ISSPA87, signal processing, theories, Implementations and Applications.
Editor: B. Boashash, Brisbane, Australia
24-28 August, 1987.

Decision Feedback Equalizers: Concepts Towards Design Guidelines*

B. D. O. ANDERSON, R. A. KENNEDY and P. R. BITMEAD
Department of Systems Engineering, Research School of Physical Sciences, Australian National University, G.P.O. Box 4, Canberra A.C.T., 2601, Australia
* Research supported by the Australian Telecommunications and Electronics Research Board (A.T.E.R.B.)

ABSTRACT: A quantitative and qualitative analysis of the stochastic dynamics of the blind adaptation of decision feedback equalizers (DFEs) operating on noiseless communication channels is presented. The overall goal, not yet acheived, is to identify channel classes for which DFE use is a practical option. Averaging analysis is invoked to predict the existence and location of undesirable attraction points to the adaptation algorithm. The parameter space in which adaptation takes place is shown to be subdivided into a large but finite set of convex polytopes when the channel can be modelled as a finite impulse response filter. With each polytope we show that the parameters describing the averaged equation are fixed. Further, for every point in a given polytope the error performance is described by a single finite state Markov process. Hence, the adaptive dynamics and the error properties of a noiseless DFEs do not depend continuously with the operating point in parameter space but rather only as a function of polytope membership. The possibility arises of the algorithm becoming hung at an inappropiate attraction point.

1 INTRODUCTION

Equalization is the process of cancelling out the distorting effect of a linear channel. Decision feedback equalization refers to two properties of a particular type of equalizer depicted in Fig.1: the cancellation is effected by feeding back the equalizer output (which is supposed to be identical to the channel input, perhaps with delay), and a slicer is used to decide whether the equalizer output should be +1 or −1. Of course such an equalizer is only of use when the a_k sequence takes values ±1. (Generalizations to M−level sequences are of course possible.)

In adaptive decision feedback equalizers (DFEs), the d_i taps are adjusted by processing in some way the outputs of the equalizer. When the input sequence a_k giving rise to the outputs is not known, which is common, the equalization is blind. The best setting of the d_i is $d_i = h_i$ (assuming $h_0 > 0$ for the moment); then if a sequence of a_k of length N is correctly decoded, equalization will thereafter be perfect, the fedback signal exactly cancelling the distortion. Of course, equalization can be disturbed by a large enough noise pulse on the channel output.

What are the key issues? First, in relation to non−adaptive equalization, one can note that even if $d_i = h_i$ for all i, noise, or incorrect initial conditions can mean that initially, $\hat{a}_k \neq a_k$. Then the key question arises: how long will it take to obtain correct equalization or recovery from noise? For practical channels, equalization seems rapid. However, one can readily contrive channel impulse responses for which, at typical bit rates and impulse response lengths, correct equalization will only occur after 10^{10} years on average [1]. So the real question is: for what classes of channels will equalization occur in an acceptable time? Some partial answers are known, [2]. Incidentally, tools for evaluating the time to equalization reasonably accurately are available [3,4]. However, there is a need for further analytic tools.

Once one turns to adaptive identification schemes, where the d_i are initially different form the h_i, the range of questions is even more broad. The most general questions are those of relating the time of convergence of the adjustable d_i to acceptable values to certain properties of the channel. Going along with this are questions of deriving analytical tools, and identifying classes of channels for which one has an acceptable overall equalization time (resulting from both d_i adjustment and recovery from noise or inappropiate initial conditions).

The adaptation algorithm for adjusting the d_i can be thought of as searching for a minimum; a noisy gradient algorithm is most common. As it turns out, there can be

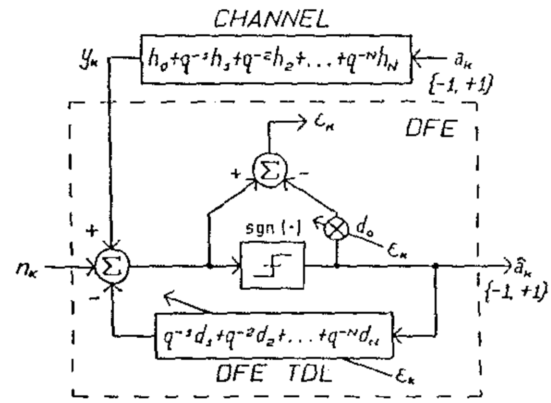

FIGURE 1: Model of Channel and Adaptive DFE.

more than one local minimum of the underlying objective function. So we have to be concerned with how long one takes to get to a local minimum, and whether the errors in estimating the gradient, interpreted as a noise disturbance, can cause movement away from one local minimum to another. Again, one issue is whether or not a particular local minimum corresponds to correct equalisation or not. This occurence of multiple local minima has been observed in other forms of equalizer, see e.g., [5−9].

The analysis of some of these questions is pursued in this paper. One of our principal tools is averaging theory, which applies when the time scale governing adjustment of the d_i is much slower than the time scale associated with the channel and non−adaptive DFE stochastic dynamics. The issue of jumping out from one local minimum to reach a better one is hardly addressed. The theoretical tools which will ultimately give typical times are those of large deviation theory including importance sampling [10], and the recent theories of [11,12] developed to study systems excited by small amounts of noise (modelling the gradient estimation errors in our problem).

Much of this paper is a consequence of the ideas presented by Jennings [13] which also addresses the problem of blind adaptation of DFEs.

The paper is structured as follows. In section 2, we provide a precise description of the adaptive DFE. In section 3, we establish that DFE performance depends on the feedback tap weights d_i in a *piecewise constant* manner. Section 4 discusses the modelling of channel and DFE using finite state Markov processes. Section 5 discusses the equilibria or local minima using averaging theory, and in section 6 we discuss qualitative features of the tap weight variations. Section 7 contains concluding remarks.

2 SYSTEM DESCRIPTION AND NOTATION

The system we consider is shown in Fig.1. The communication channel will be modelled as a finite impulse response (FIR) filter, with impulse response $h_0, h_1, \ldots, h_N$, driven by binary data $a_k \in \{-1,+1\}$ (k denotes the discrete time index). Channel noise n_k is depicted in Fig.1, however, we will regard its influence as secondary to simplify our analysis. Therefore, we make our first assumption:

(A.1) Channel noise is negligible, i.e., $n_k = 0$.

The DFE consists of a N−tap delay line, represented by weights $d_1, d_2, \ldots, d_N$ adapted to minimize the (residual) intersymbol interference (ISI). This tapped delay line is fed by past decisions or data estimates $\hat{a}_k \in \{-1,+1\}$ and this can lead to problems when past decisions are incorrect, an effect referred to as error propagation [14]. Note that in Fig.1, an additional weight d_0 is incorporated when forming an error signal ϵ_k to compensate for the non−unity channel cursor h_0. We will describe the adaptive update mechanism which uses ϵ_k, shortly.

The fundamental DFE output equation describing *non−adaptive* operation, from Fig.1 and incorporating assumption (A.1), is given by

$$\hat{a}_k = \mathrm{sgn}(h_0 a_k + \sum_{i=1}^{N} h_i a_{k-i} - \sum_{i=1}^{N} d_i \hat{a}_{k-i}). \qquad (2.1)$$

We make two further assumptions:

(A.2) The cursor h_0 is non−zero.

We will see that this assumption ensures that under ideal adaptation we have $d_i \rightarrow \mathrm{sgn}(h_0)\, h_i$ for $i = 0, 1, \ldots, N$ and thus $\hat{a}_k = \mathrm{sgn}(h_0)\, a_k$ for all k (in steady state).

(A.3) The number of DFE weights, N+1, equals the length of the channel impulse response including the cursor.

Remark: We refer to a decision of the form $\hat{a}_k = \mathrm{sgn}(h_0)\, a_k$ as *correct* even though initially it may seem nonsensical for the case $h_0 < 0$. The reader should convince themselves that *only* with this definition does the hypothesis that the last N decisions are correct imply the next and all future decisions are correct.

We introduce a vector convention to simplify the presentation. It will be necessary to consider vectors belonging to $\mathbb{R}^N$ and to $\mathbb{R}^{N+1}$. Generally, the DFE output equation (2.1) involves vectors in $\mathbb{R}^N$, whereas we will see the equations describing adaptation necessarily involve vectors in $\mathbb{R}^{N+1}$ since we are adapting N+1 weights. Our convention is as follows. Let W represent a vector in $\mathbb{R}^{N+1}$. Then $\underline{W} \in \mathbb{R}^N$ is derived from $W \in \mathbb{R}^{N+1}$ by deleting the first component (i.e., by projection). For example, we define the channel impulse response vector as $H \triangleq [h_0\ h_1\ \ldots\ h_N]^T \in \mathbb{R}^{N+1}$, then by convention $\underline{H} \triangleq [h_1\ h_2\ \ldots\ h_N]^T \in \mathbb{R}^N$ (v^T denotes the transpose of v). We also define the vector of time−varying (adaptive) tap weights at time k as $D_k \triangleq [d_0\ d_1\ \ldots\ d_N]_k{}^T$ along with its associated vector $\underline{D}_k \in \mathbb{R}^N$. Further, the vectors representing past and present data and data estimates are $A_k = [a_k\ a_{k-1}\ \ldots\ a_{k-N}]^T$ and $\hat{A}_k = [\hat{a}_k\ \hat{a}_{k-1}\ \ldots\ \hat{a}_{k-N}]^T$, respectively. In our analysis, we will also be using $\underline{A}_k \in \mathbb{R}^N$ and $\underline{\hat{A}}_k \in \mathbb{R}^N$, representing the *past only* data and data estimates, respectively.

With these definitions, (2.1) can be written sussinctly

$$\hat{a}_k = \mathrm{sgn}(h_0 a_k + \underline{A}_k{}^T \underline{H} - \underline{\hat{A}}_k{}^T \underline{D}_k). \qquad (2.2)$$

The form of (2.2) will make the partition of parameter space, treated in section 3, more transparent.

Our emphasis in this paper is with the stochastic dynamics of the *blind adaptation* of DFEs. This means we will be interested in the time−varying properties of the tap weights D_k described (typically) through the fundamental tap weight update equation

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

where

$$\epsilon_k \triangleq A_k{}^T H - \hat{A}_k{}^T D_k. \qquad (2.3b)$$

The scalar error ϵ_k represents the discrepancy between the input and the renormalized output of the slicer described by the sgn(·) function, (see Fig.1). The scalar γ in (2.3a) represents a *small* (time−invariant) adaptive gain. Typically, $\gamma = 0.01$. Later we will see that it is the small (but not necessarily infinitesimal) value taken by γ which makes an analysis of adaptation through averaging methods tractable.

Note that in equations (2.3a) and (2.3b) we use the decisions $\{\hat{a}_k\}$ rather than the true data $\{a_k\}$ for the regressor when adapting; this is referred to as *blind adaptation*. In other words, we do not use a training sequence to adapt the DFE tap weights. This results in a considerable complication to the analysis, as analogous results for the case of linear equalization have indicated [5−9].

Equation (2.3a) is in the form of a standard least mean square (LMS) adaptation algorithm [15] (with input $\{\hat{a}_k\}$ rather than the more usual $\{a_k\}$). Equally well we may consider algorithms based on normalized LMS (NLMS) or Recursive Least Squares (RLS) or variations thereof. However, for brevity, we will consider only (2.3a) in detail in this paper.

A major problem we consider in this paper is the interplay between the output equation (2.2) and the adaptive tap weight update algorithm (2.3a). In particular, we will be interested in deriving an *averaged equation* describing the mean trajectory of the DFE tap weights D_k in $\mathbb{R}^{N+1}$ (the mean is in the sense of the ensemble of possible input binary sequences). (Typically, the variation about this mean can be described by a diffusion approximation [8]). We also highlight the role played by the channel parameters H in determining whether or not blind adaptation will lead to difficulties in convergence.

We note that since the subject of our investigations is blind adaptation we need to incorporate error propagation into the analysis, i.e., we cannot assume past decisions are correct. This is the main feature which makes the analysis both difficult and interesting.

3 THE PARTITION OF PARAMETER SPACE

3.1 The General Partition

In this section we quantify the effect of the quantization property of the sgn(·) function on the behaviour of the DFE. We begin by studying the non−adaptive DFE equation (2.2).

Define two concatenated vectors: $V \triangleq [\ \underline{H}^T : -\underline{D}_k{}^T\ ]^T \in \mathbb{R}^{2N}$ representing a general parameter V−space, and

$$X_k \triangleq [\underline{A}_k{}^T : \underline{\hat{A}}_k{}^T]^T \in \mathbb{R}^{2N}, \qquad (3.1)$$

which plays the role of a state vector. With the above concatenated vector definitions we can rewrite (2.2) as

$$\hat{a}_k = \mathrm{sgn}(\ h_0 a_k + V^T X_k\ ), \qquad (3.2)$$

noting that for a fixed X_k, i.e., a given state, $V^T X_k = \pm h_0$ is the equation for a hyperplane in V−space.

Consider now, for sake of argument, $a_k = -1$ for some fixed state X_k. Then for all $V \in \mathbb{R}^{2N}$ satisfying

$$\{V \in \mathbb{R}^{2N} \mid V^T X_k = +h_0\}, \qquad (3.3)$$

we find the argument of the sgn(·) function in (3.2) is zero. Hence, the hyperplane in (3.3) partitions V−space such that on one side of this hyperplane (the side containing the origin) $\hat{a}_k = -\mathrm{sgn}(h_0)$, i.e., a correct decision, and on the other side $\hat{a}_k = +\mathrm{sgn}(h_0)$, i.e., an incorrect decision. Similarly, with $a_k = +1$ we get a hyperplane parallel to the first reflected in the origin, which can be written

$$\{V \in \mathbb{R}^{2N} \mid V^T X_k = -h_0\}. \qquad (3.4)$$

Hence for a *given* state X_k (or equivalently $-X_k$) it is natural to consider the parameter space partitioned into three regions by the two parallel hyperplanes (3.3) and (3.4).

We can repeat this analysis for all 4^N possible values of the state X_k. There are then 4^N pairs of hyperplanes subdividing parameter space. However, the total number of *distinct* hyperplanes taking into account redundancy is 4^N. This gives us our desired *complete* partition of the parameter space. The (smallest) regions thus produced are defined by the intersection of a number of half−spaces, and therefore form *convex polytopes*. We conclude: *For any two points interior to one such polytope we could not distinguish between the corresponding DFE systems on the basis of*

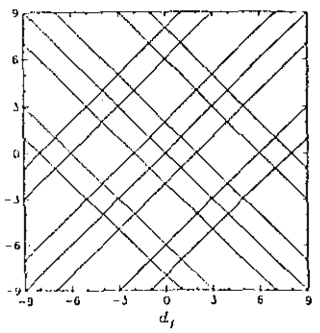

FIGURE 2: The 64 Polytopes in $\underline{D}_k$-space for the N=2 Channel $h_0=1$, $\underline{H}=[4\ 3]^T$.

observations of $\hat{a}_k$. This has wide ranging implications, e.g., during adaptation, only when the trajectory of the tap weights $\{D_k\}$ *crosses the boundary* of a polytope will we observe a change in the stochastic error dynamics of the DFE and hence the behaviour of adaptation.

Remark: The polytopes above may be thought of as lower dimensional projections of polytopes existing in the parameter space of vectors of the form $[H^T{:}D_k^T]^T \in R^{2N+2}$, called W−space, i.e., V−space with h_0 and d_0 augmented.

3.2 Post−Adaptation Error Propagation − Partition

The idea of tessellation of parameter space into polytopes was brought up in [1] which investigated the error propagation mechanism of DFEs whose tap weights had converged (other authors have recognised this discrete property of the parameter space, e.g., [16]). Such polytopes are merely the lower dimensional projections of W−space polytopes defined by imposing $\underline{D}_k = \underline{H}$, deleting d_0 and fixing $h_0 > 0$ (the latter to be compatible with [1], i.e., more generally we would fix $\underline{D}_k = \mathrm{sgn}(h_0)\ \underline{H}$ with $h_0 \neq 0$). Indeed, the derivation of the equations of the hyperplanes defining the polytopes in the previous subsection is a generalization of the derivation found in [1].

3.3 Adaptation to a Time−Invariant Channel − Partition

The channel parameters H will be fixed for a time−invariant channel. Hence, we have only to consider the space of N+1 feedback tap weights, D_k−space (which can be regarded as a function of H) when dealing with adaptation to a time−invariant channel. A typical hyperplane that subdivides D_k−space (or equivalently $\underline{D}_k$−space) into polytopes is given by

$$\{ D_k \in R^{N+1} \text{ or } \underline{D}_k \in R^N \mid \hat{\underline{A}}_k^T \underline{D}_k - A_k^T H \}, \qquad (3.5)$$

for a given state X_k (remember $X_k \triangleq [\ \underline{A}_k^T : \hat{\underline{A}}_k^T\ ]^T$) and fixed data a_k. The derivation of (3.5) is entirely analogous to that in section 3.1 and therefore omitted. We illustrate the preceding discussion and derivation by an example.

3.4 Example of Polytopes in D_k−space for N = 2

Let N = 2, $h_0 = 1$, $h_1 = 4$, and $h_2 = 3$. The $4^N = 16$ lines which partition $\underline{D}_k$−space, $[d_1\ d_2]_k \in R^2$, are given by $d_1 \pm d_2 = \zeta$, for $\zeta \in \{\pm 0, \pm 2, \pm 6, \pm 8\}$ according to (3.5). These lines are depicted in Fig.2. Notice that due to the degeneracy in this example we really only have only 14 *distinct* lines rather than the usual full quota of 16. We will return later to consider this example in more detail.

4 FINITE STATE MARKOV PROCESSES

4.1 Stochastic Modelling of the Non−Adaptive DFE

We need a mechanism whereby we can quantify the effects of a stochastic input $\{a_k\}$ on the behaviour of the DFE. Our quest for a suitable description is eased by making a realistic whiteness assumption on the input a_k:

(A.4) The input binary sequence $\{a_k\}$ is white.
(Note that (A.4) implies $E[A_k A_k^T] = I$.) We also assume a uniform distribution, viz.,

(A.5) The data a_k takes binary values with equal probability.

Then it is a standard result that a finite state Markov process (FSMP) can be developed to describe the stochastic relationship between the decisions $\hat{a}_k$ and the input a_k [1−4,14,17−19] given that the channel and DFE parameters are fixed. However, only the FSMP developed in [1] is sufficiently general (capable of handling the case $\underline{D}_k \neq \mathrm{sgn}(h_0)\ \underline{H}$) to meet our needs for this paper. We will see later that the FSMP describing the non−adaptive stochastic dynamics varies according to the channel and DFE parameters but only in a discrete fashion.

In [1] it was shown that the 4^N values taken on by the atomic state vector X_k (3.1) form Markov states in a FSMP. This set of 4^N Markov states is denoted Ω. This FSMP can simply be viewed as imposing a probabilistic structure onto a state transition diagram which describes all possible transitions between atomic states. Summarizing the development in [1] (see also [19]): Given a "binary" ordering of atomic states indexed by $i \triangleq \frac{1}{2}([2^{2N-1}\ \ldots\ 2\ 1]\ X_k + 2^{2N} - 1)$, where $i \in \{0, 1, 2, \ldots, m \triangleq 4^N - 1\}$, we associate a probability transition matrix P and time−varying (non−stationary) probability density vector π_k such that $\pi_{k+1} = P\ \pi_k$. Typically there exists a unique stationary or steady state distribution $\pi_s \triangleq [\pi_0\ \pi_1\ \ldots\ \pi_m]^T$ such that $\pi_s \triangleq P\ \pi_s$. (Unfortunately, there exist P with non−unique π_s. However, it appears that this only occurs for collections of polytopes which adaptation should sensibly avoid, see section 6.) For our purposes we wish simply to indicate how certain important statistical quantities can be calculated using this FSMP, assuming π_s is unique.

4.2 Error Probabilities and the Covariance Matrices

To make calculations tractable, we make the following assumption:

(A.6) All FSMPs considered are assumed stationary.
For definiteness we show how to calculate the various stochastic quantities of interest by considering the case N = 2. (Higher order cases are handled similarly, albeit with increased computational burden.) With N = 2 there are 16 atomic states. In the binary ordering of the last subsection, atomic state $X_k = [-1\ +1\ +1\ -1]^T$ codes as i = 6, and $X_k = [+1\ -1\ +1\ +1]^T$ codes as i = 11. Recalling the definition of the components in X_k (3.1), we can compute the error probability $P_e \triangleq \Pr(\hat{a}_k \neq \mathrm{sgn}(h_0)\ a_k)$ under stationarity by first recording the indices corresponding to either (i) those atomic states whose first component does not equal $\mathrm{sgn}(h_0)$ times the third component, i.e., an error occured one time step ago, or alternatively (ii) those atomic states whose second component does not equal $\mathrm{sgn}(h_0)$ times the fourth component, i.e., an error occured two time steps ago. Then steady state P_e can be computed by adding the components of π_s whose indices were recorded. For example, if $\mathrm{sgn}(h_0) > 0$, then in each case we get $P_e = \pi_2+\pi_3+\pi_6+\pi_7+\pi_8+\pi_9+\pi_{12}+\pi_{13}$ and $P_e = \pi_1+\pi_3+\pi_4+\pi_6+\pi_9+\pi_{11}+\pi_{12}+\pi_{14}$, respectively. Of course, the two expressions for P_e are necessarily equal and therefore imply $\pi_1+\pi_4+\pi_{11}+\pi_{14} = \pi_2+\pi_7+\pi_8+\pi_{13}$.

The second major class of statistical quantities of interest to us is covariance. In particular we will be interested in Toeplitz matrices of the form $R \triangleq E[\hat{A}_k \hat{A}_k^T]$ and $C \triangleq E[\hat{A}_k A_k^T]$ which have scalar components of the form $E[\hat{a}_k \hat{a}_{k-i}]$ and $E[\hat{a}_k a_{k-i}]$, respectively. Then using a simple extension of the argument of the P_e calculation we have, e.g., $E[\hat{a}_k a_{k-1}] = \pi_0+\pi_1+\pi_4+\pi_5+\pi_{10}+\pi_{11}+\pi_{14}+\pi_{15}-\pi_2-\pi_3-\pi_6-\pi_7-\pi_8-\pi_9-\pi_{12}-\pi_{13}$ and $E[\hat{a}_k \hat{a}_{k-1}] = \pi_0+\pi_1+\pi_6+\pi_7+\pi_8+\pi_9+\pi_{14}+\pi_{15}-\pi_2-\pi_3-\pi_4-\pi_5-\pi_{10}-\pi_{11}-\pi_{12}-\pi_{13}$. The details are not important. However what is important is that the R and C matrices are determinable from P and π_s, i.e., the FSMP. We will see in section 5.2 that R and C are crucial parameters in the description of the *adaptive* stochastic dynamics of DFEs.

5 EQUILIBRIA AND AVERAGING ANALYSIS

5.1 Wiener−Hopf Solution

Our first task is to determine the locations of the

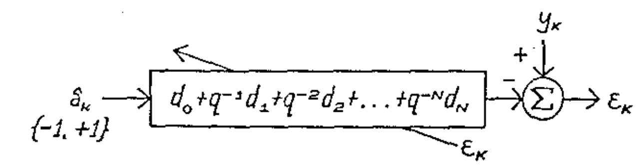

FIGURE 3: **DFE Adaptation Block Diagram — a standard LMS Indentification Problem**

attraction points (equilibrium points) in D_k−space for the weights when the algorithm seeks to minimize a least squares criteria as in the LMS algorithm (2.3a).

Consider Fig.3 which simply redraws a portion of Fig.1. (The definitions of the symbols in Fig.3 are identical to those in Fig.1.) This figure *suggests* that $\hat{a}_k$ can be interpreted as an input sequence and that y_k can be interpreted as a desired response. As is well known, the objective of the LMS algorithm (2.3a) is to minimize the mean square error defined by $\xi(D_k) = E[\varepsilon_k^2(D_k)]$ which is a quadratic (hence uni−modal) surface. The tap weight setting, D_{eq} (equilibrium), which gives the minimum mean square error is the classical Wiener−Hopf formula [15] and is given by

$$D_{eq} \triangleq E[\hat{A}_k\hat{A}_k^T]^{-1}\; E[\hat{A}_k y_k]\,;\; \det(E[\hat{A}_k\hat{A}_k^T]) \neq 0 \quad (5.1a)$$

$$= R^{-1}Q \quad \text{if } \det(R) \neq 0, \quad (5.1b)$$

where we have defined: (i) the covariance of $\hat{A}_k$ as $R \triangleq E[\hat{A}_k\hat{A}_k^T]$ (non−negative definite), and (ii) the cross−covariance vector $Q \triangleq E[\hat{A}_k y_k]$. Note, from this point on we consider only the case when R is non−singular. We will see that this restriction is justifiable when we come to section 6.3.

In reality, y_k is not a user supplied sequence but simply the channel output which in our case is identified with $y_k \triangleq A_k^T H$. Hence (5.1a) may be expanded to

$$D_{eq} = E[\hat{A}_k\hat{A}_k^T]^{-1}\; E[\hat{A}_k A_k^T]\; H \quad (5.2a)$$

$$= R^{-1}C\; H, \quad (5.2b)$$

defining the cross−covariance matrix $C \triangleq E[\hat{A}_k A_k^T]$ which is upper triangular by causality, and Toeplitz by stationarity. 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 perhaps the major problem of *blind* adaptation with DFEs, i.e., $\mathrm{sgn}(h_0)\,H$ (the desired tap weight setting) need not be a global attraction point for the adaptive algorithm.

Remark: If, during adaptation, we used a known training sequence, rather than blind equalization, then $\mathrm{sgn}(h_0)\{a_k\}$ in Fig.3, rather than $\{\hat{a}_k\}$, would be the tapped delay line input. With this change, (5.2a) degenerates to $D_{eq} = \mathrm{sgn}(h_0)\,H$, and there is no problem.

5.2 Averaged Equation Describing the Mean Trajectory

In this section we will derive an expression for the mean trajectory for the tap weights D_k as they head towards the unique equilibrium D_{eq} given by (5.2b) (assuming R and C are constant for the moment). The mean 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 gains γ.

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

$$D_{k+1} = (I - \gamma\, \hat{A}_k\hat{A}_k^T)\; D_k + \gamma\, \hat{A}_k A_k^T\; H. \quad (5.3)$$

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 (5.3) are bounded). Further, we might anticipate that the matrices $\hat{A}_k\hat{A}_k^T$ and $\hat{A}_k A_k^T$ 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, $E[D_k]$, takes the form,

$$E[D_{k+1}] = (I - \gamma E[\hat{A}_k\hat{A}_k^T])E[D_k] + \gamma E[\hat{A}_k A_k^T]H \quad (5.4a)$$

$$= (I - \gamma\, R)\; E[D_k] + \gamma\; C\; H, \quad (5.4b)$$

provided γ is sufficiently small. The formal justification that (5.4b) is the correct equation may be found in [20].

We make some observations regarding (5.4b). It is straightforward to verify that: (i) D_{eq} (5.2b) is indeed the equilibrium of the averaged (mean) equation (5.4b), and (ii) the mean equation (5.4b) is stable if and only if $\gamma\, \lambda_{max}(R) < 2$, where $\lambda_{max}(R)$ is the maximum eigenvalue of R. We use these averaging ideas in the next section to show $\{D_k\}$ can have very interesting behaviour.

6 TAP WEIGHT TRAJECTORIES

6.1 Piecewise Constant Behaviour

With this section we introduce the most important new ideas of this paper. We merge our previously (disconnected) results regarding the polytopes (section 3), finite state Markov processes (section 4), and the averaging analysis (section 5). 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.

The parameters which determine the dynamics of the mean equation (5.4b) are the covariance matrices R and C. We also met these matrices earlier in section 4 and we showed they could be evaluated with the assistance of FSMPs. Now, recall our one−to−one correspondence between the polytopes and FSMPs. Hence, in D_k−space the matrices R and C will be constant *only whilst* the tap setting D_k remains inside any one D_k−space polytope (assuming steady state). Therefore, we have shown (albeit informally)

Proposition 1: In steady state, the matrices R and C are piecewise constant functions of D_k, where the "pieces" are precisely the D_k−space polytopes defined by the hyperplanes (3.5).

Remarks: (i) To emphasize R, C and Q = CH are discrete functions of the polytope P for which $D_k \in P$, we write R(P), C(P) and Q(P).

(ii) Within each polytope the mean trajectory (5.4b) 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, deter−ministic difference equation (see the example in section 6.2)

The next property is an embellishment on proposition 1.

Proposition 2: The mean square error surface $\xi(D_k) \triangleq E[\varepsilon_k^2(D_k)]$ is a piecewise (polytope−wise) quadratic function of D_k given by

$$\xi(D_k \in P) = E[y_k^2] - 2Q(P)^T D_k + D_k^T R(P) D_k \quad (6.1a)$$

$$= H^T H - 2H^T C(P)^T D_k + D_k^T R(P) D_k. \quad (6.1b)$$

Proof: This is a trivial modification of a standard result in adaptive least squares, i.e., (6.1a) is [15, eqn (2.31)]. To obtain (6.1b), we substitute both $y_k = A_k^T H$ and $Q(P) = C(P)\, H$ into (6.1a), and invoke the whiteness of the input (A.4). Of course, we emphasize that for blind adaptation of DFEs, C(P) and R(P) are constant only within a polytope P, and generally these matrices vary from polytope to polytope. Therefore, the quadratic surface is different for different polytopes, i.e., piecewise quadratic.□

Remark: The local minimum of the mean square error $\xi(D_{eq}(P))$ associated with a polytope P, in terms of our fixed H, can be written $\xi_{min}(P) = H^T\,[I - C(P)^T R(P)^{-1} C(P)]\; H$. However, if $D_{eq}(P)$ lies outside P then this minimum is not attainable. We will see in section 6.3 that the global minimum mean square error is zero at $D^{opt} \triangleq \mathrm{sgn}(h_0)\,H$ for the polytope which contains $\mathrm{sgn}(h_0)\,H$, and is always attainable in the sense described in the following paragraph.

With each polytope P we have associated an equilibrium $D_{eq}(P) = R(P)^{-1}C(P)\; H$. To recapitulate, it is natural to

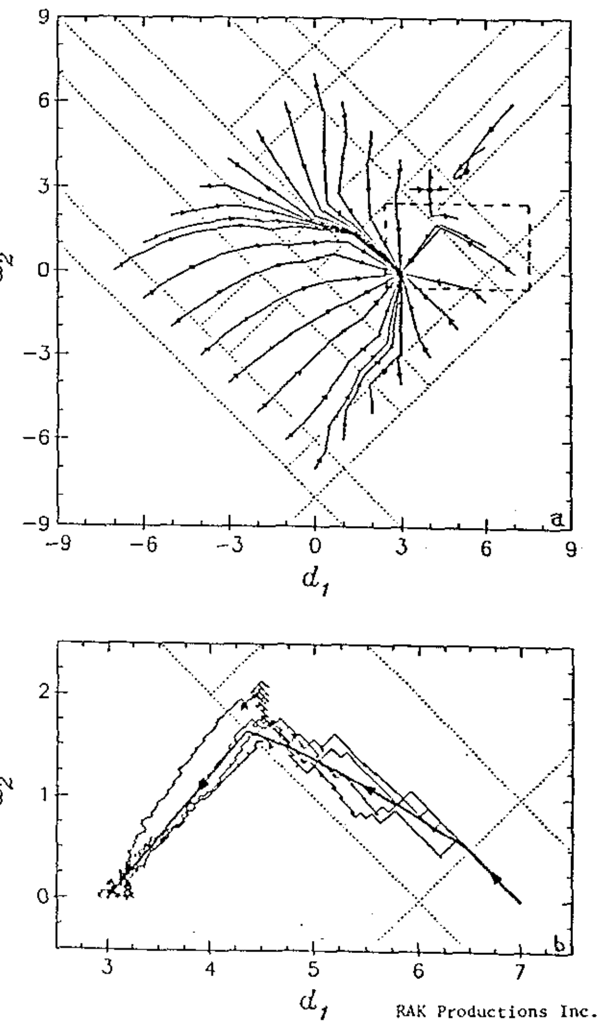

FIGURE 4: (a) Averaged trajectories describing adaptation for various starting points (all with $d_0(0)=0$) on the $N=2$ channel $h_0=1$, $\underline{H}=[4\ 3]^T$ with gain $\gamma=0.01$. Dotted lines show where there is a discontinuity in the mean square error surface in accordance with groups of polytopes having identical R(P) and C(P) matrices. (b) The averaged trajectory is shown as a heavy line with starting point $d_0=0$, $d_1=7$ and $d_3=0$. Four realizations cluster about the average converging on approximate equilibria $d_0=4$, $d_1=3$, $d_2=0$.

classify two types of equilibria.

Definition: $D_{eq}(P)$ is *attainable* if $D_{eq}(P) \in P$, otherwise it is *unattainable*.

If $D_{eq}(P)$ is attainable then $\{D_k\}$ will tend to move towards and settle down around it, (whenever $D_k \in P$). Otherwise, if $D_{eq}(P)$ is unattainable $\{D_k \in P\}$ will tend to move towards the boundary ∂P of P nearest to $D_{eq}(P)$ and thus head on into the adjacent polytope. The example in the next subsection best illustrates these ideas.

6.2 Example of Adaptation and the Averaged Trajectory

As in section 3.4 we choose the example $h_0 = 1$, $h_1 = 4$ and $h_2 = 3$. In Fig.4a we have plotted a large number of averaged trajectories according to (5.4b) with $\gamma = 0.01$, noting that the R and C matrices are now dependent on the polytopes (pictured in Fig.2). Note Fig.4a is the $[d_1\ d_2]^T$−space *projection* (2D) of D_k−space (3D), 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 evolution of d_0 during adaptation cannot be discerned in such a figure.

Figure 4b shows the precise sense in which to interpret Fig.4a. It shows an insert of Fig.4a with a single (bold) averaged trajectory and four realizations (i.e., simulations according to (5.3)) which appear to cluster about the averaged trajectory.

Note for this example there are only three attainable equilibria at $[1\ 4\ 3]^T$, $[4\ 3\ 0]^T$ and $[4.667\ 5.333\ 3.667]^T$. In Fig.4a the 2D projections of these equilibria are given by $[4\ 3]^T$, $[3\ 0]^T$ and $[5.333\ 3.667]^T$ and these appear as (local) attraction points for the mean trajectories.

6.3 Approximate and Exact Attainable Equilibria

The equilibria predicted by the theory developed are based only on mean behaviour (5.4b). However we note 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 $D_k = D^{opt} \triangleq \mathrm{sgn}(h_0)H$ (the proof is straightforward and omitted). So we say that *only* D^{opt} is an *exact equilibrium* in the sense that $\varepsilon_k(D^{opt}) = 0$. All other attainable equilibria are termed *approximate equilibria* and the physical manifestation of an attainable equilibrium with $\varepsilon_k(D_{eq}(P)) \neq 0$ is that the sequence $\{D_k\}$ is observed to jiggle about (but generally stay in the vicinity of) $D_{eq}(P)$. The questions arise: (i) regarded as a small noise can the variations in ε_k (estimating the gradient) drive $\{D_k\}$ away from $D_{eq}(P)$ such that $\{D_k\}$ hits ∂P (the boundary of P)?, and (ii) what is the expected time to do so? In the stochastic process literature [11,12] this is known as an *exit problem*. These are crucial questions because if the DFE hangs at an equilibrium $D_{eq}(P)$ corresponding to high error rates, e.g., $D_{eq} = [4.667\ 5.333\ 3.667]^T$ in Fig.4a, then an unacceptibly high exit time has serious practical consequences. These questions are the subject of current research.

6.4 Delay−type Equilibria and Their Attainability

In this section we consider those polytopes in D_k−space which yield a decision sequence which is a delay of the input (with a possible sign change) under steady state. We derive necessary (and conjecture sufficient) conditions for attainability of the attraction points of these polytopes for a broad class of adaptation algorithms in terms of the channel parameters.

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

$$\hat{a}_k = \mathrm{sgn}(h_\delta a_{k-\delta} + U_k(\delta) + V_k(\delta)),\quad 0 \le \delta \le N \quad (6.2a)$$

where

$$U_k(\delta) \triangleq \sigma(\delta) \sum_{i=\delta+1}^{N} d_{i-\delta}(a_{k-i} - \sigma(\delta)\hat{a}_{k+\delta-i}) \quad (6.2b)$$

$$V_k(\delta) \triangleq \sum_{i=0}^{\delta-1} h_i a_{k-i} + \sum_{i=\delta+1}^{N} (h_i - \sigma(\delta) d_{i-\delta}) a_{k-i} - \sum_{i=N-\delta+1}^{N} d_i \hat{a}_{k-i} \quad (6.2c)$$

We also define an upper bound on (6.2c) as

$$V_{max}(\delta) \triangleq \sum_{i=0}^{\delta-1} \left|h_i\right| + \sum_{i=\delta+1}^{N} \left|h_i - \sigma(\delta) d_{i-\delta}\right| + \sum_{i=N-\delta+1}^{N} \left|d_i\right|. \quad (6.3)$$

The reason for the curious decomposition given by (6.2a) will become clearer later. We will see that δ corresponds to a nominal time delay and $\sigma(\delta) \triangleq \mathrm{sgn}(h_\delta)$, corresponds to an associated sign change of the channel/DFE combination.

We define two proper subsets of the set of atomic (Markov) states Ω (section 4.2) parametrized by $0 \le \delta < N$. (The case $\delta = N$ needs to be treated separately but fortunately is easily disposed of.) Define,

$$\omega_+(\delta) \triangleq \{X_k \in \Omega \,|\, \hat{a}_{k-i} = +\sigma(\delta)\, a_{k-\delta-i},\ i = 1,\dots,N-\delta\}$$

$$\omega_-(\delta) \triangleq \{X_k \in \Omega \,|\, \hat{a}_{k-i} = -\sigma(\delta)\, a_{k-\delta-i},\ i = 1,\dots,N-\delta\}$$

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

Remark: 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 $i = 1, \dots, N-\delta$, and thus these subsets are in effect independent of k (as the notation suggests).

The notion of a closed subset of Ω will considerably simplify development.

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

The following statements are equivalent: (i) Suppose $X_k \in \omega_{\pm}(\delta)$, then all future ($m \geq k$) decisions are of the form $\hat{a}_m = \pm\sigma(\delta)\, a_{m-\delta}$. (ii) $\omega_{\pm}(\delta)$ is closed. Hence to investigate channel/DFE combinations yielding simple time delay behaviour we need only to determine when $\omega_{\pm}(\delta)$ are closed and reachable from within Ω. Our first proposition narrows our investigations by showing a DFE can never behave consistently in the form $\hat{a}_m = -\sigma(\delta)\, a_{m-\delta}$ 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

$$\left|h_\delta\right| > V_{max}(\delta). \qquad (6.4)$$

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

Proof: Suppose $\omega_{\pm}(\delta)$ is closed and that the system has been in $\omega_{\pm}(\delta)$ for some time. Then $\hat{a}_{k-i} = \pm\sigma(\delta)\, a_{k-\delta-i}$, $i = 1,2,\ldots,N$. Substituting into (6.2b) and (6.2c) we obtain

$$U_k^{-}(\delta) = 2\,\sigma(\delta) \sum_{i=\delta+1}^{N} d_{i-\delta}\, a_{k-i}; \qquad U_k^{+}(\delta) = 0, \quad (6.5a)$$

and

$$V_k^{\pm}(\delta) \triangleq \sum_{i=0}^{\delta-1} h_i a_{k-i} + \sum_{i=\delta+1}^{N} (h_i - \sigma(\delta) d_{i-\delta}) a_{k-i} \mp \sigma(\delta) \sum_{i=N+1}^{N+\delta} d_{i-\delta} a_{k-i} \quad (6.5b)$$

Consider first $\omega_{+}(\delta)$. If $\left|h_\delta\right| > V_{max}(\delta)$ with $U_k^{+}(\delta) = 0$ then this implies $\left|h_\delta\right| > \left|V_k^{+}(\delta)\right|$ for all $V_k^{+}(\delta)$, hence $\hat{a}_k = +\sigma(\delta)\, a_{k-\delta}$ by (6.2a). Then note that the a_k in (6.5b) are distinct and therefore the supremum of $V_k^{+}(\delta)$ over $\{a_k\}$ is just $V_{max}(\delta)$, so (6.4) is also necessary. Now consider $\omega_{-}(\delta)$. In this case we have $\hat{a}_k = \mathrm{sgn}(h_\delta a_{k-\delta} + U_k^{-}(\delta) + V_k^{-}(\delta))$ where by (6.5a,b) $U_k^{-}(\delta) + V_k^{-}(\delta)$ is independent of $a_{k-\delta}$. Hence $\Pr(\hat{a}_k = \sigma(\delta)\, a_{k-\delta}) \geq \frac{1}{2}$, contradicting closure. □

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. We formulate this as:

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

The inequality (6.4) can only hold for at most one value of δ. To prove this one assumes at least two inequalities of the form (6.4) are simultaneously satisfied (say for δ_1 and δ_2) then an application of the triangle inequality establishes a contradiction. The details of the proof are omitted. 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 outside $\omega_{+}(\delta)$. A full answer to this question is not yet known. We present the following result (without proof) and important conjecture.

Proposition 6: Let h_δ, $0 \leq \delta < N$ satisfy (6.4). Then the following alternative conditions are sufficient to guarantee that there exists an input sequence such that $N-\delta$ consecutive $\hat{a}_m = +\sigma(\delta)\, a_{m-\delta}$ decisions are made:

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

(ii) $\mathrm{sgn}(d_1) = \mathrm{sgn}(d_2) = \ldots = \mathrm{sgn}(d_{N-\delta})$.

Conjecture 7: Let h_δ, $0 \leq \delta < N$ satisfy (6.4). Then there exists an input sequence such that $N-\delta$ consecutive decisions are made of the form $\hat{a}_m = +\sigma(\delta)\, a_{m-\delta}$.

Remarks: (i) With (6.4) satisfied, and the hypothesis of Proposition 6 fulfilled, $\omega_{+}(\delta)$ is closed, indecomposable and reachable, so that $\Pr(X_k \in \omega_{+}(\delta)) \to 1$ as $k \to \infty$. Hence under stationarity the channel/DFE combination produces decisions of the form $\hat{a}_m = +\sigma(\delta)\, a_{m-\delta}$ if and only if $\left|h_\delta\right| > V_{max}(\delta)$ *with* $\omega_{+}(\delta)$ reachable. Note that the output is clearly white under such conditions.

(ii) Given a time-invariant channel H, define the following region of D_k-space (which is just another way of writing (6.4)).

$$\Xi(\delta) \triangleq \{D_k \in R^{N+1} \,|\, \rho(\delta) > \sum_{i=\delta+1}^{N} \left|h_i - \sigma(\delta) d_{i-\delta}\right| + \sum_{i=N-\delta+1}^{N} \left|d_i\right|\} \quad (6.6a)$$

where

$$\rho(\delta) \triangleq \left|h_\delta\right| - \sum_{i=0}^{\delta-1} \left|h_i\right| \qquad 0 \leq \delta \leq N \qquad (6.6b)$$

Then $\hat{a}_m = +\sigma(\delta)\, a_{m-\delta}$ under steady state conditions only if $D_k \in \Xi(\delta)$ and sometimes if (according to the reachability of $\omega_{+}(\delta)$). Note this region may also be written $\rho(\delta) > \|\underline{D}_k - \sigma(\delta) S^{\delta}(N)\underline{H}\|_1$, where S(i) denotes an $i \times i$ matrix of super-diagonal ones. Hence the projection of $\Xi(\delta)$ onto $\underline{D}_k$-space defines an l_1-ball with centre $\sigma(\delta) S^{\delta}(N)\underline{H}$ and radius $\rho(\delta)$. (Note the d_0 component of D_k does not play a role in the constraint in (6.6a).) Region $\Xi(\delta)$ is non-empty only if $\rho(\delta) > 0$. In Fig.4a, $\Xi(0)$ and $\Xi(1)$ are non-empty.

(iii) $\rho(0) = \left|h_0\right| > 0$ by (A.2) hence there is a (non-empty) collection of polytopes in D_k-space giving $\hat{a}_m = +\sigma(0)\, a_m$ (correct decisions), under steady state.

(iv) In the case $\delta = N$, it is readily apparent that condition (6.4) is necessary and sufficient for *every* decision to be of the form $\hat{a}_m = \sigma(N)\, a_{k-N}$.

Now on to the main result.

Theorem 8: A necessary condition for the LMS adaptive algorithm to have an *attainable* equilibrium corresponding to the channel/DFE combination producing decisions of the form $\hat{a}_m = +\sigma(\delta)\, a_{m-\delta}$ under steady state is $\rho(\delta) > 0$. Further, this equilibrium is given by $D_{eq}(P) \triangleq \sigma(\delta) S^{\delta}(N+1) H$. The condition is also sufficient when $\omega_{+}(\delta)$ is reachable.

Proof: First, we require $\Xi(\delta)$ to be non-empty thus $\rho(\delta) > 0$. Now let $D_k \in \Xi(\delta)$ and let steady state conditions prevail. Hence all decisions are of the form $a_m = +\sigma(\delta)\, a_{m-\delta}$ implying $\{\hat{a}_k\}$ is white. Therefore the output covariance is simply $R(P) = I$. Further, the cross-covariance is given by $C(P) = +\sigma(\delta) S^{\delta}(N+1)$ by elementary considerations. Hence for all polytopes belonging to $\Xi(\delta)$ the equilibrium (5.2b) in D_k-space is given by $D_{eq}(P) \triangleq R(P)^{-1} C(P) H = \sigma(\delta) S^{\delta}(N+1) H$. Then projecting into $\underline{D}_k$-space we deduce $\|\underline{D}_{eq}(P) - \sigma(\delta) S^{\delta}(N)\underline{H}\|_1 = 0$, i.e., $\rho(\delta) > 0$, showing $D_{eq}(P) \in \Xi(\delta)$. Hence there exists a polytope P inside $\Xi(\delta)$ with attainable equilibrium $D_{eq}(P)$. □

Remarks: (i) For $\delta = 0$, $D_{eq}(P) = D^{opt} \triangleq +\sigma(0) H$ is always attainable and achieves the global minimum mean square error of zero (recall it is the only exact equilibria).

(ii) The LMS qualifier in the theorem statement is superfluous. The same result holds for *any* adaptation algorithm which seeks to minimize the (quadratic) mean square error (6.1a,b) by searching for the Wiener-Hopf solution. Hence the result holds also for algorithms based on NLMS, RLS, etc.

6.5 White Equilibria

As we have earlier commented, when $\hat{a}_k = \sigma(\delta) a_{k-\delta}$, the $\{\hat{a}_k\}$ process is white. Let us term any equilibrium with the $\{\hat{a}_k\}$ process white as a white equilibrium. In this sub-section, we shall present further results on this class.

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 $\{\hat{a}_k\}$ is white, (under steady state). Then

$$\| H \|_1 > \| \underline{D}_k \|_1. \qquad (6.7)$$

Proof: If $\{\hat{a}_k\}$ is white then the subsequence $\{\hat{a}_k = -1;\ \hat{a}_{k-i} = -\mathrm{sgn}(d_i),\ i = 1, 2, \ldots, N\}$ occurs with non-zero probability (for some input sequence). In (2.1) this implies $-1 = \mathrm{sgn}(A_k^T H + \|\underline{D}_k\|_1)$ and (6.7) follows.□

Remarks: (i) $\|H\|_1$ is the peak excursion of the channel output when driven by a white binary input. Hence we can estimate $\|H\|_1$ by channel output measurements and thus impose during adaptation that $\{\underline{D}_k\}$ does not leave (6.7).

(ii) Closely related to the above is the following. Clearly, by the definition of D^{opt}, we have $\|D^{opt}\|_1 = \|H\|_1$ (the latter of which can be adaptively estimated). It is obvious, yet has not been suggested in the literature, that adaptation algorithms should constrain $\{D_k\}$ only to move on the l_1-ball given by $\|D_k\|_1 = \|H\|_1$.

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

Proof: If $\det(R(P)) = 0$ this implies there exists $x \triangleq [x_0\ x_1\ \ldots\ x_N]^T \neq 0$ such that $x^T R(P) x = 0$, i.e., $E[(x^T \hat{A}_k)^2] = 0$. Therefore, under steady state, we have almost surely

$$x_0 \hat{a}_k + x_1 \hat{a}_{k-1} + x_2 \hat{a}_{k-2} + \ldots + x_N \hat{a}_{k-N} = 0 \quad (6.8)$$

where at least two x_i's are non-zero.

Now consider $\|\underline{D}_k\|_1 < \|H\|_1$. Using (6.8) we may write (2.2) in the form $\hat{a}_k = \mathrm{sgn}(A_k^T H - \underline{\hat{A}}_{k-N}^T \underline{D}_k^{\#})$ where $\underline{D}_k^{\#}$ is a function of x and $\underline{D}_k$. This shows selection of A_k can be made independent of $\underline{\hat{A}}_k$. Now consider two input subsequences $\{a_{k-i} = +\mathrm{sgn}(h_i)\}$ and $\{a_{k-i} = -\mathrm{sgn}(h_i)\}$. Then in the two cases $A_k^T H = \pm\|H\|_1$ which implies (by hypothesis) that $\hat{a}_k = \pm 1$, respectively. However, this contradicts (6.8) which says that given $\underline{\hat{A}}_k$ then $\hat{a}_k$ is uniquely determined. Therefore $\|\underline{D}_k\|_1 > \|H\|_1$ as claimed. □

Remark: This justifies the earlier restriction that we should only consider polytopes P satisfying $\det(R(P)) \neq 0$, 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. There is some evidence that the same region implies the stationary atomic distribution π_S is unique, e.g., when $h_i > 0$ for all i.

One key property of white equilibria is that it is in practice possible to test for whiteness. In principle, one should verify that $\Pr(\hat{a}_{\alpha(1)}, \hat{a}_{\alpha(2)}, \ldots, \hat{a}_{\alpha(m)}) = 2^{-m}$ for all selections $\alpha(1), \alpha(2), \ldots, \alpha(m)$ of different indices. However, given certain hypotheses, testing is far easier, as indicated by the following proposition (proof is omitted).

Proposition 11: Given (i) Zero mean: $\Pr(\hat{a}_k = +1) = \frac{1}{2}$, (ii) Zero Correlation: $E(\hat{a}_k \hat{a}_{k-i}) = 0$ for all $i \neq 0$, (iii) Sign symmetry: $\Pr(\hat{a}_{\alpha(1)}, \hat{a}_{\alpha(2)}, \ldots, \hat{a}_{\alpha(m)}) = \Pr(-\hat{a}_{\alpha(1)}, -\hat{a}_{\alpha(2)}, \ldots, -\hat{a}_{\alpha(m)})$, and (iv) No parity bias: the probability that for any sample of size 2m $\{\hat{a}_i\}$ having an even number of +1's equals the probability that the same sample has an odd number of −1's, then the $\{\hat{a}_k\}$ process is white.

As already noted, delay equilibria are white. We suspect the reverse implication is also true, and formalize the statement as:

Conjecture 12: If $\{\hat{a}_k\}$ is white then $\hat{a}_k = +\sigma(\delta)\, a_{k-\delta}$ for precisely one $0 \leq \delta \leq N$.

If the conjecture held we would have a way of statistically testing the output of a DFE to prove it was inverting the channel with suitable delay. Note also for a very large class of practical channels $V_{max}(\delta) > |h_\delta|$ for all $\delta > 0$. Thus the only possibility in this case is to have $\delta = 0$ which is a highly desirable situation.

7 CONCLUSIONS

In this paper, we have described some analytical tools which should be of assistance in dealing with the fundamental question: for what classes of channel impulse responses if DFE use a practical proposition? We are not yet in a position to answer this question, and we have indicated several conjectures bearing on the use of these tools. Let us note here two directions in which considerably more work may need to be done to move in the direction of a definitive answer. First, the techniques of large deviation analysis [10,11] need to be brought to bear to consider how long taps are likely to remain in the neighbourhood of an incorrect or inappropriate minimum. Second, it would be highly desirable to find some systematic way of approximating a Markov process with a much smaller number of states (and we note here the work of [21]), or by a Markov process with continuous state space for which analytic calculation could be more easily executed (there is really no evidence to suggest this would be possible for the DFE problem, but the possibility cannot be excluded *a priori*).

REFERENCES:

[1] R.A. Kennedy, and B.D.O. Anderson, 'Recovery Times of Decision Feddback Equalizers on Noiseless Channels', IEEE Trans. Commun., (to appear).

[2] R.A. Kennedy, and B.D.O. Anderson, 'Error Recovery of Decision Feedback Equalizers on Exponential Impulse Response Channels', IEEE Trans. Commun., (August 1987).

[3] J.J. O'Reilly and A.M. de Oliveira Duarte, 'Error propagation in decision feedback receivers', IEE Proc. F, Comm., Radar and Sig. Proc., 1985, 132, (7), pp.561−566.

[4] A.M. de Oliveira Duarte and J.J. O'Reilly, 'Simplified technique for bounding error statistics for DFB receivers', IEE Proc. F, Comm., Radar and Sig. Proc., 1985, 132, (7), pp.567−575.

[5] J.E. Mazo, 'Analysis of Decision−directed Equalizer Convergence' Bell Syst. Tech. J., Vol.59, No.10, pp.1857−1876, December 1980.

[6] O. Macchi, and E. Eweda, 'Convergence Analysis of Self Adaptive Equalizers' IEEE Trans. Inform. Theory, vol.IT−30, No.2. pp.161−176, March 1984.

[7] R. Kumar, 'Convergence of a Decision−directed Adaptive Equalizer' IEEE Conf. on Decision and Control, 1983.

[8] 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. Auto. Cont., AC−25, pp.385−399, June 1980.

[9] A. Benveniste, and M. Goursat, 'Blind Equalizers' IEEE Trans. Commun., vol.COM−32, No.8. pp.871−883, August 1984.

[10] M. Cottrell, J.C. Fort, and G. Malgouyres, 'Large Deviations and Rare Events in the Study of Stochastic Algorithms', IEEE Trans. Automat. Contr., vol.AC−28, pp.907−920, September 1983.

[11] M.I. Freidlin and A.D. Wentzell, 'Random Perturbations of Dynamical Systems' Springer−Verlag New York, Inc. 1984.

[12] J. Zabczyk, 'Exit Problem and Control Theory', Systems and Control Letters, 6, pp.165−172, 1985.

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

[14] D.L. Duttweiler, J.E. Mazo, and D.G. Messerschmitt, 'An Upper Bound on the Error Probability in Decision Feedback Equalization', IEEE Trans. Inform. Theory, vol.IT−20, pp.490−497, July 1974.

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

[16] P.V. Kabaila (private communication).

[17] R.A. Kennedy, B.D.O. Anderson, and R.R. Bitmead, 'Tight Bounds on the Error Probabilities of Decision Feedback Equalizers', IEEE Trans. Commun., (to appear).

[18] A. Cantoni, and P. Butler, 'Stability of Decision Feedback Inverses', IEEE Trans. Commun., vol.COM−24, No.9, pp.1064−1075, September 1976.

[19] R.A. Kennedy, B.D.O. Anderson, 'Stochastic Analysis of Non−Adaptive Decision Feedback Equalizers', (this conference).

[20] R.Z. Khas'minskii, 'Stochastic Stability of Differential Equations', Sijthoff & Noordhoff, Alphen aan den Rijn 1980.

[21] J.B. Rohlicek, 'Aggregation and Time Scale Analysis of Perturbed Markov Systems' M.I.T. PhD Dissertation LIDS−TH−1641, January 1987.