BLIND ADAPTIVE DECISION FEEDBACK EQUALIZATION: A CLASS OF *BAD* CHANNELS

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Raúl Alejandro Casas

May 1996

© 1996 Raúl Alejandro Casas

ALL RIGHTS RESERVED

ABSTRACT

In digital communications, an equalizer that can fully cancel intersymbol interference from a linear dispersive channel with finite time span while not amplifying noise, and that can be implemented with relative ease is bound to be popular. That is the case with the decision feedback equalizer (DFE), a receiver component that has gained the attention of researchers and practitioners for over thirty years. The difficulty in analyzing the DFE, arising from its feedback structure and nonlinearity, has made it hard to asses DFE performance in comparison to other schemes, such as the linear feedforward baud and fractionally-spaced equalizers, and has hindered the development of a comprehensive DFE theory.

One of the major areas lacking understanding of the behavior of a DFE is in the blind adaptation of its equalizer parameters. Due to the inherent complexities of this nonlinear stochastic system, little is known about the convergence properties of adaptive algorithms, like the decision-directed algorithm, when used to blindly update a DFE. This thesis works on expanding the theory on decision-directed DFE (DD-DFE).

Since the literature rarely mentions the initialization of DD-DFE, and the decision-directed algorithm, used in both DFEs and linear feedforward filters, is known to

have convergence problems when there are sufficient symbol estimate errors, this work studies the convergence of DD-DFE from the most natural initialization (when no estimate of the channel dynamics is available): the origin. It is shown that certain channels guarantee that DD-DFE will converge as desired, and as the main result of the thesis, it is shown that certain channels from a dense set, classified as *bad* channels, guarantee that DD-DFE will converge to a bad minimum.

Biographical Sketch

Raúl Alejandro Casas was born in Denver, Colorado on May 23^{rd}, 1972. He lived in Brazil until 1990, when he came for his BS degree in electrical engineering at Cornell University. In 1994, he decided to pursue higher education and study at Cornell for his MS and PhD degrees.

Para papá y mamá

Acknowledgements

By far, my parents deserve my first acknowledgement. I would have to write theses for many lifetimes to match the amount of love and support that they offer me.

When I decided to continue at Cornell for graduate school, many friends and myself were afraid I had made the wrong decision. Well, until now, I do not repent having stayed for one main reason: my advisor Rick Johnson. He has made research a cultural experience, rather than just a technical exercise, by creating a healthy environment among our research group, and providing his students with a diverse spectrum of opportunities from which to learn. That is why he is second on my list.

Next, I must thank my friend and mentor Roberto for his help in my studies and on this thesis. His excellence in mathematics and engineering, as well as his ability and pleasure in teaching should be models for every engineer.

I also acknowledge the students Kenneth Andrews, Greg Conklin, Matthew Shoemake, and Jay Yun, for their insights on the adaptive DFE problem.

I thank my friends Marcia and Chichao for making Ithaca warm during the winter, and my research companions, Jenny, Tom, Fernando, James, John, Dirk, Adnan, Michael, and Rod, for a group to be proud of.

Finally, I would like to show my appreciation to the National Science Foundation and Applied Signal Technology for supporting my graduate education.

Table of Contents

1 Introduction **1**
- 1.1 Preliminary 1
- 1.2 Literature Review 3

2 Blind DFE Algorithms **8**
- 2.1 DFE Structure 8
- 2.2 DD-DFE and CMA-DFE 11
- 2.3 Occurrence of Local Minima 12

3 Bad DD-DFE Channels **19**
- 3.1 Three-Tap Channels 19
  - 3.1.1 $\mathcal{P}_0^H$, the center polytope 19
  - 3.1.2 $\mathcal{G}$, channels with guaranteed convergence 21
  - 3.1.3 $\mathcal{C}$, the set of closed-eye channels 23
  - 3.1.4 $\mathcal{S}$, channels with a bad minimum in $\mathcal{C}$ 26
  - 3.1.5 $\mathcal{B}$, the class of bad channels 28
  - 3.1.6 Example: Some cases of ill-convergence 31
- 3.2 $N+1$ Tap Channels 35
  - 3.2.1 $\mathcal{C}^N$, the corner-class 35
  - 3.2.2 Locating bad $N+1$-tap channels 40
  - 3.2.3 Example: Some channels $\mathcal{C}^4$ 42

4 Observations and Conclusions **44**

Bibliography **50**

List of Figures

2.1 Communication system using DFE 9
2.2 Markov Chain corresponding to polytope $\mathcal{P}_3$ 17
2.3 Cost function for channel $H = (1, 0.8)^t$ 17

3.1 Partition of channel parameter space (three-tap case). 23
3.2 Class of closed eye channels, $\mathcal{C} = \cup_{j=1}^{4} \mathcal{C}_j$, and center polytope $\mathcal{P}_0^H$. . 25
3.3 $\mathcal{S}$ for three-tap channels. 28
3.4 DD-DFE parameter trajectory escape from $\mathcal{P}_0^H$ 29
3.5 Ruling out minima with no guarantee of convergence. 30
3.6 $\mathcal{B}$ for three-tap channels. 31
3.7 Hyperplanes and polytopes for channel $H = (1, 0.9, 0.8)^t$. 32
3.8 DD-DFE cost function for $H = (1, 0.9, 0.8)^t$. 33
3.9 CMA-DFE cost function for $H = (1, 0.9, 0.8)^t$. 33
3.10 DD-DFE cost function for $H = (1, 0.9, -0.8)^t$. 34
3.11 CMA-DFE cost function for $H = (1, 0.9, -0.8)^t$. 34
3.12 One set of five-tap channels in $\mathcal{C}^4$ leading to bad minima in $\mathcal{P}_0^H$ . . . 43

4.1 Divisions of channel parameter space and DD-DFE convergence with initialization at origin 45
4.2 Ill-convergence of DD-DFE for $H = (1, 0.9, 0.8, 0.01)^t$. 46

Chapter 1

Introduction

1.1 Preliminary

> *"Blind adaptation for DFEs is poorly understood, so there exists a dire need to have some results in this subject area."*
>
> Rodney Kennedy, PhD Thesis, 1988

The simplicity and effectiveness of the decision feedback equalizer (DFE) as a receiver component in digital communications has attracted the attention of researchers in the area for over four decades. At the same time, the intrinsic nonlinear and feedback characteristics of the DFE make it extremely difficult to analyze, thus hindering the process of establishing a comprehensive DFE theory. For this reason, throughout these four decades much of the research on decision feedback equalization has been done using the key assumption of *no past decision errors* (which if present, can lead to error propagation), greatly facilitating analysis. This assumption is reasonable in most cases, especially when the DFE is fixed and perfectly

matched to the channel, but when this condition is not met, as is the case with a blind adaptive DFE, the key assumption is no longer valid. As research on the DFE evolves from the joint design of feedforward/feedback filters to adaptive filtering with training and then to blind adaptation through popular algorithms such as decision-direction and the Constant Modulus Algorithm, the need to resort to more complex models of the DFE which account for past decision errors increases. This work concentrates on comprehending the blind adaptation of a DFE through the use of dynamic system models.

When a training sequence and a gradient descent algorithm are used to update a DFE it is possible to obtain a cost function with a unique, global minimum; this is the case with the Least Mean Square algorithm (LMS), which eliminates the problem of past decision errors with a training sequence, and makes analysis of the algorithm's performance simple. When a training sequence is no longer available, decision errors change the shape of the algorithm's cost function, disseminating undesired, or *bad* local minima over equalizer parameter space. In that case, it is crucial to study the effects of initialization in the adaptation of the DFE, since starting an algorithm close to a bad minimum may lead to convergence to an unwanted position. Due to the analytical complexities introduced by considering past decision errors, little is found in the literature on this problem.

The aim of this thesis is to find a class of channels for which the decision feedback equalizer adapted through decision-direction and initialized at the origin converges to an undesired minimum. The first chapter overviews research on the DFE and points out gaps in the literature related to blind adaptation of DFEs. Chapter 2

contains the definitions of the structures and algorithms under consideration and illustrates a few key concepts. Chapter 3 presents our main result, which consists in defining a dense class of channels for which DD-DFE converge to a bad minimum from an initialization at the origin. Chapter 4 concludes the study.

1.2 Literature Review

The decision feedback equalizer was first formally derived by Austin [1], in 1967, as a suboptimal maximum likelihood decision device; it basically consists of a linear baud-spaced feedforward filter followed by a feedback filter. [1] Already in this work, the main features of the DFE were apparent: the DFE does not introduce noise gain and is able to perfectly equalize channels which do not exceed its length. In his derivation, Austin assumed no past decision errors were present in the DFE tapped-delay line, but he did study analytically and through simulations the problem of error propagation, which according to his findings, is only significant at low figures of signal to noise ratio and occurs in bursts. The intervening almost thirty years of work have not advanced significantly beyond these observations on these points.

Shortly after, stemming from [1], several papers defined the areas of research on decision feedback equalization. The immediate ground for investigation was design of DFEs [2]-[11]. Early work studied several design criteria, such as minimum mean squared error (MMSE) derivation for a (baud-spaced) feedforward and decision feedback equalizer [2]. Recent work ([9], [10], [11]) switches from the use of baud-

[1]The DFE is often considered as a joint feedforward and feedback structure. In this thesis, however, we term the feedforward portion as a linear equalizer and label as a "DFE" only the feedback segment.

spaced filters to the application of fractionally-spaced feedforward equalizers (FSEs). Furthermore, much research is being done on the hybridization of feedforward and feedback equalization with coding techniques ([27]-[30]).

Due to the concern with the performance of the DFE, specifically the effects of error propagation, literature swelled in an attempt to establish bounds on error probabilities and to classify channels for which a DFE may or may not behave as desired ([18]-[26]). Such an approach typically involves modeling the dynamics of a DFE through a Markov chain so that the steady state statistics of the chain can be used to estimate quantities such as the probability of making a decision error, or the mean time for recovering from an error burst. Unfortunately, the amount of work necessary to perform these calculations increases exponentially with the length of the channel. In [18], Duttweiler *et al.* managed to reduce this computational burden by manipulating the error mechanism model in order to find upper bounds on error probabilities. Work in [20], [21], [24] and [26] improved these bounds and extended the theory to cover several types of input and noise.

From a slightly different angle, [19] found a class of channels for which the DFE is *stable*, meaning that it is able to flush out past decision errors in a finite time. Similar work was done in [22] and [23], where it was shown that exponentially decaying, and, more generally, strictly positive real (SPR) channels possess this quality. These concepts are akin to the idea of classifying channels for which an adaptive DFE will converge or not, since they provide *intuition* into which types of channels the DFE may exhibit problematic behavior.

Stepping up in the complexity involved in the analysis of a DFE, we consider

the adaptive DFE. Monsen's paper on the MMSE design of a DFE [2] includes the development of a blind adaptive algorithm using past estimates to update the equalizer so as to converge to the MMSE design. For the algorithm to (always) converge, however, he had to assume a small probability of estimate errors. George *et al.* presented a similar algorithm in [12] and performed simulations that suggested that the DFE reached a MMSE solution faster than a feedforward equalizer. Unlike Monsen, they mentioned where the adaptive DFE should be initialized: at the origin (i.e. all taps are set to zero). That is actually the most logical place to start the equalizer if there is no previous channel estimate, since if there is no intersymbol interference, the optimum DFE setting would be all zeros. But, since past decision errors could lead to error propagation, there is no guarantee that the algorithm will converge to the desired solution from this initialization.

The idea of using past symbol estimates to blindly update an equalizer through a gradient descent method [2] was introduced by Lucky in [39] for the adaptive linear equalizer problem, and is usually called decision-direction, or the decision-directed algorithm (DD). Lucky was aware that it was crucial to have a sufficiently low probability of decision errors for the algorithm to converge to the desired MMSE solution. This situation is analogous to initializing the equalizer close enough to the solution to obtain good convergence. The problem of convergence of a decision-directed equalizer was analyzed in depth in three works: initially by Mazo [31], and then by Kumar [32] and Macchi and Eweda [33]. Mazo explained how the decision device tessellates equalizer parameter space into polytopes, allowing local minima in the

[2] Using the Least Mean Squares (LMS) algorithm by substituting the training sequence by a sequence of past decisions.

cost function to appear within each polytope, and presented examples of channels for which the algorithm converged to a bad local minimum from multiple initializations over channel-parameter space. Moreover, he studied how the algorithm could escape from a local bad minimum due to the use of a finite step-size. [32] and [33] deal with these issues in addition to other notions such as the stability of the decision directed algorithm around a desired minimum.

Using similar ideas, Kennedy *et al.* exposed the mechanism of convergence of blind adaptive decision feedback equalizers in [14]. Just as in the feedforward case, the decision device partitions equalizer parameter space into polytopes where the cost function may have undesired minima. This implies that neglecting initialization of a blind adaptive decision feedback equalizer algorithm, such as the decision-directed algorithm (DD-DFE), may lead to ill-convergence. In addition to DD-DFE, the paper also suggests the use of the Constant Modulus Algorithm [35], [36], for the blind update of a DFE.

Three recent conference papers are closely related to the work in [14]. The first one, [15], considers the use of soft decisions to blindly update a DFE in an attempt to *smooth out* regions with undesired minima. Clearly, this paper shows concern with respect to the problem of a multimodal cost function. [16], on the other hand, expects that CMA used to update the DFE will probably be able to open the eye in the presence of noise and a sufficiently large step-size. The paper also shows that under the key assumption of no past decision errors, and under certain length and channel zero conditions, CMA will actually converge to a perfect equalization setting. Clearly, this is an example of work that does not consider bad local minima

a serious problem.

The third paper, [17], which is the basis of this thesis, attacks the problem of potential ill-convergence of a blind DFE and considers the use of CMA as an alternative blind adaptive algorithm, which may or may not have better convergence properties than DD-DFE. This thesis extends the main result of the paper [17], which finds a class of three-tap channels for which DD-DFE converges to a bad minimum when started at the origin, to the $N + 1$-tap case. The main argument for choosing this initialization, as explained before, is that if the channel has close to no intersymbol interference (ISI) a DFE initialized at zero will begin close to the solution. If the channel is closed eye, the origin, as shown in [17], is probably not the best place for initialization, but something about the channel must be known for a more intelligent guess.

If the interest in the blind adaptation of DFEs continues to grow, the theory in this area should expand by covering more classes of channels, initializations and blind adaptive algorithms that either guarantee success or failure.

Chapter 2

Blind DFE Algorithms

The aim of this chapter is to provide background on the mechanisms involved in the blind update of a decision feedback equalizer. The first section defines the DFE structure under study and carefully lays down a set of assumptions about the communication system that aid to limit and focus the research problem. The second section introduces the stochastic system models used to analyze an adaptive DFE. [1]

2.1 DFE Structure

We are interested in the DFE structure shown in figure 2.1. A sequence of symbols $\{a_k\}$ drawn from an alphabet $\mathcal{A}$ is transmitted over a linear dispersive channel, so that its output is given by

$$y_k = \sum_{i=0}^{\infty} h_i a_{k-i} \tag{2.1}$$

[1] For a more complete description of theses mechanisms refer to [14]

where $h_i \in \mathbf{R}$ are the taps of the channel. The output of the DFE is a sequence of estimates, $\{\widehat{a}_k\}$, from a decision device. These estimates, or *past decisions* are fed back through the tapped-delay line and the result is subtracted from the channel output

$$z_k \quad = \quad y_k - \sum_{i=1}^{N} d_i \widehat{a}_{k-i} \tag{2.2}$$

where the DFE line has N-taps $d_i \in \mathbf{R}$.

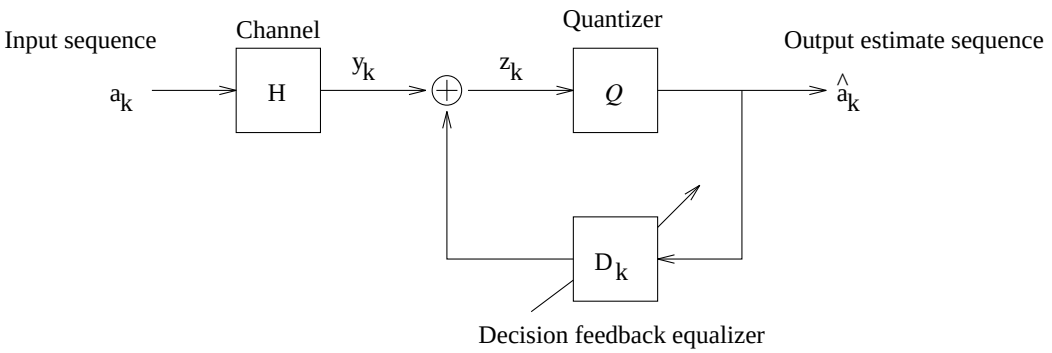

Figure 2.1: Communication system using DFE

For the purpose of simplifying the analysis, we make the following assumptions:

- **Assumption 1:** The source alphabet $\mathcal{A}$ is $\{\pm 1\}$ (BPSK) and the source symbols are independent and have equal probability of occurring.

  As we will see later, the decision device partitions parameter space into polytopes with different dynamics. A BSPK source, which requires a signum *function to make decisions, partitions parameter space into hypercubes, which are quite simple to view and understand. Even though other modulation schemes, such as Pulse Amplitude Modulation (PAM), are more used in practice, they*

partition parameter space into radically more complex polytopes, making the problem very difficult. The now popular multidimensional modulation schemes, such as Quadrature Amplitude Modulation (QAM), would make the problem hellish. BPSK results, however, may show a best case behavior; since there is always at least a 50% chance of correctly estimating a symbol, channel which indicate signs of ill-convergence for BPSK could offer severe problems with higher order constellations.

- **Assumption 2:** The channel is real (i.e. $h_i \in \mathbf{R} \quad \forall i$) and (linear) finite impulse response (FIR).

  Since BPSK is used the channel must be real. Again, due to the fact that parameter space space is partitioned into polytopes which depend on the channel and DFE, an infinite impulse response channel would imply an infinite number of polytopes. It would be possible, however, to undermodel the DFE and consider the contribution of remaining taps of the channel not matched by the equalizer as a noise term. This noise term could then be approximated as a white Gaussian disturbance. We do not attempt, however, to work out these details.

- **Assumption 3:** The channel has no precursor (i.e. $h_0 = 1$ and $|h_i| < 1$ for $i = 1, 2, \cdots$).

  We will later see that [14] has already studied the class of channels with precursor. One other argument for working in this restricted class of channels is that by adding a feedforward filter before the DFE one can mitigate the effect

of the precursor, so that the effective channel-equalizer response falls within this class.

- **Assumption 4:** The channel introduces no noise.

  Noise smoothes the boundaries of polytopes so that a global analysis of parameter space is required, not only within separate polytopes. Adding noise to the problem leads to a challenging stochastic dynamical system problem that is beyond the scope of this thesis.

2.2 DD-DFE and CMA-DFE

The focus of this thesis is on the occurrence of certain local minima in the cost function of the blind, adaptive, decision-directed DFE (DD-DFE). This algorithm minimizes the cost function

$$J_{DD} = E\{\frac{1}{2}(z_k - \hat{a}_k)^2\} \tag{2.3}$$

through the steepest descent algorithm

$$D_{k+1} = D_k + \mu(z_k - \hat{a}_k)\hat{A}_k \tag{2.4}$$

where the N-tap equalizer is given by $D = (d_1, d_2, \cdots, d_N)^t$ and the regressor of past decisions is defined as $\hat{A}_k = (\hat{a}_{k-1}, \hat{a}_{k-2}, \cdots, \hat{a}_{k-N})^t$.

We also briefly consider the use the popular Constant Modulus Algorithm for the update of a DFE (CMA-DFE). This algorithm has as its cost function

$$J_{CMA} = E\{\frac{1}{4}(z_k^2 - \gamma)^2\} \tag{2.5}$$

where $\gamma = E\{a_k^4\}/E\{a_k^2\}^2$ ($\gamma = 1$ for a BPSK source). The resulting CMA-DFE algorithm is

$$D_{k+1} = D_k + \mu(\gamma - z_k^2)z_k\widehat{A}_k. \tag{2.6}$$

2.3 Occurrence of Local Minima

[14] examines the structures underlying an adaptive DFE, and studies the *gross* convergence of the Decision Directed algorithm. The paper classifies local minima as equilibria resulting from either the DFE structure (termed *delay-type* minima) or from the adaptive algorithm itself. Throughout this thesis, we closely follow the analysis in [14]. Delay-type minima arise from channels with a precursor; for some delay δ, if

$$|h_\delta| - \sum_{i=0}^{\delta-1} |h_i| > 0 \tag{2.7}$$

the adaptive DFE may converge to the setting $D = (h_{\delta+1}, \cdots, h_N, 0, \cdots)^t$. Since [14] develops a full theory on delay-type minima, we choose to concentrate on minima resulting from the adaptive algorithm (therefore minima exposing the weakness of the algorithm).

Thus, we look at the case when the channel has no precursor (i.e. $h_0 = 1$ and $|h_i| < 1$ for $i = 1, 2, ...$). Consider the case where BPSK is being transmitted over a channel with $N + 1$ taps. The decision device is a *signum* function yielding

$$\begin{aligned} \widehat{a}_k &= sgn(\sum_{i=0}^{N} h_i a_{k-i} - \sum_{i=1}^{N} d_i(k)\widehat{a}_{k-i}) \\ &= sgn(A_k^t H - \widehat{A}_k^t D_k) \end{aligned} \tag{2.8}$$

where $A_k = (a_k, a_{k-1}, \cdots, a_{k-N})^t$ and $H = (h_0, h_1, \cdots, h_N)^t$. This non-linear function tessellates parameter space into polytopes with hyperplanes

$$\mathcal{H} \quad = \quad \{D_k \in \mathbf{R}^N : A_k^t H - \widehat{A}_k^t D_k = 0\} \tag{2.9}$$

where A_k and $\widehat{A}_k$ vary through all possible combinations of $+1$ and -1.

Define as a state of past symbols and estimates

$$X_k = (a_{k-1}, \cdots, a_{k-N} | \widehat{a}_{k-1}, \cdots, \widehat{a}_{k-N}). \tag{2.10}$$

Within each polytope, this random variable defines a Finite State Markov Process (FSMP) with transitions given by the input symbol a_k. Notice that the FSMP, or Markov chain, will be the same for any DFE setting inside any one polytope.

The Markov chain corresponding to each polytope characterizes the convergence of the adaptive algorithm. Some observations and assumptions follow:

- **Assumption 5:** The step-size is sufficiently small for the Markov chain to have time to move to its steady-state before the algorithm travels significantly within a particular polytope. The steady-state statistics of the Markov chain are valid for any DFE setting within the polytope and are thus said to be *piecewise constant* statistics.

- The steady-state statistics of the Markov chain will rule the average behavior of the adaptive DFE. In fact, the statistics determine the shape of the cost function inside each polytope (notice the expectation operator in (2.3) and (2.5) acting on terms such as $a_{k-1}\widehat{a}_{k-1}$).

- It is possible that a Markov chain has multiple steady-state behaviors. In that case, the cost function may have ambiguous shapes (one shape for each steady-state).

As seen, a single polytope can be thought of as an independent domain where convergence of an adaptive algorithm may be drastically different from another polytope. The main difficulty encountered by an adaptive algorithm is that its cost function may have a minimum inside a polytope not enclosing the desired solution, so that it may get trapped at that point and constantly produce decision errors when another setting would produce none.

Given the statistics of the FSMP inside a polytope it is possible to find the minimum for DD-DFE. Recall the DD-DFE cost function (now defined inside some polytope $\mathcal{P}$)

$$J_{DD}(\mathcal{P}) \;=\; E\{\frac{1}{2}(A_k^t H - \widehat{A}_k^t D - \widehat{a}_k)^2\}. \tag{2.11}$$

By setting its differential (well defined inside a polytope)

$$d_D J_{DD}(\mathcal{P}) \;=\; -E\{\widehat{A}_k A_k^t\}H + E\{\widehat{A}_k \widehat{A}_k^t\}D_k + E\{\widehat{a}_k \widehat{A}_k\} \tag{2.12}$$

to zero we find the possible local minimum

$$D^\star \;=\; E\{\widehat{A}_k \widehat{A}_k^t\}^{-1}[E\{\widehat{A}_k A_k^t\}H - E\{\widehat{A}_k \widehat{a}_k\}] \tag{2.13}$$

For each polytope, the quantities $E\{\cdot\}$ will vary, so that we define $R(\mathcal{P}) = E\{\widehat{A}_k \widehat{A}_k^t\}$, $C(\mathcal{P}) = E\{\widehat{A}_k A_k^t\}$, and $T(\mathcal{P}) = E\{\widehat{A}_k \widehat{a}_k\}$, resulting in

$$D^\star(\mathcal{P}) = R(\mathcal{P})^{-1}[C(\mathcal{P})H - T(\mathcal{P})]. \tag{2.14}$$

$D^\star(\mathcal{P})$ is a local minimum only if $R(\mathcal{P})$ (the Hessian at that point) is positive definite. Notice that there is only one possible local minimum for each polytope since the DD-DFE is quadratic inside each polytope. The overall cost function is said to be *piecewise quadratic*.

For CMA-DFE it is not (yet) possible to find a similar direct solution. Expanding the CMA-DFE cost function inside a polytope

$$\begin{aligned} J_{CMA}(\mathcal{P}) &= E\{\frac{1}{4}((A_k^t H - \widehat{A}_k^t D)^2 - \gamma)^2\} && (2.15) \\ &= E\{\frac{1}{4}(H^t A_k A_k^t H - H A_k^t \widehat{A}_k^t D && (2.16) \\ &\quad -D^t \widehat{A}_k A_k^t H - D^t \widehat{A}_k \widehat{A}_k^t D \\ &\quad -\gamma)^2\} \end{aligned}$$

shows that it is difficult to express and compute cross terms such as

$$E\{H^t A_k A_k^t H H^t A_k A_k^t H\} \tag{2.17}$$

and it is not possible to find a direct solution to a quartic equation for general channel and equalizer parameters.

Example: 2-Tap Channel

These concepts are made clearer through an example. Take the 2-tap channel $H = (1, 0.8)^t$. The hyperplanes tesselating parameter space into polytopes are given by

$$\mathcal{H} = \{D_k \in \mathbf{R}^N : d_1(k)\widehat{a}_{k-1} = a_k + 0.8a_{k-1}\} \tag{2.18}$$

(since there is only one parameter, hyperplanes are simply a point). There are five polytopes within the boundaries created by the hyperplanes (i.e. $\mathcal{P}_1 = \{D_k \in$

$\mathbf{R}^N n : d_1(k) < -1.8\}$, etc.). As explained before, we only deal with channels with no precursor, so that we need only consider the region $-1 < d_1 < 1$, which intersects polytopes $\mathcal{P}_0, \mathcal{P}_2$ and $\mathcal{P}_3$. The desired solution is in $\mathcal{P}_3$, for $d_1 = 0.8$.

In order to find the DD-DFE cost function for this channel, we must first find the Markov chain associated for each polytope using a transition table. The table below shows the transitions occurring in polytope $\mathcal{P}_3$.

Current State	*Next State*	
$X_k = (a_k, \widehat{a}_k)$	$X_{k-1} = (a_{k-1}, \widehat{a}_{k-1})$	
	$a_k = 1$	$a_k = -1$
(-1,-1)	(1,1)	(-1,-1)
(-1,1)	(1,-1)	(-1,-1)
(1,-1)	(1,1)	(-1,1)
(1,1)	(1,1)	(-1,-1)

The Markov chain of polytope $\mathcal{P}_3$ is shown in figure 2.2. Notice that the steady state for that chain is formed by only two states, $(1,1)$ and $(-1,-1)$; this is true for any d_1 in $\mathcal{P}_3$.

The steady state of the Markov chain provides sufficient information for gathering the statistics of the polytope. $\mathcal{P}_3$ has

$$R(\mathcal{P}_3) = E\{\widehat{a}_{k-1}\widehat{a}_{k-1}\} = 1 \tag{2.19}$$

$$C(\mathcal{P}_3) = E\{\widehat{a}_{k-1}(a_k, a_{k-1})\} = (0,1) \tag{2.20}$$

$$T(\mathcal{P}_3) = E\{\widehat{a}_k\widehat{a}_{k-1}\} = 0 \tag{2.21}$$

so that $d_1^\star(\mathcal{P}_3) = 0.8$. Continuing with finding the steady state for each polytope

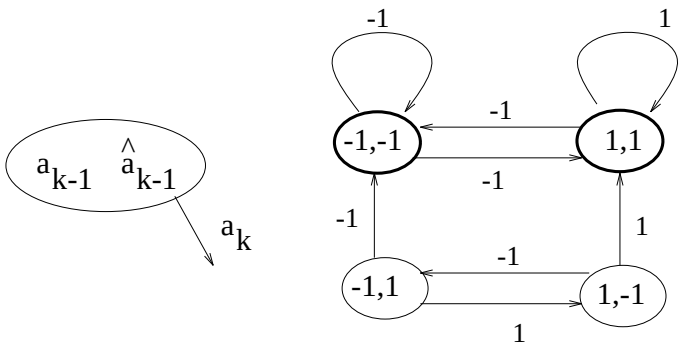

Figure 2.2: Markov Chain corresponding to polytope $\mathcal{P}_3$

and computing statistics we find the cost function in figure 2.3.

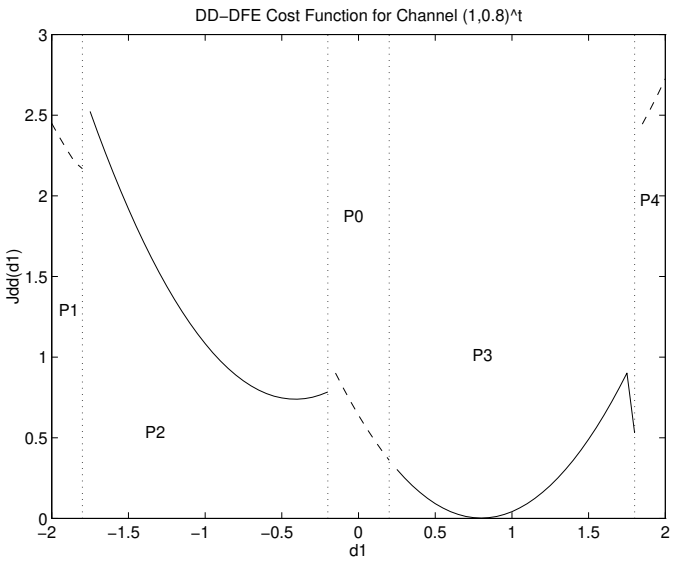

Figure 2.3: Cost function for channel $H = (1, 0.8)^t$

It is interesting to note that even this simple example has a bad local minimum, namely at $d_1 = -0.46$ (in $\mathcal{P}_2$, $R(\mathcal{P}_2) = 1$, $C(\mathcal{P}_2) = (0, 0.3)$ and $T(\mathcal{P}_2) = 0.7$). Imagine that the algorithm were initialized anywhere in $\mathcal{P}_2$. Then the algorithm would travel down the quadratic until it hit $d_1 = -0.4$ and stop, rather than con-

tinuing to the desired minimum, $d_1 = 0.8$ (assuming a sufficiently small step-size). However, for this example, initialization at the origin (i.e. $d_1 = 0$) would result in local convergence to the global minimum. As we will see in the next chapter, we will not always be so fortunate.

Chapter 3

Bad DD-DFE Channels

Using our ability to locate DD-DFE minima associated with a polytope, in this chapter we find a class of three-tap channels for which DD-DFE will not converge to the desired solution when initialized at the origin. For these channels DD-DFE will converge to an undesired solution in the polytope centered at the origin. Through simulations, we verify this result for a couple of channels and briefly compare the behavior of DD-DFE with CMA-DFE. Finally, we develop the result for three-tap channels into the general $N+1$-tap channel case.

3.1 Three-Tap Channels

3.1.1 $\mathcal{P}_0^H$, the center polytope

Let us begin by paying some attention to the initialization at the origin. Recall that the *signum* function partitions parameter space into polytopes divided by hy-

perplanes. For the three-tap channel case (i.e. $H = (h_0, h_1, h_2)^t$), [1] these borders are given by

$$\begin{aligned} \mathcal{H} &= \{D = (d_1, d_2)^t \in \mathbf{R}^2 : \\ &\quad a_k h_0 + a_{k-1} h_1 + a_{k-2} h_2 = \hat{a}_{k-1} d_1 + \hat{a}_{k-2} d_2\}. \end{aligned} \tag{3.1}$$

Define

$$\tau = \min_{a_{k-j}} |a_k h_0 + a_{k-1} h_1 + a_{k-2} h_2| \tag{3.2}$$

as the minimum magnitude combination of h_0, h_1, and h_2. Notice that the combinations of these taps are the d_1-axis (x-axis) and d_2-axis (y-axis) intercepts for the hyperplanes (lines) in (3.2). There exists a polytope centered around the origin defined by

$$\mathcal{P}_0^H = \{(d_1, d_2)^t \in \mathbf{R}^2 : |d_1| + |d_2| < \tau\} \tag{3.3}$$

(refer to figure 3.2). This polytope, which we call $\mathcal{P}_0^H$, encloses the initial DFE parameter setting $(0, 0)^t$. DD-DFE may behave in several manners given this specific initialization:

- If the center polytope also encloses a minimum of the DD-DFE cost function (i.e. $D^\star(\mathcal{P}_0^H) = (d_1, d_2)^t \in \mathcal{P}_0^H$) and it is the desired minimum (i.e. $D^\star(\mathcal{P}_0^H) = (h_1, h_2)^t$) then DD-DFE may converge as desired.

- If $\mathcal{P}_0^H$ does not enclose the desired solution, but does have a bad minimum (i.e. $D^\star(\mathcal{P}_0^H) \neq (h_1, h_2)^t$, $D^\star(\mathcal{P}_0^H) \in \mathcal{P}_0^H$), then DD-DFE may converge to this unwanted location.

[1] From now on, for better notation, we will omit the first tap h_0, which is always unity, and define a three-tap channel as $H = (h_1, h_2)^t$.

- If there is no DD-DFE cost function minimum in $\mathcal{P}_0^H$ (i.e. $D^\star(\mathcal{P}_0^H) \notin \mathcal{P}_0^H$), the algorithm will depart the polytope and further analysis must be done on the parameter trajectories to predict good or bad convergence.

Thus, convergence is highly dependent on the channel and how it partitions parameter space into polytopes.

3.1.2 $\mathcal{G}$, channels with guaranteed convergence

In order to obtain a better grasp of the types of channels being dealt with, we classify channel-parameter space into multiple regions. The first two regions are the set of open-eye channels

$$\mathcal{O} = \{(h_1, h_2)^t \in \mathbf{R}^2 : |h_1| + |h_2| < 1\}. \tag{3.4}$$

and the set of closed-eye channels

$$\mathcal{C} = \{(h_1, h_2)^t \in \mathbf{R}^2 : |h_1| + |h_2| > 1, |h_1| < 1, |h_2| < 1\}. \tag{3.5}$$

Let us first look at the set of open-eye channels, $\mathcal{O}$. For channels $H \in \mathcal{O}$, DD-DFE has a very good chance of converging to the desired solution. Starting the algorithm at the origin guarantees that no past decision errors occur at that DFE setting, and therefore no decision errors occur at any DFE setting within $\mathcal{P}_0^H$. Thus, the statistics are

$$\begin{aligned} R(\mathcal{P}_0^H) &= E\{\widehat{A}_k \widehat{A}_k^t\} = I_{2\times 2} \\ C(\mathcal{P}_0^H) &= E\{\widehat{A}_k A_k^t\} = [0_{2\times 1} \;\; I_{2\times 1}] \\ T(\mathcal{P}_0^H) &= E\{\widehat{A}_k \widehat{a}_k\} = 0_{2\times 1} \end{aligned}$$

so that the minimum associated with $\mathcal{P}_0^H$ is in fact the desired minimum, $D^\star(\mathcal{P}_0^H) = H$, which may or may not be enclosed in $\mathcal{P}_0^H$. Thus, while in $\mathcal{P}_0^H$, DD-DFE will converge to the desired minimum. When the parameter trajectory departs $\mathcal{P}_0^H$, however, convergence it not guaranteed but expected, since open-eye channels are *well behaved* and because the equalizer taps will be closer to the correct solution.

It is possible to find a subset $\mathcal{G} \subset \mathcal{O}$ of the open-eye channels for which DD-DFE will always converge. For any $H \in \mathcal{O}$, $\tau = 1 - |h_1| - |h_2|$, so that the center polytope is defined as

$$\mathcal{P}_0^H = \{(d_1, d_2)^t \in \mathbf{R}^2 : |d_1| + |d_2| < 1 - |h_1| - |h_2|\}. \tag{3.6}$$

Using this, we define

$$\mathcal{G} = \{(h_1, h_2)^t \in \mathbf{R}^2 : |h_1| + |h_2| < 1 - |h_1| - |h_2|\} \tag{3.7}$$

$$= \{(h_1, h_2)^t \in \mathbf{R}^2 : |h_1| + |h_2| < \frac{1}{2}|\} \tag{3.8}$$

as the class of channels which lie inside the center polytope $\mathcal{P}_0^H$. $\mathcal{G}$ also has the property that DD-DFE will always converge to the desired solution for any $H \in \mathcal{G}$. That is because the channel is contained in the center polytope, and because the quadratic DD-DFE cost function is symmetric, [2] so that every trajectory to the solution is a straight line, and the adapting equalizer parameters cannot leave $\mathcal{P}_0^H$.

Let us summarize our results until now (refer to figure 3.1). We have separated parameter space into the region of open-eye channels, $\mathcal{O}$, and closed-eye channels, $\mathcal{C}$. For channels $H \in \mathcal{O}$, DD-DFE has a very good chance of converging; our argument

[2] Since there are no decision errors and $R(\mathcal{P}_0^H)$ and $C(\mathcal{P}_0^H)$ are trivial, the problem becomes a channel identification problem using LMS. Analysis shows that the LMS cost function is quadratic and symmetric.

was simply intuitive. For a subset of $\mathcal{O}$, specifically the set $\mathcal{G}$, however, we were able to show guaranteed convergence.

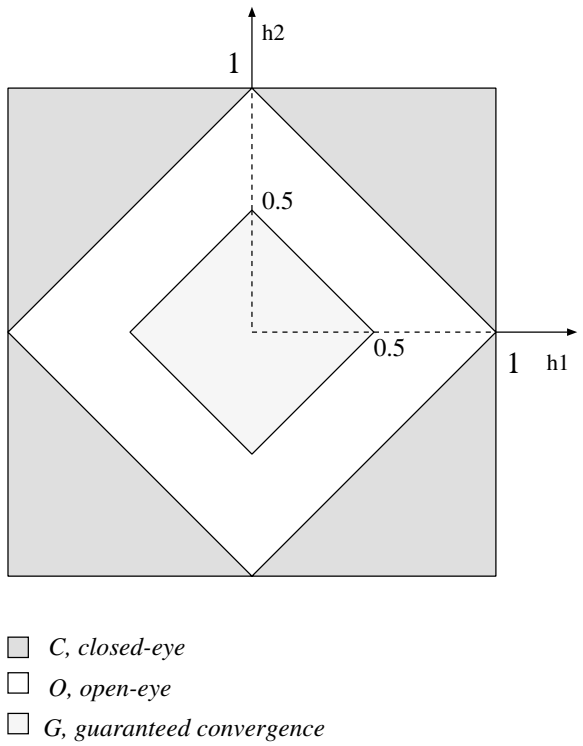

Figure 3.1: Partition of channel parameter space (three-tap case).

Now we will study the set of closed-eye channels.

3.1.3 $\mathcal{C}$, the set of closed-eye channels

For the three-tap channel case, it happens that the set of *closed-eye* channels does not intersect $\mathcal{P}_0^H$, so that any closed-eye channel is a candidate for bad convergence. We conduct our search for *bad* (three-tap) DFE channels by studying the class of closed-eye channels. For this class, $\tau = |h_1| + |h_2| - 1$, so that

$$\mathcal{P}_0^H \quad = \quad \{(d_1, d_2)^t \in \mathbf{R}^2 : |d_1| + |d_2| < |h_1| + |h_2| - 1\}. \tag{3.9}$$

We begin by showing that if a channel is closed-eye, it will not be contained in the center polytope, $\mathcal{P}_0^H$.

Lemma 1:

$$\mathcal{C} \cap \mathcal{P}_0^H = \emptyset. \tag{3.10}$$

Proof: Assume the desired DFE setting $D^\star(\mathcal{P}_0^H) = (d_1, d_2)^t = (h_1, h_2)^t \in \mathcal{P}_0^H$. Then $|d_1| + |d_2| = |h_1| + |h_2| < |h_1| + |h_2| - 1$, which is not possible. □

This lemma guarantees that if there is a local minimum in $\mathcal{P}_0^H$ it will not be the desired minimum (recall that the desired minimum is always $D = (h_1, h_2)^t$.

Given this notion, we proceed to find the minimum $D^\star(\mathcal{P}_0^H)$ for each $H = (h_1, h_2)^t \in \mathcal{C}$ and to check if it is attainable (i.e. $D^\star(\mathcal{P}_0^H) \in \mathcal{P}_0^H$). Hence, we need the statistics $R(\mathcal{P}_0^H), C(\mathcal{P}_0^H), T(\mathcal{P}_0^H)$. To find these statistics we use the next lemma.

Lemma 2: *Divide $\mathcal{C}$ into the four regions (refer to figure 3.2)*

$$\mathcal{C}_1 = \{H \in \mathcal{C} : h_1 > 0, h_2 > 0\} \tag{3.11}$$

$$\mathcal{C}_2 = \{H \in \mathcal{C} : h_1 > 0, h_2 < 0\} \tag{3.12}$$

$$\mathcal{C}_3 = \{H \in \mathcal{C} : h_1 < 0, h_2 < 0\} \tag{3.13}$$

$$\mathcal{C}_4 = \{H \in \mathcal{C} : h_1 < 0, h_2 > 0\}. \tag{3.14}$$

For all H in some $\mathcal{C}_j$, and for all $D \in \mathcal{P}_0^H$, the FSMP formed by

$$X_k = (a_{k-1}, a_{k-2} | \widehat{a}_{k-1}, \widehat{a}_{k-2}) \tag{3.15}$$

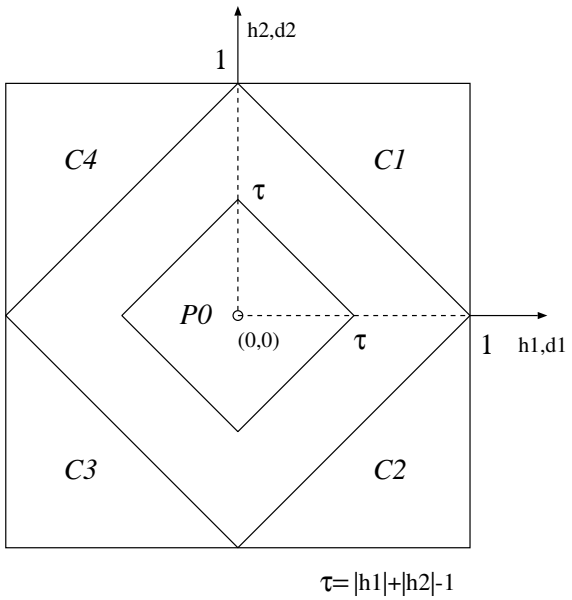

Figure 3.2: Class of closed eye channels, $\mathcal{C} = \cup_{j=1}^{4} \mathcal{C}_j$, and center polytope $\mathcal{P}_0^H$.

has transition probabilities $P(X_k|X_{k-1})$ that do not depend on the specific H or D, but depend on which class, $\mathcal{C}_j$, H belongs to.

Proof: Recall that for any DFE setting D within a polytope, the behavior of the Markov process is the same so that

$$\widehat{a}_k(D_a) = \widehat{a}_k(D_b) \;\; \forall D_a, D_b \in \mathcal{P}_0^H. \tag{3.16}$$

Therefore, w.l.o.g. we can assume $D_a = (0,0)^t$ and write

$$\widehat{a}_k = sgn(a_k + h_1 a_{k-1} + h_2 a_{k-2}). \tag{3.17}$$

For all $H \in \mathcal{C}_1$, we find

$$\widehat{a}_k \;\; = \;\; \begin{cases} a_k & \text{if } \; a_k = \widehat{a}_{k-1} = \widehat{a}_{k-2} \;\; \text{or} \\ & \widehat{a}_{k-1} = -\widehat{a}_{k-2} \\ -a_k & \text{if } \; -a_k = \widehat{a}_{k-1} = \widehat{a}_{k-2} \end{cases} \tag{3.18}$$

and conclude that $\widehat{a}_k$ is also independent of $H \in \mathcal{C}_1$. Using the same argument we find that the random process $\{\widehat{a}_k\}$ only depends of whether H is in $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$ or $\mathcal{C}_4$ for all $D \in \mathcal{P}_0^H$. □

This result tells us that for any channel in one of the four classes $\mathcal{C}_j$ of $\mathcal{C}$ there is a unique set of statistics ruling the adaptation of DD-DFE in $\mathcal{P}_0^H$. Given a set of statistics, we are ready to find which channels have a minimum in $\mathcal{P}_0^H$ and which do not.

3.1.4 $\mathcal{S}$, channels with a bad minimum in $\mathcal{C}$

Using the following proposition, we define the set of channels, $\mathcal{S}$, for which DD-DFE has a bad minimum in $\mathcal{P}_0^H$:

Proposition 1: *For all $H \in \mathcal{C}$,*

$$\begin{aligned} D^\star(\mathcal{P}_0^H) \in \mathcal{P}_0^H \quad \Longleftrightarrow \quad & H \in \mathcal{S} \quad \textit{and} \\ & x^t R(\mathcal{P}_0^H) x > 0 \quad \forall x \in \mathbf{C}^2 \end{aligned} \tag{3.19}$$

where $\mathcal{S} = \cup_{j=1}^4 \mathcal{S}_j$ and

$$\begin{aligned} \mathcal{S}_j \;=\; & \{H = (h_1, h_2)^t \in \mathcal{C}_j : \\ & |a h_1 + b h_2 + c| + |d h_1 + e h_2 + f| \\ & < |h_1| + |h_2| - 1\} \end{aligned} \tag{3.20}$$

and a, b, c, d, e, and f are given by (2.14) as

$$D^\star(\mathcal{P}_0^H) \;=\; \begin{bmatrix} a & b \\ d & e \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} + \begin{bmatrix} c \\ f \end{bmatrix} \tag{3.21}$$

Proof: By lemma 2 we know that for any $H \in \mathcal{C}_j$, $R(\mathcal{P}_0^H)$, $C(\mathcal{P}_0^H)$ and $T(\mathcal{P}_0^H)$ will be the same. In order for $D^\star(\mathcal{P}_0^H) = (d_1, d_2)^t$ to be in $\mathcal{P}_0^H$ we need

$$|d_1| + |d_2| < |h_1| + |h_2| - 1. \tag{3.22}$$

Since for each $\mathcal{C}_j$ we can write d_1 and d_2 as in (3.21) then inequality (3.22) translates into the inequality in (3.20). Finally, in order for $D^\star(\mathcal{P}_0^H)$ to be a minimum, the Hessian of J_{DD} must be positive definite, or in other words, for any $x \in \mathbf{C}^2$, $x^t R(\mathcal{P}_0^H) x > 0$. □

Using these ideas, we find the regions $\mathcal{S}_j$, $1 \leq j \leq 4$ for which DD-DFE has a bad minimum in $\mathcal{P}_0^H$ for any $H \in \mathcal{S}_j$ and plot them by taking the inequality in proposition 1, setting up a function

$$g(h_1, h_2) \;=\; |ah_1 + bh_2 + c| + |dh_1 + eh_2 + f| - |h_1| - |h_2| + 1. \tag{3.23}$$

and drawing contours of $g(h_1, h_2) < 0$. Refer to figure 3.3, and notice the four regions C_j used in lemma 2 (i.e. C_1 is the upper right hand triangle, etc.). Within each $\mathcal{C}_j$, the corresponding region of channels, $\mathcal{S}_j$, is covered by contour lines.

For any channel in $\mathcal{S}$ we can expect DD-DFE converging to a bad minimum in $\mathcal{P}_0^H$, but ill-convergence is not guaranteed. Imagine that the parabolic shape of the DD-DFE cost function (given by the statistics $R(\mathcal{P}_0^H)$, $C(\mathcal{P}_0^H)$ and $T(\mathcal{P}_0^H)$) degenerates to an almost hyperbolic shape, as shown in the contours in figure 3.4, so that the algorithm's trajectory manages to escape $\mathcal{P}_0^H$. Once the algorithm quits the center polytope, it can either return to it and converge to the bad minimum, or

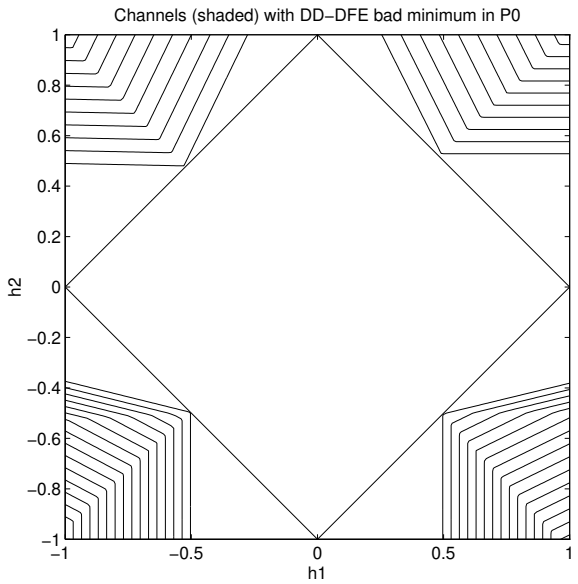

Figure 3.3: $\mathcal{S}$ for three-tap channels.

continue to other polytopes until it either stops at another bad minimum, or lands on the desired minimum.

That is why we find a class of channels, $\mathcal{B}$ for which DD-DFE will definitely converge to a bad position.

3.1.5 $\mathcal{B}$, the class of bad channels

As seen in figure 3.4, if the DFE parameter trajectory exits the center polytope, ill-convergence is not guaranteed. Since DD-DFE trajectories are bounded by a limiting L-shaped trajectory, where the parameters travel down the direction of steepest descent (the fast mode) and then across the shallow direction (the slow mode) to meet the bad minimum, we can find by construction the set of bad minima for which DD-DFE would depart $\mathcal{P}_0^H$ if the cost function had the degenerate limiting hyper-

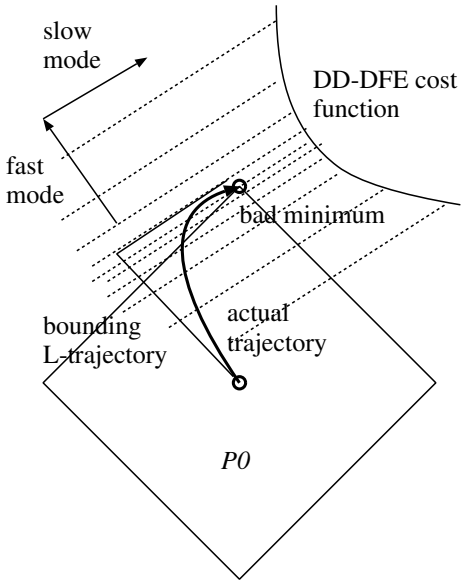

Figure 3.4: DD-DFE parameter trajectory escape from $\mathcal{P}_0^H$

bolic shape. We draw L-shaped parameter trajectories at every possible departure angle and add to this set the intersection of the slow mode trajectory and $\mathcal{P}_0^H$ (refer to figure 3.5). It becomes apparent that $\mathcal{P}_0^H$ is composed by an inscribed circle of

the shape of the cost function, and a set of corner minima, for which DD-DFE would probably converge to, but convergence would depend on the cost function shape.

Hence, we can define the class of channels for which DD-DFE would always converge to a bad minimum from an initialization at the origin as

$$\begin{aligned} \mathcal{B} &= \{H = (h_1, h_2)^t \in \mathcal{S} : \\ &\quad \sqrt{d_1^2 + d_2^2} < \frac{1}{\sqrt{2}}(|h_1| + |h_2| - 1)\} \end{aligned} \tag{3.24}$$

where $(d_1, d_2)^t = D^\star(\mathcal{P}_0^H)$. We use a procedure similar to the one used to plot $\mathcal{S}$ in

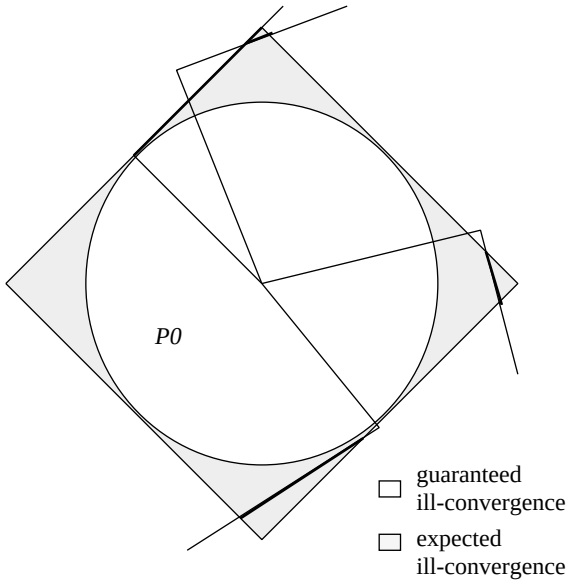

Figure 3.5: Ruling out minima with no guarantee of convergence.

figure 3.3 to plot $\mathcal{B}$ in figure 3.6.

3.1.6 Example: Some cases of ill-convergence

In this section we present two channels for which DD-DFE and CMA-DFE exhibit both similar and different behaviors.

The first channel is $H = (1, 0.9, 0.8)^t$. The polytopes associated with this channel are depicted in figure 3.7; notice that since $H \in \mathcal{C}_1 \subset \mathcal{C}$, $\mathcal{P}_0^H$ is effectively defined by $|d_1| + |d_2| < |h_1| + |h_2| = 0.7$. In this case DD-DFE has two local minima at $D = (0.9, 0.8)^t$ (the desired minimum) and at $D = (0.3667, -0.0333)^t$ (a minimum in $\mathcal{P}_0^H$, as expected). We confirm these results by evaluating the cost function for DD-DFE over a grid of points in DFE parameter space. This calculation is done by simulating the transmission of hundreds of symbols for each point, or DFE setting,

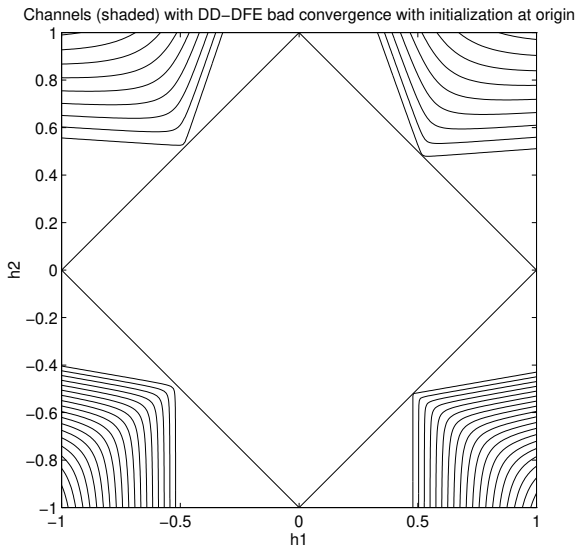

Figure 3.6: $\mathcal{B}$ for three-tap channels.

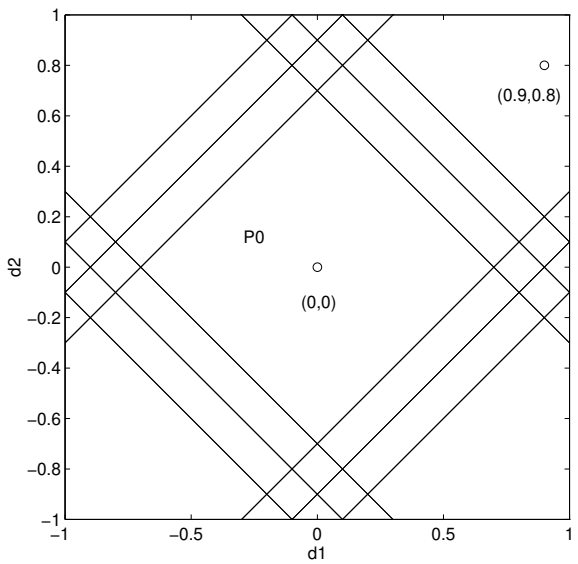

Figure 3.7: Hyperplanes and polytopes for channel $H = (1, 0.9, 0.8)^t$.

and by averaging the DD-DFE cost (and CMA-DFE cost) at each point. Verify that $H = (1, 0.9, 0.8)^t$ is indeed in the class $S_1 \subset C_1$, as shown in figure 3.3.

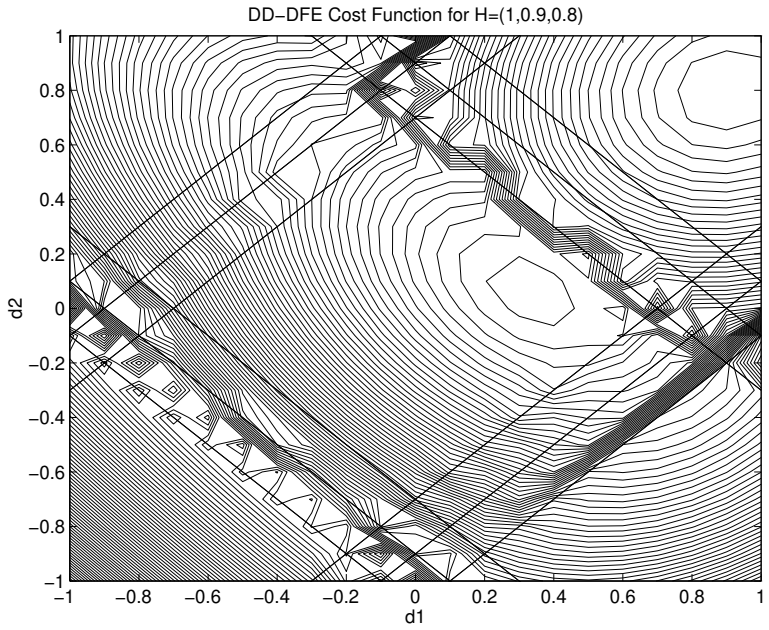

Figure 3.8: DD-DFE cost function for $H = (1, 0.9, 0.8)^t$.

For this same channel, we find via simulation that unlike DD-DFE, CMA-DFE only has one global minimum (the desired minimum) at $D = (0.9, 0.8)^t$ (refer to figures 3.8 and 3.9).

For the second channel, $H = (1, 0.9, -0.8)^t$, ($H \in S_2$, as seen in figure 3.3) the picture changes dramatically. Now both DD-DFE and CMA-DFE have bad minima, at $D = (0.05, , -0.15)^t$ for DD-DFE and at $D \approx (-0.15, -0.35)^t$ for CMA-DFE (refer to figures 3.10 and 3.11).

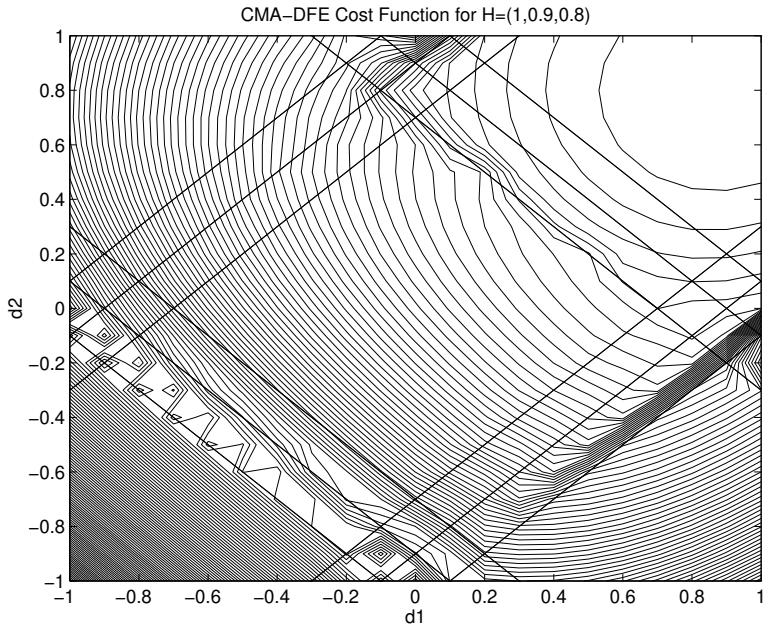

Figure 3.9: CMA-DFE cost function for $H = (1, 0.9, 0.8)^t$.

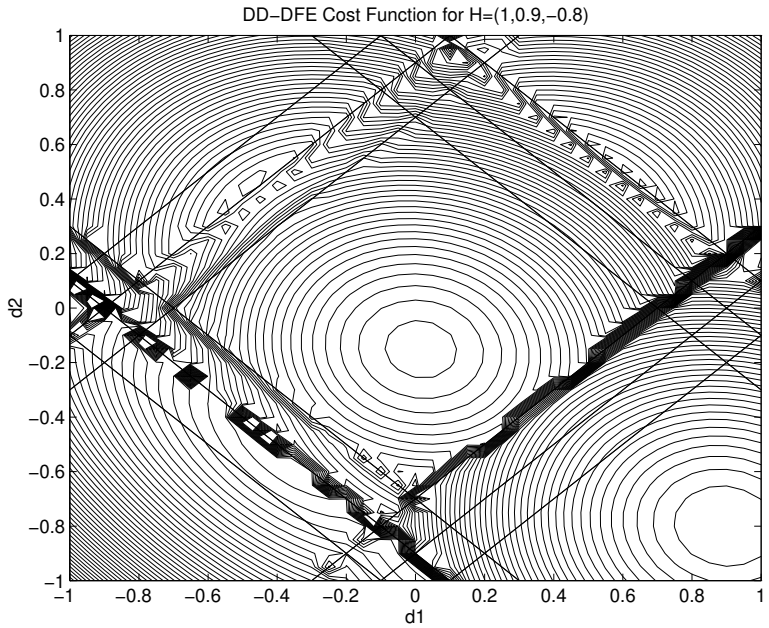

Figure 3.10: DD-DFE cost function for $H = (1, 0.9, -0.8)^t$.

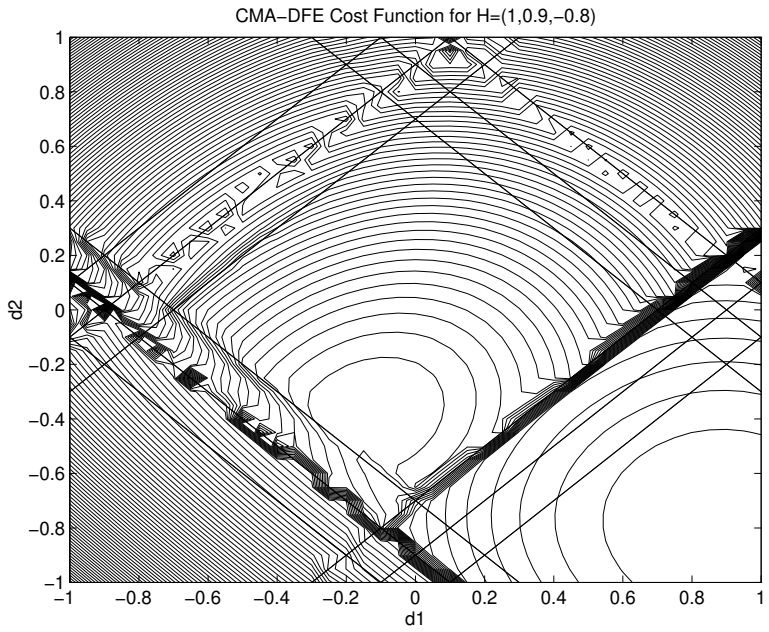

Figure 3.11: CMA-DFE cost function for $H = (1, 0.9, -0.8)^t$.

3.2 $N + 1$ Tap Channels

For $N + 1$-tap channels ($N > 1$),the results are very similar. This time, however, the class of candidate channels for bad minima in $\mathcal{P}_0^H$ is not the class of closed-eye channels, which for three dimensions, has the *nice* property that for any closed-eye channel, the Markov Chain ruling the statistics inside $\mathcal{P}_0^H$ are the same, leading to the simple statement in proposition 1. Unfortunately, in higher dimensions this is not true. For this reason, we consider a more restrictive class, the *corner-class*, given by

$$\mathcal{C}^N \quad = \quad \{(h_1, \cdots, h_N)^t \in \mathbf{R}^N : 1 - \frac{1}{N} < |h_i| < 1 \ \ \forall i\}, \tag{3.25}$$

which has the property that for any channel $H \in \mathcal{C}^N$,$H \notin \mathcal{P}_0^H$, and that inside $\mathcal{P}_0^H$, the random variable X_k does not depend on either the channel or the DFE setting (except for which corner or quadrant in channel parameter space H belongs to). Then, regions $\mathcal{S}$ of the form (3.20) contain the channels which have a minimum in $\mathcal{P}_0^H$.

3.2.1 $\mathcal{C}^N$, the corner-class

As before, we begin by showing that if a channel is in $\mathcal{C}^N$, then it is not in $\mathcal{P}_0^H$, and furthermore, if for a certain channel $\mathcal{P}_0^H$ contains a local minimum, this minimum is not the desired DFE solution.

To aid in the analysis, we assign an index for each corner in channel parameter space. A channel $H = (h_0, \cdots, h_N)^t$ is in corner $\mathcal{C}_i^N$ where i is given by

$$i = \frac{1}{2}(2^N - 1 + 2^{N-1}\sigma_1 + 2^{N-2}\sigma_2 + \cdots + \sigma_N) \tag{3.26}$$

and where $\sigma_j = sgn(h_j)$. As an example, a channel in $\mathcal{C}^3$ with all taps positive will be an element of $\mathcal{C}_7^3$ only. With this notation, we are able to write

$$\mathcal{C}^N = \cup_{j=0}^{2^N-1} \mathcal{C}_j^N \tag{3.27}$$

Lemma 3: *The polytope centered at the origin is given by*

$$\mathcal{P}_0^H = \{(d_1, \cdots, d_N)^t \in \mathbf{R}^N : \sum_{i=1}^{N} |d_i| < \min_{\nu_j} |\sum_{j=0}^{N} \nu_j h_j|\}, \tag{3.28}$$

where $\nu_j \in \{\pm 1\}$. [3] *Then,*

$$\mathcal{C}^N \cap \mathcal{P}_0^H = \emptyset. \tag{3.29}$$

[3] We use ν_j to simplify the notation (i.e. $\nu_j = a_{k-j}$).

Proof: We show that for all $H = (h_1, \cdots, h_N)^t \in \mathcal{C}^N$ $H \notin \mathcal{P}_0^H$, so that

$$\min_{\nu_j} |\sum_{j=0}^{N} \nu_j h_j| \quad < \quad \sum_{i=1}^{N} |h_i| \tag{3.30}$$

First, since $H \in \mathcal{C}^N$ so $|h_i| \geq 1 - \frac{1}{N}$, then

$$\sum_{i=1}^{N} |h_i| \quad \geq \quad N(1 - \frac{1}{N}) = N - 1. \tag{3.31}$$

Now, Since the minimum combination of $h_i's$ is always at most as large as any combination, and because $|h_i| < 1$,

$$\begin{aligned} \min_{\nu_j} |\sum_{j=0}^{N} \nu_j h_j| \quad &\leq \quad \sum_{j=1}^{N} |h_j| - 1 \\ &< \quad N - 1, \end{aligned} \tag{3.32}$$

so

$$\min_{\nu_j} |\sum_{j=0}^{N} \nu_j h_j| < N - 1 \leq \sum_{i=1}^{N} |h_i| \tag{3.33}$$

□

Following the outline established by the work done in the three-tap channel case, the next logical step is to show that for any $H \in \mathcal{C}^N$ the Markov Chain for X_k associated with $\mathcal{P}_0^H$ is the same. To do this, we state the following lemma and later explain its use.

Lemma 4: *Let $h_i \in [0, 1]$ and $\nu_i \in \{\pm 1\}$ for $0 \leq i \leq N$. Define the function*

$$g(h_1, \cdots, h_N, \nu_0, \cdots, \nu_N) \quad = \quad sgn(\nu_0 + \nu_1 h_1 \cdots + \nu_N h_N). \tag{3.34}$$

If and $1 > |h_i| > 1 - \frac{1}{N}$ for $1 \leq i \leq N$ then

$$g(h_1, \cdots, h_N, \nu_0, \cdots, \nu_N) \quad = \quad sgn(\nu_0 + \nu_1 \sigma_1 \cdots + \nu_N \sigma_N). \tag{3.35}$$

where $\sigma_i = sgn(h_i)$.

Proof: We prove the lemma for the case $h_i > 1 - \frac{1}{N} \quad \forall i$ (i.e. $\sigma_i = 1$), and then use this to prove the general case.

Let K be the number of $\nu_i's$ equal to $+1$. We take three cases:

1. If $\underline{K > \frac{N}{2}}$ then $g(h_0, \cdots, h_N, \nu_0, \cdots, \nu_N) = 1$.Let $\nu_{i_1} = \cdots = \nu_{i_K} = +1$ be the positive $\nu_i's$ and let $\nu_{j_1} = \cdots = \nu_{j_{N-K}} = -1$ be the negative $\nu_i's$. We keep the sign of ν_0 unspecified. Then we let

$$\begin{aligned} x &= \nu_0 + \nu_1 h_1 + \cdots + \nu_N h_N \qquad (3.36) \\ &= \nu_0 + \\ &\quad (h_{i_1} - h_{j_1}) + \cdots + (h_{i_{N-K}} - h_{j_{N-K}}) + \\ &\quad h_{i_{N-K+1}} + \cdots + h_{i_K}. \qquad (3.37) \end{aligned}$$

Recall that $1 > h_i > 1 - \frac{1}{N}$, implying that

$$|h_{i_l} - h_{j_l}| < \frac{1}{N} \forall i_l, j_l \qquad (3.38)$$

so that

$$\begin{aligned} x &> \nu_0 + \\ &\quad k(-\frac{1}{N})(N-K) + \\ &\quad (1 - \frac{1}{N})(K - (N-K)) = \nu_0 + 2K - N - \frac{K}{N}. \qquad (3.39) \end{aligned}$$

Since $K > \frac{N}{2}$, then $2K - N \geq 2$ so that

$$x > \nu_0 + 2 - \frac{K}{N} \geq -1 + 2 - \frac{K}{N} = 1 - \frac{K}{N} \geq 0. \qquad (3.40)$$

Thus, $x > 0$ and

$$g(h_1, \cdots, h_N, \nu_0, \cdots, \nu_N) = sgn(x) \tag{3.41}$$

$$= 1. \tag{3.42}$$

2. Similarly, if $K < \frac{N}{2}$ then $g(h_0, \cdots, h_N, \nu_0, \cdots, \nu_N) = -1$.

3. If $K = \frac{N}{2}, \quad N$ even, using the notation above, we have

$$x = \nu_0 + (h_{i_1} - h_{j_1}) + \cdots + (h_{i_K} - h_{j_K}) \tag{3.43}$$

yielding

$$|x - \nu_0| = |(h_{i_1} - h_{j_1}) + \cdots + (h_{i_K} - h_{j_K})| \tag{3.44}$$

$$\leq |h_{i_1} - h_{j_1}| + \cdots + |h_{i_K} - h_{j_K}| \tag{3.45}$$

$$< \frac{1}{N}K = \frac{1}{N}\frac{N}{2} = \frac{1}{2} \tag{3.46}$$

This implies that $|x - \nu_0| < \frac{1}{2}$, and that

$$x \in (\nu_0 - \frac{1}{2}, \nu_0 + \frac{1}{2}), \tag{3.47}$$

or in other words, x has the same sign as ν_0, so

$$g(h_1, \cdots, h_N, \nu_0, \cdots, \nu_N) = sgn(\nu_0) \tag{3.48}$$

In sum, we found that

$$g(h_1, \cdots, h_N, \nu_0, \cdots, \nu_N) = \begin{cases} 1 & K > \frac{N}{2} \\ -1 & K < \frac{N}{2} \\ sgn(\nu_0) & K = \frac{N}{2} \quad N\text{even} \end{cases} \tag{3.49}$$

which clearly tells us that

$$g(h_1, \cdots, h_N, \nu_0, \cdots, \nu_N) = sgn(\nu_0 + \cdots + \nu_N) \tag{3.50}$$

whenever $h_i > 1 - \frac{1}{N}$. Now, switching the sign of h_1 negates the contribution of ν_1. In that case

$$g(h_1, \cdots, h_N, \nu_0, \cdots, \nu_N) = sgn(\nu_0 - \nu_1 + \cdots + \nu_N) \tag{3.51}$$

$$= sgn(\nu_0 + \nu_1\sigma_1 + \cdots + \nu_N) \tag{3.52}$$

Similarly, flipping the sign of any h_i flips the contribution of ν_i so

$$g(h_1, \cdots, h_N, \nu_0, \cdots, \nu_N) = sgn(\nu_0 + \nu_1\sigma_1 \cdots + \nu_N\sigma_N). \tag{3.53}$$

□

Using this result, the next lemma guarantees that for any channel in the corner-class, $\mathcal{C}^N$, the random variable X_k is only a function of the input symbols a_k and of the particular corner, $\mathcal{C}_j^N$, in channel parameter space in which the channel is located.

Lemma 5: *For all H in $\mathcal{C}^N$, and for all $D \in \mathcal{P}_0^H$, the FSMP formed by $X_k = (a_{k-1}, \cdots, a_{k-N} | \hat{a}_{k-1}, \cdots, \hat{a}_{k-N})$ has transition probabilities $P(X_k|X_{k-1})$ that do not depend on H or D, but depend on to which corner $\mathcal{C}_j^N$ in channel parameter space H belongs.*

Proof: The value of $\hat{a}_k$ which is the determining factor in the transition from state X_{k-1} to X_k is given by

$$\hat{a}_k = sgn(a_0 + a_{k-1}h_1 + \cdots + a_{k-N}h_N$$
$$-\hat{a}_{k-1}d_1 - \cdots - \hat{a}_{k-N}h_N) \tag{3.54}$$

Since for any DFE setting within a polytope the *signum* function returns the same result, take $D = (d_1, \cdots, d_N)^t = (0, \cdots, 0)^t$, yielding

$$\widehat{a}_k \;=\; sgn(a_0 + a_{k-1}h_1 + \cdots + a_{k-N}h_N). \tag{3.55}$$

Because $H \in \mathcal{C}^N$, so that $|h_i| < 1 - \frac{1}{N}$, by lemma 4,

$$\widehat{a}_k \;=\; sgn(a_0 + a_{k-1}\sigma_1 + \cdots + a_{k-N}\sigma_N). \tag{3.56}$$

where $\sigma_i = sgn(h_i)$. Therefore, $\widehat{a}_k$ is only a function of the $a_i^l s$ and of which $\mathcal{C}_j^N$ given by the indexing of the $\sigma^l s$, the channel is located at. □

3.2.2 Locating bad $N+1$-tap channels

Our final proposition lays out the condition for a channel $H \in \mathcal{C}^N$ to lead to an undesired DD-DFE local minimum.

Proposition 2: *For all $H \in \mathcal{C}$,*

$$\begin{aligned} D^\star(\mathcal{P}_0^H) \in \mathcal{P}_0^H \quad \Longleftrightarrow \quad & H \in \mathcal{S}^N \quad and \\ & x^t R(\mathcal{P}_0^H) x > 0 \quad \forall x \in \mathbf{C}^2 \end{aligned} \tag{3.57}$$

where $\mathcal{S}^N = \cup_{j=0}^{2^N-1} \mathcal{S}_j^N$ and $\mathcal{S}_j^N$ is given by

$$\mathcal{S}_j^N \;=\; \{H = (h_1, \cdots, h_N)^t \in \mathcal{C}_j^N : \sum_{i=1}^{N} |q_i(H)| < \tau\} \tag{3.58}$$

where

$$(q_1(H), \cdots, q_N(H))^t \;=\; D^\star(\mathcal{P}_0^H) \tag{3.59}$$

$$\;=\; R^{-1}(\mathcal{P}_0^H)[C(\mathcal{P}_0^H)H - T(\mathcal{P}_0^H)]. \tag{3.60}$$

Proof: By lemma 2 we know that for any $H \in \mathcal{C}_j^N$, $R(\mathcal{P}_0^H)$, $C(\mathcal{P}_0^H)$ and $T(\mathcal{P}_0^H)$ will be the same. In order for $D^\star(\mathcal{P}_0^H) = (q_1(H), \cdots, q_N(H))^t$ to be in $\mathcal{P}_0^H$ we need

$$\begin{aligned} \sum_{i=1}^{N} |d_i| &= \sum_{i=1}^{N} |q_i(H)| && (3.61) \\ &< \tau && (3.62) \end{aligned}$$

and in order for $D^\star(\mathcal{P}_0^H)$ to be a minimum, the Hessian of J_{DD} must be positive definite, or in other words, for any $x \in \mathbf{C}^2$, $x^t R(\mathcal{P}_0^H) x > 0$. □

Finally, we wish to take channels in $\mathcal{S}^N$ and verify if their minima lie within a ball inscribed in $\mathcal{P}_0^H$. These channels, in the set

$$\mathcal{B}^N = \{H = (h_1, \cdots, h_N)^t \in \mathcal{S} : \sqrt{\sum_{i=1}^{N} d_i^2} < \frac{1}{\sqrt{N}}\tau\} \quad (3.63)$$

will lead DD-DFE to always converge to the bad minimum.

3.2.3 Example: Some channels $\mathcal{C}^4$

As a first example, we try to find four-tap channels in $\mathcal{C}^3$ that lead to bad DD-DFE minima in $\mathcal{P}_0^H$. We are surprised to find none! The main reason for this fact is that the center polytope defined as

$$\mathcal{P}_0^H = \{(d_1, d_2, d_3)^t \in \mathbf{R}^3 : |d_1| + |d_2| + |d_3| < \tau\} \quad (3.64)$$

is small in general, since τ, the minimum modulus combination of the four channel taps, is also generally small.

At the next dimension, however, we are able to find cases of five-tap channels that exhibit bad minima in $\mathcal{P}_0^H$. We point out one such set, $\mathcal{S}_3^4 \subset \mathcal{C}_3^4$ (recall the

indexing given in formula (3.26); $\mathcal{C}_3^4$ is the set of channels $H \in \mathcal{C}^4$ with $h_2, h_3, h_4 > 0$ and $h_1 < 0$). In order to visualize the set, we draw circles in figure 3.12 at the locations (h_1, h_2, h_3) in channel parameter space (we ignore the position h_4, which is constrained to $h_4 \geq 1 - \frac{1}{4}$ since it would not be possible to visualize a four-dimensional space). Note that we do not draw all sets $\mathcal{S}_j^4$. Figure 3.12 also pictures the corner blocks $\mathcal{C}_j^4$ associated with $h_4 \geq 1 - \frac{1}{4}$. Another set of eight blocks is associated with $h_4 \leq 1 - \frac{1}{4}$.

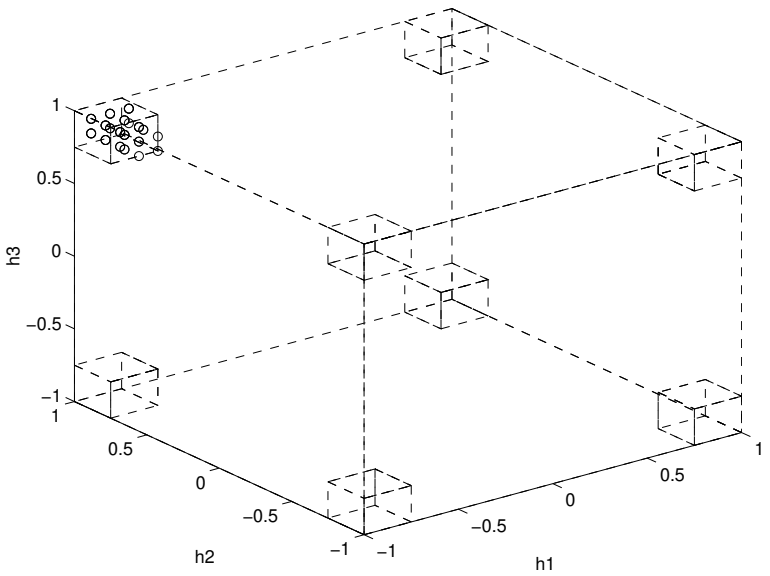

Figure 3.12: One set of five-tap channels in $\mathcal{C}^4$ leading to bad minima in $\mathcal{P}_0^H$

An intuition on why $\mathcal{C}^2$ (three-tap channels) and $\mathcal{C}^4$ contain a larger relative set of bad channels that $\mathcal{C}^3$ (which contains none) is that the relative size of $\mathcal{P}_0^H$ is larger for channels with odd lengths. Take for example, large N, so that the magnitude of the channel taps is close to 1. Recall that the boundaries of $\mathcal{P}_0^H$ are determined by the minimum magnitude of the combination of the channel taps. With this in

mind, we note that an even combination of numbers close to unity has its minimum approaching zero, while an odd combination has minimum approaching unity.

Chapter 4

Observations and Conclusions

The main result of this work was to find a class of channels for which a blind adaptive DFE using the decision-directed algorithm, and initialized at the origin converges to a bad minimum. Other classifications were also obtained in the process and are shown in figure 4.1.

Here follow some observations on these results and a few ideas for future research:

- In the three-tap channel case we managed to verify whether the DD-DFE cost function for a closed-eye channel will have a bad minimum contained in the center polytope, $\mathcal{P}_0^H$, or not. This is not a guarantee, however, that there are no other bad minima in other polytopes, but it does say when a channel will definitely have problems (assuming that all the assumptions on step-size, noise and DFE length are met). The next step in this line of research would be to look for closed-eye channels which have other bad minima along the way from a particular initialization to the polytope enclosing them.

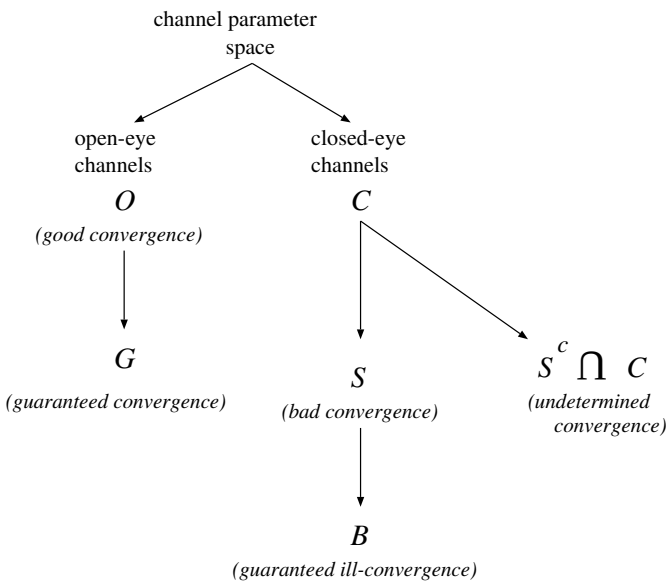

Figure 4.1: Divisions of channel parameter space and DD-DFE convergence with initialization at origin

- For $N+1$-tap channels, we are only able to consider a subset of the closed-eye channels, the set of corner channels. For large N, all channel taps must have a high magnitude (close to unity); it is doubtful that such channels exist in real applications. The importance of the result, however, is not in finding the class of *bad* channels, but in the idea of grouping channels that induce a common DFE behavior so that the statistics $R(\mathcal{P})$, $C(\mathcal{P})$ and $T(\mathcal{P})$ can be easily obtained, and establishing the existence of such a class exhibiting DD-DFE ill-convergence from a zero initialization. A natural suggestion for continuing research is to cover the whole set of closed-eye channels using the same type of argument in order to enlarge our knowledge of *bad* channels.

- There is a possibility of using low dimension results for long channels. Imagine

that the first three taps of a channel have high magnitudes and that the remaining taps are small; for example

$$H = (1, 0.9, 0.8, 0.01)^t. \tag{4.1}$$

This four-tap channel is not an element of the corner-class $\mathcal{C}^3$, but DD-DFE does exhibit convergence problems when initialized at the origin, as seen in figure 4.2, where it stops around $D = (0.39, 0.02, -0.09)^t$ a possible bad minimum inside $\mathcal{P}_0^H$. The argument for understanding this behavior is that the

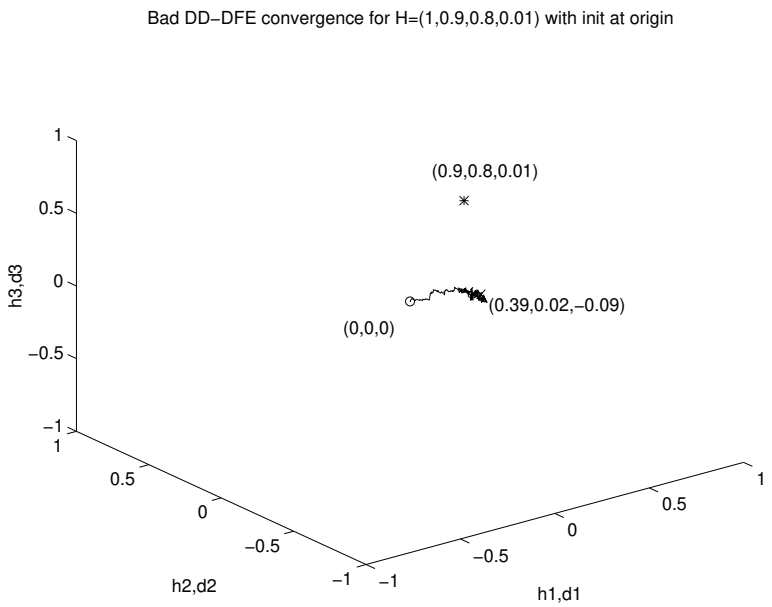

Figure 4.2: Ill-convergence of DD-DFE for $H = (1, 0.9, 0.8, 0.01)^t$.

last tap $h_3 = 0.01$ does not change the boundaries of $\mathcal{P}_0^H$ too much and does not shift the minimum $D^\star(\mathcal{P}_0^H)$ in the center polytope by much either. If the last tap increases enough, however, the algorithm does converge. Adding other small taps offers a similar *bad* behavior. Hence, it might be possible to extend

the results of this thesis for a wider class of channels.

- One more idea for continuing investigation of this type is to break the assumptions, such as transmitting a real two-element constellation, having no channel noise, and using an infinitesimal step-size. Getting rid of the first assumption to allow complex-valued signals will make analysis very difficult, but will also make the work more accessible in practice. The assumptions on noise and step-size are very interesting too. It can be shown through simulation and some analysis that when noise is injected into the received data, or a sufficiently large step-size is used for adaptation, DD-DFE wanders close to bad minima in $\mathcal{P}_0^H$ but eventually escapes the minimum to reach the desired minimum. [1] The same idea was developed by [31],[32] and [33] for the case of a linear feedforward equalizer using decision-direction. Mazo, for example, made the analogy of a particle trapped in a well with a certain energy given by the step-size and the height of the well allowing it to escape with a certain probability. Another source dealing with this issue is [38] which studies the time for escape of CMA (used on a feedforward equalizer) from a bad minimum. This area of research offers much ground for deep analysis of this complex nonlinear stochastic process.

- As an example of a different type of blind adaptive DFE, in section 3.1.6 we drew the CMA-DFE cost function for a three-tap channel. We noticed that for the first channel, while DD-DFE had a bad minimum in $\mathcal{P}_0^H$, CMA-DFE

[1] This has been done by the students Kenneth Andrews, Greg Conklin, Matthew Shoemake and Jay Yun for their final project in EE 525 at Cornell University.

did not. For the second channel, however, both algorithms had bad minima in the center polytope. One of the aims of [17] was to determine whether CMA-DFE had better convergence properties than DD-DFE. The study was inconclusive, but it showed that region of *bad* CMA-DFE channels could be different from the region of *bad* DD-DFE channels. Further study of CMA-DFE could eventually offer a better blind scheme for updating DFEs.

- One should note that it is possible to detect bad minima resulting from a white source since the output at a bad minimum will be colored. One could conceivably check the distribution of the output to detect a bad minimum and then try to jump out of that minimum by either changing the step-size or by adding an offset to the equalizer parameters.

- This thesis directly attacks the open problem of starting up a blind adaptive decision feedback equalizer. This can be done in several ways, each offering possibilities for further study:

    1. Starting the DFE anywhere (a natural position being the origin).

       The first idea is the one we have been dealing with.

    2. Opening up the eye of the channel by using a feedforward filter, which can also be blindly updated. It is possible to adapt this filter first and then turn to decision-direction with a DFE, or to adapt both DFE and feedforward equalizer at the same time.

       It is possible to show that (in the white source case) a joint MMSE feedforward/DFE combination cancels the overall channel/feedforward-filter

precursor ISI using the feedforward portion and uses the DFE to take care of the postcursor. It is possible that the adaptation methods suggested in this second idea will also fail to reach the desired joint MMSE solution from their particular initializations.

3. Estimating the channel dynamics and using these to initialize DD-DFE. *With ideas similar to the ones used to find channels with* guaranteed *convergence, one could possibly determine when DD-DFE would converge as desired for an initialization close to the best minimum.*

Even though the assumptions established are strong, and the class of channels found is severely restricted and relatively impractical, this work poses some concern with the use of blind adaptive schemes in decision feedback equalization. If the algorithm is not initialized properly, or the channel is not prefiltered by a feedforward equalizer to a close to open-eye condition, this thesis provides legitimate reasons for expecting ill-convergence of the algorithm. Therefore, careless use of such algorithms should be avoided.

Bibliography

[1] M. Austin, "Decision feedback equalization for digital communication over dispersive channels," *MIT Research Laboratory of Electronics Technical Report 461*, August 1967

[2] P. Monsen, "Feedback equalization for fading dispersive channels," *IEEE Transactions on Information Theory*, IT-17,pp.56-64, January 1971

[3] R. Price, "Nonlinearly feedback-equalized PAM vs. capacity from noisy filter channels," *Proceedings of IEEE International Conference on Communications*, Philadelphia, USA, 1972, pp.22/12-17

[4] D. Falconer and G. Foschini, "Theory of minimum mean-square-error QAM systems employing decision feedback equalization," *Bell System Technical Journal*, Vol.52, No.10, December 1973

[5] J. Salz, "Optimum mean-square decision feedback equalization," *Bell System Technical Journal*, Vol.52, no.8, October 1973

[6] D. Messerschmitt, "A geometric theory of intrasymbol interference. Part I: Zero-forcing and decision-feedback equalization," *Bell System Technical Journal*, Vol.52, No.10, pp.1483-1519, November 1973

[7] A. Salazar, "Design of transmitter and receiver filters for decision feedback equalization," *Bell System Technical Journal*, Vol.53, No.3, March 1974

[8] D. Messerschmitt, "Minimum MSE equalization of digital fiber optic systems," *IEEE Transactions on Communications*, COM-26, pp.1110-1118, 1978

[9] M. Adbulrahman and D.D.Falconer, "Cyclostationary crosstalk suppression by decision feedback equalization on digital subscriber loops," *IEEE Journal on Selected Areas in Communications*, Vol 10, No.3, April 1992, pp.640-649

[10] N. Al-Dhahir and J.M. Cioffi, "MMSE decision-feedback equalizers: finite-length results," *IEEE Transactions on Information Theory*, pp.961-975, July 1995

[11] N. Al-Dhahir and J.M. Cioffi, "Fast computation of channel-estimate based equalizers in packet data transmission," *IEEE Transactions on Information Theory*, November 1995, pp.2462-2473

[12] D. George, R. Bowen, and J. Storey, "An adaptive decision feedback equalizer," *IEEE Transactions on Communications*, Vol.COM-19, pp.281-292, June 1971

[13] C.A. Belfiore and J.H. Park, Jr., "Decision feedback equalization," *Proceedings of the IEEE*, Vol. 67, pp.1143-1156, August 1979

[14] R.A. Kennedy, B.D.O. Anderson, and R.R. Bitmead, "Blind adaptation of decision feedback equalizers: gross convergence properties," *International Journal of Adaptive Control and Signal Processing*, Vol.7, pp.497-523, 1993

[15] S.Marcos, S.Cherif, M.Jaidane, "Blind cancellation of intersymbol interference in decision feedback equalizers," *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, May 1995

[16] C.B.Papadias and A.Paulraj, "Decision-feedback equalization and identification of linear channels using blind algorithms of the bussgang type," *Proceedings of the 29th Annual Asilomar Conference on Signals, Systems, and Computers*, November 1995

[17] R. Casas, Z. Ding, R.A. Kennedy, C.R.Johnson, Jr., and R. Malamut, "Blind adaptation of decision feedback equalizers based on the constant modulus algorithm," *Proceedings of the 29th Annual Asilomar Conference on Signals, Systems, and Computers*, November 1995

[18] D.L.Duttweiler, J.E. Mazo, and D.G.Messerschmitt, "An upper bound to the error probability in decision-feedback equalization," *IEEE Transactions on Information Theory*, IT-20, pp.490-497, July 1974

[19] A.Cantoni and P.Butler, "Stability of decision feedback inverses," *IEEE Transactions on Communications*, Vol COM-24, pp.970-977, September 1976

[20] J.J. O'Reilly and A.M. de Oliveira Duarte, "Error propagation in decision feedback receivers," *Proceedings of the IEE*, pt. F, pp.561-566, December 1985

[21] A.M. de Oliveira Duarte and J.J. O'Reilly, "Simplified technique for bounding error statistics for DFB receivers," *Proceedings of the IEE*, pt. F, No.7, pp.567-575, December 1985

[22] R.A. Kennedy and B.D.O. Anderson "Error recovery of decision feedback equalizers on exponential impulse response channels," *IEEE Transactions on Communications*, vol. COM-35, pp.846-848, August 1987

[23] R.A. Kennedy, and B.D.O. Anderson, "Recovery times of decision feedback equalizers on noiseless channels," *IEEE Transactions on Communications*, Vol.COM-35, No.10, October 1987, pp.1012-1021

[24] R.A. Kennedy, B.D.O. Anderson, and R.R. Bitmead, "Tight bounds on the error probabilities of decision feedback equalizers," *IEEE Transactions on Communications*, Vol. COM-35, pp1022-1029, October 1987

[25] R.A. Kennedy, B.D.O. Anderson, and R.R. Bitmead, "Channels leading to rapid error recovery for decision feedback equalizers," *IEEE Transactions on Communications*, pp. 1867-1875, November 1989

[26] S.A.Altekar and N.C.Beaulieu, "Upper bounds to the error probability of decision-feedback equalization," *IEEE Transactions on Information Theory*, IT-39, pp.145-156, January 1993

[27] M.V. Eyuboglu and S.U.H. Quereshi, "Reduced-state sequence estimation with the set partition and decision feedback," *IEEE Transactions on Communications*, Vol.36, pp.13-20, January 1988

[28] A. Duel-Hallen and C. Heegard, "Delayed decision-feedback sequence estimation," *IEEE Transactions on Communications*, Vol.37, pp.428-436, May 1989

[29] J.M. Cioffi, P.H.Algoet and P.S. Chow, "Combined equalization and coding with finite-length decision feedback equalization," *IEEE Transactions on Communications*, Vol.37, pp.807.2.1-807.2.4, December 1990

[30] J.Cioffi, G. Dudevoir, M. Eyuboglu, and G.D. Forney, Jr., "Minimum mean-square-error decision feedback equalization and coding – Parts I and II," *IEEE Transactions on Communications*, 1995

[31] J.E. Mazo, "Analysis of decision-directed equalizer convergence," *Bell System Technical Journal*, 1980, pp. 1857-1876

[32] R. Kumar, "Convergence of a decision-directed adaptive equalizer," *Proceedings of the Conference on Decision and Control*, Long Beach, CA, 1983, IEEE New York, 1983, pp.1319-1324

[33] O.Macchi and E.Eweda, "Convergence analysis of self-adaptive equalizers," *IEEE Transactions on Information Theory*, Vol IT-30, No.2, March 1984, pp.161-176

[34] R.A. Kennedy, G. W. Pulford, B.D.O. Anderson, and R.R. Bitmead, "When has a decision directed equalizer converged?," *IEEE Transactions on Communications*, 1989, pp.879-884

[35] D.N.Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Transactions on Communications*, COM-28, pp.1867-1875, November 1980

[36] J.R. Treichler, and M.G.Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-31, pp.349-472, April 1983

[37] A. Benveniste, M.Métivier, P.Priouret, *Adaptive Algorithms and Stochastic Approximations*, Springer Verlag, New York, 1990

[38] M. Frater, R. Bitmead, C.R. Johnson, Jr., "Local Minima Escape Transients by Stochastic Gradient Descent Algorithms in Blind Adaptive Equalizers," *Automatica*, Vol. 31, No. 4, pp.637-641, 1995

[39] R. Lucky, "Automatic equalization for digital communication," *Bell System Technical Journal*, Vol.44, pp.547-588, April 1965

[40] S.U.H. Qureshi, "Adaptive equalization," *Proceedings of the IEEE*, Vol.73, No.9, September 1985