BLIND ERROR DETECTION AND CHANNEL EQUALISATION IN COMMUNICATION SYSTEMS

Kutluyıl Doğançay

B.S., Boğaziçi University, Turkey
M.Sc., Imperial College, University of London, U.K.

October 1995

A thesis submitted for the degree of Doctor of Philosophy
of the Australian National University

Telecommunications Engineering Group
Research School of Information Sciences and Engineering

The Australian National University

Declaration

The contents of this thesis are the result of original research and have not been submitted to any other university or institution for the purpose of obtaining a postgraduate degree.

Much of the work in this thesis has been published or submitted for publication as journal papers. These papers are:

J1. K. Doğançay and R. A. Kennedy, "Testing for the convergence of a linear decision directed equaliser," *IEE Proc. Vision, Image and Signal Processing*, vol. 141, no. 2, pp. 129–136, Apr. 1994.

J2. K. Doğançay and R. A. Kennedy, "Blind detection of equalisation errors in communication systems," submitted to *IEEE Trans. Inform. Theory*.

J3. K. Doğançay and R. A. Kennedy, "Least squares approach to blind channel equalisation," submitted to *IEEE Trans. Commun.*

A number of papers have been published in conference proceedings. These papers, some of which contain material partially overlapping with the publications listed above, are:

C1. K. Doğançay and R. A. Kennedy, "Testing for parameter convergence in blind adaptive channel equalisation," in *Proc. Second Int. Workshop on Intelligent Signal Processing and Communication Systems, ISPACS '93*, pp. 1–6, Sendai, Japan, Oct. 1993.

C2. J. P. LeBlanc, K. Doğançay, R. A. Kennedy and C. R. Johnson, Jr., "Effects of input data correlation on the convergence of blind adaptive equalizers," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, ICASSP '94*, vol. III, pp. 313–316, Adelaide, Australia, Apr. 1994.

C3. K. Doğançay and R. A. Kennedy, "A globally admissible off-line modulus restoral algorithm for low-order adaptive channel equalisers," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, ICASSP '94*, vol. III, pp. 61–64, Adelaide, Australia, Apr. 1994.

C4. K. Doğançay and R. A. Kennedy, "Testing equalisation performance in blind adaptation," in *Proc. 33rd IEEE Conf. on Decision and Control, CDC '94*, pp. 2817–2818, Lake Buena Vista, Florida, USA, Dec. 1994.

The following journal papers are currently in preparation:

P1. K. Doğançay and Shin-Ho Chung, "Estimation of pulse height distribution in quantal analysis".

P2. K. Doğançay, R. A. Kennedy and C. R. Johnson, Jr., "Least squares approach to fractionally spaced blind equalisation of FIR channels".

The research described in this thesis has been performed, to a large extent, in collaboration with Dr. Rodney A. Kennedy. A small portion of the work on fractionally spaced blind equalisation was carried out jointly with Professor C. Richard Johnson, Jr. However, the majority of the work, approximately 90%, is my own.

Kutluyıl Doğançay

12 October 1995

Acknowledgments

Abstract

The thesis is concerned with the development of equalisation error tests and globally admissible channel equalisation algorithms in data communication systems. Specifically, no explicit knowledge of the channel input sequence is assumed, say, in the form of a training sequence.

Blind detection of equalisation errors is studied in the framework of hypothesis testing. For a linear decision-directed equaliser (LDDE) operating on a linear channel with binary inputs, a necessary and sufficient condition for convergence to an open-eye parameter setting is shown to be identical channel input and slicer output second-order statistics under some assumptions about the channel noise distribution. A consistent threshold test is constructed based on this criterion. For dependent M-ary channel inputs, a simple convergence test criterion comparing the decision device output and channel input variances is conjectured.

In view of the restriction of the convergence test based on second-order statistics to LDDEs and binary channel inputs, a new test criterion is proposed, which is shown to be invariant to the equaliser structure and the digital modulation technique used so long as the channel is linear and time-invariant during the test interval. The criterion draws on the observation that a time variation in the parameters of the underlying linear model taking the decision device output to the equaliser input is commensurate with the presence of equalisation errors. A least squares (LS) parameter estimation approach is adopted to construct uniformly most powerful tests.

The constant modulus algorithm (CMA) is modified to develop a new fast blind equalisation algorithm for channels with approximately finite-duration impulse response (FIR) inverses and constant modulus inputs. An LS approach is employed to compute a nonlinearly transformed version of the equaliser parameters. The actual equaliser parameters are shown to be extractable from their transformed version up to a scaling factor. Unlike CMA, the algorithm is globally admissible, relatively insensitive to correlated channel inputs and "parsimonious" in its use of the channel output observations.

A fractionally spaced implementation of the above algorithm is also proposed to make it applicable to FIR channels. The issue of channel noise enhancement is addressed, and some remedies are proposed by way of a reduced rank matrix approximation. Computer simulations are used to show that the new algorithm can have a considerably faster convergence rate than fractionally spaced CMA.

Contents

List of Tables

List of Figures

Glossary

Abbreviations

a.s.	almost surely
i.i.d.	independent identically distributed
w.p.1	with probability one
AR	autoregressive
ARMA	autoregressive moving average
BPSK	binary phase shift keying
CLEM	closed-eye measure
CMA	constant modulus algorithm
DFE	decision feedback equaliser
DPSK	differential phase shift keying
FIR	finite-duration impulse response
FS-CMA	fractionally spaced CMA
FSE	fractionally spaced equaliser
FSK	frequency shift keying
IIR	infinite-duration impulse response
ISI	intersymbol interference
LDDE	linear decision-directed equaliser
LMS	least mean square
LS	least squares
LTI	linear time-invariant
LTV	linear time-varying
PAM	pulse amplitude modulation
PSK	phase shift keying
QAM	quadrature amplitude modulation
QPSK	quadrature phase shift keying
RLS	recursive least squares
SNR	signal-to-noise ratio

SVD	singular value decomposition
UMP	uniformly most powerful

Notation

A^T	transpose of A
A^*	conjugate of A
A^H	Hermitian (conjugate transpose) of A
A^-	g-inverse of A:
	$G = A^-$ if $AGA = A$
A_r^-	symmetric reflexive g-inverse of a real matrix A:
	$G = A_r^-$ if $AGA = A$, $GAG = G$ and $G^T = G$
$A^\dagger$	Moore-Penrose inverse of A:
	$G = A^\dagger$ if $AGA = A$, $GAG = G$, $(AG)^H = AG$ and $(GA)^H = GA$
$\text{col}(A)$	column space of A
$\text{rank}(A)$	rank of A
$\text{tr}(A)$	trace of A
$\lambda_{\max}(A)$	maximum eigenvalue of A
χ_d^2	central chi-square distribution with d degrees of freedom
$\chi_d^2(\rho)$	noncentral chi-square distribution with d degrees of freedom and noncentrality parameter ρ
$\mathcal{N}(\mu, \Sigma)$	(multivariate) Gaussian distribution with mean μ and covariance Σ
$\text{sgn}(\cdot)$	signum function:
	$\text{sgn}(x) = 0$ if $x = 0$, $\text{sgn}(x) = x/|x|$ if $x \neq 0$
$\delta(\cdot)$	Kronecker delta function:
	$\delta(x) = 1$ if $x = 0$, $\delta(x) = 0$ if $x \neq 0$
$E\{\cdot\}$	expectation operator
$\Im\{\cdot\}$	imaginary part
$\Re\{\cdot\}$	real part
P_D	probability of detection
P_{FA}	probability of false alarm
$Q_M(\cdot)$	M-ary quantiser

$m_{4,x}(\tau_1, \tau_2, \tau_3)$ fourth-order joint moment of $\{x(k)\}$ at lags τ_1, τ_2 and τ_3:

$$m_{4,x}(\tau_1, \tau_2, \tau_3) = E\left\{(x(k) - \mu)(x(k+\tau_1) - \mu)(x(k+\tau_2) - \mu)(x(k+\tau_3) - \mu)\right\}$$

where $\mu = E\{x(k)\}$

$R_x(\tau)$ autocorrelation of $\{x(k)\}$ at lag τ

$\widehat{R}_x(\tau)$ biased estimate of $R_x(\tau)$, i.e. $E\left\{\widehat{R}_x(\tau)\right\} \neq R_x(\tau)$

$R_{xy}(\tau)$ crosscorrelation of $\{x(k)\}$ and $\{y(k)\}$ at lag τ

$*$ convolution operator

$\circ$ Hadamard (entrywise) product

$\otimes$ Kronecker product

$\sim$ distributed according to

CHAPTER *1*

Introduction and Summary

1.1 Overview

The central theme of this thesis is blind channel equalisation in data communication systems. The following aspects of blind channel equalisation are considered: (i) the detection of decision errors at the output of a blind adaptive channel equaliser, and (ii) the development of globally admissible blind equalisation algorithms. Specifically, no direct access to the channel input is assumed, say, in the form of a training sequence, which enables the detection of equaliser decision errors or *equalisation errors* in the context of blind channel equalisation, as well as conventional channel equalisation when the training session is not in progress. The development of globally admissible blind equalisation algorithms is motivated by the observation that on-line memoryless-cost-function blind equalisation algorithms developed so far are all capable of converging to undesirable parameter settings.

The aspects of blind channel equalisation considered in this thesis are closely intertwined despite their seemingly different application areas. Whilst convergence to an unacceptable parameter setting can be detected by a blind equalisation error test, the ideas behind blind error detection can be exploited to provide a fix for the observed misbehaviour of on-line blind equalisation algorithms. It should be noted, however, that the successful development of a globally admissible blind equalisation algorithm does not necessarily detract from the importance of blind error detection since equalisation errors can still occur owing to other reasons such as high channel noise, error propagation and improper equaliser design, to name but a few.

We will purposely avoid being rigorous in our presentation in this chapter so as to provide a tutorial-level introduction to the subject matter of the thesis. The operation of channel equalisation will be described first, using a baseband equivalent model of communication systems. The equal-

isation objective and a measure of intersymbol interference (ISI) will be defined. In Section 1.3, we will explain the notion of global admissibility in blind channel equalisation. Section 1.4 motivates the need for testing the equalisation objective and points out the shortcomings of a popular experimental testing method known as the eye pattern test. The problem of testing the equalisation objective is considered in the binary hypothesis testing framework as a viable alternative to the eye pattern test. Section 1.5 summarises the recent work that is relevant to the research reported in this thesis. A chapter-by-chapter summary of the thesis is given in Section 1.6. A point summary of the major contributions of the thesis is listed in Section 1.7.

1.2   Blind Channel Equalisation

Elements of a data communication system are depicted in Fig. 1.1. An analog source signal to be transmitted is first sampled and quantised to produce an input bit stream which is a discrete-time signal with discrete amplitudes. Some information about the analog signal is lost in the process, but the accuracy can be improved by choosing finer quantisation levels at the expense of increasing the number of bits for every source symbol. The bit stream obtained from the analog source signal is encoded to produce an efficient representation for the source symbols (source coding) and to add a controlled amount of redundancy for the purpose of improving its robustness to additive channel noise (channel coding). The encoded signal is passed through a transmit filter to generate pulses with limited bandwidth so that distortion-free transmission through a passband medium can be achieved. The pulse shaping is followed by modulation in which some characteristic, viz. the magnitude, phase or frequency, of a high frequency carrier is varied in sympathy with the pulse shaped signal to make it suitable to be sent through the transmission medium. At the receiver, the signal is demodulated by removing the high frequency carrier, filtered using a knowledge of the pulse shape introduced by the transmit filter, and equalised to cancel the effects of the transmission medium. After sampled and passed through a decision device, the transmitted bit stream is finally recovered at the decoder output.

The discrete "channel", which is composed of a cascade of the transmit filter, the modulator, the actual medium, the demodulator and the receive filter, delivers a corrupted and transformed version of its input. The corruption that is incurred by the channel input is usually random in its nature and is associated with background thermal noise. The transformation introduced by the discrete channel involves time dispersion (resulting from the finite bandwidth and nonconstant phase of the channel), frequency translation, as well as nonlinear distortion. Time dispersion results in spreading of transmitted symbols beyond their duration, leading to an overlap of successive
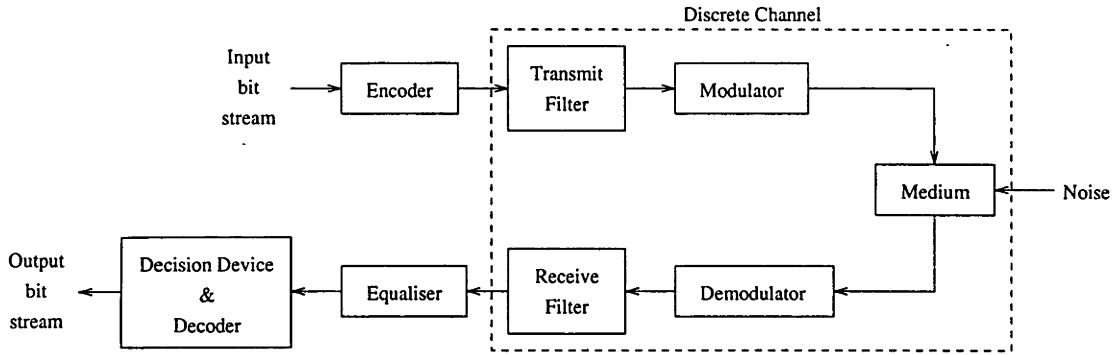
Figure 1.1 Block diagram of a data communication system.
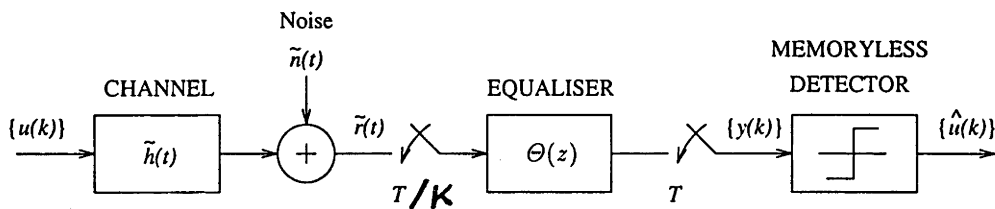
Figure 1.2 Baseband equivalent model of a communication system.

symbols. The resulting distortion is called *intersymbol interference* (ISI) which imposes a limit on reliable transmission of source symbols in high speed data communication systems. The operation of removing ISI is called *channel equalisation*, and the device used for this purpose is called a *channel equaliser*. The increasing need for high speed data communication over voice-grade telephone lines has added to the importance of channel equalisation.

While in telephone channels ISI is mostly associated with linear distortion (time dispersion) of the channel, in radio channels ISI is due to multipath fading which is characterised by transmission through several paths with different time-varying magnitudes and phase characteristics. Equalisers for radio channels are therefore required to be *adaptive* to track time variations in channel characteristics.

Throughout the thesis we will use the baseband equivalent model of a passband communication system shown in Fig. 1.2. Baseband representation preserves the characteristics of passband signals while providing a common framework to work within by removing the high frequency carrier from consideration. In other words, baseband representation is a convenient abstraction of the actual communication system. The baseband (complex-valued) channel input $u(k)$ and the passband modulated signal $\tilde{s}_P(t)$ are related according to

$$\tilde{s}_P(t) = \Re \left\{ \sum_k u(k)\tilde{a}(t - kT) \exp(j2\pi f_c t) \right\} \qquad (1.1)$$

where $\Re\{\cdot\}$ denotes real part, $\tilde{a}(t)$ is the continuous-time impulse response of the transmit filter (defining the pulse shape), T is the symbol rate, and f_c is the carrier frequency. Note that $u(k)$ corresponds to the baseband encoder output in Fig. 1.1. Likewise, the real-valued passband channel impulse response $\tilde{h}_P(t)$ is connected to its (compex-valued) baseband counterpart $\tilde{h}(t)$ through

$$\tilde{h}_P(t) = \Re\left\{\tilde{h}(t)\exp(j2\pi f_c t)\right\}. \tag{1.2}$$

The baseband channel impulse response is obtained from a cascade of the transmit filter $\tilde{a}(t)$, the baseband medium $\tilde{m}(t)$ and the receive filter $\tilde{b}(t)$:

$$\tilde{h}(t) = \tilde{a}(t) * \tilde{m}(t) * \tilde{b}(t) \tag{1.3}$$

where $*$ denotes convolution. The baseband channel noise is given by

$$\tilde{n}(t) = \tilde{b}(t) * \left(\tilde{n}_P(t)\exp(-j2\pi f_c t)\right) \tag{1.4}$$

where $\tilde{n}_P(t)$ is the passband noise. In Eq. (1.4) the receive filter is assumed to suppress high frequency components resulting from demodulation of the received signal. It should be noted that the baseband noise is usually not white even if the passband noise is.

Intersymbol interference occurs in all forms of digital communication when the bandwidth of the channel input signal is wider than half the passband of the channel. As an illustration of ISI consider the baseband channel impulse response depicted in Fig. 1.3. The received signal before sampling is given by

$$\tilde{r}(t) = \sum_{i=-\infty}^{\infty} u(i)\tilde{h}(t - iT) + \tilde{n}(t). \tag{1.5}$$

After sampling $\tilde{r}(t)$ at T-second intervals, which results in the discrete-time sequence $\{r(k)\}$, we obtain

$$r(k) = \underbrace{u(k - \Delta)h_\Delta}_{\text{desired signal}} + \underbrace{\sum_{i \neq \Delta} u(k - i)h_i}_{\text{ISI}} + n(k) \tag{1.6}$$

where Δ denotes the time instant of the *cursor* defined by $\max_k |h_k|$, and

$$h_k = \int_{-\infty}^{+\infty} \tilde{h}(t)\delta(t - kT)dt = \tilde{h}(kT) \tag{1.7a}$$

$$n(k) = \int_{-\infty}^{+\infty} \tilde{n}(t)\delta(t - kT)dt = \tilde{n}(kT) \tag{1.7b}$$

are the sampled baseband channel impulse response and the sampled baseband channel noise,
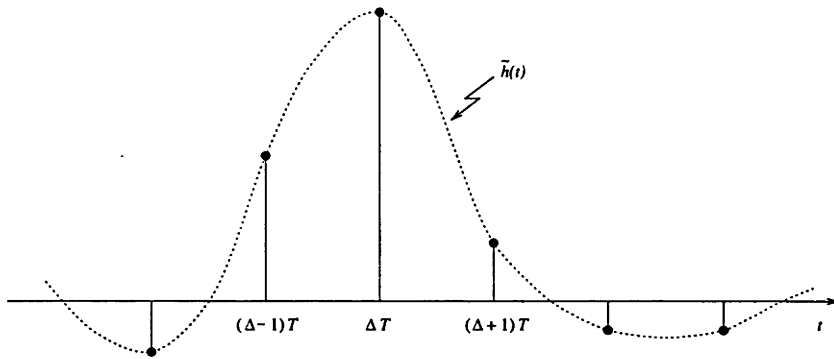
Figure 1.3 Baseband channel impulse response.

respectively. Obviously, ISI is zero if and only if all samples of the channel impulse response except one, viz. h_Δ, are zero.

An adaptive channel equaliser is shown in Fig. 1.4. The objective of the equaliser $\Theta(z)$ is to remove ISI sufficiently so that a decision device can recover the channel input without error. The equaliser is therefore designed to provide a close approximation to a possibly delayed and truncated version of the channel inverse $H^{-1}(z)$. In conventional channel equalisation a training sequence is made available at the receiver in the form of a finite-length synchronised channel input sequence. The purpose of the training sequence is to enable the equaliser to get information about the channel characteristics and tune the equaliser parameters accordingly. The training sessions are held at regular intervals to track any changes in the transmission medium characteristics. At the completion of training sessions the equaliser is usually switched to a *decision-directed* mode of operation. In decision-directed mode, the error signal is derived from the difference between the equaliser output and the decision device output.

In polled data communication systems using multipoint modems and microwave digital radio systems, the adaptive equalisers are required to have a fast start-up, which implies the use of decision-directed mode rather than any training sequence. The major reason for not using a training sequence in these applications is the severe degradation in data throughput which results from the transmission of a training sequence. The decision-directed mode ensures correct convergence if the eye is already open and thus cannot be expected to perform well on channels with severe ISI. In multipoint data communications, it is therefore desirable to have a blind start-up without the requirement of a training sequence, which has brought about the inception of a different class of equalisers called *blind channel equalisers*. In blind channel equalisation the only information utilised is the channel output sequence and *a priori* knowledge of the channel input statistics. If the dotted line in Fig. 1.4 is removed, the resultant equaliser is called a blind equaliser since the equaliser has to adapt its parameters in the absence of an explicit knowledge of the channel input.
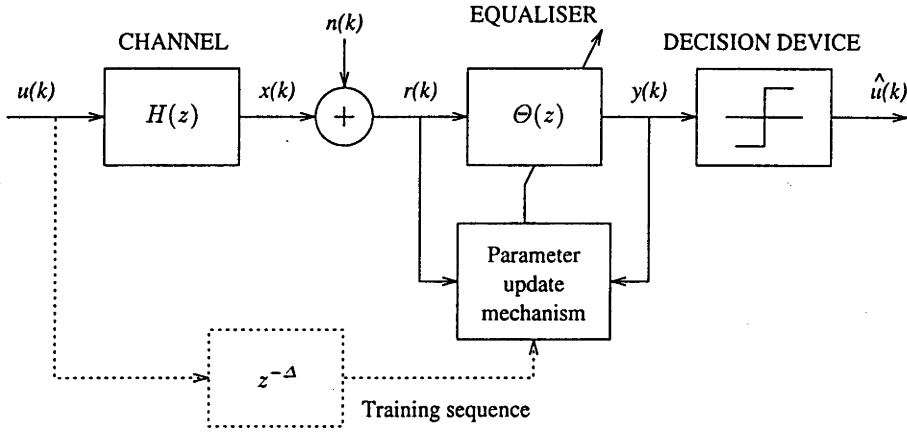
Figure 1.4 Adaptive channel equalisation. If the dotted line is removed, equaliser parameter adaptation is performed by a blind equalisation algorithm.

The *equalisation objective* can be mathematically described as the satisfaction of the following condition:

$$\hat{u}(k) = \Gamma u(k - \Delta) \quad \forall k \tag{1.8}$$

where Γ is a constant that maps a given channel input $u(k)$ to a member of the channel input constellation and Δ is the equalisation delay which is a finite nonnegative integer. The ambiguity arising from Γ is resolved by using differential encoding such as differential phase shift keying (DPSK). For an M-ary DPSK system Γ is $e^{jm2\pi/M}$ where m is an integer. If Eq. (1.8) is satisfied, then we say that *the eye is open*. A violation of Eq. (1.8) indicates that *the eye is closed*.

Assuming an automatic gain control at the decision device, the open-eye condition is characterised by the satisfaction of the inequality

$$\frac{\max_k |u(k)| \sum_{i \neq \Delta} |h_i|}{|h_\Delta|} < \frac{d_{\min}}{2}, \quad |h_\Delta| \triangleq \max_i |h_i| \tag{1.9}$$

where the quantity on the left-hand side is the maximum value that the ISI term in Eq. (1.6) can take, divided by the magnitude of the cursor, and $d_{\min}$ denotes the minimum distance between any two members of the channel input constellation. If the above inequality is satisfied the channel input can be recovered at the equaliser output with the aid of a decision device. The *closed-eye measure* (CLEM), which we shall use throughout the thesis to quantify the amount of ISI at the equaliser output, is defined by

$$\text{CLEM} \triangleq \frac{2 \max_k |u(k)| \sum_{i \neq \Delta} |h_i|}{d_{\min} |h_\Delta|}. \tag{1.10}$$

Clearly, if CLEM < 1, the eye is open. Otherwise, the decision device output contains errors

rendering the eye closed. For pulse amplitude modulated channel inputs that are drawn from the constellation set $\mathbb{S} = \{\pm 1, \pm 3, \ldots, \pm(M-1)\}$, we have $d_{\min} = 2$ and $\max_k |u(k)| = M - 1$, which results in

$$\text{CLEM} = (M-1)\frac{\sum_{i \neq \Delta} |h_i|}{|h_\Delta|}. \tag{1.11}$$

For phase shift keyed inputs drawn from the set $\mathbb{S} = \left\{ 1, e^{j2\pi/M}, e^{j4\pi/M}, \ldots, e^{j2(M-1)\pi/M} \right\}$, the minimum distance between the constellation members is given by the expression

$$d_{\min} = 2 \sin \frac{\pi}{M}. \tag{1.12}$$

Noting that all the symbols have unity modulus which implies $\max_k |u(k)| = 1$, CLEM for M-ary phase shift keying (PSK) modulation can be written as

$$\text{CLEM} = \frac{\sum_{i \neq \Delta} |h_i|}{\left(\sin \frac{\pi}{M}\right)|h_\Delta|}. \tag{1.13}$$

ISI at the equaliser output results from the time dispersion of the channel and equaliser combination. Therefore, CLEM for ISI measurements at the equaliser output involves the use of the combined channel-equaliser impulse response rather than the channel impulse response alone. Care should be exercised in interpreting CLEM if the channel noise is not negligible. Whilst for zero channel noise, the necessary and sufficient condition for the equaliser output decisions to have no errors is CLEM < 1, if the channel output is corrupted by noise, the condition CLEM < 1 does not guarantee that the eye is open.

1.3 Global Admissibility and Ill-convergence

In blind channel equalisation the lack of direct access to the channel input has resulted in the consideration of *pseudo error* signals rather than the true error between the channel input and equaliser output sequences. The use of a pseudo error in on-line blind equalisation algorithms with memoryless cost functions has been shown to create multimodal cost functions whose minimisation using simple stochastic gradient-based techniques results in a tendency to get stuck with local minima yielding inferior or unacceptable equalisation performance [13, 41]. The notion of multimodality has endowed the communication and signal processing community with some new terminology, viz. *admissibility* and *ill-convergence*.

Global admissibility is defined as the ability of an equalisation algorithm to converge to an open-eye parameter setting from *any* initialisation. Ill-convergence is the phenomenon of convergence to a closed-eye minimum while open-eye parameter settings exist. If an algorithm is

globally admissible it does not suffer from ill-convergence. That stochastic gradient-based blind equalisation algorithms proposed thus far all exhibit ill-convergence to closed-eye local minima has led to an intensive research into ways of avoiding or fixing this problem. The ill-convergence phenomenon calls for the construction of blind testing procedures to detect its occurrence, on the one hand, and the development of globally admissible algorithms, on the other hand. The latter does not necessarily dispense with the need for detecting equalisation errors due to the existence of other reasons for equalisation errors.

In an effort to alleviate the problem of ill-convergence, the so-called centre-tap initialisation has been proposed as a means to guarantee convergence to an open-eye setting with a proper choice of the centre-tap magnitude. This algorithm fix has been, however, shown to be capable of going astray in the face of channel input correlation [14].

1.4   Raison d'être for Equalisation Error Tests

In adaptive channel equalisation it is imperative that any discrepancy between the channel input and equaliser output sequences be detected (or monitored) and corrected, if possible. The problem of testing the equalisation objective is particularly acute in the case of blind adaptive equalisation. The reason is two-fold: (i) the adaptation of the parameters is carried out in the absence of a training sequence (i.e. there is no access to the channel input to make a direct comparison between the channel input and equaliser output sequences), and (ii) on-line blind equalisation algorithms in current use have multimodal (non-convex) cost functions, which render them prone to converging to closed-eye local minima. Although it has been suggested that the converged local minimum will eventually be escaped [1, 2], the required time for that to happen can be prohibitively long. Even in conventional channel equalisation where a training sequence is used to update the equaliser parameters, various extraneous influences may lead to a corruption of the equaliser output. The detection of equalisation errors is therefore of crucial importance in blind and conventional equalisation alike, and is of substantive interest in the context of channel equalisation.

Ideally, the equalisation error test should not require an explicit knowledge of the channel input, as that would reduce the practicality of the test, as well as defeat the purpose of testing for errors at arbitrary time instants. After all, if the channel input were accessible all the time there would be no need for equalisation. A well-known experimental method to monitor the performance of an adaptive equaliser is the *eye pattern test*, which makes use of the sequence at the equaliser output (i.e. at the decision device input) as the basis for deducing whether or not ISI is prevalent. One of
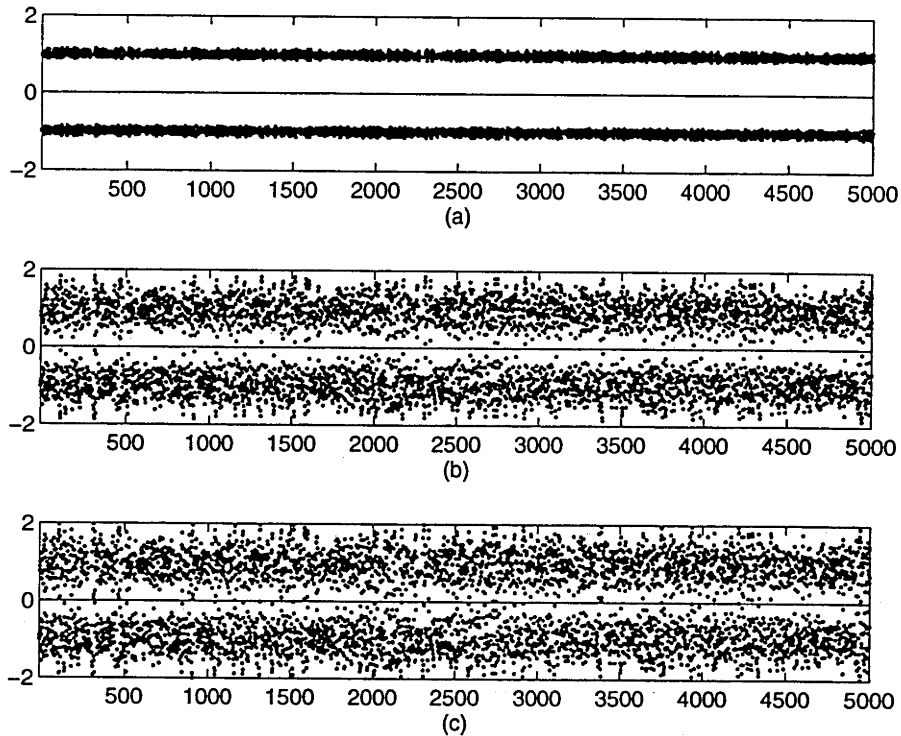
Figure 1.5 Equaliser output sequences for some channel-equaliser combination with binary inputs; in (a) and (b) the eye is open (i.e. no errors occur at the decision device output), while in (c) the eye is closed.

the problems with the eye pattern test is its complete reliance on visual perception, which renders it misleading at times. Part of the reason for this shortcoming is that while certain amount of ISI can be tolerated by the nonlinear decision device, the eye pattern closes further with increasing ISI. Additionally, it is not feasible to automate the eye pattern test so that inferences can be drawn about the shape of the eye pattern in the absence of a human interpreter. For an illustration of these assertions refer to Fig. 1.5 where various equaliser output observations are shown for some channel-equaliser combination driven by a binary input sequence. In Fig. 1.5(a) the equalisation objective is satisfied as clearly confirmed by the observed eye pattern. In Fig. 1.5(b), the observed eye pattern is severely smeared even though the equalisation objective is still satisfied. The equalisation objective is not satisfied in Fig. 1.5(c) because the observed sequence results in 20 equalisation errors at the decision device output. Note especially the difficulty associated with deducing whether or not the equalisation objective is satisfied from the eye patterns in Figs. 1.5(b) and 1.5(c). The eye pattern test is useful to detect ISI, to measure the noise margin and to determine the sensitivity to timing error, but does not always provide a precise answer to whether or not the decision device output is in error.

In view of the shortcomings of the eye pattern test and the need for a precise test for equalisation errors, we approach the problem from a completely different point of view. Since there are two

possibilities, viz. the eye is open or the eye is closed, we formulate the detection problem in the binary hypothesis testing framework. Using the terminology of detection theory, we will call the open-eye case *the null hypothesis* (denoted H_0) and the closed-eye case *the alternative hypothesis* (denoted H_1). In this framework, the objective of a test is to determine from a *probabilistic* manifestation of either H_0 or H_1 which hypothesis is true. Because of the probabilistic mechanism involved the decision made by the test is subject to errors. The process of decision making comprises choosing the hypothesis that has the highest likelihood of fitting the observations. A *test statistic* is used to reach a decision by comparing it with some threshold. Depending on the amount of *a priori* information available about the hypotheses, different statistical tests can be devised [3]. In Chapters 2 and 3 threshold tests based on various statistics are constructed for testing ill-convergence and equalisation errors.

1.5 Literature Review

Blind detection of ill-convergence or equalisation errors has not received much attention in the open literature despite its potential application in bandwidth-efficient retransmission protocols and reliability assessment in data communication systems. The first attempt at constructing a convergence test using only the decision device output observations was made in [4]. The major result of that work was the conclusion that if the channel input is an i.i.d. (independent identically distributed) binary sequence taking on values ± 1, and a linear decision-directed equaliser (LDDE) is used, the decision device output will be an uncorrelated binary sequence if the parameters of the LDDE have converged to an open-eye setting. The channel noise is assumed to be statistically independent of the decision device output. In [5, 6] we have significantly extended the results of [4] by allowing the channel input to be a correlated binary sequence. We have also provided a practical implementation for the convergence criterion in the form of a statistical test using only the decision device outputs and *a priori* knowledge of the channel input autocorrelation and fourth-order moments.

Test criteria for pulse amplitude modulation (PAM) channel inputs seem to be rather difficult to come by and construct using the channel input and decision device output statistics (see [4]). In view of this and the limited use of the test criterion for binary inputs, a completely different approach to testing for equalisation errors was presented in [7]. This new approach is quite general and can be applied to any equaliser structure as long as the channel input is a discrete-level sequence, has all its finite-length subsequences occur with nonzero probability and the channel is an approximately linear time-invariant (LTI) system.

Blind channel equalisation has been an active research area since 1970s which has resulted in the development of many blind equalisation algorithms. In his pioneering work [8], Sato proposed a blind (self-recovering) equalisation algorithm for PAM communication systems, which is very similar to the minimum mean square error (MMSE) equalisation algorithm, but uses a different average cost function. Under the assumption of uncorrelated, uniformly distributed channel input (i.e. PAM with infinitely many levels) Sato showed that his algorithm converges to a zero ISI setting from any open-eye initialisation (with CLEM < 1) provided that the equaliser is long enough. In [9] the global convergence of the Sato algorithm was proved under the assumption of sub-Gaussian channel input distribution (such as uniform distribution) and doubly infinite equaliser parameterisation. The ill-convergence of the Sato algorithm was first demonstrated in [1] for i.i.d. PAM inputs, a dynamicless channel and a finite-length equaliser. The importance of initialisation is also emphasised in the same paper. Similar results were proved in [2]. The Sato algorithm was later generalised to a family of cost functions for complex quadrature amplitude modulation (QAM), which are collectively called the Godard algorithm [10]. In [11], an alternative form and a special case of the Godard algorithm was proposed independently under the name the constant modulus algorithm (CMA). CMA was applied to the removal of multipath distortion incurred by frequency modulated (FM) signals. The key features of the Godard algorithm (or CMA) are: (i) it does not make use of the signal phase at the channel output (thereby isolating the problem of phase recovery from that of channel equalisation), and (ii) it is applicable not only to constant modulus signals using phase shift keying (PSK) or frequency shift keying (FSK), but also to nonconstant modulus signals. Godard hinted at the possibility of existence of some local minima (of the type described in [1]) on the cost surface and suggested that a centre-tap initialisation with a sufficiently large tap magnitude should be used to guarantee convergence to an open-eye setting for a doubly infinite equaliser. In [12] the Mazo-type local minima were shown to be unstable for the Godard algorithm and therefore unimportant for i.i.d. QAM channel inputs and a doubly infinite equaliser. This was also used as an argument to claim that special initialisation tactics are not necessary to guarantee convergence to open-eye parameters.

Finite equaliser parameterisation and its detrimental effects on the Godard algorithm have recently been analysed in [13]. Beneficial and detrimental effects of the channel input correlation on the convergence of CMA were documented in [14, 15]. In [14], it was noted that for certain sequential regressor probabilities the centre-tap initialisation can fail to converge to an open-eye parameter setting. In [15], using experimental results, the channel input correlation was shown to be capable of reducing the number of stable minima to two and rendering them to be closed-eye minima while open-eye parameter settings exist. An example of global convergence where only

two stable minima exist, both of which open the eye, was also presented in the same paper.

The existence of these results overshadows the universality of the centre-tap initialisation, which has yet to be proved to result in global convergence under i.i.d. channel inputs. The on-line blind equalisation algorithms using memoryless cost functions are a bane of many researchers and certainly present a whole spectrum of open problems as regards global convergence and initialisation strategies. In an effort to solve the global admissibility problem, an alternative formulation of CMA was presented in [16, 17] which is relatively insensitive to the channel input correlation and has global convergence properties. This algorithm uses a quadratic unimodal cost function in the nonlinearly transformed equaliser parameter space and has a very fast convergence rate.

The latest approach to blind channel equalisation is based on a revamp of the fractionally spaced equaliser implementation in a blind equalisation context. For i.i.d. channel inputs, fractionally spaced CMA was shown to have all its local minima correspond to optimum open-eye settings with different equalisation delays [18, 19]. Under correlated channel inputs, however, fractionally spaced CMA may exhibit ill-convergence [20].

1.6   Outline of the Thesis

In this section a chapter-by-chapter summary of the thesis is presented.

Chapter 2: Testing for the Convergence of a Linear Decision-Directed Equaliser

The material in the chapter is an extension of [4]. In addition to establishing a test criterion for correlated channel inputs, the criterion is implemented as a statistical test. The theoretical manifestation of wrong convergence at the equaliser is combined with statistical methods to construct a practical convergence test that does not require an explicit knowledge of the channel input.

Based on *a priori* knowledge of the channel input correlation, a match between the channel input and decision device output autocorrelation is shown to be a necessary and sufficient condition for the convergence of an LDDE to an open-eye parameter setting. In the derivation of the convergence criterion, statistical independence of the channel noise and decision device output sequences is assumed. The necessity of this assumption is corroborated in an example, which also serves the purpose of clarifying some aspects of the channel noise analysis in [4]. The convergence criterion is implemented in the binary hypothesis testing framework. A consistent statistical test is constructed using *a priori* knowledge of the channel input autocorrelation and

fourth-order joint moments. An extension to PAM channel inputs is conjectured by way of a heuristic variance matching criterion. A crosscorrelation-based parameter estimation technique is invoked to estimate a lower bound on the equalisation delay.

Moments of Markov chains are discussed in an appendix in order to facilitate the use of a Markov chain to generate the channel input sequence in simulation examples.

Some of the results of this chapter have been published in [5, 6].

Chapter 3: Blind Detection of Equalisation Errors

Rather than restricting the attention to the detection of ill-convergence, the issue of error detection at the decision device output is considered in a general setting. In this sense, this chapter builds upon the previous chapter, but the approach to the detection problem is completely different in order to relax the requirements for the applicability of the resulting tests. The test criterion proposed in this chapter draws on the fact that if a communication channel is equalised satisfactorily so that the decision device output does not contain any errors, the relationship of the recovered sequence to the noise-free channel output will simply be given by the channel response advanced in time by the equalisation delay. In other words, the above-mentioned relationship will be linear and time-invariant if the eye is open and the channel is an LTI system. The chapter starts with a proof of the observation that if the eye is closed, the relationship in question is no longer linear and time-invariant, but may be considered to be linear and *time-varying*. The channel input sequence is supposed to have all its finite-length subsequences occur with nonzero probability. This criterion allows the use of short observation records to determine the presence of equalisation errors.

The method of least squares (LS) is used to come by a test statistic whose probability density changes according to whether the observations of the decision device output and the equaliser input fit an LTI model with noisy output. Uniformly most powerful (UMP) tests are constructed to detect the time variance of the model, which in turn flags the presence of equalisation errors. Relations between the test parameters are established. Asymptotic analysis is carried out to get some feel for the detection performance of the tests as the sample size increases. Conditions are derived to achieve consistency. The effects of underestimated channel length and equalisation delay are also discussed and robustness of the tests to variation in these parameters is proved. As a by-product, a method for estimating the equalisation delay is proposed, which returns a lower bound on the equalisation delay if the channel includes a pure delay.

The effectiveness of the tests is illustrated with simulation examples. Given that the test statistics include a g-inverse, computationally efficient recursive and iterative methods for computing a g-inverse are presented in an appendix.

The results of this chapter have appeared in [7, 21].

Chapter 4: Least Squares Approach to Blind Channel Equalisation

A blind equalisation algorithm for constant modulus constellations is developed. The main features of the algorithm are its parsimonious use of the channel output observations to converge to, or to compute, the equaliser parameters and relative insensitivity to the channel input correlation. The communication channel to be equalised is assumed to have an approximately finite-duration impulse response (FIR) inverse.

The chapter starts with a short section explaining the reasons for the general failure of CMA. A set of nonlinear equations is derived to show the potential problem of ill-convergence and its relation to the channel output statistics which are also influenced by the channel input statistics. The effects of the channel input autocorrelation are illustrated in an example.

The new blind equalisation algorithm is developed using the ideas behind CMA with a different equaliser parameterisation. The equaliser output squared modulus is explicitly written in terms of the equaliser parameters. The coupling between the equaliser parameters is left untouched and the coupled parameters are relabelled as if they were independent parameters. This way a quadratic unimodal cost function is obtained in the transformed equaliser parameter space. The equaliser parameters are obtained by means of an LS solution to the transformed parameters and a nonunique backward transformation resulting in a phase ambiguous solution. The use of differential encoding circumvents the problem of phase ambiguity. The algorithm requires a small number of channel output observations to compute the equaliser parameters. For example, given a first-order autoregressive (AR) channel driven by an i.i.d. binary input sequence and followed by a two-tap equaliser, the number of channel output observations required to obtain the equaliser parameters is only four.

Robust estimation of the equaliser parameters is discussed using a singular value decomposition (SVD) based matrix approximation which exploits the special structure of the transformed equaliser parameters. On-line and recursive implementations of the algorithm are also developed. The effects of overparameterisation of the equaliser are investigated and the rank deficient LS problem that emerges as a result of overparameterisation is solved using some property of the nonunique LS solution. A modified version of the recursive least squares algorithm that avoids eigenvalue explosion resulting from the lack of persistent excitation is shown to be applicable to the case of overparameterisation.

Extensions to PAM constellations are indicated using a search method that may prove too time consuming and a matrix equation based on the Wiener solution for the case of constant modulus

constellations.

Publications relevant to the material in this chapter are [15, 16, 17].

Chapter 5: Fractionally Spaced Blind Equalisation of FIR Channels

The LS approach of Chapter 4 is extended to fractionally spaced equalisers which, unlike baud-rate equalisers, are capable of providing perfect equalisation for oversampled FIR channels under the so-called channel disparity condition that the resulting subchannels have no zeros common to all of them. The fractionally spaced implementation of the LS algorithm is motivated by the restriction of the baud-rate algorithm to channels with approximately FIR inverses. The salient features of the baud-rate algorithm directly carry over to the case of fractionally spaced equalisation.

The channel noise enhancement is shown to become a particularly acute problem when the channel disparity is almost but not exactly lost. Some remedies are proposed based on this observation by making the close subchannel zeros exactly identical, which leads to a loss of the channel disparity while greatly reducing the equaliser norm.

Using a modified recursive least squares implementation, the performance of the fractionally spaced LS equalisation algorithm is illustrated and compared to fractionally spaced CMA by way of simulations.

1.7   Point Summary of Contributions

Following is a list of the major contributions of the thesis.

- Proof of a test criterion for the convergence of a linear decision-directed equaliser to an open-eye parameter setting when the channel input sequence is binary and correlated, and the equaliser parameters are adapted blindly.

- Construction and analysis of a consistent threshold test based on the above test criterion.

- Development of a heuristic test criterion for the convergence of linear decision-directed equalisers when the channel input sequence is a dependent (or correlated) PAM sequence.

- Methods for estimating a lower bound on the equalisation delay introduced by the channel-equaliser combination under convergence to an open-eye parameter setting.

- Development of a general test criterion to detect the occurrence of equalisation errors without imposing any restrictions on the channel input modulation or the equaliser structure.

- Implementation of the general test criterion as uniformly most powerful tests and analysis of their optimality properties.

- Recursive and iterative computation of the test statistics.

- Illustration of the effects of channel input correlation on the convergence of CMA.

- Development of a new least squares approach to blind channel equalisation with a quadratic cost function which is comparatively parsimonious in its use of the channel output observations and is also insensitive to the channel input correlation, unlike CMA, subject to some channel input richness condition.

- Implementation of the least squares blind equalisation algorithm as a fractionally spaced equalisation algorithm for FIR channels.

- A method for curbing the channel noise enhancement encountered in fractionally spaced equalisation.

CHAPTER 2

Testing for the Convergence of a Linear Decision-Directed Equaliser

2.1 Introduction

On-line blind equalisation algorithms with memoryless cost functions are susceptible to converging to undesirable parameter settings giving rise to inferior equalisation performance. This phenomenon, which is called *ill-convergence* in the blind equalisation jargon, arises from the existence of spurious closed-eye local minima on the cost surface topology. In blind equalisation the dearth of an explicit knowledge of the channel input sequence makes it considerably difficult to establish the occurrence of ill-convergence. In this chapter we will construct a statistical test for the purpose of deciding whether or not a converged parameter setting opens the eye. The test does not resort to an explicit knowledge of the channel input, and is applicable only to linear decision-directed equalisers (LDDEs) operating on binary-input channels. An extension to M-ary channel inputs will also be considered by way of a simple heuristic test criterion.

The chapter is organised as follows. To begin with, the convergence testing problem is described and motivated by a previous work for independent binary channel inputs. In Section 2.3 a necessary and sufficient condition for the convergence of an LDDE to an open-eye setting is presented when the communication channel is an infinite-duration impulse response (IIR) system satisfying some mild condition, and driven by a binary and correlated input sequence with *a priori* known autocorrelation. In Section 2.4 a statistical test is constructed based on the convergence criterion of Section 2.3 by casting the detection problem at hand in the statistical hypothesis

testing framework. The main advantages of the proposed statistical convergence test *vis-à-vis* the eye pattern test are: (i) the statistical test has a simple construction which makes it amenable to performance analysis, and (ii) unlike the eye pattern test it does not rely on potentially deceptive visual information. The channel noise considerations and the necessity of imposing some bound on the channel noise amplitude are discussed in Section 2.5. Section 2.6 is concerned with a generalisation of the convergence test to the case of dependent M-ary channel input sequences. A simple heuristic convergence criterion based on a match between the channel input and decision device output variances is proposed for M-ary channel inputs. In Section 2.7 we briefly digress to look into the problem of equalisation delay estimation when convergence to an open-eye setting has taken place. A classical crosscorrelation-based impulse response estimation technique is invoked to estimate a lower bound on the equalisation delay. Computer simulations illustrating the application of the statistical test are presented in Section 2.8. The results of the chapter are discussed in Section 2.9. In Appendices 2.A and 2.B the main result of the chapter (Theorem 2.1) is proved, and second and fourth-order output moment expressions for Markov chains are derived, respectively.

2.2   *Problem Set-Up and Motivation*

An IIR communication channel followed by an LDDE is shown in Fig. 2.1(a). The LDDE is made up of a linear tapped-delay line with L taps and a decision device (which is a slicer in the case of binary channel inputs) connected in tandem. The adaptation mechanism for the equaliser parameters is omitted from Fig. 2.1(a) as we are only concerned with the final stage of convergence in which the tap coefficients of the equaliser wander about a local minimum in accordance with a certain probability distribution [22]. Once the adaptation mechanism is switched off, the equaliser parameters $\{\theta_i\}$ become constant.

We will assume that the magnitude of the channel noise $n(k)$ is bounded so that it does not interfere with decisions made by the slicer when the convergence has stopped. This necessarily implies that the slicer output and channel noise sequences are statistically independent when the convergence test is applied. Fig. 2.1(b) shows the equivalent channel-equaliser combination when $\{\hat{u}(k)\}$ and $\{n(k)\}$ are independent, implying that the sole source of error, if any, is *intersymbol interference* (ISI) resulting from the finite bandwidth of the channel-equaliser combination. The channel input will be assumed to be a zero-mean binary sequence taking on values ± 1 for most of the chapter. In the derivation of the test criterion, $\{u(k)\}$ will be supposed to be a wide-sense stationary sequence with *a priori* known autocorrelation $R_u(\tau) = E\{u(k+\tau)u(k)\}$ and a finite
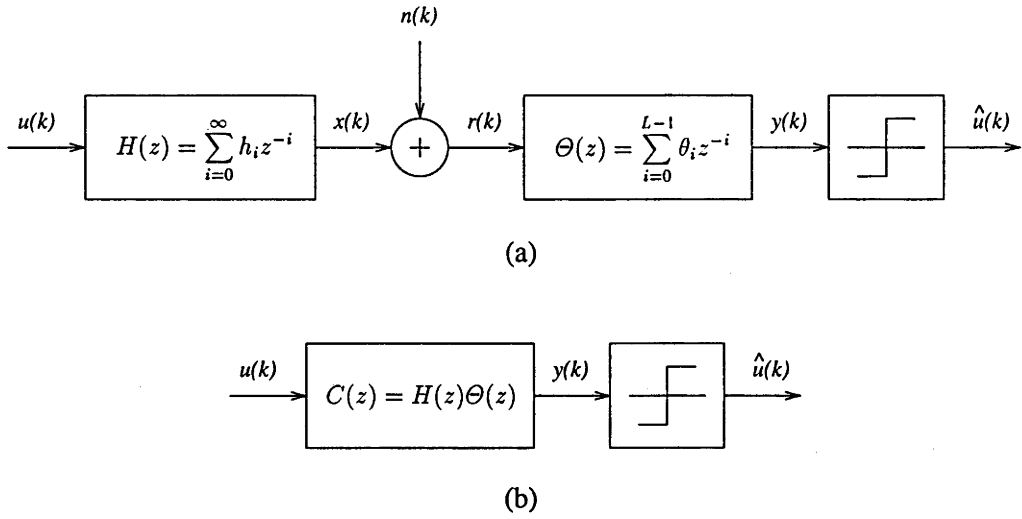
Figure 2.1 (a) IIR communication channel and LDDE models at baseband, and (b) equivalent system if $\{n(k)\}$ is statistically independent of $\{\hat{u}(k)\}$.

span of dependence. When constructing the statistical test, we will further assume that the input sequence is fourth-order weakly stationary and prior knowledge of certain fourth-order moments of $\{u(k)\}$ is available. The communication channel will be assumed to be causal so that $h_i = 0$ for $i < 0$.

Referring to Fig. 2.1(b) where $\{\hat{u}(k)\}$ and $\{n(k)\}$ are assumed to be statistically independent, the slicer output sequence $\{\hat{u}(k)\}$ can be written as

$$\hat{u}(k) = \mathrm{sgn}\left(y(k)\right) \tag{2.1}$$

with

$$y(k) = \sum_{i=0}^{\infty} c_i u(k-i) \tag{2.2}$$

where the impulse response $\{c_i\}$ of the equivalent system $C(z)$ is given by the convolution of $\{h_i\}$ and $\{\theta_i\}$ and the signum function $\mathrm{sgn}(\cdot)$ is defined by

$$\mathrm{sgn}(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0. \end{cases} \tag{2.3}$$

For binary channel input sequences the equalisation objective described in Section 1.2 takes the following form

$$\hat{u}(k) = u(k-\Delta) \quad \text{or} \quad \hat{u}(k) = -u(k-\Delta) \quad \forall k \tag{2.4}$$

where Δ is a fixed nonnegative integer representing the equalisation delay. Recall from Section 1.2 that if Eq. (2.4) holds, we say that *the eye is open*; otherwise, *the eye is closed*.

The case of finite-duration impulse response (FIR) channels with i.i.d. (independent identically distributed) binary inputs was considered in [4]. The main result of [4] can be summarised as follows. Under the assumption of statistical independence between the slicer output and channel noise sequences, an LDDE converges to an open-eye parameter setting if and only if the slicer output sequence $\{\hat{u}(k)\}$ is uncorrelated. It is not immediately obvious, however, that the same result should carry over to the case of IIR channels. Consider, for example, an IIR channel-equaliser combination that happens to be an all-pass system as a result of convergence to a closed-eye local minimum. The autocorrelation of the channel input $\{u(k)\}$ and the tapped-delay line output $\{y(k)\}$ would then be identical. The possibility of an all-pass channel-equaliser combination $C(z)$ may be thought to jeopardise the validity of the convergence test criterion in [4] for IIR channels (see Fig. 2.1(b)). As a matter of fact, this concern is unwarranted owing to the nonlinear characteristic of the slicer. As will be proved later, the sequence at the slicer input must be "signwise" independent to ensure that the binary sequence at its output is uncorrelated. If the binary sequence at the channel input is independent, signwise independence at the slicer input is assured if the overall system impulse response $\{c_i\}$ has a cursor whose absolute value is larger than the ℓ_1 norm (i.e. sum of absolute values) of the rest. This observation also highlights the importance of using the slicer output sequence rather than its input sequence (the tapped-delay line output) for convergence testing purposes.

The need for an analogous result for correlated channel input sequences is apparent at the outset since certain channel coding schemes purposely introduce correlation between successive binary symbols to be transmitted through the communication channel. In blind adaptation of the equaliser parameters, it is desirable to make a decision as to whether $\{\theta_i\}$ has converged to a setting for which Eq. (2.4) holds only by drawing upon the information available at the noisy channel output onwards, as well as *a priori* knowledge of the channel input statistics. Obviously, if Eq. (2.4) holds, the channel input sequence $\{u(k)\}$ and the slicer output sequence $\{\hat{u}(k)\}$ will have the same autocorrelation. In the next section, we will prove that the converse is also true subject to some restriction on the noise magnitude so that the tapped-delay line output can be written as in Eqs. (2.1) and (2.2).

2.3   Convergence Criterion for Correlated Binary Input Sequences

In most studies of the channel equalisation problem the channel input sequence $\{u(k)\}$ is almost invariably assumed to consist of independent symbols. The validity of this assumption is often somewhat dubious from the standpoint of practice. Certain channel coding schemes employed to improve the noise immunity of transmitted symbols inevitably lead to correlation between successive symbols, hence the need to deal with correlated input sequences.

We will assume that the combined channel-equaliser impulse response $\{c_i\}$ obeys the following inequality for some *finite* integers $0 \leq I \leq J$

$$\min_k \left| \sum_{i=I}^{J} c_i u(k-i) \right| > \sum_{i=0}^{I-1} |c_i| + \sum_{i=J+1}^{\infty} |c_i| \tag{2.5}$$

which implies that the slicer output can be expressed as

$$\hat{u}(k) = \mathrm{sgn}\left(\sum_{i=I}^{J} c_i u(k-i) \right) \quad \forall k. \tag{2.6}$$

Eq. (2.5) imposes a mild restriction on $C(z)$ which is automatically satisfied by most exponentially stable channel-equaliser combinations with a proper choice of I and J. Note that Eq. (2.5) is unconditionally satisfied for FIR channels.

Following is a formal statement of our test criterion for the convergence of an LDDE to an open-eye setting when the channel input sequence is binary and correlated.

Theorem 2.1 Let $\{u(k)\}$ be a wide-sense stationary sequence of binary, zero-mean random variables assuming values ± 1, and let $\hat{u}(k)$ be as given in Eqs. (2.1) and (2.2) where $\{c_i\}$ obeys the inequality in Eq. (2.5). Suppose that the autocorrelation $R_u(\tau)$ of the channel input sequence $\{u(k)\}$ is known and has a finite support set such that $R_u(\tau) = 0$ for $|\tau| > m$, where m is a nonnegative integer[1], and that the argument of the signum function never becomes zero[2]. Then, $R_{\hat{u}}(\tau) = R_u(\tau)$ for all τ if and only if

$$\hat{u}(k) = \mathrm{sgn}(c_\Delta) u(k - \Delta) \tag{2.7}$$

for all k and some fixed integer $\Delta \geq 0$.

Refer to Appendix 2.A for a proof of Theorem 2.1. A few remarks about the theorem are in

---

[1] In other words, $\{u(k)\}$ is an m-dependent sequence.

[2] This assumption is necessary to resolve the ambiguity arising from a decision on $\hat{u}(k)$ when $y(k) = 0$. In practice, redefining the decision rule in Eq. (2.3) to assign $x = 0$ to either 1 or -1 would dispense with the requirement $y(k) \neq 0$ without affecting any of the results.

order:

Remarks 2.1

    i. As a special case of the above theorem, if the channel input sequence is uncorrelated (or independent), i.e. $R_u(0) = 1$ and $R_u(\tau) = 0$ for $\tau \neq 0$, then the LDDE converges to an open-eye parameter setting if and only if the slicer output sequence is uncorrelated. This is simply a generalisation of the result for FIR channels given in [4] to IIR channels obeying Eq. (2.5).

    ii. The terms uncorrelatedness and independence can be used interchangeably for the binary sequences under consideration. Refer to the argument leading to Eq. (2.64) for a proof of this assertion.

    iii. If $|c_\Delta| = \max_{i=0,\dots,\infty} |c_i|$, then Eq. (2.7) holds if and only if the following inequality is satisfied

$$|c_\Delta| > \frac{1}{2}\sum_{i=0}^{\infty} |c_i| \tag{2.8}$$

which is an open-eye condition and assures that the sign of $y(k)$ is always determined by the binary input corresponding to the cursor c_Δ. The condition given in Eq. (2.8) tacitly assumes that all possible subsequences of the binary input sequence occur with nonzero probability.

    iv. Eqs. (2.4) and (2.7) are equivalent by virtue of their implication on the equalisation performance.

    v. The result of Theorem 2.1 does not readily extend to M-ary communication systems where $u(k)$ can take on more than two possible values. More will be said on this in Section 2.6.

2.4 Statistical Test for Correct Convergence

2.4.1 Preliminaries

Before a decision can be made as to whether or not the true autocorrelation of $\{\hat{u}(k)\}$ is identical to that of $\{u(k)\}$, the former needs to be estimated from a finite-length observation of the slicer output. A decision made in favour of identical autocorrelations implies that the LDDE has converged to a parameter setting that conforms to Eq. (2.7). If the decision is against identical autocorrelations, then the conclusion will be that the equaliser has failed to converge to an open-eye parameter setting.

We cast the decision problem at hand into the hypothesis testing framework by defining the null and alternative hypotheses as follows:

H_0 : Convergence to an open-eye minimum

H_1 : Convergence to a closed-eye minimum.

Theorem 2.1 allows us to write the above hypotheses in an equivalent way

$H_0 : R_u(\tau) = R_{\hat{u}}(\tau) \; \forall \tau$

$H_1 : R_u(\tau) \neq R_{\hat{u}}(\tau)$ for some τ.

For N observations $\{\hat{u}(k), \hat{u}(k+1), \ldots, \hat{u}(k+N-1)\}$ of the slicer output, a *biased* estimate of $R_{\hat{u}}(\tau)$ for $0 \leq \tau < N$ is given by

$$\widehat{R}_{\hat{u}}(\tau) = \frac{1}{N} \sum_{i=0}^{N-\tau-1} \hat{u}(k+i+\tau)\hat{u}(k+i). \tag{2.9}$$

In the statistical test, we will use the biased estimate of $R_{\hat{u}}(\tau)$ rather than its unbiased version which can easily be obtained by replacing the factor in front of the summation with $1/(N-\tau)$ in Eq. (2.9). There are a number of reasons for choosing the biased estimate. The biased estimate, unlike its unbiased counterpart, always results in a positive semidefinite covariance matrix. The variance of the biased estimate is $O(1/N)$ irrespective of the correlation lag τ, whereas the "tail" of the unbiased estimate tends to have an increased variance as τ approaches N, thereby resulting in an erratic behaviour for large τ [23, 24].

We will assume that $\{u(k)\}$ is fourth-order weakly stationary. If the overall channel-equaliser combination is time-invariant during the test interval, the slicer output sequence $\{\hat{u}(k)\}$ will also be fourth-order weakly stationary. Then, the biased estimate of $R_{\hat{u}}(\tau)$ has mean

$$E\left\{\widehat{R}_{\hat{u}}(\tau)\right\} = \left(1 - \frac{\tau}{N}\right) R_{\hat{u}}(\tau). \tag{2.10}$$

and covariance

$$\text{Cov}\left\{\widehat{R}_{\hat{u}}(\tau_1), \widehat{R}_{\hat{u}}(\tau_2)\right\} = E\left\{\widehat{R}_{\hat{u}}(\tau_1)\widehat{R}_{\hat{u}}(\tau_2)\right\} - E\left\{\widehat{R}_{\hat{u}}(\tau_1)\right\}E\{\widehat{R}_{\hat{u}}(\tau_2)\} \tag{2.11a}$$

$$= \frac{1}{N^2} \sum_{i=0}^{N-\tau_1-1} \sum_{j=0}^{N-\tau_2-1} m_{4,\hat{u}}(\tau_1, \tau_2 + j - i, j - i)$$

$$- \left(1 - \frac{\tau_1}{N}\right)\left(1 - \frac{\tau_2}{N}\right) R_{\hat{u}}(\tau_1)R_{\hat{u}}(\tau_2) \tag{2.11b}$$

where $m_{4,\hat{u}}(\ldots)$ denotes the fourth-order joint moment of $\{\hat{u}(k)\}$ defined by

$$m_{4,\hat{u}}(\tau_1, \tau_2, \tau_3) \triangleq E\{\hat{u}(k)\hat{u}(k + \tau_1)\hat{u}(k + \tau_2)\hat{u}(k + \tau_3)\}. \tag{2.12}$$

Note that the time index k has been dropped from the arguments of $m_{4,\hat{u}}(\ldots)$ under the assumption that $\{\hat{u}(k)\}$ is a fourth-order weakly stationary sequence.

2.4.2 *Asymptotic Sampling Distribution of Autocorrelation Estimates*

Under certain conditions on $\{\hat{u}(k)\}$ the central limit theorem holds for the asymptotic sampling distribution of $\widehat{R}_{\hat{u}}(\tau)$. The importance of the asymptotic normality (as $N \to \infty$) of $\widehat{R}_{\hat{u}}(\tau)$ stems from its influence on the approximate distribution of the test statistic as will be explained in the next subsection. To facilitate the asymptotic normality result, we will consider an alternative way of expressing the biased autocorrelation estimate of the slicer output sequence

$$\widehat{R}_{\hat{u}}(\tau) = \frac{1}{N} \sum_{i=0}^{a-1} S_i(\tau) \tag{2.13}$$

where

$$S_i(\tau) = \sum_{j=0}^{l-1} \hat{u}(k + ja + \tau + i)\hat{u}(k + ja + i) \tag{2.14}$$

and

$$a = \tau + w + 1 \tag{2.15}$$

with w denoting the span of dependence of $\{\hat{u}(k)\}$. The relation between the sample size N and the other parameters can be found by considering the last term in $S_{a-1}(\tau)$ corresponding to $j = l - 1$. Setting $k + (l - 1)a + \tau + a - 1 = k + N - 1$ gives

$$N = a(l + 1) + \tau. \tag{2.16}$$

Note that $S_i(\tau)$ is a sum of independent random variables since the separation between its successive terms is $(k + (j + 1)a + i) - (k + ja + \tau + i) = w + 1 > w$. Then, by the central limit theorem,

$$\lim_{l \to \infty} \Pr\left\{ \frac{1}{\sqrt{l}} \left(S_i(\tau) - E\{S_i(\tau)\} \right) < c \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{c} e^{-t^2/2} dt, \quad i = 0, 1, \ldots, a - 1 \tag{2.17}$$

where we note that $\mathrm{Var}\{S_i^2(\tau)\} = l$ for $i = 0, 1, \ldots, a - 1$. Although the $S_i(\tau)$ are sums of independent random variables, the $S_i(\tau)$ themselves are not generally independent random

variables. Nevertheless, recalling that the sum of Gaussian random variables (whether independent or not) is also a Gaussian random variable, we conclude that as $l \to \infty$ (or as $N \to \infty$) the distribution of $\widehat{R}_{\hat{u}}(\tau)$ approaches multivariate Gaussian. We have thus proved the following:

Theorem 2.2 *The asymptotic joint distribution of* $\left[\widehat{R}_{\hat{u}}(0), \widehat{R}_{\hat{u}}(1), \ldots, \widehat{R}_{\hat{u}}(K)\right]^T$, *whose entries are defined in Eq. (2.9), is multivariate Gaussian for any* K *with means and covariances given by Eqs. (2.10) and (2.11b), respectively, provided that* $\{\hat{u}(k)\}$ *is a correlated binary sequence with a finite span of dependence.*

Generalisation of the above result to arbitrary dependent sequences of stationary random variables was made in [25] under the uniformly strong mixing condition. A similar result for linear stationary processes can be found in [23, 26].

2.4.3 The Test Statistic

The uncertainty about the autocorrelation of the slicer output sequence when the eye is closed, which is a consequence of not knowing the channel parameters h_i, precludes us from determining the exact distribution of the autocorrelation estimate $\widehat{R}_{\hat{u}}(\tau)$ under the alternative hypothesis. Therefore, we have a *composite* alternative hypothesis under which the distribution of $\widehat{R}_{\hat{u}}(\tau)$ depends on the resulting $C(z)$. On the other hand, the null hypothesis is simple since the distribution of $\widehat{R}_{\hat{u}}(\tau)$ can be found as described in Theorem 2.2 by replacing $\hat{u}(k)$ with $u(k)$. We propose the following quadratic form as our test statistic:

$$\Lambda = z^T \Sigma^- z \tag{2.18}$$

where

$$z = \left[\widehat{R}_{\hat{u}}(1) - R_u(1), \widehat{R}_{\hat{u}}(2) - R_u(2), \ldots, \widehat{R}_{\hat{u}}(K) - R_u(K)\right]^T, \quad K < N \tag{2.19}$$

is a $K \times 1$ vector of the difference between the slicer output autocorrelation estimates and the channel input autocorrelation for $1 \leq \tau \leq K$, and

$$\Sigma = [\sigma_{ij}] \quad i, j = 1, 2, \ldots, K \tag{2.20}$$

is the covariance matrix of the autocorrelation estimates of $\{\hat{u}(k)\}$ under H_0, whose entries are defined by $\sigma_{ij} = \text{Cov}\{\widehat{R}_u(i), \widehat{R}_u(j)\}$. Assuming *a priori* knowledge of $R_u(\tau)$ for $\tau = 1, 2, \ldots, K$ and of $m_{4,u}(\ldots)$ for moment lags in the range of summation indices in Eq. (2.11b), the entries of Σ can be computed from Eq. (2.11b) by replacing $\hat{u}(k)$ with $u(k)$. The covariance matrix Σ need

be computed only once. In Eq. (2.18) Σ^- denotes a *g-inverse*[3] of Σ. If Σ is nonsingular, then $\Sigma^- = \Sigma^{-1}$. The choice of the maximum autocorrelation lag K is not always straightforward as it reflects an expectation for the maximum spread of the slicer output autocorrelation when the eye is closed, which is closely related to the quantity $J - I$ (see Eq. (2.5)). Determination of K is further compounded by the fact that setting K to an unduly large value will tend to impair the detection performance. The slicer output variance estimate $\widehat{R}_{\hat{u}}(0)$ is ignored in the test statistic since its expected value is invariant to the underlying hypothesis, thereby bearing no useful information.

We will allude to the following result when we compute the test threshold and make an assessment of the detection performance:

Theorem 2.3 [27] *Given a vector z distributed according to $\mathcal{N}(\mu, K)$ (i.e. multivariate Gaussian with mean μ and covariance K) where K may be singular, the statistic*

$$z^T C z \tag{2.21}$$

has a chi-square distribution with d degrees of freedom and noncentrality parameter ρ, if and only if

i. $(KC)^3 = (KC)^2$

ii. $\mathrm{col}(KC\mu) \subset \mathrm{col}(KCK)$

iii. $(C\mu)^T K(C\mu) = \mu^T C\mu$

in which case $d = \mathrm{tr}(CK)$ and $\rho = (C\mu)^T KCK(C\mu)$.

We used the notation $\mathrm{col}(\cdot)$ and $\mathrm{tr}(\cdot)$ above to denote the column space and the trace of a matrix, respectively. Theorem 2.3 provides us with useful information concerning the asymptotic distribution of the test statistic under the null and alternative hypotheses, as delineated below.

Distribution of Λ under H_0

By Theorem 2.2, under H_0, the asymptotic distribution of the vector of slicer output autocorrelation estimates

$$\left[\widehat{R}_{\hat{u}}(1), \widehat{R}_{\hat{u}}(2), \ldots, \widehat{R}_{\hat{u}}(K)\right]^T$$

is multivariate Gaussian with mean

$$[R_u(1), R_u(2), \ldots, R_u(K)]^T$$

---

[3]A g-inverse of A (denoted A^-) is given by any matrix satisfying the equality $AA^-A = A$. Refer to Appendix 3.A.1 for a detailed treatment of g-inverses.

and covariance Σ defined in Eq. (2.20). Thus, the vector z is approximately distributed according to $\mathcal{N}(0, \Sigma)$ for large N. Setting $\mu = 0$, $K = \Sigma$ and $C = \Sigma^{-}$, we see that the test statistic Λ satisfies the conditions of Theorem 2.3, which implies that Λ has an approximately central chi-square distribution for large N with $d = \text{rank}(\Sigma)$ and $\rho = 0$.

Distribution of Λ under H_1

Under H_1, the vector z in Eq. (2.19) is asymptotically Gaussian with nonzero mean and a covariance matrix which is in general different to Σ. With $\mu \neq 0$, $K \neq \Sigma$ and $C = \Sigma^{-}$, the conditions of Theorem 2.3 cannot be satisfied for every possible K. Thus, the test statistic Λ does not usually have a chi-square distribution if the alternative hypothesis is prevalent.

In view of the distinct properties of the distribution of Λ under the two hypotheses, we construct the following threshold test to detect convergence to a closed-eye minimum:

$$\Lambda \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{2.22}$$

where η is the test threshold. The probability of detection (or the test power) and the probability of false alarm are defined by $P_D = \Pr\{\Lambda > \eta \mid H_1\}$ and $P_{FA} = \Pr\{\Lambda > \eta \mid H_0\}$, respectively. The test threshold η is chosen to satisfy the equality $P_{FA} = \alpha$, where α is the significance level of the test. Once α is decided upon and fixed, η can be determined simply by referring to a chi-square distribution table, using the relationship between η and α given by $\Pr\{\chi_d^2 > \eta\} = \alpha$. The choice of η involves a trade-off between a high P_D and a small P_{FA} as these are usually conflicting objectives.

Since the distribution of Λ is not well-defined under H_1, the success of the test depends, to a large extent, on how far apart the means of Λ will be under H_0 and H_1. For the moment we will content ourselves with the intuitive result that Λ is likely to have a larger mean under H_1 than that of χ_d^2. An argument substantiating this result is given in the next subsection.

2.4.4   Consistency of the Test

A statistical test is said to be *consistent* if $P_D \to 1$ as the sample size $N \to \infty$. The proof of consistency draws on the behaviour of the expected value of the test statistic under H_0 and H_1. The fact that the approximate null hypothesis distribution of Λ is not affected by N (with an approximately constant mean value $E\{\Lambda \mid H_0\} \approx d$) as long as N is large reduces the consistency analysis to the evaluation of only $E\{\Lambda \mid H_1\}$ as a function of N.

Using the definition of the fourth-order joint cumulant of $\{\hat{u}(k)\}$

$$c_{4,\hat{u}}(\tau_1, \tau_2, \tau_3) = m_{4,\hat{u}}(\tau_1, \tau_2, \tau_3) - R_{\hat{u}}(\tau_1)R_{\hat{u}}(\tau_3 - \tau_2) - R_{\hat{u}}(\tau_3)R_{\hat{u}}(\tau_2 - \tau_1)$$
$$- R_{\hat{u}}(\tau_2)R_{\hat{u}}(\tau_3 - \tau_1) \tag{2.23}$$

and

$$E\left\{\widehat{R}_{\hat{u}}(\tau_1)\right\} E\left\{\widehat{R}_{\hat{u}}(\tau_2)\right\} \approx R_{\hat{u}}(\tau_1)R_{\hat{u}}(\tau_2) \tag{2.24}$$

the covariance expression in Eq. (2.11b) can be written approximately as

$$\text{Cov}\left\{\widehat{R}_{\hat{u}}(\tau_1), \widehat{R}_{\hat{u}}(\tau_2)\right\} \approx \frac{1}{N^2} \sum_{i=0}^{N-\tau_1-1} \sum_{j=0}^{N-\tau_2-1} c_{4,\hat{u}}(\tau_1, \tau_2 + j - i, j - i)$$
$$+ R_{\hat{u}}(j - i)R_{\hat{u}}(\tau_2 - \tau_1 + j - i) + R_{\hat{u}}(\tau_2 + j - i)R_{\hat{u}}(j - i - \tau_1). \tag{2.25}$$

As shown in [23] the above expression can be simplified by making a change of variable from i and j to $n = j - i$, resulting in

$$\text{Cov}\left\{\widehat{R}_{\hat{u}}(\tau_1), \widehat{R}_{\hat{u}}(\tau_2)\right\} \approx \frac{1}{N} \sum_{n=-(N-\tau_1-1)}^{N-\tau_2-1} \left(1 - \frac{\kappa(n) + \tau_2}{N}\right)\left(R_{\hat{u}}(n)R_{\hat{u}}(n + \tau_2 - \tau_1)\right.$$
$$\left. + R_{\hat{u}}(n + \tau_2)R_{\hat{u}}(n - \tau_1) + c_{4,\hat{u}}(\tau_1, \tau_2 + n, n)\right) \tag{2.26}$$

where

$$\kappa(n) = \begin{cases} n, & n > 0 \\ 0, & -\tau_2 + \tau_1 \leq n \leq 0 \\ -n - \tau_2 + \tau_1, & -(N - \tau_1 - 1) \leq n < -\tau_2 + \tau_1. \end{cases} \tag{2.27}$$

For large N, Eq. (2.26) can be further simplified as

$$\text{Cov}\left\{\widehat{R}_{\hat{u}}(\tau_1), \widehat{R}_{\hat{u}}(\tau_2)\right\} \approx \frac{1}{N} \sum_{n=-\infty}^{\infty} \left(R_{\hat{u}}(n)R_{\hat{u}}(n + \tau_2 - \tau_1) + R_{\hat{u}}(n + \tau_2)R_{\hat{u}}(n - \tau_1)\right.$$
$$\left. + c_{4,\hat{u}}(\tau_1, \tau_2 + n, n)\right). \tag{2.28}$$

The summation in Eq. (2.28) is invariant to N. Thus, the covariance matrix of $\widehat{R}_{\hat{u}}(\tau)$ is approximately inversely proportional to N, as N gets large. In general, the random vector z which is approximately distributed according to $\mathcal{N}(\mu, K)$ can be written approximately in terms of a constant matrix Q that does not vary with N and a vector of independent standard Gaussian random variables $n \sim \mathcal{N}(0, I)$

$$z \approx \mu + \frac{1}{\sqrt{N}} Q n$$

where Q is related to K by $K \approx (1/N)QQ^T$. Also, the covariance matrix Σ can be expressed in terms of a constant matrix S and N as $\Sigma \approx (1/N)S$. Thus, the test statistic under H_1 can be approximated as

$$z^T \Sigma^- z \approx \left(\mu + \frac{1}{\sqrt{N}} Qn \right)^T \Sigma^- \left(\mu + \frac{1}{\sqrt{N}} Qn \right) \tag{2.29a}$$

$$= N(\mu^T S^- \mu) + \sqrt{N}(\mu^T S^- Qn) + \sqrt{N}(n^T Q^T S^- \mu)$$

$$+ n^T Q^T S^- Qn. \tag{2.29b}$$

Taking the expectation of both sides above gives

$$E\{\Lambda \mid H_1\} \approx N(\mu^T S^- \mu) + \text{tr}(Q^T S^- Q) \tag{2.30}$$

where we note that the first term is equal to $\mu^T \Sigma^- \mu$.

We formally have the following result:

Theorem 2.4 *For large N, the mean of the test statistic Λ under H_1 increases monotonically with N if $\mu^T \Sigma^- \mu > 0$.*

We make the following observations:

Remarks 2.2

    i. Monotone increase in $E\{\Lambda \mid H_1\}$ with N implies that $\lim_{N \to \infty} P_D = 1$ (i.e. the test is consistent), since $\lim_{N \to \infty} E\{\Lambda \mid H_0\} = d$ and $\lim_{N \to \infty} E\{\Lambda \mid H_1\} = \infty$.

    ii. If Σ is nonsingular, then it is positive definite, meaning that $\mu^T \Sigma^- \mu = \mu^T \Sigma^{-1} \mu = 0$ if and only if $\mu = 0$. Thus, for nonsingular Σ consistency of the test is guaranteed unconditionally.

    iii. If Σ is singular, then consistency is dependent on the inequality $\mu^T \Sigma^- \mu > 0$ being satisfied for the resultant μ. A full-rank g-inverse of Σ guarantees consistency unconditionally. Refer to Appendix 3.A.1 for the details of how to compute a full-rank (positive definite) g-inverse.

2.5   Channel Noise Effects

Theorem 2.1 assumes that $\{\hat{u}(k)\}$ and $\{n(k)\}$ are statistically independent. In this section, we explain the necessity of this assumption with a simple example.

We will first quantify the contributions of the channel input $u(k)$ and the channel noise $n(k)$ to the tapped-delay line output $y(k)$. To do so, let us write $y(k)$ as

$$y(k) = y^u(k) + y^n(k) \tag{2.31}$$

where $y^u(k)$ is the contribution of the channel input to $y(k)$

$$y^u(k) = \sum_{i=0}^{\infty} c_i u(k-i) \tag{2.32}$$

and $y^n(k)$ is the contribution of the channel noise

$$y^n(k) = \sum_{i=0}^{L-1} \theta_i n(k-i). \tag{2.33}$$

Then, the statistical independence of $\{\hat{u}(k)\}$ and $\{n(k)\}$ is equivalent to

$$\Pr\left\{|y^n(k)| > \min_l |y^u(l)|\right\} = 0 \quad \forall k. \tag{2.34}$$

This inequality implicitly specifies what has to be the bound on the magnitude of $n(k)$, given a knowledge of $\{c_i\}$ and $\{\theta_i\}$. The following example illustrates the importance of Eq. (2.34) so far as Theorem 2.1 is concerned.

Example 2.1 Let $\{u(k)\}$ be an i.i.d. binary sequence and the channel noise be a linear process given by $n(k) = \sum_{i=0}^{\infty} h_i \nu(k-i)$ where $\{\nu(k)\}$ is an i.i.d. zero-mean Gaussian noise with variance σ_ν^2. We assume perfect equalisation, i.e. $\Theta(z) = H^{-1}(z)$, where the channel $H(z)$ is a stable minimum phase system.

Then, the output of $\Theta(z)$ is given by $y(k) = u(k) + \nu(k)$ which is an i.i.d. sequence with the following probability distribution function:

$$\Pr\{y(k) < c\} = \frac{1}{2\sqrt{2\pi\sigma_\nu^2}} \left(\int_{-\infty}^{c} \exp\left(-\frac{t-1}{2\sigma_\nu^2}\right) dt + \int_{-\infty}^{c} \exp\left(-\frac{t+1}{2\sigma_\nu^2}\right) dt \right). \tag{2.35}$$

Since $y(k)$ is i.i.d., so is $\hat{u}(k) = \text{sgn}(y(k))$, implying $R_{\hat{u}}(\tau) = R_u(\tau) \ \forall\tau$ although the eye is closed. To see this, consider the conditional probability

$$\Pr\{\hat{u}(k) = 1 \mid u(k) = 1\} = \frac{1}{\sqrt{2\pi\sigma_\nu^2}} \int_0^{\infty} \exp\left(-\frac{t-1}{2\sigma_\nu^2}\right) dt \tag{2.36}$$

which is neither one nor zero, thereby constituting a violation of Eq. (2.4). In order for the condition $\hat{u}(k) = u(k) \ \forall k$ to be satisfied, $\nu(k)$ must be bounded such that $-1 < \nu(k) < 1 \ \forall k$, as

stipulated by Eq. (2.34). □

This example shows that the condition $R_{\hat{u}}(\tau) = R_u(\tau)$ $\forall \tau$ does *not* necessarily assure the satisfaction of Eq. (2.4) if $\{n(k)\}$ is unbounded (e.g. Gaussian noise). The match of channel input and slicer output autocorrelations is still a necessary condition for convergence to an open-eye minimum whether or not the noise is bounded. For it to be also a sufficient condition, however, the channel noise has to comply with Eq. (2.34). This conclusion is in conflict with some of the claims made in [4] regarding the channel noise.

It is also worth noting that if the equaliser has already converged to a closed-eye parameter setting, the channel noise is not likely to cause the slicer output autocorrelation to be equal to the channel input autocorrelation. Thus, the condition in Eq. (2.34) has more significance under correct convergence than ill-convergence.

2.6 On Generalisation to Dependent PAM Inputs

The amount of information conveyed at a given time instant k can be increased by allowing $u(k)$ to take on more than two possible values. A multilevel input sequence can be obtained, for example, by means of pulse amplitude modulation (PAM) which produces a so-called M-ary input sequence drawn from the set $\mathbb{S} = \{-(M-1), -(M-3), \ldots, -3, -1, 1, 3, \ldots, M-3, M-1\}$ where $M > 2$ is an even integer. An LDDE for M-ary inputs consists of a tapped-delay line followed by an M-ary quantiser. Supposing that the channel noise $n(k)$ is bounded and sufficiently small such that $\{\hat{u}(k)\}$ and $\{n(k)\}$ are statistically independent sequences (i.e. Eq. (2.34) is satisfied), the quantiser output can be written as

$$\hat{u}(k) = Q_M \left(\sum_{i=0}^{\infty} c_i u(k-i) \right) \tag{2.37}$$

where $Q_M(\cdot)$ is a memoryless quantiser replacing the slicer, defined by

$$Q_M(x) \triangleq \sum_{i=1-M/2}^{M/2-1} \text{sgn}(x + 2i). \tag{2.38}$$

It was shown in [4] that if the channel input is an i.i.d. M-ary sequence with equiprobable symbols, then a necessary and sufficient condition for the convergence of an LDDE to an open-eye minimum is that the quantiser output be an i.i.d. M-ary sequence with equiprobable symbols. The desirability of extending this convergence criterion to dependent M-ary sequences was expressed in [4], although no attempt was made towards its fruition. This is partly due to the anticipated complexity of such criteria offering little benefit, if any, in return for the mammoth effort put into

their construction. In this section, baulking at the idea of coming up with a test criterion based on a comprehensive check of the quantiser output distribution, we propose a simple heuristic criterion for the convergence of an LDDE to an open-eye parameter setting when the channel input is a dependent M-ary sequence with equiprobable symbols. Unfortunately, the simplicity of this heuristic criterion does not carry over to i.i.d. channel inputs.

Using a minimum finite-length representation for the quantiser output in terms of the c_i

$$\hat{u}(k) = Q_M \left(\sum_{i=I}^{J} c_i u(k-i) \right) \tag{2.39}$$

which is the M-ary counterpart of Eq. (2.6), we will concentrate on the symbol probabilities at the quantiser output

$$\Pr\{\hat{u}(k) = s\} = \Pr \left\{ Q_M \left(\sum_{i=I}^{J} c_i u(k-i) \right) = s \right\}, \quad \forall s \in \mathbb{S}. \tag{2.40}$$

We will assume that the channel input is a dependent sequence such that $u(k)$ and $u(k-\tau)$ is dependent for at least one lag $|\tau| \leq J - I$. Note that if the eye is open, the channel input and quantiser output symbol probabilities will be equal; that is,

$$\Pr\{\hat{u}(k) = s\} = \Pr\{u(k) = s\} = \frac{1}{M}, \quad \forall s \in \mathbb{S}. \tag{2.41}$$

The quantiser output symbol probabilities in Eq. (2.40) can be written in terms of the union of the joint probabilities of input sequences that result in $\hat{u}(k) = s$

$$\Pr\{\hat{u}(k) = s\} = \Pr \left\{ \bigcup_{i=1}^{M^{J-I}} \mathcal{U}_s(i) \right\} \tag{2.42a}$$

$$= \sum_{i=1}^{M^{J-I}} \Pr\{\mathcal{U}_s(i)\} \tag{2.42b}$$

where $\mathcal{U}_s(i)$ denotes the ith combination of all possible input subsequences of length $J - I + 1$ $u = [u(k-I), u(k-I-1), \ldots, u(k-J)]^T$ that results in $\hat{u}(k) = s$. We assume that an Mth of the input sequences of length $J - I + 1$ yield $\hat{u}(k) = s$, hence the upper limit of the summation in Eq. (2.42b). If this were not the case, symbols at the quantiser output would not be equiprobable. The question is now whether or not Eq. (2.41) will still be true if the eye is closed or $J > I$.

Referring to the identity

$$\Pr\{u(k) = s_1\} = \sum_{s_2 \in \mathbb{S}} \cdots \sum_{s_{J-I+1} \in \mathbb{S}} \Pr\{u(k-I) = s_1, u(k-I-1) = s_2, \ldots, u(k-J) = s_{J-I+1}\}$$
(2.43)

we see that the equality $\Pr\{\hat{u}(k) = s\} = \Pr\{u(k) = s\}$ is guaranteed if one entry of $\mathcal{U}_s(i)$ in Eq. (2.42b) is constant for all i, which would mean that the eye is open[4] and is therefore ruled out. On the other hand, if $\{u(k)\}$ is an independent sequence, Eq. (2.42b) can be rewritten as

$$\Pr\{\hat{u}(k) = s\} = \sum_{i=1}^{M^{J-I}} \frac{1}{M^{J-I+1}} = \frac{1}{M} \quad \forall s \in \mathbb{S}$$
(2.44)

Hence, the quantiser output symbol probabilities can be the same as the channel input symbol probabilities for an i.i.d. input sequence even if the eye is closed, whereas the same is in general not true if $\{u(k)\}$ is a dependent sequence. In other words, if the channel input is a dependent sequence and the eye is closed, the quantiser output symbol probabilities will be

$$\Pr\{\hat{u}(k) = s\} \neq \frac{1}{M}$$
(2.45)

for some $s \in \mathbb{S}$.

Remark 2.3 In general, it is sufficient to check if $R_{\hat{u}}(0) = R_u(0)$ to determine whether or not $\Pr\{\hat{u}(k) = s\} = 1/M \ \forall s$ is true, since

$$R_{\hat{u}}(0) = E\{\hat{u}^2(k)\}$$
(2.46a)

$$= \sum_{i=1}^{M} s_i^2 \Pr\{\hat{u}(k) = s_i\}$$
(2.46b)

where the s_i are the members of $\mathbb{S}$. Thus, it is possible to test for convergence to an open-eye minimum by testing $R_u(0) = R_{\hat{u}}(0)$ against $R_u(0) \neq R_{\hat{u}}(0)$.

The following example provides an illustration of the above result.

Example 2.2 Consider a dependent 4-ary input sequence (i.e. $u(k) \in \mathbb{S} = \{-3, -1, 1, 3\}$) generated by a Markov chain with equally likely outputs and the transition probability matrix

$$\Pi = \begin{bmatrix} 0.3 & 0.2 & 0.2 & 0.3 \\ 0.2 & 0.4 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0.4 & 0.2 \\ 0.3 & 0.2 & 0.2 & 0.3 \end{bmatrix}.$$
(2.47)

---

[4]We tacitly make use of the assumption that all finite-length subsequences of $\{u(k)\}$ occur with nonzero probability and of the fact that the equaliser has a linear structure.

$u(k)$	1	1	1	1	3	3	3	3
$u(k-1)$	-3	-1	1	3	-3	-1	1	3
$\hat{u}(k)$	-1	1	1	3	1	3	3	3

Table 2.1 Mapping from channel input to quantiser output. Only eight input-output combinations are shown as the rest can be obtained by simple negation.

Dependent input sequence			Independent input sequence		
τ	$R_u(\tau)$	Sample mean of $\widehat{R}_{\hat{u}}(\tau)$	τ	$R_u(\tau)$	Sample mean of $\widehat{R}_{\hat{u}}(\tau)$
0	5.0000	4.6005	0	5.0	4.9920
1	0.1000	1.7693	1	0.0	1.8611
2	0.0200	0.0009	2	0.0	-0.0166
3	0.0040	-0.0148	3	0.0	-0.0117
4	0.0008	-0.0246	4	0.0	-0.0138
5	0.0002	-0.0250	5	0.0	-0.0130

Table 2.2 Channel input autocorrelation and sample mean of quantiser output autocorrelation estimates for dependent and independent input sequences.

Suppose that the quantiser output can be written as

$$\hat{u}(k) = Q_4\Big(u(k) + 0.8u(k-1)\Big) \quad \forall k \tag{2.48}$$

which is a closed-eye setting because the equalisation objective in Eq. (2.4) is not satisfied as can be seen from Table 2.1. Only eight input-output combinations are shown in Table 2.1 as the other eight can be obtained simply by flipping the signs of the first eight combinations.

The channel input autocorrelation and sample mean of the quantiser output autocorrelation are tabulated in Table 2.2 for dependent and independent channel input sequences. The sample size was $N = 5000$ and 50 trials were used to obtain the sample mean values. Take note of the significant difference between $R_u(0)$ and the sample mean of $\widehat{R}_{\hat{u}}(0)$ when $\{u(k)\}$ is a dependent sequence, which confirms the conclusion of Remark 2.3.   □

In [1] an example is given to illustrate the claim that checking the independence of the quantiser output sequence when the channel input sequence is i.i.d. with equiprobable symbols is not sufficient to infer convergence to an open-eye minimum. The example considers a channel-equaliser combination such that $\hat{u}(k) = Q_4(1000u(k))$, implying $\hat{u}(k) = 3\text{sgn}(u(k))$, for which the eye is closed. Although both the channel input and quantiser output sequences are i.i.d., it suffices to consider the quantiser output autocorrelation to conclude that the eye is closed as the quantiser output variance $R_{\hat{u}}(0) = 9$ is not equal to the channel input variance $R_u(0) = 5$. Thus, we remark that this particular example, which was also referred to in [4], does not serve the purpose of illustrating the need for a comprehensive check of the channel input and quantiser output distributions.

2.7 *Estimating a Lower Bound on the Equalisation Delay*

Estimation of the equalisation delay Δ in a blind equalisation setting is a problem of substantive interest with potential applications in performance analysis of multiaccess data networks. As pointed out in [4], to find the exact equalisation delay requires to have direct access to the channel input, the possibility of which is ruled out in blind equalisation. Nevertheless, it is possible to estimate some bound on Δ as we shall see below. The method described here relies upon the crosscorrelation of the noisy channel output sequence $\{r(k)\}$ and the slicer output sequence $\{\hat{u}(k)\}$ (see Fig. 2.1(a))

$$R_{r\hat{u}}(\tau) = E\{r(k)\hat{u}(k-\tau)\} \tag{2.49}$$

where

$$r(k) = \sum_{i=0}^{\infty} h_i u(k-i) + n(k). \tag{2.50}$$

When the eye is open, i.e. Eq. (2.7) holds, we have

$$R_{r\hat{u}}(\tau) = E\left\{\left(\sum_{i=0}^{\infty} h_i u(k-i) + n(k)\right) \operatorname{sgn}(c_\Delta) u(k-\tau-\Delta)\right\} \tag{2.51a}$$

$$= E\left\{\sum_{i=0}^{\infty} h_i u(k-i) \operatorname{sgn}(c_\Delta) u(k-\tau-\Delta)\right\}. \tag{2.51b}$$

Note that $E\{n(k)\operatorname{sgn}(c_\Delta)u(k-\tau-\Delta)\} = 0$ since $\{\hat{u}(k)\}$ and $\{n(k)\}$ are independent sequences with $E\{u(k)\} = 0$. Changing the order of expectation and summation in Eq. (2.51b) yields

$$R_{r\hat{u}}(\tau) = \operatorname{sgn}(c_\Delta) \sum_{i=0}^{\infty} h_i E\{u(k-i)u(k-\tau-\Delta)\} \tag{2.52a}$$

$$= \operatorname{sgn}(c_\Delta) \sum_{i=0}^{\infty} h_i R_u(i-\tau-\Delta). \tag{2.52b}$$

Defining the convolution $\xi(x) = \sum_{i=0}^{\infty} h_i R_u(x-i)$ and using the even symmetry property of $R_u(\tau)$, Eq. (2.52b) can be written as

$$R_{r\hat{u}}(\tau) = \operatorname{sgn}(c_\Delta)\xi(\tau + \Delta). \tag{2.53}$$

Assuming $c_\Delta > 0$ with no loss of generality, Eq. (2.53) takes the following form for different values of τ

$$R_{r\hat{u}}(\tau) = \begin{cases} 0, & \tau < -\Delta - m \\ h_0 R_u(m), & \tau = -\Delta - m \\ \sum_{i=0}^{\Delta+m+\tau} h_i R_u(i - \tau - \Delta), & -\Delta - m < \tau \le -\Delta + m \\ \sum_{i=\Delta-m+\tau}^{\Delta+m+\tau} h_i R_u(i - \tau - \Delta), & \tau > -\Delta + m \end{cases} \tag{2.54}$$

where we used the assumption that $\{u(k)\}$ is an m-dependent sequence, i.e. $R_u(\tau) = 0$ for $|\tau| > m$.

Referring to Eq. (2.54), if we define τ_0 as the minimum value of τ for which $R_{r\hat{u}}(\tau)$ is nonzero, the equalisation delay estimate is obtained as $\widehat{\Delta} = -\tau_0 - m$. Note that whereas $\widehat{\Delta}$ is an exact delay estimate if $h_0 \ne 0$, it becomes an underestimate of Δ by l if $\{h_i\}$ contains l leading zeros. Hence, the procedure described above essentially provides a lower bound on Δ if the restriction $h_0 \ne 0$ is lifted. If $\{u(k)\}$ is an independent sequence, Eq. (2.53) reduces to

$$R_{r\hat{u}}(\tau) = \text{sgn}(c_\Delta) h_{\tau+\Delta}. \tag{2.55}$$

In this case the lower bound on Δ is simply $\widehat{\Delta} = -\tau_0$ since $m = 0$. Eq. (2.54) can also be used to estimate the channel impulse response $\{h_i\}$ when the eye is open.

2.8 Simulation Studies

Two examples are presented to demonstrate the application of the statistical test. The first example compares estimates of the slicer output autocorrelation with the channel input autocorrelation for two different ISI levels corresponding to open-eye and closed-eye conditions. The objective is to verify the test result of Theorem 2.1. In the second example, the performance of the statistical test is studied in terms of the sample size. Consistency of the test is established by an observed monotone increase in the probability of detection with the sample size.

In the examples, the channel input sequence is obtained from the output of a Markov chain with the transition probability matrix

$$\Pi = \begin{bmatrix} 0.65 & 0.35 \\ 0.35 & 0.65 \end{bmatrix}. \tag{2.56}$$

The autocorrelation of the channel input sequence $\{u(k)\}$ is shown in Fig. 2.2. The input can be assumed to be an approximately m-dependent sequence with $m = 5$. The entries of the covariance
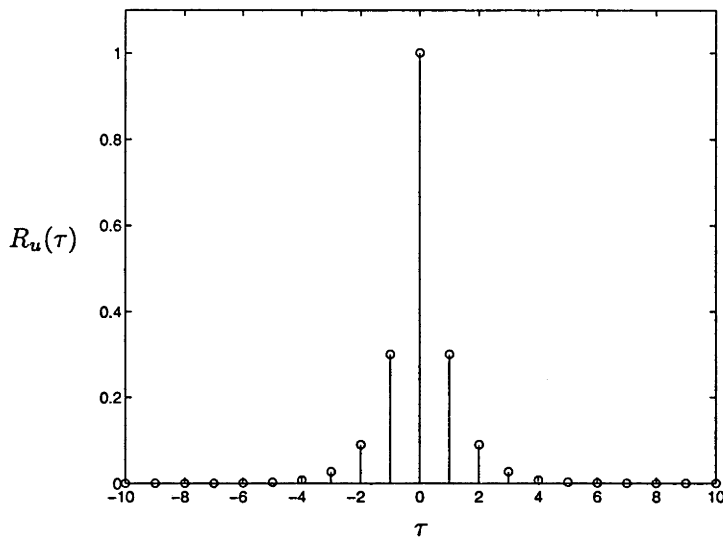
Figure 2.2 Autocorrelation of channel input sequence $\{u(k)\}$.

matrix Σ are given by the fourth-order moments of $\{u(k)\}$. Expressions for the moments of a Markov chain are derived in Appendix 2.B. The covariance matrix Σ is nonsingular. Therefore, the test statistic is of the form $\Lambda = z^T \Sigma^{-1} z$ which is distributed according to χ_K^2 under H_0. The communication channel is a nonminimum phase FIR system

$$H(z) = -0.6 + 1.1z^{-1} - 1.9z^{-2} - 0.8z^{-3}. \qquad (2.57)$$

An LDDE with 3 taps is used (i.e. $L = 3$). The channel-equaliser combination satisfies the condition in Eq. (2.5). The channel noise $\{n(k)\}$ is a zero-mean discrete-time white Gaussian process with variance $\sigma_n^2 = 0.1$ which is passed through a hard limiter to impose a bound on its magnitude in accordance with $\Pr\{|n(k)| > 0.5\} = 0 \ \forall k$. This ensures that $\{\hat{u}(k)\}$ and $\{n(k)\}$ are independent sequences, which is a condition required by Theorem 2.1. As a representative blind equalisation algorithm the constant modulus algorithm (CMA) will be used, although other algorithms for blind adaptation of the equaliser taps may also be considered.

Example 2.3 In this example, the detection of convergence to a closed-eye minimum is considered under the blind adaptation of the equaliser parameters. The scenarios of correct convergence and ill-convergence by CMA are depicted in Fig. 2.3. A stepsize of $\mu = 1.5 \times 10^{-4}$ was used to obtain the CMA parameter trajectories. The difference between the minima converged in Figs. 2.3(a) and (b) is mainly due to the choice of initialisation; in Fig. 2.3(a) the equaliser parameters were initialised to $\theta(0) = [0, 1, 0]^T$ (centre-tap initialisation) and in Fig. 2.3(b) to $\theta(0) = [1, 0, 0]^T$.

(a)

(b)
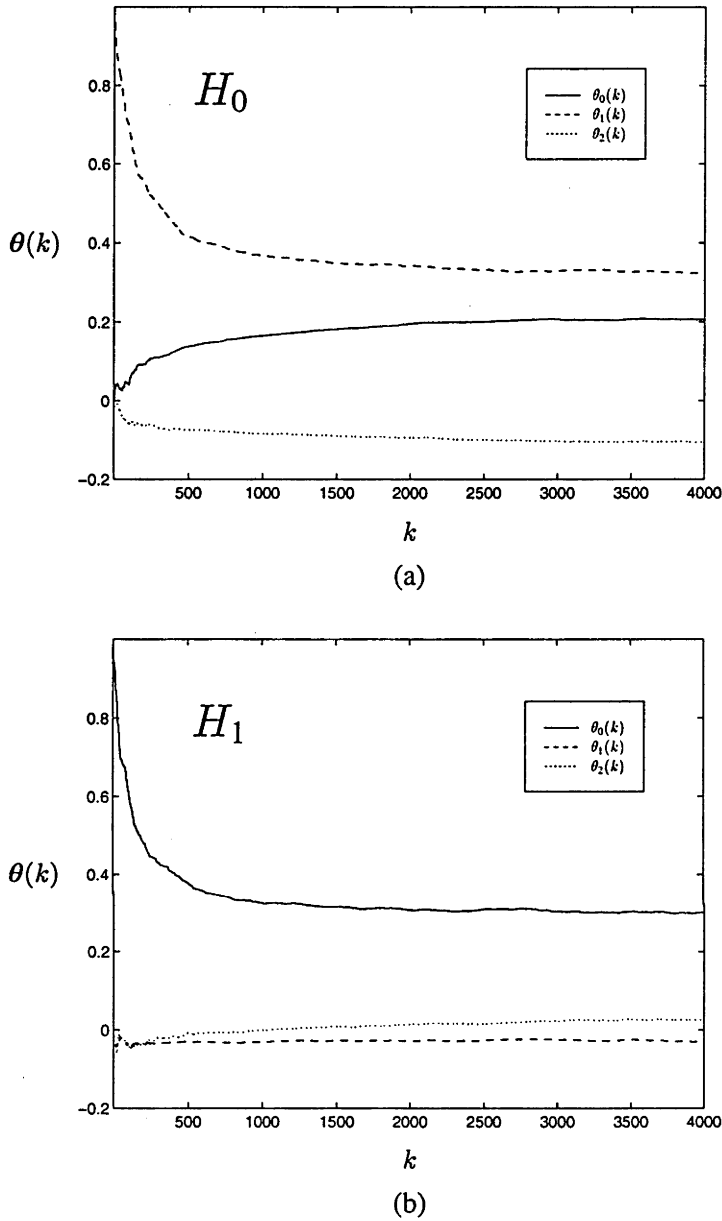
Figure 2.3 CMA adaptation of equaliser parameters resulting in convergence (a) to an open-eye minimum with CLEM = 0.3675, and (b) to a closed-eye local minimum with CLEM = 1.1741.

In Fig. 2.3(a) the equaliser parameters converge to the *open-eye* setting

$$\theta = [0.2084, 0.3281, -0.1045]^T$$

which yields an overall system impulse response with $c_\Delta = -0.9052$ and CLEM=0.3675. Using the definition of CLEM in Eq. (1.11), we obtain the minimum absolute value of $y(k)$ due to $u(k)$ as $\min_k |y^u(k)| = |c_\Delta|(1 - \text{CLEM}) = 0.5725$. The maximum absolute value of $y(k)$ due to noise, on the other hand, is $\max_k |y^n(k)| = \max_k |n(k)| \sum_{i=0}^{2} |\theta_i| = 0.3205$. Since $\min_k |y^u(k)| > \max_k |y^n(k)|$, the bound in Eq. (2.34) is satisfied (i.e. Theorem 2.1 can be used to

(a)

(b)
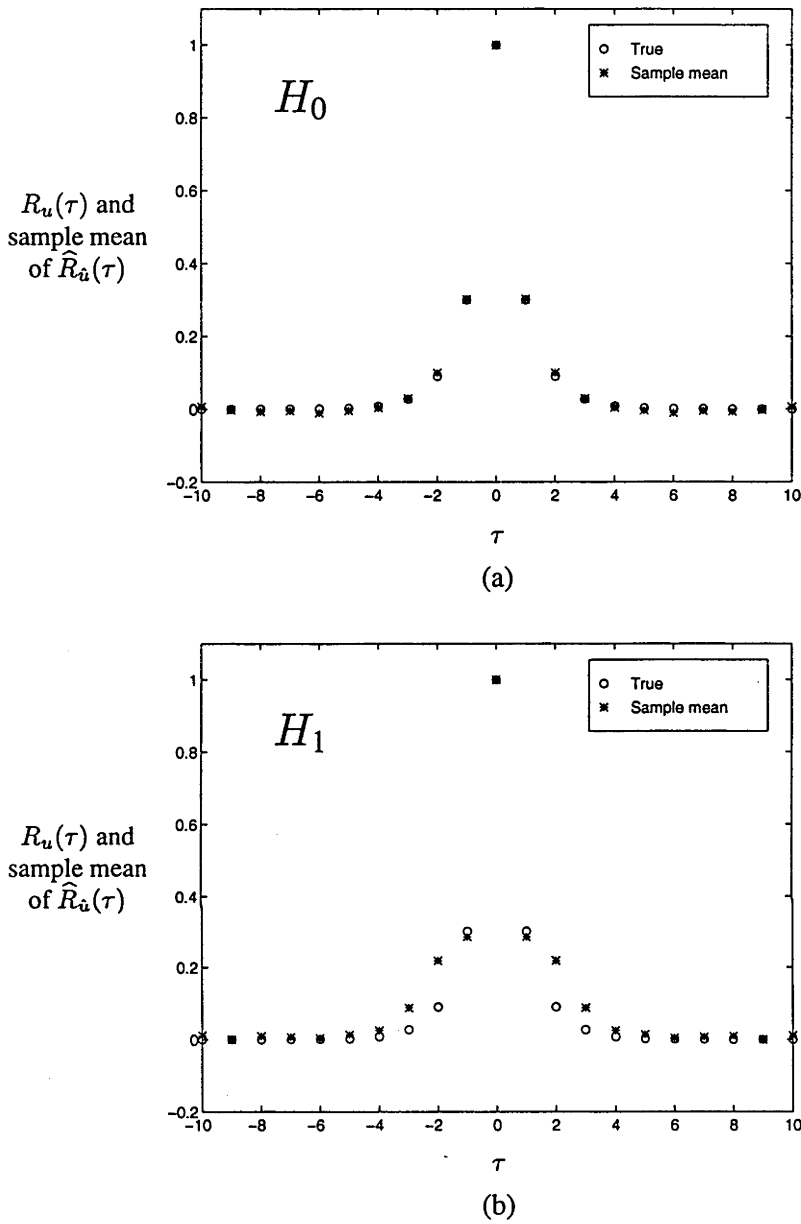
Figure 2.4 Channel input autocorrelation $R_u(\tau)$ and sample mean of slicer output autocorrelation estimates $\widehat{R}_{\hat{u}}(\tau)$ for convergence (a) to an open-eye minimum (H_0), and (b) to a closed-eye minimum (H_1).

test for convergence to an open-eye minimum). In Fig. 2.3(b) the equaliser parameters converge to the *closed-eye* setting

$$\theta = [0.3041, -0.0331, 0.0341]^T$$

for which we get CLEM=1.1741.

The slicer output autocorrelation was estimated for 10 correlation lags ($K = 10$) using Eq. (2.9) for a sample size $N = 500$. The channel input autocorrelation and the sample mean of the slicer output autocorrelation estimates are plotted in Figs. 2.4(a) and (b) after 50 trials for the cases of

	True	Estimated
$E\{\Lambda \mid H_0\}$	10	10.2
$\mathrm{Var}\{\Lambda \mid H_0\}$	20	23.2
P_{FA}	0.25	0.2670
	0.10	0.1090
	0.05	0.0545

Table 2.3 Sample statistics and P_{FA} of Λ under H_0 for $N = 300$.
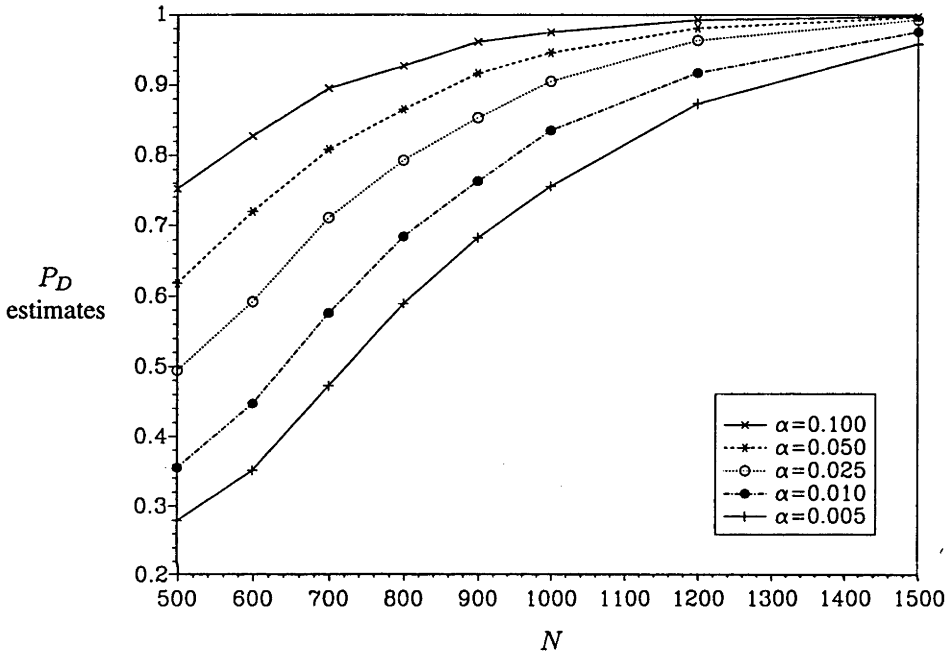
Figure 2.5 Plot of test power estimates *v.* sample size.

correct convergence and ill-convergence, respectively. Notice the significant difference between $R_u(\tau)$ and the sample mean values of $\widehat{R}_{\widehat{u}}(\tau)$ in Fig. 2.4(b), which clearly indicates the occurrence of ill-convergence. The sample mean and variance, and the estimated P_{FA} of the test statistic under the null hypothesis are listed in Table 2.3, alongside their true values, for 2000 trials and $N = 300$. The results confirm the approximate normality of $\widehat{R}_{\widehat{u}}(\tau)$ for large sample sizes. □

Example 2.4 The probability of detection estimates of the test are plotted in Fig. 2.5 as a function of the sample size N for various significance levels when the LDDE has converged to the closed-eye local minimum in Fig. 2.3(b). To estimate the values of P_D, 2000 simulation runs were used. The test statistic was computed using 10 correlation lags (i.e. $K = 10$). As is evident from Fig. 2.5, P_D is a monotonically increasing function of N, thereby confirming the result of Eq. (2.30). In this particular example the sample size N needs to be set to a value higher than 1500 in order to achieve a high test power while maintaining a relatively small probability of false alarm. □

2.9 Discussion and Conclusions

In this chapter, we have obtained a necessary and sufficient condition for the convergence of an LDDE to an open-eye minimum on the cost surface of the adaptation algorithm used when the channel input is a correlated binary sequence. The convergence test criterion is based on the correlation statistics of the slicer output sequence and *a priori* knowledge of the autocorrelation of the channel input sequence, and requires the autocorrelation of the channel input and slicer output sequences to be identical as a necessary and sufficient condition for convergence to an open-eye parameter setting, subject to the proviso that the channel noise does not have any influence on the slicer output. For dependent M-ary channel inputs, a simpler heuristic criterion was shown to be viable, which compares only the quantiser output variance to the channel input variance. For nonlinear equaliser structures such as the decision feedback equaliser, convergence criteria based on the channel input and slicer output correlation statistics have proved to be potentially impracticable, thereby defying the construction of a simple statistical test [4]. In Chapter 3 we will propose a different test criterion which is applicable to linear, as well as, nonlinear equaliser structures.

The statistical test is applicable under the following assumptions:

- The channel input sequence is binary, m-dependent and fourth-order weakly stationary.

- The autocorrelation and fourth-order moments of the channel input sequence are known.

- The channel is linear.

- The channel-equaliser combination has an impulse response that complies with Eq. (2.5).

- The equaliser has a tapped-delay line structure followed by a slicer.

If the tapped-delay line $\Theta(z)$ is not long enough to provide perfect equalisation (i.e. $y(k) \neq \pm 1$ for some k) as was the case in Example 2.3, the eye pattern test may become unreliable. The statistical test, however, ignores $y(k)$ and considers $\hat{u}(k)$ only, which provides some degree of robustness in cases where the equaliser is not capable of removing ISI completely. The role of the visual eye pattern test can therefore be replaced by the proposed convergence test in practice. The performance of the test critically depends *inter alia* on a distance measure between the channel input and slicer output autocorrelations. If the length of observations (the sample size) is sufficiently large, an almost unity test power may be achieved (subject to the condition of Theorem 2.4). The minimum sample size that would result in a near unity test power is again determined by how much $R_u(\tau)$ and $R_{\hat{u}}(\tau)$ would differ when the eye is closed, which strongly

depends on the resulting $\{c_i\}$ and also on CLEM to some degree. For certain correlation lags, it is possible to write the relationship between $R_u(\tau)$ and $R_{\hat{u}}(\tau)$ (see Eqs. (2.64) and (2.70)).

As can be gleaned from the definitions of H_0 and H_1 in Section 2.4.1, the convergence test of Theorem 2.1 does not provide a means to distinguish between different open-eye minima on the average cost surface. In other words, if a particular local minimum opens the eye with CLEM < 1, whilst there are other minima with smaller CLEM, the convergence test will not be able to identify the existence of the other minima with "better" equalisation performance. However, once the eye is open, it is possible to estimate $\{c_i\}$ from the equaliser response. The resulting CLEM can then be used to quantify the amount of residual ISI still present in $y(k)$. If there are multiple open-eye minima, the minimum that results in the smallest CLEM and the smallest ℓ_2 norm for the equaliser parameters (i.e. $\sqrt{\sum_{i=0}^{L-1} |\theta_i|^2}$) will obviously yield the most robust equalisation in the face of channel noise.

The delay estimation in a blind equalisation setting is often a difficult problem that can be solved only under certain assumptions such as a causal channel and no leading zeros in the channel impulse response. The success of the method discussed in Section 2.7 by and large depends on the magnitude of $h_0 R_u(w)$ (see Eq. (2.54)) from a practical point of view. Clearly, if $h_0 R_u(w)$ is too small, the chances of its being detected will be significantly diminished due to crosscorrelation estimation errors. A better method for estimating the equalisation delay will be proposed in the next chapter.

2.A Proof of Theorem 2.1

The autocorrelation of the slicer output sequence is defined by

$$R_{\hat{u}}(\tau) = E\{\hat{u}(k+\tau)\hat{u}(k)\}. \tag{2.58}$$

Substituting Eq. (2.7) in Eq. (2.58) readily proves the "if" part. The proof of the "only if" part is, however, not as trivial. Firstly, we establish the implication of identical channel input and slicer output autocorrelations on the channel input and slicer output conditional probabilities. The input sequence to the slicer $y(k)$ has zero mean and is symmetrically distributed about zero, since $E\{u(k)\} = 0$ and $C(z)$ is a linear system, which implies that the binary symbols at the slicer output are equally likely with $\Pr\{\hat{u}(k) = 1\} = 1/2 \ \forall k$. Note that $\{n(k)\}$ is ignored since, by assumption, it has no effect on $\{\hat{u}(k)\}$. We can express Eq. (2.58) as

$$R_{\hat{u}}(\tau) = \Pr\{\hat{u}(k+\tau)\hat{u}(k) = 1\} - \Pr\{\hat{u}(k+\tau)\hat{u}(k) = -1\} \tag{2.59a}$$

$$= 2 \Pr\{\hat{u}(k + \tau)\hat{u}(k) = 1\} - 1 \tag{2.59b}$$

where the probability of the product of two slicer outputs can be written as

$$\Pr\{\hat{u}(k+\tau)\hat{u}(k) = 1\} = \Pr\{\hat{u}(k+\tau) = 1, \hat{u}(k) = 1\} + \Pr\{\hat{u}(k+\tau) = -1, \hat{u}(k) = -1\}. \tag{2.60}$$

Writing the slicer output symbol probabilities in terms of their joint probabilities and noting that $\Pr\{\hat{u}(k) = 1\} = \Pr\{\hat{u}(k + \tau) = -1\} = 1/2$ which follows from equiprobable slicer output symbols, we have

$$\Pr\{\hat{u}(k) = 1, \hat{u}(k + \tau) = 1\} + \Pr\{\hat{u}(k) = 1, \hat{u}(k + \tau) = -1\} =$$
$$\Pr\{\hat{u}(k + \tau) = -1, \hat{u}(k) = 1\} + \Pr\{\hat{u}(k + \tau) = -1, \hat{u}(k) = -1\} \tag{2.61}$$

whence we obtain

$$\Pr\{\hat{u}(k + \tau) = 1, \hat{u}(k) = 1\} = \Pr\{\hat{u}(k + \tau) = -1, \hat{u}(k) = -1\}. \tag{2.62}$$

Thus, Eq. (2.60) can be rewritten as

$$\Pr\{\hat{u}(k + \tau)\hat{u}(k) = 1\} = 2 \Pr\{\hat{u}(k + \tau) = 1, \hat{u}(k) = 1\} \tag{2.63a}$$
$$= \Pr\{\hat{u}(k + \tau) = 1 \mid \hat{u}(k) = 1\}. \tag{2.63b}$$

Substituting Eq. (2.63b) in Eq. (2.59b) yields

$$R_{\hat{u}}(\tau) = 2 \Pr\{\hat{u}(k + \tau) = 1 \mid \hat{u}(k) = 1\} - 1. \tag{2.64}$$

Similarly, the autocorrelation of the channel input sequence can be written as

$$R_{u}(\tau) = 2 \Pr\{u(k + \tau) = 1 \mid u(k) = 1\} - 1. \tag{2.65}$$

A comparison of Eqs. (2.64) and (2.65) results at once in the following relation

$$R_{\hat{u}}(\tau) = R_{u}(\tau) \Leftrightarrow \Pr\{\hat{u}(k+\tau) = 1 \mid \hat{u}(k) = 1\} = \Pr\{u(k+\tau) = 1 \mid u(k) = 1\} \quad \forall \tau, k. \tag{2.66}$$

Let us define the integers I and J in Eq. (2.5) in such a way that the following holds:

$$\hat{u}(k) \neq \text{sgn}\left(\sum_{i=I}^{J-1} c_i u(k-i) \right) \quad \text{for some } \{u(k)\} \tag{2.67a}$$

$$\hat{u}(k) \neq \text{sgn}\left(\sum_{i=I+1}^{J} c_i u(k-i) \right) \quad \text{for some } \{u(k)\}. \tag{2.67b}$$

Thus, $\{c_I, c_{I+1}, \ldots, c_J\}$ is a *minimum-length* subsequence of the impulse response of $C(z)$ that is sufficient to know in order to determine the sign of $y(k)$ for any input sequence $\{u(k)\}$. In the remainder of the proof we show that the inequality $I < J$, which amounts to the convergence of the equaliser parameters to a closed-eye local minimum, violates Eq. (2.66) for a particular τ.

Consider the channel inputs $u(k)$ and $u(k+J-I+m)$ separated by $J-I+m$ in time. If $I < J$, $u(k)$ and $u(k+J-I+m)$ would be independent since $J-I+m > m$ and $\{u(k)\}$ is an m-dependent sequence. In what follows, we will prove that while $u(k)$ and $u(k+J-I+m)$ are independent if $I < J$, the same is not true for $\hat{u}(k)$ and $\hat{u}(k+J-I+m)$. Let us denote the time separation using a new variable $t = J-I+m$. Consider the following conditional probability

$$\Pr\{\hat{u}(k+t)=1, \hat{u}(k)=1 \mid u(k-I+m)=s_1, u(k-I)=s_2\}$$

$$= \Pr\{\hat{u}(k+t)=1 \mid \hat{u}(k)=1, u(k-I+m)=s_1, u(k-I)=s_2\}$$

$$\times \Pr\{\hat{u}(k)=1 \mid u(k-I+m)=s_1, u(k-I)=s_2\} \tag{2.68a}$$

$$= \Pr\{\hat{u}(k+t)=1 \mid u(k-I+m)=s_1\}\Pr\{\hat{u}(k)=1 \mid u(k-I)=s_2\} \tag{2.68b}$$

where $s_1, s_2 \in \mathbb{S} = \{-1, 1\}$. Simplifications in the conditional probabilities above result from a consideration of the dependence of $\hat{u}(k)$ on $u(k)$ as specified in Eq. (2.6), as well as m-dependence of $\{u(k)\}$. An expression for the joint probability of $\hat{u}(k+t)$ and $\hat{u}(k)$ can be obtained by eliminating the conditioning on the left-hand side of Eq. (2.68b) as follows

$$\Pr\{\hat{u}(k+t)=1, \hat{u}(k)=1\}$$

$$= \frac{1}{2} \sum_{s_1 \in \mathbb{S}} \sum_{s_2 \in \mathbb{S}} \Pr\{\hat{u}(k+t)=1, \hat{u}(k)=1 \mid u(k-I+m)=s_1, u(k-I)=s_2\}$$

$$\times \Pr\{u(k-I+m)=s_1 \mid u(k-I)=s_2\} \tag{2.69a}$$

$$= \frac{1}{2} \sum_{s_1 \in \mathbb{S}} \sum_{s_2 \in \mathbb{S}} \zeta_{1|s_1}(J)\epsilon_{1|s_2}(I)p_{s_1|s_2}(m) \tag{2.69b}$$

where $\zeta_{s_1|s_2}(J) \triangleq \Pr\{\hat{u}(k)=s_1 \mid u(k-J)=s_2\}$, $\epsilon_{s_1|s_2}(I) \triangleq \Pr\{\hat{u}(k)=s_1 \mid u(k-I)=s_2\}$ and $p_{s_1|s_2}(m) \triangleq \Pr\{u(k+m)=s_1 \mid u(k)=s_2\}$. Then, the conditional probability of $\hat{u}(k+t)$

given $\hat{u}(k)$ can be written as

$$\Pr\{\hat{u}(k+t) = 1 \mid \hat{u}(k) = 1\} = \sum_{s_1 \in \mathbb{S}} \sum_{s_2 \in \mathbb{S}} \zeta_{1|s_1}(J) \epsilon_{1|s_2}(I) p_{s_1|s_2}(m). \qquad (2.70)$$

Using the relations

$$p_{1|1}(m) = p_{-1|-1}(m) = \frac{1}{2}(R_u(m) + 1) \qquad (2.71\text{a})$$

$$p_{1|-1}(m) = p_{-1|1}(m) = 1 - \frac{1}{2}(R_u(m) + 1) \qquad (2.71\text{b})$$

$$\zeta_{1|-1}(J) = 1 - \zeta_{1|1}(J) \qquad (2.71\text{c})$$

$$\epsilon_{1|-1}(I) = 1 - \epsilon_{1|1}(I) \qquad (2.71\text{d})$$

Eq. (2.70) can be simplified as follows

$$\Pr\{\hat{u}(k+t) = 1 \mid \hat{u}(k) = 1\} = \frac{1}{2} + 2R_u(m)\left(\zeta_{1|1}(J) - \frac{1}{2}\right)\left(\epsilon_{1|1}(I) - \frac{1}{2}\right) \qquad (2.72\text{a})$$

$$\neq \frac{1}{2}. \qquad (2.72\text{b})$$

In other words, if $I < J$, then $R_u(t) \neq R_{\hat{u}}(t)$ because $\hat{u}(k)$ and $\hat{u}(k+t)$ are not independent whilst $u(k)$ and $u(k+t)$ are. Neither $\zeta_{1|1}(J)$ nor $\epsilon_{1|1}(I)$ can be equal to $1/2$ since $\hat{u}(k)$ depends on both $u(k-I)$ and $u(k-J)$ as can be seen from Eqs. (2.67a) and (2.67b). This completes the proof of the "only if" part. ∎

2.B   Moments of Markov Chains

We consider the second and fourth-order statistics of a homogeneous discrete-time Markov chain generating the channel input sequence $\{u(k)\}$. States of the Markov chain will be denoted by the entries of the vector $m = [m_1, m_2, \ldots, m_M]^T$ where $m_i \in \mathbb{S}, i = 1, 2, \ldots, M$. In other words, the output and the state of the Markov chain are identical. The transition probability matrix of a Markov chain is a *stochastic matrix* which has nonnegative entries with the property that the sum of its columns results in a vector of ones. The one-step transition probability matrix is given by

$$\Pi = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,M} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,M} \\ \vdots & \vdots & & \vdots \\ p_{M,1} & p_{M,2} & \cdots & p_{M,M} \end{bmatrix} \qquad (2.73)$$

where $p_{i,j}$ is the conditional probability $p_{i,j} = \Pr\{u(k) = m_j \mid u(k-1) = m_i\}$ for $i,j = 1,2,\ldots,M$. Since $\{u(k)\}$ is homogeneous, $p_{i,j}$ is independent of k. The vector of initial state probabilities is defined by

$$p_0 = [\Pr\{u(0) = m_1\}, \Pr\{u(0) = m_2\}, \ldots, \Pr\{u(0) = m_M\}]^T. \tag{2.74}$$

We use the following result to obtain the steady-state output probabilities.

Theorem 2.5 (Perron) [28] *If Π is an $M \times M$ matrix with strictly positive entries and unity spectral radius, then we have*

$$\Pi_a \triangleq \lim_{p \to \infty} \Pi^p = L \tag{2.75}$$

where $L = xy^T$, $\Pi x = x$, $\Pi^T y = y$, x and y have positive entries, and $x^T y = 1$.

The spectral radius of a stochastic matrix is always equal to one [28]. We therefore conclude that if Π has strictly positive entries, the Markov chain is asymptotically stationary with steady-state (asymptotic) state probabilities

$$\begin{bmatrix} \Pr\{u(k) = m_1\} \\ \Pr\{u(k) = m_2\} \\ \vdots \\ \Pr\{u(k) = m_M\} \end{bmatrix} = \Pi_a^T p_0 \tag{2.76}$$

whence the mean of the Markov chain output readily follows

$$E\{u(k)\} = \sum_{i=1}^{M} m_i \Pr\{u(k) = m_i\} \tag{2.77a}$$

$$= m^T \Pi_a^T p_0. \tag{2.77b}$$

The second-order moments of the Markov chain output are given by

$$E\{u(k)u(k+\tau)\} = \sum_{i=1}^{M}\sum_{j=1}^{M} m_i m_j \Pr\{u(k) = m_i, u(k+\tau) = m_j\} \tag{2.78a}$$

$$= \sum_{i=1}^{M}\sum_{j=1}^{M} m_i m_j \Pr\{u(k+\tau) = m_j \mid u(k) = m_i\}$$
$$\times \Pr\{u(k) = m_i\} \tag{2.78b}$$

$$= \sum_{i=1}^{M}\sum_{j=1}^{M} m_i m_j \left(\Pi_a^\tau\right)_{i,j} \left(\Pi_a^T p_0\right)_i \tag{2.78c}$$

where $(\cdot)_{i,j}$ denotes the (i,j)th entry and $(\cdot)_i$ the ith entry. The autocorrelation of $\{u(k)\}$ is then

obtained by subtracting the square of the mean from the second-order moments, i.e.

$$R_u(\tau) = E\{u(k)u(k+\tau)\} - E^2\{u(k)\} \tag{2.79a}$$

$$= \sum_{i=1}^{M}\sum_{j=1}^{M} m_i m_j \left(\Pi_a^\tau\right)_{i,j} \left(\Pi_a^T p_0\right)_i - \left(m^T \Pi_a^T p_0\right)^2. \tag{2.79b}$$

The fourth-order moments are computed using the expression

$$m_{4,u}(\tau_1, \tau_2, \tau_3)$$

$$= E\{u(k)u(k+\tau_1)u(k+\tau_2)u(k+\tau_3)\} \tag{2.80a}$$

$$= \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{p=1}^{M}\sum_{q=1}^{M} m_i m_j m_p m_q$$

$$\times \Pr\{u(k+\tau_3) = m_i, u(k+\tau_2) = m_j, u(k+\tau_1) = m_p, u(k) = m_q\} \tag{2.80b}$$

$$= \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{p=1}^{M}\sum_{q=1}^{M} m_i m_j m_p m_q$$

$$\times \left(\Pi_a^{(\tau_3-\tau_2)}\right)_{j,i} \left(\Pi_a^{(\tau_2-\tau_1)}\right)_{p,j} \left(\Pi_a^{\tau_1}\right)_{q,p} \left(\Pi_a^T p_0\right)_q \tag{2.80c}$$

where $0 \leq \tau_1 \leq \tau_2 \leq \tau_3$. To simplify the conditional probabilities above we used the Markov property that $\Pr\{u(k+l) = a \mid u(k+m) = b, u(k+n) = c\} = \Pr\{u(k+l) = a \mid u(k+m) = b\}$ if $l \geq m \geq n$.

CHAPTER 3

Blind Detection of Equalisation Errors

3.1 Introduction

In adaptive channel equalisation, the decisions at the equaliser output can be in error due to a number of reasons. Some of the more commonly encountered causes of erroneous equaliser output decisions or equalisation errors are ill-convergence, excessive channel noise and error propagation. The detection of equalisation errors is a particularly acute problem in blind adaptive channel equalisation because of the lack of direct access to the channel input at any time. The ill-convergence of on-line blind equalisation algorithms compounds the problem of detecting equalisation errors even further. Even in the more familiar case of conventional channel equalisation, where the equaliser parameters are adapted with the aid of a training sequence, there is no straightforward means to detect the occurrence of equalisation errors once the training session is over. It is thus desirable to be able to detect the presence of equalisation errors in a blindfolded manner without relying on an explicit knowledge of the channel input. In this chapter we will cast the problem of blind equalisation error detection into the statistical hypothesis testing framework as we did in the previous chapter. To enhance the applicability of the resulting tests, no assumption will be made about the channel input constellation, the equaliser structure and the channel noise magnitude (apart from Gaussianity). In other words, we will not restrict ourselves to testing for the convergence of an equaliser to an idealised open-eye parameter setting as we did in the previous chapter and reference [6]. One potential application of blind error tests is in bandwidth-efficient retransmission protocols in data networks.

The material in the chapter is motivated, to a larger extent, by certain disadvantages of the

convergence test discussed in Chapter 2. The convergence test of Chapter 2 is applicable only to linear decision-directed equalisers (LDDEs) for communication channels driven by binary input sequences, and requires *a priori* knowledge of second and fourth-order statistics of the channel input sequence. It also requires long observations to guarantee sure detection of equalisation errors, although the particular sample size required is very much case-dependent. On the credit side, however, the computational requirements of the convergence test are relatively low.

The test criterion proposed in this chapter is based on a relationship between the presence of errors in the decision device output sequence and time variations in the overall input-output relationship of the equaliser-decision device combination when viewed as a linear model. Unlike the convergence test of Chapter 2, no *a priori* knowledge of the channel input statistics is required. The method of least squares (LS) is invoked to construct statistical tests that use only the decision device output and equaliser input observations. To demonstrate the application of the statistical tests, two popular equaliser structures are considered, namely, the LDDE, which is a linear equaliser followed by a nonlinear decision device, and the decision feedback equaliser (DFE). It should be born in mind, however, that the consideration of these equalisers does not imply that the tests are restricted to these equaliser structures only.

The chapter is organised as follows. Section 3.2 gives a formal description of the detection problem and states the assumptions that are made throughout the chapter. Section 3.3 develops the test criterion for detecting equalisation errors. In Section 3.4 the test criterion is implemented as a statistical threshold test. Optimality properties of the test are presented and the influence of the test parameters on the detection performance studied. Alternative tests based on the same test criterion are also outlined. The case of no *a priori* knowledge of the test parameters and its ensuing effects on the test statistic are considered in Section 3.5. The test is shown to have a certain degree of "robustness" to the choice of the test parameters. A method for estimating a lower bound on the equalisation delay is proposed when the decision device output contains no errors. Section 3.6 presents simulation studies for M-ary LDDE and binary DFE, and compares the detection performance of the test used in the simulation examples to that of the convergence test in Chapter 2 by way of an example. The chapter concludes with a discussion of the results. The computational aspects of g-inverses are discussed in Appendix 3.A. In particular, an iterative method is proposed based on [29] for g-inverse computations, resulting in an appreciable reduction in the computational complexity of the test statistics.
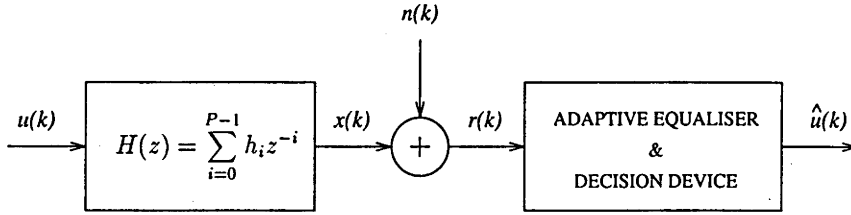
Figure 3.1 Generic channel-equaliser set-up.

3.2 Problem Formulation and Assumptions

Consider the generic channel-equaliser set-up shown in Fig. 3.1, where the channel $H(z)$ is followed by an adaptive equaliser and a decision device for the purpose of recovering the channel input sequence $\{u(k)\}$ from the noisy channel output observations $\{r(k)\}$. The channel input sequence $\{u(k)\}$ is a (not necessarily zero-mean) sequence of discrete-valued symbols drawn from a constellation $\mathbb{S}$ with finite number of members M. In the interest of keeping the exposition simple and focused, the channel input sequence will be assumed to be drawn from a pulse amplitude modulation (PAM) constellation[1] given by $\mathbb{S} = \{-(M-1), -(M-3), \ldots, -3, -1, 1, 3, \ldots, M-3, M-1\}$ with $M > 0$ an even integer. All finite subsequences of $\{u(k)\}$ are supposed to occur with nonzero probability, which subsumes the case of i.i.d. channel inputs. While the channel is assumed to be a linear and slowly time-varying system, it will be assumed to be strictly time-invariant during the test interval. The transfer function of the channel is, or can be approximated by, a causal finite-duration impulse response (FIR) system with impulse response

$$h = [h_0, h_1, \ldots, h_{P-1}]^T$$

where P is the length of the impulse response sequence and the subscript i in the h_i denotes the discrete time index. The channel noise $n(k)$ is a stationary Gaussian random process with zero mean and *known* positive definite covariance matrix Σ which is symmetric and Toeplitz.

The tests will be applied to the blind equaliser structures shown in Fig. 3.2. The LDDE is made up of an adaptive FIR filter $\Theta(z)$ with impulse response of length L

$$\theta = [\theta_0, \theta_1, \ldots, \theta_{L-1}]^T$$

followed by a memoryless *discontinuous* nonlinear device $Q_M(\cdot)$, whose purpose is to map the

---

[1]This assumption does not place any restriction on the generality of the results. It is indeed possible to extend the discussion to input sequences drawn from, say, a quadrature amplitude modulation (QAM) constellation. The only condition required is that the constellation set have a finite number of members (i.e. $M < \infty$).

(a)

(b)
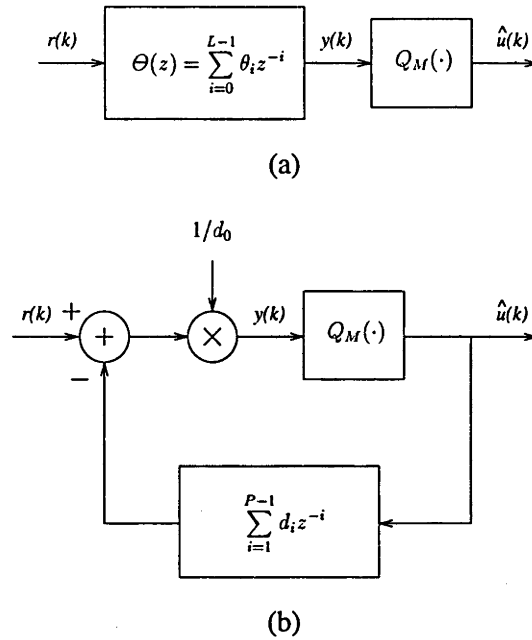
Figure 3.2 (a) Linear decision-directed equaliser (LDDE), (b) decision feedback equaliser (DFE).

continuous-valued $y(k)$ to the symbol constellation $\mathbb{S}$ using the nearest neighbour rule. The name decision-directed equaliser refers to the use of the difference between $\hat{u}(k)$ and $y(k)$ as a pseudo error in the adaptation of θ. Recall from Chapter 2 that for PAM communication systems the nonlinear decision device is defined by

$$Q_M(x) \triangleq \sum_{i=1-M/2}^{M/2-1} \mathrm{sgn}(x+2i). \tag{3.1}$$

The DFE, on the other hand, is a nonlinear recursive filter with a decision device in the forward path and an adaptive FIR filter in the feedback loop which has the transfer function $D(z) - d_0 = \sum_{i=1}^{P-1} d_i z^{-i}$ where $D(z)$ is the z-transform of the equaliser parameters

$$d = [d_0, d_1, \ldots, d_{P-1}]^T.$$

Ideally, the vector d would have the same length as h. In some implementations of the DFE, a prefilter is inserted between $r(k)$ and the adder (see Fig. 3.2(b)) for channels with slowly rising precursor.

Our objective in this chapter is to find out whether or not the following equation[2] is satisfied

$$\hat{u}(k) = \Gamma u(k - \Delta) \quad \forall k \in \mathcal{O} \tag{3.2}$$

---

[2]This equation is a finite-time version of the *equalisation objective* described in Section 1.2.

where Γ is a constant gain[3] that equals either 1 or -1, Δ a constant nonnegative integer, and $\mathcal{O}$ some finite observation interval. Recall that Δ is referred to as the equalisation delay. It appears that in order to be able to make a decision about the validity of Eq. (3.2) one needs to have access to the channel input and decision device output sequences, as well as a knowledge of Δ. Even in conventional equalisation schemes with intermittent access to the channel input in the form of a training sequence, a straightforward verification of Eq. (3.2) may not be possible at a given time instant if the training session is not in progress. Moreover, in the case of blind adaptive equalisation access to the channel input is completely barred, making the direct measurements of Δ and $\{u(k)\}$ impossible. Hence, it would be desirable to use only the observations available at the equaliser so as to test whether or not Eq. (3.2) holds.

From a parameter convergence point of view, the channel noise can play a restrictive role as far as the region of open-eye parameters is concerned. For instance, a given parameter setting may open the eye under zero channel noise and yet it may lead to an eye closure in the presence of channel noise. Likewise, in the particular case of DFE past decision errors may upset the equalisation process, thereby leading to the violation of Eq. (3.2) even under zero forcing tuning (i.e. $d = h$). In the final analysis, Eq. (3.2) plays a central role in the detection of equalisation errors because it takes into account all internal and external sources of error without discrimination. In the rest of the chapter, we will show how Eq. (3.2) can be tested by casting the problem into the hypothesis testing framework.

3.3   Test Criterion for Error Detection

As we have seen in the previous section, verification of the equalisation objective in Eq. (3.2) presents a nontrivial problem especially if direct access to the channel input is not available, which brings on the need for an alternative criterion that is commensurate with Eq. (3.2) while only making use of the information available at the equaliser. In this section we will show that such a criterion is conceivable thanks to the distinct behaviour of the equaliser-decision device combination under open-eye and closed-eye situations. In order to characterise this distinct behaviour, which can be attributed to the presence of $Q_M(\cdot)$ at the equaliser output, we shall consider the backward response of the channel-equaliser-and-decision device combination, taking into account the channel noise, as well. The backward response will ultimately enable us to utilise only the sequences available at the equaliser input onwards to detect the presence of equalisation errors. The impulse response of the backward response will be denoted by $g(k, \Omega)$ which is

---

[3]For two-dimensional (complex) constellations the constant Γ takes the form $e^{j\phi}$ where ϕ is a constant phase.
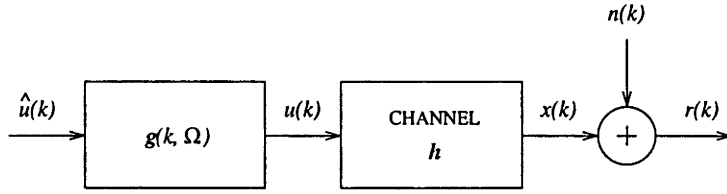
Figure 3.3 Time domain model of the system that takes the decision device output to the noisy channel output. Note that $g(k, \Omega)$ represents a linear time-varying system.

in general time-varying, noncausal and nonunique. The response from the nonlinear decision device output $\hat{u}(k)$ to the noisy channel output $r(k)$ has a time domain representation shown in Fig. 3.3. Since h is time-invariant by assumption, the only part of the model that can be affected by equalisation errors is $g(k, \Omega)$. In other words, $g(k, \Omega)$ carries all the information pertinent to the equaliser performance. The response of the channel-equaliser-and-decision device combination is in fact a nonlinear function of $u(k)$ for arbitrary channel and equaliser characteristics owing to the presence of $Q_M(\cdot)$ implicit in $g(k, \Omega)$. To see why we have chosen to represent the overall response from $\hat{u}(k)$ to $u(k)$ by a linear time-varying (LTV) system with impulse response $g(k, \Omega)$ rather than a nonlinear system, consider, for instance, the channel equalisation problem where the channel input sequence is binary (i.e. $M = 2$) and an LDDE is used with $Q_2(x) = \text{sgn}(x)$. Then, the decision device output can be written as

$$\hat{u}(k) = \text{sgn}\left(\sum_{i=0}^{L+P-2} \left(\sum_{j=0}^{P-1} h_j \theta_{i-j} \right) u(k-i) + \sum_{i=0}^{L-1} \theta_i n(k-i) \right). \tag{3.3}$$

Since $\text{sgn}(\cdot)$ is not invertible, a closed-form expression for a function from $\hat{u}(k)$ to $u(k)$ cannot be obtained for arbitrary equaliser parameter settings, the only known exception being an open-eye case. It is therefore convenient to think of the channel-equaliser-and-decision device combination as a noncausal LTV system with its input related to its output by

$$u(k) = \sum_{i \in \Omega} g_i(k) \hat{u}(k-i) \tag{3.4}$$

where $\Omega = \{i_0, i_0 + 1, \ldots, i_f\}$ is a support set of the impulse response sequence

$$g(k, \Omega) \triangleq [g_{i_0}(k), g_{i_0+1}(k), \ldots, g_{i_f}(k)]^T. \tag{3.5}$$

Note that this representation is sufficiently general but by no means unique, as it would be possible to find infinitely many $g(k, \Omega)$ satisfying Eq. (3.4) for a given subsequence $\{\hat{u}(k - i_0), \ldots, \hat{u}(k - i_f)\}$.

It follows from Eq. (3.2), however, that if the eye is open, $g(k, \Omega)$ can be reduced to a simple time advance operator with $g_i(k) = \Gamma \delta(i + \Delta) \; \forall k$ and $\Omega = \{-\Delta\}$, or

$$g(k, \Omega) = \Gamma \delta(k + \Delta) \tag{3.6}$$

where $\delta(\cdot)$ is the Kronecker delta function defined as $\delta(x) = 1$ if $x = 0$ and $\delta(x) = 0$ if $x \neq 0$ for any integer x. Thus, using Eqs. (3.4) and (3.6) we deduce that if the equalisation objective is satisfied, the system taking $\hat{u}(k)$ to $x(k)$ is *linear time-invariant* (LTI) and noncausal if $\Delta > 0$, and the relationship between the equaliser input and the decision device output is governed by

$$r(k) = \Gamma \sum_{i=0}^{P-1} h_i \hat{u}(k + \Delta - i) + n(k). \tag{3.7}$$

Next we prove the link between the time-variance of $g(k, \Omega)$ and the presence of equalisation errors.

Property 3.1 *The impulse response* $g(k, \Omega)$ *is time-invariant if and only if the eye is open (i.e. Eq. (3.2) is satisfied), provided that all finite-length subsequences of a sufficiently long channel input sequence occur with nonzero probability.*

Proof. The proof of the "if" part readily follows from Eq. (3.6). We prove the "only if" part by contradiction. Suppose that $g(k, \Omega)$ is time-invariant and has the transfer function $G(z) = \sum_{i \in \Omega} g_i z^{-i}$. Then, Eq. (3.4) becomes

$$u(k) = \sum_{i \in \Omega} g_i \hat{u}(k - i). \tag{3.8}$$

In a similar fashion the decision device output can be written as

$$\hat{u}(k) = \sum_i c_i u(k - i) \tag{3.9}$$

where $\{c_i\}$ is the impulse response of $G^{-1}(z)$, which is also time-invariant. Since no finite-length subsequences of $\{u(k)\}$ are prohibited from occurring, the decision device output $\hat{u}(k)$ can take on at least $(M - 1)l + 1$ and at most M^l distinct values where l is the number of nonzero c_i in Eq. (3.9). While the upper bound on the distinct values is obvious, the lower bound[4] is attained when the absolute values of the nonzero c_i are all equal. Thus, if $g(k, \Omega)$ is time-invariant as in Eq. (3.8), $l > 1$ contradicts the fact that both $\{u(k)\}$ and $\{\hat{u}(k)\}$ are M-ary sequences taking on M possible

---

[4]In general, if $l > 1$, the minimum number of distinct values will be larger than M irrespective of the symbol constellation.

values. In other words, the time-invariance of $g(k, \Omega)$ implies $l = 1$ with $\hat{u}(k) = c_\Delta u(k - \Delta)$ or $g_i = \delta(i + \Delta)/c_\Delta$ where $c_\Delta = \Gamma$. Hence, a comparison with Eq. (3.6) reveals that if $g(k, \Omega)$ is time-invariant, the eye is open. ∎

Remarks 3.1

i. If there exist no LTI difference equations in the form of Eq. (3.7) to represent the input-output relation of the equaliser-decision device combination, the eye is necessarily closed or the decision device output contains errors. Conversely, the equaliser-decision device combination has a nonlinear or LTV input-output relationship if the eye is closed.

ii. Since $g(k, \Omega)$ is not unique, we can represent it as

$$g_i(k) = \frac{\delta(i + \Delta)}{\gamma(k)} \qquad (3.10)$$

which implies $\Omega = \{-\Delta\}$ in Eq. (3.5). According to Eqs. (3.4) and (3.6), the form of $\gamma(k)$ is given by

$$\text{The eye is open} \Longleftrightarrow \gamma(k) = 1/\Gamma$$

$$\text{The eye is closed} \Longleftrightarrow \gamma(k) = \hat{u}(k + \Delta)/u(k)$$

Thus, in contrast to Eq. (3.7), the presence of equalisation errors results in the following (noncausal) LTV difference equation representation for the input-output relationship of the equaliser-decision device combination

$$r(k) = \sum_{i=0}^{P-1} v_i(k)\hat{u}(k + \Delta - i) + n(k) \qquad (3.11)$$

where $v_i(k) = h_i/\gamma(k - i)$. The time-variance of $\gamma(k)$ therefore carries over to the parameters of the equaliser-decision device combination, yielding the following key relationship:

$$\text{The eye is open} \Longleftrightarrow v_i(k) = \Gamma h_i, \; i = 0, 1, \dots, P - 1$$

$$\text{The eye is closed} \Longleftrightarrow v_i(k) = \left(u(k - i)/\hat{u}(k - i + \Delta)\right) h_i, \; i = 0, 1, \dots, P - 1.$$

$$(3.12)$$

iii. A comparison of the $v_i(k)$ under open and closed-eye conditions in Eq. (3.12) reveals that the time-invariance of the parameters of the linear system taking $\hat{u}(k)$ to $r(k)$ is a necessary and sufficient condition for the equalisation objective to hold since h is time-invariant. In the following section this fact will be exploited to construct a statistical test to distinguish between time-invariance and time-variance of the system parameters using only the observations available at the equaliser.

3.4   Testing For Equalisation Errors

3.4.1   Preliminaries

The one-to-one correspondence between the attainment of the equalisation objective and the time-invariance of the input-output relationship of the equaliser-decision device combination, which was proved in the previous section, can be utilised to construct statistical tests with the aim of finding out whether Eq. (3.2) is true for the duration of decision device output observations. The statistical tests constructed in this section will be based on the Neyman-Pearson lemma (see e.g. [3]) and shown to be *uniformly most powerful* (UMP). A study of the error detection performance will be presented later in the section.

In order to obtain an inference about the time-invariance of the underlying system parameters, finite-length records of decision device output and equaliser input sequences will be used. The inference will be based on an estimation of the parameters $v_i(k)$ in Eq. (3.11) using the *method of least squares*. Statistical properties of the LS estimates will be manipulated to formulate the problem of detecting time variations in the parameters as a likelihood ratio test. The choice of the method of least squares directly follows from the formulation of the problem as one of detecting time variations in the system parameters from the decision device output to the equaliser input rather than the other way around. The LS parameter estimates can be obtained from a record of supposedly correct data $\{\hat{u}(k)\}$ and noisy observations $\{r(k)\}$ by finding a minimum norm perturbation on the noisy observations that makes the observations consistent. It is also possible to obtain parameter estimates of the system from the equaliser input to the decision device output using *data least squares* [30], but the statistical properties of the resulting estimates are not easy to come by.

Consider first the case of no equalisation errors in which the input-output relationship of the equaliser-decision device combination can be represented by the following matrix equation obtained from Eq. (3.7)

$$r = Av + n \tag{3.13}$$

where r is the $N \times 1$ vector of the noisy channel output observations delayed by Δ

$$r = [r(k - \Delta), r(k - \Delta - 1), \ldots, r(k - \Delta - N + 1)]^T,$$

A is an $N \times P$ Hankel matrix of decision device outputs $(N > P)$ with full column rank

$$
A = \begin{bmatrix}
\hat{u}(k) & \hat{u}(k-1) & \cdots & \hat{u}(k-P+1) \\
\hat{u}(k-1) & \hat{u}(k-2) & \cdots & \hat{u}(k-P) \\
\vdots & \vdots & & \vdots \\
\hat{u}(k-N+1) & \hat{u}(k-N) & \cdots & \hat{u}(k-N-P+2)
\end{bmatrix},
$$

v is the $P \times 1$ parameter vector which is the same as the channel parameter vector except for a scaling factor $\Gamma \in \{-1, 1\}$

$$
v = \Gamma h = [v_0, v_1, \ldots, v_{P-1}]^T
$$

and n is the $N \times 1$ noise vector delayed by Δ

$$
n = [n(k-\Delta), n(k-\Delta-1), \ldots, n(k-\Delta-N+1)]^T.
$$

The LS cost function is defined as the squared Euclidean norm of the noise vector

$$
\mathcal{J}_{\text{LS}}(v) \triangleq \|n\|_2^2 = (r - Av)^T (r - Av). \tag{3.14}
$$

The LS estimate of v is obtained by minimising $\mathcal{J}_{\text{LS}}(v)$ over all possible v:

$$
\hat{v} \triangleq \arg\min_v \mathcal{J}_{\text{LS}}(v) = A^\dagger r \tag{3.15}
$$

where $A^\dagger$ is the *Moore-Penrose inverse* of A defined by

$$
A^\dagger \triangleq \left(A^T A\right)^{-1} A^T. \tag{3.16}
$$

Central to LS parameter estimation is the concept of orthogonal projection of r onto signal and noise subspaces. The projection onto noise subspace is given by the $N \times N$ matrix $P = I - AA^\dagger$ with rank $N - P$. The projection P has a number of important properties; some of these that will be used later are $PA = 0$ (P annihilates A), $P^T = P$ (P is symmetric) and $PP = P$ (P is idempotent). The resulting LS noise estimate (or the fitting error) is given by

$$
\hat{n} = Pr = r - A\hat{v}. \tag{3.17}
$$

The distribution of the LS noise estimate can be used for model validation purposes of Eq. (3.13). Since n is Gaussian and v is deterministic, the distribution of $\hat{n}$ conditioned on the elements of

A given by

$$\widehat{\mathcal{U}} \triangleq \{\hat{u}(k), \hat{u}(k-1), \ldots, \hat{u}(k-N-P+2)\}$$

is multivariate Gaussian with mean μ and covariance K. Note that the time indices of the elements of $\widehat{\mathcal{U}}$ define the interval $\mathcal{O}$ in Eq. (3.2).

As in the previous chapter, the detection problem will be formulated in the hypothesis testing framework with the null hypothesis (H_0) denoting an error-free recovery of the channel input at the decision device output and the alternative hypothesis (H_1) referring to the presence of errors in the decision device output sequence. The following proposition specifies the conditional mean and covariance of the LS noise estimate under the null and alternative hypotheses.

Proposition 3.1 If $n \sim \mathcal{N}(0, \Sigma)$, the conditional distribution of $\hat{n}$ given $\widehat{\mathcal{U}}$ is $\mathcal{N}(\mu, K)$. Under H_0, the conditional mean is $\mu = 0$ and, under H_1, $\mu = P(A \circ G)h$, where $\circ$ denotes the Hadamard product (entrywise multiplication) and the $N \times P$ matrix G is defined by

$$G = \begin{bmatrix} \frac{1}{\gamma(k-\Delta)} & \frac{1}{\gamma(k-\Delta-1)} & \cdots & \frac{1}{\gamma(k-\Delta-P+1)} \\ \frac{1}{\gamma(k-\Delta-1)} & \frac{1}{\gamma(k-\Delta-2)} & \cdots & \frac{1}{\gamma(k-\Delta-P)} \\ \vdots & \vdots & & \vdots \\ \frac{1}{\gamma(k-\Delta-N+1)} & \frac{1}{\gamma(k-\Delta-N)} & \cdots & \frac{1}{\gamma(k-\Delta-N-P+2)} \end{bmatrix}. \tag{3.18}$$

The conditional covariance of $\hat{n}$ is $K = P\Sigma P$ irrespective of the hypotheses.

Proof. Conditional Gaussianity of $\hat{n}$ follows from Eq. (3.17). In terms of the delayed noise-free channel output vector $x = [x(k-\Delta), \ldots, x(k-\Delta-N+1)]^T$ and the noise vector n, Eq. (3.17) can be rewritten as

$$\hat{n} = P(x + n) \tag{3.19}$$

whence we obtain

$$\mu = E\{P(x+n) \mid \widehat{\mathcal{U}}\} \tag{3.20a}$$

$$= Px \tag{3.20b}$$

since x is nonrandom once $\widehat{\mathcal{U}}$ is given and $E\{n\} = 0$. It follows from Eq. (3.13) that, under H_0, $x = Av$ and

$$\mu = PAv \tag{3.21a}$$

$$= 0 \tag{3.21b}$$

where we used the property that P annihilates A. Under H_1, however, the parameter vector v

is time-varying as is evident from Eq. (3.11) which can be concatenated in a similar fashion to Eq. (3.13) to yield

$$
x = \begin{bmatrix}
\frac{\hat{u}(k)}{\gamma(k-\Delta)} & \frac{\hat{u}(k-1)}{\gamma(k-\Delta-1)} & \cdots & \frac{\hat{u}(k-P+1)}{\gamma(k-\Delta-P+1)} \\
\frac{\hat{u}(k-1)}{\gamma(k-\Delta-1)} & \frac{\hat{u}(k-2)}{\gamma(k-\Delta-2)} & \cdots & \frac{\hat{u}(k-P)}{\gamma(k-\Delta-P)} \\
\vdots & \vdots & & \vdots \\
\frac{\hat{u}(k-N+1)}{\gamma(k-\Delta-N+1)} & \frac{\hat{u}(k-N)}{\gamma(k-\Delta-N)} & \cdots & \frac{\hat{u}(k-N-P+2)}{\gamma(k-\Delta-N-P+2)}
\end{bmatrix} h. \tag{3.22}
$$

Written in a compact matrix form, Eq. (3.22) can be seen to be equivalent to $x = (A \circ G)\, h$.

Noting that P is symmetric, the conditional covariance of $\hat{n}$ given $\widehat{\mathcal{U}}$ can be written in terms of the noise covariance Σ as

$$
K = E\left\{ (\hat{n} - \mu)(\hat{n} - \mu)^T \mid \widehat{\mathcal{U}} \right\} \tag{3.23a}
$$

$$
= E\left\{ Pnn^T P \mid \widehat{\mathcal{U}} \right\} \tag{3.23b}
$$

$$
= P\Sigma P \tag{3.23c}
$$

which completes the proof. ∎

We are now interested in detecting any departure of the conditional mean μ from the zero vector, which unravels the presence of equalisation errors. This can be formally cast into the following hypothesis testing problem:

H_0: $\mu = 0$ almost surely (the eye is open or no equalisation errors)

H_1: $\Pr\{\mu \neq 0\} > 0$ (the eye is closed or $\{\hat{u}(k)\}$ contains errors).

We next prove that the rank of K is $N - P$, which implies that the LS noise estimate $\hat{n}$ given the decision device output sequence $\widehat{\mathcal{U}}$ has a *degenerate* multivariate Gaussian distribution.

Theorem 3.1 *Given an $N \times N$ projection P with rank $N - P$ and an $N \times N$ positive definite (i.e. full rank) covariance matrix Σ, the rank of the covariance matrix $K = P\Sigma P$ is $N - P$.*

Proof. By Corollary 6.1 in [31] (extension of Sylvester's law),

$$
\text{rank}(\Sigma) + \text{rank}(P) - N \leq \text{rank}(\Sigma P) \leq \min(\text{rank}(\Sigma), \text{rank}(P)). \tag{3.24}
$$

Since $\text{rank}(\Sigma) = N$ and $\text{rank}(P) = N - P$, the inequality in Eq. (3.24) reduces to the equality $\text{rank}(\Sigma P) = N - P$. By Corollary 6.2 in [31],

$$
\text{rank}(P\Sigma P) = \text{rank}(\Sigma P) - \dim\left(\text{null}(P) \cap \text{col}(\Sigma P)\right) \tag{3.25}
$$

where null($\cdot$) denotes the null space and col($\cdot$) the column space. The null space of P is given by $A\alpha$ for any $P \times 1$ real vector α. Thus, the intersection of null(P) and col(ΣP) can be written as

$$\text{null}(P) \cap \text{col}(\Sigma P) = \{\xi \mid \xi = A\alpha = \Sigma P\beta \text{ for some } \alpha, \beta\}. \tag{3.26}$$

For the matrix equation

$$A\alpha = \Sigma P\beta \tag{3.27}$$

to be consistent so that it can be satisfied for some nonzero α and β, we must have [27, Theorem 2.3.1]

$$A\alpha = \Sigma P(\Sigma P)^{-}A\alpha \tag{3.28a}$$

$$= \Sigma P\Sigma^{-1}A\alpha \tag{3.28b}$$

which is satisfied if and only if one of the following is true

i. $\Sigma P\Sigma^{-1} = I$,

ii. $\Sigma P\Sigma^{-1} = I - P$.

Noting that condition **i.** cannot be satisfied since $\Sigma P\Sigma^{-1}$ has rank $N - P$ which is less than the rank of I, condition **ii.** remains to be the only requirement for Eq. (3.27) to be consistent. Postmultiplying $\Sigma P\Sigma^{-1}$ by ΣP in **ii.** above results in

$$\Sigma P = (I - P)\Sigma P \tag{3.29}$$

which implies that if Eq. (3.27) is consistent, then $P\Sigma P = 0$. Thus, Eq. (3.27) cannot be consistent and the intersection space in Eq. (3.26) consists of the zero vector only, implying

$$\dim\left(\text{null}(P) \cap \text{col}(\Sigma P)\right) = 0.$$

It therefore follows from Eq. (3.25) that rank($P\Sigma P$) = rank(ΣP) = $N - P$.   ∎

3.4.2   Testing the Equality of the Mean Vector to Zero

Using Proposition 3.1, the null and alternative hypotheses can be equivalently written as

H_0: $\hat{n} \sim \mathcal{N}(0, K)$

H_1: $\hat{n} \sim \mathcal{N}(\mu, K)$ where $\mu \neq 0$.

For nondegenerate multivariate Gaussian distributions with probability density function $f(\widehat{n})$, the following test statistic can be used to distinguish between H_0 and H_1 [32]

$$\int f(\widehat{n}) \ln f(\widehat{n}) d\widehat{n} - \ln f(\widehat{n}) \tag{3.30}$$

which can be shown to be equivalent to the quadratic form $\widehat{n}^T K^{-1} \widehat{n}$. In our case, however, the covariance matrix K cannot be inverted since it is rank deficient as proved in Theorem 3.1. All the same, the following test statistic can be considered as a natural extension of Eq. (3.30):

$$T = \widehat{n}^T K_r^- \widehat{n} \tag{3.31}$$

where K_r^- denotes a *symmetric reflexive g-inverse* of $K = P\Sigma P$. Refer to Appendix 3.A for a discussion of the computational aspects of g-inverses. Before proceeding any further, we will give an important result concerning the distribution of T defined in Eq. (3.31).

Theorem 3.2 [27] *If $\widehat{n} \sim \mathcal{N}(\mu, K)$, then the quadratic form $\widehat{n}^T K_r^- \widehat{n}$ has a noncentral chi-square distribution with $d = \mathrm{rank}(K)$ degrees of freedom and noncentrality parameter $\rho = \mu^T K_r^- \mu$.*

Theorem 3.2 is in fact a direct consequence of Theorem 2.3 which sets out necessary and sufficient conditions for a quadratic form to have a (noncentral) chi-square distribution. According to Proposition 3.1, and Theorems 3.1 and 3.2, the test statistic T has a *central* chi-square distribution with $N - P$ degrees of freedom under the null hypothesis and a *noncentral* chi-square distribution with $N - P$ degrees of freedom and noncentrality parameter $\rho = \mu^T K_r^- \mu$ under the alternative hypothesis. Therefore, the hypotheses of the detection problem can be written as follows

H_0: $T \sim \chi^2_{N-P}$

H_1: $T \sim \chi^2_{N-P}(\rho)$ where $\rho = \mu^T K_r^- \mu > 0$.

The conditional mean and variance of the proposed test statistic are

$$E\{T \mid H_0\} = N - P$$

$$\mathrm{Var}\{T \mid H_0\} = 2(N - P)$$

$$E\{T \mid H_1\} = N - P + \rho$$

$$\mathrm{Var}\{T \mid H_1\} = 2(N - P) + 4\rho.$$

The detection problem described at the beginning of this subsection has now been reduced to a one-sided binary hypothesis testing problem of the *simple* null hypothesis H_0: $\rho = 0$ versus the *composite* alternative hypothesis H_1: $\rho > 0$.

As a parenthetical remark, we note that if the channel noise $n(k)$ is white Gaussian with a covariance matrix of the form $\Sigma = \sigma^2 I$, the covariance matrix of $\hat{n}$ becomes $K = \sigma^2 P$ which has a symmetric reflexive g-inverse given by $K_r^- = \sigma^{-2} P$. The test statistic in Eq. (3.31) can now be written as $T = \hat{n}^T \hat{n}/\sigma^2 = r^T P r/\sigma^2$ which has the same conditional distribution as in the coloured channel noise case. This test is known as the χ^2 *goodness-of-fit test* in the literature (see e.g. [33]) and is used to verify the underlying model assumptions in LS parameter estimation problems.

Turning back to the test statistic of Eq. (3.31), we will now show that it is possible to design a threshold test with certain optimality properties.

Definition 3.1 (Uniformly most powerful tests) *A test of H_0 versus H_1 is uniformly most powerful (UMP) with significance level α if its probability of false alarm is equal to α and its power (probability of detection) is uniformly greater than the power of any other test whose significance level is less than or equal to α.*

Definition 3.2 (Monotone likelihood ratio) *The real-parameter family of densities $f_p(x)$ parameterised by p is said to have a monotone likelihood ratio if the real-valued function (likelihood ratio)*

$$\Lambda(x) = \frac{f_{p_1}(x)}{f_{p_0}(x)} \tag{3.32}$$

is a nondecreasing function of x for any $p_0 < p_1$.

Theorem 3.3 (UMP one-sided tests) [33] *Let X be a real-valued random variable whose density function is parameterised by p. If X has a monotone likelihood ratio, then the threshold test*

$$X \overset{H_1}{\underset{H_0}{\gtrless}} \eta, \quad \Pr\{X > \eta \mid p = p_0\} = \alpha \tag{3.33}$$

is the UMP test with significance level α for testing H_0: $p \leq p_0$ versus H_1: $p > p_0$.

The probability density function of noncentral chi-square distribution parameterised by its noncentrality parameter has a monotone likelihood ratio [33, 34]. Therefore, by Theorem 3.3 the threshold test

$$T \overset{H_1}{\underset{H_0}{\gtrless}} \eta \tag{3.34}$$

is the UMP test of H_0: $\rho = 0$ versus H_1: $\rho > 0$. The test threshold η is determined by setting α equal to the probability of false alarm

$$P_{FA} \triangleq \Pr\{T > \eta \mid H_0\} = \int_{\eta}^{\infty} f(T \mid H_0) dT \tag{3.35}$$

and the probability of detection (the power) of the test is accordingly given by

$$P_D \triangleq \Pr\{T > \eta \mid H_1\} = \int_{\eta}^{\infty} f(T \mid H_1) dT \tag{3.36}$$

where $f(T \mid H_i)$, $i = 0, 1$, denotes the probability density function of T under the null and alternative hypotheses, respectively.

Remark 3.2 Whilst $\rho > 0$ implies $\mu \neq 0$, the converse is not necessarily true since rank $\left(K_r^-\right) = $ rank$(K) = N - P$ by the definition of reflexive g-inverses [27]. In other words, K_r^- is positive semidefinite. Thus, the composite alternative hypotheses $\mu \neq 0$ and $\rho > 0$ are not exactly equivalent. This can be rectified by replacing the symmetric reflexive g-inverse with a positive definite (full rank) g-inverse in the definition of the test statistic T in Eq. (3.31). But then the alternative hypothesis distribution of T will no longer be chi-square because condition **iii.** of Theorem 2.3 will not be satisfied. In this case the test in Eq. (3.34) cannot be guaranteed to be UMP unless the resulting T has a monotone likelihood ratio. A method for computing positive definite g-inverses is outlined in Appendix 3.A.1.

3.4.3   Relationship of the Noncentrality Parameter to the Test Power

We will now establish the connection between the noncentrality parameter ρ and the test power P_D. If the number of degrees of freedom d of a chi-square random variable satisfies the inequality $d = N - P > 30$, the chi-square random variable can be approximated by a Gaussian random variable using the following square-root transformation, also known as Fisher's result [35, pp. 399–401],

$$\sqrt{2\chi_d^2} \stackrel{\text{approx}}{\sim} \mathcal{N}\left(\sqrt{2d-1}, 1\right) \tag{3.37}$$

which is equivalent to

$$\left(\sqrt{2\chi_d^2} - \sqrt{2d-1}\right) \stackrel{\text{approx}}{\sim} \mathcal{N}(0, 1). \tag{3.38}$$

Using Eq. (3.38) and assuming $d > 30$, the significance level of the test can be written in terms of the test threshold η as follows

$$\alpha = \Pr\{\chi_d^2 > \eta\} \tag{3.39a}$$

$$\approx \frac{1}{\sqrt{2\pi}} \int_l^\infty \exp\left(-\frac{x^2}{2}\right) dx = \mathrm{erfm}(l), \quad l = \sqrt{2\eta} - \sqrt{2d - 1} \tag{3.39b}$$

where $\mathrm{erfm}(l)$ denotes the *modified complementary error function* which is related to the more familiar error function through $\mathrm{erfm}(l) = 1/2 - (1/2)\mathrm{erf}(l/\sqrt{2})$. Thus, an approximate expression for the test threshold η is

$$\eta \approx \frac{1}{2}\left(\mathrm{erfm}^{-1}(\alpha) + \sqrt{2d - 1}\right)^2. \tag{3.40}$$

The following Gaussian approximation of a noncentral chi-square random variable [36] will be used to obtain an approximate closed-form expression for the test power P_D

$$\frac{1}{\sqrt{\mathrm{Var}\{\chi_d^2(\rho)\}}}\left(\chi_d^2(\rho) - E\{\chi_d^2(\rho)\}\right) \overset{\mathrm{approx}}{\sim} \mathcal{N}(0, 1), \quad d > 30. \tag{3.41}$$

When applied to our test statistic T under H_1, this transformation leads to the expression

$$P_D \approx \mathrm{erfm}\left(\frac{\eta - E\{\chi_d^2(\rho)\}}{\sqrt{\mathrm{Var}\{\chi_d^2(\rho)\}}}\right). \tag{3.42}$$

Substituting Eq. (3.40) in Eq. (3.42) and noting that $E\{\chi_d^2(\rho)\} = d + \rho$ and $\mathrm{Var}\{\chi_d^2(\rho)\} = 2d + 4\rho$, Eq. (3.42) can be rewritten as

$$P_D \approx \mathrm{erfm}\left(\frac{-\rho + \mathrm{erfm}^{-1}(\alpha)\sqrt{2d - 1} + \frac{1}{2}\left((\mathrm{erfm}^{-1}(\alpha))^2 - 1\right)}{\sqrt{2d + 4\rho}}\right). \tag{3.43}$$

If α is fixed, P_D is a monotone increasing function of

$$\varsigma \triangleq \frac{\rho - \mathrm{erfm}^{-1}(\alpha)\sqrt{2d - 1}}{\sqrt{2d + 4\rho}} \tag{3.44}$$

which we will call the equivalent "SNR" [36] for the detection problem at hand. It is easy to see that P_D increases monotonically with ρ if d is constant.

Channel Noise Effects

The actual SNR at the channel output affects the detection performance of the test through the equivalent SNR. The noncentrality parameter is a function of both μ and K. If the decision

device output contains errors, μ will be different than zero. The channel noise shows up in the test statistic through the covariance matrix K. Let us assume that the LS noise estimate covariance is given by $K = aC$, where $a > 0$ and C is a constant symmetric positive semidefinite matrix. Then, for given equaliser input and decision device output observations (i.e. for a fixed μ), we obtain $\rho = (1/a)\mu^T C_r^- \mu$. As a decreases, which implies an increase in the channel output SNR, ρ and therefore ς will increase, leading to a higher P_D. To recapitulate, the higher the channel output SNR, the higher the test power P_D will be. Note that this property is not shared by the convergence test of Chapter 2.

3.4.4   Relationship of Equalisation Errors to the Noncentrality Parameter

The test power increases monotonically with the noncentrality parameter as shown in the previous subsection. To complete our analysis of the detection performance, we now consider the contribution of equalisation errors to the noncentrality parameter of the test statistic. Recalling that the noncentrality parameter ρ is defined in terms of the mean vector μ (see Theorem 3.2), we first investigate the contribution of a single error in the decision device output sequence $\{\hat{u}(k)\}$ to μ. The relationship between individual equalisation errors and μ is set out in the following theorem.

Theorem 3.4 *If $\hat{u}(k - m) \in \widehat{\mathcal{U}}$ is in error, its contribution to μ will be*

$$\mu_c(m) = s(m)P\tilde{I}(m)h \tag{3.45}$$

where $s(m)$ is the error

$$s(m) = u(k - \Delta - m) - \Gamma\hat{u}(k - m) \tag{3.46}$$

and $\tilde{I}(m)$ is defined by

$$\tilde{I}(m) = \begin{bmatrix} e_1^T \\ e_2^T \\ \vdots \\ e_N^T \end{bmatrix} \tag{3.47}$$

with

$$e_i^T = [\underbrace{0,\dots,0}_{m+1-i \text{ times}}, 1, 0, \dots, 0]_{1 \times P}. \tag{3.48}$$

If $m + 1 - i < 0$ or $m + 1 - i > P$, then $e_i = 0$.

Proof. The conditional mean of $\hat{n}$ can be written as

$$\mu = P(A \circ G + \Gamma A - \Gamma A)h \tag{3.49}$$

which is equivalent to

$$\mu = P(A \circ G - \Gamma A)h \qquad (3.50)$$

since $\Gamma P A h = 0$. The difference of the matrices $A \circ G - \Gamma A$ contains all the equalisation errors in the observation interval $\mathcal{O}$. Indeed, if $\hat{u}(k-m)$ is in error $(0 \le m \le N + P - 2)$, we get

$$A \circ G - \Gamma A = \hat{u}(k - m) \left(\frac{1}{\gamma(k - \Delta - m)} - \Gamma \right) \tilde{I}(m) \qquad (3.51)$$

where $\tilde{I}(m)$ is an $N \times P$ matrix with ones on its backward diagonal corresponding to the time-varying gain $1/\gamma(k - \Delta - m)$ in G. We can obtain the contribution of a single equalisation error to μ by substituting Eq. (3.51) in Eq. (3.50), which results in Eq. (3.45). ∎

The overall contribution of equalisation errors to μ is then given by the summation of individual contributions in the interval $\mathcal{O}$, i.e. $\mu = \sum_{i=0}^{N+P-2} \mu_c(i)$. The condition for a decision device output error occurring at time $k - m$ to lead to an increase in ρ and, therefore, an incremental improvement in the detection performance, is

$$\mu^T K_r^- \mu \ge (\mu - \mu_c(m))^T K_r^- (\mu - \mu_c(m)) \iff 2\mu^T K_r^- \mu_c(m) \ge \mu_c^T(m) K_r^- \mu_c(m). \qquad (3.52)$$

Expanding the inequality on the right-hand side of Eq. (3.52) using Eq. (3.45), we obtain the following inequality, which provides additional insight into the above condition:

$$s(m) \sum_{\substack{i=0 \\ i \ne m}}^{N+P-2} s(i) h^T \tilde{I}^T(i) P K_r^- P \tilde{I}(m) h \ge -\frac{1}{2} \mu_c^T(m) K_r^- \mu_c(m). \qquad (3.53)$$

Intuitively, the detection performance should improve with increasing number of equalisation errors. However, ρ cannot be guaranteed to increase with every equalisation error unless Eq. (3.53) is satisfied. The right-hand side of Eq. (3.53) is always less than or equal to zero. However, the sign of the left-hand side of Eq. (3.53) cannot be determined without a knowledge of the quantities involved. It is therefore not possible to say anything conclusive about the relation of ρ to the number of equalisation errors. A proportional increase in μ with the number of equalisation errors is of lesser importance, albeit still desirable, if some of the equalisation errors result in comparatively large increases in μ.

3.4.5   Asymptotic Detection Analysis

It is of particular interest to analyse the detection performance of the test as the observation length approaches infinity. Asymptotic detection performance is closely related to the concept of

consistency. In this subsection, we will derive conditions that assure the consistency of the test. Although we are primarily concerned with the detection of errors in finite-length observations as implied by Eq. (3.2), we will find it useful to gain insight into the dependence of the test performance on the sample size $N + P - 1$ as it becomes large. For instance, if doubling the observation length turns out to yield substantial improvement in the error detection performance, then it may be worthwhile to increase the number of observations accordingly. Our first result is concerned with the asymptotic approximation of the LS noise estimate covariance K.

Theorem 3.5 *If the length of the channel impulse response P is finite, $\{\hat{u}(k)\}$ is second-order ergodic, $A^T A$ has full rank, and $\{n(k)\}$ has a finite span of dependence w, the covariance matrix K converges stochastically[5] to Σ as N increases.*

Proof. Let us write K explicitly in terms of A

$$
\begin{aligned}
K = \Sigma - \Sigma A \left(A^T A\right)^{-1} A^T - A \left(A^T A\right)^{-1} A^T \Sigma \\
+ A \left(A^T A\right)^{-1} A^T \Sigma A \left(A^T A\right)^{-1} A^T.
\end{aligned}
\tag{3.54}
$$

Since P is finite and $\{\hat{u}(k)\}$ is second-order ergodic, we have

$$
\frac{1}{N} A^T A \overset{N \to \infty}{\Rightarrow} E\left\{\hat{u}(k)\hat{u}^T(k)\right\} \quad \text{w.p.1}
\tag{3.55}
$$

where $\hat{u}(k) = [\hat{u}(k), \ldots, \hat{u}(k-P+1)]^T$ and $E\left\{\hat{u}(k)\hat{u}^T(k)\right\}$, which is the second-order moment matrix of the decision device output, is necessarily finite since $\{\hat{u}(k)\}$ is made up of finite-valued symbols. For large N, the matrix product $A^T \Sigma A$ may be approximately written as

$$
A^T \Sigma A \approx N\sigma_{1,1} E\left\{\hat{u}(k)\hat{u}^T(k)\right\} + 2 \sum_{i=1}^{N-1} (N - i)\sigma_{1,i+1} E\left\{\hat{u}(k)\hat{u}^T(k - i)\right\}
\tag{3.56}
$$

where $\sigma_{i,j}$ is the (i, j)th entry of the covariance matrix Σ which is Toeplitz by assumption. Under the assumption that $\{n(k)\}$ has a finite span of dependence, the matrix product $A^T \Sigma A / N$ has a finite limit as N tends to infinity:

$$
\frac{1}{N} A^T \Sigma A \overset{N \to \infty}{\Rightarrow} \sigma_{1,1} E\left\{\hat{u}(k)\hat{u}^T(k)\right\} + 2 \sum_{i=1}^{w} \sigma_{1,i+1} E\left\{\hat{u}(k)\hat{u}^T(k - i)\right\} \quad \text{w.p.1.}
\tag{3.57}
$$

---

[5]A sequence t_k is said to *converge stochastically* to θ if, for small numbers $\epsilon > 0$, $\lambda > 0$, there is some K such that

$$
\Pr\{|t_k - \theta| < \epsilon\} > 1 - \lambda, \quad k > K.
$$

Using a similar argument, the $N \times P$ matrix product ΣA can be shown to have entries with finite limiting values as N tends to infinity. Putting all the above results together, we see that, for large N, K can be approximately written as

$$
\begin{aligned}
K \approx \Sigma &- \frac{1}{N} \Sigma A \left(E \left\{ \widehat{u}(k) \widehat{u}^T(k) \right\} \right)^{-1} A^T - \frac{1}{N} A \left(E \left\{ \widehat{u}(k) \widehat{u}^T(k) \right\} \right)^{-1} (\Sigma A)^T \\
&+ \frac{1}{N} A \left[\sigma_{1,1} \left(E \left\{ \widehat{u}(k) \widehat{u}^T(k) \right\} \right)^{-1} + 2 \left(E \left\{ \widehat{u}(k) \widehat{u}^T(k) \right\} \right)^{-1} \right. \\
&\left. \times \sum_{i=1}^{w} \sigma_{1,i+1} E \left\{ \widehat{u}(k) \widehat{u}^T(k-i) \right\} \left(E \left\{ \widehat{u}(k) \widehat{u}^T(k) \right\} \right)^{-1} \right] A^T.
\end{aligned}
\tag{3.58}
$$

Thus, all terms except the first one in Eq. (3.54) vanish as N tends to infinity, thereby leading to the conclusion that the LS noise estimate covariance K asymptotically approaches the noise covariance Σ with probability one (w.p.1). Convergence with probability one implies stochastic convergence. In Eq. (3.58) we used the fact that $E \left\{ \widehat{u}(k) \widehat{u}^T(k) \right\}$ is a finite matrix and its inverse exists since $A^T A$ has full rank. Note that the dimensions of K and Σ increase linearly with N. ∎

The mean of $\widehat{n}$ can be expanded as

$$
\mu = \left((A \circ G) - A \left(A^T A \right)^{-1} A^T (A \circ G) \right) h.
\tag{3.59}
$$

Noting that $A \circ G$ is the channel input counterpart of A, under the ergodicity assumption, the matrix product $A^T(A \circ G)/N$ converges stochastically to $E \left\{ \widehat{u}(k) u^T(k - \Delta) \right\}$ as N increases, where $u(k) = [u(k), \ldots, u(k - P + 1)]^T$. Then, μ can be asymptotically written as

$$
\mu \overset{N \to \infty}{\Rightarrow} \mu_a = \left((A \circ G) - A \left(E \left\{ \widehat{u}(k) \widehat{u}^T(k) \right\} \right)^{-1} E \left\{ \widehat{u}(k) u^T(k - \Delta) \right\} \right) h \quad \text{w.p.1.} \tag{3.60}
$$

It is interesting to note that μ_a is determined not only by the difference between $u(k - \Delta)$ and $\Gamma \widehat{u}(k)$, but also the second-order statistics of the channel input and the decision device output. In Eq. (3.60) the matrix product

$$
\left(E \left\{ \widehat{u}(k) \widehat{u}^T(k) \right\} \right)^{-1} E \left\{ \widehat{u}(k) u^T(k - \Delta) \right\}
$$

is in fact an estimate of

$$
[g(k, \Omega), g(k - 1, \Omega), \ldots, g(k - P + 1, \Omega)].
$$

Considering that $(A \circ G)h$ is the noise-free channel output x and $[g(k, \Omega), g(k-1, \Omega), \ldots, g(k - P + 1, \Omega)]h$ is the convolution of $g(k, \Omega)$ and h, we immediately see that μ is asymptotically

equal to the difference between x and its estimate

$$A \left(E \left\{ \widehat{u}(k) \widehat{u}^T(k) \right\} \right)^{-1} E \left\{ \widehat{u}(k) u^T(k - \Delta) \right\} h.$$

This difference is generally nonzero under H_1.

The test is consistent if $\lim_{N \to \infty} P_D = 1$. Recalling that P_D is a monotone increasing function of ς, the consistency requires $\lim_{N \to \infty} \varsigma = \infty$ or, using Eq. (3.44),

$$\lim_{N \to \infty} \frac{\rho}{\sqrt{2(N - P) + 4\rho}} - \lim_{N \to \infty} \text{erfm}^{-1}(\alpha) \sqrt{\frac{2(N - P) - 1}{2(N - P) + 4\rho}} = \infty \qquad (3.61)$$

where the second limit takes some finite value in the interval $[0, \text{erfm}^{-1}(\alpha)]$, depending on how ρ varies with N. Therefore, the test is consistent if

$$\lim_{N \to \infty} \frac{\rho}{\sqrt{2(N - P) + 4\rho}} = \infty \qquad (3.62)$$

which, assuming P is finite, implies that the noncentrality parameter ρ must be proportional to $N^{\frac{1}{2}+\epsilon}$, $\epsilon > 0$. Thus, using the approximation for K in Theorem 3.5, we reach the conclusion that, for large N, the error detection performance improves with increasing N if $\mu_a^T \Sigma^{-1} \mu_a$ is approximately proportional to $N^{\frac{1}{2}+\epsilon}$ for some $\epsilon > 0$.

3.4.6 Testing for Time-Invariance of the Mean Vector

We can deduce from the previous analysis that the decision device output is in error if the mean vector μ is time-varying for different observation records since that indicates a time variation in the parameters of the underlying linear system from the decision device output to the equaliser input. The converse is in general not true since two LS noise estimates may have the same nonzero μ. However, the probability of that happening is very small.

Alternative equalisation error tests can be constructed by checking the difference between the successive values of μ. We will consider the following approaches:

- Compute successive LS noise estimates $\widehat{n}_i$ and test for the equality of their means,

- Use $\widehat{n}$ from one observation record and test for the equality of subvectors of its mean μ.

In what follows, we will present threshold tests based on these approaches. The detection performance of the tests will not be discussed as the analysis in Sections 3.4.3 through to 3.4.5 can be used for that purpose.

Test Statistic Using Two Successive LS Noise Estimates:

Firstly we compute the LS noise estimates $\hat{n}_i$ from successive observations r_i and A_i for $i = 1, 2$. Note that the $\hat{n}_i$ given $\widehat{\mathcal{U}}_i$ are multivariate Gaussian with means $\mu_i = P_i x_i$ and covariances $K_i = P_i \Sigma P_i$ where $P_i = I - A_i A_i^{\dagger}$ and $x_i = (A_i \circ G_i) h$. The subscript i refers to different time instants k at which the quantities involved are formed or computed. We have the following hypotheses

H_0: $\mu_1 = \mu_2$ a.s.

H_1: $\Pr\{\mu_1 \neq \mu_2\} > 0$.

The difference $\varepsilon = \hat{n}_1 - \hat{n}_2$ given $\widehat{\mathcal{U}}_1$ and $\widehat{\mathcal{U}}_2$ is distributed according to $\mathcal{N}(\mu_1 - \mu_2, C)$ where

$$C = P_1 \Sigma_{11} P_1 - P_2 \Sigma_{21} P_1 - P_1 \Sigma_{12} P_2 + P_2 \Sigma_{22} P_2. \tag{3.63}$$

The noise covariance matrices Σ_{ij} are defined by

$$\Sigma_{ij} = E\left\{ n_i n_j^T \right\}, \quad i, j = 1, 2. \tag{3.64}$$

Note that $\Sigma = \Sigma_{11} = \Sigma_{22}$ as a result of stationary channel noise. Consider the test statistic $\varepsilon^T C_r^- \varepsilon$, which has central chi-square distribution with d degrees of freedom under H_0 and noncentral chi-square distribution with d degrees of freedom and noncentrality parameter ρ under H_1, where $d = \text{rank}(C)$ and $\rho = (\mu_1 - \mu_2)^T C_r^- (\mu_1 - \mu_2)$. Then, the threshold test

$$\varepsilon^T C_r^- \varepsilon \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{3.65}$$

is the UMP test with significance level α for testing the simple hypothesis H_0: $\rho = 0$ versus the composite hypothesis H_1: $\rho > 0$. The significance level of the test is equal to the probability of false alarm, i.e. $\alpha = \Pr\{\chi_d^2 > \eta\}$. The test is UMP because the chi-square distribution has a monotone likelihood ratio. Although $\rho > 0$ implies $\mu_1 \neq \mu_2$, $\rho = 0$ does not necessarily imply $\mu_1 = \mu_2$ because C_r^- is a positive semidefinite matrix. At the expense of losing the UMP property, the reflexive g-inverse of C can be replaced with a positive definite g-inverse to render $\rho = 0$ and $\mu_1 = \mu_2$ equivalent; refer to Appendix 3.A.1 for details.

Test Statistic Using a Single Observation Record:

The equalisation objective is satisfied in a given interval $\mathcal{O}$ if and only if the mean of $\hat{n}$ is $\mu = 0$. We propose to test the hypothesis that two subvectors of $\hat{n}$ have equal means, which would automatically be satisfied if the eye were open. Although this hypothesis fails to indicate the presence of equalisation errors if μ is nonzero and has equal subvectors, the likelihood of that happening is extremely small. The resulting test is similar to Eq. (3.65), except that it does not require computation of P_i from successive observations. Let the LS noise estimate be partitioned as

$$\hat{n} = \begin{bmatrix} \hat{n}_1 \\ \hat{n}_2 \end{bmatrix} \tag{3.66}$$

which has conditional multivariate Gaussian distribution given $\widehat{\mathcal{U}}$ with mean

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \tag{3.67}$$

and covariance

$$K = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix}. \tag{3.68}$$

Suppose N is even so that $\hat{n}_1$ and $\hat{n}_2$ each have $N/2$ components. Then, $\varepsilon = \hat{n}_1 - \hat{n}_2$ given $\widehat{\mathcal{U}}$ has multivariate Gaussian distribution with mean $\mu_1 - \mu_2$ and covariance $K_{11} - K_{21} - K_{12} + K_{22}$. The following threshold test

$$\varepsilon^T (K_{11} - K_{21} - K_{12} + K_{22})_r^- \varepsilon \overset{H_1}{\underset{H_0}{\gtrless}} \eta \tag{3.69}$$

rejects the null hypothesis H_0: $\rho = 0$ with a probability of false alarm given by $\Pr\{\chi_d^2 > \eta\}$, where $\rho = (\mu_1 - \mu_2)^T (K_{11} - K_{21} - K_{12} + K_{22})_r^- (\mu_1 - \mu_2)$ is the noncentrality parameter and $d = \operatorname{rank}(K_{11} - K_{21} - K_{12} + K_{22})$ is the number of degrees of freedom of the test statistic. Eq. (3.69) is the UMP test to distinguish between H_0: $\rho = 0$ and H_1: $\rho > 0$.

3.5 Effects of Undermodelling and Incorrect Equalisation Delay Assumption

Up to now we have assumed *a priori* knowledge of the channel impulse response length P and the equalisation delay Δ that would have resulted from an error-free equalisation in the interval $\mathcal{O}$. We will now consider the effects of relaxing this assumption on the applicability of the tests. In particular, we will show that a knowledge of some relative bounds on the true parameters P

and Δ is sufficient for the tests to be applicable. This feature makes the tests *robust* in the face of parameter uncertainties. As a by-product, we will introduce a simple method for estimating a lower bound on Δ when the eye is open. The delay estimator is based on the detection of an abrupt change in the null hypothesis distribution of the test statistic T in Eq. (3.31) when Δ is underestimated. This estimation scheme may prove beneficial especially in the context of blind channel equalisation where there is no straightforward means to estimate a bound on the delay incurred by the channel input sequence.

To gain insight into what happens to the LS parameter estimates under H_0 when P and Δ are replaced with arbitrary numbers P' and Δ', respectively, we will consider the following version of Eq. (3.13)

$$r' = Bv + n' \tag{3.70a}$$

$$= A'v' + \underbrace{Bv - A'v' + n'}_{w} \tag{3.70b}$$

where r' is the $N \times 1$ vector of noisy channel outputs delayed by Δ'

$$r' = [r(k - \Delta'), r(k - \Delta' - 1), \ldots, r(k - \Delta' - N + 1)]^T,$$

A' is the $N \times P'$ full-rank Hankel matrix of decision device outputs

$$A' = \begin{bmatrix} \hat{u}(k) & \hat{u}(k-1) & \cdots & \hat{u}(k-P'+1) \\ \hat{u}(k-1) & \hat{u}(k-2) & \cdots & \hat{u}(k-P') \\ \vdots & \vdots & & \vdots \\ \hat{u}(k-N+1) & \hat{u}(k-N) & \cdots & \hat{u}(k-N-P'+2) \end{bmatrix},$$

B is a version of A time-shifted by $\Delta - \Delta'$

$$B = \begin{bmatrix} \hat{u}(k+\Delta-\Delta') & \hat{u}(k+\Delta-\Delta'-1) & \cdots & \hat{u}(k+\Delta-\Delta'-P+1) \\ \hat{u}(k+\Delta-\Delta'-1) & \hat{u}(k+\Delta-\Delta'-2) & \cdots & \hat{u}(k+\Delta-\Delta'-P) \\ \vdots & \vdots & & \vdots \\ \hat{u}(k+\Delta-\Delta'-N+1) & \hat{u}(k+\Delta-\Delta'-N) & \cdots & \hat{u}(k+\Delta-\Delta'-N-P+2) \end{bmatrix},$$

v' is the $P' \times 1$ parameter vector to be estimated, and n' is the noise vector delayed by Δ'. In contrast to the case of known P and Δ, the model noise now takes the form $w = Bv - A'v' + n'$ and its LS estimate is given by

$$\hat{w} = P'r' \tag{3.71}$$

where $P' = I - A'A'^{\dagger}$ is the projection onto noise subspace. The following result sets out the influence of the arbitrary parameters P' and Δ' on the conditional distribution of $\hat{w}$.

Proposition 3.2 *If $n' \sim \mathcal{N}(0, \Sigma)$, the LS noise estimate $\hat{w}$ for arbitrary P' and Δ' is conditionally distributed according to $\mathcal{N}(\mu', K')$ given the entries of A' and B, denoted by $\hat{\mathcal{U}}'$, where $\mu' = 0$ under H_0 if the following holds:*

    i. $\Delta' \geq \Delta$

    ii. $P' \geq P - \Delta + \Delta'$.

Otherwise, the conditional mean of $\hat{w}$ is $\mu' = P'Bv$. The conditional covariance matrix of $\hat{w}$ is $K' = P'\Sigma P'$ irrespective of the hypotheses.

Proof. The conditional Gaussianity of $\hat{w}$ is easily established from Eq. (3.71). The conditional mean of $\hat{w}$ can be written as

$$\mu' = E\left\{P'(Bv + n') \mid \hat{\mathcal{U}}'\right\} \tag{3.72a}$$

$$= P'Bv. \tag{3.72b}$$

The mean is zero if $Bv = A'v'$ since P' annihilates A' or, equivalently,

$$Bv = \begin{bmatrix} x(k - \Delta') \\ x(k - \Delta' - 1) \\ \vdots \\ x(k - \Delta' - N + 1) \end{bmatrix} = A'v'. \tag{3.73}$$

If $\Delta' \geq \Delta$ and $P' \geq P + \Delta' - \Delta$, then

$$Bv = A' \underbrace{\begin{bmatrix} 0 \\ v \\ 0 \end{bmatrix}}_{v'} \begin{matrix} \Delta' - \Delta \\ P \\ P' - P - \Delta' + \Delta \end{matrix}. \tag{3.74}$$

Thus, if the inequalities in **i.** and **ii.** are satisfied, $\mu' = 0$.

The conditional covariance of $\hat{w}$ given $\hat{\mathcal{U}}'$ is

$$K' = E\left\{(\hat{w} - \mu')(\hat{w} - \mu')^T \mid \hat{\mathcal{U}}'\right\} \tag{3.75a}$$

$$= E\left\{P'n'n'^T P' \mid \hat{\mathcal{U}}'\right\} \tag{3.75b}$$

$$= P'\Sigma P' \tag{3.75c}$$

which completes the proof.   ■

If the inequalities in Proposition 3.2 are not satisfied, the conditional mean of the LS noise estimate cannot be guaranteed to be zero because Bv will not necessarily lie in the column space of A', implying that P' will not annihilate Bv. A violation of condition **i.** in Proposition 3.2 implies incorrect equalisation delay assumption, and that of condition **ii.** undermodelling of the channel.

In the light of Proposition 3.2 we maintain that the quadratic form parameterised by Δ'

$$T(\Delta') = \widehat{w}^T \left(P'\Sigma P'\right)_r^{-} \widehat{w} \tag{3.76}$$

is central chi-square distributed with $N - P'$ degrees of freedom as long as the eye is open and the inequalities in Proposition 3.2 are satisfied. If the eye is closed, however, $T(\Delta')$ has noncentral chi-square distribution with $N - P'$ degrees of freedom and noncentrality parameter

$$\rho' = \mu'^T \left(P'\Sigma P'\right)_r^{-} \mu'. \tag{3.77}$$

Thus, if P' and Δ' satisfy the inequalities in Proposition 3.2, the statistical test in Eq. (3.34) can be replaced with

$$T(\Delta') \overset{H_1}{\underset{H_0}{\gtrless}} \eta' \tag{3.78}$$

which preserves the properties of the test in Eq. (3.34).

3.5.1 Estimation of the Equalisation Delay

A method for estimating the equalisation delay using the crosscorrelation of $\{r(k)\}$ and $\{\widehat{u}(k)\}$ was proposed in Chapter 2 and reference [6]. In this section we introduce an alternative method that exploits the nonzero conditional mean of $\widehat{w}$ when the inequality $\Delta' \geq \Delta$ is violated under H_0 (see Proposition 3.2). We will assume that P' always satisfies the inequality $P' \geq P - \Delta + \Delta'$, i.e. the channel $H(z)$ is *not* undermodelled. A requirement for the delay estimator to produce the true equalisation delay Δ is that the channel be causal with $h_0 \neq 0$. If, however, the h_i has leading zeros, the estimator will yield a lower bound on the true delay, which is a property also shared by the estimator in Section 2.7.

Under the assumption that the channel is not undermodelled, we obtain the following relationship between the equalisation delay Δ and the distribution of $T(\Delta')$ in Eq. (3.76) (see

Proposition 3.2):

$$\Delta' \geq \Delta \quad \Longrightarrow \quad T(\Delta') \sim \chi^2_{N-P'}$$

$$\Delta' < \Delta \quad \Longrightarrow \quad T(\Delta') \sim \chi^2_{N-P'}(\rho'), \quad \rho' > 0.$$

The equalisation delay estimate $\widehat{\Delta}$ is then given by the minimum Δ' for which $T(\Delta') \sim \chi^2_{N-P'}$ and $T(\Delta'-1) \sim \chi^2_{N-P'}(\rho')$ where $\rho' > 0$. The estimation procedure can be outlined as follows:

i. Start with a sufficiently large Δ' chosen on physical grounds.

ii. Perform the test in Eq. (3.78).

iii. If H_0 is decided, decrease Δ' by one and go back to Step ii. If H_1 is decided, the delay estimate is $\widehat{\Delta} = \Delta' + 1$.

Note that we accept H_1 with a probability of false alarm $P_{FA} = \Pr\{\chi^2_{N-P'} > \eta' \mid H_0\}$. The test threshold η' is determined by setting P_{FA} to a significance level α. Since varying Δ' affects only r' by virtue of the problem set-up, $T(\Delta')$ does not require a great deal of computational effort once $(P' \Sigma P')_r^-$ is computed.

3.6 Simulation Studies

In this section, the test in Eq. (3.34) and the equalisation delay estimation method in Section 3.5.1 are demonstrated by means of computer simulations. We will replace the test in Eq. (3.34) with the more practical test in Eq. (3.78) which does not require an exact knowledge of P and Δ if the inequalities in Proposition 3.2 are satisfied. The first example deals with the detection of ill-convergence of an LDDE and the estimation of the equalisation delay when convergence to an open-eye parameter setting has taken place. CMA is used to update the equaliser parameters. The channel input is a PAM sequence generated by a Markov chain. The second example demonstrates the application of the same test to the detection of error propagation in a binary DFE perfectly tuned to the channel. The final example compares the performance of the error detection test in Eq. (3.78) with that of the convergence test developed in Chapter 2, using the same set-up as in Example 2.4 except for an unclipped Gaussian channel noise $\{n(k)\}$.

Example 3.1 Consider the following nonminimum phase FIR channel

$$H(z) = 1 + 1.4z^{-1} - 0.6z^{-2} + 0.2z^{-3} + 0.5z^{-4} \tag{3.79}$$
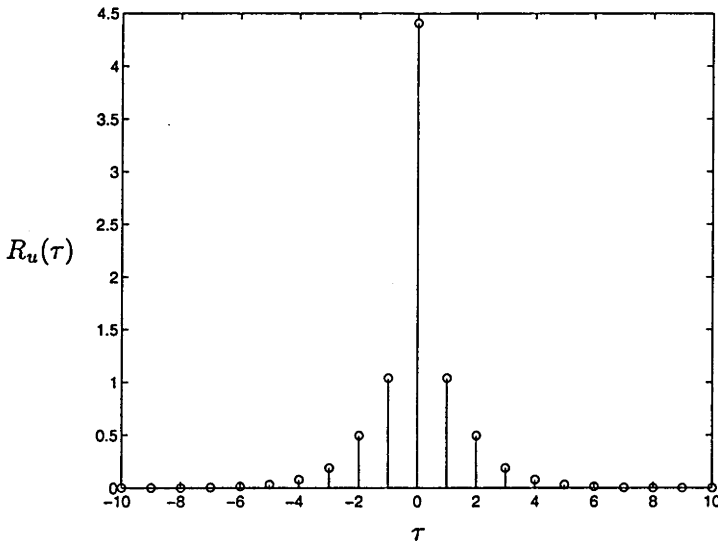
Figure 3.4 Autocorrelation of channel input sequence $\{u(k)\}$.

driven by a 4-ary PAM sequence (i.e. $M = 4$). The channel input sequence is generated by a Markov chain with state vector $m = [m_1, m_2, m_3, m_4]^T = [-3, -1, 1, 3]^T$. The output of the Markov chain is identical to its state. The transition probability matrix of the Markov chain is

$$\Pi = \begin{bmatrix} 0.100 & 0.300 & 0.200 & 0.100 \\ 0.400 & 0.519 & 0.300 & 0.100 \\ 0.300 & 0.081 & 0.100 & 0.500 \\ 0.200 & 0.100 & 0.400 & 0.300 \end{bmatrix}^T . \tag{3.80}$$

The initial states are assumed to be equiprobable, i.e. $p_0 = [0.25, 0.25, 0.25, 0.25]^T$. The mean of the channel input sequence is $E\{u(k)\} = -1.5 \times 10^{-4}$. Note that $E\{u(k)\}$ need not be equal to zero for the error detection test to be applicable. The autocorrelation of $\{u(k)\}$ is shown in Fig. 3.4. Refer to Appendix 2.B for the details of how to compute the output moments of a Markov chain.

The channel noise $\{n(k)\}$ is supposed to be a stationary coloured Gaussian process with autocorrelation $R_n(0) = 0.0011$, $R_n(\pm 1) = 0.0003$, $R_n(\pm 2) = 0.0001$, and $R_n(\tau) = 0$ for $|\tau| > 2$. Thus, the $N \times N$ covariance matrix of $\{n(k)\}$ has the following form[6]

$$\Sigma = \begin{bmatrix} R_n(0) & R_n(1) & R_n(2) & & \mathbf{0} \\ R_n(1) & \ddots & \ddots & \ddots & \\ R_n(2) & \ddots & \ddots & \ddots & R_n(2) \\ & \ddots & \ddots & \ddots & R_n(1) \\ \mathbf{0} & & R_n(2) & R_n(1) & R_n(0) \end{bmatrix} . \tag{3.81}$$

---

[6]In this case the covariance matrix has a special structure called *quintdiagonal* since $R_n(\tau) = 0$ for $|\tau| > 2$.

	True	Estimated
$E\{T(\Delta') \mid H_0\}$	10	10.03
$\mathrm{Var}\{T(\Delta') \mid H_0\}$	20	19.92
P_{FA}	0.250	0.248
	0.100	0.112
	0.050	0.054
	0.025	0.026

Table 3.1 True and estimated mean, variance and P_{FA} of $T(\Delta')$ under H_0 for $\Delta' = 15$ and $N - P' = 10$.
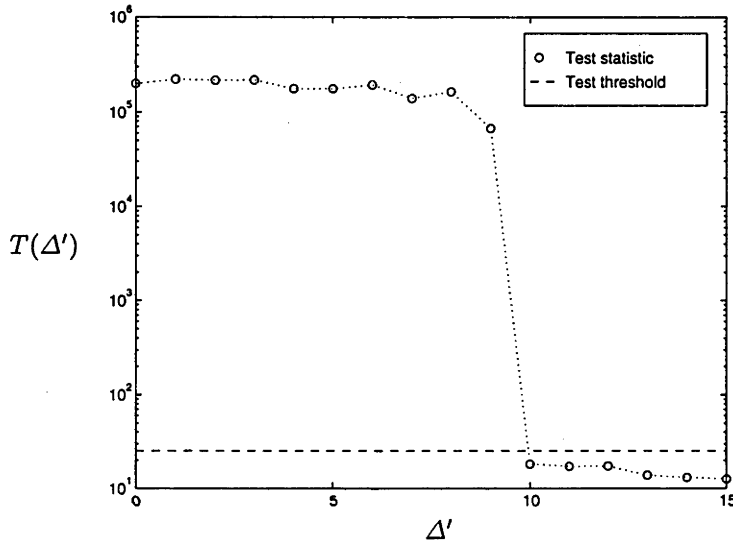
Figure 3.5 Plot of test statistic *v.* assumed equalisation delay for $\eta' = 25.2$ which corresponds to $\alpha = 0.005$.

We have simulated the case of no equalisation errors (the null hypothesis) by setting the equaliser parameters θ to a finite-length approximation of the channel inverse delayed by $\Delta = 10$. The resulting channel-equaliser combination has CLEM= 0.0897. Refer to Eq. (1.11) for the definition of closed-eye measure (CLEM). The test parameters were chosen as $N = 35$, $P' = 25$ and $\Delta' = 15$. The parameters P' and Δ' satisfy the inequalities in Proposition 3.2. The estimated mean, variance and probability of false alarm of the test statistic for 500 simulation runs are listed in Table 3.1 alongside their true values. In every simulation run $N + P' - 1 = 59$ decision device output observations were used to construct the matrix A'. The results in Table 3.1 confirm the null hypothesis distribution of $T(\Delta')$ which is supposed to be central chi-square with $N - P' = 10$ degrees of freedom.

The test statistic $T(\Delta')$ as Δ' varies in the range $15 \geq \Delta' \geq 0$ is plotted in Fig. 3.5. The test threshold for equalisation delay estimation is set to $\eta' = 25.2$ corresponding to a significance level of $\alpha = 0.005$. Observe that $T(\Delta')$ is decisively higher than η' for $0 \leq \Delta' < 10$. Thus,

the procedure in Section 3.5.1 leads us to the conclusion that the equalisation delay estimate is $\hat{\Delta} = 10$. Note that in this case $\hat{\Delta}$ is equal to the true equalisation delay Δ.

The alternative hypothesis of the detection problem was simulated using CMA [11] (or the Godard algorithm [10]) whose cost surface is known to have multiple local minima resulting in potential ill-convergence. Our objective is to illustrate the application of the test to the case where the equaliser parameters converge to a closed-eye setting as a result of ill-convergence. CMA has the following cost function[7]

$$J = \frac{1}{4}E\left\{\left(y^2(k) - R\right)^2\right\} \qquad (3.82)$$

where R is the *dispersion factor* defined by

$$R = \frac{E\{u^4(k)\}}{E\{u^2(k)\}}. \qquad (3.83)$$

The purpose of the dispersion factor is to scale the equaliser output sequence $\{y(k)\}$ in such a way that a match between the average moduli of the channel input and equaliser output is achieved. The above expression for the dispersion factor assumes that the channel input is an i.i.d. sequence. An exact expression for R when $\{u(k)\}$ is correlated requires a knowledge of the channel impulse response h, the availability of which would undoubtedly defeat the purpose of blind channel equalisation. All the same, a small deviation of R from its true value does not pose a major problem as it essentially amounts to some fluctuation in the modulus of $\{y(k)\}$ away from that of $\{u(k)\}$. Thus, we will be content to use Eq. (3.83) to compute R which gives $R = 7.9576$ for the input sequence used in this example. We set the equaliser length to $L = 21$ and used centre-tap initialisation with stepsize $\mu = 0.5 \times 10^{-5}$. CMA converged to a closed-eye parameter setting with CLEM= 1.6002 after 8000 iterations. In this case, the test parameters were chosen as $N = 35$, $P' = 25$, $\Delta' = 15$ and $\eta' = 23.2$ which corresponds to a significance level of $\alpha = 0.01$. Different observation records of A' and r' were collected to simulate the detection performance of the test. The number of equalisation errors were noted for every observation record consisting of $N + P' - 1 = 59$ decision device outputs and $N = 35$ noisy channel outputs. The test statistic $T(\Delta')$ and the number of equalisation errors in every observation record are plotted in Fig. 3.6. To see the effect of an increase in N, the same plot is repeated in Fig. 3.7 for $N = 55$. The test threshold for the case of $N = 55$ was $\eta' = 50.9$ corresponding to the same significance level as before. The increase in η' is due to the increase in $N - P'$, which is the number of degrees of freedom of the test statistic. For $N = 55$, observation records contain 79 decision device outputs and 55 noisy channel outputs. As is evident from Fig. 3.7, the increase in N has improved the

---

[7]CMA is a special case of the Godard algorithm which is defined in terms of a family of cost functions.

(a)

(b)
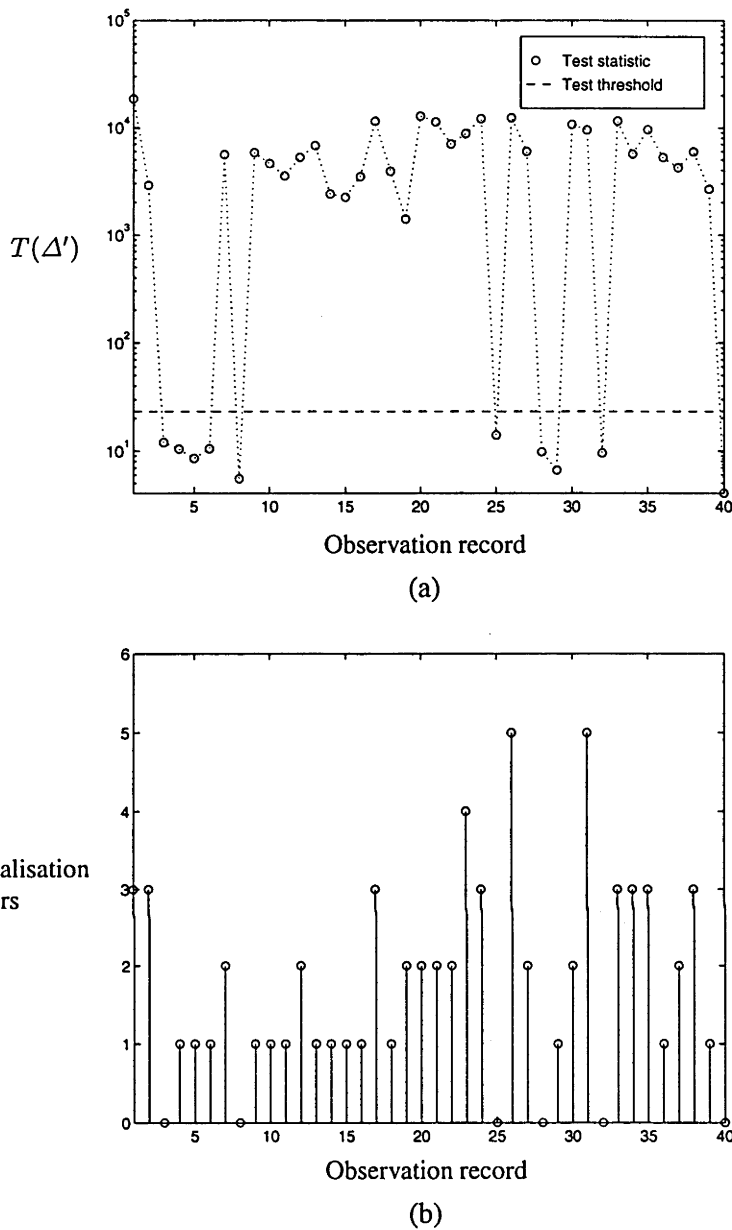
Figure 3.6 Test statistic for different observation records of A' and r'. Test parameters are $N = 35$, $P' = 25$ and $\eta' = 23.2$ (i.e. $\alpha = 0.01$). (a) Plot of test statistic $T(\Delta')$, (b) number of equalisation errors in observation records.

error detection performance of the test. Note that the test is capable of detecting the occurrence of a single error in a decision device output sequence of length 79. This is indicative of the high detection performance of the test for small sample sizes. □

Example 3.2 In this example we consider the detection of equalisation errors in a DFE structure arising from error propagation. We assume that the channel input is an i.i.d. binary sequence with $M = 2$ and the channel under consideration is

$$H(z) = 0.1 + 0.06z^{-1} + 0.13z^{-2} + 0.25z^{-3} + 0.5z^{-4} + z^{-5} + 2z^{-6} + 4z^{-7} \qquad (3.84)$$

(a)

(b)
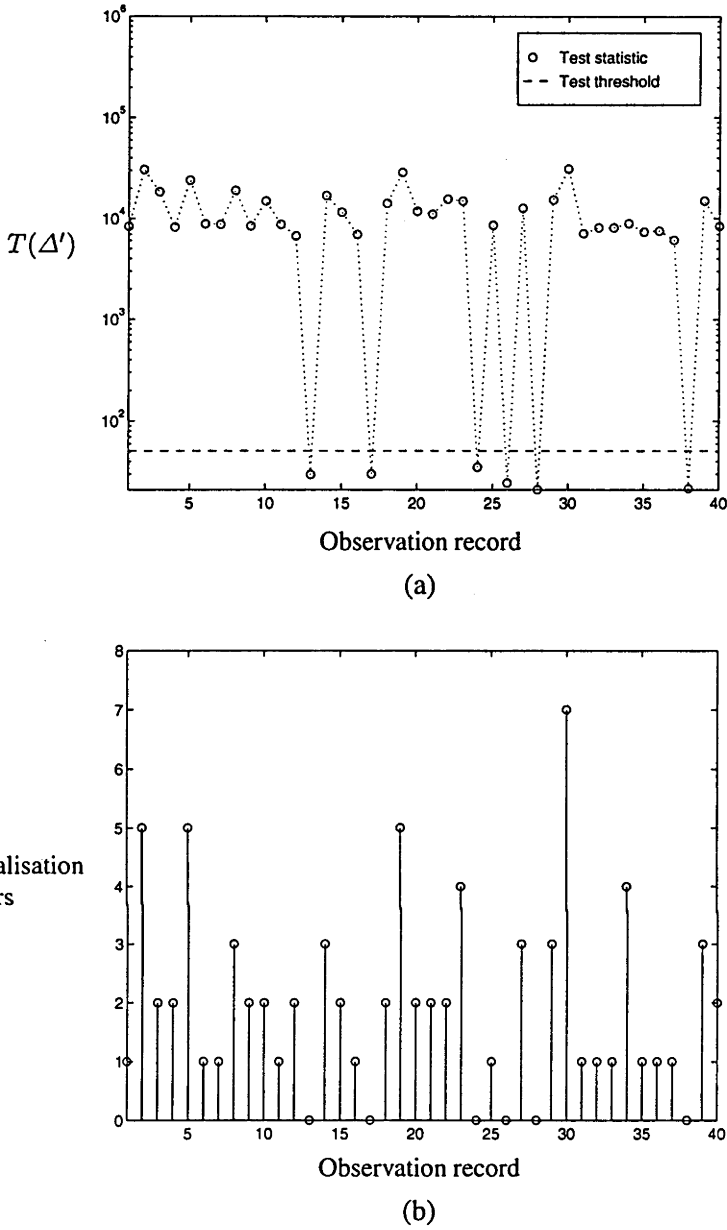
Figure 3.7 Test statistic for different observation records of A' and r'. Test parameters are $N = 55$, $P' = 25$ and $\eta' = 50.9$ (i.e. $\alpha = 0.01$). (a) Plot of test statistic $T(\Delta')$, (b) number of equalisation errors in observation records.

which has a finite-duration impulse response with exponential growth. Such channels are known to have long average recovery time from error propagation [37]. The channel noise $\{n(k)\}$ is the same as in the previous example. The equaliser parameters are perfectly tuned to the channel parameters, i.e. $D(z) = H(z)$ in Fig. 3.2(b). In this example, the test parameters were set to $N = 25$, $P' = 15$ and $\Delta' = \Delta = 0$. The test statistic $T(\Delta')$ and the number of equalisation errors for 40 successive observation records, each containing $N + P' - 1 = 39$ decision device outputs and $N = 25$ noisy channel outputs, are shown in Fig. 3.8. The test threshold was set to $\eta' = 25.2$ corresponding to a significance level of $\alpha = 0.005$. Note that as soon as the error propagation
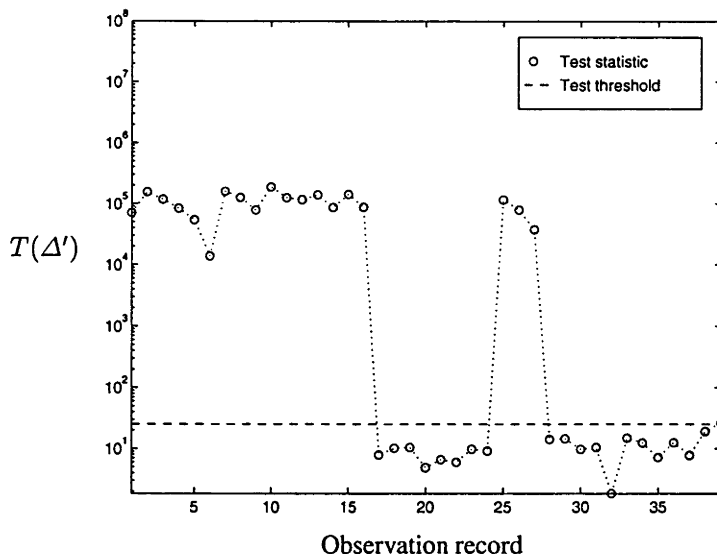
(a)
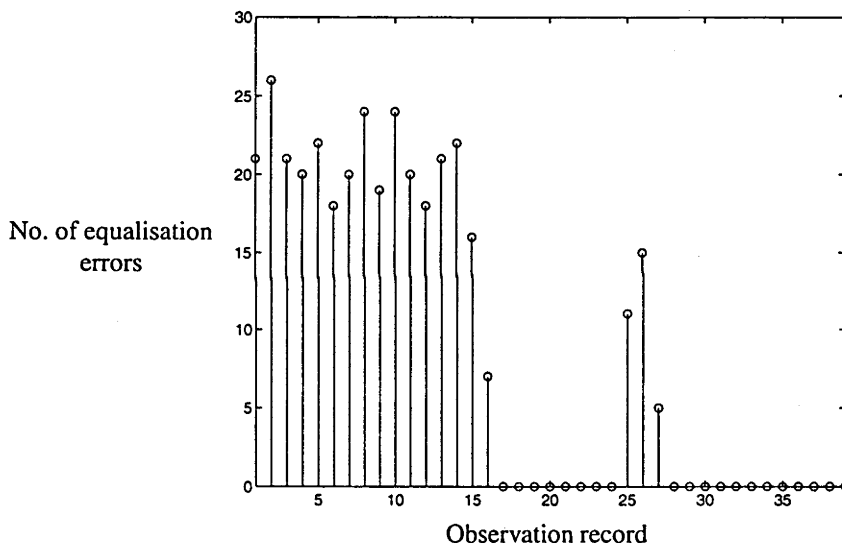
(b)

Figure 3.8 Testing for error propagation in DFE. Test parameters are $N = 25$, $P' = 15$ and $\eta' = 25.2$ (i.e. $\alpha = 0.005$). (a) Plot of test statistic $T(\Delta')$, (b) number of equalisation errors in observation records.

ceases the test statistic drops below the test threshold. Occasional errors due to excessive channel noise, as well as ensuing error propagation seem to be detected perfectly well. □

Example 3.3 To get some feel for the relative performance of the error detection test in Eq. (3.78), we now consider the channel-equaliser set-up used in Examples 2.3 and 2.4 in the previous chapter with

$$H(z) = -0.6 + 1.1z^{-1} - 1.9z^{-2} - 0.8z^{-3} \tag{3.85}$$

α		0.005	0.010	0.025	0.050	0.100
Average	$N = 30$	0.8350	0.8365	0.8380	0.8400	0.8415
P_D	$N = 40$	0.9025	0.9035	0.9075	0.9120	0.9205

Table 3.2 Average P_D of $T(\Delta')$ under H_1.

and the closed-eye equaliser parameters

$$\Theta(z) = 0.3041 - 0.0331z^{-1} + 0.0341z^{-2}. \tag{3.86}$$

The channel noise is taken to be a zero-mean white Gaussian process with variance $\sigma_n^2 = 0.1$ and, unlike in Examples 2.3 and 2.4, is *not* passed through a hard limiter.

The assumed channel length and equalisation delay are $P' = 10$ and $\Delta' = 3$, respectively. Table 3.2 lists the average probability of detection of the test in Eq. (3.78) for $N = 30$ and $N = 40$ after 2000 trials. A comparison of Table 3.2 with Fig. 2.5 reveals the superior performance of the test in Eq. (3.78) for short sample sizes. Note that the number of decision device output observations is only $N + P' - 1 = 39$ for $N = 30$ and 49 for $N = 40$. To get comparable results for the convergence test of Chapter 2, the sample size has to be over 1500. □

3.7 Discussion and Conclusions

We have presented UMP tests to detect the occurrence of equalisation errors in communication systems by using only the observations available at the equaliser. The major advantages of the tests compared to the convergence test of Chapter 2 are their high probability of detection for rather short observation durations and their wide applicability. No *a priori* knowledge of the channel input autocorrelation is required in the application of the tests. Additionally, no restrictions are imposed on the channel input constellation (so long as it is a finite set), the equaliser structure and the channel noise (apart from Gaussianity). Indeed the test criterion described in Property 3.1 can be used for any equaliser structure including the DFE as illustrated in Example 3.2. In this sense, the tests are not based on a conjecture as in [38], but on a well-proved criterion. Only disadvantage of the tests is their high computational complexity relative to the convergence test of Chapter 2. The high computational cost arising from g-inverse computations can be alleviated to some degree by using iterative methods for the computation of the test statistics as discussed in Appendix 3.A.2. In this chapter, we have also kept the promise made in Section 2.9 to devise a better method for estimating the equalisation delay and showed its feasibility in a simulation example.

To recapitulate, the applicability of the tests is subject to the following conditions:

- The channel input sequence is drawn from a finite set and has all its subsequences occur with nonzero probability.

- The channel is linear and time-invariant during the test interval and can be approximated by an FIR system.

- The equaliser is followed by, or incorporates, a discontinuous decision device.

The tests are constructed based on the method of least squares. Since the channel noise is assumed to be coloured, it is possible, and may even prove beneficial, to consider the use of *generalised least squares* [39] which solves the LS problem

$$\min_{v}(r - Av)^T \Sigma^{-1}(r - Av) \tag{3.87}$$

Although we have not explicitly mentioned the use of the generalised least squares so far, its use in the construction of the tests involves only trivial modifications.

The distribution of the test statistics has been shown to be chi-square with d degrees of freedom where $d = N - P'$. For large d, it may be cumbersome to work out the significance level of the tests. In such cases, the chi-square distribution can be approximated by a Gaussian distribution. Two transformations are in common use [35]:

i. Square-root transformation (Fisher's result), which was used in Section 3.4.3:

$$\sqrt{2\chi_d^2} \overset{\text{approx}}{\sim} \mathcal{N}\left(\sqrt{2d-1}, 1\right), \quad d > 30$$

ii. Cube-root transformation (Wilson and Hilferty's result):

$$\left(\frac{\chi_d^2}{d}\right)^{1/3} \overset{\text{approx}}{\sim} \mathcal{N}\left(1 - \frac{2}{9d}, \frac{2}{9d}\right), \quad d > 30.$$

Note that ii. is a more accurate approximation although it involves more computation in applications.

3.A   Computation of Symmetric Reflexive g-Inverses

3.A.1   Computation of a Symmetric Reflexive g-Inverse of LS Covariance Matrix Using the SVD

A g-inverse of the LS covariance matrix $K = P\Sigma P$ can be obtained from the singular value decomposition (SVD) of K

$$P\Sigma P = UDV^T \tag{3.88}$$

where U and V are orthogonal unitary matrices, which are identical since K is symmetric, and D is a diagonal matrix of the singular values of K sorted in descending order (i.e. $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{N-P}$)

$$D = \begin{bmatrix} \begin{array}{ccc|c} \sigma_1 & & 0 & \\ & \ddots & & 0 \\ 0 & & \sigma_{N-P} & \\ \hline & 0 & & 0 \end{array} \end{bmatrix}. \tag{3.89}$$

Note that $\sigma_{N-P+1} = \sigma_{N-P+2} \cdots = \sigma_N = 0$ since K has rank $N - P$ (see Theorem 3.1). A *g-inverse* of K, denoted by K^-, is defined as any matrix satisfying the equality $KK^-K = K$, which can be obtained from the SVD of K as follows [31]

$$K^- = U \underbrace{\begin{bmatrix} \begin{array}{ccc|c} 1/\sigma_1 & & 0 & \\ & \ddots & & X \\ 0 & & 1/\sigma_{N-P} & \\ \hline & Y & & Z \end{array} \end{bmatrix}}_{D^-} U^T \tag{3.90}$$

where X, Y and Z are arbitrary matrices of appropriate dimension. A full-rank g-inverse of K can be formed by choosing the matrices X, Y and Z in such a way that D^- has full rank. One such choice would be to let X and Y be zero matrices, and Z be a diagonal matrix with strictly positive diagonal entries. In this case, K^- can be decomposed as $K^- = CC^T$ where $C = U(D^-)^{1/2}$. Then, $\mu^T K^- \mu = \mu^T CC^T \mu > 0$ if $\mu \neq 0$ since C has full rank. Thus, the resulting K^- will also be *positive definite*.

A *symmetric reflexive g-inverse* of K, denoted by K_r^-, is defined through the following relations: $KK_r^-K = K$, $K_r^-KK_r^- = K_r^-$ and $\left(K_r^-\right)^T = K_r^-$. A symmetric reflexive

g-inverse of K can be shown to have the following form

$$
K_r^- = U \underbrace{\left[\begin{array}{ccc|c}
1/\sigma_1 & & 0 & \\
& \ddots & & X^T \\
0 & & 1/\sigma_{N-P} & \\
\hline
& X & & Z
\end{array} \right]}_{D_r^-} U^T
\tag{3.91}
$$

where X is an arbitrary $P \times (N - P)$ matrix and Z is a $P \times P$ symmetric matrix. Notice that the equality $KK_r^-K = K$ is satisfied for any choice of Z independent of X. The symmetry condition confines the choice of Z to symmetric matrices. The condition $K_r^- KK_r^- = K_r^-$ renders Z dependent on X. Thus, as we will show below, once X is fixed, Z is no longer an arbitrary matrix. The equality $K_r^- KK_r^- = K_r^-$ is in fact equivalent to $D_r^- DD_r^- = D_r^-$, which implies

$$
\left[\begin{array}{cccc|c}
\sigma_1 x_1 & \sigma_2 x_2 & \cdots & \sigma_{N-P} x_{N-P} & 0
\end{array} \right]
\left[\begin{array}{c}
x_1^T \\
x_2^T \\
\vdots \\
x_{N-P}^T \\
\hline
Z
\end{array} \right] = Z
\tag{3.92}
$$

where x_i denotes the ith column of X. From Eq. (3.92) we obtain the following relationship between X and Z

$$
Z = \sum_{i=1}^{N-P} \sigma_i x_i x_i^T.
\tag{3.93}
$$

While Eq. (3.91) is a general expression for the symmetric reflexive g-inverse, it can be simplified significantly by setting the arbitrary X to zero. In this case, K_r^- becomes the *Moore-Penrose inverse*, denoted by $K^\dagger$, which satisfies $KK^\dagger K = K$, $(KK^\dagger)^T = KK^\dagger$, $K^\dagger KK^\dagger = K^\dagger$ and $(K^\dagger K)^T = K^\dagger K$. Note that the Moore-Penrose inverse of K is also a symmetric reflexive g-inverse of K. The converse is of course not true.

3.A.2 Recursive Computation of a Symmetric Reflexive g-Inverse of LS Covariance Matrix

The computation of a symmetric g-inverse of the LS covariance matrix $K = P\Sigma P$ involves firstly the computation of the projection $P = I - AA^\dagger$ and then the computation of $(P\Sigma P)_r^-$. Matrix inversion and the singular value decomposition are computationally expensive operations. In this section, we show how these operations can be avoided by using iterative methods. The

inverse of the *Gram matrix* $(A^T A)^{-1}$ is needed in the computation of P. Two iterative methods are possible to compute $(A^T A)^{-1}$ and $A A^\dagger$, respectively. The first one makes use of the *matrix inversion lemma* and the second one is an iterative method based on the Schultz method [40].

Let $X(k)$ be defined by

$$X(k) \triangleq \sum_{i=1}^{k} \widehat{u}(i)\widehat{u}^T(i) \qquad (3.94)$$

where $\widehat{u}(i)$ is the ith column of A^T. Then, the Gram matrix $A^T A$ is equal to $X(N)$. Noting that $X(i) = X(i-1) + \widehat{u}(i)\widehat{u}^T(i)$ and denoting $X^{-1}(i)$ by $W(i)$, we can employ the matrix inversion lemma to obtain the following recursion

$$W(i) = W(i-1) - \frac{W(i-1)\widehat{u}(i)\widehat{u}^T(i)W(i-1)}{1 + \widehat{u}^T(i)W(i-1)\widehat{u}(i)}. \qquad (3.95)$$

Starting with $W(0) = \kappa I$ where κ is a large positive number, after N iterations $W(N)$ approximately gives $(A^T A)^{-1}$. Thus, the projection P is given by $I - A W(N) A^T$.

The iterative method for computation of P can be summarised as follows [29]. If μ is a real number such that

$$0 < \mu < \frac{2}{\lambda_{\max}(A^T A)} \qquad (3.96)$$

where $\lambda_{\max}(\cdot)$ denotes the maximum eigenvalue, the sequence

$$Z(0) = \mu A A^T$$

$$Z(i+1) = 2Z(i) - Z^2(i), \quad i = 0, 1, \ldots$$

converges to $A A^\dagger$ as $i \to \infty$. The projection P is then approximately given by $I - Z(i)$ where i is a large number.

Once P is computed, the next step is to compute $(P \Sigma P)_r^-$. Since $K^\dagger$ is also a symmetric reflexive g-inverse, the following iteration can be used to compute K_r^- [29]. If $0 < \mu < 2/\lambda_{\max}(K^T K)$, the sequence

$$Y(0) = \mu K^T$$

$$Y(i+1) = Y(i)(2I - K Y(i)), \quad i = 0, 1, \ldots$$

converges to $K^\dagger$ as $i \to \infty$.

CHAPTER 4

Least Squares Approach to Blind Channel Equalisation

4.1 Introduction

 lind channel equalisation is a signal processing operation to mitigate intersymbol interference (ISI) introduced by a communication channel without resorting to an explicit knowledge of the channel input. On-line memoryless-cost-function blind equalisation algorithms proposed so far have been shown to fall short of their purported objective of error-free recovery of the channel input sequence under practical conditions [13, 41]. The most frequently quoted reason for this failure is the *nonconvex* cost topology of these algorithms, which makes them prone to converging to local minima with inferior equalisation performance and staying at these closed-eye local minima for prolonged periods [1, 2].

It has been hypothesized that it may be possible to avoid convergence to closed-eye local minima by means of careful initialisation of the equaliser. One such initialisation strategy is the so-called centre-tap initialisation in which all equaliser tap coefficients except for the centre tap are set to zero. Despite its experimental success, there exits no rigorous proof of the effectiveness of the centre-tap initialisation; an argument based on the nonminimum phase channel inverse characteristics is often utilised to justify its success. The centre-tap initialisation may fail to lead to converge to open-eye minima if the channel input is not i.i.d. (independent identically distributed). Experimental evidence suggests that certain types of sequential correlation in the channel input sequence can render the initial channel-equaliser combination (with centre-tap initialisation) a stable closed-eye local minimum for some equalisation algorithms [14]. Hence, there is no guarantee that the centre-tap initialisation provides a fix for the existing on-line blind

equalisation algorithms under arbitrary channel input correlation. Further detrimental effects of the channel input correlation have been illustrated in [15]. The observed misbehaviour of the on-line blind equalisation algorithms brings about the need for the development of globally admissible blind equalisation algorithms that are not affected by the channel input correlation either in terms of the algorithmic complexity or convergence properties. The latter of course exclude the convergence rate since it is affected by the channel input correlation if a stochastic gradient-based implementation is adopted.

The central idea behind the new approach to blind channel equalisation draws on the philosophy of the constant modulus algorithm (CMA). A nonlinearly transformed version of the equaliser parameters is estimated by means of the method of least squares. The equaliser parameters are shown to be extractable from the transformed parameters up to a scaling factor. Unlike CMA, the resultant algorithm is globally admissible for channels whose inverse can be approximated by finite-duration impulse response (FIR) systems, and insensitive to the channel input correlation provided that all finite-length channel input subsequences occur with nonzero probability. In addition, the algorithm is "parsimonious" in its use of the channel output observations. Off-line (batch), on-line and recursive implementations of the algorithm are presented. The computational complexity of the off-line implementation is alleviated, to some extent, by the requirement of short channel output observations to arrive at the equaliser parameters. In cases where the computational complexity is of primary concern, on-line or recursive implementations can be considered as a trade-off between the computational cost and the speed of convergence.

The chapter starts with an illustration of the relationship between the CMA stationary points and the channel input correlation in order to motivate the need for an alternative approach to blind equalisation. Section 4.3 formulates the problem of channel equalisation assuming a constant modulus constellation for the channel input sequence. A least squares (LS) solution to the channel equalisation problem is presented in Section 4.4. A method for robust extraction of the equaliser parameters is proposed based on the singular value decomposition (SVD). The cases of real and complex channels are treated separately. In Section 4.5 computationally efficient on-line and recursive implementations of the algorithm are presented. Section 4.6 addresses the consequences of overparameterisation of the equaliser and proposes a way to come by the equaliser parameters from a solution to the resulting rank deficient LS problem. The iterative scheme in Appendix 3.A and a modified version of the recursive least squares (RLS) algorithm are also presented. An extension to pulse amplitude modulation (PAM) constellations is discussed in Section 4.7. The channel noise and its effects on the algorithm are considered in Section 4.8. Extensive simulation studies of the algorithm implementations are presented in Section 4.9. The chapter concludes with
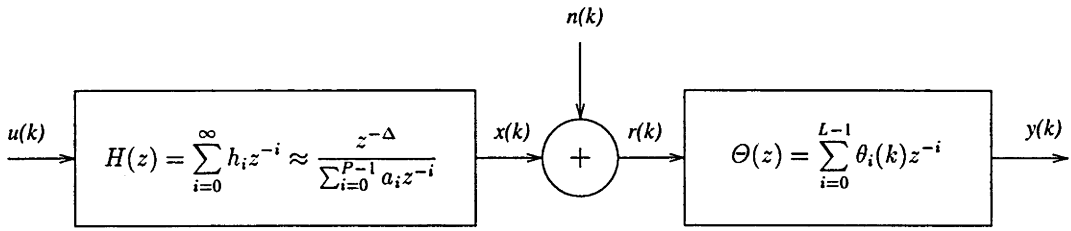
Figure 4.1 Channel and equaliser models.

a brief discussion of the results in Section 4.10.

4.2 Ill-convergence of CMA

In this section we will shed some light on the reasons for multimodality exhibited by the mean CMA cost function. The exposition should also provide general insight into the ill-convergence phenomenon observed in other on-line blind equalisation algorithms. At the end of the section we will show the relationship between the channel input correlation and the CMA stationary points by way of a simple example.

Consider the channel equalisation problem depicted in Fig. 4.1 where $\{u(k)\}$ is a zero-mean (not necessarily white) sequence of binary numbers taking on values ± 1. The channel $H(z)$ is a linear time-invariant (LTI) system and has a stable, causal and possibly nonminimum phase transfer function

$$H(z) = \sum_{i=0}^{\infty} h_i z^{-i} \tag{4.1}$$

where the channel impulse response $\{h_i\}$ will be assumed to be real for the moment. The channel noise $\{n(k)\}$ will be supposed to be negligible until Section 4.8. If the channel has an inverse that can be approximated by a (noncausal) FIR system of length P

$$H^{-1}(z) \approx z^{\Delta} \sum_{i=0}^{P-1} a_i z^{-i} \tag{4.2}$$

where Δ is a nonnegative integer representing the minimum equalisation delay, an adaptive linear tapped-delay line with transfer function

$$\Theta(z) = \sum_{i=0}^{L-1} \theta_i(k) z^{-i} \tag{4.3}$$

can be used to equalise the channel with a proper choice of L (i.e. $L \geq P$). Assuming that the tap coefficients $\theta_i(k)$ are real-valued and the channel noise $\{n(k)\}$ is negligible, the equaliser output

at time k can be written as the inner product

$$y(k) = \theta^T(k)r(k) \tag{4.4}$$

where $\theta(k)$ is the $L \times 1$ equaliser parameter vector at time k

$$\theta(k) = [\theta_0(k), \theta_1(k), \dots, \theta_{L-1}(k)]^T \tag{4.5}$$

and $r(k)$ is the $L \times 1$ regressor vector

$$r(k) = [r(k), r(k-1), \dots, r(k-L+1)]^T. \tag{4.6}$$

Central to CMA is the concept of restoring the expected modulus of the equaliser output to that of the channel input. Under the assumption that all finite-length subsequences of $\{u(k)\}$ occur with nonzero probability, restoral of the modulus is a necessary and sufficient condition for the equalisation objective to be achieved. This is a consequence of Property 3.1 in Chapter 3. For binary channel inputs the mean cost function of CMA is given by

$$\mathcal{J}_{\text{CMA}}(y(k)) = \frac{1}{4}E\left\{\left(1 - y^2(k)\right)^2\right\}. \tag{4.7}$$

Stripping off the expectation operator above and taking the gradient of the resulting expression with respect to $\theta(k)$, we obtain the *instantaneous* gradient

$$\varphi(y(k)) = \left(y^2(k) - 1\right)y(k)r(k). \tag{4.8}$$

On negation and multiplication by a small stepsize μ, Eq. (4.8) produces the update term in the CMA recursion

$$\theta(k+1) = \theta(k) - \mu\varphi(y(k)). \tag{4.9}$$

According to averaging theory, the stationary points of the instantaneous cost function are identical to the stationary points of the mean cost function $\mathcal{J}_{\text{CMA}}(y(k))$, provided that the stepsize μ in Eq. (4.9) is very small [42]. Then, any vector $\breve{\theta} = [\breve{\theta}_0, \breve{\theta}_1, \dots, \breve{\theta}_{L-1}]^T$ that satisfies $E\{\varphi(y(k))\}|_{\theta(k) = \breve{\theta}} = 0$ or

$$E\left\{\left(\left(\theta^T(k)r(k)\right)^2 - 1\right)\left(\theta^T(k)r(k)\right)r(k)\right\}\Big|_{\theta(k) = \breve{\theta}} = 0 \tag{4.10}$$

is a *stationary point* of CMA. We can rewrite Eq. (4.10) as a system of L nonlinear equations

$$
\begin{aligned}
\mathcal{P}(0) &= 0 \\
\mathcal{P}(1) &= 0 \\
&\vdots \\
\mathcal{P}(L-1) &= 0
\end{aligned}
\tag{4.11}
$$

where, assuming that $\{r(k)\}$ is fourth-order weakly stationary, the polynomials $\mathcal{P}(i)$ are defined by

$$
\mathcal{P}(i) \triangleq \sum_{p=0}^{L-1} m_{4,r}(i-p, i-p, i-p)\theta_p^3(k) + 3 \sum_{p=0}^{L-1} \sum_{\substack{q=0 \\ q \neq p}}^{L-1} m_{4,r}(i-p, i-q, i-q)\theta_p(k)\theta_q^2(k)
$$

$$
+ 6 \sum_{p=0}^{L-1} \sum_{q>p}^{L-1} \sum_{s>q}^{L-1} m_{4,s}(i-p, i-q, i-s)\theta_p(k)\theta_q(k)\theta_s(k)
$$

$$
- \sum_{p=0}^{L-1} R_r(i-p)\theta_p(k), \quad i = 0, 1, \ldots, L-1
\tag{4.12}
$$

with

$$
m_{4,r}(\tau_1, \tau_2, \tau_3) = E\{r(k)r(k+\tau_1)r(k+\tau_2)r(k+\tau_3)\}
\tag{4.13}
$$

denoting the fourth-order moment of $\{r(k)\}$ and

$$
R_r(\tau) = E\{r(k)r(k+\tau)\}
\tag{4.14}
$$

the second-order moment (autocorrelation) of $\{r(k)\}$. Even though the multimodality of the mean CMA cost function and the influence of the channel input statistics on the location and stability of stationary points can be discerned from Eq. (4.11), we will illustrate these observations in an example.

Example 4.1 Let us suppose that the channel is a single-pole system of the form $H(z) = 1/(1 - 0.6z^{-1})$ and is followed by a two-tap equaliser ($L = 2$) which is capable of equalising the channel perfectly. The channel input sequence is binary, i.e. $u(k) \in \{-1, 1\}$ $\forall k$. According to Eq. (4.12) the polynomials in Eq. (4.11) are given by

$$
\mathcal{P}(i) = m_{4,r}(i, i, i)\theta_0^3(k) + m_{4,r}(i-1, i-1, i-1)\theta_1^3(k) + 3m_{4,r}(i, i-1, i-1)\theta_0(k)\theta_1^2(k)
$$

$$
+ 3m_{4,r}(i-1, i, i)\theta_1(k)\theta_0^2(k) - R_r(i)\theta_0(k) - R_r(i-1)\theta_1(k), \quad i = 0, 1.
\tag{4.15}
$$

The channel input sequence will be considered to be either i.i.d. or correlated with autocorrelation

(a)

(b)
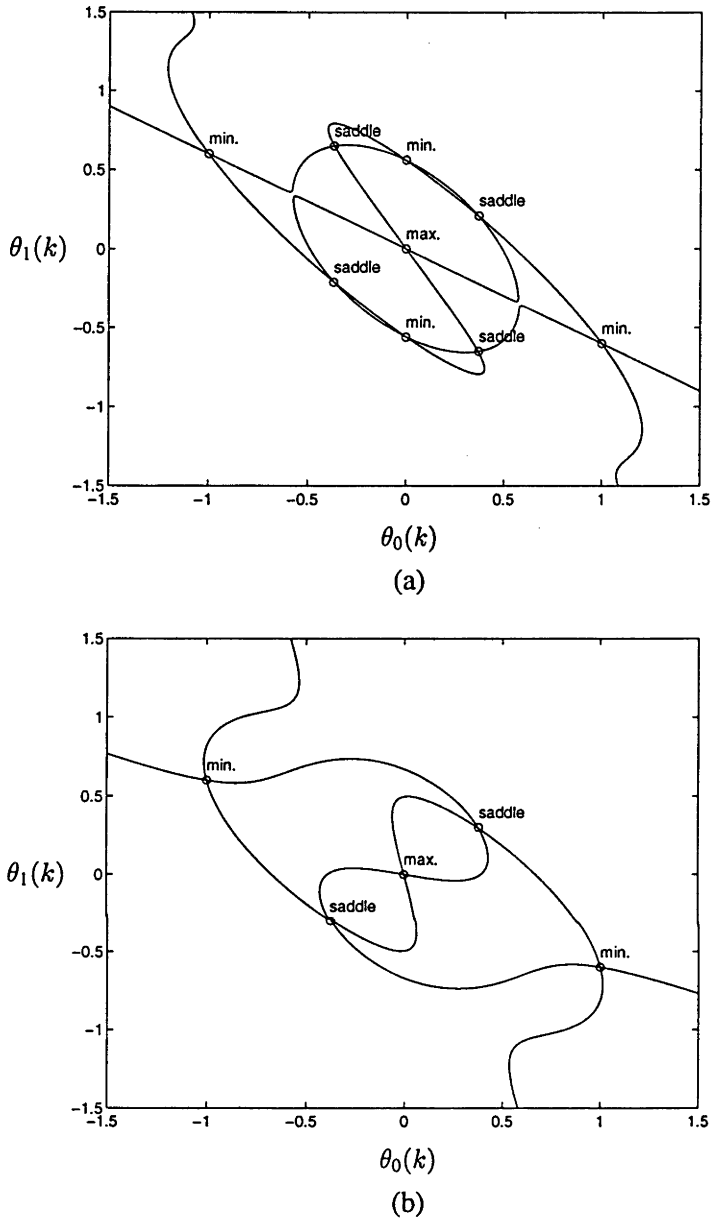
Figure 4.2 Stationary points of CMA (a) for i.i.d. input, and (b) for correlated input.

values $R_u(\pm 1) = -0.5$, $R_u(\pm 2) = 0.25$ and $R_u(\tau) = 0$ for $|\tau| > 2$. Forming Eq. (4.11) and solving the resulting system of nonlinear equations for $\theta(k)$ we get nine stationary points for the i.i.d. channel input and five stationary points for the correlated channel input. The resulting algebraic curves (cubics) $\mathcal{P}(0) = 0$ and $\mathcal{P}(1) = 0$ in the $(\theta_0(k), \theta_1(k))$ space are shown in Figs. 4.2(a) and (b) for the i.i.d. and correlated channel inputs, respectively. The stability of the stationary points is determined from the Hessian of the mean cost function $\mathcal{J}_{\text{CMA}}(y(k))$ which is defined by

$$\mathcal{H}_{\breve{\theta}} \triangleq E\left\{ \frac{\partial \varphi(y(k))}{\partial \theta(k)} \right\}\bigg|_{\theta(k) \,=\, \breve{\theta}} \qquad (4.16a)$$

$$= E\left\{\left(3\left(\boldsymbol{\theta}^T(k)\boldsymbol{r}(k)\right)^2 - 1\right)\boldsymbol{r}(k)\boldsymbol{r}^T(k)\right\}\Big|_{\boldsymbol{\theta}(k) = \breve{\boldsymbol{\theta}}}. \qquad (4.16b)$$

If $\mathcal{H}_{\breve{\theta}} > 0$ (i.e. all eigenvalues of $\mathcal{H}_{\breve{\theta}}$ are strictly positive), then $\breve{\theta}$ is a stable local minimum. For the i.i.d. channel input there are four *stable* stationary points (minima), two of which are closed-eye local minima, whereas for the correlated channel input there are only two *stable* stationary points both of which open the eye. The relationship between the number of stationary points is governed by the Morse theory. For $L = 2$ the Morse theory produces the following relations:

i. Number of maxima ≥ 0,

ii. Number of saddle points $\geq$ number of maxima,

iii. Number of minima = number of saddle points $-$ number of maxima $+$ 1.

In this example the mean cost function has only one maximum located at the origin. Relation **iii.** above therefore stipulates that the number of minima should be equal to the number saddle points. Based on the symmetry of the cost function, the number of stable minima can be shown to be even. Since two cubics can have at most nine simultaneous solutions, for $L = 2$ there exist only two possible scenarios; viz. (i) nine stationary points with four stable minima, and (ii) five stationary points with two stable minima. As illustrated in Fig. 4.2, the channel input correlation plays a central role in determining which scenario is to prevail in the convergence of CMA. The effect of the channel input correlation on the solutions of Eq. (4.11) stems from the indirect influence of $R_u(\tau)$ on $R_r(\tau)$ and $m_{4,r}(\tau_1, \tau_2, \tau_3)$, both of which determine the coefficients of polynomials in Eq. (4.11). □

Although the central idea behind CMA has intuitive appeal, it is subject to the caveat that the mean CMA cost function possesses a multimodal topology in the equaliser parameter space. This in turn makes the recursive adaptation algorithm in Eq. (4.9) prone to getting stuck with local minima which fall short of achieving desirable equalisation performance. One of the reasons for this observed misbehaviour is the averaging operation that results from the CMA recursion. The averaging operation seems to create spurious stationary points for which the mean update term in Eq. (4.9) is zero. An alternative method would be to seek for the equaliser parameters by forcing the instantaneous, rather than expected, modulus of $y(k)$ to be constant. It should be noted that in a deterministic setting the channel input correlation does not have any effect on the cost surface topology. This approach will be pursued further in the rest of the chapter.

4.3   Problem Formulation for Constant Modulus Constellations

Before we make any attempt at formulating our new approach to blind channel equalisation for constant modulus inputs, we need to clarify a few points related to the equalisation objective. In this chapter, unless otherwise stated, our objective will be to realise the equality

$$y(k) = \Upsilon e^{j\phi(k)} u(k - \Delta) \quad \forall k \tag{4.17}$$

where Υ is the equalisation gain, $\phi(k)$ is a generally time-varying phase shift resulting from a frequency offset in the demodulation process, and Δ is the equalisation delay. The particular value of Υ is immaterial as long as it is nonzero and constant, since all it does is to move the channel input constellation points in radial directions. Any time variation in Δ amounts to a failure of the equalisation objective.

In fact, Eq. (4.17) is not the most strict equalisation objective that can be imposed because it allows Υ to be different to the unity and $\phi(k)$ to be nonzero and time-varying. The cost criterion of CMA tolerates a nonzero time-varying $\phi(k)$. In blind equalisation it is often desirable to remove ISI before or during carrier phase recovery, which creates an interest in equalisation algorithms that are insensitive to phase rotations (however fast they are) such as CMA. The new algorithm has the same insensitivity to phase rotations in $\phi(k)$. Carrier phase tracking can be achieved subsequently using a decision-directed estimator [10]. The ambiguity involved in phase detection can be alleviated by using noncoherent modulation schemes such as differential phase shift keying (DPSK) or noncoherent frequency shift keying (FSK) which do not rely on absolute phase reference. Information about transmitted symbols is carried in the phase difference between successive symbols in the case of DPSK and the frequency of the received symbol in the case of FSK.

The channel input sequence will be assumed to take on values from a symbol constellation with M members, i.e. $u(k) \in \mathbb{S} = \{s_1, s_2, \ldots, s_M\}$, where the modulus of the symbols $|s_i| \triangleq \sqrt{s_i s_i^*}$ is constant. Note that the notation s_i^* denotes the complex conjugate of s_i.

4.4   Least Squares Solution

4.4.1   Closed-Form Derivation of the Solution

As we have seen in Section 4.2, the CMA cost function has multiple stable stationary points that may lead to convergence to undesirable parameter settings. We now propose an alternative deterministic approach using the method of least squares by exploiting the constant modulus

property of the channel input sequence. We will assume that the input sequence $\{u(k)\}$ has all of its finite-length subsequences occur with nonzero probability[1]. Then, it follows from Property 3.1 in Chapter 3 that the equalisation objective in Eq. (4.17) is achieved if $|y(k)|^2$ is constant for all k. With no loss of generality we will assume that the equalisation gain Υ normalises the modulus of $y(k)$ to unity. Thus, $|y(k)|^2 = 1 \; \forall k$ and Eq. (4.17) can be considered to be equivalent.

In the following discussion we will initially assume that

$$H^{-1}(z) = z^{\Delta} \sum_{i=0}^{P-1} a_i z^{-i} \tag{4.18}$$

(cf Eq. (4.2)) which necessarily implies that minimisation of $\mathcal{J}_{\text{CMA}}(y(k))$ with respect to $\theta(k)$ can be replaced by solving $\mathcal{J}_{\text{CMA}}(y(k)) = 0$ for $\theta(k)$. The assumption of strict equality in Eq. (4.18) will be relaxed later on. The number of equaliser parameters L will be assumed to be equal to P. The consideration of the case where $L > P$ will be deferred until Section 4.6.

In practice, only a finite number of equaliser output observations can be used to compute the equaliser parameters. For a sequence of equaliser output observations of length N, $\{y(k), y(k-1), \ldots, y(k-N+1)\}$, a deterministic counterpart of Eq. (4.7) may be written

$$\begin{bmatrix} |y(k)|^2 \\ |y(k-1)|^2 \\ \vdots \\ |y(k-N+1)|^2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}. \tag{4.19}$$

Noting that the equaliser output is defined as the complex inner product of the equaliser parameters and the channel output vector

$$y(k) = \theta^H(k) r(k) \tag{4.20}$$

where ^{H} denotes the Hermitian (transpose and conjugate) operator, $|y(k-i)|^2$ can be written explicitly for arbitrary i as

$$|y(k-i)|^2 = \sum_{l=0}^{L-1} \sum_{m=0}^{L-1} \theta_l^* \theta_m r(k-l-i) r^*(k-m-i) \tag{4.21}$$

where we have dropped the time index k from the equaliser parameters to signify the off-line nature of Eq. (4.19). In the off-line context, the equaliser parameters will be denoted by

$$\theta = [\theta_0, \theta_1, \ldots, \theta_{L-1}]^T. \tag{4.22}$$

---

[1]This is a more general assumption than independence and imposes a mild restriction on correlated inputs.

An on-line implementation of Eq. (4.19) will be considered later in Section 4.5.1. Using Eq. (4.21), Eq. (4.19) can be written compactly as

$$A\psi = 1 \qquad (4.23)$$

where $A = [a_{ij}]$ is an $N \times L^2$ matrix ($N \geq L^2$) with entries

$$a_{ij} = r(k - i + 1 - \text{rem}(j - 1, L))r^*(k - i + 1 - \text{mod}(j - 1, L)),$$

$$i = 1, 2, \ldots, N, \quad j = 1, 2, \ldots, L^2, \qquad (4.24)$$

the $L^2 \times 1$ vector of unknown transformed equaliser parameters ψ is defined, using the *Kronecker product* (denoted $\otimes$), by

$$\psi = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_{L-1} \end{bmatrix}^* \otimes \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_{L-1} \end{bmatrix} \qquad (4.25)$$

and 1 is the $N \times 1$ vector of ones. In Eq. (4.24) $\text{mod}(a, b)$ denotes the integer part of a/b and $\text{rem}(a, b)$ is defined as $a - b\,\text{mod}(a, b)$. The matrix A can alternatively be written as a row partitioning

$$A = \begin{bmatrix} r^T(k) \otimes r^H(k) \\ r^T(k - 1) \otimes r^H(k - 1) \\ \vdots \\ r^T(k - N + 1) \otimes r^H(k - N + 1) \end{bmatrix}. \qquad (4.26)$$

Once Eq. (4.23) is solved for ψ, the equaliser parameters θ_i can easily be obtained from Eq. (4.25). Notice that if Eq. (4.23) is consistent, i.e. there exists a ψ satisfying the equality, which is certainly the case when there is no channel noise and the channel $H(z)$ is an all-pole system with an exact FIR inverse as in Eq. (4.18), it suffices to set the number of observations N equal to L^2 in Eq. (4.23) so that A is a square matrix. If the square matrix A has full rank, the vector ψ can be obtained uniquely as $\psi = A^{-1}1$. Note that the uniqueness of ψ does not imply that of θ because of Eq. (4.25). Nevertheless, the eye will be open with zero ISI for any θ resulting in ψ that satisfies Eq. (4.23) [13].

The identification of ψ, which is a nonlinear transformation of θ (see Eq. (4.25)), is a well-defined problem in a blind equalisation setting for constant modulus channel inputs, whereas the same cannot be said for the determination of θ using the CMA cost function or other on-line criteria for that matter. So long as A has full column rank, a unique solution for ψ exists. The LS solution for ψ is obtained from the minimisation of a convex (unimodal) cost function with

no risk of spurious local minima. In this respect, the philosophy behind Eq. (4.23) is novel and its usefulness will be demonstrated in the rest of the chapter. As a parenthetical remark, note that in the formulation of Eq. (4.23) no assumption is made about the channel input correlation apart from a full-column-rank A.

Example 4.2 Consider the simple complex channel $H(z) = j/(1 - 0.6z^{-1})$ with 4-DPSK inputs, i.e. $u(k) \in \mathbb{S} = \{1, -j, -1, j\}$. If a 2-tap equaliser is used (i.e. $L = 2$) and N is set to $L^2 = 4$, the matrix A can be written explicitly as

$$A = \begin{bmatrix} r(k)r^*(k) & r(k)r^*(k-1) & r(k-1)r^*(k) & r(k-1)r^*(k-1) \\ r(k-1)r^*(k-1) & r(k-1)r^*(k-2) & r(k-2)r^*(k-1) & r(k-2)r^*(k-2) \\ r(k-2)r^*(k-2) & r(k-2)r^*(k-3) & r(k-3)r^*(k-2) & r(k-3)r^*(k-3) \\ r(k-3)r^*(k-3) & r(k-3)r^*(k-4) & r(k-4)r^*(k-3) & r(k-4)r^*(k-4) \end{bmatrix}.$$

(4.27)

Note that only five channel output observations $\{r(k), r(k-1), \ldots, r(k-4)\}$ are needed to form A. Supposing that we have made the following channel output observations $\{0.2106 - j0.2462, 0.3510 + j1.2563, 0.5850 + j0.4272, 0.9749 - j0.9547, -0.0418 - j1.5912\}$ for which A is a full-rank matrix, the solution of Eq. (4.23) can be written as

$$\psi = \underbrace{\begin{bmatrix} 0.1050 & -0.2354 - j0.3510 & -0.2354 + j0.3510 & 1.7015 \\ 1.7015 & 0.7420 + j0.5850 & 0.7420 - j0.5850 & 0.5246 \\ 0.5246 & 0.1625 + j0.9749 & 0.1625 - j0.9749 & 1.8620 \\ 1.8620 & 1.4784 + j1.5912 & 1.4784 - j1.5912 & 2.5337 \end{bmatrix}}_{A}^{-1} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

(4.28a)

$$= [1.00 \quad -0.60 \quad -0.60 \quad 0.36]^T.$$

(4.28b)

Using Eq. (4.25) to transform ψ back to θ we get $\theta = \pm[1, -0.6]^T$, for which the equalisation gain and the phase shift are $\Upsilon = 1$ and $\phi(k) = 0$ or π radians, respectively. □

If the channel has an ARMA (autoregressive moving average) model, an exact solution to Eq. (4.23) cannot be obtained using a finite-length equaliser. Assuming that the channel inverse can be approximated as in Eq. (4.2) with the equaliser length $L = P$, a reasonable approach to the channel equalisation problem will be to find a solution that minimises the squared Euclidean norm of errors

$$\mathcal{J}_{\text{LS}}(\psi) \triangleq (A\psi - 1)^H (A\psi - 1)$$

(4.29)

where the matrix A has usually more rows than columns (i.e. $N > L^2$). An LS estimate for ψ is

then given by

$$\widehat{\psi} = \arg\min_{\psi} \mathcal{J}_{\text{LS}}(\psi) \tag{4.30}$$

$$= \left(A^H A\right)^{-1} A^H \mathbf{1} \tag{4.31}$$

which is the solution to the *normal equations*

$$A^H A \widehat{\psi} = A^H \mathbf{1}. \tag{4.32}$$

After dividing through by N, Eq. (4.32) approaches the following expression as N tends to infinity

$$\underbrace{E\left\{(r^*(k) \otimes r(k))(r^*(k) \otimes r(k))^H\right\}}_{M} \psi_o = \underbrace{E\left\{r^*(k) \otimes r(k)\right\}}_{c} \tag{4.33}$$

which is analogous to the Wiener-Hopf equations. In Eq. (4.33) we assumed that the channel output sequence $\{r(k)\}$ is fourth-order weakly stationary and ergodic. The $L^2 \times L^2$ Toeplitz matrix M contains the channel output fourth-order moments and the $L^2 \times 1$ vector c is the vector of channel output autocorrelation. If M has full rank, which is guaranteed when $L = P$, the optimum solution ψ_o is unique. Thus, the LS solution of Eq. (4.23) asymptotically (as $N \to \infty$) takes the form $\psi_o = M^{-1} c$.

The use of Eq. (4.19) can result in a tremendous reduction in the number of channel output observations needed to determine the equaliser parameters. CMA, in general, requires a large number of iterations to converge to a minimum, be it open-eye or closed-eye. Eq. (4.23) is, on the other hand, comparatively parsimonious in that it uses only a small number of channel output observations of the order of L^2 to obtain the equaliser parameters without the problem of ill-convergence (see Example 4.2). The cost function $\mathcal{J}_{\text{LS}}(\psi)$ in Eq. (4.29) has convex topology if A has full column rank. This guarantees a unique $\widehat{\psi}$ if the method of least squares is applied. Since the channel $H(z)$ cannot always be modelled as an all-pole system, we will often resort to the LS solution to compute the equaliser parameters.

4.4.2   Robust Extraction of Equaliser Parameters

The LS estimates $\widehat{\psi}$ for the transformed equaliser parameters do not usually have the Kronecker product structure in Eq. (4.25) due to estimation errors. A robust method that utilises all the information in $\widehat{\psi}$ to obtain the equaliser parameters can be devised using the singular value decomposition (SVD). As in [43], we first form the following $L \times L$ *Hermitian* matrix with rank

one

$$T = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_{L-1} \end{bmatrix}^* \begin{bmatrix} \theta_0 & \theta_1 & \cdots & \theta_{L-1} \end{bmatrix}. \tag{4.34}$$

Replacing the elements of T with the corresponding entries in $\widehat{\psi} = [\widehat{\psi}_1, \widehat{\psi}_2, \ldots, \widehat{\psi}_{L^2}]^T$, we obtain

$$\widehat{T} = \begin{bmatrix} \widehat{\psi}_1 & \widehat{\psi}_2 & \cdots & \widehat{\psi}_L \\ \widehat{\psi}_{L+1} & \widehat{\psi}_{L+2} & \cdots & \widehat{\psi}_{2L} \\ \vdots & \vdots & & \vdots \\ \widehat{\psi}_{L(L-1)+1} & \widehat{\psi}_{L(L-1)+2} & \cdots & \widehat{\psi}_{L^2} \end{bmatrix}. \tag{4.35}$$

In order to gain insight into the symmetry properties of $\widehat{T}$, let us consider the LS projection of $\mathbf{1}$ onto signal subspace

$$A\widehat{\psi} = A \left(A^H A \right)^{-1} A^H \mathbf{1} \tag{4.36}$$

where the projection $A \left(A^H A \right)^{-1} A^H$ is a real matrix, implying that $A\widehat{\psi}$ is a real vector. As can be seen from the definition of A in Eq. (4.26), some columns of A are real-valued, while others are in complex conjugate pairs. Those entries of $\widehat{\psi}$ that multiply the real columns of A should be real, and the ones that multiply complex conjugate pairs in A should also be complex conjugate pairs in order for the vector inner products producing $A\widehat{\psi}$ to be real-valued. Thus, the complex conjugate pair relations in T are preserved in $\widehat{T}$, resulting in the identity $\widehat{T}^H = \widehat{T}$. This in turn implies that $\widehat{T}$ is Hermitian.

Although T has rank one, $\widehat{T}$ is in general a full-rank matrix due to estimation errors in $\widehat{\psi}$. Using the SVD of $\widehat{T}$

$$\widehat{T} = \sum_{i=1}^{L} \sigma_i v_i v_i^H \tag{4.37}$$

where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_L \geq 0$ are the singular values of $\widehat{T}$, the best rank one approximation to $\widehat{T}$ in the Frobenius norm sense is given by $\sigma_1 v_1 v_1^H$ [44]. An estimate for the equaliser parameters is obtained in a straightforward way as $\widehat{\theta} = v_1^*$. We have left the factor $\sqrt{\sigma_1}$ out of $\widehat{\theta}$, since the multiplication of the equaliser parameters by a complex constant does not affect the equaliser performance.

The robustness of this extraction method stems from its making use of all the information available in $\widehat{\psi}$ to arrive at an estimate $\widehat{\theta}$ which is in accord with the structure of ψ. In situations where the computation of the SVD is not conceivable due to increased computational complexity, the best way to extract $\widehat{\theta}$ from any estimate of ψ will be to consider the first L entries of $\widehat{\psi}$ which

correspond to the equaliser parameters multiplied by θ_0^* (see Eq. (4.25)). An estimate for the equaliser parameters is then given by

$$\widehat{\theta} = \left[\widehat{\psi}_1, \widehat{\psi}_2, \ldots, \widehat{\psi}_L\right]^T. \tag{4.38}$$

The first tap $\widehat{\theta}_0$ of the resulting estimate above is always real and positive, the other taps being rotated accordingly. This rotation is not significant as it only causes a constant shift in $\phi(k)$.

4.4.3 Real Channels

If the baseband channel impulse response is real-valued, which is typically the case when the channel input sequence is drawn from a real (one-dimensional) constellation such as binary phase shift keying (BPSK), the number of columns in the matrix A can be reduced for a given equaliser length L. This reduction in dimensionality results from the insensitivity of real numbers to complex conjugation. To emphasise the difference between the complex and real channel cases, the matrix equation of Eq. (4.23) will be written for the latter case as

$$B\vartheta = 1. \tag{4.39}$$

Bearing in mind that the equaliser parameters θ_i are also real, some simple algebra reveals that B can be written as an $N \times L(L+1)/2$ (rather than $N \times L^2$) matrix with the following submatrix partitioning

$$B = [B_0, B_1, \ldots, B_{L-1}] \tag{4.40}$$

where

$$B_i \triangleq \begin{bmatrix} \Re\left\{r(k-i)r_i^H(k)\right\} \\ \Re\left\{r(k-i-1)r_i^H(k-1)\right\} \\ \vdots \\ \Re\left\{r(k-i-N+1)r_i^H(k-N+1)\right\} \end{bmatrix}, \quad i = 0, 1, \ldots, L-1. \tag{4.41}$$

The symbol $\Re\{v\}$ denotes a vector whose entries are the real parts of v, and $r_i(k)$ is the $(L-i) \times 1$ vector

$$r_i(k) \triangleq [r(k-i), r(k-i-1), \ldots, r(k-L+1)]^T, \quad i = 0, 1, \ldots, L-1. \tag{4.42}$$

The $L(L+1)/2 \times 1$ vector of transformed equaliser parameters ϑ is given by

$$\vartheta = \left[\theta_0 \varrho_0^T, \theta_1 \varrho_1^T, \ldots, \theta_{L-2} \varrho_{L-2}^T, \theta_{L-1}^2\right]^T \qquad (4.43)$$

where the $(L-i) \times 1$ vector ϱ_i is defined by

$$\varrho_i \triangleq [\theta_i, 2\theta_{i+1}, 2\theta_{i+2}, \ldots, 2\theta_{L-1}]^T, \quad i = 0, 1, \ldots, L-2. \qquad (4.44)$$

The equaliser parameters θ_i can easily be obtained from the relationship between the entries of ϑ in Eq. (4.43). If the channel is not an all-pole system or the channel noise is not negligible, an improvement on the resulting equaliser parameters can be achieved in much the same way as in Section 4.4.2 by applying a rank one approximation to the following matrix formulation of $\widehat{\vartheta} = \left[\widehat{\vartheta}_1, \widehat{\vartheta}_2, \ldots, \widehat{\vartheta}_{L(L+1)/2}\right]^T$

$$\widehat{T} = \begin{bmatrix} \widehat{\vartheta}_1 & \frac{\widehat{\vartheta}_2}{2} & \frac{\widehat{\vartheta}_3}{2} & \cdots & \frac{\widehat{\vartheta}_L}{2} \\ \frac{\widehat{\vartheta}_2}{2} & \widehat{\vartheta}_{L+1} & \frac{\widehat{\vartheta}_{L+2}}{2} & \cdots & \frac{\widehat{\vartheta}_{2L-1}}{2} \\ \frac{\widehat{\vartheta}_3}{2} & \frac{\widehat{\vartheta}_{L+2}}{2} & \widehat{\vartheta}_{2L} & \cdots & \frac{\widehat{\vartheta}_{3L-3}}{2} \\ \vdots & & & & \vdots \\ \frac{\widehat{\vartheta}_L}{2} & \frac{\widehat{\vartheta}_{2L-1}}{2} & \frac{\widehat{\vartheta}_{3L-3}}{2} & \cdots & \widehat{\vartheta}_{L(L+1)/2} \end{bmatrix}. \qquad (4.45)$$

If the SVD proves to be computationally too expensive, a modified version of Eq. (4.38) can be considered. Using only the first L entries of $\widehat{\vartheta}$, we can estimate the equaliser parameters as follows

$$\widehat{\theta} = \left[\widehat{\vartheta}_1, \frac{\widehat{\vartheta}_2}{2}, \frac{\widehat{\vartheta}_3}{2}, \ldots, \frac{\widehat{\vartheta}_L}{2}\right]^T. \qquad (4.46)$$

Example 4.3 Assume that the communication channel has the single-pole transfer function $H(z) = 1/(1 - 0.6z^{-1})$ and is driven by binary inputs. The equaliser has two taps $(L = 2)$ and we set N equal to $L(L+1)/2$, which implies that B is the 3×3 matrix

$$B = \begin{bmatrix} r^2(k) & r(k)r(k-1) & r^2(k-1) \\ r^2(k-1) & r(k-1)r(k-2) & r^2(k-2) \\ r^2(k-2) & r(k-2)r(k-3) & r^2(k-3) \end{bmatrix}. \qquad (4.47)$$

It is sufficient to collect only four channel output observations $\{r(k), r(k-1), r(k-2), r(k-3)\}$ to form B and solve Eq. (4.39) for ϑ. Supposing that we have made the following output observations $\{0.1251, 1.8751, 1.4585, 0.7642\}$ for which B is a full-rank matrix, the solution of

Eq. (4.39) is simply given by

$$\vartheta = \underbrace{\begin{bmatrix} 0.0156 & 0.2345 & 3.5161 \\ 3.5161 & 2.7349 & 2.1273 \\ 2.1273 & 1.1146 & 0.5840 \end{bmatrix}}_{B}^{-1} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1.00 \\ -1.20 \\ 0.36 \end{bmatrix} \qquad (4.48)$$

whence the equaliser parameters are obtained as $\theta = \pm[1, -0.6]^T$ with $\Upsilon = 1$ and $\phi(k) = 0$ or π radians.   □

4.5   On-line and Recursive Implementations of the Algorithm

The number of columns in A increases quadratically with the equaliser length L. For large L, a direct solution of Eq. (4.31) or Eq. (4.33) becomes computationally expensive because of a large matrix inversion involved. In what follows, we propose the normalised least-mean-square (LMS) and recursive least squares (RLS) implementations of the off-line LS algorithm. Both implementations avoid direct matrix inversion and are, therefore, computationally efficient as explained below. They may, however, require longer channel output observations to converge to a solution. We will use the notation $\widehat{\psi}(k)$ (or $\psi(k)$) to denote the time-dependent equaliser parameters.

4.5.1   Normalised Least-Mean-Square Algorithm

Given the transformed equaliser parameter vector $\psi(k)$, the error between the squared modulus of the equaliser output and its desired value is given by

$$e(k) = \left(r^*(k) \otimes r(k)\right)^H \psi(k) - 1. \qquad (4.49)$$

The squared modulus of the error is

$$|e(k)|^2 = e^*(k)e(k) \qquad (4.50a)$$

$$= \left(\psi^H(k)\left(r^*(k) \otimes r(k)\right) - 1\right)\left(\left(r^*(k) \otimes r(k)\right)^H \psi(k) - 1\right) \qquad (4.50b)$$

$$= \psi^H(k)\left(r^*(k) \otimes r(k)\right)\left(r^*(k) \otimes r(k)\right)^H \psi(k) - \psi^H(k)\left(r^*(k) \otimes r(k)\right)$$
$$- \left(r^*(k) \otimes r(k)\right)^H \psi(k) + 1. \qquad (4.50c)$$

The mean-squared error can now be written as

$$\mathcal{J}_{\text{MSE}}(\psi(k)) = E\left\{|e(k)|^2\right\} \tag{4.51a}$$

$$= \psi^H(k)M\psi(k) - \psi^H(k)c - c^H\psi(k) + 1 \tag{4.51b}$$

where the fourth-order moment matrix M and the correlation vector c are defined in Eq. (4.33). The gradient of $\mathcal{J}_{\text{MSE}}(\psi(k))$ with respect to $\psi(k)$ is given by

$$\nabla\mathcal{J}_{\text{MSE}}(\psi(k)) = 2M\psi(k) - 2c. \tag{4.52}$$

Using the gradient vector, a steepest descent algorithm can be obtained as follows

$$\psi(k+1) = \psi(k) - \frac{\mu}{2}\nabla\mathcal{J}_{\text{MSE}}(\psi(k)) \tag{4.53a}$$

$$= \psi(k) + \mu\left(c - M\psi(k)\right) \tag{4.53b}$$

where μ is a small positive stepsize. Replacing the moment matrix M and the correlation vector c in Eq. (4.53b) with their *instantaneous* estimates at time k

$$\widehat{M}(k) = \left(r^*(k) \otimes r(k)\right)\left(r^*(k) \otimes r(k)\right)^H \tag{4.54a}$$

$$\widehat{c}(k) = r^*(k) \otimes r(k) \tag{4.54b}$$

we obtain the instantaneous estimate of the gradient vector as

$$\widehat{\nabla}\mathcal{J}_{\text{MSE}}\left(\widehat{\psi}(k)\right) = 2\left(r^*(k) \otimes r(k)\right)\left(r^*(k) \otimes r(k)\right)^H \widehat{\psi}(k) - 2\left(r^*(k) \otimes r(k)\right). \tag{4.55}$$

Substituting $\widehat{\nabla}\mathcal{J}_{\text{MSE}}\left(\widehat{\psi}(k)\right)$ for $\nabla\mathcal{J}_{\text{MSE}}(\psi(k))$ in Eq. (4.53a) yields the LMS implementation

$$\widehat{\psi}(k+1) = \widehat{\psi}(k) - \mu\left(r^*(k) \otimes r(k)\right)\underbrace{\left(\left(r^*(k) \otimes r(k)\right)^H \widehat{\psi}(k) - 1\right)}_{\widehat{e}(k)}. \tag{4.56}$$

The LMS algorithm is known to suffer from the so-called *gradient noise amplification* when $r^*(k) \otimes r(k)$ is large [45]. A way to circumvent this problem is to use a time-varying stepsize which is obtained by normalising the correction term in Eq. (4.56) with respect to the squared Euclidean norm of $r^*(k) \otimes r(k)$. This results in the so-called *normalised LMS algorithm*

$$\widehat{\psi}(k+1) = \widehat{\psi}(k) - \frac{\tilde{\mu}}{\epsilon + \|r^*(k) \otimes r(k)\|_2^2}\left(r^*(k) \otimes r(k)\right)\widehat{e}(k) \tag{4.57}$$

where $\epsilon > 0$ is a small number whose purpose is to avoid numerical problems that may arise from the division of $\tilde{\mu}$ by $\|r^*(k) \otimes r(k)\|_2^2$ when $r^*(k) \otimes r(k)$ is small. The stepsize should satisfy the inequality $0 < \tilde{\mu} < 2$. The normalised LMS has a faster convergence rate than LMS in certain situations [46]. The algorithm is usually initialised to $\widehat{\psi}(0) = 0$. The eigenspread of M tends to affect the convergence rate of the algorithm. Overparameterisation of the equaliser, which will be shown to cause rank deficiency in M in Section 4.6, can therefore lead to poor convergence in certain cases.

4.5.2   Recursive Least Squares Algorithm

The explicit matrix inversion in Eq. (4.31) can be dispensed with by computing $\left(A^H A\right)^{-1}$ recursively with the aid of the *matrix inversion lemma* (see e.g. [45]). An exponentially weighted LS cost can be written for the observation interval $1 \leq i \leq k$ as

$$\mathcal{J}_{\text{RWLS}}(\psi(k)) = \sum_{i=1}^{k} \lambda^{k-i} \left| (r^*(i) \otimes r(i))^H \psi(k) - 1 \right|^2 \tag{4.58}$$

where λ is the forgetting factor ($0.9 < \lambda \leq 1$) and the quantity weighted by λ^{k-i} is the squared modulus of the error between the squared modulus of the equaliser output and its desired value (see Eq. (4.49). The optimum value of $\psi(k)$ that minimises $\mathcal{J}_{\text{RWLS}}(\psi(k))$ is obtained from the time-varying normal equations

$$\widehat{M}(k)\widehat{\psi}(k) = \widehat{c}(k) \tag{4.59}$$

where $L^2 \times L^2$ time-varying Gram matrix $\widehat{M}(k)$ is defined by

$$\widehat{M}(k) = \sum_{i=1}^{k} \lambda^{k-i} q(i) q^H(i) \tag{4.60}$$

and the $L^2 \times 1$ correlation vector $\widehat{c}(k)$ is defined by

$$\widehat{c}(k) = \sum_{i=1}^{k} \lambda^{k-i} q(i) \tag{4.61}$$

with

$$q(i) \triangleq \begin{bmatrix} r(i) \\ r(i-1) \\ \vdots \\ r(i-L+1) \end{bmatrix}^* \otimes \begin{bmatrix} r(i) \\ r(i-1) \\ \vdots \\ r(i-L+1) \end{bmatrix}. \tag{4.62}$$

In the definition of $\widehat{M}(k)$ and $\widehat{c}(k)$ above, the channel output vector $q(i)$ has been assumed to be prewindowed so that $q(i) = 0$ for $i < 1$. It follows from Eq. (4.59) that the LS estimate for the transformed equaliser parameters at time k is given by

$$\widehat{\psi}(k) = \widehat{M}^{-1}(k)\widehat{c}(k). \tag{4.63}$$

Defining $P(k) = \widehat{M}^{-1}(k)$ and following the same line of development as in the recursive least squares (RLS) algorithm, we finally obtain the following algorithm to compute $\widehat{\psi}(k)$ recursively without any matrix inversion:

$$g(k) = \frac{\lambda^{-1}P(k-1)q(k)}{1 + \lambda^{-1}q^H(k)P(k-1)q(k)}$$

$$\alpha(k) = 1 - \widehat{\psi}^H(k-1)q(k)$$

$$\widehat{\psi}(k) = \widehat{\psi}(k-1) + g(k)\alpha^*(k)$$

$$P(k) = \lambda^{-1}P(k-1) - \lambda^{-1}g(k)q^H(k)P(k-1).$$

The above algorithm is initialised by setting $\widehat{\psi}(0) = 0$ and $P(0) = \kappa I$ where κ is a large positive number.

Global convergence of the RLS algorithm is dependent upon persistent excitation at the channel output. The existence of a unique solution to Eq. (4.63) requires $q(i)$ to be such that $\widehat{M}(k)$ is invertible. Such a $q(i)$ is termed *sufficiently exciting*. Persistent excitation is achieved if $q(i)$ satisfies the following condition

$$aI < \frac{1}{l} \sum_{i=j+1}^{j+l} q(i)q^H(i) < bI \tag{4.64}$$

for all j, some fixed l, some $a > 0$, and some $b < \infty$. The inequality signs in Eq. (4.64) imply that the eigenvalues of the matrix $(1/l)\sum_{i=j+1}^{j+l} q(i)q^H(i)$ are constrained to lie on the interval (a, b). The lower bound a in Eq. (4.64) is of considerable importance because, if a is very close (or equal) to zero, as would be the case if M were rank deficient, at least one eigenvalue of $P(k)$ will explode, causing severe numerical problems [46].

4.6 Effects of Overparameterisation

4.6.1 Some Preliminaries

Thus far we have assumed that the matrix A (or B in the case of real channels) has full column rank. This assumption guarantees that there exists a *unique* solution to Eq. (4.23). The full-column-rank assumption is not, however, valid if the equaliser length L is larger than the channel inverse length P. This results in multiple solutions for the transformed equaliser parameters. If multiple solutions exist, the best strategy is in general to single out the solution that has the minimum ℓ_2 norm. In our case, however, the norm of the solution is of little concern because the equalisation gain Υ is allowed to be an arbitrary number in the equalisation objective (see Eq. (4.17)). While A is rank deficient for all-pole channels if $L > P$, it becomes ill-conditioned[2] for ARMA channels if L is larger than the approximate channel inverse length P in Eq. (4.2).

Proposition 4.1 *If the channel is an all-pole (AR) system and $L > P$, then the maximum column rank of the $N \times L^2$ matrix A is $L^2 - (L - P)$.*

Proof. Eq. (4.23) is consistent since the channel is an all-pole system with an inverse as in Eq. (4.18) and the equaliser length L is larger than P. Shifting the true channel inverse parameters

$$a = [a_0, a_1, \ldots, a_{P-1}]^T \tag{4.65}$$

in the $L \times 1$ equaliser parameter vector θ and filling in the rest of θ with zeros produces

$$\theta_i = \begin{bmatrix} 0_{i \times 1} \\ a \\ 0_{(L-P-i) \times 1} \end{bmatrix} \quad i = 0, 1, \ldots, L - P \tag{4.66}$$

which are the linearly independent equaliser parameter vectors with $0_{i \times j}$ denoting the $i \times j$ matrix of zeros. Substituting the θ_i in Eq. (4.25) gives the corresponding linearly independent solutions of Eq. (4.23), which we shall denote by ψ_i for $i = 0, 1, \ldots, L - P$. Using the fact that the matrix A in Eq. (4.23) has to be rank deficient by the number of linearly independent solutions of Eq. (4.23) minus one, we conclude that for arbitrary channel output observations A has the maximum column rank $L^2 - (L - P)$. Note that we have not ruled out the possibility of having more than $L - P + 1$ linearly independent solutions, hence the maximum rank qualifier. ∎

---

[2]A matrix is called ill-conditioned if the ratio of its maximum singular value to its minimum singular value (also known as the *condition number*) is large.

If A is rank deficient and Eq. (4.18) holds, which ensures that Eq. (4.23) is consistent, all the solutions of Eq. (4.23) are given by [27]

$$\psi = A^- 1 + (\mathbf{I} - A^- A)z \qquad (4.67)$$

where A^- is a *g-inverse* of A which satisfies the relation $AA^- A = A$ and z is an arbitrary $L^2 \times 1$ vector. Eq. (4.67) reveals that ψ is not unique if A is rank deficient. As a consequence, ψ is not necessarily related to θ by Eq. (4.25) if the equaliser is overparameterised. Nevertheless, a certain relationship between the entries of ψ still exists as shown below, which allows for the recovery of the equaliser parameters θ from any ψ obtained from Eq. (4.67).

Proposition 4.2 *If the channel is an all-pole system, the equaliser is overparameterised (i.e.* $L > P$) *and* A *has the maximum column rank* $L^2 - (L - P)$, *then the equaliser parameters are given by*

$$\theta = c\Big[\psi_1, \psi_2, \dots, \psi_P, \underbrace{0, \dots, 0}_{L-P \text{ times}}\Big]^T \qquad (4.68)$$

where c *is some constant and the* ψ_i, $i = 1, 2, \dots, P$ *are the first* P *entries of* ψ *obtained from Eq. (4.67).*

Proof. Under the conditions stated in the proposition, any solution of Eq. (4.23) can be written as a linear combination of the independent solutions defined in the proof of Proposition 4.1

$$\psi = \sum_{i=0}^{L-P} c_i \psi_i \qquad (4.69)$$

where the c_i obey the relation $\sum_{i=0}^{L-P} c_i = 1$, since the right-hand side of Eq. (4.23) is the unity vector. Noting that the first L entries of the ψ_i are zero for $i = 1, 2, \dots, L - P$, which can easily be seen by substituting Eq. (4.66) in Eq. (4.25), we have

$$[\psi_1, \psi_2, \dots, \psi_L]^T = c_0 [\psi_{0,1}, \psi_{0,2}, \dots, \psi_{0,L}]^T \qquad (4.70)$$

where $\psi_{0,j}$ denotes the jth entry of ψ_0. Using Eqs. (4.25) and (4.66), we finally obtain

$$[\psi_1, \psi_2, \dots, \psi_L]^T = c_0 a_0^* [a_0, a_1, \dots, a_{P-1}, 0, \dots, 0]^T \qquad (4.71)$$

whence Eq. (4.68) is obtained with $c = 1 / (c_0 a_0^*)$. ∎

Remarks 4.1

    i. The exact solution for the equaliser parameters in Eq. (4.68) yields the minimum equali-
sation delay Δ.

    ii. As we have remarked earlier, the unknown constant c can be ignored by incorporating
it in the equaliser parameters, which will cause some scaling and rotation of the equaliser
output without any violation of the equalisation objective in Eq. (4.17).

4.6.2 Least Squares Solution Revisited

Whether or not the equaliser is overparameterised and the channel is an AR system, the minimum-
norm LS solution for the transformed equaliser parameters is given by [27]

$$\widehat{\psi} = A^{\dagger} \mathbf{1} \tag{4.72}$$

where $A^{\dagger}$ is the Moore-Penrose inverse of A. In Section 4.4.1 we were motivated by the possible
inconsistency of Eq. (4.23) to derive the LS solution. Although using the same logic we can
arrive at the LS solution in Eq. (4.31) for overparameterised equalisers as well, caution should be
exercised since an ill-conditioned A can cause earnest numerical problems and a large LS error
$\|A\widehat{\psi} - \mathbf{1}\|_2^2$ [47]. The advantage of using Eq. (4.72) is that any ill-conditioning of A becomes
immediately evident when the singular values of A are computed. Although for ARMA channels
none of the singular values will be exactly zero, some of them may be very small if $L > P$. In
computing the Moore-Penrose inverse we take the reciprocal of the nonzero singular values of
A (refer to Appendix 3.A.1). To avoid numerical problems, a decision has to be made between
taking the reciprocal of small singular values and setting them to zero. It is usually advantageous
to adopt the latter and continue with g-inverse computations by pretending A to be rank deficient.

    In Eq. (4.72) the matrix A is required to have a finite condition number if the channel is an
ARMA system. This can be achieved by selecting an appropriate value for N, the channel input
correlation permitting. In this case, a robust extraction of θ from $\widehat{\psi}$ can be achieved by means of
a reduced rank matrix approximation to $\widehat{T}$ in Eq. (4.35) which is this time constructed using the
entries of $\widehat{\psi}$ obtained from Eq. (4.72). For perfect AR channels the approximation to $\widehat{T}$ will have
rank $L - P + 1$ as can be seen from Eq. (4.69). In view of numerical problems that may arise
from ill-conditioning of A we will look for an approximation to $\widehat{T}$ with rank $L - P + 1$. The best
rank $L - P + 1$ approximation to $\widehat{T}$ in the Frobenius norm sense is given by the truncated SVD

of $\widehat{T}$ [44]

$$\widehat{T}_{\text{trunc}} = \sum_{i=1}^{L-P+1} \sigma_i v_i v_i^H. \tag{4.73}$$

An estimate for the equaliser parameters is then given by $\widehat{\theta} = \sum_{i=1}^{L-P+1} \sigma_i v_{i,1} v_i^H$ where $v_{i,1}$ denotes the first entry of v_i.

Rather than using the SVD, the equaliser parameters may be extracted using Proposition 4.2, which results in Eq. (4.38) without any modifications. Recall that Eq. (4.38) uses only the first L entries of $\widehat{\psi}$ rather than the entire $\widehat{\psi}$ to obtain an estimate for the equaliser parameters.

4.6.3 Iterative Computation of the Minimum-Norm Least Squares Solution

For large L, the computation of the Moore-Penrose inverse in Eq. (4.72) may become prohibitively expensive. To alleviate the computational cost involved to some degree, the iterative scheme discussed in Appendix 3.A can be invoked to compute the Moore-Penrose inverse of A without computing the SVD. The Moore-Penrose inverse is incidentally a special type of g-inverse. For the sake of convenience, we have reproduced the procedure below. If $0 < \mu < \frac{2}{\lambda_{\max}(A^H A)}$, the sequence

$$Y(0) = \mu A^H$$
$$Y(i+1) = Y(i)(2I - AY(i)), \quad i = 0, 1, \dots$$

converges to $A^\dagger$ as $i \to \infty$. The computed Moore-Penrose inverse can also be substituted for A^- in Eq. (4.67) to obtain an exact solution ψ if the channel is an AR system. Once $\widehat{\psi}$ is computed using the resulting $A^\dagger$, one of the methods suggested at the end of Section 4.6.2 can be applied to arrive at the equaliser parameters.

4.6.4 Modified Recursive Least Squares Algorithm

When the equaliser is overparameterised (i.e. $L > P$), the matrix of channel output fourth-order moments M will be rank deficient by $L - P$, which follows from the definition of M in Eq. (4.33)

$$M = \lim_{N \to \infty} \frac{1}{N} A^H A.$$

If the channel is an ARMA system and the equaliser is overparameterised, M will be ill-conditioned. This implies that the $q(i)$ will no longer be persistently exciting since the lower bound in Eq. (4.64) will either be $a = 0$ if $H(z)$ is an AR system or $a \approx 0$ if $H(z)$ is ARMA. This

can cause severe numerical problems if the RLS algorithm is used as it stands. In this section, we consider a modified version of RLS proposed in [48], which imposes upper and lower bounds on the eigenvalues of $P(k)$. The resultant algorithm consists of the following iterations:

$$\widehat{\psi}(k) = \widehat{\psi}(k-1) + \frac{\alpha P(k-1)q(k)}{1 + q^H(k)P(k-1)p(k)}\left(1 - \widehat{\psi}^H(k-1)q(k)\right) \qquad (4.74a)$$

$$P(k) = \frac{1}{\lambda}P(k-1) - \frac{\alpha P(k-1)q(k)q^H(k)P(k-1)}{1 + q^H(k)P(k-1)q(k)} + \beta I - \delta P^2(k-1) \qquad (4.74b)$$

where α, β, λ and δ are some constants. It is shown in [48] that if these constants satisfy the following constraints

$$\begin{aligned}
0 &< \gamma < \alpha < 1 \\
(\gamma - \alpha)^2 &+ 4\beta\delta < (1 - \alpha)^2 \\
\beta &> 0, \quad \delta > 0 \\
\bar{\sigma}I &\leq P(0) \leq \bar{\nu}I
\end{aligned} \qquad (4.75)$$

where

$$\gamma \triangleq \frac{1 - \lambda}{\lambda} \qquad (4.76)$$

$$\bar{\sigma} \triangleq \frac{\alpha - \gamma}{2\delta}\left(\sqrt{1 + \frac{4\beta\delta}{(\alpha - \gamma)^2}} - 1\right) \qquad (4.77)$$

$$\bar{\nu} \triangleq \frac{\gamma}{2\delta}\left(\sqrt{1 + \frac{4\beta\delta}{\gamma^2}} + 1\right) \qquad (4.78)$$

then the matrix $P(k)$ has the following properties:

i. $\bar{\sigma}I \leq P(k) \leq \bar{\nu}I$ for all k

ii. if $q(k) = 0$ for all k, then $P(k) \to \bar{\nu}I$.

Property i. ensures that the eigenvalues of $P(k)$ will not explode even if the matrix M is not invertible. Property ii. is the exponential resetting property of the algorithm. The constants of the algorithm are chosen on the following intervals: $\alpha \in [0.1, 0.5]$, $\beta \in (0, 0.01]$, $\lambda \in [0.9, 0.99]$ and $\delta \in (0, 0.01]$. The initial value of $P(k)$ is set to $P(0) = \kappa I$ where $\bar{\sigma} < \kappa < \bar{\nu}$. In order to improve the accuracy of the resulting equaliser parameter estimates and to avoid sluggish convergence to the solution, the bounds on the eigenvalues of $P(k)$, i.e. the quantities $\bar{\sigma}$ and $\bar{\nu}$, should be selected in accordance with the expected eigenspread of M. For instance, if $\bar{\nu}$ is chosen too small, slow convergence can result, not to mention gross inaccuracies in the converged parameters.

4.7 Extension to PAM Constellations

We have so far assumed that the channel input sequence has a constant modulus. This can impose a harsh limitation on the applicability of the algorithm since there is a wealth of digital modulation techniques with so-called *higher-order constellations*, such as pulse amplitude modulation (PAM) and quadrature amplitude modulation (QAM). CMA can be made to work for higher-order constellations although its cost function is somewhat counterintuitive. The first evidence of a criterion similar to CMA, which works for PAM constellations, was provided in [8]. In this section, we investigate the possibility of extending the channel equalisation approach expounded in the previous sections to such constellations after some modifications. For the sake of simplicity, we will concentrate on PAM communication systems only; extension to other higher-order constellations should be straightforward.

Consider the PAM (or M-ary) constellation defined as $\mathbb{S} = \{\pm 1, \pm 3, \ldots, \pm(M-1)\}$ where $M > 0$ is an even integer. The baseband channel $H(z)$ for PAM systems is necessarily real since PAM is a one-dimensional modulation technique as opposed to PSK or FSK which have two-dimensional (complex) constellations. This said, it should be clear that the matrix equation $B\vartheta = 1$ does not hold for pulse amplitude modulated channel inputs because the vector on the right-hand side would consist of members of the set $\mathbb{Y} = \{1, 3^2, \ldots, (M-1)^2\}$ rather than just ones. This immediately raises the question as to whether or not it is possible to solve the following matrix equation by searching for the "right" vector of squared equaliser outputs d whose entries are in $\mathbb{Y}$

$$B\vartheta = d \tag{4.79}$$

where both ϑ and d are unknown. If B is a full-column-rank matrix with more rows than columns (i.e. $N > L(L+1)/2$) and the channel is an AR system, the consistency of Eq. (4.79) hinges upon the satisfaction of the matrix equation [27]

$$BB^\dagger d = d \tag{4.80}$$

or, equivalently,

$$Pd = 0 \tag{4.81}$$

where $B^\dagger = \left(B^T B\right)^{-1} B^T$ is the Moore-Penrose inverse of B and $P = I - BB^\dagger$ is the LS projection of d onto noise subspace. Incidentally, the equaliser can be overparameterised without causing any rank deficiency in B as ϑ has to be unique for a given d unlike the case of $d = 1$. Eq. (4.81) has multiple solutions for d since P has column rank $N - L(L+1)/2$. On the other

hand, the vector d is constrained to have all of its entries in $\mathbb{Y}$. Thus, the solution of Eq. (4.81) can be obtained by trying out all possible d until one is found that satisfies the equation. The search problem at hand can be formally written as

Search for $d = [d_1, d_2, \ldots, d_N]^T$ such that $Pd = 0$ and $d_i \in \mathbb{Y}$, $i = 1, 2, \ldots, N$.

The maximum number of trials to find the $N \times 1$ vector d that solves Eq. (4.81) is $(M/2)^N$ where $N > L(L+1)/2$. Even if N is set to its minimum value $L(L+1)/2 + 1$, the number of trials needed to find the right d can be extremely large if L is large. Therefore, the applicability of the search scheme[3] is limited to equalisers with short length. The procedure is illustrated below.

Example 4.4 Consider the channel $H(z) = 1/(1 - 0.6z^{-1})$ driven by a PAM sequence with $M = 4$. A full column rank $(L(L+1)/2 + 1) \times L(L+1)/2$ matrix B for this channel is given by

$$B = \begin{bmatrix} 0.0024 & -0.0769 & 2.5150 \\ 2.5150 & 6.8348 & 18.5743 \\ 18.5743 & 9.4082 & 4.7654 \\ 4.7654 & 4.3040 & 3.8873 \end{bmatrix} \tag{4.82}$$

where $L = P = 2$. The number of different d vectors to be tried out is $2^4 = 16$. All possible d of length 4 and the resulting squared ℓ_2 norm when they are substituted in Eq. (4.81) are listed in Table 4.1. The right d vector is obviously $[1, 1, 9, 1]^T$. Solving Eq. (4.79) for ϑ using $d = [1, 1, 9, 1]^T$ gives

$$\vartheta = B^{\dagger} \begin{bmatrix} 1 \\ 1 \\ 9 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1.2 \\ 0.36 \end{bmatrix} \tag{4.83}$$

whence the equaliser parameters are obtained as $\theta = \pm[1, -0.6]^T$.   □

Notice that the procedure demonstrated in the above example is essentially a search for a vector that makes Eq. (4.79) consistent subject to certain constraints on its entries. The method works only if the matrix B has more rows than columns because underdetermined matrix equations for which the data matrix B has less rows than columns are either consistent with infinitely many solutions or do not have any solutions. Therefore, if Eq. (4.79) is underdetermined, it may be consistent for any choice of d, making the above procedure inapplicable. This fact restricts the consideration to overdetermined systems that have data matrices with more rows than columns, hence the lower limit on the value of N with respect to L. We have ignored the inconsistencies in

---

[3]The search problem in question is in fact *NP-hard*.

d^T	$\|Pd\|_2^2$	d^T	$\|Pd\|_2^2$
[1, 1, 1, 1]	1.0812	[9, 1, 1, 1]	50.2872
[1, 1, 1, 9]	35.0504	[9, 1, 1, 9]	143.3264
[1, 1, 9, 1]	0	[9, 1, 9, 1]	36.6214
[1, 1, 9, 9]	23.8198	[9, 1, 9, 9]	119.5111
[1, 9, 1, 1]	0.2854	[9, 9, 1, 1]	30.4405
[1, 9, 1, 9]	18.8902	[9, 9, 1, 9]	108.1152
[1, 9, 9, 1]	2.4776	[9, 9, 9, 1]	20.0481
[1, 9, 9, 9]	10.9329	[9, 9, 9, 9]	87.5733

Table 4.1 List of all possible d vectors and resulting squared ℓ_2 norms of the consistency equation.

Eq. (4.79), as well as the channel noise. A consideration of these will prompt a different search criterion such as the minimum ℓ_2 norm rather than strict equality to zero, leading to

$$\widehat{d} = \arg\min_{d_i \in \mathbf{Y}} \|Pd\|_2^2 \qquad (4.84)$$

where $\widehat{d}$ is the norm minimising vector.

Having pointed out the possibility of finding the equaliser parameters by an exhaustive search method, we will now investigate a more practical approach which does not require as much computational effort and simply draws on the asymptotic Wiener solution in Eq. (4.33) for constant modulus channel inputs. If we had a knowledge of the channel input sequence, the normal equations could be written as

$$B^T B \widehat{\vartheta} = B^T d_m \qquad (4.85)$$

where d_m is the $N \times 1$ vector of squared channel inputs delayed by m

$$d_m = \left[u^2(k-m), u^2(k-m-1), \ldots, u^2(k-m-N+1) \right]^T. \qquad (4.86)$$

As $N \to \infty$, the solution of the normal equations divided through by N stochastically converges to that of the following matrix equation

$$\underbrace{E\{v(k)v(k)^T\}}_{M_{\text{PAM}}} \vartheta_o = \underbrace{E\{u^2(k-m)v(k)\}}_{c_{\text{PAM}}} \qquad (4.87)$$

where $v(k)$ is a $L(L+1)/2 \times 1$ vector defined in terms of $r_i(k)$ in Eq. (4.42) as

$$v(k) \triangleq \left[r(k)r_0^T(k), r(k-1)r_1^T(k), \ldots, r(k-L+1)r_{L-1}^T(k) \right]^T. \qquad (4.88)$$

In Eq. (4.87) M_{PAM} contains fourth-order moments of the channel output, and c_{PAM} is the vector of fourth-order crossmoments of the channel input and output. One of the difficulties with Eq (4.87) is that we cannot compute ϑ_o without access to the channel input. Nor can we circumvent this problem by appealing to the constant modulus of $u(k)$ as we did before. Another problem is that the solution is sensitive to the equalisation delay inevitably introduced by nonminimum phase channels. Indeed, we require $m \geq \Delta$ and $L \geq P + m - \Delta$, so that ϑ_o can accommodate the equaliser parameters.

Next we conjecture that it is possible to extract the equaliser parameters from the solution of Eq. (4.87) by replacing c_{PAM}, whose *a priori* knowledge is unavailable in a blind equalisation setting, with $E\{u^2(k)\}E\{v(k)\}$ which is the real-channel counterpart of the correlation vector c in Eq. (4.33) scaled by the channel input variance. This method is applicable only if the channel can be equalised without any equalisation delay, i.e. the channel is a minimum phase system.

The basic idea is as follows. If m is such that $u^2(k - m)$ and the entries of $v(k)$ are statistically independent, and $\{u(k)\}$ is weakly stationary, then Eq. (4.87) can be written, using only the channel output moments and the channel input variance, as follows

$$M_{\mathrm{PAM}}\vartheta = E\{u^2(k)\}E\{v(k)\}. \tag{4.89}$$

Since m that leads to the above equation is not unique, the vector ϑ consists of a weighted sum of shifted parameter vectors ϑ_i

$$\vartheta = \sum_i c_i' \vartheta_i \tag{4.90}$$

where the ϑ_i are computed from the corresponding θ_i using Eq. (4.43). For minimum phase channels, the θ_i are defined by

$$\theta_i = \begin{cases} \left[0_{i\times 1}, a^T, 0_{(L-P-i)\times 1}\right]^T & \text{if } 0 \leq i \leq L - P \\ \left[0_{i\times 1}, a_0, \ldots, a_{L-i-1}\right]^T & \text{if } L - P < i \leq L - 1 \end{cases} \tag{4.91}$$

where a is defined in Eq. (4.65). A finite-length ϑ will therefore result in a finite range for the summation index in Eq. (4.90). Note that Eq. (4.89) is not sensitive to delays in the equaliser output, which is the reason why θ_0 contains a without any delay. For a minimum phase channel, the scaled equaliser parameters are easily obtained from the first L entries of $\vartheta = M_{\mathrm{PAM}}^{-1}E\{u^2(k)\}E\{v(k)\}$ as follows

$$\theta = \frac{1}{c_0'\theta_0}\left[\vartheta_1, \frac{\vartheta_2}{2}, \ldots, \frac{\vartheta_L}{2}\right]^T \tag{4.92}$$

where ϑ_i denotes the ith entry of ϑ. The matrix M_{PAM} normally has full rank even if the equaliser
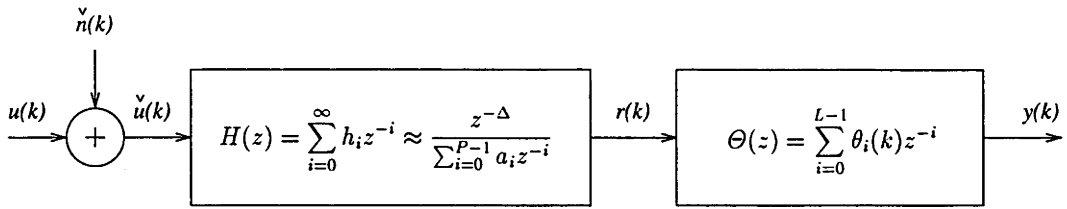
Figure 4.3 Equivalent channel and equaliser models. The channel input $\check{u}(k)$ is a summation of $u(k)$ and $\check{n}(k)$ where $\check{n}(k)$ is the equivalent channel noise at the channel input.

is overparameterised. PAM systems are sensitive to the gain factor Υ. Gain identification is therefore needed to compensate for the unknown factor $1/\left(c_0'\theta_0\right)$. This can be achieved by matching the equaliser output statistics to channel input statistics after the equaliser parameters are set to $c_0'\theta_0\theta$.

4.8 Channel Noise Considerations

The discussion up to this point has assumed zero channel noise. Needless to say, the solution of Eq. (4.23) is sensitive to any disturbance at the channel output. The effects of the channel noise can be best grasped by relegating it to the channel input and considering the resultant communication system to be noise-free. This approach results in a distorted channel input constellation that does not possess the constant modulus property any more. To see this, consider the noisy channel output

$$r(k) = \sum_{i=0}^{\infty} h_i u(k - i) + n(k) \tag{4.93}$$

which can be rewritten as

$$r(k) = \sum_{i=0}^{\infty} h_i \underbrace{\left(u(k - i) + \check{n}(k - i)\right)}_{\check{u}(k - i)} \tag{4.94}$$

where $\check{n}(k)$ is the equivalent noise at the channel input defined through

$$n(k) = \sum_{i=0}^{\infty} h_i \check{n}(k - i) \tag{4.95}$$

We denote the *distorted* channel input by $\check{u}(k)$ which is simply equal to the channel input $u(k)$ plus the equivalent noise $\check{n}(k)$. Replacing $u(k)$ with $\check{u}(k)$, the channel equalisation problem in Fig. 4.1 can now be modelled as shown in Fig. 4.3.

With no prior knowledge of the channel noise, the method of least squares strives to find an estimate for ψ such that $|y(k)|^2$ is more or less constant. This results in errors in $\widehat{\theta}$ relative to the

zero channel noise case. The magnitude of errors will certainly depend on $|\tilde{n}(k)|^2$. Ideally, for nonzero channel noise Eq. (4.19) should be modified as follows to make the equaliser parameter estimates invariant to the channel noise

$$
\begin{bmatrix} |y(k)|^2 \\ |y(k-1)|^2 \\ \vdots \\ |y(k-N+1)|^2 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}}_{\mathbf{1}} + \underbrace{\begin{bmatrix} |\varepsilon(k-\Delta)|^2 \\ |\varepsilon(k-\Delta-1)|^2 \\ \vdots \\ |\varepsilon(k-\Delta-N+1)|^2 \end{bmatrix}}_{\varepsilon}
\tag{4.96}
$$

where $|\varepsilon(k)|^2 = 2\Re\{u(k)\tilde{n}^*(k)\} + |\tilde{n}(k)|^2$. The error resulting from the use of Eq. (4.31) will therefore be

$$
\left(A^H A\right)^{-1} A^H \mathbf{1} - \left(A^H A\right)^{-1} A^H(\mathbf{1}+\varepsilon) = -\left(A^H A\right)^{-1} A^H \varepsilon.
\tag{4.97}
$$

Asymptotically (as $N \to \infty$), the Wiener solution can be written as

$$
\psi_o = M^{-1}\left(c + E\left\{|\varepsilon(k-\Delta)|^2 r^*(k) \otimes r(k)\right\}\right).
\tag{4.98}
$$

It is obvious from Eqs. (4.96) and (4.98) that the channel noise affects the solution of Eq. (4.23) in proportion to its variance. If the signal-to-noise ratio (SNR) at the channel output is below 20dB, it is advisable to choose a large N (compared with L^2) so that the resultant estimate of ψ does not suffer from a large variance. The 20dB threshold for the channel output SNR is a common requirement for most equalisation techniques [49]. We finally note that a perfect identification of the channel inverse does not always guarantee error-free recovery of the channel input if the channel output is corrupted by noise. Formulation of an equalisation criterion in the noisy channel case is not therefore a well-posed problem.

4.9   Simulation Studies

Various implementations of the proposed LS approach to blind channel equalisation will be demonstrated by means of computer simulations. In the following simulation examples, the communication channel will be assumed to have the form

$$
H(z) = z^{-\Delta} \sum_{i=-\Delta}^{p} b_i z^{-i}
\tag{4.99a}
$$

$$
\approx z^{-\Delta} B^{-1}(z)
\tag{4.99b}
$$

with $\Delta = 20, p = 9$ and

$$B(z) = (0.4 - 1.2j) + (0.6 + j)z^{-1} + (0.7 + 0.4j)z^{-2}. \qquad (4.100)$$

Note that $B(z)$ is a nonminimum phase system and, therefore, its inverse has a divergent right-sided z-transform. In order to obtain a stable inverse for $B(z)$, one has to consider the double-sided (noncausal) z-transform expansion. Accordingly, $B^{-1}(z)$ in Eq. (4.99b) denotes the double-sided z-transform of the inverse of $B(z)$. The delay $z^{-\Delta}$ is required to make the resulting truncated system causal. Since $H(z)B(z) \approx z^{-20}$, the minimum equaliser length is equal to the length of the impulse response of $B(z)$, i.e. $P = 3$. The real and imaginary parts of the channel impulse response are shown in Fig. 4.4. In Example 4.9 (extension to PAM constellations) we will make an exception and assume a real AR channel. The channel input will be assumed to be an i.i.d. sequence drawn from the 8-DPSK constellation $\mathbb{S} = \left\{1, e^{j\frac{\pi}{4}}, e^{j2\frac{\pi}{4}}, \ldots, e^{j7\frac{\pi}{4}}\right\}$, except for the simulation example dealing with PAM, in which the channel input will instead be assumed to be drawn from an 8-PAM constellation. The equaliser will be overparameterised with $L = 5$ except when the RLS algorithm is used. Parameter adaptation plots will be provided to illustrate the evolution of the equaliser parameters. Since the equaliser parameters are complex-valued, the plots will include only the absolute values of the first L entries of $\widehat{\psi}(k)$

$$\left[|\widehat{\psi}_1(k)|, |\widehat{\psi}_2(k)|, \ldots, |\widehat{\psi}_L(k)|\right]^T$$

which is in accordance with Proposition 4.2. The equaliser parameters are estimated from the final $\widehat{\psi}$ using Eq. (4.38).

To quantify the amount of ISI at the equaliser output $y(k)$ we will use the closed-eye measure (CLEM) defined in Eq. (1.13). Refer to Eq. (1.13) for comments on the significance and interpretation of CLEM.

Example 4.5 (Normalised LMS) The normalised LMS implementation does not suffer from overparameterisation to a large extent as will be illustrated in this example. It should be noted, however, that slow convergence may occur in certain situations due to a large eigenspread of the fourth-order moments matrix M. The parameter adaptation curves of the normalised LMS implementation in Eq. (4.57) is shown in Fig. 4.5. A stepsize of $\tilde{\mu} = 1$ was used. After 1000 iterations the equaliser parameters converged to

$$\widehat{\theta} = [0.524, -0.316 + 0.367j, -0.067 + 0.327j, -0.001, 0.001]^T$$
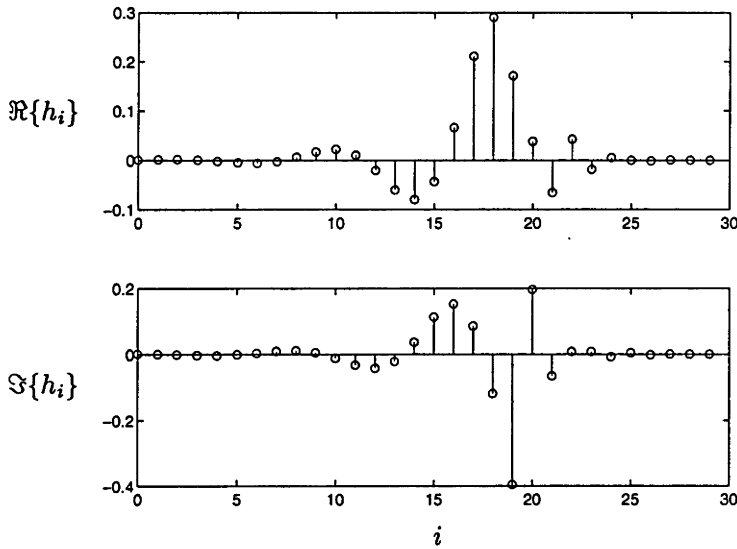
Figure 4.4 Real and imaginary parts of the channel impulse response used in the simulations.
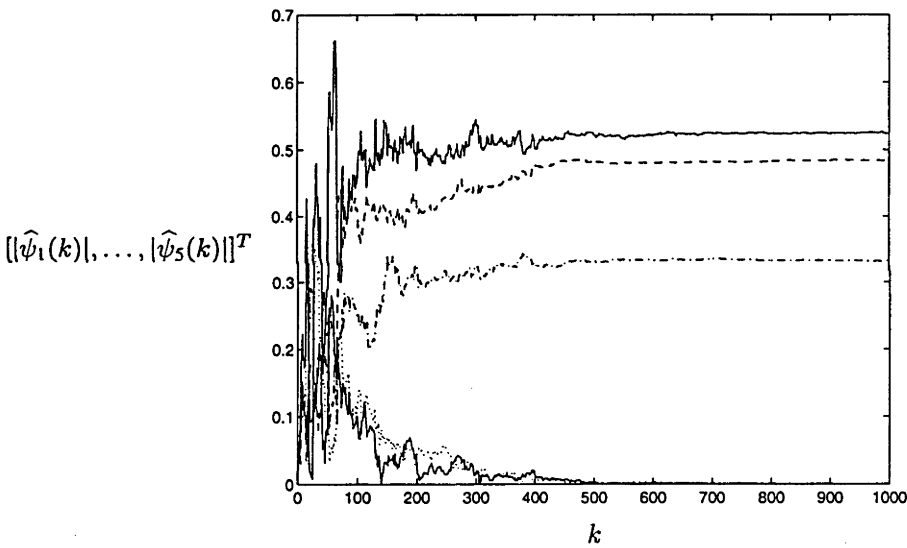
Figure 4.5 Parameter adaptation of the normalised LMS algorithm.

for which we get CLEM=0.031. The resulting equalisation gain and phase shift in Eq. (4.17) are $\Upsilon \approx 0.414$ and $\phi(k) \approx 0.398\pi$ radians, respectively. $\square$

Example 4.6 (Least Squares) In this example we use the off-line minimum-norm LS solution in Eq. (4.72) to estimate the equaliser parameters. A plot of the singular values of A constructed from a record of channel outputs of length $N + L - 1 = 54$ is shown in Fig. 4.6. As predicted by Proposition 4.1, the smallest two singular values of A are very close to zero since the equaliser is overparameterised by $L - P = 2$. For the observed channel output sequence, the equaliser parameters are given by

$$\widehat{\theta} = [0.576, -0.346 + 0.403j, -0.071 + 0.359j, 0.001j, 0.001]^T$$

Figure 4.6 Plot of the singular values of A. Note that A is *ill-conditioned* because $L > P$.
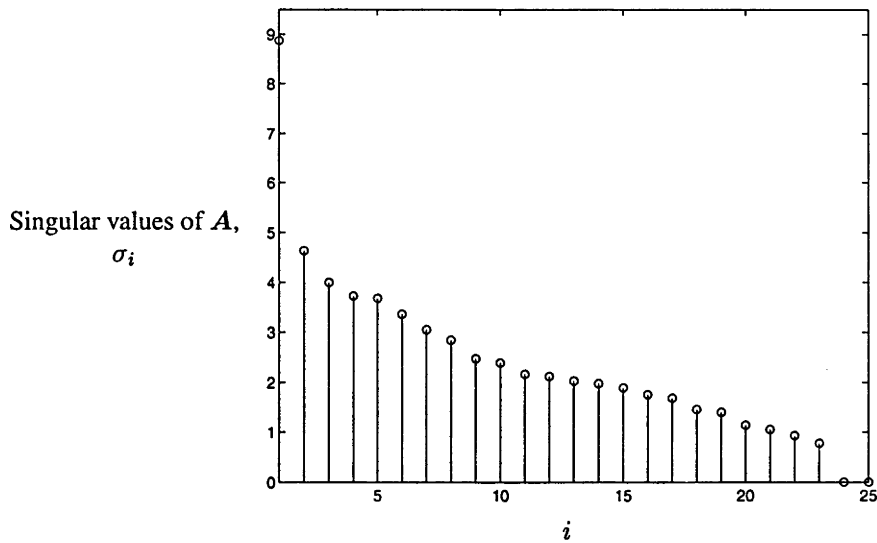
for which CLEM=0.024. The iterative scheme in Section 4.6.3 has also been applied with $\mu = 0.0111$. The sequence $Y(i)$ converged to $A^\dagger$ after 26 iterations. The resulting gain and phase shift are $\Upsilon \approx 0.455$ and $\phi(k) \approx 0.398\pi$ radians, respectively. $\qquad\qquad\qquad\square$

Example 4.7 (RLS Algorithms) The modified RLS algorithm described in Eqs. (4.74a) and (4.74b) has been simulated with $\alpha = 0.5$, $\beta = 10^{-3}$, $\lambda = 0.96$ and $\delta = 10^{-8}$, which results in the bounds $\bar{\sigma} = 2.20 \times 10^{-3}$ and $\bar{\nu} = 4.17 \times 10^6$ for the eigenvalues of $P(k)$. The matrix $P(k)$ is accordingly initialised to $P(0) = 10^3 I$. The equaliser parameter adaptation curves of the modified RLS algorithm are shown in Fig. 4.7. After 300 iterations the equaliser parameters converged to

$$\widehat{\theta} = [0.524, -0.315 + 0.368j, -0.066 + 0.327j, 0, 0]^T$$

which results in CLEM=0.018. The converged equalisation setting yields $\Upsilon \approx 0.414$ and $\phi(k) \approx 0.398\pi$ radians.

We have also implemented the ordinary RLS algorithm with the exponential weighting factor $\lambda = 0.96$ and the initial $P(k)$ set to $P(0) = 10^5 I$ under the assumption of exact equaliser parameterisation (i.e. $L = P = 3$). The adaptation of the equaliser parameters for the ordinary RLS is shown in Fig. 4.8. Note the fast convergence of the algorithm (which takes about $L^2 = 9$ iterations) in comparison with the other implementations. The algorithm converged to

$$\widehat{\theta} = [1.603, -0.960 + 1.119j, -0.201 + 1.001j]^T$$

with CLEM=0.018, $\Upsilon \approx 1.267$ and $\phi(k) \approx 0.398\pi$ radians. $\qquad\qquad\qquad\square$
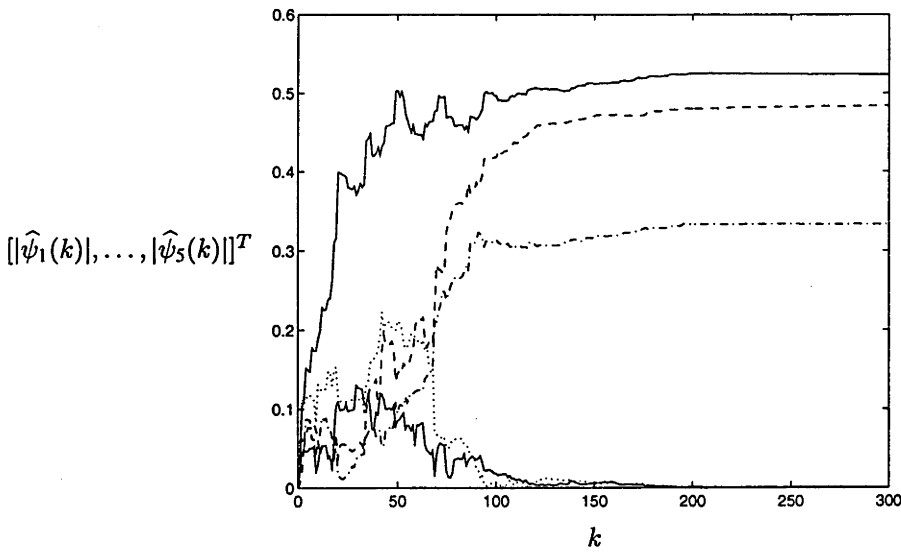
Figure 4.7 Parameter adaptation of the modified RLS algorithm.
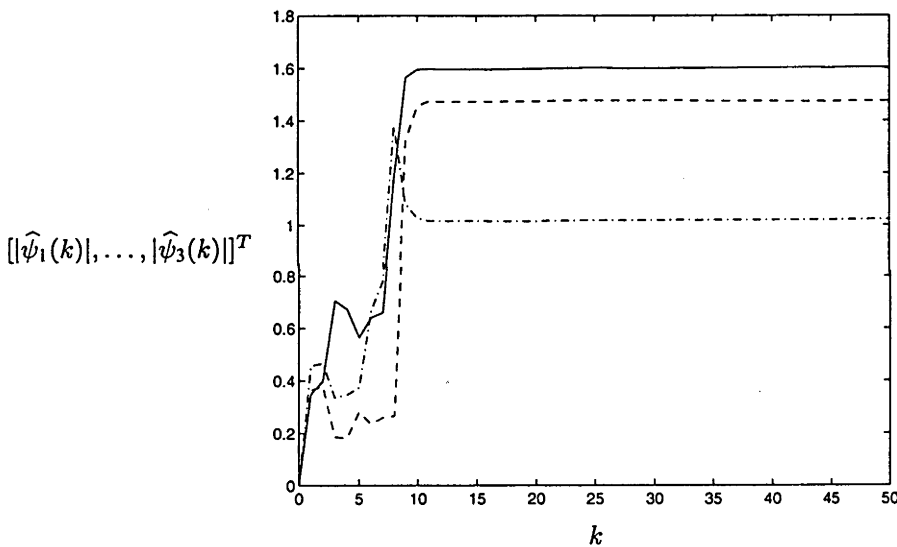
Figure 4.8 Parameter adaptation of the RLS algorithm with $L = P = 3$.

Example 4.8 (CMA) For DPSK modulation CMA takes the following form

$$\widehat{\theta}(k+1) = \widehat{\theta}(k) + \mu \left(1 - |y(k)|^2\right) y^*(k) r(k).   \tag{4.101}$$

Using a stepsize of $\mu = 0.01$ and centre-tap initialisation with $\widehat{\theta}(0) = [0, 0, 1, 0, 0]^T$, CMA converged to the following open-eye parameters with CLEM= 0.010

$$\widehat{\theta} = [0.004, -0.915 - 0.873j, 1.161 - 0.116j, 0.660 - 0.463j, 0]^T.$$

The equaliser parameter adaptation for this case is shown in Fig. 4.9(a). The resulting equalisation gain and phase shift are $\Upsilon \approx 1.0$ and $\phi(k) \approx -0.360\pi$ radians, respectively.

The initialisation $\widehat{\theta}(0) = [0, 0, 0, 0, 1]^T$ on the other hand led to an ill-convergence to the
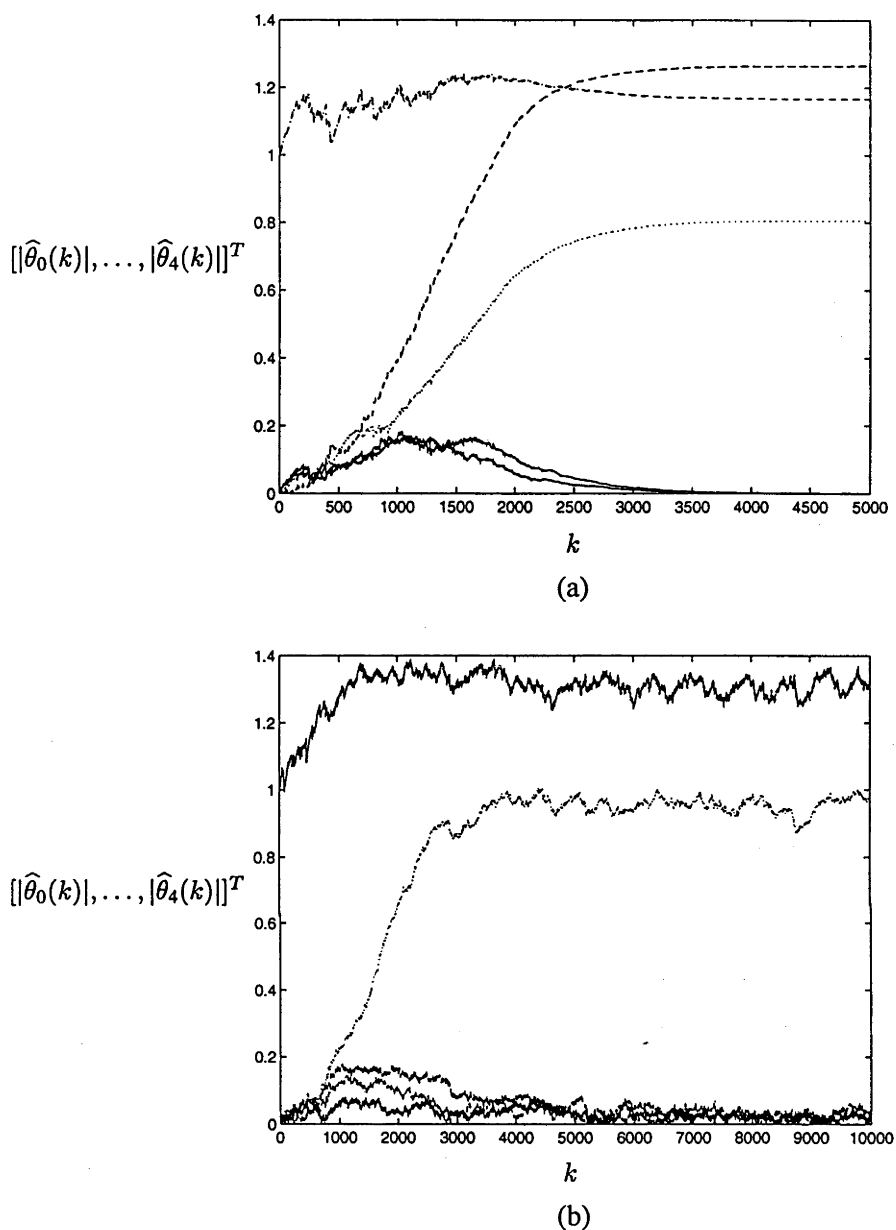
Figure 4.9 (a) Correct convergence and (b) ill-convergence of CMA.

following parameter setting with CLEM= 2.871 after 10000 iterations

$$\widehat{\theta} = [-0.010 + 0.024j, 0.012 + 0.016j, 0.032 - 0.027j, -0.694 - 0.695j, 1.326 + 0.007j]^{T}.$$

Fig. 4.9(b) shows the adaptation of the equaliser parameters resulting in ill-convergence. This example is included to illustrate the potential tendency of CMA to converge to closed-eye local minima depending on the initialisation. As we have pointed out before, the ill-convergence phenomenon is not relevant to the LS blind equalisation algorithm. □

Example 4.9 (Extension to PAM) In this example the communication channel is the all-pole

Equaliser parameters	$\widehat{\theta}_0$	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\theta}_3$	$\widehat{\theta}_4$
Sample mean	1.0000	−0.6014	0.4018	0.1960	0.0041
Sample stand. deviation	0.0000	0.0150	0.0165	0.0181	0.0170

Table 4.2 Sample mean and standard deviation of the equaliser parameter estimates for 8-PAM channel input after 100 trials.

system

$$H(z) = \frac{1}{1 - 0.6z^{-1} + 0.4z^{-2} + 0.2z^{-3}}$$

driven by an i.i.d. sequence drawn from the 8-PAM constellation $\mathbb{S} = \{\pm1, \pm3, \pm5, \pm7\}$. The solution of Eq. (4.89) was simulated after replacing the moments with their sample average estimates given by the normal equations with $L = 5$ and $N = 3000$. The sample mean and standard deviation of the equaliser parameter estimates for 100 trials are given in Table 4.2 after the normalisation of the first tap of $\Theta(z)$ to unity (i.e. $\widehat{\theta}_0 = 1$).   □

Example 4.10 (Channel noise) Consider the problem of equalising the channel in Eq. (4.99a) when its output is corrupted by additive complex Gaussian noise. We will use the LS solution in Eq. (4.72) with $L = 5$ (i.e. the equaliser is overparameterised) and $N = 100$. For 20dB SNR at the channel output, the resulting equaliser output is shown in Fig. 4.10 where $\Re\{\cdot\}$ and $\Im\{\cdot\}$ denote the real and imaginary parts, respectively. For the observed noisy channel output sequence, Eq. (4.72) produces the following equaliser parameters

$$\widehat{\theta} = [0.590, -0.336 + 0.398j, -0.053 + 0.366j, 0.036 - 0.016j, 0.017 + 0.037j]^T$$

for which CLEM=0.548. The algorithm seems to cope with small channel noise as suggested by Eq. (4.97). However, if the SNR at the channel output is 10dB or lower, the algorithm fails to return an equaliser parameter setting that opens the eye. Similar observations have been made with CMA when it is applied to noisy channels. The chief reason for this observed failure is the inappropriateness of the CMA criterion for noisy channels as pointed out in Section 4.8.   □

4.10   Conclusions and Discussion

In this chapter we have presented a novel least squares approach to blind channel equalisation that results in globally admissible algorithms in contrast to CMA. The resulting algorithm is parsimonious in the sense of the required number of channel output observations and insensitive to the channel input correlation subject to the proviso that all finite-length channel input subsequences occur with nonzero probability.
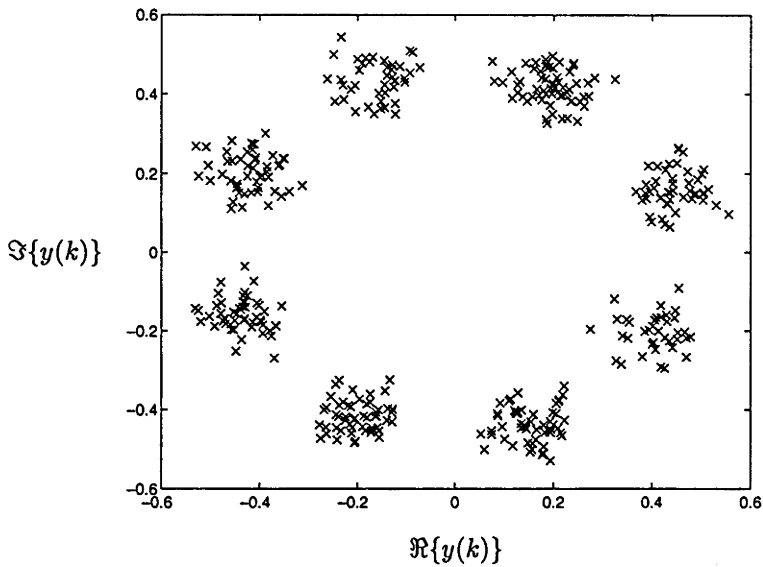
Figure 4.10 The equaliser output for SNR=20dB at the channel output. The LS solution is used to estimate the equaliser parameters.

The computational complexity of the off-line implementation can become formidable if a long equaliser is used because of a large matrix inversion involved. Stochastic gradient-based and recursive methods have been proposed based on the normalised LMS and the RLS algorithm, respectively, to overcome the matrix inversion problem. The computational complexity of the off-line LS algorithm is offset, to a large extent, by the requirement of relatively few channel output observations to compute the equaliser parameters, thereby justifying its viability in comparison with the "data-hungry" CMA which generally requires far longer channel output observations to converge to an open-eye or closed-eye minimum (see Example 4.8). Then, a question naturally arises as to the benefit of using an on-line method when it has a very slow convergence rate. Stochastic gradient-based algorithms using incremental parameter updates such as CMA generally have a slow convergence rate as they respond to local sets of data rather than averaging over longer data. The pitfalls of this local behaviour can be alleviated by means of smoothing. However, smoothing involves off-line or batch processing of data, which essentially takes away all the advantages associated with on-line computation. The on-line and recursive implementations discussed in this chapter seem to have a considerably faster convergence rate than CMA. This improvement in convergence speed is achieved at a computational complexity comparable with CMA and with no drawback of ill-convergence.

The algorithm is mainly applicable to communication systems using constant modulus channel input constellations, although possible modifications have been suggested to make it suitable for PAM systems. This should help increase the potential utility of the algorithm since digital

modulation techniques with nonconstant modulus constellations find common use in modern data communication systems.

Throughout the chapter we have assumed that the communication channel has an approximately FIR inverse. From the standpoint of practice, this may prove to be a serious restriction on the applicability of the algorithm. In addition, the algorithm may fail for some FIR channels that require long equalisers to achieve equalisation. Since the algorithm is geared towards perfect equalisation using an FIR equaliser, fractionally spaced equaliser implementation can be considered for arbitrary FIR channels. In the next chapter we tackle the problem of blind equalisation of FIR channels using the ideas presented in this chapter in the context of fractionally spaced equalisation.

Finally we note that a minimisation approach for Eq. (4.23) subject to the nonlinear constraint that the entries of ψ lie on the manifold in Eq. (4.25) reduces the algorithm to CMA. Such a constraint creates dependence between the entries of ψ by bringing down the number of "independent" entries to L. The constrained minimisation problem will therefore have the same cost function as CMA.

CHAPTER 5

Fractionally Spaced Blind Equalisation of FIR Channels

5.1 Introduction

Our objective in this chapter is to develop a fractionally spaced implementation of the blind equalisation algorithm presented in the previous chapter. We are motivated, in part, by the restricted applicability of the baud-rate least squares (LS) algorithm only to channels with approximately finite-duration impulse response (FIR) inverses. Further motivation comes from the fact that equalisers currently being built are fractionally spaced and an FIR system serves the purpose of modelling most of the practical channels in wireless communications better than an all-pole system. One of the attractive features of fractionally spaced equalisers (FSEs) is their ability to equalise oversampled FIR channels perfectly using a finite number of parameters. The same does not apply to baud-rate equalisers which typically require more taps than the channel length to approximate the channel inverse. This distinct ability of FSEs has been shown to depend on a certain *channel disparity condition* for the resulting subchannels of the oversampled communication channel (see e.g. [54]). One observed problem that results from an approximate loss of channel disparity is the channel noise enhancement which could degrade the performance of an FSE even when the channel noise is small.

The constant modulus algorithm (CMA), which is known to suffer from ill-convergence in baud-rate channel equalisation, has recently been applied to fractionally spaced blind equalisation with considerable success [18, 19, 50]. The effects of the channel input correlation, the channel disparity and the equaliser length on the performance of fractionally spaced CMA (FS-CMA) remain to be open problems, although some of them have been partially investigated with the aid
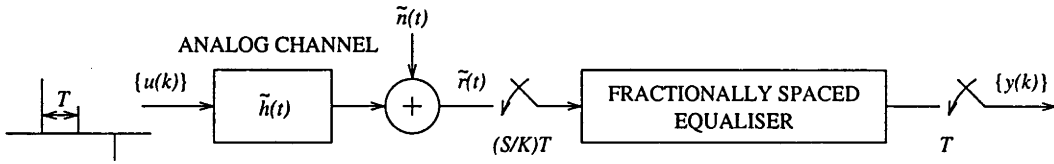
Figure 5.1 Fractionally spaced equalisation. K and S are positive integers such that $K > S$.

of computer simulations. In this chapter, we propose an alternative LS-based blind equalisation algorithm for FSEs. The proposed algorithm directly draws on the LS approach of the previous chapter and is not therefore affected by the channel input correlation so long as all finite-length subsequences of the channel input sequence occur with nonzero probability. The approximate loss of channel disparity is handled by setting small singular values to zero in the LS solution which results in a smaller equaliser norm, thereby alleviating the channel noise enhancement. The case of an overparameterised equaliser is dealt with in exactly the same way as in the previous chapter.

To start with, we give an overview of fractionally spaced equalisation of FIR channels and show how perfect equalisation is achieved using a finite number of parameters. In Section 5.3 we develop the new blind equalisation algorithm for fractionally spaced implementation both for constant modulus and PAM constellations. The effects of overparameterisation of the equaliser are also mentioned. The channel disparity condition, its implications in terms of the channel noise enhancement and an improved pseudoinversion technique to cope with it are discussed in Section 5.4. Simulation studies illustrating the application of the new algorithm and comparing it with FS-CMA are presented in Section 5.5. The chapter concludes with a brief summary and discussion of the results.

5.2 Feasibility of Fractionally Spaced Equalisers on FIR Channels

An FSE is shown in Fig. 5.1. The output of the analog channel $\tilde{r}(t)$ is oversampled at the rate of $(S/K)T$ and the equaliser output is downsampled to the symbol rate T. The positive integers K and S are chosen such that $K > S$. The equaliser taps are spaced at an interval of $(S/K)T$ which is less than, or a fraction of, the symbol interval T. This is in contrast to baud-rate equalisation in which the equaliser taps are spaced at T-second intervals. For the sake of notational convenience, we will assume $S = 1$. The integer K is typically chosen such that no aliasing occurs at the equaliser input. The FSE can combine the characteristics of a matched filter and a baud-rate equaliser, thereby synthesising the optimum receive filter in a linear modulation system [51]. The FSE has been shown to offer significant advantages in dealing with passband channels with markedly different band edge characteristics [51]. Its performance is also insensitive to the choice
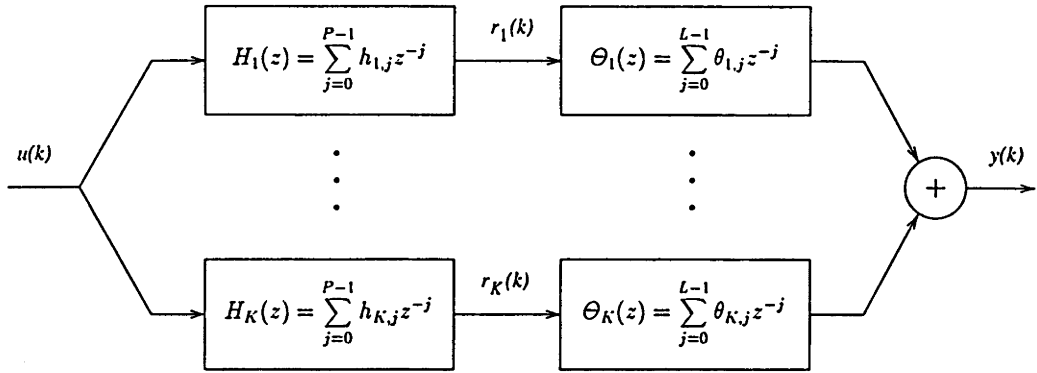
Figure 5.2 Multichannel representation for fractionally spaced equalisation.

of sampler phase. Although an FSE with the same number of taps as a baud-rate transversal filter has an impulse response that spans a shorter time interval, it has been shown to perform as well as, if not better than, a baud-rate equaliser for essentially every channel and significantly better than a baud-rate equaliser for channels with severe band-edge delay distortion [51].

The FSE shown in Fig. 5.1 can be considered in the framework of a multichannel T-spaced (baud-rate) model which is shown in Fig. 5.2. Assuming $S = 1$, the oversampled channel output sequence can be written as

$$\tilde{r}(kT/K) = \sum_{i=-\infty}^{+\infty} u(i)\tilde{h}(kT/K - iT - t_0) + \tilde{n}(kT/K) \tag{5.1}$$

where $\tilde{h}(t)$ is the baseband channel impulse response, t_0 is some delay accounting for the channel delay and the sampler phase, and $\tilde{n}(t)$ is the baseband channel noise. Throughout this chapter the channel noise will be assumed to be negligible. Then, assuming a finite support for $\tilde{h}(t)$ (i.e. $\tilde{h}(t) = 0$ if $t < 0$ or $t > t_d$ where $t_d \in \mathbb{R}^+$) the subchannels in Fig. 5.2 can be defined as follows

$$H_i(z) = \sum_{j=0}^{P-1} h_{i,j}z^{-j}, \quad i = 1, 2, \ldots, K. \tag{5.2}$$

The output of the ith subchannel is accordingly given by

$$r_i(k) = \sum_{j=0}^{P-1} h_{i,j}^* u(k - j), \quad i = 1, 2, \ldots, K. \tag{5.3}$$

The oversampled channel output sequence in Eq. (5.1) can be reconstructed by interleaving $r_i(k)$ for $i = 1, 2, \ldots, K$ at every discrete-time symbol interval k.

According to the multichannel representation of Fig. 5.2, the FSE consists of K subequalisers

each of length L

$$\Theta_i(z) = \sum_{j=0}^{L-1} \theta_{i,j} z^{-j}, \quad i = 1, 2, \ldots, K. \tag{5.4}$$

The sum of subequaliser outputs produces the FSE output

$$y(k) = \sum_{i=1}^{K} \sum_{j=0}^{L-1} \theta_{i,j}^* r_i(k-j). \tag{5.5}$$

The $KL \times 1$ regressor vector for the FSE is given by

$$r(k) = \left[r_1^T(k), r_2^T(k), \ldots, r_K^T(k) \right]^T \tag{5.6}$$

where r_i is the regressor vector for the subequaliser $\Theta_i(z)$ and is defined by

$$r_i(k) = \left[r_i(k), r_i(k-1), \ldots, r_i(k-L+1) \right]^T, \quad i = 1, 2, \ldots, K. \tag{5.7}$$

In the framework of Fig. 5.2, the equalisation problem can be posed as solving the polynomial equation

$$T(z) = [H_1(z), \ldots, H_K(z)] \begin{bmatrix} \Theta_1(z) \\ \vdots \\ \Theta_K(z) \end{bmatrix} = z^{-\Delta} \tag{5.8}$$

for the $\Theta_i(z)$, where $0 \leq \Delta \leq L + P - 2$ is the equalisation delay. Eq. (5.8) is a Bezout-type equation which has been studied extensively (see e.g. [52, 53]).

The overall impulse response from $u(k)$ to $y(k)$ can be written as

$$t = H\theta \tag{5.9}$$

where t is the $(L + P - 1) \times 1$ impulse response vector of $T(z)$, H is the $(L + P - 1) \times KL$ channel convolution matrix

$$H = \begin{bmatrix} h_{1,0} & 0 & h_{2,0} & 0 & & h_{K,0} & 0 \\ \vdots & \ddots & \vdots & \ddots & & \vdots & \ddots \\ h_{1,P-1} & \ddots & h_{1,0} & h_{2,P-1} & \ddots & h_{2,0} & \cdots & h_{K,P-1} & \ddots & h_{K,0} \\ & \ddots & \vdots & & \ddots & \vdots & & & \ddots & \vdots \\ 0 & & h_{1,P-1} & 0 & & h_{2,P-1} & \cdot & 0 & & h_{K,P-1} \end{bmatrix} \tag{5.10}$$

and $\theta = \left[\theta_1^T, \ldots, \theta_K^T \right]^T$ is the FSE parameter vector with

$$\theta_i = [\theta_{i,0}, \theta_{i,1}, \ldots, \theta_{i,L-1}]^T \tag{5.11}$$

denoting the impulse response of the subequaliser $\Theta_i(z)$. Note that H is a *block Toeplitz* matrix (also known as Sylvester matrix). Eq. (5.9) admits a solution for the FSE parameters θ for any given t, if (i) it is a square or underdetermined system (i.e. $KL \geq L + P - 1$ or equivalently $L \geq (P - 1)/(K - 1)$) and (ii) H has full row rank.

Lemma 5.1 [54] *If the subchannels $H_i(z)$ are such that*

    i. *$h_{i,0} \neq 0$ for some i in the range $1 \leq i \leq K$,*

    ii. *$h_{i,P-1} \neq 0$ for some i in the range $1 \leq i \leq K$,*

    iii. *the subequaliser length $L \geq (P - 1)/(K - 1)$*

and there are no zeros common to all the subchannels, then H has full row rank.

Lemma 5.2 [55] *If the length conditions i., ii. and iii. in Lemma 5.1 are met, the rank of the channel convolution matrix H is given by $L + P - 1 - Z_0$ where Z_0 is the number of zeros common to all the subchannels.*

If $L = (P - 1)/(K - 1)$ and the rank condition is satisfied, the FSE parameters are simply given by

$$\theta = \left(H^T H \right)^{-1} H^T [0, \ldots, 0, \underbrace{1}_{(\Delta+1)\text{th entry}}, 0, \ldots, 0]^T. \tag{5.12}$$

Remarks 5.1

    i. Conditions i., ii. and iii. in Lemma 5.1 are collectively called the *length conditions*. If the channel has an infinite-duration impulse response, then P should be sufficiently long.

    ii. The existence of no common zeros to all the subchannels is known as the *channel disparity condition*, and its violation renders perfect channel equalisation using a finite-length equaliser impossible as H will not have full row rank.

    iii. We have assumed zero channel noise. If the channel noise is not negligible, Eq. (5.8) should be modified.

    iv. It is sufficient to let $K = 2$ to achieve perfect equalisation for FIR channels.

v. Under the length and channel disparity conditions, the multichannel is minimum phase even if some of its subchannels may not be. Thus, methods based on second-order channel output statistics can be used to achieve perfect equalisation [55].

We now illustrate the length and channel disparity conditions with an example.

Example 5.1 Consider an FIR channel oversampled at the rate $T/2$ (i.e. $K = 2$), which has the following T-spaced subchannels representing the "odd" and "even" samples of the oversampled channel impulse response in the multichannel representation of Fig. 5.2

$$H_1(z) = h_{1,0} + h_{1,1}z^{-1} + h_{1,2}z^{-2}$$
$$H_2(z) = h_{2,0} + h_{2,1}z^{-2} + h_{2,2}z^{-2}.$$

Since $P = 3$, it is sufficient to let $L = 2$ to satisfy the length conditions so long as conditions **i.** and **ii.** of Lemma 5.1 are satisfied. Thus, Eq. (5.9) becomes

$$\underbrace{\begin{bmatrix} t_0 \\ t_1 \\ t_2 \\ t_3 \end{bmatrix}}_{t} = \underbrace{\begin{bmatrix} h_{1,0} & 0 & h_{2,0} & 0 \\ h_{1,1} & h_{1,0} & h_{2,1} & h_{2,0} \\ h_{1,2} & h_{1,1} & h_{2,2} & h_{2,1} \\ 0 & h_{1,2} & 0 & h_{2,2} \end{bmatrix}}_{H} \underbrace{\begin{bmatrix} \theta_{1,0} \\ \theta_{1,1} \\ \theta_{2,0} \\ \theta_{2,1} \end{bmatrix}}_{\theta} \qquad (5.13)$$

where H is a square matrix. If $L > 2$, Eq. (5.13) becomes underdetermined. In either case Eq. (5.13) is consistent, i.e. any given t on the left hand-side can be realised with a proper choice of θ, provided that H has full row rank. The full row rank condition for H is tantamount to the channel disparity condition (see Lemmas 5.1 and 5.2). Therefore, the oversampled channel convolution matrix H in Eq. (5.13) will have full rank if $H_1(z)$ and $H_2(z)$ have no common zeros. If the subchannels have common zeros, the FSE cannot achieve perfect equalisation. If $L = 1$, Eq. (5.13) is an overdetermined system, and therefore is not consistent for arbitrary t. In this case the FSE cannot be guaranteed to achieve perfect equalisation.

Consider the following channel

$$\begin{aligned} H_1(z) &= 1 + z^{-1} - 6z^{-2} \\ H_2(z) &= 1 - 4z^{-1} + 3z^{-2}. \end{aligned} \qquad (5.14)$$

for which Eq. (5.13) is a full-rank square system if $L = 2$, admitting the following solutions:

$$
\begin{aligned}
t &= [1,0,0,0]^T &\Rightarrow&\quad \theta = [19/8, -21/8, -11/8, -21/4]^T \\
t &= [0,1,0,0]^T &\Rightarrow&\quad \theta = [7/8, -9/8, -7/8, -9/4]^T \\
t &= [0,0,1,0]^T &\Rightarrow&\quad \theta = [3/8, -5/8, -3/8, -5/4]^T \\
t &= [0,0,0,1]^T &\Rightarrow&\quad \theta = [5/24, -11/24, -5/24, -7/12]^T.
\end{aligned}
\tag{5.15}
$$

If $L = 1$, Eq. (5.13) is not consistent for the candidate t vectors $[1,0,0]^T$, $[0,1,0]^T$ or $[0,0,1]^T$, as the 3×3 augmented matrix $[H, t]$ where

$$
H = \begin{bmatrix} 1 & 1 \\ 1 & -4 \\ -6 & 3 \end{bmatrix}
$$

will have rank$([H, t]) = 3$.

Suppose that oversampling of the continuous-time channel output results in the following subchannels with a common zero at $z = 2$

$$
H_1(z) = 1 + z^{-1} - 6z^{-2}
$$
$$
H_2(z) = 1 - 3z^{-1} + 2z^{-2}.
$$

In this case, for $L = 2$ the channel convolution matrix H will be rank deficient by one, i.e. rank$(H) = 3$ (see Lemma 5.2). If H is augmented with the candidate t vectors in Eq. (5.15), we find out that rank$([H, t]) = 4$, which implies that the t vectors in Eq. (5.15) are not in the column space of H and hence Eq. (5.13) is not consistent for any of them. Therefore, perfect equalisation by a finitely parameterised FSE is not plausible. Our interest in the consistency of Eq. (5.13) stems from the fact that if Eq. (5.13) is consistent, the channel is perfectly identifiable, otherwise it is not. □

5.3   A New Fractionally Spaced Blind Equalisation Algorithm

We will now apply the LS approach introduced in the previous chapter to FSEs. To start with, we present the explicit form of the blind equalisation algorithm for constant modulus channel inputs. Then, the use of the algorithm for PAM constellations will be explored after some modifications.

Before we go any further, a few words about FS-CMA will be in order. FS-CMA updates the equaliser parameters once per symbol using the recursion

$$
\widehat{\theta}(k+1) = \widehat{\theta}(k) + \mu r(k) y^*(k) \left(R - |y(k)|^2 \right)
\tag{5.16}
$$

where μ is a small positive stepsize, $r(k)$ is the fractionally spaced regressor vector and R is the CMA dispersion factor. Although all the stable stationary points of the FS-CMA cost function have been shown to be globally optimal under the length and channel disparity conditions, white channel input and zero channel noise, algorithm misbehaviours have been observed when the channel input is correlated [20]. As explained in the previous chapter, the major advantages of the LS approach are its relative insensitivity to correlated channel inputs and requirement of short channel output observations to compute the equaliser parameters. Rather than expounding various implementations of the LS approach, we will content ourselves with a development of the basic off-line LS algorithm and description of its properties. The off-line LS algorithm provides the basis for other recursive implementations.

5.3.1 Constant Modulus Channel Input Constellations

We will assume an FIR channel with T-spaced subchannels of length P and an oversampling rate $K = 2$. For constant modulus channel inputs which have the property that $|u(k)|$ is constant (assumed to be equal to unity) for all k, the FSE achieves its objective when its output has constant modulus for all k. This is of course dependent on the occurrence of all finite-length subsequences of the channel input with nonzero probability. Then, the equalisation objective is attained if

$$|y(k)|^2 = \left| r^T(k) \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}^* \right|^2 = 1, \quad \forall k \tag{5.17}$$

which can be rewritten for a finite time span $\{k-N+1, k-N+2, \ldots, k-1, k\}$ as the following matrix equation

$$A\psi = 1 \tag{5.18}$$

where the $N \times (2L)^2$ matrix $(N \geq (2L)^2)$ is defined in terms of the oversampled channel outputs as

$$A = \begin{bmatrix} r^T(k) \otimes r^H(k) \\ r^T(k-1) \otimes r^H(k-1) \\ \vdots \\ r^T(k-N+1) \otimes r^H(k-N+1) \end{bmatrix} \tag{5.19}$$

and the $(2L)^2 \times 1$ vector ψ of transformed FSE parameters is given by

$$\psi = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}^* \otimes \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}. \tag{5.20}$$

Unfortunately, Eq. (5.18) has multiple solutions due mainly to the freedom of selecting Δ
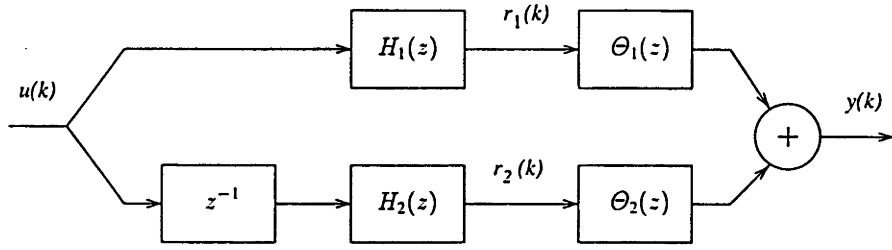
Figure 5.3 Modified multichannel representation.

in the range $0 \leq \Delta \leq L + P - 2$. This implies that the matrix A is rank deficient. Recall from the previous chapter that our ability to extract the equaliser parameters from any solution ψ to Eq. (5.18) depends on the structure of the multiple solutions. If the linearly independent solutions for the equaliser parameters can be written in such a way that only one of them has a nonzero first entry and the others all have zero first entries, then the FSE parameters θ_1 and θ_2 are extractable (refer to the proof of Proposition 4.2). It turns out that the manner in which the subchannel outputs are treated dictates whether or not the above-mentioned relation holds. As we have seen in Example 5.1, the FSE parameters for the multichannel representation in Fig. 5.2 do not obey this relation. To guarantee extractability of the FSE parameters from any solution of Eq. (5.18), we propose a slightly different multichannel structure to the one shown in Fig. 5.2. This structure involves delaying the second subchannel output by the symbol period T, so that the subchannel outputs are generated in proper synchronism with the subequalisers. According to this new structure whose benefit will become clear shortly, the regressor for $\Theta_2(z)$ is given by

$$r_2 = [r_2(k), r_2(k-1), \ldots, r_2(k-L+1)]^T \tag{5.21}$$

where (cf Eq. (5.3))

$$r_2(k) = \sum_{j=0}^{P-1} h_{2,j}^* u(k-1-j). \tag{5.22}$$

The resulting multichannel representation is shown in Fig. 5.3.

The $(L+P) \times 2L$ channel convolution matrix for the fractionally spaced equalisation problem

in Fig. 5.3 has the following form

$$
H = \begin{bmatrix}
h_{1,0} & & \mathbf{0} & 0 & \cdots & 0 \\
\vdots & \ddots & & h_{2,0} & \ddots & \vdots \\
h_{1,P-1} & \ddots & h_{1,0} & \vdots & \ddots & 0 \\
0 & \ddots & \vdots & h_{2,P-1} & \ddots & h_{2,0} \\
\vdots & \ddots & h_{1,P-1} & & \ddots & \vdots \\
0 & \cdots & 0 & \mathbf{0} & & h_{2,P-1}
\end{bmatrix}.
\tag{5.23}
$$

Observation 5.1 *Solutions of the matrix equation $t = H\theta$, where H is defined in Eq. (5.23) and*

$$
t = [0, \ldots, 0, \underbrace{1}_{(\Delta+1)\text{th entry}}, 0, \ldots, 0]^T, \quad 0 \le \Delta \le L + P - 1
\tag{5.24}
$$

have the following property

$$
\theta_{1,0} = \begin{cases} 1/h_{1,0} & \text{if } \Delta = 0 \\ 0 & \text{if } 0 < \Delta \le L + P - 1 \end{cases}
\tag{5.25}
$$

if $h_{1,0} \ne 0$ and $t = H\theta$ is consistent.

Proof. The first row in $t = H\theta$ can be written as

$$
h_{1,0}\theta_{1,0} = t_0.
\tag{5.26}
$$

Since $t = H\theta$ is consistent, the first entry of θ is

$$
\theta_{1,0} = \frac{t_0}{h_{1,0}}.
\tag{5.27}
$$

Eq. (5.25) follows straightforwardly from Eq. (5.27).   ∎

Remark 5.2 The condition $h_{1,0} \ne 0$, which is key to the extractability of θ from a solution of Eq. (5.18), can be dispensed with if the unit delay before $H_2(z)$ is replaced with z^{-P}. Then, θ can be extracted from ψ if the resulting $(L + 2P - 1) \times 2L$ H matrix has full row rank.

For the multichannel representation in Fig. 5.3 the length conditions can be listed as

i. $h_{1,0} \ne 0$ and $h_{2,P-1} \ne 0$,

ii. $L \ge P$.

Condition ii. becomes $L \ge P - 1$ if $h_{2,P-1} = 0$ and $h_{1,P-1} \ne 0$. The channel disparity condition remains unchanged since the multichannel representation in Fig. 5.3 can be obtained from the one

in Fig. 5.2 by increasing the length of the subequalisers by one, which allows for a unit delay to be placed on the path of the second subequaliser.

A solution to Eq. (5.18) can be found using the Moore-Penrose inverse of the rank deficient A

$$\psi = A^{\dagger} \mathbf{1}. \tag{5.28}$$

This will also yield the minimum-norm LS solution if Eq. (5.18) is not consistent due to the channel noise or the FIR approximation of the channel. Adoption of the channel model in Fig. 5.3 enables the extraction of θ from Eq. (5.28) using

$$\theta = c\,[\psi_1, \psi_2, \ldots, \psi_{2L}]^T \tag{5.29}$$

where $c = 1/\left(c_0\theta_{1,0}^{*}\right)$ and c_0 is the weight of ψ_0 in the following sum of linearly independent solutions ψ_i of Eq. (5.28)

$$\psi = \sum_{i=0}^{L+P-1} c_i\psi_i. \tag{5.30}$$

As will be explained in the next subsection, the maximum column rank of A is $(2L)^2 - L - P + 1$, which can be achieved if the equaliser is not overparameterised.

5.3.2   Overparameterised FSE

Eq. (5.18) is a rank deficient matrix equation even if the subequalisers have the same time span as the subchannels (i.e. $L = P$). The easiest way to see this is to consider the channel and the FSE solutions in Eqs. (5.14) and (5.15) in Example 5.1. The solutions in Eq. (5.15) are all linearly independent since they are the columns of a full-rank square matrix H^{-1}. Since Eq. (5.18) will be satisfied for any of the solutions, it is a rank deficient system. The matrix A can be seen to be rank deficient by the number of independent solutions minus one or, equivalently, the number of rows in H minus one (i.e. $L + P - 1$) provided the length and zero conditions are satisfied.

The experimental evidence suggests that under the length and zero conditions the rank of A is given by $(L + P)(L + P - 1) + 1$ if $L \geq P$. This rank relation clearly indicates that as the difference $L - P$ increases, the matrix A becomes more and more rank deficient relative to its number of columns.

If the conditions of Observation 5.1 hold, the FSE parameters can easily be obtained from any solution of Eq. (5.28) using Eq. (5.29) as explained in Proposition 4.2.

5.3.3 Extension to PAM Constellations

For nonconstant modulus constellations the algorithm needs to be modified as in Section 4.7. As a representative nonconstant modulus constellation we will consider pulse amplitude modulation (PAM). Since PAM is a one-dimensional (real) constellation, it is sufficient to consider the vector

$$v(k) = \left[p_1(k)p_1^T(k), p_2(k)p_2^T(k), \ldots, p_{2L}(k)p_{2L}^T(k) \right]^T \tag{5.31}$$

where $p_i(k)$ is the ith entry of the regressor $r(k)$ written in the form

$$r(k) = \begin{bmatrix} r_1(k) \\ r_2(k) \end{bmatrix} = [p_1(k), p_2(k), \ldots, p_{2L}(k)]^T$$

and the $(2L - i + 1) \times 1$ vector $p_i(k)$ is defined by

$$p_i(k) \triangleq [p_i(k), p_{i+1}(k), \ldots, p_{2L}(k)]^T, \quad i = 1, 2, \ldots, 2L.$$

In much the same way as in Section 4.7, the FSE parameters can be shown to be extractable from

$$\vartheta = E\left\{ u^2(k) \right\} \left(E\left\{ v(k)v(k)^T \right\} \right)^\dagger \mathbf{E}\left\{ \mathbf{v(k)} \right\} \tag{5.32}$$

using the relationship between the transformed parameters and the FSE parameters themselves

$$\theta = \frac{1}{c_0'\theta_0} \left[\vartheta_1, \frac{\vartheta_2}{2}, \ldots, \frac{\vartheta_{2L}}{2} \right]^T \tag{5.33}$$

where $c_0'\theta_0$ has to be identified using *a priori* knowledge of channel input statistics. Note that the matrix $E\left\{ v(k)v^T(k) \right\}$ is singular if $L > P$, hence the need for the Moore-Penrose inverse in Eq. (5.32) in case the equaliser is overparameterised.

5.4 Channel Disparity Considerations

If the subchannels $H_1(z)$ and $H_2(z)$ have common zeros, perfect equalisation cannot be attained by a finitely parameterised FSE. On the other hand, if the subchannels have zeros very close to each other, perfect equalisation is achievable under the length conditions, albeit at the expense of an increase in the norm of FSE parameters, resulting in a large noise enhancement at the equaliser output.

The loss of channel disparity is best dealt with by factoring out the common zeros, denoted $H_0(z)$, from the subchannels as shown in Fig. 5.4. This approach has been suggested in [50].
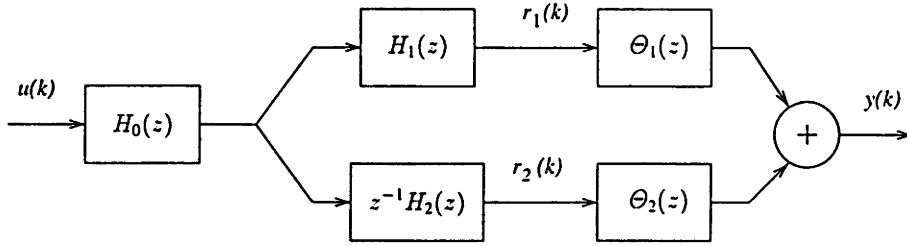
Figure 5.4 Multichannel representation with common subchannel zeros.

Now the task of the FSE is to achieve

$$H_0(z)\left(H_1(z)\Theta_1(z) + z^{-1}H_2(z)\Theta_2(z)\right) = z^{-\Delta} \tag{5.34}$$

or

$$\left(H_1(z)\Theta_1(z) + z^{-1}H_2(z)\Theta_2(z)\right) = z^{-\Delta}H_0^{-1}(z) \tag{5.35}$$

which reveals that unless $\Theta_1(z)$ and $\Theta_2(z)$ are of infinite length, as would be required in T-spaced equalisation, perfect equalisation is not achievable. Note that $H_0^{-1}(z)$ denotes the double-sided (noncausal) z-transform in anticipation of a nonminimum phase $H_0(z)$. However, if $L+P$ is large enough to approximate the (delayed) inverse of $H_0(z)$, Eq. (5.35) can be realised approximately. Thus, the presence of common zeros turns the fractionally spaced equalisation problem into a baud-rate equalisation problem. In this sense, Eq. (5.35) provides further insight into the zero condition that we have stipulated to achieve perfect equalisation. An important caveat is that FS-CMA may exhibit ill-convergence if $H_0(z) \neq 1$ because of the baud-rate nature of Eq. (5.35) [56].

If the subchannels have zeros that are very close but not strictly equal to each other, solutions to the consistent system of equations in Eq. (5.9) exhibit increased ℓ_2 norm. A large equaliser norm presents a particularly acute problem in the face of channel noise as it tends to amplify the noise variance at the equaliser output. This is illustrated in the following example.

Example 5.2 Consider the following channel

$$
\begin{aligned}
H_1(z) &= \left(1 + 0.8z^{-1}\right)\left(1 + (2+j)z^{-1}\right) \\
H_2(z) &= \left(1 + (0.8 + \epsilon)z^{-1}\right)\left(1 - j3z^{-1}\right)
\end{aligned}
\tag{5.36}
$$

where ϵ is a small parameter controlling the distance between the zeros of $H_1(z)$ and $H_2(z)$. Perfect equalisation can be achieved using the FSE as long as $\epsilon \neq 0$. One way of doing this is to solve Eq. (5.9) for θ. To comply with the formulation of the equalisation algorithm, H in Eq. (5.23) will be used as the channel convolution matrix in Eq. (5.9). For $L = 3$, which satisfies

ϵ	0.200	0.050	0.040	0.030	0.020
$\|\theta\|_2^2$	48.0	111.2	145.8	216.5	406.0
ϵ	0.010	0.008	0.005	0.002	0
$\|\theta\|_2^2$	1355.3	2044.2	4966.7	29495.9	6.6

Table 5.1 Squared ℓ_2 norm of FSE parameters as a function of ϵ.

the length conditions, H is a 6×6 square matrix. Thus, if $\epsilon \neq 0$, a perfect equalisation solution can be obtained from $\theta = H^{-1}t$. If $\epsilon = 0$, however, H is rank deficient by the degree of $H_0(z)$ (see Lemma 5.2) and, therefore, a minimum-norm LS solution should be sought using $\widehat{\theta} = H^{\dagger}t$. The vector t is the desired impulse response of the channel and equaliser combination. For various values of ϵ, the resulting squared ℓ_2 norm of the FSE parameters is tabulated in Table 5.1. Note that for $\epsilon = 0$, $\|\theta\|_2^2$ is the squared ℓ_2 norm of the minimum-norm LS solution.

Even if ϵ is very small, approximating it with zero changes the nature of the problem drastically. While perfect equalisation is attainable for $\epsilon \neq 0$, the equaliser norm increases as ϵ gets smaller due to ill-conditioning of H. On the other hand, if $\epsilon = 0$, perfect equalisation is not achievable, but can only be approximated with the aid of sufficiently long subequalisers. This time, however, the solution for FSE parameters does not incur an enormously large norm. Therefore, the trade-off between perfect equalisation with large norm and approximate equalisation with small norm should be considered carefully. In the presence of channel noise, the latter option sounds certainly better.

Consider the case of $\epsilon = 0.002$ in which the matrix H has the following singular values:

$$\sigma_1 = 7.0302, \sigma_2 = 4.8343, \sigma_3 = 3.3239, \sigma_4 = 2.1074, \sigma_5 = 0.3637, \sigma_6 = 0.0012.$$

Noting that σ_6 is very small, we can set it equal to zero to avoid numerical problems arising from ill-conditioning of H. Then, the minimum-norm LS solution gives an estimate for θ with squared ℓ_2 norm $\|\widehat{\theta}\|_2^2 = 6.6$. The resulting t which is very close to the t vector for the $\epsilon = 0$ case is plotted in Fig. 5.5. Clearly, a longer equaliser is needed to better approximate the inverse of the common zero $H_0(z)$.   □

In view of the conclusions reached in the above example, it may be necessary to set small singular values of A to zero if close subchannel zeros are detected. This way the equalisation problem is converted into the one shown in Fig. 5.4 where $H_0(z)$ has degree equal to the number of singular values set to zero. The advantage of doing this is to bring down the equaliser norm while avoiding numerical instabilities that may arise from ill-conditioning of A. Note that if H is ill-conditioned, so is A. Thus any increase in the norm of θ obtained from Eq. (5.9) directly
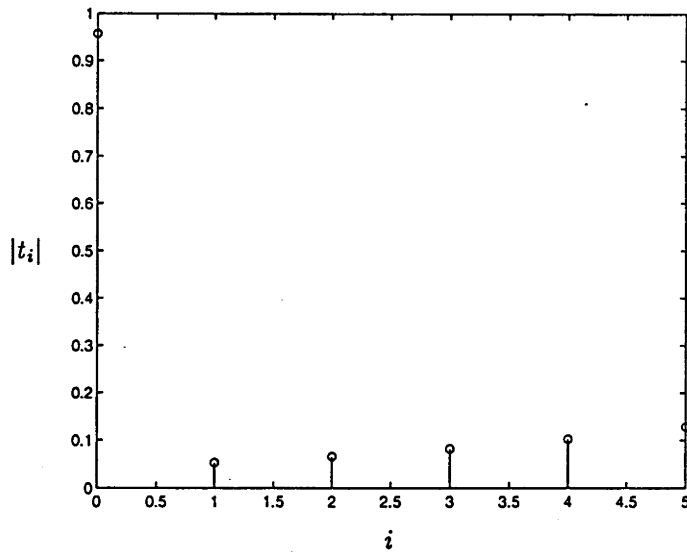
Figure 5.5 Overall impulse response for small-norm FSE parameters.

affects the norm of ψ.

5.5   Simulation Examples

In this section, we will demonstrate the new fractionally spaced blind equalisation algorithm with the help of computer simulations. The algorithm will be implemented using the modified recursive least squares algorithm in Section 4.6.4. The performance of FS-CMA will be compared to the new algorithm developed in this chapter. The extension to PAM inputs will be illustrated. The equaliser norm reduction ideas proposed in Example 5.2 will be extended to the solution of Eq. (5.18).

Example 5.3 In this example the communication channel will be assumed to have the following multichannel representation

$$
\begin{aligned}
H_1(z) &= 1 + (-2 + j)z^{-1} + (0.3 - j1.4)z^{-2} \\
H_2(z) &= -1.4 - j0.7 + (0.5 - j0.2)z^{-1} + 1.3z^{-2}
\end{aligned}
\tag{5.37}
$$

for which the channel disparity condition is preserved. The channel input is an i.i.d. sequence drawn from a 4-DPSK source which has the same constellation as quadrature phase shift keying (QPSK). We have implemented the fractionally spaced blind equalisation algorithm of Section 5.3 using the modified recursive least squares algorithm which avoids an explicit computation of the Moore-Penrose inverse. Since the matrix A is rank deficient even when $L = P$, the ordinary RLS algorithm cannot be used (see Sections 4.5.2 and 4.6.4). The parameters of the modified RLS algorithm were chosen as follows: $\alpha = 0.5$, $\beta = 10^{-3}$, $\lambda = 0.98$, $\delta = 10^{-8}$. This results in
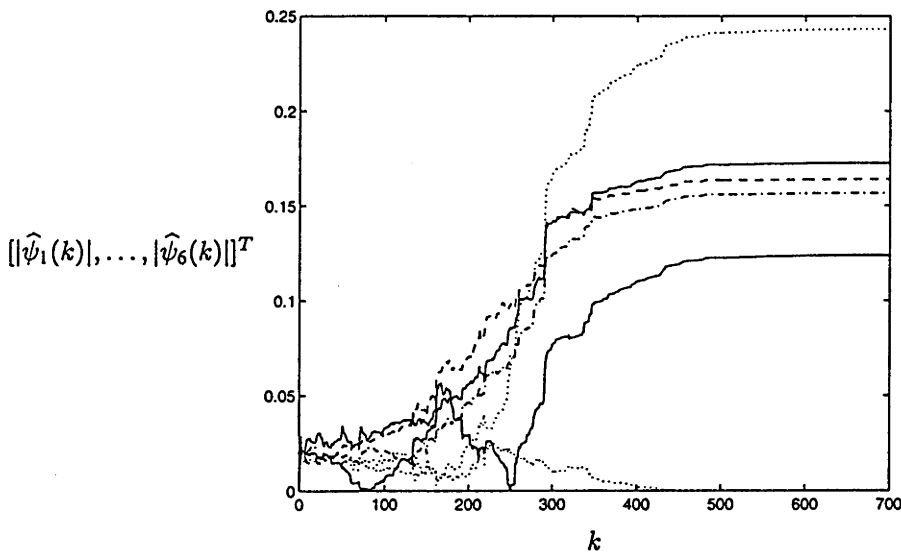
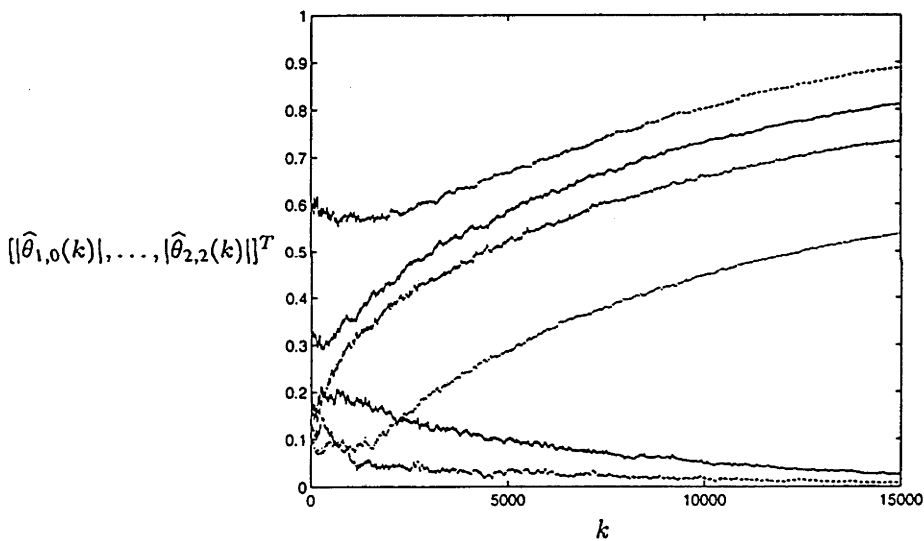**Figure 5.6** Parameter adaptation of the modified RLS algorithm.

Figure 5.7 Parameter adaptation of FS-CMA.

the bounds $\bar{\sigma} = 0.0021$ and $\bar{\nu} = 2.04 \times 10^6$ for the eigenvalues of $P(k)$. The matrix $P(k)$ was accordingly initialised to $P(0) = 50I$. The high upper bound ensures that $P(k)$ is reasonably close to its true value, thereby minimising inaccuracies in the resulting FSE parameter estimates. Adaptation of the FSE parameters is plotted in Fig. 5.6 for $L = 3$. Almost perfect equalisation is achieved with $\Delta = 0$ and CLEM ≈ 0 after 500 iterations.

We have simulated FS-CMA given in Eq. (5.16) on the same channel using a stepsize of $\mu = 0.003$ and the initial FSE parameters $\widehat{\theta}(0) = [0, 1, 0, 0, 0, 0]^T$. The resulting parameter adaptation curves are shown in Fig. 5.7. Note especially the comparatively slow convergence of FS-CMA. The stepsize was chosen in such a way that any noticeable increase in its value leads to an instability of FS-CMA. This ensures a convergence rate close to the maximum achievable.

Subequaliser parameters	$\widehat{\theta}_{1,0}$	$\widehat{\theta}_{1,1}$	$\widehat{\theta}_{1,2}$	$\widehat{\theta}_{1,3}$
Sample mean	1	−0.8327	−0.8897	0.3873
Sample stand. deviation	0	0.0075	0.0108	0.0069
Subequaliser parameters	$\widehat{\theta}_{2,0}$	$\widehat{\theta}_{2,1}$	$\widehat{\theta}_{2,2}$	$\widehat{\theta}_{2,3}$
Sample mean	−0.8414	−1.2875	0.5807	0.0017
Sample stand. deviation	0.0118	0.0080	0.0055	0.0087

Table 5.2 Sample mean and standard deviation of FSE parameter estimates after 50 trials.

Regarding the equalisation performance, after 15000 iterations FS-CMA converged to a setting with CLEM= 0.1059 and $\Delta = 2$. □

Example 5.4 Suppose that the channel has the following multichannel representation

$$
\begin{aligned}
H_1(z) &= 1 + 2.6z^{-1} + 4.5z^{-2} \\
H_2(z) &= 2.1 - 1.5z^{-1} - 3z^{-2}
\end{aligned}
\tag{5.38}
$$

which is driven by a 4-ary PAM sequence (i.e. $M = 4$). The channel disparity (or the zero condition) is satisfied. For $L = 4$, which satisfies the length conditions given $P = 3$, Eq. (5.32) has been simulated by replacing $\left(E\left\{ v(k)v(k)^T \right\} \right)^\dagger$ with its sample average $B^\dagger$ where B is the $N \times L(2L + 1)$ matrix defined in Eq. (4.40) which was adapted to the fractionally spaced equalisation case. The sample mean and standard deviation of the equaliser parameter estimates were estimated for $N = 2000$ and 50 independent trials and are listed in Table 5.2 after the normalisation of the first tap of $\Theta_1(z)$ to unity. The ISI measure for the averaged FSE parameters is CLEM=0.0424. □

Example 5.5 The purpose of this example is to illustrate the relationship between the singular values of H and those of A when the subchannels have close zeros. The channel in Example 5.2 is used with $\epsilon = 0.002$. The channel input is a 4-DPSK sequence. A subequaliser length of $L = 6$ will be assumed in order to provide a reasonable approximation for the inverse of $H_0(z)$ when the small eigenvalues of A are set to zero. For one realisation of the oversampled channel output, the smallest 15 nonzero eigenvalues of the 150×144 matrix A are plotted in Fig. 5.8. The squared ℓ_2 norm of the normalised FSE parameters is 2735.7. Although the rank of A is 73 when $\epsilon \neq 0$, setting $\epsilon = 0$ results in a rank reduction in both H and A. It turns out that if $\epsilon = 0$, we get rank$(A) = 64$. In the light of this observation, we set the singular values of A in the range σ_{65} to σ_{72} to zero before taking the Moore-Penrose inverse of A. Doing this introduces a loss of channel disparity, but produces FSE parameters with squared ℓ_2 norm $\|\widehat{\theta}\|_2^2 = 8.1$ after the normalisation of t_0 to unity. The FSE output for reduced norm equaliser parameters is plotted in Fig. 5.9. The ISI measure is CLEM=0.5051. □
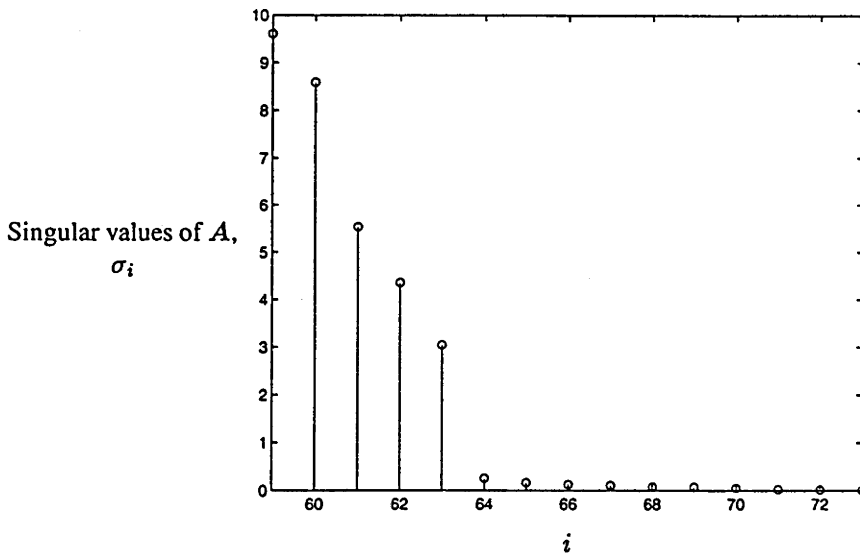
Figure 5.8 The smallest 15 nonzero singular values of A.
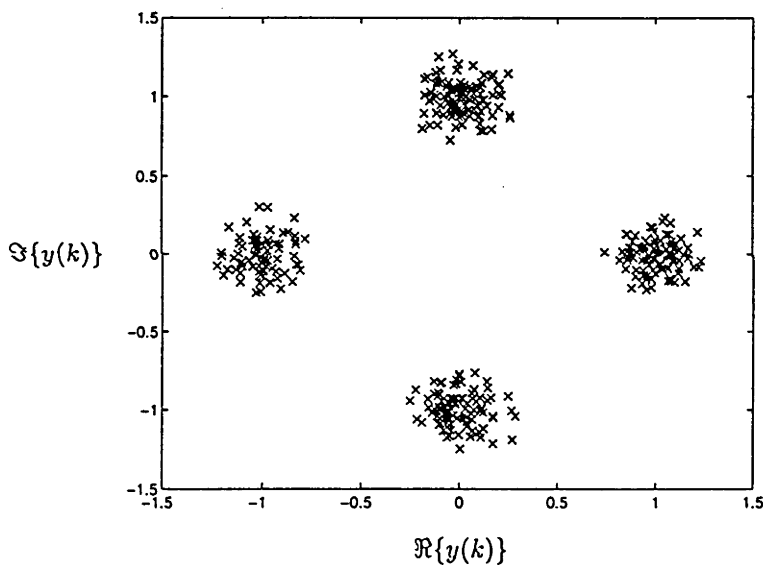
Figure 5.9 The FSE output for reduced norm equaliser parameters.

5.6 Conclusions

A new fractionally spaced blind equalisation algorithm has been presented. The algorithm enjoys most of the properties of the LS approach proposed in the previous chapter, since it is based on the same philosophy. The most important features of the algorithm are its independence of the channel input correlation under some mild richness condition and fast convergence compared with FS-CMA.

The channel disparity condition is investigated in relation to close subchannel zeros. The distinction between the exact and approximate loss of the channel disparity has been explored using numerical examples. It has been found out that forcing close subchannel zeros equal to each

other results in improvements in combatting with the channel noise enhancement at the expense of implicitly transforming the fractionally spaced equalisation problem to a baud-rate equalisation problem. These ideas were incorporated into the new algorithm when it is used as an off-line method. It should be noted that in practical channels an exact loss of the channel disparity is out of the question, hence the practical importance of reduced equaliser norm.

CHAPTER 6

Conclusions and Future Work

6.1 Conclusions

We have addressed the following topics in this thesis: (i) blind detection of equalisation errors which may result from ill-convergence of the adaptation algorithm used, excessive channel noise, etc., and (ii) blind equalisation of communication channels with no risk of ill-convergence. In this final chapter we will briefly outline the contributions of the thesis, and state some of the more interesting open problems that have been identified while working on the main research topics mentioned above.

In Chapters 2 and 3 the problem of blind detection of equalisation errors is cast into the binary hypothesis testing framework. While Chapter 2 is a natural extension of [4], Chapter 3 includes completely new results with wider application. Statistical threshold tests have been constructed to detect equaliser output decision errors, using only the observations available at the equaliser. The tests were shown to be (asymptotically) optimal in some sense. Although the test statistics were derived from formulas requiring off-line computation, alternative recursive and iterative methods for the computation of the test statistics were also obtained at the end of Chapter 3.

Chapters 4 and 5 develop a blind channel equalisation algorithm based on the constant modulus algorithm (CMA). A different equaliser parameterisation is used to transform the blind equalisation problem to a well-posed linear least squares problem. While Chapter 4 proposes a baud-rate implementation of the algorithm for channels with approximately FIR inverses, Chapter 5 considers the application of the algorithm to fractionally spaced equalisers which are shown to be capable of providing perfect equalisation for FIR channels under some mild conditions. The observed deficiency of global admissibility of current on-line blind equalisation algorithms with memoryless cost functions has motivated most of the material in Chapters 4 and 5. The developed algorithm

has the added advantage of fast convergence in terms of the required number of channel output observations, and is also indifferent to any channel input correlation so long as all finite-length subsequences of the channel input occur with nonzero probability, a requirement that has to be met by other on-line memoryless-cost-function blind equalisation algorithms, as well.

We should also point out the intriguing "duality" between the blind detection of equalisation errors and blind channel equalisation. This duality arises from the possibility of using the blind detection ideas as the basis for developing new blind equalisation algorithms which would not hopefully suffer from ill-convergence. We have not explicitly pursued this dual approach in Chapters 4 and 5. This notwithstanding, the particular test criterion in Chapter 2 can be considered for the purpose of designing new candidate algorithms. Since this is a major departure from the methodology used in this thesis, substantive research needs to be carried out into an investigation of this approach.

6.2   Future Work

Research reported in this thesis has uncovered a number of additional research problems some of which have proved to be quite challenging, while others are extensions of what has already been done. In this section we outline some of these research problems with a view to pursuing them in a future work.

6.2.1   Convergence Tests for Dependent PAM Sequences

In [4] it was shown that if the channel input is an i.i.d. M-ary sequence and the channel noise is sufficiently small with limited magnitude, then a necessary and sufficient condition for an LDDE to converge to an open-eye minimum is that the output of the decision device be an i.i.d. M-ary sequence. The extension of this criterion to the case of dependent M-ary channel input has turned out to be nontrivial. A convergence criterion that follows suit from the i.i.d. input case can be conjectured:

Conjecture 6.1 Let the channel input $\{u(k)\}$ be a dependent M-ary sequence (drawn from a PAM constellation set $\mathbb{S}$). Suppose that an LDDE is used to perform channel equalisation and the decision device output sequence $\{\hat{u}(k)\}$ is statistically independent of the channel noise $\{n(k)\}$. Then the eye is open if and only if the following is true:

$$\Pr\{\hat{u}(k+\tau) = s_1 \mid \hat{u}(k) = s_2\} = \Pr\{u(k+\tau) = s_1 \mid u(k) = s_2\} \quad \forall k, \tau \text{ and } \forall s_1, s_2 \in \mathbb{S}. \quad (6.1)$$

Although it would be useful to establish whether or not the above conjecture is indeed true, the result will not go a long way in view of its complexity. Besides, we have already developed an alternative, simple-to-implement criterion in Chapter 3 which is capable of handling the case of dependent M-ary channel inputs with no assumption on the channel noise magnitude.

We showed in Chapter 2 that the statistical dependence of symbols in an M-ary channel input sequence can be exploited to arrive at a very simple test for the convergence of an LDDE to an open-eye setting. The resultant criterion is, however, somewhat heuristic in that we were not able to rule out the existence of channel-equaliser combinations, if any, that may pass the test while leading to an eye closure. Therefore, a complete understanding of the heuristic test of variance matching is necessary. The construction of a statistical test based on the properties of variance estimators can be considered as a natural extension.

6.2.2 Sequential (On-line) Error Detection

Chapters 2 and 3 were devoted to the design of off-line tests which require block processing of the equaliser input and output observations to decide over whether or not the decision device output sequence is in error. We will now briefly discuss the possibility of sequential testing in which the observations are processed as they become available without imposing a fixed sample size on the observation length.

The RLS algorithm with exponential weighting can be considered in lieu of the method of least squares that was used in Chapter 3. The advantage of using RLS is that it lends itself readily to the on-line detection problem thanks to its recursive nature. The exponential weighting is essential since the test objective is to detect changes in the underlying model parameters with time. The basic idea is to analyse and characterise the probabilistic behaviour of the parameter updates generated by RLS. The detection test is then required to distinguish between the null hypothesis of no change in the parameters with time and the alternative hypothesis of time-varying parameters. Recalling from Chapter 3 that any time variation in the underlying linear model from the decision device output to the noisy channel output is directly linked to the presence of errors at the equaliser output, the test of equalisation errors is reduced to a test of time variation in the model parameters.

On-line processing of data is attractive mainly because of its lesser computational demand. Ideally, an on-line test should be able to detect equalisation errors with a small delay as they occur and with few false alarms. Sequential detection of abrupt changes in digital signals has been studied in the literature; see e.g. [32, 57, 58]. The general approach is to formulate some cumulative sum and compare it to a threshold, assuming that the model before and after the change is known. If that is not the case, the underlying model parameters are first estimated using different data

blocks and then a distance measure between the parameter estimates obtained from two different data blocks is used to decide whether or not a change in the model parameters has occurred. Here we pursue a completely different approach which is easier to build, but considerably more difficult to analyse.

The notation used in this section is based on the notation of Chapter 3, the only difference being the addition of time index k to matrix quantities in order to emphasise the recursive nature of the RLS algorithm. The objective of the RLS algorithm is to obtain the parameter estimates by minimising the weighted cumulative square error

$$\mathcal{J}_{\text{RWLS}}(v(k)) \triangleq \sum_{i=\Delta+1}^{k} \lambda^{k-i} \left(r(i - \Delta) - v^T(k)\widehat{u}(i) \right)^2 \tag{6.2}$$

where λ is the *exponential weighting factor* $(0.9 \leq \lambda < 1)$, $v(k) = [v_0(k), v_1(k), \ldots, v_{P-1}(k)]^T$ represents the unknown parameters of the underlying linear model from the decision device output to the equaliser input, and $\widehat{u}(k) = [\widehat{u}(k), \widehat{u}(k - 1), \ldots, \widehat{u}(k - P + 1)]^T$ is the decision device output vector. When λ equals one, the RLS algorithm becomes equivalent to the ordinary method of least squares. In order to be able to track time variations in the underlying model parameters, we will set λ to a value smaller than one. The RLS algorithm produces the following time update relation for the parameter estimates

$$\widehat{v}(k) = \widehat{v}(k - 1) + g(k)\alpha(k) \tag{6.3}$$

where $g(k)$ is the Kalman gain and $\alpha(k)$ is the innovation. After an initial transition period for the parameter estimates to settle down, we compute the squared ℓ_2 norm of the update term in Eq. (6.3) and use it as an on-line test statistic

$$T(k) = \|g(k)\alpha(k)\|_2^2. \tag{6.4}$$

As an example consider the channel $H(z) = 0.7 + 0.6z^{-1} + 0.5z^{-2}$ and a linear equaliser $\Theta(z)$ yielding CLEM = 0.4734. The channel input is an i.i.d. 4-ary PAM sequence and the channel output is corrupted by zero-mean white Gaussian noise with variance $\sigma_n^2 = 0.015$. Note that the presence of channel noise results in equalisation errors even though CLEM < 1. The on-line test statistic in Eq. (6.4) was simulated for $\lambda = 0.94$. In Fig. 6.1 the test statistic $T(k)$ is plotted along with equalisation error locations for some equaliser output observations. Note that equalisation errors can be detected by comparing $T(k)$ to an appropriately selected threshold. The resulting delay in detecting errors is of the order of a few samples. This on-line test has promising
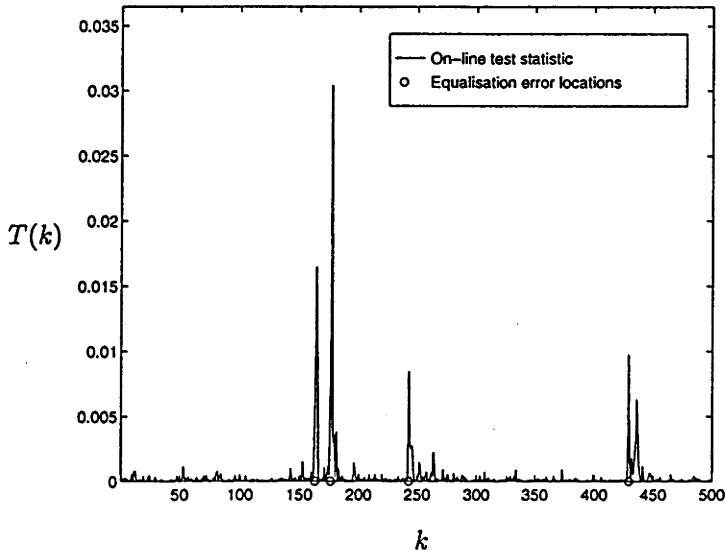
Figure 6.1 Plot of on-line test statistic $T(k)$ and equalisation error locations.

application in detecting rare equalisation errors in the presence of high channel noise.

There are two fundamental issues relating to the construction of an on-line test based on $T(k)$ or other similar test statistics. Firstly a test threshold has to be determined from whatever *a priori* information available about the hypotheses tested. To be able to compute the test threshold requires a knowledge of at least the null hypothesis probability density function of the test statistic in order for a threshold test to be designed under the assumption of a monotone likelihood ratio. Unfortunately, despite the simplicity of the on-line test in Eq. (6.4), the distribution of $T(k)$ under the null and alternative hypotheses is not easy to come by. Other test statistics may be considered to simplify the problem of computing a test threshold. The second issue is concerned with the detection performance analysis. It would certainly be useful to have some tentative idea about the probability density of the test statistic under the alternative hypothesis so that an estimate for the test power can be obtained with reference to the test threshold.

The material in [59] may be a starting point for further work although most of the simplifying assumptions made in that paper cannot be applied to the detection problem at hand.

6.2.3   *Nonparametric and Robust Testing of Equalisation Errors*

The statistical test in Chapter 2 assumes that the channel input autocorrelation is known. In Chapter 3 we assumed that the channel noise is additive Gaussian and a prior knowledge of the channel noise covariance is available. In many situations in practice, these assumptions may not be applicable. For instance, the channel noise may be impulsive and an exact knowledge of its covariance may not be available. The former situation calls for a nonparametric test in which a

test statistic is sought to produce a constant false alarm probability for a broad class of channel noise. If the noise covariance is subject to errors, then a robust test should be considered, which imposes limits on the likelihood ratio. These tests are somewhat beyond the scope of this thesis, but may well be taken into consideration in a future work.

6.2.4 Effects of Channel Input Correlation on Global Admissibility of CMA

We touched on the effects of channel input correlation on the convergence behaviour of CMA in Chapter 4. This interesting research problem has in fact many facets. For instance, its study can result in a better understanding of the CMA cost surface topology, not to mention practical design guidelines regarding channel input coding.

Initial motivation came from some experimental evidence that pointed to a drastic change in the CMA cost surface topology as a function of the channel input correlation. Assume that the channel input is a binary sequence generated by a Markov chain with state vector $m = [-1, 1]$ and one-step transition probability matrix of the form

$$\Pi = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}. \tag{6.5}$$

Assume further that the Markov chain states are initially equally likely. Obviously, the parameter p affects the channel input correlation. To illustrate the relationship between p and the global admissibility of CMA we have considered two channels. The first one was the one-pole channel $H(z) = 1/(1 - 0.6z^{-1})$ followed by a two-tap equaliser. The cubics $\mathcal{P}(0) = 0$ and $\mathcal{P}(1) = 0$ in the $(\theta_0(k), \theta_1(k))$ space, and the channel input autocorrelation are shown in Fig. 6.2 for $p = 0.60$ and $p = 0.65$, respectively. For $p = 0.60$ CMA is *not* globally admissible as its cost surface has four stable minima of which two correspond to closed-eye parameter settings. Setting $p = 0.65$, on the other hand, renders CMA globally admissible with two open-eye stable minima. The transition from inadmissibility to admissibility (and vice versa) occurs at roughly $p = 0.62$. For $0 < p < 0.62$ which includes the i.i.d. input case, CMA exhibits ill-convergence, whereas for $0.62 < p < 1$ CMA is globally admissible.

The second channel that was considered was an FIR system given by $H(z) = 1 - 0.8z^{-1} - 0.6z^{-2}$ which can be equalised (albeit not perfectly) again by a two-tap equaliser. The cubics and the resulting channel input autocorrelation are shown in Fig. 6.3 for $p = 0.45$ and $p = 0.30$, respectively. This time for $0 < p < 0.37$ CMA is globally admissible and for $0.37 < p < 1.00$ ill-convergence is observed due to the existence of two spurious closed-eye local minima in addition to open-eye local minima.
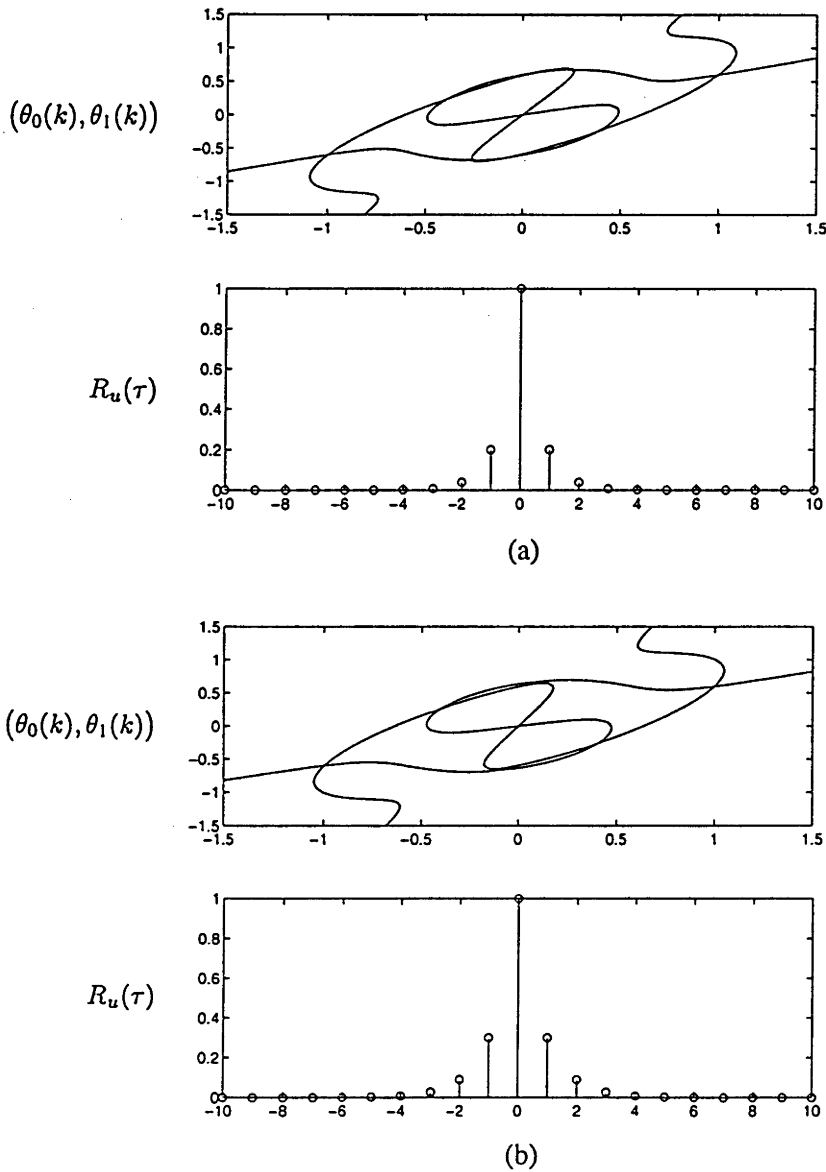
$(\theta_0(k), \theta_1(k))$

$R_u(\tau)$

(a)

$(\theta_0(k), \theta_1(k))$

$R_u(\tau)$

(b)

Figure 6.2 Plot of CMA cost surface cubics and channel input autocorrelation for (a) $p = 0.60$ and (b) $p = 0.65$.

It seems that there is some relationship between the form of the channel impulse response and the choice of p that guarantees global admissibility. Further research is needed to establish this relationship firmly.

6.2.5 Relocation of Closed-Eye Local Minima on the CMA Cost Surface

A more malicious application of the channel input correlation will be to find an input correlation for which the centre-tap initialisation is within the region of attraction of a closed-eye local minimum. Our ability to come up with such a correlation is reliant, to a large extent, on our understanding of how the CMA stationary points are relocated. Analytical tools from topology and algebraic geometry may be invoked to gain an insight into the behaviour of the CMA stationary points
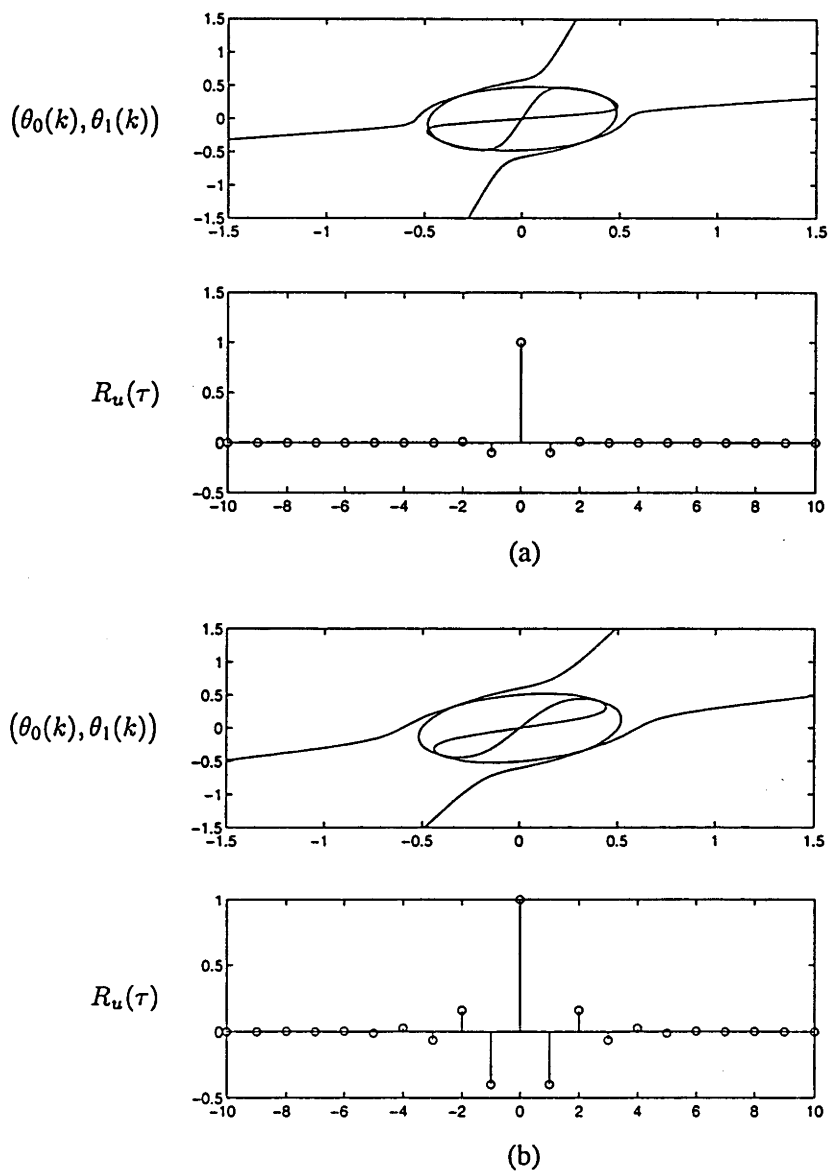
Figure 6.3 Plot of CMA cost surface cubics and channel input autocorrelation for (a) $p = 0.45$ and (b) $p = 0.30$.

obtained from the solutions of the set of nonlinear equations in Eq. (4.11).

Bibliography

[1] J. E. Mazo, "Analysis of decision-directed equalizer convergence," *Bell Syst. Tech. J.*, vol. 59, pp. 1857–1876, Dec. 1980.

[2] O. Macchi and E. Ewada, "Convergence analysis of self-adaptive equalizers," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 162–176, Mar. 1984.

[3] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, vol. I. New York: Wiley, 1969.

[4] R. A. Kennedy, G. Pulford, B. D. O. Anderson, and R. R. Bitmead, "When has a decision-directed equalizer converged?," *IEEE Trans. Commun.*, vol. COM-37, pp. 879–884, Aug. 1989.

[5] K. Doğançay and R. A. Kennedy, "Testing for parameter convergence in blind adaptive channel equalisation," in *Proc. Second Int. Workshop on Intelligent Signal Processing and Communication Systems, ISPACS '93*, (Sendai, Japan), pp. 1–6, Oct. 1993.

[6] K. Doğançay and R. A. Kennedy, "Testing for the convergence of a linear decision directed equaliser," *IEE Proc. Vision, Image and Signal Processing*, vol. 141, pp. 129–136, Apr. 1994.

[7] K. Doğançay and R. A. Kennedy, "Testing equalisation performance in blind adaptation," in *Proc. 33rd IEEE Conf. on Decision and Control, CDC '94*, (Florida, USA), pp. 2817–2818, Dec. 1994.

[8] Y. Sato, "A method of self-recovering equalization for multilevel amplitude modulation systems," *IEEE Trans. Commun.*, vol. COM-23, pp. 679–682, June 1975.

[9] A. Benveniste, M. Goursat, and G. Ruget, "Robust identification of a nonminimum phase system: blind adjustment of a linear equalizer in data communications," *IEEE Trans. Automat. Control*, vol. AC-25, pp. 385–399, June 1980.

[10] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. COM-28, pp. 1867–1875, Nov. 1980.

[11] J. R. Treichler and B. G. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 459–472, Apr. 1983.

[12] G. J. Foschini, "Equalizing without altering or detecting data," *AT&T Tech. J.*, vol. 64, pp. 1885–1911, Oct. 1985.

[13] Z. Ding, R. A. Kennedy, B. D. O. Anderson, and C. R. Johnson, Jr., "Ill-convergence of Godard blind equalizers in data communication systems," *IEEE Trans. Commun.*, vol. 39, pp. 1313–1327, Sept. 1991.

[14] C. R. Johnson, Jr., J. P. LeBlanc, and V. Krishnamurthy, "Godard blind equalizer misbehavior with correlated sources: two examples," *J. Marocain d'Automatique, d'Informatique et de Traitement du Signal*, pp. 1–39, June 1993.

[15] J. P. LeBlanc, K. Doğançay, R. A. Kennedy, and C. R. Johnson, Jr., "Effects of input data correlation on the convergence of CMA," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, ICASSP '94*, vol. III, (Adelaide, Australia), pp. 313–316, Apr. 1994.

[16] K. Doğançay and R. A. Kennedy, "A globally admissible off-line modulus restoral algorithm for low-order adaptive channel equalisers," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, ICASSP '94*, vol. III, (Adelaide, Australia), pp. 61–64, Apr. 1994.

[17] K. Doğançay and R. A. Kennedy, "Least squares approach to blind channel equalisation," submitted to *IEEE Trans. Commun.*

[18] I. Fijalkow, F. L. de Victoria, and C. R. Johnson, Jr., "Adaptive fractionally spaced blind equalization," in *Proc. 6th IEEE DSP Workshop*, (Yosemite), pp. 257–260, Oct. 1994.

[19] Y. Li and Z. Ding, "Global convergence of fractionally spaced Godard equalizers," in *Proc. 28th Asilomar Conf. Signals, Syst. Comput.*, Nov. 1994.

[20] J. P. LeBlanc, I. Fijalkow, B. Huber, and C. R. Johnson, Jr., "Fractionally spaced CMA equalizers under periodic and correlated inputs," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, ICASSP '95*, (Detroit, MI), May 1995.

[21] K. Doğançay and R. A. Kennedy, "Blind detection of equalisation errors in communication systems," submitted to *IEEE Trans. Inform. Theory*.

[22] J. A. Bucklew, T. G. Kurtz, and W. A. Sethares, "Weak convergence and local stability properties of fixed step size recursive algorithms," *IEEE Trans. Inform. Theory*, vol. IT-39, pp. 966–978, May 1993.

[23] M. B. Priestley, *Spectral Analysis and Time Series*, vol. I. London, UK: Academic Press, 1981.

[24] G. M. Jenkins and D. G. Watts, *Spectral Analysis and its Applications*. California: Holden-Day, 1968.

[25] I. A. Ibragimov, "A note on the central limit theorem for dependent random variables," *Theory Prob. Appl.*, vol. XX, no. 1, pp. 135–141, 1975.

[26] T. W. Anderson and A. M. Walker, "On the asymptotic distribution of the autocorrelations of a sample from a linear stochastic process," *Ann. Math. Stat.*, vol. 35, pp. 1296–1303, 1964.

[27] C. R. Rao and S. K. Mitra, *Generalized Inverse of Matrices and its Applications*. New York: Wiley, 1971.

[28] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge: Cambridge University Press, 1985.

[29] A. Ben-Israel and D. Cohen, "On iterative computation of generalized inverses and associated projections," *J. SIAM Numer. Anal.*, vol. 3, no. 3, pp. 410–419, 1966.

[30] R. D. DeGroat and E. M. Dowling, "The data least squares problem and channel equalization," *IEEE Trans. Signal Processing*, vol. 41, pp. 407–411, Jan. 1993.

[31] G. Marsaglia and G. P. H. Styan, "Equalities and inequalities for ranks of matrices," *Linear and Multilinear Algebra*, vol. 2, pp. 269–292, 1974.

[32] J. Segen and A. C. Sanderson, "Detecting change in a time-series," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 249–255, Mar. 1980.

[33] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Massachusetts: Addison-Wesley, 1991.

[34] E. L. Lehmann, *Testing Statistical Hypotheses*. New York: Chapman & Hall, 2nd ed., 1994.

[35] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, vol. 1. London: Charles Griffin, 4th ed., 1977.

[36] M. J. Hinich, "Testing for gaussianity and linearity of a stationary time series," *J. Time Series Anal.*, vol. 3, no. 3, pp. 169–176, 1982.

[37] R. A. Kennedy and B. D. O. Anderson, "Recovery times of decision feedback equalizers on noiseless channels," *IEEE Trans. Commun.*, vol. COM-36, pp. 1012–1021, Oct. 1987.

[38] R. A. Kennedy, B. D. O. Anderson, and R. R. Bitmead, "Blind adaptation of decision feedback equalisers: gross convergence properties," *Int. J. Adaptive Contr. and Signal Processing*, vol. 7, pp. 497–523, 1993.

[39] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore: Johns Hopkins University Press, 2nd ed., 1989.

[40] A. S. Householder, *The Theory of Matrices in Numerical Analysis*. New York: Blaisdell, 1964.

[41] C. R. Johnson, Jr., "Admissibility in blind adaptive channel equalization," *IEEE Contr. Syst. Mag.*, vol. 11, pp. 3–15, Jan. 1991.

[42] C. R. Johnson, Jr., S. Dasgupta, and W. A. Sethares, "Averaging analysis of local stability of a real constant modulus algorithm adaptive filter," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-36, pp. 900–910, June 1988.

[43] S. A. Alshebeili, A. N. Venetsanopoulos, and A. E. Çetin, "Cumulant based identification approaches for nonminimum phase FIR systems," *IEEE Trans. Signal Processing*, vol. 41, pp. 1576–1588, Apr. 1993.

[44] J. A. Cadzow, "Total least squares, matrix enhancement, and signal processing," *Digital Signal Processing*, vol. 4, pp. 21–39, 1994.

[45] S. Haykin, *Adaptive Filter Theory*. New Jersey: Prentice-Hall, 2nd ed., 1991.

[46] C. R. Johnson, Jr., *Lectures on Adaptive Parameter Estimation*. New Jersey: Prentice Hall, 1988.

[47] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge: Cambridge University Press, 1990.

[48] M. E. Salgado, G. C. Goodwin, and R. H. Middleton, "Modified least squares algorithm incorporating exponential resetting and foregetting," *Int. J. Control*, vol. 47, no. 2, pp. 477–491, 1988.

[49] E. H. Satorius and J. J. Mulligan, "An alternative methodology for blind equalization," *Digital Signal Processing*, no. 3, pp. 199–209, 1993.

[50] I. Fijalkow, J. R. Treichler, and C. R. Johnson, Jr., "Fractionally spaced blind equalization: loss of channel disparity," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, ICASSP '95*, (Detroit, MI), May 1995.

[51] S. U. H. Qureshi, "Adaptive equalization," *Proc. IEEE*, vol. 73, pp. 1349–1387, Sept. 1985.

[52] T. Kailath, *Linear Systems*. New Jersey: Prentice Hall, 1980.

[53] R. R. Bitmead, S.-Y. Kung, B. D. O. Anderson, and T. Kailath, "Greatest common divisors via generalized Sylvester and Bezout matrices," *IEEE Trans. Automat. Control*, vol. 23, no. 6, pp. 1043–1047, 1978.

[54] Y. Li and Z. Ding, "Blind channel identification based on second order cyclostationary statistics," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, ICASSP '93*, vol. IV, (Mineapolis, MN), pp. 81–84, Apr. 1993.

[55] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second order statistics: a time domain approach," *IEEE Trans. Inform. Theory*, vol. 40, pp. 340–349, Mar. 1994.

[56] I. Fijalkow, C. E. Manlove, and C. R. Johnson, Jr., "Adaptive fractionally spaced blind CMA equalization," submitted to *IEEE Trans. Signal Processing*.

[57] M. Basseville and A. Benveniste, "Sequential detection of abrupt changes in spectral characteristics of digital signals," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 709–724, Sept. 1983.

[58] M. Basseville and A. Benveniste, *Detection of Abrupt Changes in Signals and Dynamical Systems*. New York: Springer-Verlag, 1986.

[59] E. Eleftheriou and D. D. Falconer, "Tracking properties and steady-state performance of RLS adaptive filter algorithms," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1097–1110, Oct. 1986.

[60] Alle-Jan van der Veen and A. Paulraj, "Analytical solution to the constant modulus factorization problem," *Proc. 1994 Asilomar Conf. on Signals, Systems, and Computers*, pp. 1433–1437.

[61] J.-F. Cardoso, "Iterative techniques for blind source separation using only fourth order cumulants," *Proc. EUSIPICO-92*, pp. 739–742, vol. 2, 1992.