STOCHASTIC DYNAMICS OF BLIND DECISION FEEDBACK EQUALIZER ADAPTATION

Rodney A. Kennedy, Brian D.O. Anderson, and Robert R. Bitmead.

Department of Systems Engineering, Australian National University
G.P.O. Box 4, Canberra, A.C.T. 2601, Australia.

Abstract. An outline of the analysis of the stochastic dynamics of the blind adaptation of decision feedback equalizers (DFEs) operating on noiseless, binary communication channels is presented. The general mechanism whereby the blind adaptation is flawed in the sense that there exist attraction points in parameter space which lead to unequalized systems is defined and clarified with an example. This behaviour is predicted by defining the implicit error function whose gradient defines the direction of the drift of the adapting parameters. This analysis is complemented by an averaging analysis which also graphically illustrates the drift mechanism.

Keywords. Adaptive Systems; Digital systems; Markov processes; Nonlinear systems; Stochastic systems; Telecommunications.

1. Introduction

Adaptation is employed in equalization when a channel, over which data is sent, is unknown and time-invariant or the channel is slowly time-varying and needs to be tracked [1]. Standard identification schemes normally require channel input a_k and output b_k measurements to identify the channel (see Fig.1). Of course if we had complete knowledge of the real data a_k available at the receiver then this defeats the purpose of equalization. So what is standard in practice is to send a known training sequence $\{a_k\}$ for a limited time duration during which the channel parameters may be learnt. After training, with the equalizer correctly tuned, unknown data is sent and recovered at the equalizer output $\hat{a}_k$ with a sufficiently low probability of error. Typically the adaptation during this time is left on to track slow channel variations. However, in the absence of the (knowledge of the) real data a_k, the data estimates $\hat{a}_k$ are used in the adaptive algorithm.

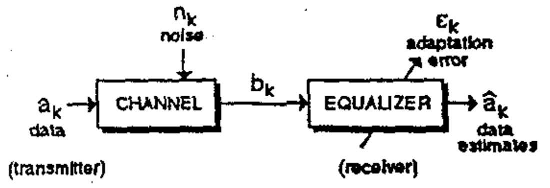

Fig.1 Adaptive Channel Equalization.

Blind adaptation, in the sense that we define, represents something quite different [2]. As in the post-training phase of standard adaptation, the data estimates $\hat{a}_k$ are used in lieu of the real data a_k, but never is a training sequence used. So channel identification is based only on signals b_k and $\hat{a}_k$. Therefore blind adaptation concerns *global convergence* issues when the input is unavailable and we have unreliable data estimates because the equalizer need not be tuned. There are a number of situations where blind adaptation is important in practice such as during a break in multipoint communications, see [3].

A popular type of equalizer structure is the decision feedback equalizer (DFE) which is a *non-linear* recursive filter which utilizes past outputs $\hat{a}_{k-1}$, $\hat{a}_{k-2}$ ..., in its filter structure [1]. The literature is thin on the subject of blind adaptation of DFEs [4]—the principle reason being the difficulty of incorporating error propagation effects [5,6] into the analysis of adaptation. The problem is most acute when the DFE is poorly tuned, then decision errors are common (the error probability may be high) and its recursive nature ensures continuing poor error performance. In this situation the effects of errors will be to distort the adaptation.

The other important component of any adaptation scheme, particularly in the blind case, is the choice (or design) of the algorithm. In the blind case, because the statistics of the equalizer output govern the dynamics of adaptation rather than the actual input data, one needs to characterize the attraction points of the algorithm carefully. In this work we will be interested in some of the gross properties of the dynamics of blind adaptation for DFEs.

2. System Description and Notation

2.1 Channel and Equalizer Models

The channel we consider is shown in Fig.2. It will be modelled as a filter, with impulse response $\{h_0, h_1, \ldots\}$, driven by binary data $a_k \in \{-1, +1\}$ (where k denotes the discrete time index). Additive zero-mean channel noise n_k is depicted in Fig.2 but in this work we will regard its influence as negligible relative to error propagation effects.

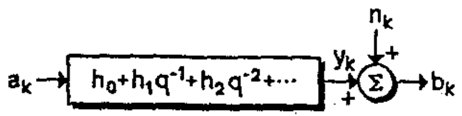

Fig.2 Channel Model.

The DFE structure we study, shown in Fig.3, consists of an N-tap delay line, represented by weights $\{d_i\}$ adapted to minimize the (residual) intersymbol interference (ISI). The tapped delay line is fed by past decisions $\hat{a}_k \in \{-1, +1\}$ which can lead to problems when past decisions are incorrect. This is the error propagation mechanism [5]. Note that in Fig.3, an additional weight d_0 is incorporated when forming an adaptation error signal ϵ_k to compensate for the non-unity channel term h_0.

A more typical and indeed more general DFE structure usually consists of a FIR filter followed by the structure given in Fig.3. We consider the simpler structure because our aim is to understand the effects of error propagation on adaptation.

The fundamental DFE output equation describing the *non-adaptive* operation, from Fig.2 and Fig.3, is given by

$$\widehat{a}_k = \mathrm{sgn}(\sum_{i=0}^{\infty} h_i a_{k-i} - \sum_{i=1}^{N} d_i \widehat{a}_{k-i} + n_k). \qquad (2.1)$$

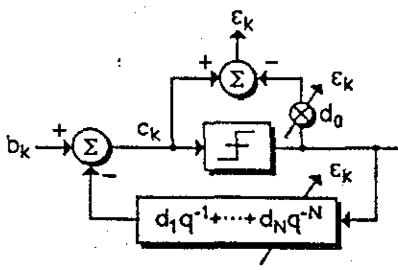

Fig.3 Blind Adaptive DFE.

Remarks:

(i) The size of N is chosen sufficiently large to model adequately the ISI present in the channel to ensure the problem is well-posed.

(ii) We refer to a decision of the form $\widehat{a}_k = \mathrm{sgn}(h_\delta)\, a_{k-\delta}$ as *correct* with the agreed convention that δ is some well defined and fixed delay.

We introduce a vector convention to simplify the presentation: if we define the truncated channel impulse response vector as $H \triangleq (h_0,\ h_1,\ \ldots,\ h_N)' \in \mathbb{R}^{N+1}$, then by convention $\underline{H} \triangleq (h_1,\ h_2,\ \ldots,\ h_N)' \in \mathbb{R}^N$ We also define the vector $D_k \triangleq (d_0,\ d_1,\ \ldots,\ d_N)_k' \in \mathbb{R}^{N+1}$ of time-varying tap weights at time k along with its associated vector $\underline{D}_k \in \mathbb{R}^N$. Further, the vectors representing the past and present data is $A_k \triangleq (a_k,\ a_{k-1},\ \ldots,\ a_{k-N})' \in \mathbb{R}^{N+1}$. In our analysis, we will also be using $\underline{A}_k \in \mathbb{R}^N$, $\widehat{A}_k \in \mathbb{R}^{N+1}$ and $\widehat{\underline{A}}_k \in \mathbb{R}^N$ with obvious definitions.

With these definitions, (2.1) can be written succinctly as

$$\widehat{a}_k = \mathrm{sgn}(A_k'H - \widehat{\underline{A}}_k'\underline{D}_k + Q_k) \qquad (2.2)$$

where Q_k represents a perturbation to the ideal DFE system which we will take to be zero.

2.2 Blind Adaptation Schemes

Our emphasis in this paper is with how error propagation in a DFE interacts with and distorts the stochastic dynamics of adaptation. The simplest *blind adaptation* scheme (Fig.3) that can be employed which updates the taps D_k, effectively identifying the channel, takes the form

$$D_{k+1} = D_k + \gamma\, \epsilon_k\, \widehat{A}_k \qquad (2.3a)$$

where

$$\epsilon_k \triangleq A_k'H - \widehat{A}_k'D_k. \qquad (2.3b)$$

This is a blind algorithm because we use past decisions from $\{\widehat{a}_k\}$ rather than the true data $\{a_k\}$ as the components in the regressor vector. The scalar error ϵ_k represents the discrepancy between the input and the renormalized output of the slicer (see Fig.3). The scalar γ in (2.3a) represents a small (time-invariant) adaptive gain.

3. Parameter Space Partition

In this section we quickly review some known results for DFEs regarding the relationship between regions in parameter space and finite state Markov processes (FSMPs). For a fixed channel H, we define the following hyperplanes in D_k-space (note d_0 is unconstrained):

$$\{D_k \in \mathbb{R}^{N+1}\colon\ A_k'H = \widehat{\underline{A}}_k'\underline{D}_k\}. \qquad (3.1)$$

as we vary across all possible values taken by $A_k \in \mathbb{Z}^{N+1}$ and $\widehat{\underline{A}}_k \in \mathbb{Z}^N$. From (2.2) it is clear that these hyperplanes define the manifolds in $\mathbb{R}^{N+1}$ for which the argument of the signum function in (2.2) is potentially small. These planes act as switching surfaces in the following sense. Define an atomic state [7]

$$X_k \triangleq (\underline{A}_k',\ \widehat{\underline{A}}_k')' \in \mathbb{Z}^{2N}. \qquad (3.2)$$

Then from (2.2) whenever the channel parameters satisfy (3.1) there exists an atomic state X_k and current input a_k such that an arbitrarily small perturbation in $\underline{D}_k$ (d_0 has no effect) can change a decision $\widehat{a}_k = +1$ to $\widehat{a}_k = -1$ (or vice-versa). Before describing the significance of these hyperplanes to the stochastic modelling we give an example.

Example: *Let $N = 2$, $h_0 = 1$, $h_1 = 4$, and $h_2 = 3$. The $4^N = 16$ lines which partition D_k-space are given by $d_1 \pm d_2 = \zeta$, for $\zeta \in \{0, \pm 2, \pm 6, \pm 8\}$ according to (3.1). These lines are depicted in Fig.4. (The point $\underline{D} = \underline{H}$ is indicated by a cross.)*

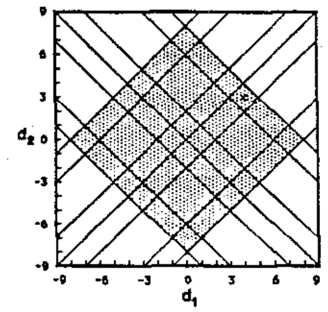

Fig.4 Polytopes for the H=(1,4,3)' Channel.

The hyperplanes above partition D_k-space into a collection of polytopes. The property of polytopes which we need here is that one cannot distinguish between any two DFE D_k parameter settings within a given polytope based on observations of the output $\{\widehat{a}_k\}$ alone. Two DFEs in the same polytope with the same initial conditions and input sequences produce identical outputs $\widehat{a}_k$. Next we move on to some statistical modelling.

Assumption: *The input binary sequence $\{a_k\}$ forms an equi-probable i.i.d. binary sequence.*

Then we have a finite state Markov process describing the stochastic dynamics of the DFE, with 4^N states given by all possible values of X_k (3.2), see [7]. One can verify there is a one-to-one correspondence between polytopes and sets of FSMPs. So the conceptual picture is that as we drift (slowly) through parameter space the underlying FSMP, which governs the full joint statistics of the input a_k and output $\widehat{a}_k$ of a fixed DFE, changes (abruptly) only when we cross polytope boundaries. Inside a given polytope we model the $\{\widehat{a}_k\}$ process by the stationary behaviour of its associated FSMP. When the input inpendence assumption above does not hold and we have input correlation it is possible to still use a FSMP as an approximation device [8,9]. Alternatively we may, in most cases, extend the Markov state space and retain an exact description.

Note that in principle the FSMP provides sufficient information to calculate, in our case, stationary entities of the form

$$R \triangleq E\{\widehat{A}_k\widehat{A}_k'\} \quad \text{and} \quad C \triangleq E\{\widehat{A}_k A_k'\} \qquad (3.3)$$

which are expressible in terms of an invariant probability measure via an unilluminating calculation.

4. Equilibria and Averaging Analysis

4.1 Wiener-Hopf Solution

Our first task is to determine the locations of the attraction (equilibrium) points in D_k-space for the weights when the algorithm (2.3) seeks to minimize some error criterion.

Consider Fig.5 which simply redraws a portion of Fig.2 and Fig.3. This figure suggests that $\widehat{a}_k$ can be ***interpreted*** as an input sequence and that b_k can be interpreted as a noiseless desired response. Thus the problem now looks clasical. As is well known, the objective of the "LMS" algorithm (2.3) is to minimize the mean square error defined by $\xi(D_k) = E\{\epsilon_k^2(D_k)\}$ which is a quadratic function of D_k and hence uni-modal. The tap weight setting, D_{EQ} (equilibrium), which gives the minimum mean square error is given by the classical discrete time Wiener-Hopf formula [10]:

$$D_{EQ} \triangleq E\{\widehat{A}_k\widehat{A}_k'\}^{-1}E\{\widehat{A}_k b_k\} \qquad (4.1)$$

where, from (3.3) we have assumed $\det(\boldsymbol{R}) \neq 0$. Note, from this point on we consider only the case when $\boldsymbol{R}$ is

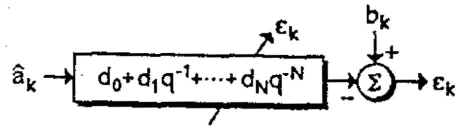

Fig.5 Equivalent Wiener Filter Problem.

non-singular. We will see that this restriction is justifiable when we come to §5.5.

In reality, b_k is not a user supplied sequence but simply the noiseless channel output which in our case is identified with $b_k \triangleq A_k'H$ (Fig.2). Hence (4.1a) may be expanded to

$$D_{EQ} \triangleq E\{\widehat{A}_k\widehat{A}_k'\}^{-1}E\{\widehat{A}_k A_k'\}\ H \qquad (4.2a)$$

$$= \boldsymbol{R}^{-1}\boldsymbol{C}\ H. \qquad (4.2b)$$

This formula for D_{EQ} makes a qualitative and quantitative analysis more accessible. Specifically, it is clear that $D_{EQ} = \mathrm{sgn}(h_0)\ H$ only under special circumstances and this highlights a major problem of blind adaptation with DFEs, i.e., $\mathrm{sgn}(h_0)\ H$ (which we will see is the desired tap weight setting [5]) need not be a global attraction point for the adaptation algorithm.

The mean square error surface is quadratic with $\boldsymbol{R} = \boldsymbol{I}$ in the *training sequence* case, because the input is i.i.d., thus uncorrelated. With blind adaptation we will see that *locally* the mean square error surface is quadratic (although $\boldsymbol{R} \neq \boldsymbol{I}$ in general) given stationarity of the output $\{\widehat{a}_k\}$ process. The above analysis assumes matrices $\boldsymbol{R}$ and $\boldsymbol{C}$ are constant. We qualify this assumption for the blind DFE case later, but first we move onto another result.

4.2 Averaged Equation Trajectory

To characterize the attraction points for the blind algorithm (2.3) it is sufficient to define the error function implicit in the adaptation. A close parallel exists here with the work of Verdú [11], except he treats linear equalization, in this regard. However because of the conceptual aid which it affords it is desirable to complement this error surface analysis with averaging theory to describe the mean drift of the D_k parameters.

In this section we will give an expression for the mean trajectory for the tap weights D_k as they drift towards the equilibrium D_{EQ} given by (4.2b), assuming $\boldsymbol{R}$ and $\boldsymbol{C}$ are constant for the moment. The mean trajectory is in the sense of the ensemble of input sequences $\{a_k\}$. However, the mean trajectory is also valuable in that individual realizations will tend to cluster closely about this mean, at least for sufficiently small gain γ as we will see from an example.

We now analyze the adaptation equations (2.3). Substituting the expression for the error ϵ_k (2.3b) into the LMS tap weight update equation (2.3a) we obtain

$$D_{k+1} = (I - \gamma\ \widehat{A}_k\widehat{A}_k')\ D_k + \gamma\ \widehat{A}_k A_k' H. \qquad (4.3)$$

If γ is sufficiently small, then the increment in going from D_k to D_{k+1} will also be small (noting that all quantities in (4.3) are bounded). Further, we might anticipate that the matrices $\widehat{A}_k\widehat{A}_k'$ and $\widehat{A}_k A_k'$ take on a large number of (statistically) different values whilst $\{D_k\}$ evolves very little with time. Hence we might predict that the deterministic equation describing the mean tap trajectory, $\{\widetilde{D}_k)\}$, takes the form,

$$\widetilde{D}_{k+1} = \left(I - \gamma E\{\widehat{A}_k\widehat{A}_k'\}\right)\ \widetilde{D}_k + \gamma E\{\widehat{A}_k A_k'\}\ H \quad (4.4a)$$

$$= (I - \gamma\boldsymbol{R})\ \widetilde{D}_k + \gamma\boldsymbol{C}\ H. \qquad (4.4b)$$

The formal justification that (4.4b) is indeed the correct equation as $\gamma \to 0$ may be found in the literature [12,13] and further study shows that the trajectories of (4.3) cluster closely about the solutions of (4.4b). This property is apparent in our later example in §5.2.

We make some observations regarding (4.4b). It is straightforward to verify that: (a) D_{EQ} (4.2b) is indeed the equilibrium of the averaged (mean) equation (4.4b), and (b) the mean equation (4.4b) is stable if and only if $\gamma\ \lambda_{MAX}(\boldsymbol{R}) < 2$, where $\lambda_{MAX}(\boldsymbol{R})$ is the maximum eigenvalue of $\boldsymbol{R}$.

5. Tap Trajectories During Adaptation

5.1 Piecewise Constant Behaviour

With this section we bring together our previous, largely disconnected, results regarding the polytopes (§3), finite state Markov processes (§3 and §4.1, see also [7]), and the averaging analysis (§4.2). We will demonstrate that the blind LMS algorithm (2.3a) can be (but is not necessarily) attracted to undesirable regions of D_k-space where the channel is not correctly equalized and the error rates are unacceptably high (even in the absence of noise).

The parameters which determine the dynamics of the mean equation (4.4b) are the covariance matrices $\boldsymbol{R}$ and $\boldsymbol{C}$. We also met these matrices earlier in §3 and we showed they could be evaluated with the assistance of FSMPs. Now, recall our one-to-one correspondence between the polytopes and FSMPs (or more generally state transition diagrams, §3). Hence, in D_k-space the matrices $\boldsymbol{R}$ and $\boldsymbol{C}$ will be constant only whilst the tap setting D_k remains inside any one of the D_k-space polytopes (assuming steady state of the underlying FSMP). Therefore, we have shown (albeit informally):

Proposition 1: *In steady state, the matrices* $\boldsymbol{R}$ *and* $\boldsymbol{C}$ *are piecewise constant functions of* D_k, *where the pieces are precisely the* D_k*-space polytopes bounded by the hyperplanes (3.1).*

Remarks:

(i) To emphasize $\boldsymbol{R}$ and $\boldsymbol{C}$ are functions only of the polytope $\mathcal{P}$ for which $D_k \in \mathcal{P}$, we write $\boldsymbol{R}(\mathcal{P})$ and $\boldsymbol{C}(\mathcal{P})$.

(ii) Within each polytope the mean trajectory (4.4b) is determined by a constant coefficient, linear, deterministic difference equation. Hence, over the whole D_k-space, the averaged trajectory describing the complete adaptation is determined by a piecewise constant coefficient, linear, deterministic difference equation (see the example in §5.2), possibly with boundary conditions.

The next property is an embellishment of Proposition 1, and the omitted proof is a simple variant of [21, eqn (2.31)].

Proposition 2: The error surface $\xi(D_k) \triangleq E\{\epsilon_k^2(D_k)\}$ is a piecewise (polytope-wise) quadratic function of D_k given by

$$\xi(D_k \in \mathcal{P}) = H'H - 2H'C(\mathcal{P})D_k + D_k'R(\mathcal{P})D_k. \quad (5.1)$$

Remarks:

(i) The minimum of the mean square error $\xi(D_{EQ}(\mathcal{P}))$ associated with a polytope $\mathcal{P}$, in terms of H, can be written $\xi_{MIN}(\mathcal{P}) = H'(I - C(\mathcal{P})'R(\mathcal{P})^{-1}C(\mathcal{P}))\ H$. However, if $D_{EQ}(\mathcal{P})$ lies outside $\mathcal{P}$ then this minimum need not be achievable.

(ii) Whenever $h_0 \neq 0$ we will see in §5.3 that the global minimum mean square error at $D^{OPT} \triangleq \text{sgn}(h_0)\ H$ is zero for the polytope which contains D^{OPT}, and is locally attainable, in the sense described in the following paragraph.

(iii) The training sequence adaptation error surface has a unique minimum. Blind DFE adaptation has potentially as many equilibria as there are polytopes!

In review, with each polytope $\mathcal{P}$ we have associated an equilibrium (potential attraction point for the blind adaptation algorithm) given by $D_{EQ}(\mathcal{P}) = R(\mathcal{P})^{-1}C(\mathcal{P})\ H$. It is natural to classify two types of equilibria according to whether or not the following property holds:

Definition: $D_{EQ}(\mathcal{P})$ is locally attainable if $D_{EQ}(\mathcal{P}) \in \mathcal{P}$.

If $D_{EQ}(\mathcal{P})$ is locally attainable then $\{D_k\}$ will tend to move towards and settle down around it, whenever $D_k \in \mathcal{P}$. Otherwise, $\{D_k \in \mathcal{P}\}$ will tend to move towards the boundary $\partial\mathcal{P}$ of $\mathcal{P}$ nearest to $D_{EQ}(\mathcal{P})$ and thus head on into an adjacent polytope. So all locally attainable equilibria are real attraction points for the blind algorithm. The example in the next subsection best illustrates these ideas.

5.2 Example of Averaged Trajectory

As in §3 we choose the example $h_0 = 1$, $h_1 = 4$ and $h_2 = 3$. In Fig.6 we have plotted a large number of averaged trajectories according to (4.4b) with $\gamma = 0.01$, noting that the R and C matrices are now dependent on the polytopes (pictured in Fig.5). Note Fig.6 is a two-dimensional projection of D_k-space, therefore some of the averaged trajectories only appear to cross. The starting d_0-component for all trajectories was arbitrarily selected at zero. Naturally the mean evolution of d_0 during adaptation cannot be discerned in such a figure. Clearly a predictable refraction phenomenon is indicated as we pass across polytope boundaries.

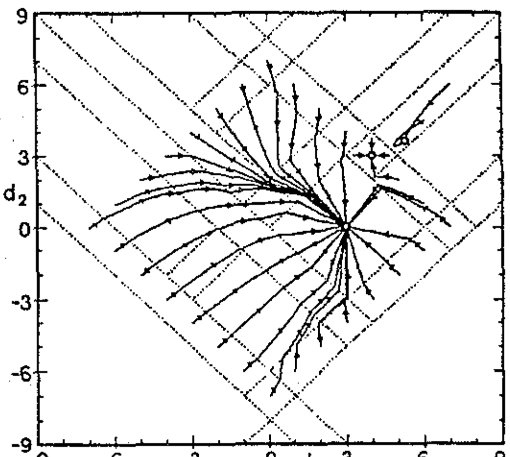

Fig.6 Refracting Averaged Trajectories.

Figure 7 shows the precise sense in which to interpret Fig.6. It shows an insert of Fig.6 with a single (bold) averaged trajectory (plucked from Fig.6) and four realizations initialized from $(0,7,0)'$ (i.e., simulations according to (4.3) generated via a random number generator) which appear to cluster about the averaged trajectory. Note for this example there are only three locally attainable equilibria at $(1,4,3)'$, $(4,3,0)'$ and $(4.667, 5.333, 3.667)'$. In Fig.6 the 2D projections of these equilibria (depicted as small circles) are given by $(4,3)'$, $(3,0)'$ and $(5.333, 3.667)'$ and these appear as (local) attraction points for the mean trajectories. The last equilibrium shows blind adaptation based on the blind algorithm (2.3) may be flawed in the sense that it does not correspond to an equalized system (whereas the first two do, corresponding to delay-0 and dealy-1 systems, respectively).

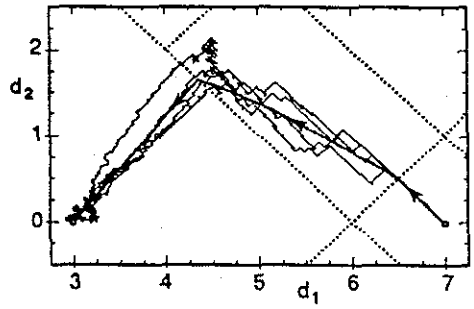

Fig.7 Averaged Trajectory with 4 Realizations.

5.3 Approximate and Exact Equilibria

The equilibria predicted by the theory developed are implicitly based only on mean behaviour (4.4b). However we note that part of the driving term of the adaptation equation (2.3a) is given by ϵ_k (2.3b). Error ϵ_k in turn is only identically zero (for all k) when $D_k = D^{OPT} \triangleq \text{sgn}(h_0)\ H$ (given $h_0 \neq 0$; the proof is straightforward and omitted). So we say that only D^{OPT} is an *exact* equilibrium in the sense that $\epsilon_k(D^{OPT}) = 0$. All other locally attainable equilibria are termed *approximate* equilibria and the physical manifestation of a locally attainable equilibrium with $\epsilon_k(D_{EQ}(\mathcal{P})) > 0$ is that the sequence $\{D_k\}$ is observed to jiggle about (but generally stay in the vicinity of) $D_{EQ}(\mathcal{P})$. This phenomenon is apparent in the lower left of Fig.8. The questions arise [14]:

(a) Regarded as a small noise can the variations in ϵ_k (estimating the gradient) drive $\{D_k\}$ away from $D_{EQ}(\mathcal{P})$ such that $\{D_k\}$ hits the boundary $\partial\mathcal{P}$?

(b) What is the expected time to do so?

In the stochastic process literature [15] this is known as an exit problem. These are crucial questions because if the DFE hangs at an equilibrium $D_{EQ}(\mathcal{P})$ corresponding to high error rates, e.g., $D_{EQ} = (4.667, 5.333, 3.667)'$ in Fig.6, then an unacceptably high exit time has serious practical consequences. On an equal footing we may ask about exit times from regions of correct equalization and consider the consequences. Also channel noise can have precisely the same effect as is well known. These questions are the subject of current research and answers are not presented in this paper.

5.4 Delay-type Equilibria Local Attainability

In this section we consider those aggregations of polytopes in D_k-space which yield a decision sequence which is a delay of the input (with an associated possible sign change) under steady state, i.e., $\hat{a}_k = \text{sgn}(h_\delta)\,a_{k-\delta}, \forall k$. We saw this behaviour was possible in our example in Fig.6. We derive necessary (and conjecture sufficient) conditions for attainability of the attraction points of these groups of polytopes for blind algorithms in terms of the channel parameters.

Let $\sigma_\delta \triangleq \operatorname{sgn}(h_\delta)$, then rewriting (2.1) we have,

$$\widehat{a}_k = \operatorname{sgn}\left(h_\delta a_{k-\delta} + U_k(\delta) + V_k(\delta)\right), \quad 0 \le \delta \le N \quad (5.2a)$$

$$\text{where} \quad U_k(\delta) \triangleq \sigma_\delta \sum_{i=\delta+1}^{N} d_{i-\delta}(a_{k-i} - \sigma_\delta \widehat{a}_{k+\delta-i}) \quad (5.2b)$$

$$V_k(\delta) \triangleq \sum_{i=0}^{\delta-1} h_i a_{k-i} + \sum_{i=\delta+1}^{N} (h_i - \sigma_\delta d_{i-\delta}) a_{k-i} - \sum_{i=N-\delta+1}^{N} d_i \widehat{a}_{k-i}. \quad (5.2c)$$

We also define an upper bound on (5.2c)

$$V_{MAX}(\delta) \triangleq \sum_{i=\delta+1}^{N} |h_i| + \sum_{i=\delta+1}^{N} |h_i - \sigma_\delta d_{i-\delta}| + \sum_{i=N-\delta+1}^{N} |d_i|. \quad (5.3)$$

The reason for the curious decomposition given by (5.2a) will become clearer later. We will see that δ corresponds to a nominal time delay and $\sigma_\delta \in \{-1, +1\}$, corresponds to an associated sign change of the channel-DFE combination.

We define two proper subsets of the set Ω of all 4^N atomic states X_k (3.2), parametrized by $0 \le \delta < N$. (The case $\delta = N$ needs to be treated separately but fortunately is easily disposed of.) Define,

$$\omega_+(\delta) \triangleq \{X_k \in \Omega\colon \widehat{a}_{k-i} = +\sigma_\delta a_{k-\delta-i},\ i = 1, \ldots, N-\delta\}$$

$$\omega_-(\delta) \triangleq \{X_k \in \Omega\colon \widehat{a}_{k-i} = -\sigma_\delta a_{k-\delta-i},\ i = 1, \ldots, N-\delta\}$$

both of which consist of collections of $2^{N+\delta}$ atomic states where (precisely) the $N-\delta$ most recent decisions are of the form $\widehat{a}_m = +\sigma_\delta a_{m-\delta}$ and $\widehat{a}_m = -\sigma_\delta a_{m-\delta}$, respectively.

Note the definitions of $\omega_\pm(\delta)$ simply express that the member atomic state vectors have their $N+i$th component equal to $\pm\sigma_\delta$ times the $\delta+i$th component for $i = 1, 2, \ldots, N-\delta$, and thus these subsets are in effect independent of k (as the notation suggests). We are aiming for conditions under which only the atomic states in $\omega_+(\delta) \in \Omega$ are recurrent. The standard notion of a closed subset of Ω will considerably simplify development.

Definition: *A subset of Ω is **closed** if any transition from any one atomic state in the subset is only to another atomic state within the subset.*

The following statements are equivalent: (a) Suppose $X_k \in \omega_\pm(\delta)$, then all future ($m \ge k$) decisions are of the form $\widehat{a}_m = \pm\sigma_\delta a_{m-\delta}$. (b) $\omega_\pm(\delta)$ is closed. Hence to investigate channel-DFE combinations yielding simple time delay behaviour we need only to determine when a $\omega_\pm(\delta)$ is closed and reachable from arbitrary states within Ω. The following proposition narrows our investigations by showing a DFE can never behave consistently according to the law $\widehat{a}_m = -\sigma_\delta a_{m-\delta}$ when in a steady state stochastic environment, and it also gives necessary and sufficient conditions for $\omega_+(\delta)$ closure.

Proposition 3: *(a) $\omega_+(\delta)$ is closed if and only if*

$$|h_\delta| > V_{MAX}(\delta). \quad (5.4)$$

(b) $\omega_-(\delta)$ is never a closed subset.

If (5.4) holds, then contriving an input sequence which visits all atomic states in $\omega_+(\delta)$ when the initial state is in $\omega_+(\delta)$ is straightforward. This shows no proper subset of $\omega_+(\delta)$ is closed assuming $\omega_+(\delta)$ itself is closed (i.e., $\omega_+(\delta)$ is a set of recurrent states). We formulate this as:

Proposition 4: *$\omega_+(\delta)$ is indecomposable.*

The inequality (5.4) can only hold for at most one value of δ. We state this result as Proposition 5.

Proposition 5: *$\omega_+(\delta)$ is closed for at most one δ.*

Proposition 5 can be viewed as a special case of a more general problem, now considered. Having established that only under suitable conditions $\omega_+(\delta)$ is closed and indecomposable, the crucial question arises as to whether it can be reached from an arbitrary atomic state $X_k \in \Omega \setminus \omega_+(\delta)$ by at least one input sequence. Then there are a number of side issues related to this, e.g.; (a) the expected capture time by $\omega_+(\delta)$, (b) which channels have an acceptable capture time, etc., (see [7]). A full answer to this question is not yet known. We present the following result (Proposition 6) and important conjecture (Conjecture 7).

Proposition 6: *Let $|h_\delta| > V_{MAX}(\delta)$ for some $0 \le \delta < N$. Then the following alternative conditions are sufficient to guarantee that there exists an input sequence such that $N-\delta$ consecutive $\widehat{a}_m = +\sigma_\delta a_{m-\delta}$ decisions are made:*

(i) $\delta = 0, 1, N-2, N-1, N$ (and thus cases $N = 1, 2, 3, 4$).

(ii) $\eta \operatorname{sgn}(d_1) = \eta^2 \operatorname{sgn}(d_2) = \ldots = \eta^{N-\delta} \operatorname{sgn}(d_{N-\delta})$ for $\eta \in \{+1, -1\}$.

Conjecture 7: *Let $|h_\delta| > V_{MAX}(\delta)$ for some $0 \le \delta < N$. Then there exists an input sequence such that $N-\delta$ consecutive decisions are made of the form $\widehat{a}_m = +\sigma_\delta a_{m-\delta}$.*

Remarks:

(i) With (5.4) satisfied, and the hypothesis of Proposition 6 fulfilled, $\omega_+(\delta)$ is closed, indecomposable and reachable, so that $Pr(X_k \in \omega_+(\delta)) \to 1$ exponentially fast as $k \to \infty$. Hence under stationarity the channel-DFE combination produces decisions of the form $\widehat{a}_m = +\sigma_\delta a_{m-\delta}$ if and only if $|h_\delta| > V_{MAX}(\delta)$ with $\omega_+(\delta)$ reachable.

(ii) Given a time-invariant channel H, define the following regions (at least one does exist) of D_k-space, by rewriting (5.4):

$$\mathcal{J}(\delta) \triangleq \left\{\rho_\delta > \sum_{i=\delta+1}^{N} |h_i - \sigma_\delta d_{i-\delta}| + \sum_{i=N-\delta+1}^{N} |d_i|\right\} \quad (5.5a)$$

$$\rho_\delta \triangleq |h_\delta| - \sum_{i=0}^{\delta-1} |h_i|, \ 0 \le \delta \le N. \quad (5.5b)$$

(These regions for $\delta > 0$ generalize a condition derived by Jennings [4].) Then $\widehat{a}_m = +\sigma_\delta a_{m-\delta}$ under steady state conditions only if $D_k \in \mathcal{J}(\delta)$ and sometimes if (according to the reachability of $\omega_+(\delta)$ and the initial conditions). Note this region may also be written $\rho_\delta > \|\underline{D}_k - \sigma_\delta S_N^\delta \underline{H}\|_1$, where S_i denotes an $i \times i$ matrix of super-diagonal ones. Hence the projection of $\mathcal{J}(\delta)$ onto $\underline{D}_k$-space defines an l_1-ball with centre $\sigma_\delta S_N^\delta \underline{H}$ and radius ρ_δ. (Note the d_0 component of D_k does not play a role in the constraint in (5.5a).) Region $\mathcal{J}(\delta)$ is non-empty only if $\rho_\delta > 0$. Note for our example in Fig.6, $\mathcal{J}(0)$ and $\mathcal{J}(1)$ are non-empty because $\rho_0 = |h_0| = 1 > 0$ and $\rho_1 = |h_1| - |h_0| = 3 > 0$, but $\mathcal{J}(2)$ is empty because $\rho_2 = |h_2| - |h_1| - |h_0| = -2 < 0$.

Now on to the main result which shows it is simple to check for the existence of delay-like attraction points for the blind algorithms for the adaptive DFE.

Theorem 8: *A necessary condition for the "LMS" blind adaptive algorithm*

$$D_{k+1} = D_k + \gamma\, \epsilon_k\, \widehat{A}_k$$

where

$$\epsilon_k \triangleq A_k' H - \widehat{A}_k' D_k$$

to have a locally attainable equilibrium corresponding to the channel-DFE combination producing delay decisions of the form $\hat{a}_m = +\sigma_\delta a_{m-\delta}$ under steady state is $p_\delta > 0$ (5.5b). The equilibrium is given by $D_{EQ}(\mathcal{P}) \triangleq \sigma_\delta S^\delta_{N+1} H$, i.e., a simple shift of H with a possible sign flip. The condition is also sufficient when $\omega_+(\delta)$ is reachable (from all atomic states in $\Omega \setminus \omega_+(\delta)$).

Remarks:

(i) For $\delta = 0$, $D_{EQ}(\mathcal{P}) = D^{OPT} \triangleq +\sigma(0)\ H$ is always locally attainable whenever $h_0 \neq 0$ and achieves the global minimum mean square error of zero, i.e., is an exact equilibrium. (If $h_0 = 0$ then trivially $D_{EQ}(\mathcal{P}) = +\sigma_1\ S_{N+1}\ H$ is always locally attainable and exact, and so on.)

(ii) The "LMS" qualifier in the theorem statement is superfluous. The same result holds for any adaptation algorithm which seeks to minimize the mean square error (2.3b).

5.5 White Equilibria

As we have earlier commented, when $\hat{a}_k = +\sigma_\delta a_{k-\delta}$, the $\{\hat{a}_k\}$ process is composed of a sequence of independent equi-probable binary random variables. Let us term any equilibrium with the $\{\hat{a}_k\}$ process white a *white equilibrium*. In this subsection, we shall present further results on this class and indicate some open problems.

We now give two closely related propositions which imply that adaptation should be restricted to a well defined region of D_k-space.

Proposition 9: *Suppose $\{\hat{a}_k\}$ forms an i.i.d. binary random sequence (under steady state). Then*

$$\|H\|_1 > \|\underline{D}_k\|_1. \qquad (5.6)$$

Remarks:

(i) As an example this region is shown shaded as a diamond in Fig.4. Here $H = (1,4,3)'$ and we need $|d_1| + |d_2| < 8$; the output of the DFE can be independent only whilst $(d_1, d_2)'$ lies within this diamond.

(ii) $\|H\|_1$ is the peak excursion of the noiseless channel output when driven by an independent binary input. Hence we can estimate $\|H\|_1$ by channel output measurements and thus impose during adaptation the requirement that $\{D_k\}$ not leave (5.6).

Proposition 10: *If $\det(R(\mathcal{P})) = 0$ then $\|\underline{D}_k\|_1 > \|H\|_1$.*

Remarks:

(i) This justifies the earlier restriction that we should only consider polytopes $\mathcal{P}$ satisfying $\det(R(\mathcal{P})) \neq 0$, because otherwise we would be considering a region of D_k-space which is complementary to the l_1-ball which, by Proposition 9, contains the only polytopes of interest and to which adaptation is sensibly constrained.

(ii) When $N = 1$ and $N = 2$, or when all d_i are zero (which occurs in decision directed equalization), we can show that the only way the output $\{\hat{a}_k\}$ can be white is for the DFE to produce decisions of the form $\hat{a}_m = +\sigma_\delta a_{m-\delta}$, for some δ (only one).

Our analysis leads to the following conjecture:

(DFE) Conjecture 11: *Let $\{a_k\}$ be an independent sequence of random variables taking values in $\{-1, +1\}$ with equal probability. Suppose that*

$$\hat{a}_k = \operatorname{sgn}\Big(\sum_{i=0}^{N} h_i a_{k-i} - \sum_{i=1}^{N} d_i \hat{a}_{k-i}\Big)$$

and the $\{\hat{a}_k\}$ is independently distributed. Then for some $\delta \in \{0, \ldots, N\}$, there holds

$$\hat{a}_k = \sigma_\delta\, a_{k-\delta}\ \ \forall\{a_k\}$$

where $\sigma_\delta \triangleq \operatorname{sgn}(h_\delta)$.

If the conjecture held we would have a way of statistically testing the output of a DFE to prove it was correctly equalizing the channel up to a delay. The corresponding question for the simpler DDE is solved in [16].

6. Conclusion

We have used averaging theory and the theory of finite state Markov processes to describe the stochastic dynamics of blind adaptation of DFEs. Many open problems remain but we have been successful in characterizing those classes of equilibria leading to correct equalization. Hopefully our tools may be extended to provide deeper insights into the operation of blind adaptive DFEs.

References

[1] S.U.H. Qureshi, "Adaptive Equalization," *Proc. IEEE*, vol.73, No.9, pp.1349-1387, September 1985.

[2] A. Benveniste, M. Goursat, and G. Ruget, "Robust Identification of a Non-minimum Phase System: Blind Adjustment of Linear Equalizer in Data Communications," *IEEE Trans. on Auto. Control*, vol.AC-25, No.6, pp.385-399, June 1980.

[3] D.N. Godard, "Self Recovering Equalization and Carrier Tracking in Two-dimensional data Communication Systems," *IEEE Trans. on Communications*, vol.COM-28, pp.1867-1875, November 1980.

[4] A. Jennings, "Analysis of the Adaption of Decision Feedback Equalizers with Decision Errors," Internal Report Telecom Aust. Research Lab., July 1985.

[5] D.L. Duttweiler, J.E. Mazo, and D.G. Messerschmitt, "An Upper Bound on the Error Probability in Decision Feedback Equalizers," *IEEE Trans. on Information Theory*, vol.IT-20, pp.490-497, July 1974.

[6] A. Cantoni, and P. Butler, "Stability of Decision Feedback Inverses," *IEEE Trans. on Communications*, vol.COM-24, pp.1064-1075, September 1976.

[7] R.A. Kennedy, and B.D.O. Anderson, "Recovery Times of Decision Feedback Equalizers on Noiseless Channels," *IEEE Trans. on Communications*, vol.COM-35, pp.1012-1021, October 1987.

[8] J.J. O'Reilly, and A.M. de Oliveira Duarte, "Error Propagation in Decision Feedback Receivers," *Proc. IEE Proc. F, Commun., Radar and Signal Process.*, vol.132, no.7, pp.561-566, 1985.

[9] A.M. de Oliveira Duarte, and J.J. O'Reilly, "Simplified Technique for Bounding Error Statistics for DFB Receivers," *Proc. IEE Proc. F, Commun., Radar and Signal Process.*, vol.132, no.7, pp.567-575, 1985.

[10] B. Widrow, and S.D. Stearns, "Adaptive Signal Processing ," Prentice Hall Inc., Englewood Cliffs, N.J., 1985.

[11] S. Verdú, "On the Selection of Memoryless Adaptive Laws for Blind Equalization in Binary Communication," *Proc. Sixth Intl. Conf. on Analysis and Optimization of Systems*, Nice, France, June 1984.

[12] M. Cottrell, J.C. Fort, and G. Malgouyres, "Large Deviations and Rare Events in the Study of Stochastic Algorithms," *IEEE Trans. on Auto. Control*, vol.AC-28, No.9, pp.907-920, September 1983.

[13] B.D.O. Anderson, R.R. Bitmead, C.R. Johnson, Jr., P.V. Kokotovic, R.L. Kosut, I.M.Y. Mareels, L. Praly, and B.D. Riedle, "Stability of Adaptive Systems: Passivity and Averaging Analysis," MIT Press, Cambridge, Massachusetts 1986.

[14] J.E. Mazo, "Analysis of Decision-Directed Equalizer Convergence," *Bell Syst. Tech. J.*, Vol.59, No.10, pp.1857-1876, December 1980.

[15] M.I. Freidlin, and A.D. Wentzell, "Random Perturbations of Dynamical Systems," Springer-Verlag New York Inc., 1984.

[16] R.A. Kennedy, G. Pulford, B.D.O. Anderson, and R.R. Bitmead, "When has A Decision-Directed Equalizer Converged?," *IEEE Trans. on Communications*, (accepted for publication).