Forty-Fourth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
Sept 27-29, 2006

WID.170

LMS versus Wiener Filter for a Decision Feedback Equalizer

Veeraruna Kavitha and Vinod Sharma

Abstract— **In this paper we study an LMS-DFE. We use the ODE framework to show that the LMS-DFE attractors are close to the true DFE Wiener filter (designed considering the decision errors) at high SNR. Therefore, via LMS one can obtain a computationally efficient way to obtain the true DFE Wiener filter under high SNR. We also provide examples to show that the DFE filter so obtained can significantly outperform the usual DFE Wiener filter (designed assuming perfect decisions) at all practical SNRs. In fact, the performance improvement is very significant even at high SNRs (up to 50%), where the popular Wiener filter designed with perfect decisions, is believed to be closer to the optimal one.**

Keywords : ODE approximation, LMS-DFE, Convergence analysis.

I. INTRODUCTION

In any communication system a channel equalizer is an important component. Equalizers are usually FIR filters designed either as Linear Equalizers (LE) or as Decision Feedback Equalizers (DFE) (Figure 1). One often uses the mean square error (MSE) criterion proposed by Widrow ([21]) to obtain an equalizer. A minimum MSE (MMSE) DFE usually outperforms the MMSE LE ([1], [15], [17]) and has interesting optimality properties ([4]). Thus it is preferred over LE.

A MMSE equalizer (also called the Wiener filter) is given by,

$$\underline{\theta}^* = \arg\min_{\underline{\theta}} E\left[\underline{\theta}^t \underline{X} - s\right]^2, \tag{1}$$

where the vector $\underline{X}$ includes channel outputs and decisions (decisions are included only in case of a DFE) while s represents the channel input and $\underline{\theta}^t$ is the transpose of vector $\underline{\theta}$.

In case of an LE it is straightforward to perform the above optimization and the Wiener filter is given by,

$$\underline{\theta}^* = R_{XX}^{-1} R_{Xs}, \tag{2}$$

where $R_{XX} = E\left[\underline{X}\underline{X}^t\right]$ is the auto-correlation matrix of the channel outputs and $R_{Xs} = E\left[\underline{X}s\right]$ is the cross correlation vector.

For a DFE, vector $\underline{X}$ includes previous decisions along with channel outputs and hence its distribution depends upon the parameter $\underline{\theta}$. Thus, it is very difficult to directly compute a Wiener filter. One gets around this problem by assuming perfect decisions (see, e.g., [5], [11], [19]) and design a Wiener filter (which is once again given by (2), the only difference being the vector $\underline{X}$ now includes channel outputs as well as the previous inputs). For convenience, for the rest of the paper, we will call it IDFE (Ideal DFE). The IDFE often outperforms the Linear Wiener filter significantly ([1], [15], [17]). But it is generally believed that the true optimal DFE Wiener filter (considering the decision errors, we will call it simply DFE Wiener filter) can significantly outperform even this. However, so far the performance of such a filter is not known.

One way to solve this problem is to replace the feedback filter at the receiver by a precoder at the transmitter ([4], [15]). This way one can indeed obtain the optimal filter but this requires the knowledge of the channel at the transmitter. However for wireless channels, which are time varying, this is often not an attractive solution ([11] [15]). Some research has been done to deal with the decision errors. Either the distribution of the decisions errors were approximated in designing an MSE optimal filter (IDFE being one such example) or some other appropriate criterion was used to get the optimal filter considering the errors in decisions. For example, in [20], authors approximated the errors in decisions with an additive noise uncorrelated with the input data and AWGN and obtained the DFE Wiener filter. But as is stated in the paper this approximation is not realistic. In [6], the authors obtain an H^∞ optimal DFE considering decision errors. However it is not compared to the DFE Wiener filter.

We address this problem directly via the iterative algorithm, LMS (Least Mean Square). LMS is a very popular, easily implementable and a widely used iterative algorithm ([2], [8], [17]) usually designed to converge to the MMSE solution. It has also been used to implement the DFE (called LMS-DFE); see [11], [15]. However LMS is only used as a convenient tool to implement the DFE (this is also used for tracking purpose but that aspect is not considered here) and the performance of the LMS-DFE has not been compared with the DFE Wiener filter. We show that the LMS-DFE converges to a limit that is close to the DFE Wiener filter at high SNRs. In fact, we show using examples that an LMS-DFE attractor will be close to the DFE Wiener filter at practical SNRs itself. Further, these examples show that unlike what is commonly believed ([19]) the IDFE even under high SNR conditions can have much worse performance.

LMS is given by the following iteration,

$$\underline{\theta}_{k+1} = \underline{\theta}_k - \mu_k \underline{X}_k \left(\underline{X}_k^t \underline{\theta}_k - s_k\right), \tag{3}$$

where $\{\mu_k\}$ is a decreasing sequence. In case of an LE there

Veeraruna Kavitha is a PhD student in ECE Dept, IISc, Bangalore, India kavitha@pal.ece.iisc.ernet.in

Vinod Sharma is a Professor at ECE Dept, IISc, Bangalore, India vinod@ece.iisc.ernet.in

This research is partially supported by DRDO-IISc program on advanced research in Mathematical Engineering.

are many results showing the convergence of LMS to the Wiener filter, $\underline{\theta}^*$ ([2], [12]). Hence it is known that for an LE, the performance of the limiting LMS is not degraded much with respect to the Wiener filter as the length of the training sequence increases. For a DFE, such results are not available.

Performance analysis of any DFE is complicated because of the feedback loop. Existence of a hard decoder inside the feedback loop, makes the study all the more difficult (DFE mainly exploits finite alphabet structure of the hard decoder output ([6], [11]) and hence one cannot get away with the hard decoder). One can study attractors of an adaptive algorithm by first approximating its trajectory with the solution of an ODE and then studying the ODE's attractors. We follow the same approach for analyzing LMS-DFE. Trajectory of the LMS-DFE algorithm without a hard decoder in the feedback loop has been approximated by an ODE in [10]. Beneveniste et al. ([2]) have shown the ODE approximation of the LMS-DFE with a hard decoder inside the feedback loop. However the ODE obtained by them does not give clear insights into the behavior of the LMS-DFE algorithm. Furthermore, they do not relate the attractors of this ODE to the DFE Wiener filter. We show the existence of a unique stationary distribution for the vector $\underline{X}$ for every $\underline{\theta}$ and using this stationary distribution we equate the ODE obtained in [2] with a more tractable ODE,

$$\dot{\underline{\theta}}\,(t) = -E_{\underline{\theta}}\left[\nabla_{\underline{\theta}}\left[\underline{\theta}^t\underline{X} - s\right]^2\right]. \tag{4}$$

The attractors of the LMS-DFE will be the zeros of the RHS of the above ODE while the DFE Wiener filter will be a zero of the gradient (if it exists) of the RHS of equation (1). Under certain conditions these two terms can be related by,

$$\begin{aligned}\nabla_{\underline{\theta}}E_{\underline{\theta}}\left[\left[\underline{\theta}^t\underline{X} - s\right]^2\right] \;&=\; E_{\underline{\theta}}\left[\nabla_{\underline{\theta}}\left[\underline{\theta}^t\underline{X} - s\right]^2\right] + \\ &\quad E\left[\left[\underline{\theta}^t\underline{X} - s\right]^2\nabla_{\underline{\theta}}\pi_{\underline{\theta}}\right], \quad (5)\end{aligned}$$

where $\pi_{\underline{\theta}}$ is the stationary density of $\underline{X}$ when the DFE $\underline{\theta}$ is used. One can expect the LMS-DFE attractors to be close to the DFE Wiener filter if the second term in the RHS of (5) is close to zero. We show in this paper that it is true at high SNRs for a decoder that is slightly perturbed from the original one. The perturbation of the decoder is required because with the original decoder $\pi_{\underline{\theta}}$ may not be differentiable. We show that the DFE Wiener filter and an LMS-DFE attractor of this perturbed decoder converge to that of the original hard decoder as the level of perturbation tends to zero. Then we analyze this perturbed decoder and show that the LMS-DFE attractors of this decoder are close to the DFE Wiener filters of this decoder at high SNR.

The paper is organized as follows. In Section II we define the model and specify our assumptions. In Section III we show that the trajectory of the LMS-DFE can be approximated by the solution of an Ordinary Differential Equation (ODE). We use this ODE for further analysis of the LMS algorithm. We study the differentiability of the stationary distribution of the system in Section IV. In Section V we show that the LMS attractors are close to that of the DFE Wiener filters at high SNRs. Section VI provides some examples while Section VII concludes the paper. Proofs are contained in the appendices.

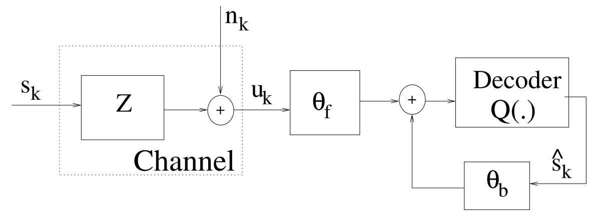

Fig. 1. Block Diagram of Decision Feedback Equalizer (DFE)

II. The model and notations

We consider the system shown in Figure 1. Inputs $\{s_k\}$ enter a time invariant finite impulse response channel $\{z_l\}_{l=0}^{L-1}$ and are corrupted by additive white Gaussian noise $\{n_k\}$ with variance σ^2. The channel output, u_k, at any time k, is given by,

$$u_k = \sum_{l=0}^{L-1} s_{k-l}z_l + n_k.$$

We use a DFE with forward filter $\underline{\theta}_f$ and feedback filter $\underline{\theta}_b$. The decisions are made by hard decoding. We make the following assumptions (some of these can be easily extended) and notations.

- We assume BPSK modulation, i.e., inputs $s_k \in \{+1, -1\}$.
- Sequences $\{s_k\}$ and $\{n_k\}$ are IID (independent, identically distributed) and independent of each other. The inputs s_k are uniformly distributed.
- $f_{\mathcal{N}}(\underline{y})$ is the N_f dimensional standard IID Gaussian density,

$$f_{\mathcal{N}}(\underline{y}) = \frac{1}{(2\pi)^{N_f/2}}exp^{-\frac{|\underline{y}|^2}{2}}.$$

- The equalizer forward filter is given by $\{\theta_{f_l}\}_{l=0}^{N_f-1}$, while the feedback filter is given by $\{\theta_{b_l}\}_{l=1}^{N_b}$. Also, $N_L \stackrel{\triangle}{=} N_f + L - 1$.
- The decisions are obtained after hard decoding. Hence decision $\hat{s}_k$ is given by,

$$\begin{aligned}\hat{s}_k \;&=\; Q\left(\sum_{l=0}^{N_f-1}\theta_{f\,l}u_{k-l} + \sum_{l=1}^{N_b}\theta_{bl}\hat{s}_{k-l}\right), \text{ where} \\ Q(x) \;&:=\; \begin{cases} +1 & if \quad x \geq 0, \\ -1 & if \quad x < 0. \end{cases}\end{aligned} \tag{6}$$

- The following vector notations are used throughout.

$$\begin{aligned}
\underline{S}_k &\triangleq [\ s_k \quad s_{k-1} \quad . \quad . \quad s_{k-N_L+1}\ ]^T, \\
\underline{N}_k &\triangleq [\ n_k \quad n_{k-1} \quad . \quad . \quad n_{k-N_f+1}\ ]^T, \\
\underline{U}_k &\triangleq [\ u_k \quad u_{k-1} \quad . \quad . \quad u_{k-N_f+1}\ ]^T, \\
\underline{\hat{S}}_k &\triangleq [\ \hat{s}_k \quad \hat{s}_{k-1} \quad . \quad . \quad \hat{s}_{k-N_b+1}\ ]^T, \\
\underline{X}_k &\triangleq [\ \underline{U}_k^T \quad \underline{\hat{S}}_{k-1}^T\ ]^T, \\
\underline{G}_k &\triangleq [\ \underline{S}_k^T \quad \underline{\hat{S}}_{k-1}^T \quad \underline{N}_k^T\ ]^T, \\
\underline{\theta}_f &\triangleq [\ \theta_{f0} \quad \theta_{f1} \quad . \quad . \quad \theta_{f\,N_f-1}\ ]^T, \\
\underline{\theta}_b &\triangleq [\ \theta_{b1} \quad \theta_{b2} \quad . \quad . \quad \theta_{bN_b}\ ]^T, \\
\underline{\theta} &\triangleq [\ \theta_f{}^T \quad \theta_b{}^T\ ]^T.
\end{aligned}$$

- With $\mathcal{S} := \{+1, -1\}$, for a fixed $\underline{\theta}$, G_k is a Markov chain taking values in $\mathcal{S}^{N_L} \times \mathcal{S}^{N_b} \times \mathcal{R}^{N_f}$, where $\mathcal{R}$ is the set of real numbers. We represent throughout this paper the current and previous state values of this Markov chain by the ordered pairs $(i, \underline{y})$, $(j, \underline{y}')$ respectively. Here i, j are taking values from the discrete part of the state space, $\mathcal{S}^{N_L} \times \mathcal{S}^{N_b}$, while $\underline{y}, \underline{y}'$ are taking values in $\mathcal{R}^{N_f}$.
- For any vector, $\underline{x}$, we use x_l to represent its l^{th} component. $\underline{x}_l^k$, $l \leq k$, represents the vector $\left[\ x_k \quad x_{k-1} \quad \cdots \quad x_l\ \right]^T$.
- $Z\underline{\theta} = \{z\theta_l\}_{l=0}^{N_L-1}$ represents the convolution of the channel $\{z_l\}$and forward filter $\underline{\theta}_f$.
- The input to the hard decoder for a given state of the Markov chain is represented by,

$$e_{\underline{\theta}}(i, \underline{y}) := \sum_{l=0}^{N_L-1} z\theta_l s_{k-l} + \sum_{l=0}^{N_f} \theta_{f\,l} n_{k-l} + \sum_{l=1}^{N_b} \theta_{bl} \hat{s}_{k-l}.$$

  Note that $\hat{s}_{k-1} = Q(e_{\underline{\theta}}(j, \underline{y}'))$.
- The equalizer output without noise, $e_{\underline{\theta}}(i, 0) \neq 0$ for all values of i at the LMS attractor. Without this assumption the LMS algorithm makes more errors than the correct decisions.
- $B(\underline{\theta}, \delta)$, $\bar{B}(\underline{\theta}, \delta)$ are the open and closed balls respectively with center $\underline{\theta}$ and radius δ.

One can easily extend this theory to the case when $\{s_k\}$ belongs to any finite alphabet with arbitary distribution. Currently the theory to follow, considers an optimal equalizer for delay 0. However, the entire theory will go through for any arbitrary delay. Indeed in Section VI, an example with an optimal equalizer for delay 1, is presented.

III. LMS-DFE

As mentioned earlier, the DFE Wiener filter, given by (1), is difficult to compute. The difficulty is in computation of the distribution of the decisions. It is usual practice to assume perfect decisions (i.e. $\underline{\hat{S}}_k = \underline{S}_{k-N_b+1}^k$) and design a Wiener filter, which we called as IDFE. The IDFE given by (2) is reproduced here for convenience,

$$\underline{\theta}_{IDFE} = \left(E\left[\underline{X}_k \underline{X}_k{}^t\right]\right)^{-1} E\left[\underline{X}_k s_k\right].$$

This computation may be expensive because of matrix inversion and one uses the famous iterative algorithm LMS (3) as an alternative. Our claim is that in case of a DFE, apart from being computationally efficient the LMS algorithm also outperforms the popularly used Wiener filter, IDFE, $\underline{\theta}_{IDFE}$. We achieve this goal by showing that the LMS-DFE attractors are close to that of the DFE Wiener filter (1) at least for high SNRs (later in Section VI we show via examples that this covers the practically used SNRs).

We first show the following useful result, which will also be used later.

Theorem 1: For every $\underline{\theta}$, Markov chain $\{G_k\}$ has a unique stationary probability distribution $\Pi_{\underline{\theta}}$ with density $\pi_{\underline{\theta}}$ (with respect to the measure $f_{\mathcal{N}}(\underline{y})d\underline{y}$). This stationary distribution (density) is continuous with respect to $\underline{\theta}$ in total variation norm (L_1 norm). Also the MSE, the cost in the RHS of (1), under stationarity is continuous in $\underline{\theta}$.

Proof : Please refer to Appendix A.

Because of the continuity of MSE as a function of $\underline{\theta}$, by confining our search to a compact region, we obtain the existence of the Wiener filter. Next we consider the LMS attractors. For this we first approximate the trajectory of the LMS-DFE with an ODE and then study the ODE's attractors.

DFE with a hard decoder has been approximated by an ODE in [2]. We start our LMS-DFE analysis with this ODE. Towards this goal, as a first step the LMS-DFE algorithm (3) is rewritten to fit in the setup of [2], p. 276,

$$\begin{aligned}
\underline{\xi}_k &:= \left[\ \underline{S}_k^t \quad \underline{X}_k^t\ \right]^t, \\
H(\underline{\theta}, \xi) &:= \underline{X}^t\left(\underline{\theta}^t \underline{X} - s\right), \\
\underline{\theta}_k &= \underline{\theta}_{k-1} - \mu_{k-1} H(\underline{\theta}_{k-1}, \xi_k).
\end{aligned}$$

Let $\underline{\theta}(t, t_0, a)$ denote the solution of the following ODE with initial condition $\underline{\theta}(t_0) = a$,

$$\begin{aligned}
\dot{\underline{\theta}}\,(t) &= -h(\underline{\theta}) \text{ where} \\
h(\underline{\theta}) &= \lim_{n\to\infty} P_{\underline{\theta}}^n H_{\underline{\theta}}(j, \underline{y}'),
\end{aligned} \tag{7}$$

where $P_{\underline{\theta}}^n$ is the n-step transition function of the Markov chain G_k with DFE $\underline{\theta}$, and $P_{\underline{\theta}}^n H_{\underline{\theta}}(j, \underline{y}')$ is the expectation of the function $H(\underline{\theta}, \xi)$ using the conditional measure $P_{\underline{\theta}}^n(.|j, \underline{y}')$ (note that ξ_k is a fixed function of G_k). The above limit will be independent of the initial condition $(j, \underline{y}')$ ([2], p. 252).

It is easy to see that the LMS algorithm satisfies all the required hypothesis of Theorem 13, p. 278, [2] and hence one can approximate its trajectory on any finite time scale with the solution of the ODE (7). We reproduce the precise result below.

For any initial condition $\underline{\theta}_0$ and for any finite time T, with $t(r) := \sum_{k=0}^r \mu_k$, $m(n, T) := max_{r \geq n}\{t(r) - t(n) \leq T\}$

$$\sup_{\{n \leq r \leq m(n,T)\}} |\underline{\theta}_r - \underline{\theta}\,(t(r),\ t(n),\ \underline{\theta}_0)| \xrightarrow{p} 0$$

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 if $\liminf_k \frac{\mu_{k+r}}{k} > 0$ for every integer r.

We will show below that the RHS of the ODE (7) is same as that of the ODE (4) and hence equate the ODE (7) with more tractable ODE (4).

For each $\underline{\theta}$, G_k is a Markov chain taking values in a general state space. Its transition function is given by (the definitions of $\tilde{\delta}(.,.), \bar{\delta}(.,.)$ and details are given in [9]),

$$P_{\underline{\theta}}(i, \underline{y} \in B | j, \underline{y}') = \tilde{\delta}(i,j)\bar{\delta}(\underline{y}, \underline{y}')P(i_1)P(\underline{y}_1 \in B_{\underline{y}'}) P_{\underline{\theta}}(i_{N_L}|j, \underline{y}'), \quad (8)$$

The only component of the transition function (8) that depends upon $\underline{\theta}$ is, $P_{\underline{\theta}}(i_{N_L}|j, \underline{y}')$ given by,

$$P_{\underline{\theta}}(i_{N_L} = 1 | j, \underline{y}') = 1_{\{e_{\underline{\theta}}(j,\underline{y}')>0\}}.$$

By IID nature of the input s_k and noise n_k one can choose n_0 large enough such that the n step transition function $P^n_{\underline{\theta}}(i, \underline{y} \in B|j, \underline{y}')$ is absolutely continuous with respect to $f_{\mathcal{N}}(\underline{y})d\underline{y}$ for all $n \geq n_0$. Further, n_0 is chosen bigger than N_L to ensure $\underline{S}_k$, $\underline{S}_{k-n_0}$ are independent. Fix one such n. The corresponding density (Raydon-Nikodyn derivative)

$$p^n_{\underline{\theta}}\left(i, \underline{y}|j, \underline{y}'\right) = \sum_l \int_{\underline{v}} P(\underline{S}^{k+n}_{k+1})\Pi^n_{q=1}P_{\underline{\theta}}(\hat{s}_{k+n-q}|\underline{x}(q))f_{\mathcal{N}}(\underline{v}) \ d\underline{v}, \quad (9)$$

where (the following notations are used throughout),

$$\begin{aligned} l &:= (\underline{S}^{k+n-N_L}_{k+1}, \underline{\hat{S}}^{k+n-1-N_b}_{k}), \\ \underline{v} &:= \underline{N}^{k+n-N_f}_{k+1}, \\ \underline{x}(q) &:= \left(\underline{S}^{k+n-q}_{k+n-q-N_L+1}, \underline{\hat{S}}^{k+n-q-1}_{k+n-q-N_b}, \sigma\underline{N}^{k+n-q}_{k+n-q-N_f+1}\right). \end{aligned}$$

It is easy to see that the density of the $n-$step transition function $p^n_{\underline{\theta}}(i, \underline{y}|j, \underline{y}') \leq 1$, for all values of $i, \underline{y}$, $j, \underline{y}'$. Using Theorem 1, boundedness of the function H and Lemma 1 in Appendix E of [9], we get (for any initial condition $(j, \underline{y}')$),

$$\begin{aligned} h(\underline{\theta}) &= \lim_{n\to\infty} P^n_{\underline{\theta}} H_{\underline{\theta}}(j, \underline{y}') = \Pi_{\underline{\theta}} H(\underline{\theta}, \xi), \\ &= E_{\underline{\theta}}\left[\nabla_{\underline{\theta}}\left[\underline{\theta}^t\underline{X} - s\right]^2\right]. \end{aligned}$$

Thus the trajectory of the LMS-DFE can be approximated in any finite time by the solution of the ODE (4), reproduced here for convenience,

$$\dot{\underline{\theta}}(t) = -E_{\underline{\theta}}\left[\nabla_{\underline{\theta}}\left[\underline{\theta}^t\underline{X} - s\right]^2\right].$$

One can expect that the attractors of the above ODE will be LMS-DFE attractors. The ODE attractors will be zeros of the RHS of the above equation. While the DFE Wiener filters will be the zeros of the gradient (if it exists) of the MSE (the cost in the RHS of (1)). Under certain conditions (which will be discussed in Section V) we get,

$$\nabla_{\underline{\theta}} E_{\underline{\theta}}\left[\left[\underline{\theta}^t\underline{X} - s\right]^2\right] = E_{\underline{\theta}}\left[\nabla_{\underline{\theta}}\left[\underline{\theta}^t\underline{X} - s\right]^2\right] + \sum_{\underline{S}, \underline{\hat{S}}} E_{f_{\mathcal{N}}}\left[\left[\underline{\theta}^t\underline{X} - s\right]^2 \nabla_{\underline{\theta}}\pi_{\underline{\theta}}\right],$$

where $\nabla_{\underline{\theta}}\pi_{\underline{\theta}}$ represents the gradient of the stationary density. One can easily see that the LMS-DFE attractors will exactly be (may be close to) the DFE Wiener filters if in addition the gradient of the stationary density equals zero (is close to zero). This shows that to get the connection between LMS-DFE attractors and the DFE Wiener filter one needs to look at the differentiability of the stationary density.

IV. Differentiability of the Stationary density with respect to $\underline{\theta}$.

One can see from equation (9) that it is very difficult to comment on differentiability of the n-step transition density itself. It will be all the more difficult to talk about differentiability of the stationary density. To proceed further with the analysis, we perturb the hard decoder Q such that the n-step transition density and the stationary density become differentiable. Next we show that the LMS attractors and the DFE Wiener filter of this perturbed decoder converge to that of the original decoder as the level of perturbation tends to zero. Finally we study the DFE using these perturbed decoders in Section V.

We alter the decoder function $Q(x)$ to,

$$Q_{\epsilon_0}(x) = \begin{cases} 1 \quad with \quad prob \quad 1 & x > \epsilon_0 \\ -1 \quad with \quad prob \quad 1 & x < -\epsilon_0 \\ 1 \quad with \quad prob \quad \frac{1}{2}\left[cos\left(\frac{(x-\epsilon_0)\pi}{2\epsilon_0}\right) + 1\right] & |x| \leq \epsilon_0, \end{cases}$$

where ϵ_0 is a small constant. Observe that the perturbed decoder is also a hard decoder. With the perturbed decoder $Q_{\epsilon_0}(x)$, the $\underline{\theta}$ dependent component of the transition function, $P^{\epsilon_0}_{\underline{\theta}}(i_{N_L} = 1|j, \underline{y}')$, equals

$$\frac{1_{\{|e_{\underline{\theta}}(j,\underline{y}')|\leq\epsilon_0\}}}{2}\left[cos\left(\frac{(e_{\underline{\theta}}(j, \underline{y}') - \epsilon_0)\pi}{2\epsilon_0}\right) + 1\right] + 1_{\{e_{\underline{\theta}}(j,\underline{y}')\geq\epsilon_0\}}.$$

The partial derivative, $\frac{\partial P^{\epsilon_0}_{\underline{\theta}}(i_{N_L}=1|j,\underline{y}')}{\partial\underline{\theta}}$ exists everywhere and equals,

$$\frac{-1_{\{|e_{\underline{\theta}}(j,\underline{y}')|\leq\epsilon_0\}}\pi}{4\epsilon_0} sin\left(\frac{(e_{\underline{\theta}}(j, \underline{y}') - \epsilon_0)\pi}{2\epsilon_0}\right)\frac{\partial e_{\underline{\theta}}(j, \underline{y}')}{\partial\underline{\theta}}. \quad (10)$$

By the uniform upper bound on the derivative (10) and by the bounded convergence theorem one can see that the n-step transition density (9) (with $n \geq n_0$) becomes differentiable (more details are given in [9]) and equals (using the notations of equation (9)),

$$\frac{\partial p^{\epsilon_0, n}_{\underline{\theta}}}{\partial\underline{\theta}}\left(i, \underline{y}|j, \underline{y}'\right) = \sum_l \int_{\underline{v}} \sum_{m=1}^{n} \Pi^n_{\substack{q=1 \\ q\neq m}} P^{\epsilon_0}_{\underline{\theta}}(\hat{s}_{k+n-q}|\underline{x}(q))\frac{\partial P^{\epsilon_0}_{\underline{\theta}}(\hat{s}_{k+n-m}|\underline{x}(m))}{\partial\underline{\theta}} P(\underline{S}^{k+n}_{k+1})f_{\mathcal{N}}(\underline{v}) \ d\underline{v}. \quad (11)$$

For these perturbed decoders we show that the stationary density also becomes differentiable. Furthermore, we get a bound on the norm of this gradient using an Implicit function theorem.

Theorem 2: For every $\epsilon_0 > 0$, for every $\underline{\theta}_0$, the Markov chain G_k has a unique stationary density (with respect to N_f dimensional Gaussian IID vector), $\pi_{\underline{\theta}}^{\epsilon_0}$. Also, $\pi_{\underline{\theta}}^{\epsilon_0}$ is continuously differentiable in L_2 norm. Furthermore, there exists a $\delta > 0$, $\sigma_0^2 > 0$ such that for all $\underline{\theta} \in B(\underline{\theta}_0, \delta)$, $\sigma^2 \leq \sigma_0^2$,

$$\left|\nabla_{\underline{\theta}} \pi_{\underline{\theta}}^{\epsilon_0}\right|^2 \quad \leq$$
$$C\left(\sum_i P\left(\left|\underline{S}_k^t \underline{Z\theta} + \underline{\theta}_b^t \underline{\hat{S}}_{k-1} + \underline{\theta}_f^t N_k\right| \leq \epsilon_0\right) + \sigma^2\right)$$

for some positive constant $C < \infty$.
Proof : Please refer to Appendix B.

We conclude this section by showing that the DFE Wiener filters and the LMS-DFE attractors of the perturbed decoder converge to that of the original decoder. Let $\underline{\theta}_n^*$ and $\underline{\theta}_n^{LMS}$ denote the DFE Wiener filter and LMS-DFE attractor for perturbation ϵ_{0n}.

Theorem 3: There exists a sequence $\epsilon_{0n} \to 0$, $\underline{\theta}_*$, a DFE Wiener filter of the original decoder, and $\underline{\theta}^{LMS}$ an LMS-DFE attractor of the original decoder, such that,

$$\begin{aligned} \underline{\theta}_n^* &\to \underline{\theta}^*, \\ \underline{\theta}_n^{LMS} &\to \underline{\theta}^{LMS}, \end{aligned}$$

for any fixed σ^2.
Proof : Please refer to Appendix C.

Hence forth, we analyze the perturbed decoder to draw important conclusions.

V. LMS ATTRACTORS VERSUS WIENER FILTER AT HIGH SNRS

In this section we would like to understand the connection between an LMS attractor and a DFE Wiener filter for a perturbed decoder. Since the former is a zero of RHS of equation (4) while the later is the zero of the gradient of the MSE, the cost in RHS of equation (1), we study the connection between these two.

Fix an $\epsilon_0 > 0$. With the error defined by, $err_{\underline{\theta}}(G_k) := \left(s_k - e_{\underline{\theta}}(G_k)\right)$ (note that i defined in notations represents, $(\underline{S}_k, \underline{\hat{S}}_{k-1})$, the discrete part of the state value of the Markov chain, G_k),

$$\begin{aligned} \nabla_{\underline{\theta}} E_{G_k(\underline{\theta})}\left[err_{\underline{\theta}}(G_k)^2\right] &\stackrel{a}{=} \sum_i \nabla_{\underline{\theta}} E_{f_{\mathcal{N}}}\left[err_{\underline{\theta}}(G_k)^2 \pi_{\underline{\theta}}^{\epsilon_0}(G_k)\right] \\ &\stackrel{b}{=} \sum_i E_{f_{\mathcal{N}}} \nabla_{\underline{\theta}}\left[err_{\underline{\theta}}(G_k)^2 \pi_{\underline{\theta}}^{\epsilon_0}(G_k)\right] \\ &= \sum_i E_{f_{\mathcal{N}}}\left[\nabla_{\underline{\theta}}\left(err_{\underline{\theta}}(G_k)^2\right) \pi_{\underline{\theta}}^{\epsilon_0}(G_k)\right] \\ &\quad + \sum_i E_{f_{\mathcal{N}}}\left[err_{\underline{\theta}}(G_k)^2 \nabla_{\underline{\theta}} \pi_{\underline{\theta}}^{\epsilon_0}(G_k)\right] \\ &= E_{G_k(\underline{\theta})}\left[\nabla_{\underline{\theta}}\left(err_{\underline{\theta}}(G_k)^2\right)\right] \\ &\quad + \sum_i E_{f_{\mathcal{N}}}\left[err_{\underline{\theta}}(G_k)^2 \nabla_{\underline{\theta}} \pi_{\underline{\theta}}^{\epsilon_0}(G_k)\right]. \end{aligned} \quad (12)$$

Here equality a follows by the existence of the stationary density $\pi_{\underline{\theta}}^{\epsilon_0}$ with respect to the Gaussian measure $f_{\mathcal{N}}(\underline{y})d\underline{y}$. Equality b follows by Cauchy Schewartz inequality and bounded convergence theorem, because of the following two consequences (see in [9]). We have shown in the previous section that the derivative $\nabla_{\underline{\theta}} \pi_{\underline{\theta}}$, exists in L_2 norm and is bounded uniformly in a neighborhood of $\underline{\theta}$. Secondly $\nabla_{\underline{\theta}} err_{\underline{\theta}}(G_k)$ exists and is bounded uniformly in a neighborhood of $\underline{\theta}$ by an integrable function of only the sample value of G_k. The above equality (12) is true for any $\epsilon_0 > 0$ and for any σ^2.

We will show that the DFE Wiener filter will be close to the limiting LMS-DFE if the second term on the right hand side of (12) is small.

We have assumed that $\underline{S}_k^t \underline{Z\theta} + \underline{\theta}_b^t \underline{\hat{S}}_{k-1} \neq 0$ at an LMS attractor. By continuity, we choose an ϵ_1 small enough such that

$$0 \notin [\underline{S}_k^t \underline{Z\theta} + \underline{\theta}_b^t \underline{\hat{S}}_{k-1} - \epsilon_1, \underline{S}_k^t \underline{Z\theta} + \underline{\theta}_b^t \underline{\hat{S}}_{k-1} + \epsilon_1],$$

for all $(\underline{S}_k, \underline{\hat{S}}_{k-1})$ and for all $\underline{\theta}$ in a small neighborhood of an LMS attractor. Choose $\epsilon_0 \leq \epsilon_1$. By Chebyshev's inequality, if $0 \notin [c - \epsilon_0, c + \epsilon_0]$ and if n is a Gaussian random variable with mean zero and variance σ^2,

$$P(|c + n| \leq \epsilon_0) \leq P(|n| \geq \min\{|c - \epsilon_0|, |c + \epsilon_0|\}) \to 0$$

as $\sigma^2 \to 0$. Thus, from Theorem 2, for any fixed $\epsilon_0 \leq \epsilon_1$,

$$\left|\nabla_{\underline{\theta}} \pi_{\underline{\theta}}\right| \to 0 \text{ as } \sigma^2 \to 0.$$

Thus by Cauchy Schewartz inequality as $\sigma^2 \to 0$,

$$\begin{aligned} &\left|\nabla_{\underline{\theta}} E_{\underline{\theta}}\left[err_{\underline{\theta}}(G_k)^2\right] - E_{\underline{\theta}}\left[\nabla_{\underline{\theta}}\left(err_{\underline{\theta}}(G_k)^2\right)\right]\right| \\ &\qquad = \left|\sum_i E_{f_{\mathcal{N}}}\left[err_{\underline{\theta}}(G_k)^2 \nabla_{\underline{\theta}} \pi_{\underline{\theta}}(G_k)\right]\right| \quad \to 0. \end{aligned}$$

Now we show that this implies that the LMS-DFE attractors will be close to the DFE Wiener filters. In general two functions f_1, f_2 can be close to each other at each point, but their zeros may be far apart, i.e., if x_1 is a zero of f_1 then $f_2(x_1)$ will be close to zero but the zero of f_2 closest to x_1 may still be far away. It is useful to rule out this possibility in our scenario. We show this using the following theorem. Define,

$$\begin{aligned} f(\underline{\theta}, \sigma^2) &:= E_{G_k(\underline{\theta})}\left[\nabla_{\underline{\theta}}\left(err_{\underline{\theta}}(G_k)^2\right)\right] \text{ and} \\ g(\underline{\theta}, \sigma^2, \eta) &:= f(\underline{\theta}, \sigma^2) + \eta. \end{aligned}$$

Theorem 4: There exists an ϵ_2 with $0 < \epsilon_2 \leq \epsilon_1$ such that for any $\epsilon_0 \leq \epsilon_2$, there exists a continuous function $q : B(0, \delta) \subset \mathcal{R} \times \mathcal{R} \mapsto \mathcal{R}^{N_f}$, such that,

$$g\left(q\left(\sigma^2, \eta\right), \sigma^2, \eta\right) = 0.$$

Proof : The proof is given in the technical report [9].

For any fixed $\epsilon_0 \leq \epsilon_2$, the gradient of the stationary density, near an LMS attractor, is tending to zero as $\sigma^2 \to 0$. Thus there exists a small enough σ_0 such that for all $\sigma^2 \leq \sigma_0^2$,

$$\left|\left(\sigma^2, \ \nabla_{\underline{\theta}} E_{\underline{\theta}} err_{\underline{\theta}}(G_k)^2 - E_{\underline{\theta}} \nabla_{\underline{\theta}} err_{\underline{\theta}}(G_k)^2\right)\right| \quad \leq \quad \delta.$$

For these $\sigma^2 \leq \sigma_0^2$, the LMS attractors, $q(\sigma^2, 0)$, will be close to that of the Wiener filters, $q\left(\sigma^2,\ \left(\nabla_{\underline{\theta}} E_{\underline{\theta}} err_{\underline{\theta}}(G_k)^2 - E_{\underline{\theta}} \nabla_{\underline{\theta}} err_{\underline{\theta}}(G_k)^2\right)\right)$.

It is clear from the above Theorem that at high SNRs for very small ϵ_0 (close to the practical decoder), the LMS attractor is close to the DFE Wiener filter. Since, IDFE $\underline{\theta}_{IDFE}$, is designed with an improper assumption (like perfect decisions), there is a good chance of these filters to be inefficient in comparison to the LMS attractors. We will see this in the examples of the next section.

VI. Examples

In this section we reinforce the theory developed so far using some examples. We take a few examples of channels obtained from previous studies and show the proximity of the DFE Wiener filter and the LMS attractor for practical values of SNRs. We also show that in many cases, the IDFE performs much worse than the DFE Wiener filter and the LMS attractor. Probability of Error (P_e) is used to compare the various equalizers. We used Monte-Carlo simulations to estimate P_e using one million samples of data.

DFE Wiener filter, $\underline{\theta}^*$ for these examples was obtained by running a gradient descent type of algorithm on the cost function (1) itself where the gradient was approximated at each point by finite difference approximation via a large number of samples. Details are provided in [9].

In Table 1, we have used an interesting example (significant part of the raised cosine channel of [7], p. 199) to show that the LMS attractors will be very close to the Wiener filters at practical SNRs. Its coefficients are provided in the table. We also provide P_e in this table. One can see an improvement up to 25% in P_e in LMS in comparison with the IDFE. In fact this improvement is more at high SNRs (where the IDFE is assumed to have lesser problem because of error propagation). We can also see that the P_e of the DFE Wiener filter is very close to that of the LMS attractor.

We have developed the entire theory for an equalizer with delay zero. One can easily extend these results to the equalizer with any arbitrary delay. In fact, the channel in Table 2 is one such example. Here the equalizer with delay 1 will be the best one. The channel of example 2 is very widely used (see p. 414 of [8], p. 165 of [7]) We can see once again a huge improvement (up to 50%) in P_e of LMS with respect to $\underline{\theta}_{IDFE}$. We can also see that the LMS attractors are very close to the DFE Wiener filter, $\underline{\theta}^*$ for all practical SNRs.

TABLE I

Example 1 : Comparison of Various Equalizers Channel = $[0.45\ \ 0.59\ \ 0.43\ \ 0.11\ \ -0.22\ \ -0.32\ \ -0.27\ \ 0\ \ 0.11\ \ 0.11]$ $N_f = 5\ N_b = 10$

SNR	$\underline{\theta}^*$		$\underline{\theta}_{IDFE}$			$\underline{\theta}_{LMS}$		
	MSE	P_e	Dist from $\underline{\theta}^*$	MSE	P_e	Dist from $\underline{\theta}^*$	MSE	P_e
16.7	.21	.024	1.1	.26	.027	.035	.21	.024
14.5	.38	.089	1.7	.64	.10	.037	.38	.091
12.5	.49	.150	1.6	.94	.184	.027	.49	.151
11.5	.54	.176	1.5	1.0	.215	.023	.54	.177
4.5	.81	.311	.64	.94	.33	.021	.80	.311
1.5	.87	.353	.37	.93	.364	.023	.87	.353

TABLE II

Example 2: Comparison of Various Equalizers $N_f = 2\ N_b = 2$ Channel = [0.41 .82 0.41]

SNR	$\underline{\theta}^*$		$\underline{\theta}_{IDFE}$			$\underline{\theta}_{LMS}$		
	MSE	P_e	Dist from $\underline{\theta}^*$	MSE	P_e	Dist from $\underline{\theta}^*$	MSE	P_e
16.7	.11	.0027	.13	.12	.0035	.014	.11	.0028
14.5	.16	.01	.26	.18	.015	.021	.16	.011
12.5	.23	.03	.35	.26	.037	.025	.23	.032
11.5	.27	.047	.40	.30	.055	.031	.28	.050
4.5	.54	.184	.43	.59	.2	.009	.54	.184
1.5	.65	.235	.32	.69	.25	.008	.65	.235

VII. Conclusions

Obtaining MSE optimal filter for DFE is a long-standing problem. Precoding provides one practical solution but can at best be used for slowly varying channels. Otherwise one commonly uses the optimal Wiener filter obtained assuming perfect past decisions.

In this paper we show via ODE analysis, that LMS itself can provide the optimal Wiener filter. We show it by proving that the attractors of the LMS are close to that of the optimal DFE at high SNRs. Proofs become nontrivial partly because of the non-differentiability of the hard decoder. Next, we show by examples that the SNRs need not be very high, i.e., in fact practically used SNRs can be sufficient. We also show that the BER of the commonly used Wiener filter, designed assuming perfect past decisions, can be up to 50% higher than the optimal Wiener filter even at high SNRs (where the former is believed to be closer to the later).

Of course, one advantage of LMS over Precoding is that it can be used at the receiver and has tracking capability when the channel is time varying. In future we will be studying the tracking performance of LMS-DFE.

Appendix A

Proof of Theorem 1 : We provide rough sketch of the proof here. The detailed proof is in [9].

We prove the existence and continuity of the stationary distribution of the Markov chain, G_k, using the results on Markov chains and Stochastic stability ([14]).

For any $\underline{\theta}_0$ and for any $\epsilon > 0$,

$$M_1 \stackrel{\triangle}{=} \min_{\underline{S}_k, \underline{\hat{S}}_{k-1}, \underline{\theta} \in \bar{B}(\underline{\theta}_0, \epsilon)} \underline{Z\theta}^t \underline{S}_k + \underline{\theta}_b^t \underline{\hat{S}}_{k-1}.$$

Then for the decoder (6),

$$\left\{ \underline{N}_k : \min_{\underline{\theta} \in \bar{B}(\underline{\theta}_0, \epsilon)} \underline{\theta}_f^t \underline{N}_k > -M_1 \right\} \tag{13}$$

is a subset of noise vectors, for which the decoder outputs 1 for any input, for any past decisions and for any $\underline{\theta} \in \bar{B}(\underline{\theta}_0, \epsilon)$. By continuity of the map $(\underline{\theta}, \underline{N}_k) \mapsto \underline{\theta}_f{}^t \underline{N}_k$, if required one can choose a smaller $\delta \leq \epsilon$ and an open set C such that for all $\underline{\theta} \in \bar{B}(\underline{\theta}_0, \delta)$, for all inputs, past decisions and for all noise vectors in C the decoder outputs 1, i.e., for any $\underline{\theta} \in \bar{B}(\underline{\theta}_0, \delta)$,

$$\begin{aligned} P(\underline{\hat{S}}_k = [\ 1 \quad .. \quad 1\ ]) &\geq P\left(\cap_{l=k-N_b-1}^{k}\{\underline{N}_l \in C\}\right), \\ &\geq P(\underline{N}_{k-N_b-N_f+1}^{k} \in C_1 \times C_2), \end{aligned}$$

where sets $C_1 \in \mathcal{R}^{N_b}$, $C_2 \in \mathcal{R}^{N_f}$ are selected such that their respective Lesbague measures are not equal to zero and such that the open set,

$$\cap_{l=k-N_b-1}^{k}\{\underline{N}_l \in C\} \supset C_1 \times C_2.$$

Define $\mathcal{G} := [\ 1 \quad .. \quad 1\ ] \times [\ 1 \quad .. \quad 1\ ] \times \mathcal{R}^{N_f}$. For any $n_0 > max\{N_b + N_f + 1, N_L\}$, for any G_{k-n_0}, for any $E = [\ 1 \quad .. \quad 1\ ] \times [\ 1 \quad .. \quad 1\ ] \times B_E \subset \mathcal{G}$, and for any $\underline{\theta} \in \bar{B}(\underline{\theta}_0, \delta)$,

$$P(G_k \in E | G_{k-n_0}) \quad \geq \quad \alpha P(\underline{N}_k \in B_E \cap C_2)$$

where $\alpha := P(\underline{S}_k = [\ 1 \quad .. \quad 1\ ]) P(\underline{N}_{k-N_b+N_f}^{k-N_f} \in C_1)$. Thus for any $\underline{\theta} \in \bar{B}(\underline{\theta}_0, \delta)$, and for any initial condition G_{k-n_0}, the n_0-step conditional measure is majorized :

$$P_{\underline{\theta}}(G_k \in E | G_{k-n_0}) \quad \geq \quad \nu_{n_0}(E \cap \mathcal{G}),$$

where the measure $\nu_{n_0}()$ is defined by,

$$\begin{aligned} \nu_{n_0}([\ 1 \quad .. \quad 1\ ] \times [\ 1 \quad .. \quad 1\ ] \times B_E) \quad &:= \\ &\alpha P(\underline{N}_k \in B_E \cap C_2). \end{aligned}$$

Thus the entire state space $\mathcal{S}^{N_L} \times \mathcal{S}^{N_b} \times \mathcal{R}^{N_f}$ is $\nu_{n_0}-$small (hence also a petite set) for all the Markov chains, $\{G_k\}$, parameterized by $\underline{\theta} \in \bar{B}(\underline{\theta}_0, \delta)$. Thus using Proposition 9.1.7, p. 206 and Theorem 10.01, p. 230, [14] one obtains the existence and uniqueness of the probability stationary distribution, $\Pi_{\underline{\theta}}$ for each $\underline{\theta}$.

Define $\rho = 1 - \nu_{n_0}(\mathcal{G})$. Further, by Theorem 16.2.4 in page 392 of [14], for all $\underline{\theta} \in \bar{B}(\underline{\theta}_0, \delta)$, we get the following uniform convergence in total variation norm,

$$\left| P_{\underline{\theta}}^{n}(.|j, \underline{y}') - \Pi_{\underline{\theta}} \right| \leq \rho^{\frac{n}{n_0}} \text{ for all initial conditions } (j, \underline{y}').$$

This along with the continuity of transition function, establishes the continuity of the stationary distribution $\Pi_{\underline{\theta}}$ under total variation norm at $\underline{\theta}_0$ (see [9]).

The stationary distribution, $\Pi_{\underline{\theta}}$, has discrete and continuous components. The continuous component of $\Pi_{\underline{\theta}}$, is absolutely continuous with respect to the measure $f_{\mathcal{N}}(\underline{y})d\underline{y}$ for every $\underline{\theta}$. Hence the stationary density, $\pi_{\underline{\theta}}$ for G_k exists. Continuity in total variation norm of the stationary distribution implies the continuity of the stationary densities in L_1 norm (Theorem 8.2, p. 110, [18]). It is also easy to see that the stationary density $\pi_{\underline{\theta}}(i, \underline{y}) \leq 1$ for all $(i, \underline{y})$.

MSE, the cost in RHS of equation (1), can be rewritten as,

$$E_{\underline{\theta}}\left[\underline{\theta}^t \underline{X} - s\right]^2 = \sum_{\underline{S}, \underline{\hat{S}}} E_{f_{\mathcal{N}}}\left[\left(\underline{\theta}^t \underline{X} - s\right)^2 \pi_{\underline{\theta}}\right].$$

Lemma 1 in Appendix E of the technical report [9], now gives the continuity of the MSE with respect to $\underline{\theta}$. $\square$

Appendix B

Proof of Theorem 2 : We provide rough sketch of the proof here. The detailed proof is in [9].

The existence and continuity of the stationary density $\pi_{\underline{\theta}}^{\epsilon_0}$ for every ϵ_0 is achieved in a similar way as in the proof of the Theorem 1. We leave superscript ϵ_0 to simplify the notation in the rest of the proof.

We use an implicit function theorem to prove differentiability. For that, we will need to prove the following results. Consider the Banach spaces :

- $X = \mathcal{R}^{N_f+N_b}$ with Euclidean norm.
- $Y = \{g : \mathcal{S}^{N_L+N_b} \times X \to \mathcal{R}; |g| < \infty\}$ with L_2 norm defined by,

$$|g| := \frac{1}{|\mathcal{S}|}\sum_i \left(\int_{\underline{y}} \left|g(i, \underline{y})\right|^2 f_{\mathcal{N}}(\underline{y}) d\underline{y}\right)^{1/2},$$

where $|\mathcal{S}|$ represents the cardinality of set $\mathcal{S}^{N_L+N_b}$.

Fix $n_0 > max\{N_f + N_b, N_L\}$. We consider the following continuous map $f : X \times Y \mapsto Y$,

$$f(\underline{\theta}, \pi) = g(\underline{\theta}, \pi) - \pi + \left(\sum_j \int_{\underline{y}'} \pi(j, \underline{y}') f_{\mathcal{N}}(\underline{y}') d\underline{y}' - 1\right),$$

where,

$$g(\underline{\theta}, \pi)(i, \underline{y}) \quad := \quad \sum_j \int_{\underline{y}'} p_{\underline{\theta}}^{n_0}(i, \underline{y}|j, \underline{y}') \pi(j, \underline{y}') f_{\mathcal{N}}(\underline{y}') d\underline{y}'.$$

Observe that $(\underline{\theta}, \pi_{\underline{\theta}})$ is a zero of f.

In the following we prove that f is differentiable with respect to π and the derivative is a homeomorphism.

Note that the above function is affine linear in the second variable $\pi \in Y$ and hence,

$$\left.\frac{\partial f}{\partial \pi}\right|_{(\underline{\theta}, \hat{\pi})}(\pi) = g(\underline{\theta}, \pi) - \pi + \left(\sum_j \int_{\underline{y}'} \pi(j, \underline{y}') f_{\mathcal{N}}(\underline{y}') d\underline{y}'\right).$$

We have shown that this map is one-one through contradiction (the details are in [9]). Also this partial derivative equals $I - T$, where T is a compact operator (see [9]). Then by Riesz-Schauder Theory (Theorem 1, p. 283, [22]), the fact that $\frac{\partial f}{\partial \pi}$ is one-one implies that it is onto also and further that the inverse is bounded. Hence it is a linear homeomorphism. Furthermore, the mapping $(\sigma^2, \underline{\theta}) \mapsto \left|\left[\left.\frac{\partial f}{\partial \pi}\right|_{(\underline{\theta}, \pi_{\underline{\theta}})}\right]^{-1}\right|$ is continuous (see [9]). Thus one can get the following uniform upper bound,

$$\left|\left[\left.\frac{\partial f}{\partial \pi}\right|_{(\underline{\theta}, \pi_{\underline{\theta}})}\right]^{-1}\right| \leq C_0 \text{ for all } \underline{\theta} \in B(\underline{\theta}_0, \delta), \sigma^2 \leq \sigma_0^2, \quad (14)$$

for some δ, σ_0^2.

From (11), by bounded convergence theorem, (see [9]),

$$\left.\frac{\partial f}{\partial \theta}\right|_{\theta, \pi}(i, \underline{y}) \quad = \quad \sum_j \int_{\underline{y}'} \frac{\partial p_{\theta}^{n}}{\partial \underline{\theta}}(i, \underline{y}|j, \underline{y}') \pi(j, \underline{y}') f_{\mathcal{N}}(\underline{y}') d\underline{y}'$$

Also (see [9]),

$$\left|\left.\frac{\partial f}{\partial \underline{\theta}}\right|_{(\underline{\theta},\pi_{\underline{\theta}})}\right|^2 \leq C_1 \sum_i P\left(\left|\underline{S}_k^t \underline{Z\theta} + \underline{\theta}_b^t \underline{\hat{S}}_{k-1} + \underline{\theta}_f^t N_k\right| \leq \epsilon_0\right) + C_2\sigma^2, \quad (15)$$

where C_1, C_2 are constants independent of $\underline{\theta}, \sigma^2$.

Using similar logic one can easily show that both the partial derivatives of f are continuous in $(\underline{\theta}, \pi)$. Hence by Implicit function theorem on Banach spaces, (Theorem 3.1.10 and corollary 3.1.11, p. 115, [3]), the map $\underline{\theta} \mapsto \pi_{\underline{\theta}}$ is continuously differentiable and the derivative is given by,

$$\nabla_{\underline{\theta}} \pi_{\underline{\theta}} = -\left[\left.\frac{\partial f}{\partial \pi}\right|_{(\underline{\theta},\pi_{\underline{\theta}})}\right]^{-1} \left.\frac{\partial f}{\partial \underline{\theta}}\right|_{(\underline{\theta},\pi_{\underline{\theta}})}.$$

The theorem follows by upper bounding this gradient using the upper bounds (14), (15). $\square$

APPENDIX C

Proof of Theorem 3 : Let

$$\begin{aligned} f_1(\underline{\theta}, \epsilon_0) &:= E^{\epsilon_0}_{G_k(\underline{\theta})}\left[err_{\underline{\theta}}(G_k)^2\right], \\ f_2(\underline{\theta}, \epsilon_0) &:= \left|E^{\epsilon_0}_{G_k(\underline{\theta})} \nabla_{\underline{\theta}}\left[err_{\underline{\theta}}(G_k)^2\right]\right|. \end{aligned}$$

Note that for any fixed ϵ_0, LMS attractors will be zeros, i.e., minima of $f_2(., \epsilon_0)$ while the DFE Wiener filters are the minima of the MSE cost function $f_1(., \epsilon_0)$. Also note that $\epsilon_0 = 0$ corresponds to the original decoder.

Let $\{\epsilon_{0n}\}$ be any sequence converging to 0. Let $\Omega = \{\epsilon_{0n}\}$. Take a compact set C large enough such that the Wiener filter is inside it (as $\underline{\theta}$ is increased to infinity, eventually MSE will start increasing and will tend to infinity).

One can follow steps similar to Theorem 1 and show that the stationary density $\pi^{\epsilon_{0n}}_{\underline{\theta}_n}$ converges to $\pi^0_{\underline{\theta}}$ as $(\epsilon_{0n}, \underline{\theta}_n) \to (0, \underline{\theta})$. Similarly, one can also show that both functions f_1, f_2 are jointly continuous in $(\underline{\theta}, \epsilon_0) \in C \times \Omega$.

The domain of the parameter $\underline{\theta}$ for every ϵ_0, say $D(\epsilon_0)$, is the same compact set C and hence the correspondence $\epsilon_0 \mapsto D(\epsilon_0)$ is compact and continuous (definitions are in [16]). Then by the maximum theorem (p. 235, [16]),

$$\begin{aligned} D_{1n}^* &:= \arg\min_{\underline{\theta} \in C} f_1(\underline{\theta}, \epsilon_{0n}), \\ D_{2n}^* &:= \arg\min_{\underline{\theta} \in C} f_2(\underline{\theta}, \epsilon_{0n}), \end{aligned}$$

are compact valued upper semi-continuous correspondences on Ω. Then by Proposition 9.8, p. 231, [16] there exists a subsequence of LMS attractors $\underline{\theta}^{LMS}_{n_k}$ converging to an LMS attractor of the original decoder, $\underline{\theta}^{LMS}_0$. Once again by the same proposition there exists a further subsequence such that the DFE Wiener filters $\underline{\theta}^*_{n_{k_l}}$ converge to a DFE Wiener filter of the original decoder, $\underline{\theta}^*_0$. Thus there exists a sequence (after renaming) $\epsilon_{0n} \to 0$ such that

$$\begin{aligned} \underline{\theta}^{LMS}_n &\to \underline{\theta}^{LMS}_0 \text{ and} \\ \underline{\theta}^*_n &\to \underline{\theta}^*_0. \quad \square \end{aligned}$$

REFERENCES

[1] C. A. Belfiore, J. H. Park Jr, "Decision feedback equalization", Proc. IEEE, Vol. 67, No. 8, Aug 1979, 1143 - 1156.
[2] A. Benveniste, M. Mietivier, P. Priouret, "Adaptive Algorithms and Stochastic Approximation", Springer-Verlag, April 1990.
[3] M. S. Berger, "Nonlinearity and Functional Analysis", Academic Press, New York, 1977.
[4] J. M. Cioffi, C. P. Dudevoir, M. V. Eyuboglu, G. D. Forney Jr, "MMSE Decision-Feedback Equalizers and Coding-Part I: General Results", IEEE Trans. Comm., Vol 43, 1995, 2582-2594.
[5] N. A. Dhahir, A. H. Sayed,"The Finite-Length Multi-Input Multi-Output MMSE-DFE", IEEE Trans. Signal Processing, Vol. 48, No. 10, October 2000 pp 2921-2936
[6] A. T. Erdogan, B. Hassibi, T. Kailath, "MIMO Decision Feedback Equalization from an H^∞ Perspective". IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 52, NO. 3, MARCH 2004, pp 734-745.
[7] G. B. Giannakis, Y. Hua, P. Stoica, L. Tong, "Signal Processing Advances in Wireless and Mobile Communications", Trends in Channel estimation and equalization. Vol.1, Prentice Hall, Upper Saddle River, NJ 2000.
[8] S. Haykin, "Adaptive Filter theory", Third Edition, Prentice-Hall International inc, 1996.
[9] V. Kavitha, V. Sharma, "LMS Versus Wiener filter for a Decision feedback equalizer", Technical report no: TR-PME-2006-06, DRDO-IISc program on mathematical engineering, ECE Dept., IISc, Bangalore, June 2006. (downloadable from http://www.pal.ece.iisc.ernet.in/PAM/tech_rep06.html).
[10] H. Kushner, G. George Yin, "Stochastic Approximation Algorithms and Applications", Springer, 1997.
[11] E. A. Lee, D. G. Messerschmitt, "Digital communications", 2nd Edn., Kluwer Academic Publishers, 1994.
[12] O. Macchi, E. Eweda, "Convergence Analysis of Self-Adaptive Equalizers", IEEE Trans on Information Theory, Vol.30, No.2, March 1984, pp 161-176.
[13] D. G. Manolakis, V. K. Ingle, S. M. Kogon, "Statistical and adaptive signal processing: spectral estimation, signal modeling, adaptive filtering" 0 - McGraw-Hill , 2000.
[14] S. P. Meyn, R. L. Tweedie "Markov chains and Stochastic Stability", Communications and Control Engineering Series Springer-Verlag, London, New York, 1993.
[15] J. G. Proakis, "Digital Communications", New York: McGraw-Hill, 2000.
[16] R. K. Sundaram, "A First Course in Optimization Theory", New York: Cambridge Univ. Press, 1996.
[17] D. P. Taylor, G. M. Vitetta, B. D. Hart, A. Mammela, "Wireless channel equalization", Eur. Trans. Telecom., Vol. 9,1998, 117-143.
[18] H. Thorisson, "Coupling, Stationarity, and Regeneration", Probability and its Applications Springer-Verlag, New York, 2000.
[19] C. Tidestav, A. Ahln, M. Sternad,"Realizable MIMO Decision Feedback Equalizers: Structure and Design", IEEE Trans. on Signal Processing, Vol. 49, No. 1, Jan 2001, pp 121 - 133.
[20] M. Sternad, A. Ahln, and E. Lindskog, "Robust decision feedback equalizers", in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process, vol. 3, Minneapolis, MN, Apr. 1993, pp. 555-558.
[21] B. Widrow "Adaptive filters I : Fundamentals", Report SEL-66- 126, TR 6764-6, Stanford Electronics Laboratories, Stanford, USA, 1966.
[22] K. Yoshida, "Functional Analysis" Springer-Verlag, Heidelberg, 1995.