Analysis of an LMS Linear Equalizer for Fading Channels in Decision Directed mode

Veeraruna Kavitha and Vinod Sharma
Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore, 560012, India.
kavitha@pal.ece.iisc.ernet.in vinod@ece.iisc.ernet.in

Abstract—We consider a time varying wireless fading channel, equalized by an LMS linear equalizer in decision directed mode (DD-LMS-LE). We study how well this equalizer tracks the optimal Wiener equalizer. Initially we study a fixed channel. For a fixed channel, we obtain the existence of DD attractors near the Wiener filter at high SNRs using an ODE, approximating the DD-LMS-LE. We also show, via examples, that the DD attractors may not be close to the Wiener filters at low SNRs. Next we study a time varying fading channel modelled by an Auto-regressive (AR) process of order 2. The DD-LMS equalizer and the AR process are jointly approximated by the solution of a system of ODEs (ordinary differential equations). Using these ODEs, via some examples, the LMS equalizer is shown to track the instantaneous Wiener filter when the SNR is high, while it may not track the same at low SNRs.

Key words: Fading channels, LMS, Decision-Directed mode, tracking performance, ODE approximation.

I. INTRODUCTION

A channel equalizer is an important component of a communication system and is used to mitigate the ISI (inter symbol interference) introduced by the channel. The equalizer depends upon the channel characteristics. In a wireless channel, due to multipath fading, the channel characteristics change with time. Thus it may be necessary for the channel equalizer to track the time varying channel in order to provide reasonable performance.

An equalizer is most commonly designed using the Minimum Mean Square Error (MMSE) criterion ([5], [9], [13]). The optimal MSE (MMSE) equalizer, also called the Wiener filter (WF), is either calculated directly using the training sequence or indirectly using a training based channel estimate. The WF often involves a matrix inverse computation. Hence a computationally simpler iterative algorithm, the Least Mean Square (LMS), is commonly used as an alternative.

A Least Mean Square linear equalizer (LMS-LE), designed using training sequence, is a simple equalizer and is extensively used ([5], [9], [13]). For a fixed channel its convergence to the Wiener filter has been studied in [1], [11] (see also the references therein). For a time varying channel, theoretical tracking behavior (how well an LMS-LE tracks the instantaneous Wiener filter) has been studied in [7] (its tracking behavior is also studied via simulations, approximations and upper bounds on probability of error in [5], [8], [14]).

To study the tracking behavior theoretically, one needs to have a theoretical model of the fading channel. Auto Regressive (AR) processes have been shown to model such channels quite satisfactorily ([10], [15]). Infact, it is sufficient to model the fading channel by an AR(2) process ([8], [10], [15]). In [7] a time varying wireless channel has been modeled by an AR(2) process. It is shown that for a stable/unstable channel (the poles are either inside or outside the unit circle) the LMS-LE tracks the instantaneous WF. It is also shown that for a marginally stable channel (the poles are on the unit circle), the distance between the LMS-LE and the instantaneous WF remains bounded.

A training based LMS-LE becomes inefficient in a wireless scenario. Due to time varying nature of the wireless channel, the training based LMS-LE, needs frequent transmission of the training sequence. Therefore, a significant (∼18% in GSM) fraction of the channel capacity is consumed by the training sequence. The usual blind equalization techniques have also been found to be inadequate [4] due to their slow convergence and/or high computational complexity. In [6] it is shown using information theoretic arguments that a semi-blind method can be a better alternative for a time varying channel. In such scenarios, the decision directed LMS-LE (the training sequence is replaced by the decisions of the symbols after some time and hence such a DD-LMS-LE can also be viewed as a semi blind algorithm) may prove to be a good alternative ([11]).

However, one needs a theoretical understanding of the DD-LMS-LE prior to its use. In [11], it has been shown that the DD-LMS-LE for a fixed channel converges to WF almost surely, if the initializer is sufficiently close to the WF. But they assume bounded channel output and perfect equalizability. These assumptions are not satisfied in most of the practical channels, e.g., an AWGN channel with ISI. In [11], the authors also deal with the AWGN noise and observe that the DD attractors are away from the WFs when the noise is non zero. However, they restrict themselves to a single tap equalizer. But ISI can be mitigated only with equalizers of length greater than one.

In this paper we first study a DD-LMS-LE on a fixed channel. We obtain an ODE approximation for its trajectory and show that the ODE’s attractors are close to the corresponding...
Wiener filters as the noise variance tends to zero. We also show via some examples that for large noise variances (i.e., at low SNRs) the DD attractors will not be close to the WFs.

Next we consider a DD-LMS-LE tracking a time varying wireless channel modelled by an AR(2) process. We use the ODE approximation of the AR(2) process (obtained by us in [7]) and obtain an ODE approximation for a DD-LMS-LE tracking an AR(2) process. Using this ODE approximation we illustrate via some examples that a DD-LMS-LE can indeed track an AR(2) process reasonably (the DD-LMS-LE trajectory is quite close to the instantaneous WFs) as long as the SNR is high. With increase in noise variance the DD algorithm loses out. We are not aware of any theoretical study on the tracking behavior of a DD-LMS-LE.

The paper is organized as follows. In Section II we explain our model. Section III studies the decision directed (DD) algorithm on a fixed channel. Section IV obtains the ODE approximation for a time varying channel. Section V provides examples to demonstrate the ODE approximations and the proximity of the DD attractors to that of the WFs. Section VI concludes the paper. The appendices contain some details on the proofs.

II. System Model, Notations and Assumptions

We consider a system consisting of a time varying (wireless) channel followed by an adaptive linear equalizer. The input of the channel $s_k$ comes from a finite alphabet and forms a zero mean IID (independent, identically distributed) process. The channel is a time varying finite impulse response (FIR) linear filter $\{Z_k\}$ of length $L$ followed by additive white Gaussian noise $\{n_k\}$. We assume $n_k \sim \mathcal{N}(0, \sigma^2)$. We also assume that $\{s_k\}$ and $\{n_k\}$ are independent of each other. The channel output at time $k$ is

$$u_k = \sum_{i=0}^{L-1} Z_{k,i} s_{k-i} + n_k,$$

where $Z_{k,i}$ is the $i^{th}$ component of $Z_k$. At the receiver the channel output $u_k$ passes through a linear equalizer $\theta_k$ and then through a hard decoder $Q$. The output of the hard decoder at time $k$ is $\hat{s}_k$.

In this paper we consider a DD-LMS-LE. For this system the LE $\theta_k$, of length $M$ at time $k$, is initially updated using a training sequence. After a while, the training sequence is replaced by the decisions made at the receiver about the current input symbol $s_k$. This is the decision directed (DD) mode.

The output of the hard decoder $Q \hat{s}_k = Q(\theta_k U_k)$, where $S_k, N_k, U_k$ are the appropriate length input, noise and channel output vectors respectively. We assume $E[S_k S_k^T] = I$. Note that, $U_k = \pi_s S_k + N_k$, where the convolutional matrix $\pi_s$ depends upon the channel co-efficients $Z_k \cdots Z_{k-M+1}$ and is given by,

$$
\begin{bmatrix}
Z_{k,1} & Z_{k,2} & \cdots & Z_{k,L} & 0 & \cdots & 0 \\
0 & Z_{k-1,1} & \cdots & Z_{k-1,L-1} & 0 & \cdots & 0 \\
0 & \vdash & \cdots & 0 & Z_{k-M+1,1} & \cdots & Z_{k-M+1,L}
\end{bmatrix}
$$

In this paper we assume the input to be BPSK, i.e., $s_k \in \{+1, -1\}$. This assumption is made to simplify the discussions and can easily be extended to any finite alphabet source. For BPSK, $Q(x) = 1(x>0) - 1(x\leq0)$.

In DD mode the LE is updated using $(\hat{s}_k(\theta) = Q(\theta^T U_k))$,

$$\theta_{k+1} = \theta_k - \mu_k U_k (\theta_k^T U_k - \hat{s}_k(\theta_k)) \quad (1)$$

where $\mu_k$ is a positive sequence of step-sizes.

Initially we study the DD system when the channel is fixed, i.e., $Z_k = Z$ for all $k$. Later on, we consider a time varying channel when the channel is modelled by an AR(2) process:

$$Z_{k+1} = d_1 Z_k + d_2 Z_{k-1} + \mu W_k$$

where $W_k$ is an IID sequence, independent of the processes $\{s_k\}, \{n_k\}$. An AR(2) process can approximate a wireless channel quite realistically ([8], [15]) and has been approximated by an ODE in [7]. Using this ODE approximation we obtain the required tracking performance analysis.

The fixed channel is studied in Section III while the time varying in Section IV.

III. DD-LMS-LE For a Fixed Channel

In this section, we assume that the channel is fixed, i.e., $Z_k = Z$ for all $k$. We first obtain an ODE approximation for it when the step-sizes $\mu_k \to 0$. We obtain the existence of DD attractors (ODE) near the corresponding Wiener filters at high SNRs. We show that as noise variance $\sigma^2$ tends to zero, these DD attractors tend to the corresponding WFs.

A. ODE approximation

DD-LMS-LE for a fixed channel has been approximated by an ODE in [1]. We start our analysis with this ODE. Towards this goal, as a first step the DD-LMS-LE algorithm (1) is rewritten to fit in the setup of [1], p. 276,

$$\xi_k := \left[ S_k^T U_k^T \hat{s}_k \right]^T,$$

$$H(\theta, \xi) := U^T (\theta^T U - \hat{s}),$$

$$\theta_k = \theta_{k-1} - \mu_{k-1} H(\theta_{k-1}, \xi_k).$$

Let $\theta(t, t_0, a)$ denote the solution of the following ODE with initial condition $\theta(t_0) = a$ ($\pi_Z$ is the convolutional matrix $\pi_k$ of the previous section for a fixed channel $Z$),

$$\dot{\theta}(t) = -R_{uu} \theta(t) + R_{us} \theta(t),$$

$$R_{uu} = \pi_Z \pi_Z^T + \sigma^2 I,$$

$$R_{us} = E \left[ U Q (U^T \theta) \right].$$

It is easy to see that the Markov chain $\{\xi_k\}$ has a unique stationary distribution for every $\theta$ and that the DD-LMS satisfies all the required hypothesis of Theorem 13, p. 278, [1]. Hence one can approximate its trajectory on any finite time scale with the solution of the above ODE. We reproduce the precise result below.

For any initial condition $\theta_0$ and for any finite time $T$, with $t(r) := \sum_{k=0}^r \mu_k$, $m(n, T) := \max_{r \geq n} \{t(r) - t(n) \leq T\}$,

$$\sup_{\{r \leq m(n, T)\}} \left| \theta_r - \theta(t(r), t(n), \theta_0) \right| = 0$$
as \( n \to \infty \), whenever \( \sum_k \mu_k^{1+\delta} < \infty \) for some \( \delta < 0.5 \), \( \mu_k \leq 1 \) for all \( k \) and \( \liminf_k \mu_k^{1+\delta} > 0 \) for all \( r \).

As in Lemma 1 of Appendix C one can show that, the above ODE has a unique global bounded solution for any finite time. We will also show the existence of attractors for this ODE, near WF, at least at high SNRs in the next subsection.

The ODE approximation suggests that, if the decision directed mode of the system is started in the region of attraction of an attractor of the ODE, the DD-LMS-LE will converge to that attractor. We will show below that under high SNR, an attractor of the above ODE will be close to the WF. Thus, the DD mode should be started when the LE is within the region of attraction of this attractor (e.g., when the ‘eye’ has opened as in [11]). To reach the region of attraction, one starts with a ‘good’ initial condition and then uses a training sequence. The region of attraction of a desired attractor depends upon the channel \( Z \), the input distribution and \( \sigma^2 \). However, for a given set of parameters it may be computed via the various available methods ([13]).

B. Relation between DD attractors and WFs

In the following we study the desired attractors in more detail.

Using implicit function theorem ([12]), we show that the DD-LMS attractors are close to the WFs at high SNRs. Let (note that \( R_{uu} \), \( R_{us} \) depend on \( \sigma^2 \)),

\[
f(\theta, \sigma^2) \triangleq -R_{uu} \theta + R_{us} \theta, \\
\theta^{*}(\sigma^2) \triangleq R_{uu}^{-1} R_{us}, \text{ where } R_{uu} = E[U^2].
\]

Note that \( \theta^{*}(\sigma^2) \) represents the WF at noise variance \( \sigma^2 \), while a DD attractor is a zero of the function \( f \). At \( \sigma^2 = 0 \), by invertibility of \( \pi \pi_Z, R_{us}(\theta^{*}(0)) \) equals \( R_{us} \). Hence \( \theta^{*}(0) \), the WF at zero noise variance, also becomes a DD attractor. Thus, \( \theta^{*}(0,0) \) is a zero of the function \( f(\cdot, \cdot) \).

One can easily verify the following :

- \( f(\theta, 0) = -R_{uu} \theta + R_{us} \theta \) whenever \( \theta \in B(\theta^{*}(0), \epsilon) \) for some \( \epsilon > 0 \), where \( B(x, r) = \{ y : |x - y| \leq r \} \).
- Thus, \( \frac{\partial f}{\partial \theta}(\theta^{*}(0), 0) = -R_{uu} \) and \( R_{uu} \) is invertible.

- By Lemma 2, \( f \) is continuously differentiable .

By implicit function theorem (Theorem 3.1.10, p. 115, [2]), there exists a \( \delta > 0 \) and a unique differentiable function \( g \) of \( \sigma^2 \) such that, for all \( 0 < \sigma^2 \leq \delta \),

\[
f(g(\sigma^2), \sigma^2) = 0.
\]

Since \( \frac{\partial f}{\partial \theta}(\theta^{*}(0), 0) = -R_{uu} \) is negative definite and \( \frac{\partial f}{\partial \sigma^2} \) is continuous at \( (\theta^{*}(0), 0) \), \( \frac{\partial f}{\partial \sigma^2} \) is negative definite on a small neighborhood around \( (\theta^{*}(0), 0) \). Thus zeros, \( g(\sigma^2) \) are DD attractors for all \( \sigma^2 \) small enough. We represent these DD attractors at noise variance \( \sigma^2 \), by \( \theta^{*}_d(\sigma^2) \).

We will now relate the DD attractors, \( \theta^{*}_d(\sigma^2) = g(\sigma^2) \), to the corresponding WFs, \( \theta^{*}(\sigma^2) \) when \( \sigma^2 \) is close to zero. Define \( h(\sigma^2) = R_{us}(\theta^{*}_d(\sigma^2)) \). Using dominated convergence theorem and continuity of the map \( g \), one can see that \( h(\sigma^2) \to h(0) = R_{us} \) whenever \( \sigma^2 \to 0 \). Define

\[
m(\theta, \sigma^2, \eta) = -R_{us} \theta + R_{us} + \eta.
\]

At any noise variance, \( \sigma^2, m(\theta^{*}(\sigma^2), \sigma^2, 0) = 0 \) as \( \theta^{*}(\sigma^2) \) is the unique WF at noise variance \( \sigma^2 \). Also, the function \( m \) is \( C^\infty \) (infinitely differentiable) in all parameters (note that \( R_{us} \) is a fixed vector independent of all the parameters whenever input is IID). Hence once again using implicit function theorem at any noise variance, \( \sigma^2 \) there exist \( \alpha, \beta > 0 \) and a continuous function \( \gamma(\cdot, \cdot) \) such that,

\[
m(\gamma(\sigma^2, \eta), \sigma^2, \eta) = 0 \text{ when } |\eta| \leq \beta, \text{ and } |\sigma^2 - \sigma^2_0| \leq \alpha.
\]

Hence by continuity of the functions \( \gamma \) and \( h \), the WF (which is also given by \( \gamma(\sigma^2, 0) \)) will be close to the DD attractor, \( \gamma(\sigma^2, [R_{us} - R_{us}(\theta^{*}_d(\sigma^2))] \)) at low noise variances.

IV. DD-LMS-LE TRACKING AN AR(2) PROCESS

In this section we present the ODE approximation for the linear equalizer (1) in decision directed mode when the channel is modeled as an AR(2) process (2). Here we set the step-size \( \mu_k = \mu \) for all \( k \), to facilitate tracking of the time varying channel. We use Theorem 3 in [7] to obtain the required ODE approximation.

We will show below that the trajectory \( (\theta_k, Z_k) \) given by equations (1), (2) can be approximated by the solution of the following system of ODEs,

\[
\begin{align*}
\dot{Z}(t) &= \frac{1}{1 + d_2} [\eta Z(t) + E(W_1)], \\
\dot{\theta}(t) &= -R_{us}(Z(t)) \theta(t) + R_{us}(\theta(t), Z(t)),
\end{align*}
\]

\[
\eta \triangleq \frac{d_1}{d_1 + d_2 - 1},
\]

\[
R_{us}(Z) \triangleq E_Z \left[ U(Z) U(Z)^T \right] = (\pi \pi_Z + \sigma^2 I),
\]

\[
R_{us}(\theta, Z) \triangleq E_Z \left[ U(Z) \hat{\theta}(\theta) \right].
\]

By Lemma 1, the above system of ODEs have unique bounded global solutions for any finite time. Let \( Z(t, t_0, Z), \theta(t, t_0, \theta) \) represent the solutions of the ODEs (3), (4) with initial conditions \( Z(t_0) = Z, \theta(t_0) = \theta \).

Let \( V^{\triangle}(Z, \theta) \). With \( d_2 \in (-1, 1) \), we prove Theorem 1.

Theorem 1: For any finite \( T > 0 \), for all \( \delta > 0 \) and for any initial condition \( (G, \theta, Z) \), with \( d_2 Z_{-1} + d_1 Z_0 = Z \) and \( \theta_0 = \theta \),

\[
P_{G,Z,\theta} \left\{ \sup_{k : t(k) \leq T} \left| V_k - V(t(k), 0, (Z, \theta)) \right| \geq \delta \right\} \to 0,
\]

as \( \mu \to 0 \), uniformly for all \( (Z, \theta) \in Q \), if \( Q \) is contained in the bounded set containing the solution of the ODEs (3), (4) till time \( T \).

Proof: Please see the Appendix A.

We can obtain the ODE approximation for \( d_2 = -1 \) case also as in [7]. The only difference here is that the channel-approximating ODE is a second order ODE

\[
\frac{d^2 Z(t)}{dt^2} = E(W_1) + \eta Z(t).
\]
The above theorem again holds, if we replace $V(k\mu, 0, (Z, \theta))$ in the statement of the above theorem with the ordered pair, $(Z(k\mu, 0, Z), \theta(k\mu, 0, \theta))$ (more details are in [7]).

One can easily see that the solution of the channel (AR(2) process) ODE is,

$$Z(t) := \begin{cases} 
C_1 \frac{E(W)}{\eta t^2} - \frac{E(W)}{\eta}, & \eta \neq 0, d_2 \in (-1, 1], \\
E(W) \frac{E(W)}{1-d_2^2 t} + C_1, & \eta = 0, d_2 \in (-1, 1], \\
C_1 \cosh(\sqrt{\eta} t) - \frac{E(W)}{\eta}, & \eta > 0, d_2 = -1, \\
C_1 \cos(\sqrt{\eta} t) + \frac{E(W)}{\eta}, & \eta < 0, d_2 = -1, \\
E(W) \frac{E(W)}{t^2} + C_1, & \eta = 0, d_2 = -1.
\end{cases}$$

The approximating ODE (4) suggests that, its instantaneous attractors will be same as the DD-LMS-LE attractors obtained in the previous section when the channel is fixed at the instantaneous value of the channel ODE (3). We have shown in the previous section that these attractors are close to the WF at high SNRs. We will verify the same behavior for tracking, using some examples, in the next section.

One of the uses of the above ODE approximation is that, one can study the tracking behavior of the DD-LMS (e.g., proximity of its trajectory to the instantaneous WFs) using this ODE. This is done in the next section. Further, one can also obtain instantaneous theoretical performance measures (approximate) like BER, MSE.

V. Examples

In this section we illustrate the theory developed so far using some commonly used examples.

We first consider a fixed channel, $Z = [.41, .82, .41]$ in Figure 1. The channel of this example is very widely used (see p. 414, [5] and p. 165, [4]). We use a two tap linear equalizer. We plot the DD-LMS-LE, its ODE approximation and the Wiener filter for two values of noise variances $\sigma^2 = 0.01, 1$ in this Figure. We can see that the ODE approximation is quite accurate for all the coefficients. We can also see that the DD-LMS coefficients as well as their ODE approximations converge to the DD attractor for both the noise variances The ODE approximation thus confirms that with high probability the realizations of the DD-LMS trajectory (the DD-LMS trajectory in the figure being one such realization) converge to the attractor. One can see from this example that the DD-LMS attractors are close to the corresponding Wiener filters at high SNRs ($\sigma^2 = 0.01$) as is shown theoretically in Section III, but are away from the same at low SNRs ($\sigma^2 = 1$).

We next consider two examples of a time varying channel equalized by a four tap equalizer. We consider stable and marginally stable channels in Figure 2, 3 respectively. For both the examples, the mean channel is a constant multiple of the commonly used raised cosine channel, $[-0.4, 1, 0.6, -0.3, 0.1]$. The AR parameters, $d_1 = .497$, $d_2 = 0.5$ and $\mu = 0.0007$ is used for the stable channel, while the same parameters for the marginally stable channel are set at 1.99999, $-1$ and 0.000001 respectively (these parameter are chosen appropriately to obtain a suitable period for oscillations as from (5) the period of oscillation is controlled by $\sqrt{1-d_1-d_2}$). Both the examples are run under high SNR conditions ($\sigma^2$ equals 0.1, 0.05 for marginally stable and stable channels respectively). In both the examples, the DD-LMS and the ODE are started with the initial value of the WF. One can see from both the figures that the ODE once again approximates the DD-LMS quite accurately. Also, the DD-LMS and the ODE track the instantaneous WF quite well. We further plot the instantaneous BER of the DD-LMS, the ODE and the WF in Figures 4, 5 respectively for the stable and the marginally stable channels. One can see that the performance of the DD-LMS and ODE are quite close to that of the WFs throughout the time axis. The proximity of the ODE solution and the BER once again indicate that with high probability the realizations of DD-LMS track the instantaneous WFs.

Next we plot the DD-LMS, the ODE and the instantaneous WFs at two different noise variances in Figure 6 for a marginally stable channel. It is evident from the figure that the LMS-LE in DD mode, can track the channel variations at high SNR ($\sigma^2 = 0.05$), while it loosen out at low SNRs ($\sigma^2 = 1$).
VI. CONCLUSIONS

We obtain theoretical performance analysis of an LMS linear equalizer in decision directed mode. We first study a decision directed LMS-LE for a fixed channel. We show the existence of DD attractors in the vicinity of the WFs at high SNR's, using an ODE, which approximates the LMS-LE trajectory in decision directed mode. The same conclusion is also illustrated using some examples in Section V. We also show via examples that, a DD attractor may be away from the Wiener filter at low SNRs. We thus conclude that at high SNR's, one can update the LMS-LE in decision directed mode, track the instantaneous WF at high SNRs. We also show that, LMS-LE in decision directed mode, can also be used to form a good equalizer.

APPENDIX A

Proof of Theorem 1 : Defining \( G_{k+1} = [S_k^T, U_k^T]^T \), one can rewrite the AR process and the DD equalizer adaptation as,

\[
Z_{k+1} = (1 - d_2)Z_k + d_2 Z_{k-1} + \mu(W_k + \eta Z_k)
\]

\(
\theta_{k+1} = \theta_k + \mu H_1(\theta_k, G_{k+1})
\)

\( H_1(\theta_k, G_{k+1}) \triangleq -U_k(\theta_k^T U_k - \hat{s}_k)
\)

\( = -U_k(\theta_k^T U_k - Q(U_k^T \theta_k)). \)

This is similar to the general system (5), (6) of Appendix B. Hence by Theorem 2, it suffices to show that our system satisfies the assumptions A.1-4, B.1-4 of Appendix B (which are reproduced from [7]) and that the above system of ODEs has a bounded solution for any finite time.

The AR(2) process \( \{Z_k\} \) in (2) clearly satisfies the assumptions A.1 - A.4 as is shown in [7]. Assumption B.2 is satisfied as for any compact set \( Q \) and for any \( \theta \in Q \),

\[
|H_1(\theta, G)| \leq 2 \left[ \max_{\theta \in Q} \left\{ 1, \sup_{\theta \in Q} |\theta| \right\} \right] (1 + |G|^2).
\]

Fixing channel \( Z_k = Z \) for all \( k \), we obtain the transition kernel \( \Pi_Z(\cdot; \cdot) \) for \( \{G_k\} \) which is a function of \( Z \) alone. Thus condition B.1 is satisfied. It is easy to see that \( G_k(Z) \) has a stationary distribution given by,

\[
\Psi_Z([s_1, s_2, \cdots, s_n] \times A_1) = \text{Prob}(S = [s_1, s_2, \cdots, s_n])
\]

\[
\text{Prob}(N \in A_1 - \pi Z[s_1, s_2, \cdots, s_n]^T),
\]

where \( \pi_Z \) is the \( M \times M + L - 1 \) length convolutional matrix formed from vector \( Z \) (as in the fixed channel case) and \( S, N \) are the input and noise vectors of length \( M + L - 1, M \) respectively. Define,

\[
h_1(\theta, Z) \triangleq E_Z(H_1(\theta, G(Z)) = -R_{uu}(Z)\theta + R_{uv}(\theta, Z),
\]

\[
\nu_{\theta, Z}(G) \triangleq \sum_{k \geq 0} \Pi_Z^k(H_1(\theta, G) - h_1(Z, \theta)).
\]

By Lemma 2 of Appendix C, \( h_1 \) is locally Lipschitz. Thus conditions B.3 a, b are met.

We now prove condition B.3c using Proposition 10, p. 270, [1]. \( G_k \) is a linear dynamic process depending upon the channel realization \( Z \) and it can be written as,

\[
G_{k+1} = A(Z)G_k + B(Z)W_{k+1}, \text{ where,}
\]

\[
A(Z) = \begin{bmatrix}
J_L & P \\
0_{L+M-1 \times L} & J_{L+M-1}
\end{bmatrix},
\]

\[
B(Z) = \begin{bmatrix}
Z_1 & 0 & \cdots & 0 \\
1 & 0 & \cdots & 0
\end{bmatrix}^T,
\]

\[
W_{k+1} = [s_k, n_k].
\]

In the above definitions, \( 0_n \) is a \( n \times n \) zero matrix. The matrix \( J_n \) is an \( n \times n \) shift matrix, and \( P \) a \( L \times L + M - 1 \) matrix and these are given by,

\[
J_n = \begin{bmatrix}
0_{1 \times n} \\
I_{n \times n} & 0_{n-1 \times 1}
\end{bmatrix},
\]

\[
P = \begin{bmatrix}
Z_2 & Z_3 & \cdots & Z_L \\
0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0
\end{bmatrix}.
\]

It is easy to see that, \( A^n(Z) = 0 \) for all \( n \geq \max\{L, L + M - 1\} \) for all \( Z \) as it involves the powers of shift matrices \( J_L, J_{L+M-1} \), which satisfy \( J_n^n = 0 \). By Lemma 3, the function \( \Pi_{\theta, Z} H(\theta, G) \) is \( L_i(R^n) \). Now all other conditions of Proposition 10, p. 270, [1] are satisfied trivially (because \( A(Z) \) and \( B(Z) \) are linear in \( Z \) ) and hence Proposition 10 holds and therefore, B.4.c holds with \( \lambda = 1 \).

The condition B.4 is trivially met as for any \( n > M + L - 1 \), the expectation does not depend upon the initial condition \( G \) but is bounded based on the compact set \( Q \) and because of the Gaussian random variable \( N \) and discrete random variable \( S \).

By Lemma 1, the DD-LMS ODE has a unique bounded solution for any finite time.
APPENDIX B

We consider the following general system,

\[ Z_{k+1} = (1 - d_2)Z_k + d_2Z_{k-1} + \mu H(Z_k, W_k), \]  
\[ \theta_{k+1} = \theta_k + \mu H_1(Z_k, \theta_k, G_{k+1}), \]

where \( \{W_k\} \) and \( H \) satisfy the conditions A.1-A.4 and the equation (6) satisfies the assumptions B.1 to B.4 given below.

We will show that the above equations can be approximated by the solution of the ODEs

\[ \dot{Z}(t) = \frac{1}{1 + d_2} E[H(Z(t), W(t))], \]  
\[ \dot{\theta}(t) = h_1(Z(t), \theta(t)), \]

where the function \( h_1(\cdot, \cdot) \) is defined in the assumptions given below.

We make the following assumptions, which are similar to that in [1] and [7]. Let \( D \) be an open subset of \( \mathbb{R}^d \).

A.1 \( W_k \) is an IID sequence.
A.2 \( h(Z) \) is a \( C^1 \) function.
A.3 For any compact set \( Q \) there exists a constant \( C(Q) \) such that,

\[ E_W |H(Z, W)|^4 \leq C(Q), \]

for all \( Z \in Q \).
A.4 \( d_2 \in (-1, 1] \).

B.1 There exists a family \( \{\Pi_{Z, \theta}\} \) of transition probabilities \( \Pi_{Z, \theta}(G, A) \) such that, for any Borel subset \( A \) we have

\[ P[G_{n+1} \in A | F_n] = \Pi_{Z_n, \theta_n}(G_n, A) \]

where \( F_k = \sigma(\theta_0, Z_0, Z_1, W_1, W_2, ..., W_k, G_0, G_1, ..., G_k) \). This in turn implies that the tuple \( (G_k, \theta_k, Z_k, Z_{k-1}) \) will form a Markov chain.

B.2 For any compact set \( Q \) of \( D \), there exist constants \( C_1, q_1 \) such that for all \( (Z, \theta) \in D \) we have

\[ |H_1(Z, \theta, G)| \leq C_1(1 + |G|^{q_1}). \]

B.3 There exists a function \( h_1 \) on \( D \), and for each \( Z, \theta \in D \) a function \( \nu_{Z, \theta}(\cdot, \cdot) \) such that

a) \( h_1 \) is locally Lipschitz on \( D \).

b) \((I - \Pi_{Z, \theta})\nu_{Z, \theta}(G) = H_1(Z, \theta, G) - h(Z, \theta)\).

c) For all compact subsets \( Q \) of \( D \), there exist constants \( C_3, C_4, q_3, q_4 \) and \( \lambda \in [0.5, 1] \), such that for all \( Z, \theta, Z', \theta' \in Q \)

i) \[ |\nu_{Z, \theta}(G)| \leq C_3(1 + |G|^{q_3}), \]

ii) \[ |\Pi_{Z, \theta'}\nu_{Z', \theta'}(G) - \Pi_{Z, \theta'}\nu_{Z', \theta'}(G)| \leq C_4(1 + |G|^{q_4}) |(Z, \theta) - (Z', \theta')|^\lambda. \]

B.4 For any compact set \( Q \) in \( D \) and for any \( q > 0 \), there exists a \( \mu_q(Q) < \infty \), such that for all \( n, G, A = (Z, \theta) \in \mathbb{R}^d \)

\[ E_{G, A} \{ I(Z_k, \theta_k \in Q, k \leq n) (1 + |G_{n+1}|^q) \} \leq \mu_q(Q)(1 + |G|^q), \]

where \( E_{G, A} \) represents the expectation taken with \( G_0, Z_0, \theta_0 = G, Z, \theta \).

Let \( Z(t, t_0, Z), \theta(t, t_0, \theta) \) represent the solutions of the ODEs (7), (8) with initial conditions \( Z(t_0) = Z, \theta(t_0) = \theta \).

Let \( Q_1 \) and \( Q_2 \) be any two compact sets of \( D \), such that \( Q_1 \subset Q_2 \) and we can choose \( T > 0 \) such that there exists \( \delta_0 > 0 \) satisfying

\[ d((Z(t, 0, Z), \theta(t, 0, \theta)), Q_2^c) \geq \delta_0, \]

for all \( (\theta, Z) \in Q_1 \) and for all \( t \leq T \). With \( V^\tau(Z, \theta) \), representing the ordered pair \( Z, \theta \), we reproduce the statement of Theorem 3, [7] (same theorem in the corresponding technical report).

**Theorem 2:** With Assumptions A.1–A.4 and B.1–B.4, for any compact sets \( Q_1 \subset Q_2 \), for all \( T > 0 \) satisfying condition (9), for all \( \delta \) and for any initial conditions \((G, \theta, Z)\), with \( d_2Z_{-1} + d_1Z_0 = Z \) and \( \theta_0 = \theta \),

\[ P_{G,Z,\theta} \left\{ \sup_{1 \leq k \leq \lfloor T \sigma \rfloor} |V_k - V(k\mu, 0, (Z, \theta))| \geq \delta \right\} \to 0 \]

uniformly for all \((Z, \theta) \in Q_1\) as \( \mu \to 0 \).

APPENDIX C

**Lemma 1:** The ODE (4) has a unique solution which satisfies,

\[ |\theta(t)| \leq c_0 + c_1 e^{-\sigma^2 t}, \]

for appropriate positive constants \( c_0 \) and \( c_1 \).

**Proof:** For convenience, we reproduce the ODE (4),

\[ \dot{\theta}(t) = -R_{au}(Z(t))\theta(t) + R_{as}(\theta(t), Z(t)). \]

The matrix \( R_{au}(Z(t)) \) is positive definite for all \( t \), and it’s minimum eigen value is greater than \( \sigma^2 \) for all \( t \). Also, \( |R_{as}(\theta(t), Z(t))| \leq C |Z(t)| \) for all \( t \) for some constant \( C > 0 \). Using (5), \( |R_{as}(\theta(t), Z(t))| \leq C(T) \) for all \( t \leq T \) for any finite time \( T \) for some positive constant \( C(T) \) depending only on \( T \). Thus, for any vector \( \theta \), the inner product,

\[ \left\langle \dot{\theta}(t), \theta \right\rangle \leq -\sigma^2 |\theta|^2 + C(T) |\theta| \]

\[ = \left[ -\sigma^2 |\theta|^2 + C(T) \right] |\theta|. \]

Therefore by Global existence theorem (pp 169 - 170 of [12]), the ODE (4), has a unique solution for any finite time and the solution is bounded by the solution of the following scalar ODE (after choosing the initial conditions properly),

\[ \dot{k}(t) = -\sigma^2 k(t) + C(T), \]

whose solution is given by, \( k(t) = c_1 e^{-\sigma^2 t} + C(T) \), for some appropriate constant \( c_1 \).

**Lemma 2:** The function \( R_{as}(\theta, Z) \) is continuously differentiable in \((\theta, Z)\), \( \sigma^2 \) and hence is locally Lipschitz.
Proof: With $f_{\mathcal{N}}(\sigma^2, N)$ representing the $M$ dimensional Gaussian density with variance $\sigma^2$,

$$R_{\text{us}}(\theta, Z) = E[(\pi Z S + N)Q(\theta^t(\pi Z S + N))]$$

$$= \sum_S \int_{\{N: \theta^t(\pi Z S + N) > 0\}} (\pi Z S + N) f_{\mathcal{N}}(\sigma^2, N) dN$$

$$- \int_{\{N: \theta^t(\pi Z S + N) < 0\}} (\pi Z S + N) f_{\mathcal{N}}(\sigma^2, N) dN.$$ 

We make the following change of variable,

$$Y = A(\theta)(\pi Z S + N)$$

where matrix

$$A(\theta) \triangleq \begin{bmatrix}
\theta_1 & \theta_2 & \ldots & \theta_M \\
0 & 1 & \ldots & 0 \\
& & \vdots & \\
0 & 0 & \ldots & 1
\end{bmatrix}.$$ 

With $|B|$ representing the determinant of the matrix $B$,

$$R_{\text{us}}(\theta, Z) = \sum_S \int_{\{Y: Y_1 > 0\}} A(\theta)^{-1} Y |A(\theta)^{-1}| f_{\mathcal{N}}(\sigma^2, A(\theta)^{-1} Y - \pi Z S) dY$$

$$- \int_{\{Y: Y_1 < 0\}} A(\theta)^{-1} Y |A(\theta)^{-1}| f_{\mathcal{N}}(\sigma^2, A(\theta)^{-1} Y - \pi Z S) dY,$$

which is continuously differentiable by dominated convergence theorem and because the terms inside the integral are $C^\infty$.

Lemma 3: Let $P_{\theta,Z}(\cdot, \cdot)$ represent the transition function of the Markov chain $G_k(\theta, Z)$ (when the channel and equalizer are fixed at $(\theta, Z)$). The function $P_{\theta,Z} H_0(G)$ is locally Lipschitz.

Proof: Note that,

$$P_{\theta,Z} H_0(G_0) = E[H_1(\theta, G_1)|G_0 = (U_0, S_0)]$$

$$= E[H_1(\theta, (A(Z)G_0 + B(Z)W_1))].$$

For all $(\theta, Z)$ in a compact set, one can get positive constant $C_1$ only depending upon the compact set $Q$ such that,

$$|P_{\theta,Z} H_0(G_0) - P_{\theta,Z} H_0(G_0')|$$

$$\leq E \left| \theta^t U_1 - \theta^t U_1' \right| + 2E |U_1 - U_1'| + C_1 E \left| Q(\theta^t U_1) - Q(\theta^t U_1') \right|,$$

where $U_1 \triangleq A(Z)G_0 + B(Z)W_1$, $U_1' \triangleq A(Z')G_0' + B(Z')W_1$. Suffices to show Lipschitz continuity for the last term. Now, $E \left| Q(\theta^t U_1) - Q(\theta^t U_1') \right|$

$$= 2P(Q(\theta^t U_1) \neq Q(\theta^t U_1'))$$

$$\leq 2P(\|\theta^t U_1\| \neq \|\theta^t U_1' - \theta^t U_1\|)$$

$$\leq 2P(\|c_1 + \theta^t n_1\| \neq c_2 \|G_0, Z, \theta - (G_0', Z', \theta')\| (1 + |n_1|))$$

$$\leq C \|G_0, Z, \theta - (G_0', Z', \theta')\|.$$ 

The last inequality follows by Lemma 12, p.275, [1]. All the constants are appropriately defined. The random variable $n_1$ is the first co-ordinate of the Gaussian noise vector $N_1$. 

References:

[6] V. Kavitha and V. Sharma, " Comparison of training, blind and semiblind equalizers in MIMO fading systems using capacity as measure", ICASSP 05, USA.