INTERNATIONAL JOURNAL OF ADAPTIVE CONTROL AND SIGNAL PROCESSING, VOL. 7, 497–523 (1993)

BLIND ADAPTATION OF DECISION FEEDBACK EQUALIZERS: GROSS CONVERGENCE PROPERTIES*

RODNEY A. KENNEDY

Department of Engineering, Faculty of Engineering, Australian National University, Canberra ACT 0200, Australia

AND

BRIAN D. O. ANDERSON AND ROBERT R. BITMEAD

Department of Systems Engineering, Research School of Physical Sciences and Engineering, Australian National University, Canberra ACT 0200, Australia

SUMMARY

An analysis of the stochastic dynamics of the blind adaptation of decision feedback equalizers is presented. The analysis accounts for the presence of decision errors which, under feedback, are propagated. A number of blind algorithms are presented and a theory is developed to explain gross convergence properties observed through simulations. The possibility of and mechanism behind undesirable local minima are highlighted and a detailed case study is given. The potential capture by local minima shows the importance of good initialization. These results superficially resemble those obtained for blind adaptation applied to linear equalizers.

KEY WORDS Blind adaptive equalization Blind decision feedback equalization

1. INTRODUCTION

1.1. Background

Adaptation is employed in tuning an equalizer 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 u_k and output y_k measurements to identify the channel (see Figure 1). Of course, if we had complete knowledge of the real data u_k available at the receiver, then this would defeat the purpose of equalization. Thus what is standard in practice is to send a known *training sequence* $\{u_k\}$ for a limited time duration during which the channel parameters may be reliably learnt. After training with the equalizer correctly tuned, unknown data are sent and recovered at the equalizer output $\hat{u}_k$ 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 u_k,

* A condensed version of this paper was presented at the IFAC Symposium on Adaptive Systems in Control and Signal Processing, Glasgow, U.K., April 1989.

This paper was recommended for publication by editor M. J. Grimble

0890–6327/93/060497–27$18.50

Received 24 February 1993
Revised 10 June 1993

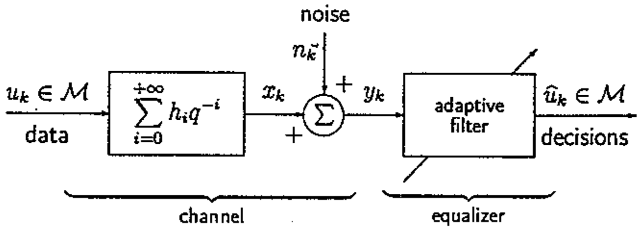

Figure 1. Linear channel and equalizer model

the data estimates $\hat{u}_k$ are used in the adaptive algorithm in place of u_k and this is the so-called *decision-directed equalizer* (DDE).[2,3]

Blind adaptation, according to our definition, represents something stronger.[4] As in the post-training phase of standard adaptation, (some function of) the data estimates $\hat{u}_k$ are used in lieu of the real data u_k, but never is a training sequence used. Thus channel (inverse) identification is based only on signals y_k and/or $\hat{u}_k$. Therefore blind adaptation concerns *global convergence* issues when the explicit knowledge of the input data is unavailable and we have unreliable data estimates because the equalizer need not be tuned (equivalently when we have arbitrary initializations). There are a number of situations where blind adaptation is important in practice, such as during a break in multipoint communications.[5]

The simplest equalizer structure that we could use on the channel in Figure 2 is the *linear equalizer* in its various guises. Typically it consists of an FIR filter cascaded with the channel[6] and usually incorporates a decision device at its output. Another variation uses an IIR filter, in which case it is called a recursive (linear) equalizer.[7] The theory analysing the blind adaptaton for linear equalizers is well developed, though incomplete, and contains many interesting results, including those above, to which we shall have cause to refer (see also References 8 and 9). However, this work deals with another important but non-linear equalizer which we now introduce.

The decision feedback equalizer (DFE) is a *non-linear* recursive equalizer which utilizes past outputs $\hat{u}_{k-1}, \hat{u}_{k-2}, \ldots$ in its filter structure[1] (see Figure 2). Such equalizers find their place in applications because of the limitations of linear equalization. For example, when the channel has zeros on the unit circle it is fundamentally impossible to find a stable linear approximation to the inverse of the linear channel.[4] Also, with zeros close to the unit circle there is the related noise enhancement problem.[1] Fewer restrictions apply to the DFE structure, meaning DFEs can be effectively used on a channel with a transfer function such as $1 + 2q^{-1} + q^{-2}$, although, because of their recursive structure, they may suffer from excessive error propagation if the channel is too pathological.[10]

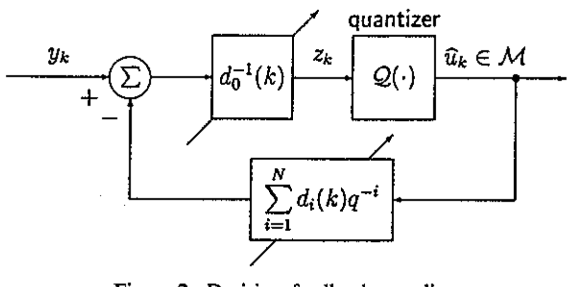

Figure 2. Decision feedback equalizer

1.2. Blind equalization literature

1.2.1. Linear equalization. In the linear case the traditional approach to blind equalization has been to use stochastic gradient algorithms based on generalized non-convex cost functions, beginning with Sato.[11] However, only under restrictive assumptions, e.g. when one assumes that the input distribution is sub-Gaussian (being a family of continuous distributions with the uniform distribution as a limiting case), are global convergence results available, i.e. can it be shown that the only stable points of certain gradient descent algorithms correspond to the channel–equalizer combination acting as a pure delay.[4] Recently, these results have been shown to be strictly valid only when the equalizer is doubly infinite in its parametrization.[12,13] For example, simple channels have been constructed showing ill-convergence of the popular algorithms developed in References 4, 5 and 11 whenever the equalizer is finite-dimensional and the input u_k takes either discrete or continuous values. As a consequence, the recent global convergence proofs in References 14 and 15 which depend critically on doubly infinite equalizer parametrizations are largely idealistic and need careful interpretation.

Thus the value of investigating blind adaptation using detailed case studies for simple channels where closed-form analysis is possible has been established in the linear case. Our work here, which actually predates the negative results for the linear case, makes a similar contribution for the DFE but requires different and new tools.[16,17]

1.2.2. Decision feedback equalization. The literature is very thin on the subject of blind adaptation of DFEs[18] in both algorithm development and subsequent analysis, the principal reason being the difficulty of incorporating error propagation effects[10,19] into the analysis of adaptation. Naturally an analysis of DFE blind adaptation which excludes the effects of decision errors is not well-posed, because decision errors and error propagation will be present given lack of tuning. In this situation the effects of errors will be to distort adaptation relative to the training sequence case (where there are no errors). We will draw on the error propagation theory developed in Reference 19 to provide a basis to quantify such a distortion. A second major barrier to obtaining results on the blind adaptation of DFEs is the non-linearity of the structure in Figure 2, meaning that the elegant results in Reference 4 and those in Reference 9, which rely on linearity, cannot be applied nor easily modified.

1.3. Aims of study

An 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 rather than the actual input data govern the dynamics of adaptation, the general task is to characterize the attraction points of the algorithm carefully. By developing a suitable theoretical framework and constructing revealing examples, we will see that, as in the linear case, ill-convergence of the blind DFE algorithms can result.

Explicitly, the general aims of this work are

(i) to compare theoretically the blind adaptation of linear equalizers with the blind adaptation of DFEs, with emphasis on the latter (the points of contact and departure between these two types of equalizer will be examined)
(ii) to understand the effects of error propagation on blind DFE adaptation
(iii) to show that a well-dimensioned blind DFE adapted according to a Sato-like algorithm exhibits ill-convergence to undesirable local minima

(iv) to find necessary and sufficient conditions on the channel parameters to ensure the existence of desirable minima leading to the eye diagram being opened.

2. SYSTEM DESCRIPTION

2.1. Channel and equalizer model

The channel shown in Figure 1 will be modelled by the impulse response $\{h_0, h_1, \ldots\}$ driven by real M-level data

$$u_k \in \mathcal{M} \triangleq \{1-M, 3-M, \ldots, M-1\}$$

with M even and where k denotes the discrete time index. Complex data as in quadrature amplitude modulation (QAM) and complex-valued channels could also be considered using similar techniques. Additive zero-mean channel noise n_k is also depicted in Figure 1. However, mostly we will regard its influence as secondary to simplify the analysis, as is standard in the blind adaptation literature (although the effect of noise will be briefly considered).

The DFE structure shown in Figure 2 consists of an N-tap delay line represented by weights $\{d_1(k), d_2(k), \ldots, d_N(K)\}$ ideally tuned to minimize the (residual) intersymbol interference (ISI). This tapped delay line is fed by past decisions or data estimates $\hat{u}_k$; even if the d_i are correctly adjusted, this can lead to problems when past decisions are incorrect, and this is the error propagation mechanism.[19] Note that in Figure 2 the additional weight $d_0(k)$ is incorporated to provide gain adjustment especially when $M > 2$.

A more typical and indeed more general DFE structure usually consists of an FIR filter followed by the structure given in Figure 2. We have three reasons for considering the simpler structure. Firstly, our application demand is for subscriber–local exchange twisted pair telephone lines whose measurements have proved to possess little precursor intersymbol interference, thus obviating the need for the FIR cascade.[20] Secondly, one of our aims is to understand the effects of error propagation on adaptation (relative to the well-understood training sequence case) and Figure 2 is the minimal non-trivial structure to study these effects unambiguously. Further, in the more general structure it is usually possible to separate the adaptation of the FIR section from the DFE section. Then, with a preliminary FIR adaptation completed, the cascade of the channel and the FIR section consists of a linear system describable by the impulse response $\{h_0, h_1, \ldots\}$, forming an effectively new channel.

The fundamental M-ary DFE output equation from Figures 1 and 2 is given by

$$\begin{aligned} \hat{u}_k &= \mathcal{Q}\left(d_0^{-1}(k)\left(y_k - \sum_{i=1}^{N} d_i(k)\hat{u}_{k-i}\right)\right) \\ &= \mathcal{Q}\left(d_0^{-1}(k)\left(\sum_{i=0}^{\infty} h_i u_{k-i} - \sum_{i=1}^{N} d_i(k)\hat{u}_{k-i} + n_k\right)\right) \end{aligned} \tag{1}$$

where $\mathcal{Q}(\cdot)$ is a nearest-neighbour M-ary quantizer which is conveniently defined by

$$\mathcal{Q}(x) \triangleq \sum_{k=1-M/2}^{M/2-1} \operatorname{sgn}(x+2k) \tag{2}$$

with $\operatorname{sgn}(x) \triangleq +1$ for $x > 0$, $\operatorname{sgn}(x) \triangleq -1$ for $x < 0$ and $\operatorname{sgn}(0) = 0$.

Comments

(i) The size of N is chosen sufficiently large so as to model adequately the majority of the

ISI present in the channel in the sense that $\Sigma_{i=N+1}^{\infty} | h_i |$ needs to be sufficiently small. This ensures that the problem is well-posed.

(ii) We refer to a decision of the form $\hat{u}_k = \text{sgn}(h_\delta) u_{k-\delta}$ as *correct* with the agreed convention that δ is some well-defined and fixed delay. The conditions under which all this makes sense will be given later. Note that this delay should not be confused with a bulk delay representing gross propagation effects and the like, which (without loss of generality and as is standard) has been implicitly removed.

(iii) Throughout this paper we disallow the argument of any sgn(·) function in (2) to be precisely zero to simplify the discussion.

2.2. *Algorithm development*

In Reference 18 a simple blind algorithm for adjusting a binary DFE was proposed. This algorithm actually corresponds to running the adaptive DFE in a *decision-directed* mode, where data decisions $\hat{u}_k$ are used in lieu of the real data u_k in a conventional recursive gradient scheme with quadratic cost. This algorithm can also be thought of as a translation of the Sato *blind* algorithm[11] originally developed for the linear equalizer. (When the data are binary ($M = 2$), there is essentially no need to distinguish between the Sato blind algorithm and the decision-directed algorithm, as will be clearer later.)

We now define two classes of algorithms: (i) decision-directed algorithms; (ii) constant modulus algorithms.[21] Both these classes include the algorithm in Reference 18 as a special case.

Let

$$D \triangleq (d_0, d_1, \ldots, d_N)^{\mathrm{T}} \in \mathbb{R}^{N+1}$$

denote the vector of DFE parameters. Then the gradient descent algorithm that we consider is given by

$$D(k+1) = D(k) - \mu \left. \frac{\partial J(z_k)}{\partial D} \right|_{D=D(k)} \tag{3}$$

where μ is the step size, z_k is the DFE quantizer input (Figure 2) and $J(\cdot)$ is some memoryless cost function to be minimized. This same approach works for an arbitrary equalizer structure, not just for the DFE or linear equalizer, where D represents the adjusted parameters. The signal z_k in all cases is to be thought of as the signal to be ideally driven to u_k (or more generally $u_{k-\delta}$ or $-u_{k-\delta}$, where δ is some fixed delay). Since $J(\cdot)$ operates on z_k which is available at the receiver, the algorithm satisfies the conditions of a blind algorithm.

2.2.1. *Decision-directed algorithms.* If we define the memoryless cost function

$$J(z_k) \triangleq \tfrac{1}{2} d_0^2(k)(z_k - \hat{u}_k)^2, \qquad \hat{u}_k = \mathcal{Q}(z_k) \tag{4}$$

then (3) gives after a little calculation,

$$D(k+1) = D(k) + \mu \varepsilon_k R_k \tag{5}$$

where

$$R_k \triangleq (\hat{u}_k, \hat{u}_{k-1}, \ldots, \hat{u}_{k-N})^{\mathrm{T}} \tag{6}$$

is the regressor, noting (see Figure 2)

$$z_k = d_0^{-1}(k)\left(y_k - \sum_{i=1}^{N} d_i(k) \hat{u}_{k-i}\right)$$

and the error is given by

$$\varepsilon_k \triangleq d_0(k)(z_k - \mathcal{Q}(z_k))$$
$$= y_k - \sum_{i=0}^{N} d_i(k)\hat{u}_{k-i}$$
$$= \sum_{i=0}^{N} (h_i u_{k-i} - d_i(k)\hat{u}_{k-i}) = \underbrace{\sum_{i=N+1}^{N} h_i u_{k-i} + n_k}_{Q_k} \tag{7}$$

Note that the term $d_0^2(k)$ in (4) is important to give the algorithm regressor in the form (5) and to achieve the simple and desirable expression for the error (7). (However, in the context of the next class of algorithms this term is better omitted.) As is depicted in Figure 3(a), this error (7) may be formed by scaling the difference between the input and output of the decision quantizer. Also note that the substitution of $\hat{u}_j$ by u_j throughout (4)–(7) leads to the conventional adaptation rule corresponding to the use of a training sequence. We have been able to imbed the decision-directed philosophy into a gradient descent framework by determining the appropriate cost (4). Finally, we note that the term Q_k in (7) represents unmodelled channel parameters and noise.

2.2.2. Constant modulus algorithms. A binary alphabet is an elementary example of a constant modulus signal, meaning that an adaptation employed to drive the signal z_k to a constant modulus is a sensible objective. In the linear equalizer case it can be concluded that this is a sufficient condition to guarantee the recovery of binary input data at the point z_k (perhaps with delay and sign inversion).[4] However, even in the M-ary case such an objective has been shown to be reasonably effective.[11,14] Such constant modulus ideas generalize naturally to complex signals, particularly phase-shift-keyed (PSK) alphabets.[15,21]

In analogy with the linear case,[21,22] define the set of cost functions indexed by parameter p and constant γ_p, i.e.

$$J(z_k) \triangleq \frac{1}{2p}(|z_k|^p - \gamma_p)^2 \tag{8}$$

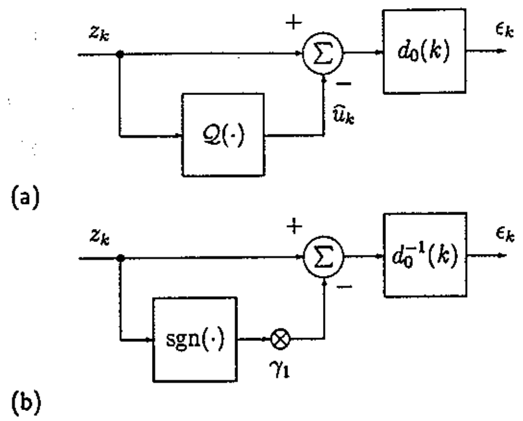

Figure 3. Error signals: (a) decision-directed; (b) Sato ($p = 1$)

Then (3) becomes

$$D(k+1) = D(k) + \mu \varepsilon_k R_k \tag{9}$$

where

$$R_k \triangleq (z_k, \hat{u}_{k-1}, \hat{u}_{k-2}, \ldots, \hat{u}_{k-N})^{\mathrm{T}} \tag{10}$$

is the regressor, which differs in the first component from R_k in (5), and

$$\varepsilon_k = d_0^{-1}(k)\ \mathrm{sgn}(z_k) | z_k |^{p-1} (| z_k |^p - \gamma_p) \tag{11}$$

is the error. When the index $p = 1$, we have an algorithm for the DFE which is analogous to the blind linear algorithm of Sato,[11] leading to a simpler error

$$\varepsilon_k = d_0^{-1}(k)(z_k - \gamma_1\ \mathrm{sgn}(z_k)) \quad (p = 1)$$

This Sato error signal is shown in Figure 3(b).

The constant γ_p in (8) is selected to ensure that the desired DFE parameter vector corresponding to correct equalization is a stationary point of the algorithm (the idea is the same as in the linear case[4,11] but the details are different). To be an average stationary point of (9) whenever we achieve the ideal objective $z_k = u_k$ (see Section 2.2) requires the average update in (5) to be zero, i.e.

$$\text{choose } \gamma_p \text{ s.t. } E\{\varepsilon_k R_k\}|_{z_k = u_k} = 0 \quad \forall k$$

which implies in particular

$$(1, 0, 0, \ldots, 0) E\{\varepsilon_k R_k\}|_{z_k = u_k} = 0 \quad \forall k$$

This reduces to the scalar equation

$$E\{\mathrm{sgn}(z_k) | z_k |^{p-1} (| z_k |^p - \gamma_p) z_k\}|_{z_k = u_k} = 0$$

or

$$\gamma_p = \frac{E\{| u_k |^{2p}\}}{E\{| u_k |^p\}} \tag{12}$$

which (by construction in the selection of cost (8)) is the same constant which needs to be employed in the linear equalizer case.[22] Note that for binary $\{u_k\}$ one has $\gamma_p = 1$.

2.3. *Comparison with linear equalization*

We have developed two classes of closely related blind algorithms to adapt the parameters of the non-linear DFE structure in Figure 2. With careful definitions of the non-convex cost functions (4) and (8) we were able to translate popular linear equalizer blind algorithms, to the DFE context. More general algorithms could also have been considered (e.g. those found in Reference 4). Note that the regressors which are derived in the DFE case have quantized components which are generated in a recursive non-linear manner, highlighting the departure from the linear case and the need for new analysis tools. For example, we will need to understand the stochastic generation mechanism of the signals $\hat{u}_k$ to predict the mean convergence behaviour of the above algorithms. This is a major problem for DFEs, because we shall need to incorporate error propagation effects into the analysis—a situation which has no analogue in the linear case.

Our approach now will be to set up the theoretical framework to analyse the convergence behaviour of the above algorithms. We shall focus principally on the decision-directed algorithm (5), with specific examples to verify our theory which predicts accurately both desirable and undesirable local minima for blind DFE adaptation.

3. GENERAL CONVERGENCE ANALYSIS

3.1. Minima

Some of the potential convergence points for the adaptive algorithm (5) are those $\tilde{D}$ which satisfy

$$E\left\{\frac{\partial J(z_k)}{\partial D}\right\}\bigg|_{D=\tilde{D}} = E\{\varepsilon_k R_k\}|_{D=\tilde{D}} = 0 \tag{13}$$

being the condition for a zero average update of the blind algorithm. To be also a minimum, such a point must satisfy an additional stability condition for which a sufficient condition is positive definiteness of the Hessian:

$$\frac{\partial}{\partial D} E\left\{\frac{\partial J(z_k)}{\partial D}\right\}\bigg|_{D=\tilde{D}} > 0 \tag{14}$$

A more general minimum at $\tilde{D} \triangleq (\tilde{d}_0, \tilde{d}_1, \ldots, \tilde{d}_N)^{\mathrm{T}}$ need not be smooth and for the mean surfaces that we will need to consider it will satisfy

$$\lim_{\varepsilon \to 0^+} \Delta_i E\left\{\frac{\partial J(z_k)}{\partial d_i}\right\}\bigg|_{d_i = \tilde{d}_i + \varepsilon\Delta_i} > 0, \quad i \in \{0, 1, \ldots, N\}, \quad \forall \Delta \tag{15}$$

where $\Delta \triangleq (\Delta_0, \Delta_1, \ldots, \Delta_N)^{\mathrm{T}}$ denotes an arbitrary direction. We will mostly focus on smooth minima satisfying (13) and (14) in this paper.

The general convergence attributes of an algorithm can be classified by determining the complete set of minima. Such minima represent the parameter values to which the blind algorithm may converge. When all minima correspond to satisfactory parameter settings which each achieve an open eye pattern, then the algorithm is termed admissible.[9]

3.1.1. Decision-directed case. Smooth equilibria $\tilde{D}$ for the decision directed DFE which utilizes error ε_k (7) and regressor R_k (6) satisfy

$$E\{\varepsilon_k R_k\}|_{D(k)=\tilde{D}} = E\{(y_k - R_k^{\mathrm{T}} D(k)) R_k\}|_{D(k)=\tilde{D}} = 0$$

Solving formally, we obtain

$$\tilde{D} = E\{R_k R_k^{\mathrm{T}}\}^{-1} E\{y_k R_k\} \tag{16}$$

which resembles the classical Wiener–Hopf formula. However, the precise interpretation of (16) requires a deeper understanding of the statistics of $\{\hat{u}_k\}$ and the joint statistics of $\{u_k\}$ and $\{\hat{u}_k\}$. For example, it is clear that the tuning of the DFE and therefore the parameter values taken by taps $D(k)$ affect the decisions $\hat{u}_k$ and thereby the regressor R_k. Hence without further qualification (16) represents only an *implicit* definition of $\tilde{D}$. We will move later to justify (16) more fully.

3.1.2. Constant modulus case. Smooth equilibria $\tilde{D}$ for the constant modulus algorithms which utilize error ε_k (11) and regressor R_k (10) satisfy (13) which reduces to

$$\frac{E\{|z_k|^{2p}\}}{E\{|z_k|^p\}} = \gamma_p \tag{17}$$

and

$$E\{\text{sgn}(z_k)|z_k|^{p-1}(|z_k|^p - \gamma_p)u_{k-i}\} = 0, \quad i \in \{1, 2, \ldots, N\} \tag{18}$$

Explicit expressions for the equilibrium $\tilde{D}$, as for the linear equalizer constant modulus algorithms,[5] are only available for limited sets of channels.[23] Because of the analytical limitations of the constant modulus algorithms, we will, unless explicitly stated otherwise, concentrate on the decision-directed algorithm.

3.2. Parameter space partition

To compute explicitly the statistics which affect the location of the potential convergence points of the blind algorithm (16)–(18), we draw on and generalize the resuts of Reference 19. These results establish the relationship between regions in parameter space D called polytopes and finite state Markov processes (FSMPs) which completely specify the relevant statistics. This relationship exists *independently* of the blind algorithm under study and so these results do not restrict themselves to the decision-directed or constant modulus cases.

Define the following hyperplanes in D-space:

$$\left\{D: \sum_{i=0}^{N} h_i u_{k-i} - \sum_{i=1}^{N} d_i \hat{u}_{k-i} \in \{(2-M)d_0, (4-M)d_0, \ldots, (M-2)d_0\}\right\} \tag{19}$$

as we vary across all possible values taken by $u_k, \ldots, u_{k-N} \in \mathcal{M}$ and $\hat{u}_k, \ldots, \hat{u}_{k-N} \in \mathcal{M}$. (Each choice of u_{k-i}, $\hat{u}_{k-i}$ and coefficient $(2-M)$, $4-M)$, ..., $(M-2)$ gives one hyperplane.) From (7) it is clear that whenever the residual Q_k is zero, these hyperplanes define the manifolds in $\mathbb{R}^{N+1}$ for which the argument of the quantizer $\mathscr{Q}(\cdot)$ in (2) lies on a decision threshold. These hyperplanes act as switching surfaces in the following sense. After Reference 19 define an atomic state

$$X_k \triangleq (u_{k-1}, u_{k-2}, \ldots, u_{k-N}, \hat{u}_{k-1}, \hat{u}_{k-2}, \ldots, \hat{u}_{k-N})^{\mathrm{T}} \tag{20}$$

Then from (1) whenever the channel and equalizer parameters satisfy (19) there exist an atomic state X_k and a current input u_k such that an arbitrarily small perturbation in D can change a $\hat{u}_k = +1$ *decision to a* $\hat{u}_k = -1$ *decision* (or vice versa). Before describing the significance of these hyperplanes for stochastic DFE modelling, we give an example.

Example. Let $M = 2$, $N = 2$, $h_0 = 1$, $h_1 = 4$ and $h_2 = 3$. The $2^{2N} = 16$ lines which partition D-space are given by $d_1 \pm d_2 = \zeta$ (with d_0 unconstrained in this case since the input is binary) for $\zeta \in \{0, \pm 2, \pm 6, \pm 8\}$ according to (19). These lines are depicted in Figure 4. In this figure the point $(4, 3)^{\mathrm{T}}$, representing the location of the channel tail, has been indicated by a small cross (the shading can be ignored for the moment). Notice that owing to the degeneracy in this example (ζ can take the value zero), we really only have only 14 distinct lines rather than the generic 16. We will return later to consider this example in more detail.

Remarks

(i) The effect of non-zero Q_k in (7) is to blur the boundaries defined in (19). This means

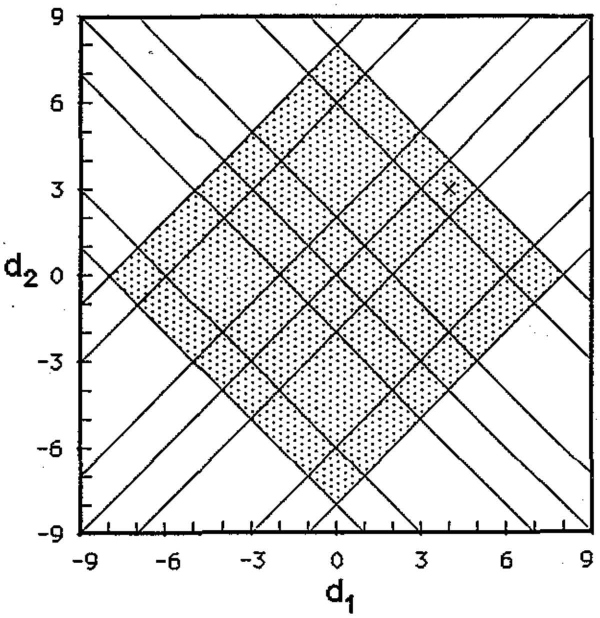

Figure 4. Polytopes for $h_0 = 1$, $h_1 = 4$ and $h_2 = 3$

that our modelling when the parameters are near the boundaries is non-ideal given the presence of channel noise or the effects of an undermodelled channel tail.

(ii) The remaining analysis will initially assume $Q_k = 0$ to simplify the presentation, keeping in mind the significance of a non-zero Q_k. We will show ill-convergence of the DFE blind algorithm in this ideal case and establish that this behaviour remains when $Q_k \neq 0$.

The hyperplanes above partition D-space into a collection of polytopes. The key property of polytopes which we need here and which is an immediate consequence of the definition of the partition is that *one cannot distinguish between any two DFE D-parameter settings within a given polytope based on observations of the output* $\{\hat{u}_k\}$ *alone.*

Now we shall use this property in understanding some statistical properties. First we restrict the input class.

Data assumption

The input data $\{u_i\}$ form an equiprobable independent and identically distributed (i.i.d.) M-ary sequence.

Then we have a finite state Markov process (FSMP) describing the stochastic dynamics of the DFE, with M^{2N} states given by all possible values of X_k (20).[19] One can verify as an immediate consequence of the just-mentioned property that there is a one-to-one correspondence between polytopes and sets of FSMPs. Thus the conceptual picture is that as we drift through parameter space (under adaptation), the underlying FSMP, which governs the full joint statistics of the input u_k and output $\hat{u}_k$ changes (abruptly) only when we cross polytope boundaries. Inside a given polytope the process $\{\hat{u}_k\}$ can be modelled by the stationary behaviour of its associated FSMP, which is independent of variations in D within the polytope. When the input independence assumption above does not hold and we have input correlation, it is possible to still use an FSMP as an approximation device.[24,25] Alternatively we may in most cases extend the Markov state space and retain an exact description. Both paths could only distract our main aim to describe *gross* properties of DFE blind adaptation on the simplest non-trivial system.

Note that from the FSMP, for a frozen DFE parameter setting, it is easy to calculate, in principle, stationary entities of the form

$$E\{\hat{u}_k\hat{u}_{k-i}\} \quad \text{and} \quad E\{u_k\hat{u}_{k-i}\} \tag{21}$$

which are expressible in terms of an invariant probability measure via an unilluminating calculation.[16] What is important here is that we will see that quantities like (21) provide us with sufficient information to characterize the gross convergence properties of the blind adaptation algorithm (5). More generally, the FSMP furnishes moments relevant to arbitrary blind algorithms.

3.3. *Piecewise constant statistics*

Above we indicated a straightforward one-to-one correspondence between the polytopes and FSMPs (or, more generally state transition diagrams). Hence in D-space the joint statistics of u_k and $\hat{u}_k$ will be invariant to variations in $D(k)$ that are constrained within a given polytope denoted $\mathscr{P}$. Therefore define

$$\mathbf{A}(\mathscr{P}) \triangleq E\{R_kR_k^{\mathrm{T}}\}|_{D(k)\in\mathscr{P}}, \qquad \mathbf{C}(\mathscr{P}) \triangleq E\{R_kU_k^{\mathrm{T}}\}|_{D(k)\in\mathscr{P}} \tag{22}$$

where

$$U_k \triangleq (u_i, u_{k-1}, \ldots, u_{k-N})^{\mathrm{T}}$$

which will be useful later (also $\hat{U}_k$ will denote the analogous vector for the data estimates).

Now with each polytope $\mathscr{P}$ we may associate an equilibrium not necessarily inside $\mathscr{P}$ (potential attraction point for the blind adaptation algorithm); this is obtained as follows. Under the negligible residual assumption $Q_k = 0$ we may write the channel output equation as

$$y_k = U_k^{\mathrm{T}}H$$

where U_k is as defined in (22) and $H \triangleq (h_0, h_1, \ldots, h_N)^{\mathrm{T}}$. Then the decision-directed equilibria (16) become

$$\begin{aligned}\tilde{D}^{(\mathscr{P})} &= E\{R_kR_k^{\mathrm{T}}\}^{-1}E\{R_kU_k^{\mathrm{T}}\}H \\ &= \mathbf{A}(\mathscr{P})^{-1}\mathbf{C}(\mathscr{P})H\end{aligned} \tag{23}$$

The Hessian (14) associated with this equilibrium is non-negative or positive definite according to the definiteness of $\mathbf{A}(\mathscr{P})$ in (22) (see Section 3.4). It is natural to classify two types of equilibria according to whether or not the following property holds.

Definition

$\tilde{D}^{(\mathscr{P})}$ *is locally attainable* if $\tilde{D}^{(\mathscr{P})} \in \mathscr{P}$.

Figure 5 shows the geometrical difference between a locally attainable equilibrium and an unattainable equilibrium (where $\tilde{D}^{(\mathscr{P})} \notin \mathscr{P}$). If $\tilde{D}^{(\mathscr{P})}$ is locally attainable, then $\{D(k)\}$ will *tend* to move towards and settle down around it whenever $D(k) \in \mathscr{P}$. Otherwise $\{D(k)\}$ will tend to move towards the boundary $\partial\mathscr{P}$ of $\mathscr{P}$ nearest to $\tilde{D}^{(\mathscr{P})}$ and thus head on into an adjacent

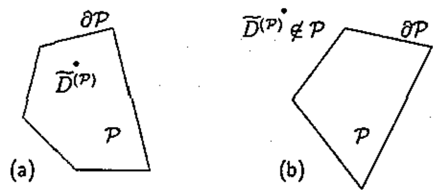

Figure 5. (a) Locally attainable equilibrium. (b) Unattainable equilibrium

polytope. Thus all locally attainable equilibria are real attraction points for the blind algorithm. The example in Section 4.2 best illustrates these ideas.

3.4. Mean error surface

For the decision-directed DFE blind algorithm the gradient descent within a polytope $\mathscr{P}$ is relative to the mean error surface defined through

$$\mathscr{J}(D(k) \in \mathscr{P}) \triangleq E\{J(z_k)\}\,|_{D(k)\in\mathscr{P}} = E\{\tfrac{1}{2}\varepsilon_k^2\}|_{D(k)\in\mathscr{P}} \tag{24}$$

substituting (7) into (4). Note that $\varepsilon_k = y_k - R_k(\mathscr{P})^{\mathrm{T}}D(k) = U_k^{\mathrm{T}}H - R_k(\mathscr{P})^{\mathrm{T}}D(k)$ under the assumption $Q_k = 0$. Therefore

$$\begin{aligned}\mathscr{J}(D(k) \in \mathscr{P}) &= \tfrac{1}{2}E\{y_k^2\} - E\{y_kR_k(\mathscr{P})^{\mathrm{T}}\}D(k) + \tfrac{1}{2}D(k)^{\mathrm{T}}E\{R_k(\mathscr{P})R_k(\mathscr{P})^{\mathrm{T}}\}D(k)\\ &= \tfrac{1}{2}H^{\mathrm{T}}H - H^{\mathrm{T}}\mathbf{C}(\mathscr{P})^{\mathrm{T}}D(k) + \tfrac{1}{2}D(k)^{\mathrm{T}}\mathbf{A}(\mathscr{P})D(k)\end{aligned} \tag{25}$$

where we have invoked the independence assumption of input data. Therefore within poly topes the mean error surface is *quadratic* with minimum at (23) and Hessian $\mathbf{A}(\mathscr{P})$. The complete picture is formed by concatenating polytopes to reveal the mean error surface to be *piecewise quadratic* with potentially as many equilibria as there are polytopes. However, only the locally attainable equilibria represent actual minima which represent actual convergence points (depending on the initialization) for the algorithm.

3.5. Averaging theory within polytopes

The objective in this subsection is to develop some equations which tie, in a probabilistic way, the behaviour of the blind algorithm under adaptation to an averaged deterministic differential equation under a reasonable set of assumptions. This simpler ordinary differential equation enables us to predict the likely evolution of the blind algorithm within a given polytope $\mathscr{P}$ of interest. The complete picture can be obtained by piecing together the trajectories in the polytopes (with due regard for initial conditions on the polytope boundaries) as will be illustrated later in a simulation example.

We rely on existing results to make this characterization. We follow Reference 26 in this work. However, closely analogous analysis methods which could alternatively be applied are treated in the work of Benveniste *et al.*[27]

Changing the co-ordinates of algorithm (5) such that the equilibrium $\tilde{D}^{(\mathscr{P})}$ associated with polytope $\mathscr{P}$ (23) becomes the origin gives an error system

$$W_{k+1} = W_k + \mu H(W_k, y_k, R_k) \tag{26}$$

valid whenever $D(k) \in \mathscr{P}$, where

$$W_k \triangleq D(k) - \tilde{D}^{(\mathscr{P})}$$

and

$$H(W_k, y_k, R_k) \triangleq y_k R_k - R_k R_k^{\mathrm{T}} \tilde{D}(\mathscr{P}) - R_k R_k^{\mathrm{T}} W_k$$

recalling that $\varepsilon_k = y_k - R_k^{\mathrm{T}} D(k)$ and we have removed the explicit dependence of R_k on $\mathscr{P}$ which is understood. Restricting motion to within a polytope implies that R_k is independent of W_k and $\tilde{D}(\mathscr{P})$ is constant, ensuring that $H(\cdot,\cdot,\cdot)$ is a (piecewise) linear function of W_k.

We form a time-scaled continuous representation of the process W_k as follows:

$$W_\mu(t) = W_{[t/\mu]}$$

where $[x]$ represents the integer part of x. The ODE to be defined will represent an approx imation to the process $W_\mu(t)$ and hence, through a suitable scaling, to the original discrete time process W_k.

Next we define the maximum time interval over which our ODE representation will be valid, corresponding to the first exit time on the polytope of interest, namely

$$\tau_\mu^{\partial\mathscr{P}} \triangleq \inf\{t\colon\ |W_\mu(t)| \in \partial\mathscr{P}\} \tag{27}$$

Here we have implicitly restricted our attention to bounded polytopes. The stopping time in (27) can be easily modified in the instance that the polytope is infinite in extent, in which case we need to additionally ensure boundedness, by further restricting the time interval if necessary, in the form $|W_\mu(t)| < K$ for some finite K.[26]

Let ν_Z represent the joint distribution of $Z_k \triangleq \{y_k, R_k\}$ which is a function of the input process. Define

$$\hat{H}(w) \triangleq E_Z\{H(w, y_k, R_k)\}$$

as the averaged quantity in the ODE equation

$$W(t) = w_0 + \int_0^t \hat{H}(W(s))\,\mathrm{d}s \tag{28}$$

For our system $\hat{H}(w)$ evaluates to be

$$\begin{aligned}\hat{H}(w) &\triangleq E\{y_k R_k\} - E\{R_k R_k^{\mathrm{T}}\}\tilde{D}(\mathscr{P}) - E\{R_k R_k^{\mathrm{T}}\}w\\ &= -E\{R_k R_k^{\mathrm{T}}\}w\\ &= -\mathbf{A}(\mathscr{P})w\end{aligned}$$

where we have used the property that the process R_k is independent of the $D(k)$ parameter setting and hence the parameter error W_k (provided that, as we have assumed, the $D(k)$ trajectory remains inside $\mathscr{P}$). Later when we illustrate the trajectories of the ODE using simulations, we adopt a straightforward difference equation approximation to (28).

Then we have the following convergence-in-probability result under suitable initial conditions:[26] for every $T > 0$,

$$\lim_{\mu\to 0} P\left(\sup_{0\leqslant t\leqslant \tau_\mu^{\partial\mathscr{P}}\wedge T} |W_\mu^{\tau_\mu^{\partial\mathscr{P}}}(t) - W(t)| > \varepsilon\right) = 0,$$

where $W_\mu^{\tau_\mu^{\partial\mathscr{P}}}(t)$ denotes the 'stopped process' (defined by freezing the continuous time process if it reaches the polytope boundary $\partial\mathscr{P}$) given by

$$W_\mu^{\tau_\mu^{\partial\mathscr{P}}}(\cdot) \triangleq W_\mu(\cdot \wedge \tau_\mu^{\partial\mathscr{P}})$$

where $\wedge$ denotes the binary operation of minimization. This result connects the evolution of W_k defined in (26) with the trajectories of the ODE in a limiting fashion as the step size is made sufficiently small.

The final indication that we made concerns characterizing the density of the process $W_\mu(t)$ when the equilibrium point $\tilde{D}^{(\mathscr{P})}$ is locally attainable (we have already indicated that such an equilibrium is a minimum). For the error system, local attainability means that the origin lies inside the polytope of interest. Thus in this case we will have a density with zero mean. Again we may use central limit theory results in Reference 26 to characterize this density, noting that

$$\frac{\partial}{\partial w}\hat{H}(w) = -\mathbf{A}(\mathscr{P})$$

is Lipshitz continuous. We refer the interested reader to Reference 26 for these central-limit-theory-type results, which also show how to incorporate noise and modelling errors into the analysis.

4. DETAILED CASE STUDIES

4.1. Background

The previous sections described the mechanisms behind equilibria for DFE blind adaptation. The theory accounts for propagated decision errors by incorporating finite state Markov processes to model the various relevant statistics. Whilst these results are of theoretical interest, we have yet to indicate their applicability for practical systems. This section seeks to establish that the theory does have significant predictive power by looking at detailed case studies. Section 5 will continue the theoretical investigations (based on observations of this section) to describe various classes of equilibria that may arise and the conditions for their existence.

One of the difficulties of analysing blind adaptation is to trace down the cause for any malperformance observed, particularly when the basis of the study relies on simulations. This is not quite the situation here, since the basic theory has been developed. However, we are conscious *initially* not to introduce extraneous components not covered by the previous theory which may cloud interpretation in our first example. Thus we wish to study what might be described as an idealistic situation which is

- (i) *dimensionally well-posed*, meaning that the tapped delay line of the DFE is neither overparametrized nor underparametrized, where for the latter it is easy to construct malconverging examples
- (ii) *noiseless*, noting that with noise eventual escape from minima is possible via a large-deviation mechanism[28] or conversely with noise the domain of attraction of an undesirable equilibrium may deepen and increase in size as has been shown for the blind linear equalizer adaptation in Reference 23
- (iii) *binary*, in preference to M-ary, to avoid a failure mechanism exposed by Mazo[8] in the linear decision-directed case which might appear in the multilevel blind DFE.

That such a system is not necessarily derived from practice is acknowledged. However, our example reveals that blind adaptation using the proven and popular decision-directed design philosophy nonetheless *exhibits convergence problems in the form of undesirable local minima.* Later we show that in the non-ideal case such behaviour is maintained, meaning that in practice such behaviour can be expected.

Our results do not imply that our formulation of blind adaptation for the non-linear DFE is poor. Precisely the same qualitative results exist for most recursive gradient descent blind algorithms developed for linear equalizers. Results in both cases highlight the importance of good *initialization strategies* when using these algorithms and the futility of trying to establish that these types of non-convex gradient descent algorithms converge globally to desirable minima for finite-dimensional equalizers.

4.2. Example

As in Section 3.2 we chose the example $h_0 = 1$, $h_1 = 4$ and $h_2 = 3$ (here, as before, we are considering the case of binary inputs). In Figure 6 we have plotted a large number of steepest descent trajectories according to a finite difference equation version of (28) with $\mu = 0{\cdot}01$, noting that the matrices **R** and **C** are now dependent on the polytopes (pictured in Figure 4). Note that Figure 6 is a two-dimensional projection of D-space. Therefore some of the 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.

Figure 7 shows the precise sense in which to interpret Figure 6. It shows a section of Figure 6 with a single (bold) steepest descent (averaged) trajectory (plucked from Figure 6) and four realizations initialized from $(0, 7, 0)^T$ (i.e. simulations according to (5) generated via a random number generator) which appear to cluster about the steepest descent (averaged) trajectory. Note for this example that there are only three locally attainable equilibria at $(1, 4, 3)^T$, $(4, 3, 0)^T$ and $(3{\cdot}792, 4{\cdot}833, 3{\cdot}292)^T$. In Figure 6 the 2D protections of these equilibria (depicted as small circles) are given by $(4, 3)^T$, $(3, 0)^T$ and $(4{\cdot}833, 3{\cdot}667)^T$ and these appear as

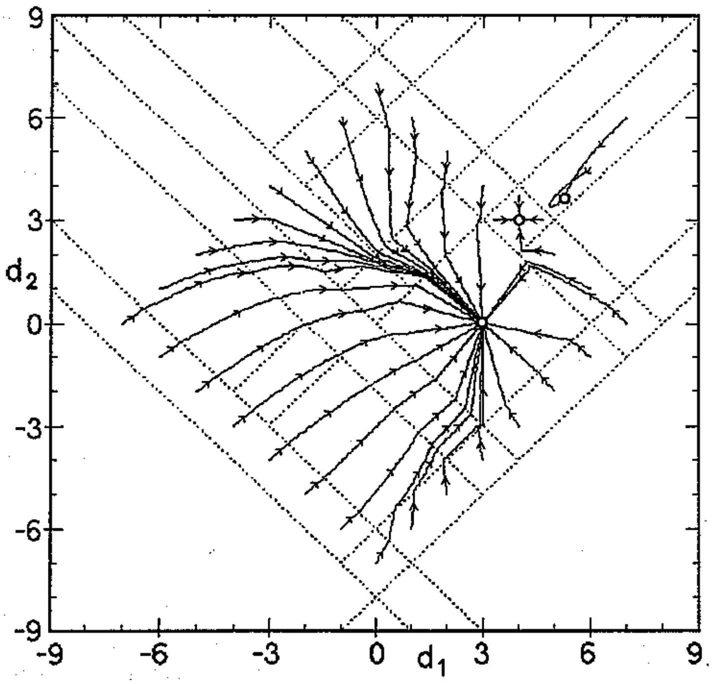

Figure 6. Steepest descent trajectories

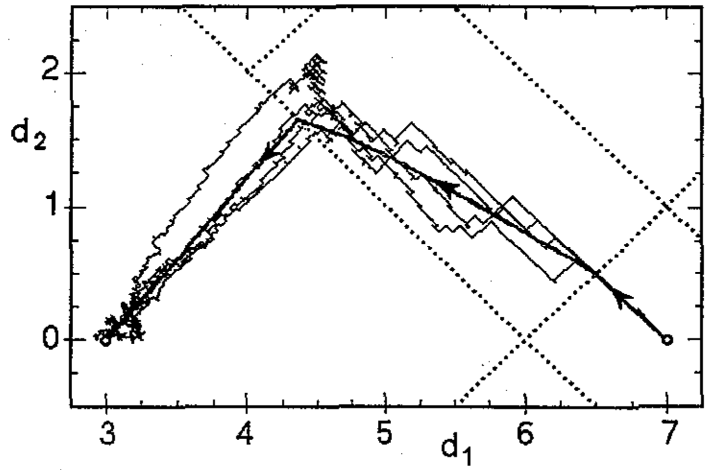

Figure 7. An averaged trajectory with four realizations

(local) attraction points for the mean trajectories. The last equilibrium shows that blind adaptation based on the Sato algorithm may be flawed in the sense that it does not correspond to an equalized system as we will see later (whereas the first two do—at the first, $\hat{u}_k = u_k$, and at the second, $\hat{u}_k = u_{k-1}$). A similar clear example for the simpler linear case is given in Reference 23.

In Figure 6 we have aggregated polytopes from Figure 4 whenever adjacent polytopes have the same correlation statistics (more precisely, when neighbouring polytopes have isomorphic sets of recurrent atomic states). In a sense this indicates that aggregations of polytopes are more important objects than the polytopes themselves for investigation and this is a lead in to some of our later results.

5. CLASSIFYING LOCALLY ATTAINABLE EQUILIBRIA

5.1. Background

The above example established unambiguously that the decision-directed blind algorithm can exhibit local minima with various properties. In this section we seek to classify and understand these crucial points in the parameter space because their properties will affect directly the success or otherwise of the blind strategy adopted.

Our explicit aims are to determine necessary and sufficient conditions for the existence of desirable equilibria. We shall also have something to say about undesirable equilibria, which are more difficult to study.

5.2. Delay-type equilibria local attainability

Firstly we consider those aggregations of polytopes in D-space which yield a decision sequence which is a delay of the input (with an associated possible sign change) under steady state, i.e. $\hat{u}_k = \text{sgn}(h_\delta)u_{k-\delta}\ \forall k$. We derive necessary (and conjecture sufficient) conditions for attainability of the attraction points of these groups of polytopes for Sato algorithms in terms of the channel parameters. (Note in the following development that all the results relate to non-adaptive properties and are therefore algorithm-independent.)

Let $\sigma_\delta \triangleq \text{sgn}(h_\delta)$, then, rewriting (1), we have (recall that we are restricting attention to binary signalling)

$$\hat{u}_k = \text{sgn}(h_\delta u_{k-\delta} + U_k(\delta) + V_k(\delta)), \quad 0 \leqslant \delta \leqslant N \tag{29}$$

where

$$U_k(\delta) \triangleq \sigma_\delta \sum_{i=\delta+1}^{N} d_{i-\delta}(u_{k-i} - \sigma_\delta \hat{u}_{k+\delta-i}) \tag{30}$$

$$V_k(\delta) \triangleq \sum_{i=0}^{\delta-1} h_i u_{k-i} + \sum_{i=\delta+1}^{N} (h_i - \sigma_\delta d_{i-\delta}) u_{k-i} - \sum_{i=N-\delta+1}^{N} d_i \hat{u}_{k-i} \tag{31}$$

We also define an upper bound on (31),

$$V_{\text{MAX}}(\delta) \triangleq \sum_{i=0}^{\delta-1} |h_i| + \sum_{i=\delta+1}^{N} |h_i - \sigma_\delta d_{i-\delta}| + \sum_{i=N-\delta+1}^{N} |d_i| \tag{32}$$

The reason for the curious decomposition given by (29) 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 of the decoded symbols through the channel–DFE combination.

We define two subsets of Ω of 4^N atomic states X_k (20) parametrized by $0 \leqslant \delta < N$. (The case $\delta = N$ needs to be treated separately, but fortunately is easily disposed of.) Define

$$\omega_+(\delta) \triangleq \{X_k \in \Omega\colon \hat{u}_{k-i} = +\sigma_\delta u_{k-\delta-i}, i = 1, 2, \ldots, N-\delta\} \tag{33}$$

$$\omega_-(\delta) \triangleq \{X_k \in \Omega\colon \hat{u}_{k-i} = -\sigma_\delta u_{k-\delta-i}, i = 1, 2, \ldots, N-\delta\} \tag{34}$$

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

Remarks

(i) Note that 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).

(ii) We will see that whereas in the linear case[4] $\hat{u}_k = +u_{k-\delta}$ $\forall\{u_k\}$ and $\hat{u}_k = -u_{k-\delta}$ $\forall\{u_k\}$ are *both* possible for some δ for a fixed channel, the situation for the DFE is different because only $\hat{u}_k = +\sigma_\delta u_{k-\delta}$ $\forall\{u_k\}$ will be possible (σ_δ being fixed by the channel).

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 $(m > k)$ decisions are of the form $\hat{u}_m = \pm\sigma_\delta u_{m-\delta}$; (b) $\omega_\pm(\delta)$ is closed. Hence to investigate channel–DFE combinations yielding simple time delay behaviour, we need only determine when a set

$\omega_{\pm}(\delta)$ is closed and reachable from arbitrary states within Ω. The following proposition narrows our investigations by showing that a DFE can never behave consistently according to the law $\hat{u}_m = -\sigma_\delta u_{m-\delta}$ when in a steady state stochastic environment, and it also gives necessary and sufficient conditions for $\omega_+(\delta)$ closure.

Proposition 1

(a) $\omega_+(\delta)$ is closed if and only if

$$|h_\delta| > V_{\mathrm{MAX}}(\delta) \tag{35}$$

(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 $\hat{u}_{k-i} = \pm\sigma_\delta u_{k-\delta-i}$ for all $i \in \{1, 2, \ldots, N\}$. Substituting into (30) and (31), we obtain

$$U_k^-(\delta) \triangleq 2\sigma_\delta \sum_{i=\delta+1}^{N} d_{i-\delta} u_{k-i}, \qquad U_k^+(\delta) \triangleq 0 \tag{36}$$

$$V_k^{\pm}(\delta) \triangleq \sum_{i=0}^{\delta-1} h_i u_{k-i} + \sum_{i=\delta+1}^{N} (h_i - \sigma_\delta d_{i-\delta}) u_{k-i} \mp \sigma_\delta \sum_{i=N+1}^{N+\delta} d_{i-\delta} u_{k-i} \tag{37}$$

Consider first $\omega_+(\delta)$. If $|h_\delta| > V_{\mathrm{MAX}}(\delta)$ with $U_k^+(\delta) = 0$, then this implies $|h_\delta| > |V_k^+(\delta)|$ for all $V_k^+(\delta)$, hence $\hat{u}_k = +\sigma_\delta u_{k-\delta}$ by (29). Then note that the u_k in (30) are distinct and therefore the supremum of $V_k^+(\delta)$ over $\{u_k\}$ is just $V_{\mathrm{MAX}}(\delta)$, so (35) is also necessary.

Now consider $\omega_-(\delta)$. In this case we have $\hat{u}_k = \mathrm{sgn}(h_\delta u_{k-\delta} + U_k^-(\delta) + V_k^-(\delta))$, where by (30) and (31) $U_k^-(\delta) + V_k^-(\delta)$ s independent of $u_{k-\delta}$ and has a distribution symmetric about zero. Hence $Pr(\hat{u}_k = +\sigma_\delta u_{k-\delta}) \geqslant \frac{1}{2}$, contradicting closure (noting that if $\omega_-(\delta)$ were closed, then $Pr(\hat{u}_k = +\sigma_\delta u_{k-\delta}) = 0$). □

If (35) 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 that no proper subset of $\omega_+(\delta)$ is closed assuming that $\omega_+(\delta)$ itself is closed (i.e. $\omega_+(\delta)$ is a set of recurrent states). We formulate this as follows.

Proposition 2

$\omega_+(\delta)$ is indecomposable.

The inequality (35) can only hold for at most one value of δ (for fixed parameter values). To prove this, one assumes that at least two inequalities of the form (35) 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 3.

Proposition 3

$\omega_+(\delta)$ is closed for at most one $\delta \in \{0, 1, \ldots, N\}$.

Proposition 3 can be viewed as a special case of a more general problem, now considered. Having established that only under suitable conditions is $\omega_+(\delta)$ closed and indecomposable, the crucial question arises as to whether it can be reached from an arbitrary atomic state

$X_k \in \Omega \backslash \omega_+(\delta)$ 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.[19] A full answer to this question is not yet known. We present the following result (Proposition 4) and important conjecture (Conjecture 1).

Proposition 4

Let $|h_\delta| > V_{\mathrm{MAX}}(\delta)$ for some $0 \leqslant \delta < N$. Then the following alternative conditions are sufficient to guarantee that, no matter what the atomic state is, there exists an input sequence such that, commencing at some possible later time than the time of commencement of the input sequence, $N-\delta$ consecutive $\hat{u}_m = +\sigma_\delta u_{m-\delta}$ decisions are made:

(i) $\delta = 0, 1, N-2, N-1, N$ (and thus cases $N = 1, 2, 3, 4$)
(ii) $\eta \operatorname{sgn}(d_1) = \eta^2 \operatorname{sgn}(d_2) = \cdots = \eta^{N-\delta} \operatorname{sgn}(d_{N-\delta})$ for $\eta \in \{+1, -1\}$
(iii) $\eta \operatorname{sgn}(h_{\delta+1}) = \eta^2 \operatorname{sgn}(h_{\delta+2}) = \cdots = \eta^{N-\delta} \operatorname{sgn}(h_N)$ for $\eta \in \{+1, -1\}$.

We prove this result for $\delta = 0$ and $\delta = 1$. The remaining cases are easier to prove and the proofs have been omitted.

Proof. Let $\mathscr{F}_k = \sigma(u_k, u_{k-1}, \ldots, \hat{u}_k, \hat{u}_{k-1}, \ldots)$ denote the sigma algebra generated by data and decisions up to and including time k.

(a) $\delta = 0$. From (30) we note that $U_k(0)$ is $\mathscr{F}_{k-1}$-measurable. Hence selecting

$$u_k = +\sigma_0 \operatorname{sgn}(U_k(0))$$

gives $\hat{u}_k = +\sigma_0 u_k$ by (35) with $\delta = 0$. Apply this rule for N consecutive k.

(b) $\delta = 1$. Suppose that for some fixed k the DFE has been driven by the homing sequence $\{u_{k-1} = +\sigma_1; u_{k-j} = +\sigma_1 \operatorname{sgn}(d_{i-j}), i = 2, 3, \ldots, N\}$. (We shall explain how, u_k, u_{k+1}, etc. are to be chosen so that $\hat{u}_{k+j} = +\sigma_1 u_{k+j-1}\ \forall j \geqslant 0$.) Then from (30) we have

$$U_k(1) = \sum_{i=2}^{N} |d_{i-1}| - \sum_{i=2}^{N} d_{i-1}\hat{u}_{k+1-i} \geqslant 0 \tag{38}$$

Thus $\hat{u}_k = +1$ independently of u_k, because, substituting into (29), we have (i) $h_1 u_{k-1} + U_k(1) \geqslant |h_1|$ and (ii) $|h_1| > |V_k(1)|$ from (35). Note that the $\hat{u}_k$ decision satisfies $\hat{u}_k = +\sigma_1 u_{k-1}$ (with $u_{k-1} = +\sigma_1$), i.e. is one decision of the desired form. We need to make the next $N-2$ decisions also of this form. Now, since $\hat{u}_k = +\sigma_1 u_{k-1}$ is guaranteed independently of u_k, we conclude that $U_{k+1}(1)$ is $\mathscr{F}_{k-1}$-measurable. Thus we may set $u_k = +\sigma_1 \operatorname{sgn}(U_{k+1}(1))$ showing that $h_1 u_k + U_{k+1}(1) = \operatorname{sgn}(U_{k+1}(1))(|h_1| + |U_{k+1}(1)|)$, which dominates $V_{k+1}(1)$ by (35), leading to $\hat{u}_{k+1} = \operatorname{sgn}(U_{k+1}(1)) = +\sigma_1 u_k$ independently of u_{k+1}. Following this, we conclude that $U_{k+2}(1)$ is $\mathscr{F}_k$-measurable, etc. and the recipe is clear. □

Conjecture 1

Let $|h_\delta| > V_{\mathrm{MAX}}(\delta)$ for some $0 \leqslant \delta < N$. Then there exists at least one input sequence such that $N-\delta$ consecutive decisions are made of the form $\hat{u}_m = +\sigma_\delta u_{m-\delta}$.

Remarks

(i) With (35) satisfied and the hypothesis of Proposition 4 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 $\hat{u}_m = +\sigma_\delta u_{m-\delta}$ if and only if $|h_\delta| > V_{\text{MAX}}(\delta)$ with $\omega_+(\delta)$ reachable. Note that the output forms an independent sequence under such conditions.

(ii) Given a time-invariant channel H, define the following regions (at least one does exist) of D-space by rewriting (35):

$$\mathscr{J}(\delta) \triangleq \left\{ D \in \mathbb{R}^{N+1}: \rho_\delta > \sum_{i=\delta+1}^{N} |h_i - \sigma_\delta . d_{i-\delta}| + \sum_{i=N-\delta+1}^{N} |d_i| \right\} \tag{39}$$

where

$$\rho_\delta \triangleq |h_\delta| - \sum_{i=0}^{\delta-1} |h_i|, \quad 0 \leqslant \delta \leqslant N \tag{40}$$

(These regions for various $\delta > 0$ generalize a condition derived by Jennings.[18]) Then $\hat{u}_m = +\sigma_\delta u_{m-\delta}$ under steady state conditions only if $D \in \mathscr{J}(\delta)$ and sometime if (according to the reachability of $\omega_+(\delta)$ and the initial conditions). Region $\mathscr{J}(\delta)$ is non-empty only if $\rho_\delta > 0$. Note for our example in Figure 4 that $\mathscr{J}(0)$ and $\mathscr{J}(1)$ are non-empty because $\rho_0 = |h_0| = 1 > 0$ and $\rho_1 = |h_1| - |h_0| = 3 > 0$, but $\mathscr{J}(2)$ is empty because $\rho_2 = |h_2| - |h_1| - |h_0| = -2 < 0$.

(iii) It can be shown that each region $\mathscr{J}(\delta)$ is a *union* of polytopes whose sets of recurrent FSMP states are isomorphic.

(iv) In the case $\delta = N$ it is readily apparent that condition (35) is necessary and sufficient for every decision to be of the form $\hat{u}_m = +\sigma_N u_{m-N}$.

Now we state the main result, which shows that it is simple to check for the existence of delay-like attraction points for the decision-directed blind algorithm for the adaptive DFE.

Theorem 1

A necessary condition for the decision-directed blind adaptive algorithm

$$D(k+1) = D(k) + \gamma \varepsilon_k \hat{U}_k \tag{41}$$

where

$$\varepsilon_k \triangleq U_k^{\mathrm{T}} H - \hat{U}_k^{\mathrm{T}} D(k) \tag{42}$$

to have a locally attainable equilibrium corresponding to the channel–DFE combination producing decisions of the form $\hat{u}_m = +\sigma_\delta u_{m-\delta}$ under steady state is

$$\rho_\delta \triangleq |h_\delta| - \sum_{i=0}^{\delta-1} |h_i| > 0 \tag{43}$$

Further, this equilibrium is given by

$$D_{\text{EQU}}(\mathscr{P}) \triangleq +\sigma_\delta (h_\delta, h_{\delta+1}, \ldots, h_N, 0, \ldots, 0)^{\mathrm{T}}$$

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 $\Omega \backslash \omega_+(\delta)$).

Proof. Suppose $\hat{u}_m = +\sigma_\delta u_{m-\delta}\ \forall m$ under steady state. Then it follows that $\omega_+(\delta)$ is closed (by definition). With $\omega_+(\delta)$ closed it is necessary that $|h_\delta| > V_{\text{MAX}}(\delta)$ by Proposition 1. In particular, this implies

$$|h_\delta| > \sum_{i=0}^{\delta-1} |h_i|$$

by (29). Hence $\rho_\delta > 0$ as in (43).

Now suppose $\rho_\delta > 0$. Then by (39) $\mathscr{J}(\delta)$ is non-empty. Let $D(k) \in \mathscr{J}(\delta)$, in which case $\omega_+(\delta)$ is closed. Now only if $\omega_+(\delta)$ is reachable from all initial atomic states (Proposition 4) can we say that transitions within $\omega_+(\delta)$ completely define the steady state behaviour, i.e. only if $\omega_+(\delta)$ is the only closed subset of Ω (Conjecture 1). Therefore only with $\omega_+(\delta)$ reachable does it follow from the closure property that $\hat{u}_m = \sigma_\delta u_{m-\delta}\ \forall m$ (given $\rho_\delta > 0$). Thus if we assume that $\omega_+(\delta)$ is reachable, then it remains to be shown that there is an equilibrium for $\{D(k)\}$ which is locally attainable (it turns out to be unique). Now, because $\hat{u}_m = +\sigma_\delta u_{m-\delta}\ \forall m$, it follows that (i) $\mathbf{R}(\mathscr{P}) = \mathbf{I}$ and (ii) $\mathbf{C}(\mathscr{P}) = +\sigma_\delta \mathbf{S}_{N+1}^\delta$, where $\mathbf{S}$ is a zero matrix except for the δth superdiagonal of ones. Hence we have $D_{\mathrm{EQU}}(\mathscr{P}) = +\sigma_\delta \mathbf{S}_{N+1}^\delta H$ for all polytopes $\mathscr{P}$ which make up $\mathscr{J}(\delta)$. Then it is trivial to show from (39) that $D_{\mathrm{EQU}}(\mathscr{P}) \in \mathscr{J}(\delta)$. Indeed, in a very real sense $D_{\mathrm{EQU}}(\mathscr{P})$ is the 'centre' of $\mathscr{J}(\delta)$; see Figure 6.) Hence there exists a (particular) polytope $\mathbb{P}^* \in \mathscr{J}(\delta)$ (say) such that $D_{\mathrm{EQU}}(\mathscr{P}^*) = D_{\mathrm{EQU}}(\mathscr{P} \in \mathscr{J}(\delta)) \in \mathscr{P}^*$, i.e. $D_{\mathrm{EQU}}(\mathscr{P}^*)$ is a locally attainable equilibrium. □

For $\delta = 0$, $D_{\mathrm{EQU}}(\mathscr{P}) = D^{\mathrm{OPT}} \triangleq +\sigma_0 H$ 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 $D_{\mathrm{EQU}}(\mathscr{P}) = +\sigma_1 \mathbf{S}_{N+1} H$ is always locally attainable and exact, and so on for more degenerate cases.)

5.3. White equilibria

As we have commented earlier when $\hat{u}_k = +\sigma_\delta u_{k-\delta}$, the process $\{\hat{u}_k\}$ is composed of a sequence of independent equiprobable binary random variables. Let us term any equilibrium with the process $\{\hat{u}_k\}$ 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-space.

Proposition 5

Suppose that $\{\hat{u}_k\}$ forms an independent, equiprobable binary random sequence (under steady state). Then

$$D(k) \in \left\{ D \in \mathbb{R}^{N+1} \colon \|H\|_1 > \sum_{i=1}^{N} |d_i| \right\} \tag{44}$$

Proof. If $\{\hat{u}_k\}$ is an independent, equiprobable binary random sequence, then the subsequence $\{\hat{u}_k = -1;\ \hat{u}_{k-i} = -\mathrm{sgn}(d_i),\ i = 1, 2, \ldots, N\}$ occurs with non-zero probability (for some input sequence). In (1) this implies $-1 = \mathrm{sgn}(U_k^T H + \Sigma_{i=1}^N |d_i|)$ and so (44) follows, noting that $\|H\|_1 \geqslant |U_k^T H|\ \forall \{u_k\}$. □

Remarks

(i) Taking our previous example, this region (44) is shown shaded as a diamond in Figure

4. Here $H = (1, 4, 3)^T$ and we need $|d_1| + |d_2| < 8$; the output of the DFE can be independent only whilst $(d_1, d_2)^T$ lies within this diamond.

(ii) The expression $\|H\|_1$ 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 (44).

(iii) Closely related to the above is the following. Clearly, by the earlier definition of D^{OPT} as $\text{sgn}(h_0)\, H$, we have $\|D^{OPT}\|_1 = \|H\|_1$ (and $\|H\|_1$ 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 $\|D(k)\|_1 = \|H\|_1$ (or progressive estimates thereof).

Proposition 6

If $\det(\mathbf{R}(\mathscr{P})) = 0$, then $\mathscr{P} \subset \{D \in \mathbb{R}^{N+1}: \Sigma_{i=1}^{N} |d_i| > \|H\|_1\}$.

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

$$x_0\hat{u}_k + x_1\hat{u}_{k-1} + x_2\hat{u}_{k-2} + \cdots + x_N\hat{u}_{k-N} = 0 \tag{45}$$

where at least two x_i are non-zero. In particular, (45) implies that $\{\hat{u}_k\}$ is periodic, because $\hat{u}_k$ can take on only a finite number of values.

Now, to obtain a contradiction, suppose that $\Sigma_{i=1}^{N} |d_i| < \|H\|_1$ and consider the two input subsequences $\{u_{k-i} = +\text{sgn}(h_i)\}$ and $\{u_{k-i} = -\text{sgn}(h_i)\}$. Then in the two cases $U_k^T H = +\|H\|_1$ and $U_k^T H = -\|H\|_1$ imply (by hypothesis) that $\hat{u}_k = +1$ and $\hat{u}_k = -1$ respectively. However, this contradicts the periodicity of $\{\hat{u}_k\}$. Therefore $\Sigma_{i=1}^{N} |d_i| > \|H\|_1$ as claimed. □

Remarks

(i) This justifies the earlier restriction that we should only consider polytopes $\mathscr{P}$ satisfying $\det(\mathbf{R}(\mathscr{P})) \neq 0$, because otherwise we would be considering a region of D-space which is complementary to the l_1-ball which, by Proposition 5, contains the only polytopes of interest and to which adaptation is sensibly constrained. There is some evidence leading to the conjecture that condition (44) implies that the stationary atomic distribution of the FSMP is unique. (For example, if (44) holds and $h_i > 0$ for all i, then it is provably unique). However, the general conjecture (which incidentally implies Conjecture 1) remains open.

(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 $\{\hat{u}_k\}$ can be white is for the DFE to produce decisions of the form $\hat{u}_m = +\sigma_\delta u_{m-\delta}$ for one particular δ.

Our analysis leads to the following conjecture.

Conjecture 2

Let $\{u_k\}$ be an independent sequence of random variables taking values in $\{-1, +1\}$ with equal probability. Suppose that

$$\hat{u}_k = \operatorname{sgn}\left(\sum_{i=0}^{N} h_i u_{k-i} - \sum_{i=1}^{N} d_i \hat{u}_{k-i}\right) \tag{46}$$

and the sequence $\{\hat{u}_k\}$ is independently distributed. Then for some $\delta \in \{0, 1, \ldots, N\}$ there holds

$$\hat{u}_k = +\sigma_\delta u_{k-\delta} \quad \forall \{u_k\} \tag{47}$$

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 that it was correctly equalizing the channel up to a delay. The corresponding question for the simpler linear equalizer is solved in Reference 29 and these issues are investigated later.

5.4. *Practical examples*

In this subsection we use our developed theory to examine the behaviour of two practical examples of channels. The first channel we consider is an example which belongs to the class of channels for which a DFE is well suited.[20]

Sampling the twisted pair cable channel found in Reference 20 one obtains the approximate response

$$H = (4, 2, 1{\cdot}6, 1{\cdot}2, 0{\cdot}6, 0{\cdot}4, 0{\cdot}3, 0{\cdot}2, 0{\cdot}1, 0{\cdot}05)^{\mathrm{T}}$$

where we have stopped at $N = 9$. Our theory tells us that there is only one delay equilibrium, because for only one δ, i.e, $\delta = 0$, is $\rho_\delta > 0$. This equilibrium has a large domain of attraction given by $\rho_0 = 4$ by (40). When simulating this system, convergence of the taps D to H was always observed, which is good from a practical perspective but uninteresting or at least unconvincing theoretically. We can say from our theory that if there are any other stable minima, then they are undesirable.

Our second example presents a more interesting behaviour and we will give a more detailed investigation of this case. This second practical example is from the seminal paper by Benveniste *et al.*,[4] which describes a typical French telephone channel. We restrict attention to binary signalling. The real part of the channel impulse response is given by

$$H = (-1, 0, 3, 0{\cdot}8, -29, -2.5, 52, 0{\cdot}8, 18, 1, 5, 0{\cdot}8, 1, 0{\cdot}5)^{\mathrm{T}}$$

and in simulations white Gaussian noise n_k with variance $E\{n_k^2\} = 2 \times 10^{-2}$ should be added as in Reference 4. We form Table I.

Table I. Delay equilibria classification of channel

δ	0	1	2	3	4	5	6	7	8	9	10
ρ_δ	+1·0	−1·0	+2·0	−3·2	−24·2	−31·3	+15·7	−87·5	−71·1	−106·1	−103·1
Rank	4	—	3	—	1	—	2	—	—	—	—

In Table I we have not bothered to give details for $\delta > 10$, because equilibria with delays of this order are impossible (since $\rho_\delta < 0 \;\; \forall \delta > 10$). Indeed, only delay equilibria of order $\delta \in \{0, 2, 4, 6\}$ are theoretically possible, as the second row shows. The ranking given in the third row of the table is according to the size of the domain of attraction, which is proportional to ρ_δ. The existence of local undesirable minima is not known for this system.

5.5. Noise considerations

With the addition of channel noise (Figure 1) additional issues come to light. First we note that it is standard in the linear equalization literature dealing with blind adaptation to work with the noiseless situation or at least deal meaningfully with the high signal-to-noise-ratio limit. (This is despite the realization that at frequencies where the channel attenuation is high a linear equalizer will result in excessive noise enhancement.[1] The problem remains, of course, that even the noiseless case is not completely understood and is very difficult to analyse,[4] so our digression for the analytically more intractable DFE case will be brief.

When at a δ-delay equilibrium and assuming sufficient correct decisions have been made (in a delay sense), we obtain from (29) the equation

$$\hat{u}_k = \operatorname{sgn}\left(h_\delta u_{k-\delta} + \sum_{i=0}^{\delta-1} h_i u_{k-i} + n_k\right)$$

where n_k is some zero-mean, additive channel noise (Figure 1). Then, taking the worst case of past decisions $\{u_{k-i},\ i=0,1,\ldots,\delta-1\}$ (which maximizes the ISI), we see that one measure of the worst-case signal-to-noise ratio is

$$\frac{\left(|h_\delta| - \sum_{i=0}^{\delta-1} |h_i|\right)^2}{E\{n_k^2\}} = \frac{\rho_\delta^2}{E\{n_k^2\}}$$

showing that the delay equilibrium with the largest domain of attraction (l_1-norm radius ρ_δ) *is also the best in terms of the margin against noise-induced decision errors* (and subsequent error propagation). Thus with noise present the exact equilibrium need not be the best delay equilibrium, but rather the delay equilibrium with the largest domain of attraction is preferable. Heuristically we could argue that adaptation initialized near the origin would tend to select the equilibrium with the largest domain of attraction as in Figure 6. Not only is it a larger target for the drifting taps but it is also closer to the origin for increasing δ because its centre is at an l_1-norm distance of $\| D_{\mathrm{EQU}} \|_1 = \Sigma_{i=\delta}^{N} | h_i |$ (Theorem 1). The marriage of these observations leads to a favourable interpretation for the blind adaptation for practical systems.

5.6. Appropriate blind algorithms

Here we wish to discuss both the linear and DFE (blind) equalizer problems. Conjecture 2 raised the question as to whether a delay equilibrium had been reached simply by performing an independence test on the output. The corresponding problem for the linear equalizer has been solved in Reference 29, including the M-ary problem. Denoting by $(l_0, l_1, \ldots, l_R)$ the coefficients representing the convolution of an FIR channel impulse response with the linear equalizer, i.e. the transfer functon from u_k (Figure 1) to z_k (Figure 2), the relevant theorem in Reference 29 in the binary case takes the following form.

Theorem 2 (linear equalizer)

Let $\{u_k\}$ be an independent sequence of random variables taking values in $\{-1, +1\}$ with equal probability. Suppose that for some constants $\{l_0, l_1, \ldots, l_R\}$

$$\hat{u}_k = \operatorname{sgn}\left(\sum_{i=0}^{R} l_i u_{k-i}\right) \tag{48}$$

and $E\{\hat{u}_k\hat{u}_{k-j}\}=0$ for $j \in \{1,2,\ldots,R\}$. Then for some $\delta \in \{0,,1,\ldots,R\}$ there holds

$$\hat{u}_k = \mathrm{sgn}(l_\delta)u_{k-\delta} \quad \forall\{u_k\} \tag{49}$$

Clearly the two-dimensional (i.e. correlation) output test implied by Theorem 2 is simpler than a full independence test (found in Conjecture 2 but probably stronger than what is required). Theorem 2 would seem to have important consequences for the linear equalizer blind global convergence problem.

This theorem gives strong indications that only two-dimensional and not higher output statistics are required and that a sufficiently clever algorithm will lead to a global convergence to only delay equilibria in the vein of results for the sub-Gaussian and super-Gaussian cases derived by Benveniste *et al.*[4] Of course, it is of interest to find the form of the simplest algorithms (computationally) which achieve this in the spirit of (5) for the one-dimensional (memoryless) case. Naturally, any binary globally converging blind linear equalization algorithm would be a significant advance building on Reference 4 (despite the good performance of the Sato algorithm reported in practice).

Parallel questions exist for the DFE and the first step is resolving Conjecture 2, which is the analogue (and a generalization) of the linear case Theorem 2 above. We have shown that the basic Sato-like algorithm (5) is flawed (Figure 6) and it is not hard to believe that more general memoryless cost functions than (4) will fail also. Hence the conjectures raised in our analysis seem to have a heightened significance and are a possible precursor to more robust algorithms.

6. CONCLUSIONS

6.1. Summary

Before broadening the discussion, we note the main features of the work here. The blind DFE adaptation problem was reduced to an analysis of a gradient descent adaptation on a piecewise quadratic mean cost surface. We have been able to predict the potential for undesirable convergence of the blind algorithms considered through a simple mechanism. This potential behaviour was shown to be realized by a simple three-tap channel.

Various potential convergence points for the blind DFE algorithm were classified, focusing on the most important δ-delay equilibria where the DFE output sequence is a δ time sample delay of the input with an associated possible sign inversion. These results differ rom the blind linear equalization case where all possible delays are possible and both signs are possible. With a DFE only a finite number of delays (generally much less than the number of tap parameters) are possible and only one sign is associated with each delay. The conditions under which delay equilibria can exist were presented in a theorem and are interpreted as a simple condition on the channel impulse response values.

6.2. Discussion

We have demonstrated that an analysis of blind adaptation in decision feedback equalization is possible, the principal difficulty being how to incorporate the effects of decision errors into the picture. We have shown that it is possible to build a conceptual model for the behaviour of the equalizer during blind adaptation, and based on this we can understand why the taps may hang at equilibria, leading to poor performance. These investigations more than anything build and aid our intuition; for example, we can view standard blind adaptation as evolving on a mean square error surface composed of a tiling (the polytopes) of quadratic functions.

When it so happens that the minimum of one of these quadratic functions for a given tile lies within the tile, we have a locally attainable minimum, i.e. an attraction point for the adaptation. Also, we can use our theory to predict with sufficient precision the adaptive behaviour of a blind DFE on a given channel, admittedly via tedious calculations.

However, the difficulty remains of gleaning useful practical information from these sorts of investigations. Here we have limited success, e.g. a simple condition on the channel parameters provides information on the *possibility* of the algorithm being found at an equilibrium giving delay-like behaviour. Still, the likelihood of obtaining broad sweeping statements relating to the guarantee of ideal blind convergence attributes seems a distant goal (and the same problem exists for linear equalization, although we have the remarkable results in Reference 4 for special cases). Our work has value in showing just how difficult such a general theory giving useful practical information would need to be and perhaps in highlighting what simplifying assumptions might be valid.

The results in this paper stand in contrast to the simplicity of the systems and algorithms proposed and under study. In fact, it becomes apparent from our work that if one tries to simplify the algorithm from a practical viewpoint, e.g. using the sign of the error (rather than full precision), then the theory can very easily get more complicated (and obscure). Our useful results in this case show that there is a general tendency for the adapting taps to stay in the vicinity of a delay equilibrium (for both the standard and sign error algorithms).

We note that to secure our theory, we need to make a stationarity assumption on the underlying FSMPs. This is a reasonably strong assumption. It is equivalent to a time-scale separation idea, meaning that given a non-ideal initial starting distribution of the FSMP (arbitrary initial conditions in the DFE), we need to stipulate that the adapting taps not move too far before the invariant distribution is established (with reasonable probability). Thus a fast time-scale is associated with the transient dynamics of the FSMP and a slow time-scale is associated with the adaptation algorithm, i.e. the adaptive gain γ needs to be sufficiently small. Clearly a theory which could deal with both time scales being of comparable order would be ambitious to say the least.

Finally, we have some comments on the system under study. Our general aim was to show how decision errors, which may be common, distort adaptation relative to the training sequence case. To achieve this understanding, we chose, sensibly, the simplest non-trivial system. This system can be used on channels exhibiting limited precursor intersymbol interference and has limited but non-empty practical application.[20] A more general DFE structure requires the use of a linear equalizer as the first stage and its taps may be simultaneously (blindly) adapted with the simpler structure we have studied. This is a different, more complicated problem and our techniques may be partially employed in performing an analysis of this more complex system. We are confident that such an analysis is possible but have made no attempt in this direction. A second comment is that blind adaptation has restricted practical application. In the case of rapidly time-varying channels training sequences would appear to be indispensable and more robust techniques of channel identification, perhaps using coding, etc., would be required.

ACKNOWLEDGEMENTS

This work was partially supported by the Australian Telecommunications and Electronics Research Board, Australian Research Council, ANU Centre for Information Science Research and CRC Centre for Robust and Adaptive Systems.

REFERENCES

1. Qureshi, S. U. H., 'Adaptive equalization', *Proc. IEEE*, **73**, 1349–1387 (1985).
2. Macchi, O., and E. Eweda, 'Convergence analysis of self-adaptive equalizer', *IEEE Trans. Info. Theory*, **IT-30**, 162–176 (1983).
3. Kumar, R., 'Convergence of a decision-directed adaptive equalizer', *Proc. Conf. on Decison and Control*, Long Beach, CA, 1983, IEEE New York, 1983, pp. 1319–1324.
4. Benveniste, A., M. Goursat and G. Ruget 'Robust identification of a nonminimum phase system: blind adjustment of a linear equalizer in data communications', *IEEE Trans. Automatic Control*, **AC-25**, 385–399 (1980).
5. Godard, D. N., 'Self-recovering equalization and carrier tracking in two-dimensional data communication systems', *IEEE Trans. Commun.*, **COM-28** 1867–1875 (1980).
6. Benveniste, A., and M. Goursat, 'Blind equalizers', *IEEE Trans. Commun.*, **COM-32**, 871–883 (1984).
7. Benveniste, A., 'Recursive blind equalizers', *IEEE Trans. Commun.*, **COM-32**, 871–883 (1984).
8. Mazo, J. E., 'Analysis of decision-directed equalizer convergence', *Bell Syst. Tech. J.*, 1857–1876 (1980).
9. Verdú, S., 'On the selection of memoryless adaptive laws for blind equalization in binary communications', *Proc. 6th Int. Conf. on Analysis and Optimization of Systems*, Nice, June 1984, pp. 239–249.
10. Kennedy, R. A., B. D. O. Anderson and R. R. Bitmead, 'Channels leading to rapid error recovery for decision feedback equalizers', *IEEE Trans. Commun*, **COM-37**, 1146–1155 (1989).
11. Sato, Y., 'A method of self-recovering equalization for multi-level amplitude modulation', *IEEE Trans. Commun.*, **COM-23**, 679–682 (1975).
12. Kennedy, R. A., B. D. O. Anderson, Z. Ding and C. R. Johnson, Jr., 'On the failure of proposed recursive identification schemes based on the Sato algorithm and its generalizations', *Proc. Second Int. Symp. on Signal Processing and its Applications, ISSPA '90*, Gold Coast, August 1990, pp. 267–270.
13. Kennedy, R. A., B. D. O. Anderson, Z. Ding and C. R. Johnson Jr., 'Local stable minima of the Sato recursive identification scheme', *Proc. 29th IEEE Conf. on Decision and Control, CDC '90*, Honolulu, HI, December 1990, IEEE, New York, 1990.
14. Foschini, G. J., 'Equalization without altering or detecting data', *AT&T Tech. J.*, **64**, 1885–1911 (1985).
15. Shalvi, O., and E. Weinstein, 'New criteria for blind deconvolution of non-minimum phase systems (channels)', *IEEE Trans. Info. Theory*, **IT-36**, 312–321 (1990).
16. Kennedy, R. A., 'Some difficulties with blind regressor based channel equalizers in digital communications', *Proc. 1989 Aust. Symp. on Signal Processing and Applications, ASSPA '89*, Adelaide, April 1989, pp. 206–210.
17. Kennedy, R. A., B. D. O. Anderson and R. R. Bitmead, 'Stochastic dynamics of blind decision feedback equalizer adaptation', *Proc. IFAC Symp. on Adaptive Svstems in Control and Signal Processing*, Glasgow, April 1989, pp. 65–70.
18. Jennings, A., 'Analysis of the adaption of decision feedback equalizers', *Internal Report*, Telecom Australia Research Laboratories, Clayton Victoria, July 1985.
19. Kennedy, R. A., and B. D. O. Anderson, 'Recovery times of decision feedback equalizers on noiseless channels', *IEEE Trans. Commun.*, **COM-35**, 1012–1021 (1987).
20. Clarke, B. R. 'The time-domain response of minimum phase networks', *IEEE Trans. Cir. Syst.*, **CAS-32**, 1187–1189 (1985).
21. Treichler, J. R., and M. G. Agee, 'A new approach to multipath correction of constant modulus signals', *IEEE Trans. Acoust., Speech, Signal Process.*, **ASSP-31**, 349–472 (1983).
22. Treichler, J. R., and M. G. Larimore, 'New processing techniques based the constant modulus adaptive algorithm', *IEEE Trans. Acoust., Speech, Signal Process.*, **ASSP-33**, 420–431 (1985).
23. Ding, Z., R. A. Kennedy, B. D. O. Anderson and C. R. Johnson Jr., 'Ill-convergence of Godard blind equalizers in data communications', *IEEE Trans. Commun.*, **COM-39**, 1313–1327 (1991).
24. O'Reilly, J. J., and A. M. de Oliveira Duarte, 'Error propagation in decision feedback receivers', *Proc. IEE*, **132**, pt. F, 561–566 (1985).
25. de Oliveira Duarte, A. M., and J. J. O'Reilly, 'Simplified technique for bounding error statistics for DFB receivers', *Proc. IEE*, **132**, pt. F, 567–575 (1985).
26. Bucklew, J. A., T. G. Kurtz and W. A. Sethares, 'Weak convergence and local stability properties of fixed step size recursive algorithms', *IEEE Trans. Inf. Theory*, **IT-30**, 966–978 (1993).
27. Benveniste, A., M. Metivier and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*, Springer, New York, 1990.
28. Friedlin, M. I., and A. D. Wentzell, *Random Perturbations of Dynamical Systems*, Springer, New York, 1984.
29. Kennedy, R. A., G. W. Pulford, B. D. O. Anderson and R. R. Bitmead, 'When has a decision directed equalizer converged?', *IEEE Trans. Commun.*, **COM-37**, 879–884 (1989).