonal is not filled with zeros. As long as this property is satisfied, the matrix is regular and its inverse exists.

With such reformulation we find now the update equation to be

$$\begin{aligned}\tilde{\mathbf{f}}_k &= \tilde{\mathbf{f}}_{k-1} - 2\mu_k \mathbf{P}\mathbf{s}_k(\mathbf{s}_k^H \mathbf{H}^H \tilde{\mathbf{f}}_{k-1} + \tilde{w}_{\mathrm{ZF},k}) \\ &= \tilde{\mathbf{f}}_{k-1} - 2\mu_k \mathbf{s}_{u,k}(\mathbf{s}_{u,k}^H \mathbf{H}_u^H \tilde{\mathbf{f}}_{k-1} + \tilde{w}_{\mathrm{ZF},k}). \\ \tilde{w}_{\mathrm{ZF},k} &= [\mathbf{s}_{l,k}^H \mathbf{H}_l^H + \mathbf{v}_k^H]\tilde{\mathbf{f}}_{k-1} + \mathbf{s}_k^H \mathbf{v}_{\mathrm{ZF}} - \mathbf{v}_k^H \mathbf{f}_{\mathrm{ZF}}. \end{aligned} \tag{32}$$

The compound noise now takes on an additional component that is also dependent on the parameter error vector.

Comparing the algorithm with a standard LMS algorithm with matrix step-size [17]–[20]

$$\tilde{\mathbf{w}}_k = \tilde{\mathbf{w}}_{k-1} - 2\mu_k \mathbf{G}\mathbf{x}_k(\mathbf{x}_k^H \tilde{\mathbf{w}}_{k-1} + \tilde{w}_{\mathrm{ZF},k}) \tag{33}$$
$$= \tilde{\mathbf{w}}_{k-1} - 2\mu_k \mathbf{G}\mathbf{x}_k \tilde{e}_{a,k} \tag{34}$$

we identify $\mathbf{x}_k = \mathbf{H}_u \mathbf{s}_{u,k}$ and $\mathbf{G}\mathbf{x}_k = \mathbf{s}_{u,k}$. As $\mathbf{H}_u$ is an upper triangular matrix we can expect it to be regular, provided that the cursor position is chosen correctly and thus $\mathbf{G} = \mathbf{H}_u^{-1}$. The parameter error vector is simply given by $\tilde{\mathbf{w}}_k = \tilde{\mathbf{f}}_k$. The distorted *a priori* error term

$$\tilde{e}_{a,k} = e_{a,k} + v_k \tag{35}$$

comprises of an undistorted *a priori* term $e_{a,k} = \mathbf{x}_k^H \tilde{\mathbf{w}}_{k-1}$ and additive noise $v_k = \tilde{w}_{\mathrm{ZF},k}$. We follow here mostly the notation of [15].

Alternatively, we can premultiply (32) by $\mathbf{H}_u^H$ from the left and use the substitution $\tilde{\mathbf{g}}_k = \mathbf{H}_u^H \tilde{\mathbf{f}}_k$, resulting in an update form in $\tilde{\mathbf{g}}_k$ rather than in $\tilde{\mathbf{f}}_k$. In this case we identify $\mathbf{x}_k = \mathbf{s}_{u,k}$ and $\mathbf{G} = \mathbf{H}_u^H$. This alternative form will be even more useful as typically driving process s_k is white and thus corresponding terms become much easier to compute. Details will follow further ahead in Section VI.

We have now reformulated Lucky's original ZF algorithm into an LMS algorithm with two unusual features:

1) a nonsymmetric matrix step-size $\mu_k \mathbf{G}$;
2) a noise term $\tilde{w}_{\mathrm{ZF},k}$ that depends on the parameter error vector itself.

In order to proceed with the algorithmic analysis, we first have to address both effects.

IV. Adaptive Gradient Algorithms With Nonsymmetric Matrix Step-Size

As the content of this section can be treated independently of the equalizer context, we will present the results in different notation. Later, we simply substitute the terms as indicated in the last discussion of the previous section. We thus will deliberately select now $\tilde{\mathbf{w}}_k$ as parameter error vector (and not $\tilde{\mathbf{f}}_k$) as later linearly transformed versions will be applied to describe the ZF algorithm leading to $\tilde{\mathbf{f}}_k$ as well as the true compound noise $\tilde{w}_{\mathrm{ZF},k}$ rather than the additive noise v_k.

We start with an LMS algorithm with matrix step-size $\mu_k \mathbf{G} \in \mathbb{C}^{M \times M}$ in the context of system identification for which we assume a reference system exists with additively disturbed output

$$d_k = \mathbf{x}_k^H \mathbf{w}_o + v_k, \tag{36}$$

d_k denoting the observed output of the reference $\mathbf{w}_o \in \mathbb{C}^{M \times 1}$, and $\mathbf{x}_k^T = [x_k, x_{k-1}, \ldots, x_{k-M+1}] \in \mathbb{C}^{M \times 1}$ being its input (or sometimes called regression) vector, $x_k \in \mathbb{C}$ the driving sequence and $v_k \in \mathbb{C}$ additive noise. We subtract the true solution $\mathbf{w}_o$ from its estimate and use only the parameter error vector $\tilde{\mathbf{w}}_k = \mathbf{w}_o - \hat{\mathbf{w}}_k$ from now on:

$$\tilde{\mathbf{w}}_k = \tilde{\mathbf{w}}_{k-1} - 2\mu_k \mathbf{G}\mathbf{x}_k(d_k - \mathbf{x}_k^H \hat{\mathbf{w}}_{k-1}) \tag{37}$$
$$= \tilde{\mathbf{w}}_{k-1} - 2\mu_k \mathbf{G}\mathbf{x}_k \underbrace{(\mathbf{x}_k^H \tilde{\mathbf{w}}_{k-1} + v_k)}_{\tilde{e}_{a,k} = e_{a,k} + v_k}. \tag{38}$$

While there exists convergence results in the mean-square sense [19], [20] and in the l_2 sense [17] for symmetric matrix step-sizes, there is none for nonsymmetric matrix step-sizes. In this section, we will address this issue with various novel ideas (Method A,B,C). We follow here a classical l_2 analysis path [13]–[15], [17] and show some weak so-called *local* robustness properties when considering the adaptation from time instant $k-1$ to k. However, it will turn out that such algorithms are not as robust as the LMS algorithm with symmetric matrix step-size. We will therefore at a certain point of the analysis have to leave this path of robustness and employ statistical properties of the input signals and noise. We thus will not be able to provide strict l_2 stability conditions for the algorithm but instead convergence in the mean square sense. Note that MSE convergence could also be shown with simpler techniques (for example, [21]) but it would be tedious, require a lot more of simplifying assumptions, and only provide loose step-size bounds, while our analysis is much more rigorous and will provide very tight bounds as we will show by simulation examples. We will even prove further ahead in the following section that the ZF equalizer algorithm indeed does not belong to the class of robust algorithms.

A. Analysis Method A

We introduce an additional square matrix $\mathbf{F} \in \mathbb{C}^{M \times M}$ that we multiply from the left to obtain a modified update equation

$$\mathbf{F}\tilde{\mathbf{w}}_k = \mathbf{F}\tilde{\mathbf{w}}_{k-1} - 2\mu_k \mathbf{F}\mathbf{G}\mathbf{x}_k \tilde{e}_{a,k}; k = 1, 2, \ldots. \tag{39}$$

A straightforward idea is now to compute parameter vector error energy $\|\tilde{\mathbf{w}}_k\|_2^2$ in the light of such matrix $\mathbf{F}$ that is

$$\|\tilde{\mathbf{w}}_k\|_{\mathbf{F}}^2 = \tilde{\mathbf{w}}_k^H \mathbf{F}^H \mathbf{F} \tilde{\mathbf{w}}_k. \tag{40}$$

This weighted square Euclidean norm requires that $\mathbf{F}^H\mathbf{F} > 0$ is being positive definite or in this case equivalently $\mathbf{F}^H\mathbf{F}$ is being of full rank, which is a first condition and restriction on $\mathbf{F}$. We thus obtain

$$\|\tilde{\mathbf{w}}_k\|_{\mathbf{F}}^2 = \|\tilde{\mathbf{w}}_{k-1}\|_{\mathbf{F}}^2 - 2\mu_k \bar{e}_{A,k}^* \tilde{e}_{a,k} - 2\mu_k \bar{e}_{A,k} \tilde{e}_{a,k}^* + 4\frac{\mu_k^2}{\bar{\mu}_{A,k}} |\tilde{e}_{a,k}|^2 \tag{41}$$

where we employed the following notation:

$$\bar{e}_{A,k} = \mathbf{x}_k^H \mathbf{G}^H \mathbf{F}^H \mathbf{F} \tilde{\mathbf{w}}_{k-1} \tag{42}$$

$$\bar{\mu}_{A,k} = \frac{1}{\mathbf{x}_k^H \mathbf{G}^H \mathbf{F}^H \mathbf{F} \mathbf{G} \mathbf{x}_k}. \tag{43}$$

We introduce a proportionality factor $\lambda_{A,k}$ such that

$$\bar{e}_{A,k} = \lambda_{A,k}^* e_{a,k} \tag{44}$$

which allows the simplification of previous (41) into

$$\|\tilde{\mathbf{w}}_k\|_{\mathbf{F}}^2 = \|\tilde{\mathbf{w}}_{k-1}\|_{\mathbf{F}}^2 + 4\frac{\mu_k^2}{\bar{\mu}_{A,k}} |\tilde{e}_{a,k}|^2 - 2\mu_k \left[\lambda_{A,k} e_{a,k}^* \tilde{e}_{a,k} + \lambda_{A,k}^* e_{a,k} \tilde{e}_{a,k}^* \right]. \tag{45}$$

Next to additive noise v_k we can now form a new variable u_k:

$$|v_k|^2 = |\tilde{e}_{a,k} - e_{a,k}|^2 \tag{46}$$
$$= |\tilde{e}_{a,k}|^2 + |e_{a,k}|^2 - \tilde{e}_{a,k} e_{a,k}^* - \tilde{e}_{a,k}^* e_{a,k}$$
$$|u_k|^2 = |\lambda_{A,k} \tilde{e}_{a,k} - e_{a,k}|^2 \tag{47}$$
$$= |\lambda_{A,k} \tilde{e}_{a,k}|^2 + |e_{a,k}|^2 - \lambda_{A,k} \tilde{e}_{a,k} e_{a,k}^* - \lambda_{A,k}^* \tilde{e}_{a,k}^* e_{a,k} \tag{48}$$

which allows to reformulate (45)

$$\|\tilde{\mathbf{w}}_k\|_{\mathbf{F}}^2 = \|\tilde{\mathbf{w}}_{k-1}\|_{\mathbf{F}}^2 + 4\frac{\mu_k^2}{\bar{\mu}_{A,k}} |\tilde{e}_{a,k}|^2 - 2\mu_k [|u_k|^2 - |\lambda_{A,k}|^2 |\tilde{e}_{a,k}|^2 - |e_{a,k}|^2]. \tag{49}$$

We can further bound $|u_k|^2$ by

$$|u_k|^2 = |\lambda_{A,k} \tilde{e}_{a,k} - e_{a,k}|^2 \tag{50}$$
$$= |\lambda_{A,k} \tilde{e}_{a,k} - \tilde{e}_{a,k} + v_k|^2 \tag{51}$$
$$\leq (1+\gamma)|v_k|^2 + \frac{1+\gamma}{\gamma} |1 - \lambda_{A,k}|^2 |\tilde{e}_{a,k}|^2 \tag{52}$$

for some positive value $\gamma > 0$ which, in turn, allows now to write

$$\|\tilde{\mathbf{w}}_k\|_{\mathbf{F}}^2 + 2\mu_k |e_{a,k}|^2 \leq \|\tilde{\mathbf{w}}_{k-1}\|_{\mathbf{F}}^2 + 4\frac{\mu_k^2}{\bar{\mu}_{A,k}} |\tilde{e}_{a,k}|^2 + 2\mu_k (1+\gamma)|v_k|^2 \tag{53}$$
$$+ 2\mu_k \left[\frac{1+\gamma}{\gamma} |1 - \lambda_{A,k}|^2 - |\lambda_{A,k}|^2 \right] |\tilde{e}_{a,k}|^2.$$
$$= \|\tilde{\mathbf{w}}_{k-1}\|_{\mathbf{F}}^2 + 2\mu_k (1+\gamma)|v_k|^2 + 2\mu_k \delta_{A,k} |\tilde{e}_{a,k}|^2 \tag{54}$$

$$\delta_{A,k} = \frac{2\mu_k}{\bar{\mu}_{A,k}} + 1 - 2\Re\{\lambda_{A,k}\} + \frac{1}{\gamma} |1 - \lambda_{A,k}|^2. \tag{55}$$

If term $\delta_{A,k}$ is negative or equivalently $0 < \frac{\mu_k}{\bar{\mu}_{A,k}} = \alpha < \Re\{\lambda_{A,k}\} - \frac{1}{2} - \frac{1}{2\gamma}|1 - \lambda_{A,k}|^2$, the last term in (54) can simply be dropped and we obtain a first local stability condition relating the update from time instant $k-1$ to k:

Lemma 4.1: The adaptive gradient type algorithm with Update (38) exhibits the following local robustness properties from its inputs $\{\tilde{\mathbf{w}}_{k-1}, \sqrt{2\mu_k(1+\gamma)} v_k\}$ to its outputs $\{\tilde{\mathbf{w}}_k, \sqrt{2\mu_k} e_{a,k}\}$:

$$\frac{\|\tilde{\mathbf{w}}_k\|_{\mathbf{F}}^2 + 2\mu_k |e_{a,k}|^2}{\|\tilde{\mathbf{w}}_{k-1}\|_{\mathbf{F}}^2 + 2(1+\gamma)\mu_k |v_k|^2} \leq 1 \tag{56}$$

as long as μ_k can be selected so that $0 < \frac{\mu_k}{\bar{\mu}_{A,k}} = \alpha < \Re\{\lambda_{A,k}\} - \frac{1}{2} - \frac{1}{2\gamma}|1 - \lambda_{A,k}|^2$ for some $\gamma > 0$, and $\mathbf{F}^H\mathbf{F} > 0$.

Such a local robustness property however is only useful if it can be extended towards a global property. To this end we sum up the energy terms over a finite horizon from $k = 1, \ldots, N$ and compute norms:

$$\sqrt{\sum \mu_k |e_{a,k}|^2} \leq \sqrt{\frac{\|\tilde{\mathbf{w}}_0\|_{\mathbf{F}}^2}{2}} + \sqrt{\sum \mu_k (1+\gamma)|v_k|^2}. \tag{57}$$

The expression makes sense as long as $\delta_{A,k} < 0$. However we can extend the result even for $\delta_{A,k} < 1$. To show this property, we start with summing up (54) under the condition that $0 < \delta_{A,k} < 1$, remembering that $\tilde{e}_{a,k} = e_{a,k} + v_k$ and obtain

$$\sqrt{\sum \mu_k |e_{a,k}|^2} \leq \sqrt{\frac{\|\tilde{\mathbf{w}}_0\|_{\mathbf{F}}^2}{2}} + \sqrt{\sum \mu_k (1+\gamma)|v_k|^2} + \sqrt{\sum \mu_k \delta_{A,k} |e_{a,k}|^2} + \sqrt{\sum \mu_k \delta_{A,k} |v_k|^2} \tag{58}$$

$$\min(1 - \delta_{A,k}) \sqrt{\sum \mu_k |e_{a,k}|^2} \leq \sqrt{\frac{\|\tilde{\mathbf{w}}_0\|_{\mathbf{F}}^2}{2}} + \max(1 - \delta_{A,k}) \sqrt{\sum \mu_k (1+\gamma)|v_k|^2}, \tag{59}$$

for which both terms $\min(1 - \delta_{A,k})$ and $\max(1 - \delta_{A,k})$ remain positive and bounded. We thus can conclude on global robustness:

Lemma 4.2: The adaptive gradient type algorithm with Update (38) exhibits a global robustness from initial uncertainties $\|\tilde{\mathbf{w}}_0\|_{\mathbf{F}}$ and additive noise energy sequence $\{\sqrt{\mu_k(1+\gamma)} v_k\}_{k=1,2,\ldots,N}$ to its *a priori* error sequence $\{\sqrt{\mu_k} e_{a,k}\}_{k=1,2,\ldots,N}$ if the normalized step-size $0 < \alpha = \frac{\mu_k}{\bar{\mu}_{A,k}} < \Re\{\lambda_{A,k}\} - \frac{1}{2\gamma}|1 - \lambda_{A,k}|^2$ for some $\gamma > 0$, and $\mathbf{F}^H\mathbf{F} > 0$.

While such statement ensures the LMS algorithm with non-symmetric matrix step-size $\mathbf{G}$ to be l_2–stable, it actually is based on the condition that $\Re\{\lambda_{A,k}\} > 0$. This brings us back to the choice of $\lambda_{A,k}$ which we will have to analyze further. Recall that we defined $\bar{e}_{A,k} = \lambda_{A,k}^* e_{a,k}$ that is we relate $e_{a,k} = \mathbf{x}_k^H \tilde{\mathbf{w}}_{k-1}$ and $\bar{e}_{A,k} = \mathbf{x}_k^H \mathbf{G}^H \mathbf{F}^H \mathbf{F} \tilde{\mathbf{w}}_{k-1}$. As these inner vector products, defining $e_{a,k}$ as well as $\bar{e}_{A,k}$, can take on every arbitrary value, independent of each other, there is no relation in form of a bound from one to the other and as a consequence a strict l_2 stability analysis must end here. Note however, if the relations of the previous lemma hold for any signal they also hold for random processes following some statistics. Thus, placing the expectation operation over all energy terms results in correct statements even though somewhat restricted now by the imposed statistics. Note further that even if $e_{a,k}$ and $\bar{e}_{A,k}$ is hard to be related for general signals, from a statistical point of view the two signals are related. This can be seen when we compute their average energy, that is

$$\begin{aligned}\mathbb{E}[e_{a,k}^*\bar{e}_{A,k}] &= \mathbb{E}[\tilde{\mathbf{w}}_{k-1}^H\mathbf{x}_k\mathbf{x}_k^H\mathbf{G}^H\mathbf{F}^H\mathbf{F}\tilde{\mathbf{w}}_{k-1}] \\ &= \mathbb{E}[\tilde{\mathbf{w}}_{k-1}^H\mathbf{R}_{\mathrm{xx}}\mathbf{G}^H\mathbf{F}^H\mathbf{F}\tilde{\mathbf{w}}_{k-1}].\end{aligned}$$

Starting with (41), taking expectations on both side and solving for steady-state, that is $\mathbb{E}[\|\tilde{\mathbf{w}}_k\|_{\mathbf{F}}^2] = \mathbb{E}[\|\tilde{\mathbf{w}}_{k-1}\|_{\mathbf{F}}^2] = \mathbb{E}[\|\tilde{\mathbf{w}}_\infty\|_{\mathbf{F}}^2]$, we find

$$\mathbb{E}[|e_{a,\infty}|^2] = \frac{\mathbb{E}\left[\frac{\mu_k}{\bar{\mu}_{A,k}}\right]\sigma_v^2}{\bar{\lambda}_A - \mathbb{E}\left[\frac{\mu_k}{\bar{\mu}_{A,k}}\right]} \quad (60)$$

$$\bar{\lambda}_A = \frac{\Re\left\{\mathbb{E}[\tilde{\mathbf{w}}_\infty^H\mathbf{R}_{\mathrm{xx}}\mathbf{G}^H\mathbf{F}^H\mathbf{F}\tilde{\mathbf{w}}_\infty]\right\}}{\mathbb{E}[\tilde{\mathbf{w}}_\infty^H\mathbf{R}_{\mathrm{xx}}\tilde{\mathbf{w}}_\infty]} \quad (61)$$

where we applied the independence assumption [15, Ch. 9] on regression vectors $\mathbf{x}_k$ with autocorrelation matrix $\mathbf{R}_{\mathrm{xx}} = \mathbb{E}[\mathbf{x}_k\mathbf{x}_k^H]$ of driving process x_k and corresponding parameter error vectors $\tilde{\mathbf{w}}_{k-1}$. The so defined $\bar{\lambda}_A$ can be interpreted as the mean of $\lambda_{A,k}$. The term $\mathbb{E}\left[\frac{\mu_k}{\bar{\mu}_{A,k}}\right]$ takes on a particular simple form ($=\alpha$) when a normalized step-size is applied: $\mu_k = \alpha\bar{\mu}_{A,k}$. The steady-state solution can be a means for defining a step-size bound: $\alpha < \bar{\lambda}_A$. As $\tilde{\mathbf{w}}_\infty$ is typically unknown, it would be difficult to evaluate $\bar{\lambda}_A$. A conservative bound however is simple to derive by the Rayleigh factor of a Hermitian matrix[1]:

$$\bar{\lambda}_{A,\min} \le \frac{\mathbf{w}^H\mathbf{R}_{\mathrm{xx}}^{\frac{1}{2}}\left[\mathbf{G}^H\mathbf{F}^H\mathbf{F} + \mathbf{F}^H\mathbf{F}\mathbf{G}\right]\mathbf{R}_{\mathrm{xx}}^{\frac{H}{2}}\mathbf{w}}{2\mathbf{w}^H\mathbf{w}} \le \bar{\lambda}_{A,\max}.$$

Let us summarize the previous considerations in the following theorem.

Theorem 4.2: The adaptive filter with Update (38) with non-symmetric step-size matrix $\mathbf{G}$, some square matrix $\mathbf{F}$ that satisfies the condition $\mathbf{F}^H\mathbf{F} > 0$, and normalized step-size $\alpha = \frac{\mu_k}{\bar{\mu}_{A,k}}$ guarantees convergence in the mean square sense of its parameter error vector $\tilde{\mathbf{w}}_k$ if the step-size

$$0 < \alpha < \bar{\lambda}_{A,\min} \le \bar{\lambda}_A \quad (62)$$

under the independence assumption of regression vectors $\mathbf{x}_k$ with $\mathbf{R}_{\mathrm{xx}} = \mathbb{E}[\mathbf{x}_k\mathbf{x}_k^H]$.

If the minimum Rayleigh factor $\bar{\lambda}_{A,\min}$ is negative we cannot conclude convergence. If the step-size is larger than $\bar{\lambda}_{A,\max}$ we expect divergence.

[1]We use the short notation $\mathbf{G}^{-H} = (\mathbf{G}^H)^{-1} = (\mathbf{G}^{-1})^H$. Similarly $\mathbf{G}^{-T} = (\mathbf{G}^T)^{-1} = (\mathbf{G}^{-1})^T$ and $\mathbf{G}^{-*} = (\mathbf{G}^*)^{-1} = (\mathbf{G}^{-1})^*$. Moreover for positive definite Hermitian matrices, we use $\mathbf{R} > 0$ to denote positiveness and $\mathbf{R} = \mathbf{R}^{\frac{1}{2}}\mathbf{R}^{\frac{H}{2}} = \mathbf{R}^{\frac{H}{2}}\mathbf{R}^{\frac{1}{2}}$.

Example A: Let us use $\mathbf{F} = \mathbf{I}$ and $\mathbf{R}_{\mathrm{xx}} = \mathbf{I}$. In this case we find

$$\bar{\mu}_{A,k} = \frac{1}{\mathbf{x}_k^H\mathbf{G}^H\mathbf{G}\mathbf{x}_k} \quad (63)$$

and convergence in the mean square sense for $0 < \alpha < \frac{1}{2}\min\{\mathrm{eig}(\mathbf{G} + \mathbf{G}^H)\}$.

B. Analysis Method B

We now modify the previous method by the following idea. Let us assume again an additional matrix $\mathbf{F}$ that is multiplied from the left. However, now we will not compute the norm in $\mathbf{F}^H\mathbf{F}$ but the inner vector product including $\mathbf{F}$ only. We repeat the process with $\mathbf{F}^H$ and obtain so the conjugate complex of the first part. Adding both terms results in the following:

$$\begin{aligned}\|\tilde{\mathbf{w}}_k\|_{\mathbf{F}+}^2 &= \|\tilde{\mathbf{w}}_{k-1}\|_{\mathbf{F}+}^2 \\ &\quad + 4\frac{\mu_k^2}{\bar{\mu}_{B,k}}|\tilde{e}_{a,k}|^2 - 2\mu_k\bar{e}_{B,k}^*\tilde{e}_{a,k} - 2\mu_k\bar{e}_{B,k}\tilde{e}_{a,k}^* \quad (64)\end{aligned}$$

with the new abbreviations

$$\|\tilde{\mathbf{w}}_k\|_{\mathbf{F}+}^2 = \tilde{\mathbf{w}}_k^H(\mathbf{F} + \mathbf{F}^H)\tilde{\mathbf{w}}_k \quad (65)$$

$$\bar{e}_{B,k} = \mathbf{x}_k^H\mathbf{G}^H[\mathbf{F}^H + \mathbf{F}]\tilde{\mathbf{w}}_{k-1} \quad (66)$$

$$\bar{\mu}_{B,k} = \frac{1}{\mathbf{x}_k^H\mathbf{G}^H[\mathbf{F}^H + \mathbf{F}]\mathbf{G}\mathbf{x}_k} \quad (67)$$

$$= \frac{1}{\|\mathbf{G}\mathbf{x}_k\|_{\mathbf{F}+}^2}. \quad (68)$$

From here, the derivation follows the same path as before, we thus will present the important highlights so that the reader can follow easily. Note that the norm in which we require convergence of the parameter error vector is in $\sqrt{\|\cdot\|_{\mathbf{F}+}^2}$ which makes Method B distinctively different to the previous one.

As in Method A we employ the same method and arrive at

$$\begin{aligned}\|\tilde{\mathbf{w}}_k\|_{\mathbf{F}+}^2 + 2\mu_k|e_{a,k}|^2 & \quad (69) \\ \le \|\tilde{\mathbf{w}}_{k-1}\|_{\mathbf{F}+}^2 &+ 2\mu_k(1+\gamma)|v_k|^2 \\ &+ 2\mu_k\delta_{B,k}|\tilde{e}_{a,k}|^2 \\ \lambda_{B,k} = 2\frac{\mu_k}{\bar{\mu}_{B,k}} &+ 1 - 2\Re\{\lambda_{B,k}\} + \frac{1}{\gamma}|1 - \lambda_{B,k}|^2. \quad (70)\end{aligned}$$

This allows for a first local stability condition:

Lemma 4.3: The adaptive gradient type algorithm with Update (38) exhibits the following local robustness properties from its inputs $\{\tilde{\mathbf{w}}_{k-1}, \sqrt{2\mu_k(1+\gamma)}v_k\}$ to its outputs $\{\tilde{\mathbf{w}}_k, \sqrt{2\mu_k}e_{a,k}\}$:

$$\frac{\|\tilde{\mathbf{w}}_k\|_{\mathbf{F}+}^2 + 2\mu_k|e_{a,k}|^2}{\|\tilde{\mathbf{w}}_{k-1}\|_{\mathbf{F}+}^2 + 2(1+\gamma)\mu_k|v_k|^2} \le 1 \quad (71)$$

as long as μ_k can be selected so that $0 < \frac{\mu_k}{\bar{\mu}_{B,k}} = \alpha < \Re\{\lambda_{B,k}\} - \frac{1}{2} - \frac{1}{2\gamma}|1 - \lambda_{B,k}|^2$ for $\gamma > 0$, and $\mathbf{F} + \mathbf{F}^H > 0$.

Following the same method as before, we find the following global statement:

Lemma 4.4: The adaptive gradient type algorithm with Update (38) exhibits a global robustness from initial uncertainty $\tilde{\mathbf{w}}_0$ and additive noise sequence $\{\sqrt{2\mu_k(1+\gamma)}v_k\}_{k=1,2,\ldots,N}$ to its *a priori* error sequence $\{\sqrt{2\mu_k}e_{a,k}\}_{k=1,2,\ldots,N}$ if the normalized step-size $0<\alpha=\frac{\mu_k}{\bar{\mu}_{B,k}}<\Re\{\lambda_{B,k}\}-\frac{1}{2\gamma}|1-\lambda_{B,k}|^2$ for $\gamma>0$ and $\mathbf{F}^H+\mathbf{F}>0$.

This lemma offers similar properties than Lemma 4.2 of Method A and thus the problem of the in general unknown $\lambda_{B,k}$. We thus also follow the steady-state computation as in the previous A and find

$$\bar{\lambda}_{B,\min}\leq\frac{\mathbf{w}^H\mathbf{R}_{\mathrm{xx}}^{\frac{1}{2}}\left[\mathbf{G}^H[\mathbf{F}^H+\mathbf{F}]+[\mathbf{F}^H+\mathbf{F}]\mathbf{G}\right]\mathbf{R}_{\mathrm{xx}}^{\frac{H}{2}}\mathbf{w}}{2\mathbf{w}^H\mathbf{w}} \leq\bar{\lambda}_{B,\max}. \tag{72}$$

Theorem 4.3: The adaptive filter with Update (38) with nonsymmetric step-size matrix $\mathbf{G}$, some square matrix $\mathbf{F}$ that satisfies the condition $\mathbf{F}^H+\mathbf{F}>0$, and normalized step-size $\alpha=\frac{\mu_k}{\bar{\mu}_{B,k}}$ guarantees convergence in the mean square sense of its parameter error vector $\tilde{\mathbf{w}}_k$ if the step-size

$$0<\alpha<\bar{\lambda}_{B,\min}\leq\bar{\lambda}_B \tag{73}$$

under the independence assumption of regression vectors $\mathbf{x}_k$ with $\mathbf{R}_{\mathrm{xx}}=\mathbb{E}[\mathbf{x}_k\mathbf{x}_k^H]$.

If the minimum Rayleigh factor $\bar{\lambda}_{B,\min}$ is negative we cannot conclude convergence. If the step-size is larger than $\bar{\lambda}_{B,\max}$ we expect divergence.

1) Example B: Let us use $\mathbf{F}=\frac{1}{2}\mathbf{I}$ and $\mathbf{R}_{\mathrm{xx}}=\mathbf{I}$. In this case, we find

$$\bar{\mu}_{B,k}=\frac{1}{\mathbf{x}_k^H\mathbf{G}^H\mathbf{G}\mathbf{x}_k} \tag{74}$$

and convergence for $0<\alpha<\frac{1}{2}\min\{\mathrm{eig}(\mathbf{G}+\mathbf{G}^H)\}$. Thus, for this choice methods A and B coincide (compare to Example A).

C. Analysis Method C

We now continue in a similar way as in previous Method B but assume that $\mathbf{F}=\mathbf{G}^{-1}$ exists. We find the following inner vector product:

$$\tilde{\mathbf{w}}_k^H\mathbf{G}^{-1}\tilde{\mathbf{w}}_k=\tilde{\mathbf{w}}_{k-1}^H\mathbf{G}^{-1}\tilde{\mathbf{w}}_{k-1}+4\mu_k^2\mathbf{x}_k^H\mathbf{G}\mathbf{x}_k|\tilde{e}_{a,k}|^2 -2\mu_k[\tilde{\mathbf{w}}_{k-1}^H\mathbf{x}_k\tilde{e}_{a,k}+\mathbf{x}_k^H\mathbf{G}^H\mathbf{G}^{-1}\tilde{\mathbf{w}}_{k-1}\tilde{e}_{a,k}^*]$$

which we complement by its conjugate complex part just as in previous Method B. However, now some terms compensate as $\mathbf{G}\mathbf{G}^{-1}=\mathbf{I}$. We now introduce

$$\|\mathbf{x}_k\|_{\mathbf{G}+}^2=\mathbf{x}_k^H[\mathbf{G}^H+\mathbf{G}]\mathbf{x}_k \tag{75}$$

$$\bar{\mu}_{C,k}=\frac{1}{\|\mathbf{x}_k\|_{\mathbf{G}+}^2} \tag{76}$$

$$\bar{e}_{C,k}=\mathbf{x}_k^H\mathbf{G}^H\mathbf{G}^{-1}\tilde{\mathbf{w}}_{k-1} \tag{77}$$

$$=\lambda_{C,k}^*e_{a,k} \tag{78}$$

$$\delta_{C,k}=2\frac{\mu_k}{\bar{\mu}_k}-|1+\lambda_{C,k}|^2+\frac{1+\gamma}{\gamma}|\lambda_{C,k}|^2. \tag{79}$$

Note that $\delta_{C,k}$ now takes a slightly different form compared to the values in Methods A and B, leading to much tighter bounds.

Lemma 4.5: The adaptive gradient type algorithm with Update (38) exhibits the following local robustness properties from its input values $\left\{\sqrt{\|\tilde{\mathbf{w}}_{k-1}\|_{\mathbf{G}^{-1}+}^2},\sqrt{2\mu_k(1+\gamma)}v_k\right\}$ to its output values $\left\{\sqrt{\|\tilde{\mathbf{w}}_k\|_{\mathbf{G}^{-1}+}^2},2\sqrt{\mu_k}e_{a,k}\right\}$

$$\frac{\|\tilde{\mathbf{w}}_k\|_{\mathbf{G}^{-1}+}^2+4\mu_k|e_{a,k}|^2}{\|\tilde{\mathbf{w}}_{k-1}\|_{\mathbf{G}^{-1}+}^2+2(1+\gamma)\mu_k|v_k|^2}\leq 1$$

as long as μ_k can be selected so that $0<\frac{\mu_k}{\bar{\mu}_{C,k}}=\alpha<\Re\{\lambda_{C,k}\}-\frac{1}{2\gamma}|\lambda_{C,k}|^2$ for some $\gamma>0$ and as long as the matrix $\mathbf{G}+\mathbf{G}^H$ is positive definite.

Summing up the energy terms and computing norms we obtain the global robustness property:

Lemma 4.6: The adaptive gradient type algorithm with Update (38) exhibits a global robustness from initial uncertainties $\sqrt{\tilde{\mathbf{w}}_0^H[\mathbf{G}^{-1}+\mathbf{G}^{-H}]\tilde{\mathbf{w}}_0}$ and additive noise energy sequence $\{\sqrt{2\mu_k(1+\gamma)}v_k\}_{k=1,2,\ldots,N}$ to its *a priori* error energy sequence $\{\sqrt{4\mu_k}e_{a,k}\}_{k=1,2,\ldots,N}$ if $0<\alpha<\Re\{\lambda_{C,k}\}+\frac{1}{2}-\frac{1}{2\gamma}|\lambda_{C,k}|^2$ for some $\gamma>0$ and $\mathbf{G}^{-1}+\mathbf{G}^{-H}>0$.

Note that this analysis method compared to the previous two methods delivers a stronger argument when compared to Methods A and B. Here the step-size bound could become positive and it might be even possible to guarantee l_2–stability in some scenarios.

Following the stochastic approach as before, we compute the steady-state to be

$$\mathbb{E}[|e_{a,\infty}|^2]=\frac{2\mathbb{E}\left[\frac{\mu_k}{\bar{\mu}_{C,k}}\right]\sigma_v^2}{\bar{\lambda}_C+1-2\mathbb{E}\left[\frac{\mu_k}{\bar{\mu}_{C,k}}\right]}$$

$$\bar{\lambda}_C=\frac{\Re\left\{\mathbb{E}[\tilde{\mathbf{w}}_\infty^H\mathbf{R}_{\mathrm{xx}}\mathbf{G}^H\mathbf{G}^{-1}\tilde{\mathbf{w}}_\infty]\right\}}{\mathbb{E}[\tilde{\mathbf{w}}_\infty^H\mathbf{R}_{\mathrm{xx}}\tilde{\mathbf{w}}_\infty]}. \tag{80}$$

We find the mean $\bar{\lambda}_C$ of $\lambda_{C,k}$ to be bounded by

$$\bar{\lambda}_{C,\min}\leq\frac{\mathbf{w}^H\mathbf{R}_{\mathrm{xx}}^{\frac{1}{2}}\left[\mathbf{G}^H\mathbf{G}^{-1}+\mathbf{G}^{-H}\mathbf{G}\right]\mathbf{R}_{\mathrm{xx}}^{\frac{H}{2}}\mathbf{w}}{2\mathbf{w}^H\mathbf{w}}\leq\bar{\lambda}_{C,\max}.$$

Theorem 4.4: The adaptive filter with Update (38) with nonsymmetric step-size matrix $\mathbf{G}$, satisfying $\mathbf{G}+\mathbf{G}^H>0$ and normalized step-size $\alpha=\frac{\mu_k}{\bar{\mu}_{C,k}}$ guarantees convergence in the mean square sense of its parameter error vector $\tilde{\mathbf{w}}_k$ if the step-size

$$0<\alpha<\frac{1}{2}+\frac{1}{2}\bar{\lambda}_{C,\min}\leq\frac{1}{2}+\frac{1}{2}\bar{\lambda}_C \tag{81}$$

under the independence assumption of regression vectors $\mathbf{x}_k$ with $\mathbf{R}_{\mathrm{xx}}=\mathbb{E}[\mathbf{x}_k\mathbf{x}_k^H]$. Alternatively, the so normalized algorithm also converges if the matrix $\mathbf{G}+\mathbf{G}^H$ is negative definite.

Note that due to the normalization of the step-size by terms in $\mathbf{G}$, replacing $\mathbf{G}$ by $-\mathbf{G}$ causes a positive definite matrix $\mathbf{G}+\mathbf{G}^H$ to become negative definite so that the product $\bar{\mu}_{C,k}\mathbf{G}$ remains positive. Also due to the products $\mathbf{G}^H\mathbf{G}^{-1}$ effects compensate each other. The positive upper bound for the normalized step-size is thus not changed by this. The derivation simply requires in this case to define $\|\tilde{\mathbf{w}}_k\|_{\mathbf{G}^{-1}+}=-\tilde{\mathbf{w}}_k^H[\mathbf{G}^{-1}+\mathbf{G}^{-H}]\tilde{\mathbf{w}}_k$ to be a norm.

TABLE II
VARIOUS ALGORITHMIC NORMALIZATIONS BASED ON THE PROPOSED METHODS A, B, C WITH CORRESPONDING CONDITION $\bar{\lambda}_{\min} > 0$ (ALGORITHM 2: $\bar{\lambda}_{\min} > -1$) FOR MEAN-SQUARE SENSE CONVERGENCE UNDER WHITE EXCITATION PROCESSES. $\mathbf{A} > 0$ STANDS FOR POSITIVE DEFINITENESS

Name	$1/\bar{\mu}_k$	$\bar{\lambda}_{\min} = \frac{1}{2}\min\mathrm{eig}\{\cdot\}$	comment
Alg. 1	$\mathbf{x}_k^H\mathbf{G}^H\mathbf{G}\mathbf{x}_k$	$\mathbf{G}+\mathbf{G}^H$	Example A+B
Alg. 2	$\mathbf{x}_k^H[\mathbf{G}+\mathbf{G}^H]\mathbf{x}_k$	$\mathbf{G}^H\mathbf{G}^{-1}+\mathbf{G}^{-H}\mathbf{G}$	Method C
Alg. 3	$\mathbf{x}_k^H\mathbf{x}_k$	$(\mathbf{G}^H\mathbf{G})^{-1}\mathbf{G}^H+\mathbf{G}(\mathbf{G}^H\mathbf{G})^{-1}$	Method A, $\mathbf{F}=(\mathbf{G}^H\mathbf{G})^{-1}\mathbf{G}^H$
Alg. 4	$\mathbf{x}_k^H(\mathbf{G}^H\mathbf{G})^2\mathbf{x}_k$	$\mathbf{G}\mathbf{G}^H\mathbf{G}+\mathbf{G}^H\mathbf{G}\mathbf{G}^H$	Method A, $\mathbf{F}=\mathbf{G}^H$
Alg. 5	$\mathbf{x}_k^H[\mathbf{G}+\mathbf{G}^H]\mathbf{x}_k$	$\mathbf{G}^H(\mathbf{G}^H\mathbf{G})^{-1}\mathbf{G}^H+\mathbf{G}^H(\mathbf{G}^H\mathbf{G})^{-1}\mathbf{G}+2\mathbf{I}$	Method B, $\mathbf{F}=(\mathbf{G}^H\mathbf{G})^{-1}\mathbf{G}^H$
Alg. 6	$\mathbf{x}_k^H[\mathbf{G}^H(\mathbf{G}^H\mathbf{G})^{-2}\mathbf{G}]\mathbf{x}_k$	$\mathbf{G}^H(\mathbf{G}^H\mathbf{G})^{-2}+(\mathbf{G}^H\mathbf{G})^{-2}\mathbf{G}$	Method A, $\mathbf{F}^H\mathbf{F}=(\mathbf{G}^H\mathbf{G})^{-2}$

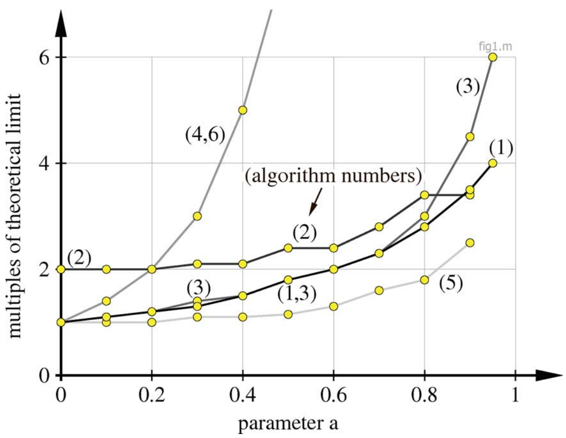

Fig. 1. Convergence bound α_l over parameter a.

D. Consequences

A further consequence worth stating is:

Corollary 4.1: Consider the three update equations:

$$\hat{\mathbf{w}}_k = \hat{\mathbf{w}}_{k-1} + 2\alpha\bar{\mu}_k\mathbf{G}\mathbf{x}_k\tilde{e}_{a,k} \tag{82}$$
$$\hat{\mathbf{w}}_k = \hat{\mathbf{w}}_{k-1} + 2\alpha\bar{\mu}_k\mathbf{G}^H\mathbf{x}_k\tilde{e}_{a,k} \tag{83}$$
$$\hat{\mathbf{w}}_k = \hat{\mathbf{w}}_{k-1} + \alpha\bar{\mu}_k[\mathbf{G}+\mathbf{G}^H]\mathbf{x}_k\tilde{e}_{a,k} \tag{84}$$

with $\tilde{e}_{a,k} = \mathbf{x}_k^H\tilde{\mathbf{w}}_{k-1} + v_k$ and $\bar{\mu}_k = \mathbf{x}_k^H[\mathbf{G}+\mathbf{G}^H]\mathbf{x}_k$. All three algorithms converge in the mean square sense as long as $\alpha\bar{\mu}_k[\mathbf{G}+\mathbf{G}^H]$ is positive definite for sufficiently small step-size α. Note that this can even include that $[\mathbf{G}+\mathbf{G}^H]$ is negative definite.

Furthermore, the steady-state of such algorithms can also be computed. Starting from (41) we compute the expectation of the energy terms considering a fixed start value $\tilde{\mathbf{w}}_0$ as well as random excitation $\mathbf{x}_k$ and additive noise v_k. For steady-state, we find that $\mathbb{E}[\|\tilde{\mathbf{w}}_k\|_2^2] = \mathbb{E}[\|\tilde{\mathbf{w}}_{k-1}\|_2^2] = \mathbb{E}[\|\tilde{\mathbf{w}}_\infty\|_2^2]$, and we obtain

$$\bar{\lambda}\mathbb{E}\|\tilde{\mathbf{w}}_\infty\|_2^2 = \mathbb{E}\left[\frac{\mu_k}{\bar{\mu}_k}\right]\left(\mathbb{E}\|\tilde{\mathbf{w}}_\infty\|_2^2 + N_o\right) \tag{85}$$

which immediately leads to the desired result for normalized step-sizes $\alpha = \frac{\mu_k}{\bar{\mu}_k}$:

$$S_{\mathrm{rel}} = \frac{\mathbb{E}\|\tilde{\mathbf{w}}_\infty\|_2^2}{\|\mathbf{w}_o\|_2^2} = \frac{\alpha N_o}{\bar{\lambda}-\alpha}. \tag{86}$$

The only difference to other LMS algorithms shows in the value of $\bar{\lambda}$ that takes on the value two in a standard NLMS. However, the actual value of $\bar{\lambda}$ is difficult to compute. For white driving processes x_k, its bounds are

$$\bar{\lambda}_{\min} = \frac{1}{2}\min\{\mathrm{eig}(\mathbf{G}+\mathbf{G}^H)\} \le \bar{\lambda} \le \frac{1}{2}\max\{\mathrm{eig}(\mathbf{G}+\mathbf{G}^H)\} = \bar{\lambda}_{\max}.$$

E. Validation

In a Monte Carlo experiment, we run simulations (20 runs for each parameter setup) for filter order $M = 50$ with a noise variance of $N_o = 0.0001$. Excitation signals are white symbols from a QPSK alphabet. The experiment applies the matrix

$$\mathbf{G} = \begin{pmatrix} 1 & a & a^2 & \dots & a^{M-1} \\ 0 & 1 & a & \dots & a^{M-2} \\ 0 & 0 & 1 & \dots & a^{M-3} \\ \vdots & & & & \\ 0 & 0 & \dots & 1 & a \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix} \tag{87}$$

where we vary a from zero to one.[2] Independent of the value a the matrix is always regular. We are interested in correctness and the precision of our derived bounds. We thus use the normalized step-sizes and normalize them w.r.t. their bounds, that is $\mu_k = \alpha\bar{\mu}_k\bar{\lambda}_{\min}$. We thus expect to find converging algorithms for $\alpha < 1$. Table II depicts a list of choices. Fig. 1 exhibits the observed bounds for $\max\alpha$ from Algorithm 1 to 6 when ranging $0 \le a < 1$. Compare Algorithm 2 and Algorithm 5, being identical but with different bounds, the bound of Algorithm 2 being about twice as large as that of Algorithm 5. Algorithm 1 and Algorithm 3 as well as Algorithm 4 and Algorithm 6 show almost identical behavior, respectively. Above all, only Algorithm 3 is of practical interest if matrix $\mathbf{G}$ is not known beforehand.

V. GRADIENT ALGORITHMS WITH VARIABLE NOISE

We now return to the second problem defined at the end of Section III – that is adaptive algorithms in which the noise part has a component depending on the parameter error vector itself. We thus consider the following update form (noisy LMS algorithm):

$$\hat{\mathbf{w}}_k = \hat{\mathbf{w}}_{k-1} + 2\mu_k\mathbf{x}_k[e_{a,k} + \tilde{v}_k] \tag{88}$$
$$\tilde{v}_k = n_k + c + \gamma\mathbf{v}_k^H\tilde{\mathbf{w}}_{k-1} \tag{89}$$

[2]Note that for all examples and simulations the corresponding Matlab code is available online at https://www.nt.tuwien.ac.at/downloads/featured-downloads.

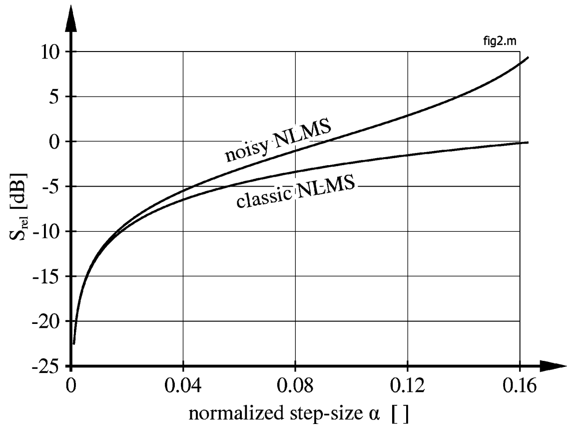

Fig. 2. System mismatch over step-size α for classical NLMS versus noisy NLMS.

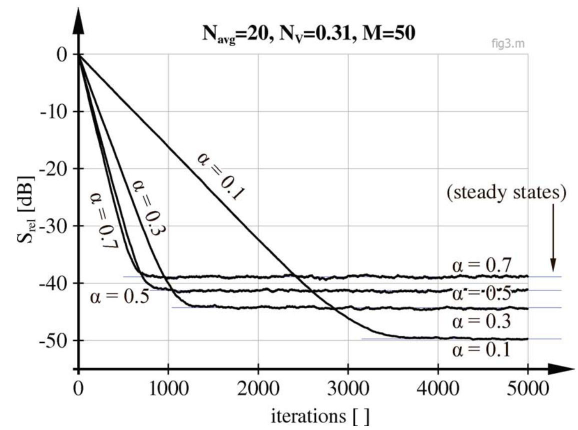

Fig. 3. System mismatch over adaptation steps for various values of α in noisy NLMS algorithm.

in which we incorporated classical noise v_k, a constant c and finally a term depending on the parameter error vector. Assuming white noise sequences n_k, v_k with variance N_n and N_v, respectively, we find

$$\sigma^2_{\tilde{v}_k} = N_n + c^2 + \gamma^2 N_v \mathbb{E}\|\tilde{\mathbf{w}}_{k-1}\|_2^2 \tag{90}$$

where we assumed noise n_k and $\mathbf{v}_k \in \mathbb{C}^{1\times M}$ to be uncorrelated. Although in later considerations this is not true, we will neglect the correlation term as it is typically small. For white driving processes and a normalized step-size α with normalization $\bar{\mu}_k = \frac{1}{\|\mathbf{x}_k\|_2^2}$ (NLMS), it is well known [11], [15] that the relative system mismatch at time instant k is given by

$$S_{\mathrm{rel},k} = \frac{\mathbb{E}\|\tilde{\mathbf{w}}_{k-1}\|_2^2}{\|\mathbf{w}_o\|_2^2} = \frac{\sigma^2_{\tilde{v}_k}}{\|\mathbf{w}_o\|_2^2}\frac{\alpha}{2-\alpha}\gamma_x. \tag{91}$$

Factor γ_x accounts for the correlation in driving sequence x_k. For uncorrelated processes, $\gamma_x = 1$. Substituting the noise variance we obtain at equilibrium

$$S_{\mathrm{rel},\infty} = [N_n + c^2]\frac{\alpha}{2-\alpha[1+\gamma^2 N_v]}\gamma_x. \tag{92}$$

Now the relative system mismatch is no longer proportional to N_v but also impacts the stability bound. We find now a reduced stability bound at

$$\alpha < \frac{2}{1+\gamma^2 N_v} \tag{93}$$

that is the higher the noise variance, the smaller the step-size bound. Fig. 2 with $\gamma = 1$, $c = 1$, $N_n = N_v = 10$ depicts the dependency of the relative system mismatch on the normalized step-size α. Only for small step-sizes we approximately find the previous behavior.

A. Interpretation

Note that such Update (88) with noise term (89) can also be interpreted differently. We can equivalently formulate the source for the parameter error vector dependent noise as part of the gradient term:

$$\hat{\mathbf{w}}_k = \hat{\mathbf{w}}_{k-1} + 2\mu_k \mathbf{x}_k [e_{a,k} + \mathbf{v}_k^H \tilde{\mathbf{w}}_{k-1}] \tag{94}$$

$$\begin{aligned} &= \hat{\mathbf{w}}_{k-1} + 2\mu_k \mathbf{x}_k [\mathbf{x}_k^H \tilde{\mathbf{w}}_{k-1} + \mathbf{v}_k^H \tilde{\mathbf{w}}_{k-1}] \\ &= \hat{\mathbf{w}}_{k-1} + 2\mu_k \mathbf{x}_k [[\mathbf{x}_k + \mathbf{v}_k]^H \tilde{\mathbf{w}}_{k-1}] \\ &= \hat{\mathbf{w}}_{k-1} + 2\mu_k [\mathbf{r}_k - \mathbf{v}_k][\mathbf{r}_k^H \tilde{\mathbf{w}}_{k-1}]. \end{aligned} \tag{95}$$

Thus, we can interpret such algorithm equivalently as an algorithm with a disturbed gradient. In [16], such algorithm is proven to be non l_2–stable although it can behave well in average. We thus can conclude from here that algorithms with such property are guaranteed not to be robust but can behave convergent in the mean square sense as long as the step-size is sufficiently small.

B. Validation

Simulation runs for such scenario with white excitation showed that this description is indeed very accurate. Fig. 3 shows typical simulation runs with various step-sizes comparing the steady-state with the predicted values according to (92) at high noise level $N_n = 0.01$, $N_v = 1.26$. The agreement is obviously excellent.

VI. Adaptive ZF Gradient Algorithm

We are now ready to analyze our original problem, the adaptive ZF algorithm. We briefly summarize the previous steps [(22) and (34)]:

$$\hat{\mathbf{f}}_k = \hat{\mathbf{f}}_{k-1} + 2\mu_k \mathbf{s}_{u,k}[s^*_{k-\tau} - \mathbf{r}_k^H \hat{\mathbf{f}}_{k-1}] \tag{96}$$

$$\tilde{\mathbf{f}}_k = \tilde{\mathbf{f}}_{k-1} - 2\mu_k \mathbf{s}_{u,k}[\mathbf{s}_{u,k}^H \mathbf{H}_u^H \tilde{\mathbf{f}}_{k-1} + \tilde{w}_{\mathrm{ZF},k}] \tag{97}$$

$$\tilde{w}_{\mathrm{ZF},k} = [\mathbf{s}_{l,k}^H \mathbf{H}_l^H + \mathbf{v}_k^H]\tilde{\mathbf{f}}_{k-1} + \mathbf{s}_k^H \mathbf{v}_{\mathrm{ZF}} - \mathbf{v}_k^H \mathbf{f}_{\mathrm{ZF}}^H.$$

Note that in literature [1], [10], it is argued that a steepest descent algorithm of this form must converge to a global minimum. It is further conjectured that noise may have an impact on the convergence of the algorithm. In the following we will show that there is indeed conditions on the channel required for global stability. Moreover, we will specify the qualitative as well as quantitative impact of noise on step-size choice, convergence condition and finally steady-state behavior.

A. Analysis

In order to analyze the behavior of Update (97), we premultiply by $\mathbf{H}_u^H$ first, $\mathbf{H}_u^H \hat{\mathbf{f}}_k = \hat{\mathbf{g}}_k$, and obtain

$$\tilde{\mathbf{g}}_k = \tilde{\mathbf{g}}_{k-1} - 2\mu_k \mathbf{H}_u^H \mathbf{s}_{u,k}[\mathbf{s}_{u,k}^H \tilde{\mathbf{g}}_{k-1} + \tilde{v}_k]. \tag{98}$$

We thus recognize our adaptive algorithm of the previous section with nonsymmetric step-size matrix $\mathbf{G} = \mathbf{H}_u^H$. In the next Section VI-B we will apply our derived conditions for typical example channels and test whether we can satisfy the required conditions and thus find step-size bounds for mean square convergence. Based on the previous analysis we can now derive the convergence of the algorithm for a normalization $\bar{\mu}_k = \frac{1}{\|\mathbf{x}_k\|_2^2}$ and a normalized step-size $0 < \alpha < \sqrt{\min \mathrm{eig}(\mathbf{H}_u^H + \mathbf{H}_u)^{-1}}$.

The second problem we encounter is the noise being dependent on the parameter error vector. Let us thus now investigate the noise term. We can write

$$\begin{aligned}\tilde{w}_{\mathrm{ZF},k} &= [\mathbf{s}_{l,k}^H \mathbf{H}_l^H + \mathbf{v}_k^H]\tilde{\mathbf{f}}_{k-1} + \mathbf{s}_k^H \mathbf{v}_{\mathrm{ZF}} - \mathbf{v}_k^H \mathbf{f}_{\mathrm{ZF}}^H \\ &= [\mathbf{s}_{l,k}^H \mathbf{H}_l^H + \mathbf{v}_k^H]\mathbf{H}_u^{-H}\tilde{\mathbf{g}}_{k-1} + \mathbf{s}_k^H \mathbf{v}_{\mathrm{ZF}} - \mathbf{v}_k^H \mathbf{f}_{\mathrm{ZF}}^H. \end{aligned} \tag{99}$$

The noise variance at time instant k of this can be computed to

$$\begin{aligned}\sigma_{\tilde{w}_k}^2 = &\|\mathbf{v}_{\mathrm{ZF}}\|_2^2 + N_o\|\mathbf{f}_{\mathrm{ZF}}\|_2^2 \\ &+ \mathbb{E}[\tilde{\mathbf{g}}_{k-1}^H \mathbf{H}_u^{-1}\left[\mathbf{H}_l\mathbf{H}_l^H + N_o\mathbf{I}\right]\mathbf{H}_u^{-H}\tilde{\mathbf{g}}_{k-1}] \end{aligned} \tag{100}$$

under the condition that we have white noise v_k as well as a white excitation signal s_k and that both random processes are uncorrelated. Note that we have selected not to apply the expectation operators (that is, $\mathbb{E}[\|\mathbf{v}_{\mathrm{ZF}}\|_2^2]$ and $\mathbb{E}[\|\mathbf{f}_{\mathrm{ZF}}\|_2^2]$) in the equation above as we are considering a given thus fixed channel. In case a set of channels with random selection is considered, the expectation operator may be applied. We obviously obtain a noise variance that depends on time and on the state of the adaptive filter. At steady-state the movement of the adaptive filter is expected to have reached an equilibrium (in the mean square). We then obtain

$$\begin{aligned}\sigma_{\tilde{w}_\infty}^2 = &\|\mathbf{v}_{\mathrm{ZF}}\|_2^2 + N_o\|\mathbf{f}_{\mathrm{ZF}}\|_2^2 \\ &+ \mathbb{E}\left[\tilde{\mathbf{g}}_\infty^H \mathbf{H}_u^{-1}\left[\mathbf{H}_l\mathbf{H}_l^H + N_o\mathbf{I}\right]\mathbf{H}_u^{-H}\tilde{\mathbf{g}}_\infty\right]. \end{aligned} \tag{101}$$

Problematic is the last term that we denote by

$$\mathbb{E}\left[\tilde{\mathbf{g}}_\infty^H \mathbf{H}_u^{-1}[\mathbf{H}_l\mathbf{H}_l^H + N_o\mathbf{I}]\mathbf{H}_u^{-H}\tilde{\mathbf{g}}_\infty\right] = \bar{\nu}\mathbb{E}\|\tilde{\mathbf{g}}_\infty\|_2^2.$$

Parameter $\bar{\nu} = \bar{\nu}(N_o)$ is bounded from below by the smallest eigenvalue (which can be as small as zero) and from above by the largest eigenvalue of $\mathbf{H}_u^{-1}[\mathbf{H}_l\mathbf{H}_l^H + N_o\mathbf{I}]\mathbf{H}_u^{-H}$ which can further be bounded by $\mathrm{trace}\{\mathbf{H}_u^{-1}[\mathbf{H}_l\mathbf{H}_l^H + N_o\mathbf{I}]\mathbf{H}_u^{-H}\}$.

Applying the considerations from the previous section we can now compute the relative system mismatch to be

$$S_{\mathrm{rel},\infty} = \alpha\frac{\|\mathbf{v}_{\mathrm{ZF}}\|_2^2 + N_o\|\mathbf{f}_{\mathrm{ZF}}\|_2^2}{\bar{\lambda} - \alpha(1+\bar{\nu})} \tag{102}$$

and the upper step-size bound is eventually

$$\alpha < \frac{\bar{\lambda}_{\min}}{1 + \bar{\nu}(N_o)}. \tag{103}$$

Note that not only $\bar{\nu}$ can have a wide range but also $\bar{\lambda}_{\min} < \bar{\lambda} < \bar{\lambda}_{\max}$. Different to MMSE estimation, we did not include a correction factor γ_s to reflect the correlation of driving sequence s_k as we expect white data sequences only.

If we approximate $\bar{\nu}(N_o) = \bar{\nu}_o + \gamma^2 N_o$, we can turn around the last equation and obtain

$$\gamma^2 N_o < \frac{\bar{\lambda}}{\alpha} - (1 + \bar{\nu}_o). \tag{104}$$

This formulation clearly shows that we expect an upper limit for noise when employing a fixed normalized step-size α. Once this limit is exceeded, the algorithm can become unstable.

B. Simulation Examples

In this section we provide some simulation results to verify our findings. For this purpose, we consider a set of seven channel impulse responses of finite length

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

We select the length of the channel to be $M = 50$ for which the first four impulse responses have decayed considerably. In all simulations relatively strong noise was added of $N_o = 0.01$. We will show all graphic results based on channel $h_k^{(4)}$ and report for which cases we found significant changes. Note that channels $h_k^{(1)}$ and $h_k^{(2)}$ perfectly fit to the example in (87) for $a = 0.8$ and $a = 0.9$, respectively.

Figs. 4 and 5 depict the ZF and MMSE solutions of both channels as well the corresponding convolutions of channel and equalizer, respectively. Both channels appear to be relatively similar in terms of ZF and MMSE performance but can be clearly differentiated due to the relatively strong noise term N_o. We selected the cursor position τ to be in the middle of the equalizer, well knowing that this may not be the optimal position. As the resulting convoluted systems are rather symmetrical it is not expected that other positions change the results dramatically. As we average the Monte Carlo runs only over different noise and transmit symbols, 20 runs were performed which already provide sufficiently smooth and accurate curves.

1) MMSE Equalizer: In our first equalizer experiment we show the classical adaptive MMSE equalizer for channel $h^{(4)}$. Fig. 6 depicts the l_2 norm of the parameter error vector (relative system distance) over iteration numbers for various normalized step-sizes $\alpha = \mu_k\|\mathbf{r}_k\|_2^2$ following (21). As expected, the stability is guaranteed for $0 < \alpha < 2$. The learning behavior is

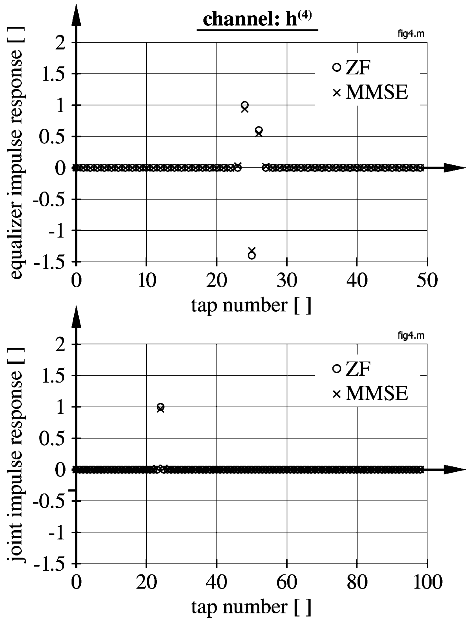

Fig. 4. Upper: ZF and MMSE equalizer impulse response for $h^{(4)}$. Lower: Corresponding convolution of channel and equalizer for $h^{(4)}$.

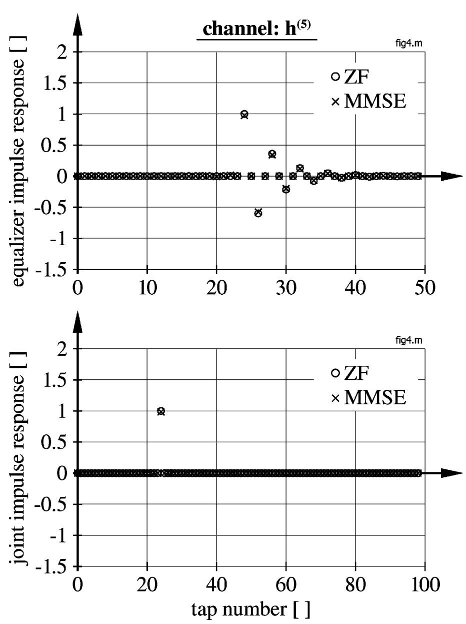

Fig. 5. Upper: ZF and MMSE equalizer impulse response for $h^{(5)}$. Lower: Corresponding convolution of channel and equalizer for $h^{(5)}$.

very much as expected. If we compare the theoretically derived values

$$S_{\mathrm{rel},k} = \frac{\mathbb{E}\|\mathbf{f}_{\mathrm{MMSE}} - \hat{\mathbf{f}}_k\|_2^2}{\|\mathbf{f}_{\mathrm{MMSE}}\|_2^2} = \frac{\alpha\gamma_x\sigma_{\bar{v}_k}^2}{2-\alpha}\frac{1}{\|\mathbf{f}_{\mathrm{MMSE}}\|_2^2}, \quad (105)$$

$$S_{\mathrm{rel},\infty} = \frac{\alpha\gamma_x}{2-\alpha}\left(\frac{\|\mathbf{v}_{\mathrm{MMSE}}\|_2^2}{\|\mathbf{f}_{\mathrm{MMSE}}\|_2^2} + N_o\right) \quad (106)$$

with the simulations in Fig. 6, we find excellent agreement. As expected the MMSE equalizer does not differentiate between various channels and behaves perfectly robust. The step-size bound for $\alpha = 2$ is tight and holds for various sequences independent of the channel and the noise. For all seven channels we obtained very similar results.

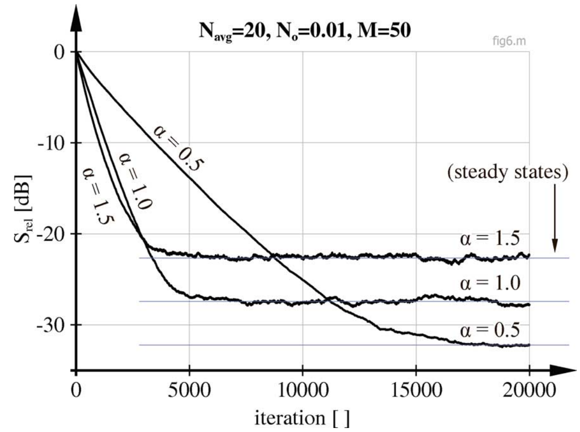

Fig. 6. MMSE equalizer with normalized step-size on channel $h^{(4)}$.

2) ZF Equalizer: We will pick now Algorithm 3 as it is the only algorithm that can work without knowing the channel in form of the triangular matrix $\mathbf{H}_u$. We identify $\mathbf{G} = \mathbf{H}_u^H$. For this algorithm on the other hand we have to find the smallest eigenvalue of $(\mathbf{H}_u\mathbf{H}_u^H)^{-1}\mathbf{H}_u + \mathbf{H}_u^H(\mathbf{H}_u\mathbf{H}_u^H)^{-1}$ in order to find the upper step-size bound which is practically impossible without knowing the matrix.

$$S_{\mathrm{rel},k} = \frac{\mathbb{E}\|\mathbf{g}_{\mathrm{ZF}} - \hat{\mathbf{g}}_k\|_2^2}{\|\mathbf{g}_{\mathrm{ZF}}\|_2^2} \quad (107)$$

$$= \frac{\mathbb{E}\|\mathbf{H}_u^H(\mathbf{f}_{\mathrm{ZF}} - \hat{\mathbf{f}}_k)\|_2^2}{\|\mathbf{H}_u^H\mathbf{f}_{\mathrm{ZF}}\|_2^2}. \quad (108)$$

We apply the adaptive ZF equalizer on channel $h^{(4)}$. We again employed the normalized step-size $\alpha = \frac{\mu_k}{\bar{\mu}_k}$ with $\bar{\mu}_k = \frac{1}{\|\mathbf{s}_k\|_2^2}$ to speed up convergence. Note that due to the QPSK symbols for s_k the norm is constant, $\bar{\mu} = \frac{1}{M}$ and the algorithm can also be interpreted as a fixed step-size algorithm. The results are displayed in Fig. 7. Compared to the adaptive MMSE filter we find a significantly higher convergence speed which is certainly due to the fact that the MMSE algorithm is driven by a strongly correlated signal while the ZF equalizer only "sees" the white data sequence. According to the theoretical derivation the algorithm cannot be guaranteed to converge as the smallest eigenvalue becomes slightly negative (see Table III). However, the derived bound is not tight and the stability bound is found for larger step-sizes at around 0.50. For all other channels, convergence in the mean square sense was expected as $\bar{\lambda}_{\min} > 0$ and the derived step-size bounds showed to be typically conservative.

Using relation (103), we can also derive a step-size bound based on $\bar{\lambda}$ and $\bar{\nu}$.

$$S_{\mathrm{rel},\infty} = \frac{\alpha\sigma_{\bar{v}}^2}{\bar{\lambda} - \alpha(1+\bar{\nu})}\frac{1}{\|\mathbf{H}_u^H\mathbf{f}_{\mathrm{ZF}}\|_2^2} = \frac{\alpha}{\bar{\lambda} - \alpha(1+\bar{\nu})}\frac{\|\mathbf{v}_{\mathrm{ZF}}\|_2^2 + N_o\|\mathbf{f}_{\mathrm{ZF}}\|_2^2}{\|\mathbf{H}_u^H\mathbf{f}_{\mathrm{ZF}}\|_2^2}. \quad (109)$$

We used a least-squares fit of the S_{rel} plots for various step-sizes to fit to the two unknown values $\bar{\lambda}$ and $\bar{\nu}$. They are typically

TABLE III
VARIOUS CHARACTERISTICS FOR THE SEVEN CHANNELS UNDER CONSIDERATION

channel	bound (exp)	$\frac{1}{2}\min\text{eig}\{\cdot\}$	$\frac{1}{2}\max\text{eig}\{\cdot\}$	$\frac{\bar{\lambda}_{\min}}{1+\bar{\nu}_{\max}}$	$\frac{\bar{\lambda}}{1+\bar{\nu}}$	D_0
$\mathbf{h}^{(1)}$	0.55	0.10	1.89	0.018	0.52	0.0058
$\mathbf{h}^{(2)}$	0.65	0.20	1.79	0.054	0.68	$1.4\ 10^{-5}$
$\mathbf{h}^{(3)}$	0.57	0.25	2.5	0.096	0.63	$5.9\ 10^{-8}$
$\mathbf{h}^{(4)}$	0.50	-0.005	3.0	0.018	0.53	$4.3\ 10^{-6}$
$\mathbf{h}^{(5)}$	0.62	0.62	2.41	0.259	0.65	0.0021
$\mathbf{h}^{(6)}$	0.40	0.50	13	0.019	0.46	1
$\mathbf{h}^{(7)}$	0.40	0.50	25.5	0.0098	0.32	1

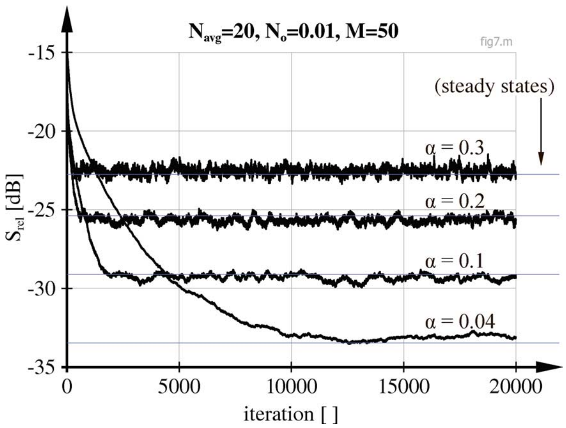

Fig. 7. ZF equalizer with normalized step-size on channel $h^{(4)}$.

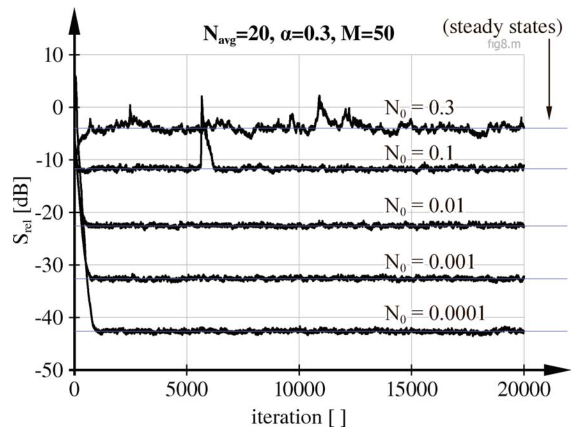

Fig. 8. ZF equalizer with fixed normalized step-size on channel $h^{(4)}$ when varying additive noise variance N_o.

around one ($\bar{\lambda}$) and 0.5 ($\bar{\nu}$) and thus the corresponding bound around 0.65 which is a good agreement with our observations.

It is worth studying the entries of Table III. On the second column we filled in the step-size bound that we found experimentally. This is to compare with the third column which provides a conservative step-size bound, the fourth column which provides a bound for which we certainly expect divergence as well as the fifth column with an extreme conservative bound and the sixth column with a bound derived from our least squares fit. The later is tedious and difficult to be used in algorithmic design but rather accurate. For the first five channels we find as good agreement as we had in the adaptive MMSE algorithm. For the last two channels however, prediction turns out to be more difficult. Even our conservative bounds from the third column turn out to be a bit too high. An explanation can be found in the matrix $\mathbf{H}_u\mathbf{H}_u^H$ that is very ill conditioned. In the last column of Table III, we list the distortion measure

$$D_0 = \frac{\sum |c_k|}{|c_\tau|} - 1,$$

c_k being the convolution of the adaptive filter and the channel impulse response. In [10] it is argued that the adaptive ZF algorithm only works for values smaller than one. What we found indeed is that the algorithm still works however, loses a lot of the estimation quality that we find in other channels. The quality of the adaptive ZF equalizer is thus very much dependent on the actual channel.

Remember that the adaptive ZF algorithm is not robust. When switching from random QPSK data sequences to worst case QPSK sequences, we found the algorithm to be non convergent, as expected. We could not find a step-size small enough to ensure convergence under such worst case sequences for any of the channels.

We repeated the experiment for a fixed normalized step-size $\alpha = 0.3$ but varied the noise variance. The result is shown in Fig. 8. As predicted in (104), the stability bound varies with the noise and in our example for noise variances larger than one, the algorithm indeed became unstable.

VII. CONCLUSION

The adaptive MMSE equalizer has shown to be robust, guaranteeing l_2–stability for a fixed range of step-sizes independent of additive noise or the channel itself. Its steady-state quality was derived analytically and showed excellent agreement with the simulation examples.

Novel criteria have been found to ensure convergence of a well-known adaptive ZF receiver. Different to the general belief these criteria strongly depend on the channel that is to be equalized as well as on the additive noise that is present. Simulation results verified the correctness of these findings. We were able to derive explicit steady-state results of adaptive ZF receivers. As a side result, a class of adaptive gradient type algorithms with arbitrary but time-invariant matrix step-size can now be treated in terms of mean square stability. Note however that the conditions we need to apply cannot be satisfied for all kind of channels. According to Table III, certain criteria on the channel are to be satisfied. We observed that although the adaptive ZF equalizer algorithm behaves stable under random sequences, a strict l_2–stability does not hold and worst case sequences under which the algorithm become unstable even under very small step-sizes can be found.

ACKNOWLEDGMENT

The author would like to thank the anonymous reviewers for their constructive help and S. Caban for preparing the figures as well as R. Dallinger for providing a Matlab code framework.

REFERENCES

[1] R. W. Lucky, "Automatic equalization for digital communication," *Bell Syst. Tech. J.*, vol. 44, pp. 547–588, Apr. 1965.

[2] R. W. Lucky, "Techniques for adaptive equalization of digital communication systems," *Bell Syst. Tech. J.*, vol. 45, pp. 255–286, Apr. 1966.

[3] A. Gersho, "Adaptive equalization in highly dispersive channels for data transmission," *Bell Syst. Tech, J.*, vol. 48, pp. 55–70, 1969.

[4] G. Forney, "Maximum likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inf. Theory*, vol. 18, no. 3, pp. 363–378, 1972.

[5] J. M. Cioffi, G. Dudevoir, M. Eyuboglu, and G. D. F. Jr, "MMSE decision feedback equalization and coding-Part I," *IEEE Trans. Commun.*, vol. 43, pp. 2582–2594, Oct. 1995.

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

[7] J. R. Treichler, I. Fijalkow, and C. R. Johnson Jr., "Fractionally spaced equalizer. How long should they really be?," *IEEE Signal Process. Mag.*, vol. 13, pp. 65–81, May 1996.

[8] T. Rappaport*, Wireless Communications*. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[9] E. A. Lee and D. G. Messerschmitt*, Digital Communications*. Norwell, MA: Kluwer, 1994.

[10] J. Proakis*, Digital Communications*. New York: McGraw-Hill, 2000.

[11] S. Haykin*, Adaptive Filter Theory*, 4 ed. Englewood Cliffs, NJ: Prentice-Hall, 2002.

[12] M. Rupp and A. Burg, "Algorithms for equalization in wireless applications," in *Adaptive Signal Processing: Application to Real-World Problems*. New York: Springer, 2003.

[13] M. Rupp and A. H. Sayed, "A time-domain feedback analysis of filtered-error adaptive gradient algorithms," *IEEE Trans. Signal Process.*, vol. 44, pp. 1428–1440, Jun. 1996.

[14] A. H. Sayed and M. Rupp, "Robustness issues in adaptive filters," in *The DSP Handbook*. Boca Raton, FL: CRC Press, 1998.

[15] A. H. Sayed*, Fundamentals of Adaptive Filtering*. New York: Wiley, 2003.

[16] R. Dallinger and M. Rupp, "A strict stability limit for adaptive gradient type algorithms," in *Conf. Rec. 43rd Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 2009, pp. 1370–1374.

[17] A. H. Sayed and M. Rupp, "Error-energy bounds for adaptive gradient algorithms," *IEEE Trans. Signal Process.*, vol. 44, no. 8, pp. 1982–1989, Aug. 1996.

[18] M. Rupp and J. Cezanne, "Robustness conditions of the LMS algorithm with time-variant matrix step-size," *Signal Process.*, vol. 80, no. 9, pp. 1787–1794, Sep. 2000.

[19] S. Makino, Y. Kaneda, and N. Koizumi, "Exponentially weighted step-size NLMS adaptive filter based on the statistics of a room impulse response," *IEEE Trans. Speech Audio Process.*, vol. 1, pp. 101–108, Jan. 1993.

[20] D. L. Duttweiler, "Proportionate normalized least mean square adaptation in echo cancellers," *IEEE Trans. Speech Audio Process.*, vol. 8, pp. 508–518, Sep. 2000.

[21] A. Feuer and E. Weinstein, "Convergence analysis of LMS filters with uncorrelated Gaussian data," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, pp. 222–230, Feb. 1985.

Markus Rupp (M'03–SM'06) received the Dipl.-Ing. degree from the University of Saarbrücken, Saarbrücken, Germany, in 1988 and the Dr.Ing. degree from Technische Universität Darmstadt, Darmstadt, Germany, in 1993, where he worked with E. Hänsler on designing new algorithms for acoustical and electrical echo compensation.

From November 1993 to July 1995, he held, with S. Mitra, a postdoctoral position with the University of California, Santa Barbara, where he worked with A. H. Sayed on a robustness description of adaptive filters with impact on neural networks and active noise control. From October 1995 to August 2001, he was a Member of Technical Staff with the Wireless Technology Research Department, Bell Laboratories, Crawford Hill, NJ, where he worked on various topics related to adaptive equalization and rapid implementation for IS-136, 802.11, and the Universal Mobile Telecommunications System. Since October 2001, he has been a Full Professor of digital signal processing in mobile communications with the Vienna University of Technology, where he founded the Christian Doppler Laboratory for Design Methodology of Signal Processing Algorithms, at the Institute of Communications and Radio-Frequency Engineering, in 2002. He served as Dean from 2005 to 2007. He is the author or a coauthor of more than 350 papers and is the holder of 15 patents on adaptive filtering, wireless communications, and rapid prototyping, as well as automatic design methods.

Dr. Rupp is currently an Associate Editor for the EURASIP *Journal of Advances in Signal Processing* and the EURASIP *Journal on Embedded Systems*. He was an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2002 to 2005. He has been an Administrative Committee Member of EURASIP since 2004 and served as the President of EURASIP from 2009 to 2010.