Adaptive Echo Cancellation in Telecommunications

John Homer

BSc (Hons I), University of Newcastle

April 1994

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

Department of Systems Engineering
Research School of Information Sciences and Engineering
The Australian National University

Declaration:

The contents of this thesis are the results of original research, and have not been submitted for a higher degree at any other university or institution.

A number of papers resulting from this work have been submitted to or are under preparation for submission to refereed journals:

[JS1] John Homer, Iven Mareels and Robert Bitmead, "Analysis and Control of the Signal Dependent Performance of Adaptive Echo Cancellers in 4-Wire Loop Telephony", accepted for publication, *I.E.E.E. Transactions on Circuits and Systems.*

JS2] John Homer, Robert Bitmead and Iven Mareels, "Quantifying the Effects of Dimension on the Convergence Rate of the LMS Adaptive FIR Estimator", submitted for publication, *I.E.E.E. Transactions on Information Theory.*

[JS3] John Homer and Iven Mareels, "Echo Canceller Performance Analysis in 4-Wire Loop Systems with Correlated AR Subscriber Signals", accepted for publication, *I.E.E.E. Transactions on Information Theory.*

[JS4 John Homer, Bo Wahlberg, Iven Mareels and Fredrik Gustafsson, "Adaptive Detection and Estimation for Acoustic Echo Cancellation", under preparation.

A number of papers have been presented at or have been submitted to conferences. Some of the material covered in these papers overlaps with that covered in the publications listed above:

[C1] John Homer, Iven Mareels and Robert Bitmead, "Signal Dependent Performance of Echo Cancellers", *Proceedings of International Symposium on Adaptive Systems in Control and Signal Processing (ACASP'92)*, 1-3 July 1992, Grenoble, France, pp. 621-626.

[C2] John Homer, Iven Mareels and Robert Bitmead, "Signal Conditioning for Adaptive Echo Cancellers", *Proceedings of The Third International Symposium on Signal Processing and its Applications (ISSPA'92)*, 16-21 August 1992, Gold Coast, Australia, pp. 545-548.

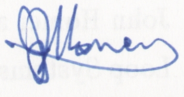
[C3] John Homer and Robert Bitmead, "Effects of Dimension on LMS Adaptive Filter Dynamics", *Proceedings of European Control Conference (ECC'93)*, 28 June-1 July 1993, Groningen, The Netherlands, pp. 2203-2206.

[C4] John Homer, "Curse of Dimension on the Learning Rate of the LMS Adaptive FIR Filter", *Proceedings of International Conference on Acoustic, Speech and Signal Processing* (ICASSP'94), 19-22 April 1994, Adelaide, Australia, Vol III pp. 405-408.

[C5] John Homer, Bo Wahlberg, Fredrik Gustafsson, Iven Mareels and Robert Bitmead, "LMS Estimation of Sparsely Parametrized Channels via Structure Detection", submitted to *IEEE Conference on Decision and Control* (CDC '94).

The work described in this thesis has been carried out in collaboration with a number of people. They are: Dr Robert R. Bitmead, Dr Iven M.Y. Mareels, Professor Bo Wahlberg and Dr Fredrik Gustafsson. However, the majority of the work is my own.

April 28, 1994



John Homer,
Department of Systems Engineering,
Research School of Information Sciences and Engineering,
The Australian National University,
CANBERRA ACT 0200, Australia.

Acknowledgements

I owe a great deal of thanks to my supervisors Bob Bitmead and Iven Mareels. They make a great supervisory team, having an incredible knack for steering the research into directions of fruition as well as providing heuristic insight and assistance with problems at hand.

I am also very grateful for Bob and Iven's encouragement. However, the greatest source of encouragement came from Kerry and for this I am very grateful. Without Kerry's love and support I would be still struggling through the PhD studies. I am also very grateful to Kerry for helping me relax and enjoy my leisure time - an important part of the last three years. To the Homers and Bradburys, I am also grateful for their love and support.

I am very grateful to the friends I have made during my PhD studies, for the friendship, support and good times they have given me.

I want to thank also Bo Wahlberg and Fredrik Gustafsson for working with me in the final stages of my PhD studies. The results of the research we carried out added significantly to my thesis. To Wolfgang Sauer, Vincent and Collette Wertz, Marc Moonen and Bob Stewart I am grateful for their hospitality and the opportunity they gave me to visit their respective Departments during my month long stay in Europe in 1993.

Finally I wish to thank the various people from CSIRO Radiophysics, Sydney, who have assisted me in my PhD studies, through discussion and provision of experimental data, as well as to acknowledge the funding of the activities of the Cooperative Research Centre for Robust and Adaptive Systems by the Australian Commonwealth Government under the Cooperative Research Centres Program.

Abstract

The focus of this thesis is the suppression of echoes within speech transmission telecommunication networks via the LMS adaptive FIR echo canceller. Poor performance of this technique has been reported, particularly when the required FIR tap length is large and the input signal is highly autocorrelated speech. The first aim of this thesis is to quantify the weaknesses of the LMS adaptive FIR filter, particularly the weaknesses relevant to echo cancellation. The second aim is to develop techniques which reduce the effects of these weaknesses and, consequently, enhance performance.

We begin with a brief review of alternatives to the LMS/FIR based echo canceller, which emphasizes that such alternatives are, in many ways, inferior. We then carry out a rigorous dynamical analysis of the LMS adaptive FIR filter connected in parallel with an unknown channel. We consider the case in which the adaptation stepsize μ is 'small'. The analyses focus on quantifying the adverse effects on transient performance of the autocorrelation level of the input signal and the filter parameter dimension (FIR tap length). The analytical results indicate conclusively that transient performance deteriorates with increasing filter dimension and input autocorrelation. A review of asymptotic analyses indicates that asymptotic performance also deteriorates with increasing filter dimension.

Dynamical analyses are then conducted on a system more representative of an echo cancellation network - a closed loop with an adaptive filter/unknown channel pair at each end and driven by signals entering from within the unknown channels. The analytical results indicate that the (transient and asymptotic) performance generally deteriorates with increasing filter dimension, while it is improved by either whitening the driving signals or whitening the input signal to each adaptive filter/unknown channel.

Motivated by the analytical results, we examine two types of signal conditioning methods. One involves using digital scramblers to whiten the subscriber signals (driving

signals) of the telecommunication network. The other assumes an autoregessive (AR) model of the input signals (typically used for speech) and whitens the input signals by filtering with AR estimates. To avoid distortion of received signals, indirect AR filtering methods are explored.

Finally, in a bid to tackle the adverse effect of dimension, we develop a low computational on-line technique which enables the detection of the nonzero or 'active' taps of an FIR modelled unknown channel. Based on this technique, a modified LMS adaptive FIR algorithm is proposed, which essentially estimates only the 'active' taps of the unknown channel. Such an 'active' tap detection-LMS estimation algorithm is particularly useful for echo cancellation since the (time-domain) impulse response of an echo path typically shows zero or 'inactive' regions.

Contents

Chapter 1

Introduction

An echo, in any context, is a delayed and perhaps distorted version of a previously transmitted signal. In telecommunication systems, the occurrence of echoes tends to reduce quality of transmission. In data transmission, echoes of sufficient magnitude result in an increase in error rates and may lead to retransmission being required. In speech transmission, the perceived quality reduction depends on the echo delay as well as the magnitude (and spectral distortion) [1]. In particular, if the echo delay is sufficiently short ($\approx 10ms$) then the echo is not noticeable and for this reason echo problems in speech transmission, in the past, were only encountered in international calls [1], particularly with satellite systems. The introduction and growing use of digital processing in speech transmission, however, has lead to the occurrence of delays in national networks comparable with those of international networks [2]. The occurrence of echoes with the potential to disrupt telephone conversations, and the need for echo suppression techniques has consequently grown considerably.

The need for the development of echo suppression techniques has received extra impetus with the advent of hands free telephony and teleconferencing. In particular, teleconferencing, which involves audio conferencing via the telephone network, has experienced increasing interest because of the considerable savings (time, money) it can provide in comparison to air travel to conference venues. These acoustic telecommunication systems are susceptible not only to electric echoes generated within the telephone circuit, but also to acoustic echoes, which arise due to acoustic coupling between microphone (transmitter) and loudspeaker (receiver) of the hands free telephone or teleconferencing system. The characteristics of acoustic echoes are somewhat different from those of circuit echoes, and, in some ways, are more difficult to suppress.

One approach to controlling echoes within telecommunication systems is to place an

attenuation device in each direction of transmission. Due to the nature of echo generation, this causes the echoes to be attenuated twice as much as the transmitted signals. This approach, however, causes unsatisfactorily low signal levels at the receiver for circuits longer than about 3000km [1]. A more sophisticated approach is to insert an attenuation device into the receiving or transmitting circuit according to which is least active. In speech transmission systems, this approach is known as voiced controlled switching. It was very popular for controlling circuit echoes [1], [2], [13], [6] and is currently the most commonly used technique for acoustic echo control [20], [6]. A major drawback of this technique is that during full duplex transmission - that is simultaneous transmission by the subscribers at both ends of the network - one of the transmitted signals is necessarily attenuated.

A popular alternative is that of echo cancellation which is illustrated in - Figure 1.1. This technique involves connecting a digital filter, the echo canceller, next to and in parallel with the echo source or echo path. The echo canceller samples the signal $u(k)$ feeding the echo path and, if correctly designed, outputs a signal $\hat{z}(k)$ which replicates the echo $z(k)$. This echo replica is then subtracted from the echo containing signal $v(k)$. To enable suppression of echoes generated by time varying echo paths, the echo canceller is typically adaptive. The use of adaptation also allows, to some extent, the echo canceller to be given arbitrary initial conditions, and thus reduces the costs due to initial echo path measuring.

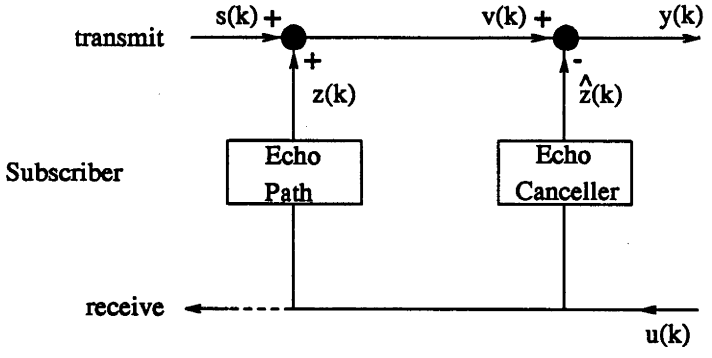


Figure 1.1: Echo suppression via echo cancellation.

The popularity of echo cancellation as a means of suppressing echoes in telecommunication networks mainly is due to its parallel configuration which leads to:

- negligible delay of the transmitted subscriber signals $s(k)$;
- full duplex transmission capability.

Of course, once the transients have died out, adaptive echo cancellation also leads to negligible distortion/attenuation of the transmitted subscriber signals.

The basic requirements of an adaptive echo canceller are:

- rapid suppression of echoes or, in other words, good transient and asymptotic performance;
- low computational complexity.

Usually, there is a trade-off between these two objectives. The meeting of these two objectives depends on a suitable choice of the following design factors:

- filter structure;
- adaptation algorithm.

Commercially made echo cancellers typically involve [3], [5], [6], [7] Finite Impulse Response (FIR) filters and the Least Mean Square (LMS) adaptive algorithm (or the closely related Normalized LMS (NLMS) algorithm). The major advantages of this echo canceller type are those of relatively low computational complexity, good stability properties, relatively good robustness against implementation errors [8] and that its behaviour is relatively easy to understand. Importantly, in a great many cases this popular echo canceller type provides adequate echo suppression with low computational complexity.

In a bid to achieve better echo cancellation under a greater range of circumstances, alternative echo cancellers to the conventional LMS adaptive FIR filter type have been and are being considered. These involve:

- modifications of the LMS adaptive FIR filter
- different filter structures and/or adaptation algorithms.

A review of the various types of echo cancellers is presented in Chapter 2.

Throughout the thesis we focus on suppressing echoes within speech transmission. In many ways, the analyses and echo cancelling techniques presented in this thesis can be also applied to the control of echoes in data transmission. However, there are shortcomings.

- The new techniques assume that the echo path is linear.
- Echo cancellation in data transmission typically is required to suppress a small but important nonlinear component (-40dB relative to that of the linear component) of the echo path. This nonlinear component, which is introduced by the circuit components of the digital transceiver [9], can be usually ignored in speech transmission. In data transmission, however, it must be considered because of the need for greater echo suppression [26].

The outline of the remainder of this introductory chapter is as follows. In Section 1.1 we give an introduction to the causes of echo, a description of the echo path characteristics and of the requirements of the echo canceller. We consider circuit echoes and acoustic echoes separately. In Section 1.2 we provide motivation for continued research into adaptive echo cancellation and for the approach taken in this thesis. An outline of the thesis is presented in Section 1.3 followed in Section 1.4 by a summary of our original contributions.

1.1 Echo Path Characteristics and Echo Canceller Requirements

1.1.1 Circuit Echoes

To understand how circuit echoes are generated and how they may be controlled consider the simplified telephone network of Figure 1.2. In this network the two subscribers are linked to a common central 4-wire loop facility via local subscriber 2-wire lines. The local 2-wire lines allow communication in either direction. In contrast, each of the 2-wire lines of the central 4-wire loop facility allow transmission in one direction only. Such unidirectional transmission enables the use of amplifiers and also of multiplexing *i.e* the sharing of one transmission channel by a number of calls. Generally, these advantages of unidirectional transmission are only necessary for long circuits and, consequently, for circuits shorter than about 60kms, the central 4-wire loop facility is not included [1].

The 4-wire loop facility is connected to the local subscriber 2-wire lines through a device known as a 2 to 4-wire hybrid. Ideally such devices permit all of the signal, transmitted by the subscriber at the far end, to pass into the connected local 2-wire line, as well as prevent any passage of the signal through to the opposite channel of the 4-wire loop. This is achievable by including in the hybrid, which is basically

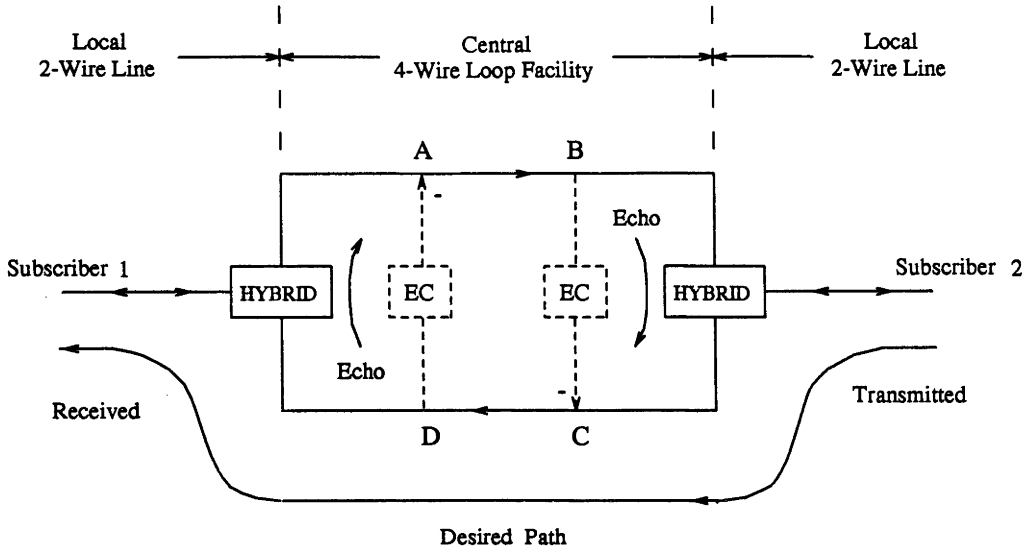


Figure 1.2: Central 4-wire loop telephone network

a circuit bridge, a balancing impedance which is set to match the input impedance of the connected 2-wire line. However, for multiplexing purposes, a hybrid may be connected to any one of a large number of different local 2-wire lines. Due to variations in the length, type and gauge of wire and number of phone extensions of the 2-wire lines connected [1] it is not economically possible to ensure perfect, or even near perfect, impedance matching [1], [2], [10]. As a result, the far end transmitted signal typically 'leaks' through the hybrid into the opposite 4-wire loop channel. This results in an echo being received by the far end (*i.e.* the original) subscriber - Figure 1.3a. Poorly terminated circuits may also lead to the receiving of echoes by the near end subscriber as indicated in Figure 1.3b and/or the observation of multiple echoes separated by time intervals equal to the round trip delay. It is important to add that because impedance matching of the distributed local 2-wire circuit is attempted with a lumped network, the echo is not just an attenuated, but a filtered version of the far end transmitted signal [1].

Depending on the delay of the echo, good quality communication may require the echo power level to be $50dB$ below the original voice level [1]. The attenuation provided by the hybrid on the echo signal leakage, however, can be as little as $10dB$ [6], [11]. Clearly, techniques which suppress the echo power by at least $40dB$ are required. The circuit echo path is typically well modelled by a linear filter [6]. Most circuit echo paths over the duration of a telephone call are time invariant [12], [13] and, more generally, are only slowly time varying [6], [11]. Furthermore, the impulse response of the echo path (through the hybrid near subscriber 1) from point D to point A in

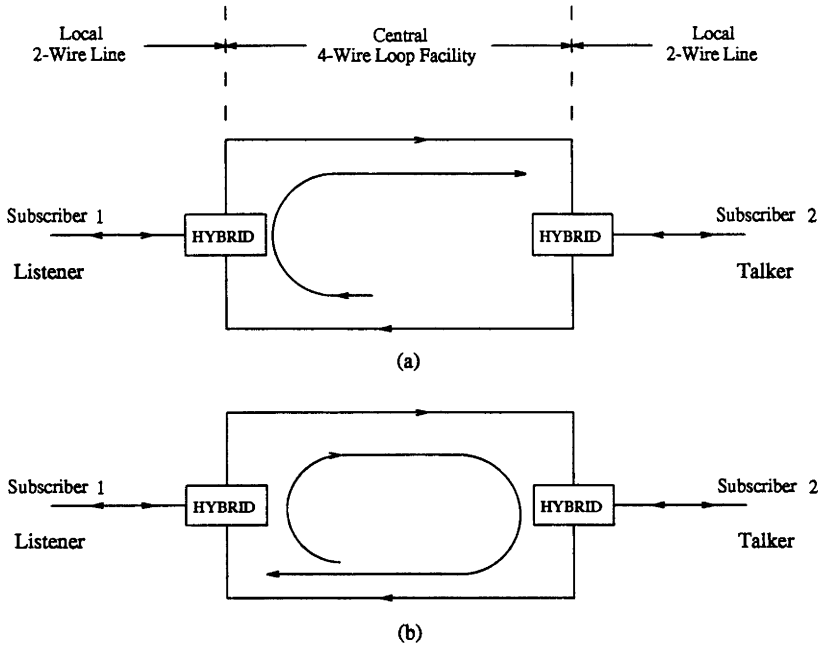


Figure 1.3: Echoes within the 4-wire loop telephone network:
(a) speaker echo, (b) listener echo

Figure 1.2 is typically only 15-16ms long [14], [5], [1] with the first 10ms typically being zero or 'flat' [1].

Combining the above characteristics of the circuit echo path with the fact that an 8kHz sampling rate is typically used for telephone speech processing, then adequate suppression ($40dB$) of circuit echoes via the technique of echo cancellation (Figure 1.1) should be achieved by using a digital filter having an impulse response length of 128 taps. To ensure adequate echo suppression for most telephone calls, however, a tap length of 256 taps is suggested [14]. So as to allow the echo canceller to have arbitrary initial conditions, the echo canceller is made adaptive. Since the echo path is time invariant or, at worst, slowly time varying then the algorithm used to adapt the adaptive echo canceller does not need to have particularly good tracking characteristics.

1.1.2 Acoustic Echoes

The typical setup for hands free telephony or teleconferencing, shown in Figure 1.4, involves two acoustic enclosures (*e.g.* the inside of a motor vehicle, or a conferencing room) connected via a telephone network. We will ignore the details of the telephone network and assume that any echoes being generated within it are adequately sup-

pressed. Each acoustic enclosure includes a loudspeaker for the receiving of signals transmitted from and a microphone for transmitting to the far end acoustic enclosure. The receiving of a signal at the loudspeaker results in, after some delay, an acoustic echo at the microphone via a direct path and paths involving reflections (on walls, furniture and persons). This acoustic coupling between loudspeaker and microphone is called the acoustic echo path.

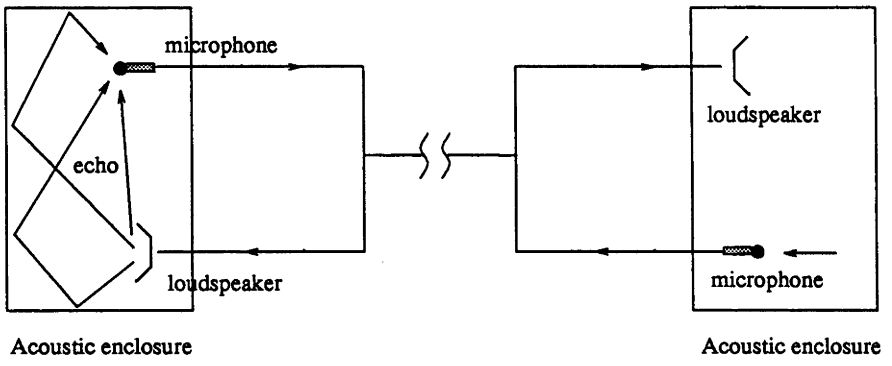


Figure 1.4: The typical setup for hands free telephony or teleconferencing

Due to the relatively low speed of acoustic signals in air, acoustic echo paths tend to have relatively long impulse responses compared to electric echo paths. The attenuation offered by acoustic echo paths is also relatively small. To be more specific: in hands free telephony, the impulse response is typically several tens of milliseconds while the attenuation may be $0dB$ or worse (that is amplifying rather than attenuation) [15], [16]; in teleconferencing the impulse response length tends to be several hundred ms [15] [13], [17], [18], [19] and the attenuation 6-$10dB$ [15], [19]. Furthermore, CCITT recommendations for hands free telephony, indicate that the acoustic echo should be suppressed to at least $45dB$ below the level of the original speech [15], [2]. In the case of teleconferencing, where high audio quality is sought, similar if not greater suppression is required.

To achieve such suppression the acoustic echo canceller needs to model essentially all of the impulse response of the acoustic echo path. At an 8kHz sampling rate, acoustic echo cancellers are therefore required to have the equivalent of an FIR tap length of several hundred to several thousand taps. This requirement is made worse in teleconferencing, which for quality reasons, uses a 16kHz sampling rate so as to to cope with wide band speech (7kHz bandwidth) [13], [18]. In addition, acoustic echo path impulse responses tend to be time varying because of movements of people and objects within the enclosure [13], [17], [5].

A summary of the echo canceller requirements for suppression of circuit and acoustic echoes is given in Table 1.1.

Table 1.1: Echo Canceller requirements

Characteristic	Circuit	Acoustic
Sampling Rate	$8kHz$	$8kHz$, prefer $16kHz$
Impulse Response Tap Length	200-300	200-4000 at $(8kHz)$ 400-8000 at $(16kHz)$
Tracking Requirement	Negligible	Important

1.2 Motivation for Research and Thesis Approach

Adaptive echo cancellation based on the LMS (or NLMS) adaptive FIR filter is now the preferred method for suppressing circuit echoes in 4-wire loop telephony. Typically this approach yields relatively fast, essentially complete echo suppression with relatively low computational complexity. However, occasional poor performance such as signal bursting and/or slow, incomplete echo suppression has been observed [21], [22], [23], [24], [25]. To avoid such undesirable behaviour, there is a need to quantify the causes of such behaviour so as to enable the development of performance enhancing techniques for the LMS/NLMS adaptive FIR echo canceller or, alternatively, to motivate the use of more sophisticated echo cancelling techniques.

In contrast, the use of the LMS/NLMS adaptive FIR echo canceller for echo suppression in acoustic telecommunication systems has not, in general, found a lot of success. This is due to the need for 'long' FIR filters in order to model acoustic echo paths adequately so as to achieve adequate asymptotic echo suppression. This requirement leads to high computational complexity and has been reported to lead to poor transient performance [20], [26], [27].

As a result of these problems, alternative echo cancelling techniques have been and are being examined. Most of these techniques, however, tend either to:

- improve transient performance at the expense of greatly increased computational complexity;
- reduce computational complexity but worsen transient performance;
- introduce signal distortion or delay in return for improved transient performance and/or reduced computational complexity.

To date, the acoustic echo cancellation techniques developed do not provide an implementable 'satisfactory' solution [20], [4].

Possible directions for research into improving echo suppression within acoustic telecommunication systems include:

1. exploring alternatives to echo cancellation;
2. continuing to explore alternatives to LMS/NLMS adaptive FIR echo cancellation;
3. developing performance enhancing techniques for LMS/NLMS adaptive FIR echo cancellation.

The first direction was the focus of research prior to the development of echo cancellation. The more successful alternatives (see [20] for a greater range of such alternatives) and which are still in use today include: (a) voiced controlled switching and (b) studio environments with highly directional loudspeakers and microphones and sound absorbing materials [20]. There are obvious inadequacies with both of these alternatives. As mentioned previously, alternative (a), during periods of double talk, (*i.e.* speech is being transmitted by both acoustic enclosures) causes significant attenuation of one of the speech signals. On the other hand, alternative (b) is not very practical for most situations and, generally, requires minimal movement of the talker(s) in each acoustic enclosure.

The second possible direction of research is still the main focus of research today. The third approach, however, is that which we follow throughout this thesis. This decision is motivated by the following reasons.

- The LMS/NLMS adaptive FIR filter is the most popular adaptive estimation technique [8], [28] and, is likely to remain so in the foreseeable future.
- The dynamics of the LMS/NLMS adaptive FIR filter are relatively 'simple' and, generally, considerably simpler than those of the alternative echo canceller types. This may allow the 'weaknesses' of the LMS/NLMS adaptive FIR filter to be quantified through dynamical analysis and, in turn, enable possible performance enhancing modifications to be developed and analysed.

In this thesis we focus on the LMS rather than the NLMS adaptive FIR filter. We do this firstly because the LMS algorithm is (slightly) easier to analyse and secondly

because, the adaptation stepsize, μ, used in practice is typically 'small' and under this condition, the LMS and NLMS algorithms show similar dynamical behaviour [29], [8]. Motivated by the above discussion, the aims of the work leading to this thesis have been as follows.

1. Through dynamical analysis attempt to quantify the weaknesses of the LMS adaptive FIR filter for the case in which the adaptation stepsize, μ, is 'small'. This needs to be carried out not only for the open loop case of Figure 1.5, in which we consider an isolated adaptive filter/unknown channel pair but also for the closed loop case of Figure 1.6 which is more representative of echo cancellation networks.

2. Based on these weaknesses and the characteristics of echo cancellation networks, develop performance enhancing techniques for the LMS adaptive FIR echo canceller

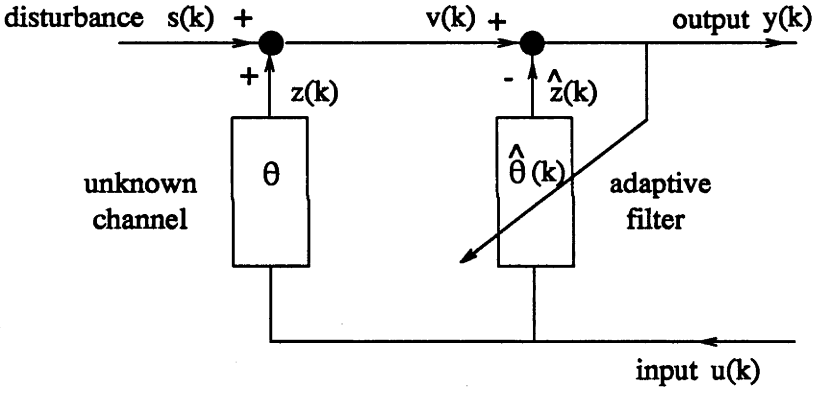


Figure 1.5: Open loop adaptive system - adaptive filter in parallel with unknown channel

1.3 Thesis Outline

Chapter 2

In this chapter we present a review of the different adaptive filtering techniques which have been proposed for echo cancellation. The main aim of this review is to highlight the advantages and disadvantages of each technique. In addition to the LMS adaptive FIR filter, we consider the Normalized LMS adaptive FIR filter, the RLS adaptive technique, lattice filters, IIR filters, frequency domain filtering and sub-band filtering. The chapter is concluded with a brief review of non-standard approaches to adaptive

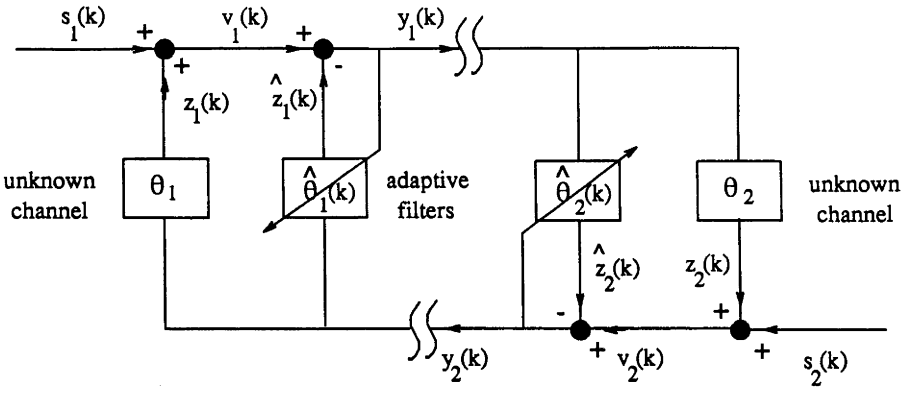


Figure 1.6: Closed loop adaptive system - an adaptive filter/unknown channel pair at each end of a closed loop

echo cancellation. The majority of these involve modifications to the LMS/FIR echo canceller.

Chapter 3

In this chapter we carry out quantitative analyses on the LMS adaptive FIR filter in the open loop configuration of Figure 1.5. We assume the tap length of the adaptive filter matches that of the unknown channel. Due to the existence of a number of useful results on asymptotic performance, we restrict our analysis to transient performance. We begin by developing a cost function which provides a quantitative measure of the convergence rate of the LMS adaptive FIR filter. Analysis of this cost function is then carried out for the case in which the LMS adaptation stepsize, μ is fixed, irrespective of filter tap length or signal characteristics. Particular attention is given to input signals, such as speech, which are well modelled as autoregressive processes. We conclude the analysis by considering the case in which μ is adjusted to maintain asymptotic performance. In short, the analyses quantify the adverse effects of (i) high autocorrelation levels of the signals input to the adaptive filter and (ii) large dimensions or FIR tap lengths of the adaptive filter.

Chapter 4

In this chapter we carry out dynamical analyses for the closed loop configuration of Figure 1.6. We begin by carrying out a semi-formal analysis on closed loop systems having echo paths/cancellers of arbitrary dimension. This provides a semi-quantitative understanding of the effects of the signal characteristics and filter dimension on performance. After invoking a number of simplifying assumptions, we then conduct quantitative analyses. The results indicate that an increase in the correlation levels within and between the driving signals (s_1, s_2 of Figure 1.6) leads to a deteri-

oration in the transient and/or asymptotic performance of the double adaptive filter closed loop system. An increase in dimension of the adaptive filters accentuates this effect. Furthermore, when the channels, which link the unknown channel/adaptive filter pairs, impose a sufficiently long delay, the closed loop system dynamics simplify to the dynamics of a pair of uncoupled open loop systems.

Chapter 5

The adverse effects of subscriber signal autocorrelation levels and adaptive filter dimension are of particular importance to echo cancellation because speech is typically highly autocorrelated and the impulse responses of echo paths are relatively long. In Chapter 5 we present two schemes which essentially whiten the echo canceller input signals. One of the schemes involves low computational nonlinear filtering/defiltering and may be only applied to circuit echo cancellation. The other involves higher computational linear filtering, but may be applied to both circuit and acoustic echo cancellation.

Chapter 6

In Chapter 6 we tackle the problem of reducing the adverse effects (performance, computational cost) of large echo canceller dimensions. Through analyses we quantify that, subject to white input signals (or input signals whitened through the application of the signal conditioning schemes presented in Chapter 5), performance improvements can be achieved by adapting only those taps in the echo canceller which correspond to active/nonzero regions of the echo path impulse response. We then develop a low computational cost scheme which enables the detection of such active regions. A clever combination of this scheme with the LMS algorithm leads to an on-line scheme for adapting only those taps which correspond to active regions of the echo path impulse response. The expected performance improvements this detection-LMS estimation algorithm can achieve are substantiated by simulations.

Chapter 7

A conclusion to the thesis is given in Chapter 7 together with a discussion of extensions and future work.

1.4 Summary of Original Contributions

Open loop adaptive filter system - LMS adaptive FIR filter connected next to and in parallel with time invariant FIR modelled unknown channel of equal tap length:

• Novel cost function developed which, for sufficiently small LMS adaptation stepsize μ, provides a quantitative measure of the expected convergence rate of the LMS adaptive FIR filter to the unknown channel.

• For autoregressive (AR) input signals, an explicit relationship is obtained between the cost function, the AR coefficients, filter dimension and μ.

• Quantification of the influence of FIR tap length (filter parameter dimension), input signal characteristics and μ on the expected convergence cost function.

Closed loop double adaptive filter system - an LMS adaptive FIR filter and unknown channel pair at each end of the loop; driving signals entering each unknown channel:

• Semi-formal to rigorous analysis conducted for small μ case. Considerable extension of the current quantitative understanding of the effects of driving signal correlation levels and filter parameter dimension on asymptotic and transient performance.

Signal conditioning to enhance performance of LMS adaptive FIR echo cancellers in speech transmission telecommunication networks:

• Proposed a scheme which uses digital scramblers at each end of the network to enhance echo canceller performance in digital 4-wire loop networks.

• Proposed modifications to existing schemes for whitening speech/AR modelled input signals of echo cancellers.

Dimension reduced LMS/FIR echo cancellation - based on the observation of inactive/zero tap regions within impulse response of typical echo paths; white input signals assumed:

• Using a least squares approach, a measure of the activity/inactivity of each tap of an FIR modelled unknown channel is developed.

• Based on this activity measure, a low computational cost algorithm is developed for determining the lag position of the 'active' or nonzero taps of an unknown channel/echo path.

• Proposed a modified LMS algorithm which uses the 'active' tap detection algorithm to reduce the adaptive filter parameter dimension and, hence provide improved asymptotic and/or transient performance.

Chapter 2

Adaptive Filtering and Echo Cancellation - A Review

2.1 Introduction

Adaptive echo cancellation is based on the use of adaptive filtering via the parallel configuration of Figure 2.1 to estimate the unknown echo path. In choosing an adaptive filtering scheme, one needs to consider the filter structure and adaptive algorithm. The choice of the filter structure depends largely on the assumed structure of the echo path. A suitable filter structure is a necessity for good asymptotic performance. To a lesser extent, the asymptotic performance also depends on the adaptive algorithm. On the other hand, the transient performance is controlled largely by the adaptive algorithm and to a lesser extent by the filter structure. As in most applications, the choice of an adaptive filtering scheme for adaptive echo cancellation involves a trade-off between performance and computational complexity. The ability to select the scheme which provides the best compromise requires an understanding of the weaknesses/strengths of each adaptive filtering scheme and of the relevance of these to echo cancellation.

The objective of this chapter is to provide such an understanding. We begin by reviewing a collection of different types of adaptive filtering schemes which may be used for estimating an unknown channel via the parallel configuration of Figure 2.1. In particular, we emphasize the weaknesses and strengths of each scheme. We use as a yardstick, the LMS adaptive FIR filtering scheme, which is that currently used in commercially produced adaptive echo cancellers. After stating the requirements and characteristics of echo cancellation, we then highlight the advantages and disad-

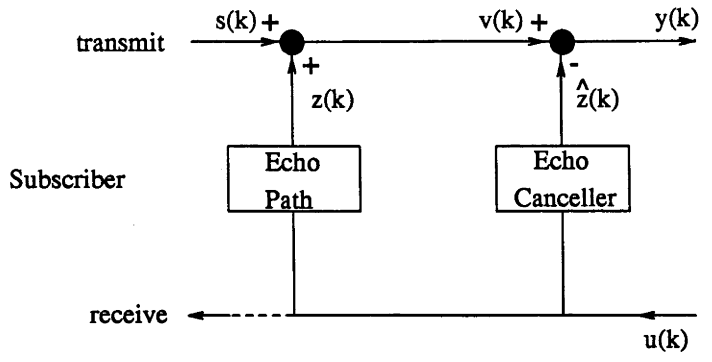


Figure 2.1: Echo suppression via echo cancellation.

vantages of each scheme for this adaptive estimation application. We complete the chapter with a brief review of non-standard approaches to adaptive echo cancellation. The majority of these involve modifications to the LMS/FIR echo canceller. Remarks on double talk are provided in the appendix at the end of this chapter.

We begin by considering the LMS adaptive FIR filter and follow this with schemes which involve alternatives to the LMS adaptive algorithm and/or the FIR filter structure.

2.2 LMS adaptive FIR filter

The LMS adaptive FIR filter, as the name implies, involves using a finite impulse response (FIR) filter to model the parallel unknown channel, Θ, and the least mean square (LMS) adaptive algorithm to enable the filter to converge to and track the channel. The discrete-time/digital FIR filter is a tap delay line:

$$\hat{\Theta}(q^{-1}) = \hat{\theta}_0 + \hat{\theta}_1 q^{-1} + ...\hat{\theta}_{n-1} q^{n-1} \tag{2.1}$$

where q^{-1} is the sample delay operator. This filter structure is particularly popular because of its simplicity and its inherent stability. Of course the usefulness of this structure for parallel adaptive estimation depends on the unknown channel being adequately modelled by an FIR structure:

$$\Theta(q^{-1}) = \theta_0 + \theta_1 q^{-1} + ...\theta_{m-1} q^{m-1} \tag{2.2}$$

The LMS algorithm is based on the idea of adjusting the coefficients of the FIR filter/estimator $\hat{\Theta}_k(q^{-1})$ (where the subscript k indicates time variation due to adaptation) so as to minimize the expectation of the squared error $\epsilon = E[y(k)^2]$, where

$$y(k) = \Theta(q^{-1})u(k) - \hat{\Theta}_k(q^{-1})u(k) + s(k) \tag{2.3}$$

is the output of the adaptive filter/unknown channel system of Figure 2.1 with input $u(k)$ and additive disturbance $s(k)$. This criterion is a quadratic function of the adaptive FIR tap coefficients $\hat{\theta}_i(k)$. The minimum point of the paraboloid formed by plotting ϵ against $\hat{\theta}_i$, $i = 0, 1, ...n-1$ corresponds to the optimum solution $\hat{\theta}^{opt}$. To adapt the tap coefficient vector $\hat{\theta}(k)$ towards this optimum, one approach is to use the method of steepest descent:

$$\hat{\theta}(k+1) = \hat{\theta}(k) + \frac{\mu}{2}(-\nabla_k) \tag{2.4}$$

where μ is a step gain and

$$\nabla_k = \frac{d\epsilon^2}{d\hat{\theta}(k)}$$

is the gradient of the mean-squared error surface. Instead of using the mean squared error, which leads to a high computational load and the need for a large amount of memory [30], [13], the LMS algorithm uses the instantaneous squared error $y(k)^2$ to provide an estimate of the gradient:

$$\hat{\nabla}_k = \frac{dy(k+1)^2}{d\hat{\theta}(k)} \tag{2.5}$$

In particular, the LMS algorithm for the n-tap adaptive FIR filter system of Figure 2.1 is:

$$\hat{\theta}(k+1) = \hat{\theta}(k) + \mu y(k) U(k) \tag{2.6}$$

where $U(k) = (u(k)\ u(k-1)\ ...\ u(k-n+1))^T$. If we let

$$\theta = (\theta_0\ \theta_1\ ...\ \theta_{n-1})^T \tag{2.7}$$

be the vector containing the first n tap coefficients of the FIR modelled unknown channel, then an equivalent form of (2.6) is:

$$\theta - \hat{\theta}(k+1) = (I - \mu U(k) U(k)^T)(\theta - \hat{\theta}(k+1)) - \mu U(k) e(k) \tag{2.8}$$

where $e(k)$ includes the additive disturbance $s(k)$ and any additive disturbances due to undermodelling of the unknown channel.

The LMS adaptive FIR filter suffers from a number of problems, which are discussed next.

2.2.1 Input Signal Autocorrelation Effects

Equation (2.8) indicates that the transient performance of the LMS adaptive FIR filter measured by, for example, the rate of convergence of the expectation

$$E[||\theta - \hat{\theta}(k)||_2], \quad \text{where } ||.||_2 = \text{Euclidean norm}$$

to some asymptotic value, is governed largely by the eigenvalues of $(I-\mu E[U(k)U(k)^T])$. Thus, the transient performance is determined largely by the eigenvalues of the $n \times n$ input signal autocorrelation matrix,

$$E[U(k)U(k)^T] \triangleq R_n.$$

These eigenvalues are essentially the power of the different orthogonal components of the input signal in n-dimensional space [6].

Therefore, the convergence rate generally has n different modes [7], [6], [30], [70], [28]. The slowest mode of convergence is determined by the minimum eigenvalue of the input signal autocorrelation matrix, R_n. For a given input signal power, a greater spread of eigenvalues of R_n, therefore, will have one or more slower modes of convergence. Generally, because the slower modes of convergence will eventually dominate, a greater spread of eigenvalues of the input signal autocorrelation matrix causes slower convergence [7], [28], [6]. This relationship suggests that convergence rate improves as the input signal spectrum becomes flatter [7] and has lead many authors (*e.g.* [13], [26], [12], [42], [19], [35]) to suggest that convergence rate deteriorates with higher input signal autocorrelation levels. Intuitively, the slower convergence rate with increased input autocorrelation results from a greater interaction among the adaptive coefficients, $\hat{\theta}_i(k)$.

This link between input signal autocorrelation characteristics and convergence rate of the LMS adaptive FIR filter is addressed in considerably more detail in Chapter 3.

2.2.2 Effect of Input Signal Power

The convergence rate and stability of the LMS adaptive FIR filter are directly dependent on the value of μ times the power of the input signal [7], [8], [6], [26]. This dependence, which is suggested by (2.8), can be particularly problematic if the power of the input signal varies greatly. In particular, if the input power becomes insignificant, then adaptation of the LMS adaptive FIR filter will essentially stop. On the other hand, if the input power becomes sufficiently large then the LMS/FIR filter can become unstable.

2.2.3 Effect of FIR Tap Length

To enable good asymptotic performance of the LMS adaptive FIR filter, the tap length of the FIR filter should be chosen to cover the impulse response of the un-

known channel. That is, no undermodelling, or, with reference to (2.1) and (2.2), $n \geq m$. This requirement leads to many filter parameters and, consequently, high computational complexity of the LMS/FIR filter when the unknown channel has a long impulse response. The large number of adaptive filter parameters may also lead to poor convergence rates, particularly in the presence of highly autocorrelated input signals [20], [26], [27]. This adverse effect of increasing filter dimension on convergence rate is quantified in Chapter 3. Intuitively, it is due to an increasing number of interactions between the different modes of convergence.

2.3 NLMS Adaptive Algorithm

As noted in the previous section, the value of $\mu\sigma_u^2$, where σ_u^2 is the variance or power of the input signal, directly affects the convergence rate and stability of the LMS adaptive FIR filter. One effective approach to overcoming this dependence is to normalize the update stepsize with an estimate of the input signal variance, $\hat{\sigma}_u^2(k)$. This is known as the Normalized LMS algorithm:

$$\hat{\theta}(k+1) = \hat{\theta}(k) + \frac{\mu}{n\hat{\sigma}_u(k)^2} y(k)U(k) \tag{2.9}$$

where n is the tap length of the adaptive FIR filter and the estimate $\hat{\sigma}_u(k)^2$ can be obtained from one of various formulas *e.g.*:

$$\hat{\sigma}_u(k)^2 = \left\{ \begin{array}{l} \sum_{j=k_0}^{k} U(j)^T U(j)/[n(k-k_0+1)] \\ \alpha + U(k)^T U(k)/n \\ \rho\hat{\sigma}_u(k-1)^2 + (1-\rho)u(k)^2 \end{array} \right\}$$

where n is the length of the vector $U(k)$, α is a small positive constant and $0 < \rho < 1$. A comparison between the LMS and NLMS algorithms has been carried out by a number of authors *e.g.* [29], [31], [8]. In [29] it is shown that, under the popular assumption that the input signal vectors $U(k)$ are Gaussian, the LMS and NLMS algorithms behave quite similarly for small stepsizes μ, where the stepsizes for the different algorithms are related by

$$\mu_{NLMS} = \frac{\mu_{LMS}}{n\hat{\sigma}_u^2}. \tag{2.10}$$

On the other hand, [31], [8] show that, under similar assumptions, by choosing μ so as to optimize the convergence rates of the algorithms, the NLMS algorithm converges more quickly than the LMS algorithm.

Remark:

1. In practice the stepsize μ is chosen to be 'small'. Consequently, in practice, the LMS and NLMS algorithms, with the stepsizes related by (2.10), behave similarly for a given filter dimension n. Note that for a given LMS stepsize, μ_{LMS}, the NLMS algorithm, as compared to the LMS algorithm, shows an additional dependence on dimension. However, as is shown in Chapter 3, Section 3.7, in order to maintain the same asymptotic performance of systems of different dimensions, μ_{LMS} needs to be reduced linearly as dimension increases. In this case, the LMS and NLMS algorithms show a similar dependence on dimension.

2. The analyses conducted in [29], [31], [8] assume that the input signal vectors $U(k)$ are i.i.d. The errors due to the use of this assumption, which is not valid in the case of an LMS adaptive FIR filter system (since U_k and U_{k-1} have $n-1$ elements in common), are reported in [71] to be relatively small when μ is 'small'. This, however, leads one to question the validity of the optimal μ convergence results, which in general do not involve small μ.

2.4 RLS Algorithm

A popular alternative to the LMS/NLMS algorithms which does not show the same dependence of input signal autocorrelation characteristics is the weighted Recursive Least Squares (RLS) algorithm. This algorithm is based on choosing $\hat{\theta}(k)$ so as to minimize the weighted Least Squares cost function:

$$V_k = \sum_{j=1}^{k} y(j)^2 w_{k,j} \tag{2.11}$$

where $y(k)$ is the residual signal output by the estimation filter/unknown channel system at time k and $w_{k,j}$ is a weighting function. A common choice is the exponential weighting function:

$$w_{k,j} = (1-\lambda)^{k-j}, \quad \text{where } 0 \leq \lambda < 1 \tag{2.12}$$

which causes the influence of the past samples to fade out exponentially.

The exponentially weighted least squares solution $\hat{\theta}(k)$ for an FIR estimator is obtained via the exponentially weighted RLS algorithm (see Appendix C for a derivation of this algorithm):

$$\hat{\theta}(k+1) \;=\; \hat{\theta}(k) + R(k)^{-1}U(k)y(k) \tag{2.13}$$

$$R(k) = (1-\lambda)R(k-1) + U(k)U(k)^T \tag{2.14}$$

where $U(k) = (u(k)\ u(k-1)\ ...\ u(k-n+1))^T$ is the input signal vector to the n-tap FIR estimator.

The basic difference between the LMS (and NLMS) and the exponentially weighted RLS algorithms is that the scalar stepsize, μ, in the LMS algorithm is replaced by λ times the inverse of $R(k)$, where $R(k)$ is the short term input signal autocorrelation matrix. The 'normalization' with $R(k)^{-1}$ essentially normalizes the adaptation in each eigenvector direction by the signal power in that direction. This leads to the convergence rate of the RLS algorithm being independent of both the input signal power and autocorrelation characteristics. It is claimed that the exponentially weighted RLS algorithm, for non-white input signals, results in faster convergence and better tracking than the LMS/NLMS algorithms [41],[42]. This superior performance, however, depends largely on an appropriate choice of the forgetting factor λ. For example, if $\lambda = 0$ then the RLS algorithm has little tracking capability and poor transient performance [96].

A major disadvantage of the RLS algorithm is the high computational complexity required to compute the inverse matrix $R(k)^{-1}$. Using the matrix inversion lemma leads to a recursive formula for computing $R(k)^{-1}$ (see Appendix C), but this still requires $O(2n^2)$ multiplications per sample interval, where n is the number of filter coefficients. In comparison, the LMS algorithm requires only $2n$ multiplications per sample interval.

Reduced computational versions of the RLS algorithm such as the FRLS (or FKF) algorithm, have been developed, which make use of the property that the input signal vector $U(k)$ is a one step shifted version of $U(k-1)$ with a new sample on the 'top'. These fast algorithms, however, still require $O(10n)$ multiplications per sample interval [43], [12], [6].

Besides relatively high computational complexity, the FRLS algorithm also suffers from instability problems [20]. One source of instability is the propagation of numerical errors due to finite wordlength representation. Considerable advances, however, have been recently made towards reducing this problem [44], [45], [42]. In particular, one modified version, the FTF algorithm, is reported to achieve these improvements by feedback of the numerical errors.

Another source of instability of the FRLS and, in general, the RLS algorithms arises

when the input signal is non-persistently exciting. This may cause the short term input autocorrelation matrix $R(k)$ to be singular, frequently. Such singularity may lead to instability of the update equation:

$$\theta - \hat{\theta}(k+1) = [I - R(k)^{-1}U(k)U(k)^T](\theta - \hat{\theta}(k)) - R(k)^{-1}U(k)e(k) \qquad (2.15)$$

particularly, as a result of 'bursting' of the noise term

$$R(k)^{-1}U(k)e(k).$$

This behaviour can be avoided by using sufficiently long time windows for updating the autocorrelation matrix or by reducing λ sufficiently closely to zero. However, this necessarily reduces the tracking capability and convergence rate.

2.5 Lattice Filters

Another approach to reducing the dependence of the LMS adaptive FIR filter on the input signal autocorrelation characteristics is to use a lattice filter [7], [46], [47], as shown in Figure 2.2. The lattice filter structure consists of a set of $n-1$ prefilter stages with internal or 'reflection' coefficients k_j, $1 \leq j \leq n-1$. The lattice filter is constructed so that, by appropriate choice of these reflection coefficients, the signals $e_b(i|j)$ output by the lattice stages are uncorrelated with each other. The signal output by the lattice based estimator is obtained from a weighted sum of these uncorrelated signals. Effectively, the lattice filter whitens the input signal so that improved convergence is obtained.

As reported in [7], the weights b_j and reflection coefficients k_j can be adapted using the LMS algorithm. The complexity is about $5n$. Good performance is reported in [48], [49] for small filter dimensions $n \sim 10$ and with stationary input signals. However, for nonstationary input signals, such as speech, [6],[26] and large n, the performance deteriorates. Under such circumstances [6] suggests that an LS based lattice filter should be used, but this results in an increase in complexity to $O(15n) - O(35n)$.

2.6 IIR Filters

Some unknown channels are more suitably modelled by a combination of poles and zeros as opposed to just zeros. For the estimation of such unknown channels, an IIR

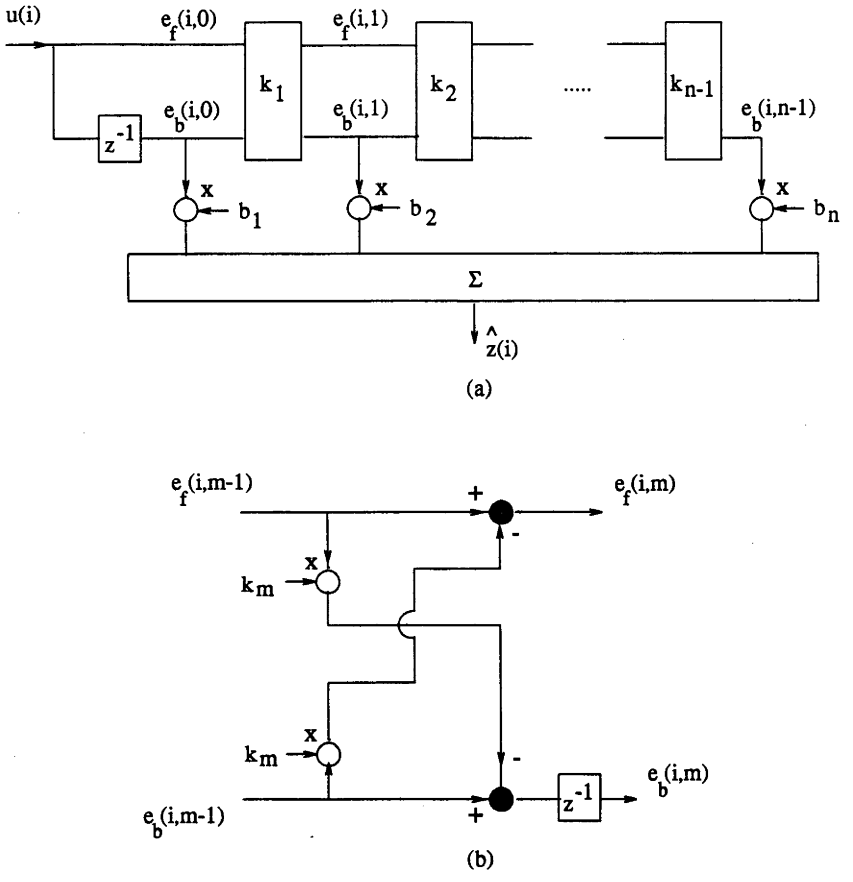


Figure 2.2: Lattice filter:(a) full structure, (b) m^{th} prefilter stage.

(infinite impulse response) filter structure:

$$\hat{\Theta}(q^{-1}) = \frac{B(q^{-1})}{A(q^{-1})}$$

where $B(q^{-1})$ and $A(q^{-1})$ are each FIR filters, seems an appropriate choice for the adaptive filter. In particular, this approach has the potential for greatly reducing the number of adaptive parameters which, in turn, can lead to reduced computational complexity and possibly improved transient performance.

Two configurations of the adaptive IIR filter are shown in Figure 2.3. The series-parallel structure shown in Figure 2.3a has the following advantages:

- lower computational complexity than an FIR filter (assuming the unknown channel has an IIR structure);
- the LMS algorithm may be used for adaptation [26];
- for sufficiently small μ and/or sufficiently large IIR order, stability is guaranteed [51], [50];

and the following disadvantages:

- the performance is limited by disturbances $s(k)$ within the unknown channel [26];
- overspecification of the IIR model order is necessary in order to achieve improvements in asymptotic performance over the adaptive FIR filter [51], [50];
- the convergence rate tends to be slower than the FIR filter, particularly for large IIR model orders [51], [50] - it is suggested in [50] that convergence problems occur for model orders larger than two.

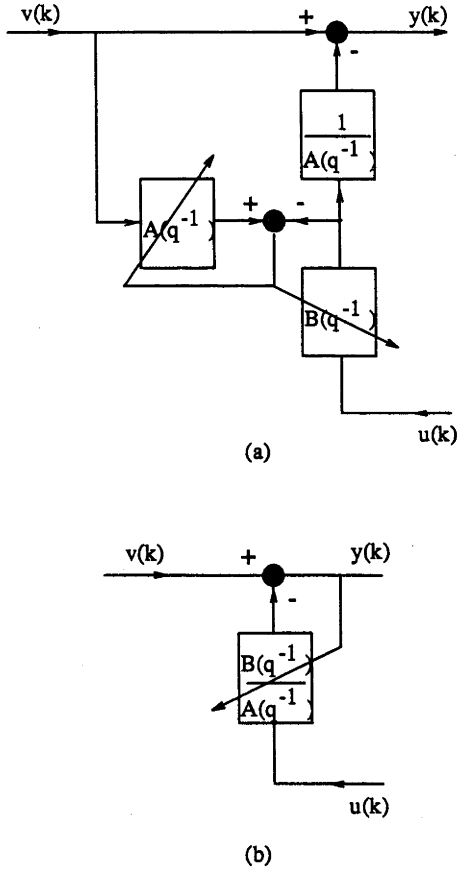


Figure 2.3: IIR adaptive filtering: (a) series-parallel, (b) parallel configuration.

Unlike the series-parallel IIR filter, the convergence characteristics of the parallel IIR configuration of Figure 2.3b are not affected by disturbances. However, the LMS algorithm is not suitable for this filter structure. In general, the adaptation of the parallel IIR filter involves considerably greater computational costs than the LMS/FIR filter [54]. In addition, this IIR configuration suffers from the following problems [26]:

- may converge to a local minimum;
- convergence rate is very slow;
- stability testing is required.

It is evident that in many applications, the weaknesses of either of these adaptive IIR filtering schemes would outweigh their strengths.

2.7 Frequency Domain Filtering

The frequency domain approach involves adapting the filter (usually FIR based) and generating its output in the frequency domain, typically using consecutive, nonoverlapping blocks of data. The number of frequency bins used is chosen to match the (assumed) time domain tap length of the unknown channel. The data block lengths are determined by the number of frequency bins.

Frequency domain filtering can be carried out by either of the two configurations shown in Figure 2.4. The advantages of this approach are as follows.

- Considerably reduced computational complexity [56] for sufficiently large filter lengths, $n > 30 - 50$. This is due to the fact that:

  - the convolution of a pair of time domain sample blocks (which is required in order to generate the echo replica) is equivalent in the frequency domain to simple multiplication of the corresponding frequency domain coefficients;

  - highly efficient fast fourier transform (FFT) methods are available to ensure that transform of signals to and from the frequency domain does not introduce too much extra computational complexity.

  In particular, for a filter size of $n = 1024$, the computational advantage of frequency domain filtering is [56]

  $$\frac{\text{time domain complexity}}{\text{freq. domain complexity}} = 15 \; to \; 50$$

  depending on the FFT method used. The advantage increases/decreases as the filter size increases/decreases.

- The potential for improved convergence rates as measured per adaptation or block of data. This is due to the fact that the signals from one frequency bin to the next are approximately uncorrelated, if the collection window is

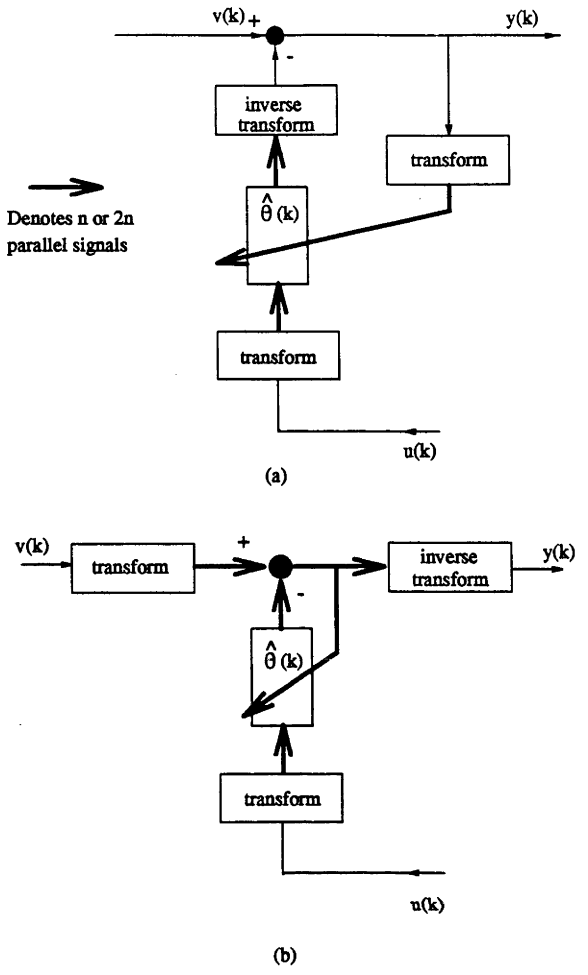


Figure 2.4: Frequency domain adaptive filtering with error signal computed in (a) time domain, (b) frequency domain.

sufficiently large [57], [6]. In effect, the signal power in each frequency bin is a measure of the signal power of each of the orthogonal components of the input, or, equivalently, is a measure of each of the eigenvalues of the input signal autocorrelation matrix [58], [59]. The adverse effect on LMS convergence rate of a wide eigenvalue spread (of the input signal autocorrelation matrix) can, therefore, be minimized by using a stepsize in each frequency bin which is inversely proportional to the signal power level in that bin.

Disadvantages of the LMS/FIR block frequency domain filter include the introduction of delay [56] and possibly a reduction in the stable range of μ [97]. For nonstationary signals, the tracking capability is also generally inferior [56] to the LMS/FIR time domain filter.

Another serious disadvantage of block frequency domain adaptive filtering is that

adaptation only takes place once per block of data. In general, this results in considerably slower convergence in real time than the LMS adaptive FIR filter [56], [98]. An approach to improve the real time convergence rate is to use a sliding window to compute the fourier transform of the block with each new input sample. This enables the adaptive filter to be updated every sample interval and results in the convergence rate approaching that of the time domain RLS adaptive FIR. However, it also leads to a considerable increase in the computational complexity.

2.8 Sub-Band Filtering

Sub-band echo cancelling involves decomposing the input signal and unknown channel output, sampled at say FkHz, into M frequency sub-bands, and providing each sub-band with its own LMS adaptive FIR filter (other adaptive filters could also be used) - as shown in Figure 2.5. In addition, the sub-band signals are downsampled by a factor $L \leq M$ so that the sampling rate in each sub-band is F/L kHz. Generally, polyphase filters [5], [60], [61] or quadrature mirror filters [62], [15] are used to efficiently implement the analysis and synthesis filter banks.

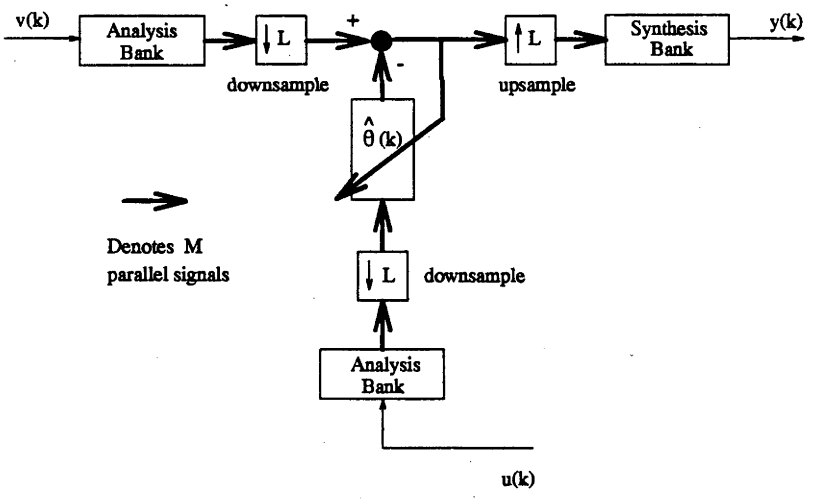


Figure 2.5: Subband adaptive filtering.

The advantages of sub-band filtering are as follows.

- Potentially, a considerable reduction in computational complexity. This is due to, primarily, the reduction in the sampling rate in each sub-band, which enables a reduction in the number of taps in each adaptive filter and a reduction in the required adaptation rate by a factor of L. Thus, neglecting the additional

requirements of the analysis and synthesis filter banks, the saving in computational complexity is by a factor of L^2/M [20]. A further reduction in the computational complexity can be achieved by adjusting the tap length of the filters in each sub-band to match the tap length of the unknown channel in the corresponding sub-band [5]. The extra computational complexity introduced by the analysis and synthesis depends on the specific approach taken. Polyphase filters cause only a small increase in computation [5]. More specifically, one approach [17], which employs two polyphase networks of order 321 and two FFTs for analysis and synthesis and a decimation factor of 32, has an overall complexity which is about 1/8 of the equivalent 4000 tap LMS/FIR filter.

- Potentially, an increase in convergence rate [20], [17], [5]. This is due to

  - adjusting the stepsize μ in each sub-band according to the input signal power in that sub-band [15] (which achieves a similar decorrelating effect of the input as that obtained with frequency domain filtering);

  - the reduction in the number of taps in each sub-band LMS/FIR filter by the factor L (as quantified in Chapter 3, the convergence rate of the LMS/FIR filter improves as the number of taps reduce).

  Superior transient performance of the sub-band filter over the LMS adaptive FIR filter has been reported by [20], [17], [5].

With regard to the latter advantage, it should be remembered that the sampling rate within the sub-band filter is reduced by a factor of L and, consequently, in real time, the convergence rate, in general, may not be superior to that of the conventional LMS/FIR.

There are a number of disadvantages with sub-band echo cancellation.

- A delay in the signal $y(k)$ output by the adaptive filter/unknown channel system caused by the analysis and synthesis filterbanks [20], [15], [5], [17]. This delay increases with the number of sub-bands [5]. This problem can be overcome by placing the synthesis filter banks immediately after the adaptive filter [63], [64] so as to avoid the analysis and synthesis of the unknown channel output signal. In addition to avoiding output delay, this modification provides reduced computational complexity. However, it does not provide the convergence benefits of the standard approach and, because of the delay in the echo replica, tends to worsen tracking ability [20].

- Poor asymptotic performance due to aliasing or cross-talk between sub-bands. This occurs with polyphase filter banks as well as QMF filter banks. The inclusion of the latter may be surprising since QMFs are designed so that aliasing is compensated for during synthesis, However, full compensation, requires that the sub-band signals all experience the same processing, which is not the case in general. To remove this aliasing, adaptive filters can be used between neighbouring sub-bands [65]. Alternatively, oversampling (*i.e.* $M > L$) can be used [60], [15]. Both alternatives result in an increase in computational complexity and limit the computational benefits of increasing the number of sub-bands.

2.9 Adaptive Filtering for Echo Cancellation

In this section we begin by making a comparison of the suitability of each of the adaptive filtering schemes discussed above for adaptive echo cancellation in speech transmission telecommunication systems. We then move onto presenting a number of non-standard adaptive schemes which have been proposed for this application.

In speech transmission echo cancelling systems, the unknown channel is the echo path, the disturbance signal is the near end subscriber signal and the input signal is comprised largely of the far end subscriber signal. The subscriber signals are basically speech, although they may also contain noise.

The characteristics of this adaptive filtering application which are of particular importance are:

- the input signal

(i) is highly autocorrelated,
(ii) on the short term shows a lack of persistent excitation, and
(iii) is approximately stationary only over a 20ms interval;

- the impulse response of the unknown channel is moderately to very long and, particularly in acoustic echo cancellation, may be time varying.

A primary requirement of an adaptive filtering scheme for this application is low computational complexity.

The NLMS/LMS adaptive FIR filter is attractive because of its inherent stability and relatively low computational costs. In acoustic echo cancellation, however, the computational requirement becomes somewhat demanding. The major drawback of

this adaptive filtering technique is the sensitivity of its transient performance to input signal autocorrelation characteristics. In particular, various authors [13], [26], [12] report poor transient performance with speech input signals particularly when the FIR filter is 'long'.

The transient performance of the RLS adaptive FIR filter, in contrast, is independent of the input signal autocorrelation characteristics. However, the stability problems and/or high computational cost of this adaptive filtering scheme with speech inputs and 'long' unknown channels detracts considerably from its appeal as an echo canceller in speech transmission systems.

Lattice based filter structures under the conditions of echo cancellation - speech inputs and long unknown channels - need to be RLS adapted for performance reasons. The computational cost of such an adaptive filtering scheme is prohibitively large.

Due to its lower computational requirements, the LMS adaptive series-parallel IIR filter may be a potential alternative to the LMS adaptive FIR filter for acoustic echo cancellation. Its applicability, however, depends on minimal double talk and on the echo path being adequately modelled by a low order IIR filter. These are serious limitations. In particular, one study [52] reported finding 80 maxima and minima in the transfer function of acoustic echo paths in the frequency range of $0 - 4$kHz. This implies [20] that an IIR filter having an order of greater than 80 is required to model acoustic echo paths. This, in turn, suggests that the series-parallel IIR filter is not suitable for acoustic echo cancellation

The parallel IIR filter does not appear to be a possible option for echo cancellation. This is supported by studies in [54] which demonstrate that such a filter provides very little, if any, performance improvement but causes considerable increase in adaptation complexity. In contrast, good performance is reported in [53] using the parallel IIR structure. The approach assumes an AR model for the disturbance signal and employs a recursive prediction error method (based on Least Squares) for adaptation. This claim is supposedly substantiated through simulations. The relevance of the simulations to echo cancellations is highly questionable since only low order IIR and AR filters are considered.

LMS/FIR frequency domain filtering, depending on whether it is conducted in consecutive data blocks or with sliding windows, offers computational cost advantages or convergence rate advantages, respectively, over time domain LMS/FIR filtering. However, each advantage usually comes at the expense of the other. The poorer

tracking ability (in comparison to time domain filtering) of frequency domain filtering with nonstationary inputs as well as the introduction of delay (which is particularly unattractive in speech transmission) detracts further from its appeal.

The main attraction of sub-band filtering is its lower computational costs in comparison to the LMS adaptive FIR filter, particularly as the number of sub-bands increases. However, along with this advantage comes the delay introduced by the analysis and synthesis filter banks, which also increases with the number of sub-bands. Modifications to avoid this serious limitation result in other problems such as reduced convergence rates and tracking capability.

The above discussion suggests that the presented alternatives to the NLMS/LMS adaptive FIR are not necessarily better suited for speech transmission echo cancellation and, in various ways, are poor substitutes. Of course, in some situations one or more of the alternatives may out-perform the LMS/FIR filter. However, more work is needed to enable such improvements and the conditions under which they occur to be quantified.

Possibly as a consequence of the problems these standard adaptive filtering techniques suffer, non-standard approaches to adaptive echo cancellation have been and are being explored. Some of these are discussed below. Note that the majority of these are modified versions of the NLMS/LMS adaptive FIR filter.

- In [55] a two stage echo canceller is proposed, in which the first stage is an LMS/FIR filter of 20-40 taps. This stage is used to model the front of the echo path. The second stage is an adaptively weighted linear combination of orthogonal IIR filters which is used to model the echo path tail. The orthogonality is ensured by basing the z-transform of the jth IIR filter on the z-transform of the jth order Laguerre function:

$$L_j(z) = \frac{z^{-1} - r)^j}{(1 - rz^{-1})^{j+1}}, \quad r = e^{-p} \tag{2.16}$$

  where $p > 0$ is a constant to be determined. Good performance and substantial complexity reduction is reported for a variety of echo paths. A major drawback with this approach is that the benefits depend on the Laguerre parameter p being optimized for each echo path, FIR tap length and number of Laguerre IIR filters used. Furthermore, this optimization is carried out off line.

- It is proposed in [32] to insert uncorrelated noise bursts into the input signal of the LMS adaptive FIR echo canceller during initialization of the connection.

The noise burst insertion leads to a whitening effect of the input signal and consequently, may provide a boost to the initial convergence rate of the echo canceller. An alternative approach, proposed in [33], is to insert the noise into the input signal $u(k)$ and residual output signal $y(k)$ only in the adaptation algorithm. This approach has the advantage that the noise insertion can continue past the initialization stage. Intuitively, however, for a given period of noise insertion, this latter approach, as compared to the former approach, produces a smaller whitening effect for the same level of noise inserted.

- In another bid to reduce the dependence of the LMS/FIR echo canceller on the autocorrelation characteristics of the input signal, it is proposed in [34],[35], [19] to update the echo canceller using only that component of $U(k)$ which is orthogonal to or uncorrelated with $U(k-1)$. This is carried out using affine projection algorithms, the second order version [35] of which is:

$$\begin{aligned} c(k) &= U(k)^T U(k-1)/(U(k-1)^T U(k-1)) \\ Z(k) &= U(k) - c(k)U(k-1) \\ \hat{\theta}(k+1) &= \hat{\theta}(k) + \mu Z(k) y(k)(1/U(k)^T Z(k)) \end{aligned}$$

- Voiced speech is typically well modelled as an autoregressive process:

$$u(k) = \frac{1}{A(q^{-1})} w(k) \tag{2.17}$$

where:

$A(q^{-1}) = 1 + a_1 q^{-1} + a_2 q^{-2} + ... + a_p q^{-p}$,

$w(k)$ is a discrete white zero mean signal.

Clearly, the filter $A(q^{-1})$ is a whitening filter for $u(k)$. By using linear prediction methods on the input signal, an estimate of the filter $A(q^{-1})$ is obtained and used to whiten the input signal [36], [37]. This approach is discussed in more detail in Chapter 5. Note that due to the nonstationary characteristics of speech, the estimate of the autoregressive filter needs to be updated regularly - every $20ms$ [20].

- Under a number of assumptions including that of the input signal being white, analyses carried out in [18] suggest that convergence rate improvements of the LMS/FIR echo canceller are achieved by assigning to each echo canceller tap $\hat{\theta}_i(k)$ its own update stepsize, $\mu_i = \mu\theta_i^2$, where θ_i is the corresponding echo path tap coefficient. Combining this with measurements of the impulse responses of acoustic echo paths suggests that the sequence of individual stepsizes, $\{\mu_i\}$,

should decay exponentially with tap index or sample lag i [18]. The difficulty lies in determining a suitable exponential decay rate. Procedures of various computational complexity are suggested.

- A similar approach is suggested in [38] to improve the tracking of time variations of the echo path - different stepsizes are chosen for each FIR echo canceller tap depending on whether the corresponding tap is slow varying or fast varying.

- Recent suggestions [39], [40] for LMS/FIR based echo cancellers include procedures which take into account the possibility of inactive areas (regions of nonzero taps) in the echo path impulse response. Ignoring these areas may lead to a reduction in the number of adaptive FIR taps, which in turn provides computational savings. It is also reported to lead to convergence rate improvements. This approach is discussed in more detail in Chapter 6.

Each of these non-standard approaches, as indicated, has its merits. The general approach of developing modifications for the LMS/FIR filter, however, has an extra appeal. This is due to the dynamical simplicity of the LMS/FIR filter, which enables the effects of the modifications to be determined, at least intuitively. However, before attempting to develop modifications, this author believes that the system dynamics should be quantitatively analysed. This avoids the risk of chasing modifications in an ad-hoc fashion and enables the usefulness of modifications to be quantified and compared. In particular, it avoids the need to estimate performance improvements solely from simulations.

The work leading to this thesis has been based on this idea:

- quantify the weaknesses of the LMS adaptive FIR filter through dynamical analyses;

- based on the analytical results, develop performance improving modifications.

2.10 Conclusion

In this chapter we briefly examined the weaknesses and strengths of the various adaptive filtering schemes which have been and are being considered for echo cancellation in speech transmission systems. The LMS adaptive FIR filter, because of its relatively low computational costs and inherent stability is popular. However, in this application, with speech inputs and large tap lengths, the LMS/FIR filter suffers

from poor convergence rates and demanding computational requirements, particularly in acoustic echo cancellation. The alternative schemes presented, such as that involving RLS adaptation or sub-band filtering, generally provide some advantages over the LMS/FIR filter, but these come at the expense of other disadvantages, such as greater computational cost, poorer performance, instability and/or transmission delay. In general, the alternatives to the LMS/FIR filter are not necessarily better suited for speech transmission echo cancellation and, in many ways, are poor substitutes. Another approach is that of developing modifications to the LMS/FIR filter. This approach is particularly appealing because the dynamical simplicity of this filter enables the effects of the modifications to be determined intuitively. To enable this approach to be followed methodically, the influence of the various system parameters on LMS/FIR filter performance should be first quantified.

Chapter 2 Appendix: Remarks on Double Talk

The LMS adaptive FIR echo canceller, like many of the alternatives suggested, suffers from misadjustment or incorrect adaptation of the echo canceller coefficients during periods of double talk - that is, simultaneous transmission from each end of the network. This is due to the fact that during such periods the near end signal may be interpretted as part of the echo, $z(k)$.

The conventional way of avoiding misadjustment, which requires fitting of a double talk detector, is to stop adaptation of the echo canceller during double talk [13], [6]. Using the frozen echo canceller tap coefficients, the output of an echo replica, however, continues as does transmission of the near end signal. Other approaches to the double talk problem involve more sophisticated, but computationally expensive, stepsize control schemes. For example, [66], [35], [67], [68],

$$\begin{aligned} \mu(k) &= 1/[1+\nu(k)/(||\theta-\hat{\theta}(k)||_2^2)] \\ \nu(k) &= E[s(k)^2]/(||U(k)||_2^2/n) \end{aligned}$$

where $E[.]$ denotes expectation operation. To implement this approach, however, additional methods and computation must be employed to obtain estimates of $E[s(k)^2]$ and $||\theta-\hat{\theta}(k)||_2^2$.

The possible misadjustment of the adaptive echo canceller due to double talk will not be addressed directly in this thesis. However, as will become apparent through quantitative analyses of the LMS/FIR echo canceller, under certain assumptions - such as echo path time-invariance, signal stationarity, zero cross correlation between input and disturbance signals, as well as a sufficiently small stepsize, μ - the LMS adaptive FIR echo canceller does not suffer from long term misadjustment due to double talk. It seems intuitive that, for sufficiently small μ (which is a basic assumption in this thesis), this result should also hold when the assumption on stationarity is relaxed.

Chapter 3

Quantitative Analysis of the LMS Adaptive FIR Filter

3.1 Introduction

In this chapter we carry out rigorous analyses of the LMS adaptive FIR filter used to estimate a time invariant unknown channel via the parallel configuration of Figure 3.1. The main objective of the analyses is to quantify the link between the adaptive filter performance and system properties such as input signal autocorrelation level and FIR tap length or parameter dimension, n. The effects of these factors, which have been suggested by other analyses, are particularly relevant to echo cancellation in which the input signal typically is highly autocorrelated speech and the filter dimension is moderate to very large. Meeting this objective will enable the development of techniques which enhance the performance of the LMS adaptive FIR echo canceller.

We separate the analyses into those of transient and asymptotic performance. Of the numerous studies previously conducted, many provide a quantitative link between asymptotic performance, signal characteristics and filter dimension. The review of these asymptotic results is left until Section 7. Our analyses, focus on transient performance. We begin in Section 2 with a system description which includes assumptions and notation as well as the LMS adaptive system equation. In Section 3, using the framework introduced in Section 2, we review relevant results on transient performance. The results rely heavily on an invalid input signal independence assumption, the effect of which is quantifiably negligible when the LMS adaptation stepsize, μ, is sufficiently small. We take a different approach. Under the assumption that μ is sufficiently small, Averaging Theory is applied in Section 4 to obtain an 'averaged' system

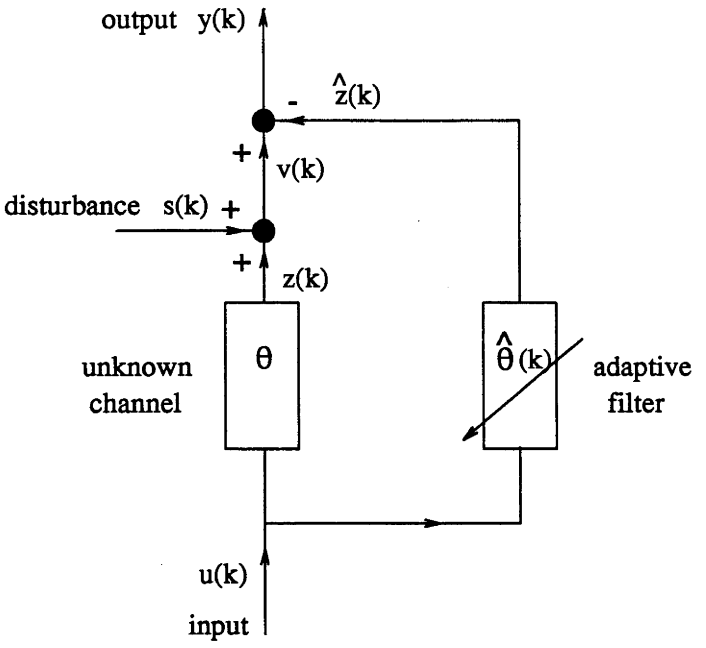


Figure 3.1: LMS adaptive FIR filter in parallel with time invariant unknown channel.

equation, the dynamics of which approximate closely and are considerably simpler to analyze than those of the original system equation. In Section 5 we propose a cost function which provides a suitable measure of the expected convergence rate of the averaged system. In Section 6, we analyse the dependence of this cost function on filter parameter dimension and input signal autocorrelation under the condition that μ is fixed. Particular attention is given to input signals described by autoregressive (AR) processes, since such processes are typically used to model voiced speech. The cost function is shown to depend explicitly on the AR parameters. In Section 7, after the review of asymptotic results, we extend the results of our transient performance analyses to the case in which μ is adjusted to maintain asymptotic performance.

So as to complete the analyses, we have also analysed the effects of filter dimension and signal characteristics on the error arising due to the averaging approximation. This averaging error analysis is given in the appendix at the end of this chapter.

Note: For easy reference, a listing of the assumptions made in this and other chapters is given in Appendix I.

3.2 System Description

Throughout this chapter we consider the LMS adaptive FIR filter in the parallel configuration of Figure 3.1. This is the typical set-up for echo cancellation. In the context of echo cancellation, the unknown channel represents the circuit or acoustic echo path. The adaptive filter represents the echo canceller being used to estimate the echo path. At sampling instant k: the input signal $u(k)$ represents the signal received from the far end of the network; the disturbance signal $s(k)$ represents the signal transmitted by the near end subscriber; the output $y(k)$ represents the signal transmitted to the far end subscriber. The signals $u(k)$ and $s(k)$ may include speech and/or noise. The signal $u(k)$ may also include echo generated by the far end subscriber.

We make the following assumption regarding the unknown channel and adaptive filter.

Assumption 1 *(i) The unknown channel is time invariant and is adequately modelled by an n-tap digital FIR filter with tap coefficient vector*

$$\theta = (\theta_0 \ \theta_1 \ \theta_2 \ ... \ \theta_{n-1})^T. \tag{3.1}$$

(ii) The LMS adaptive FIR filter also has a tap length of n and at sampling instant k has the tap coefficient vector

$$\hat{\theta}(k) = (\hat{\theta}_0(k) \ \hat{\theta}_1(k) \ \hat{\theta}_2(k) \ ... \ \hat{\theta}_{n-1}(k))^T. \tag{3.2}$$

(iii) The tap coefficients of the LMS adaptive FIR filter are initially set to zero

$$\hat{\theta}_i(0) = 0, \quad i = 0, 1, 2, ..., n-1.$$

Remark:

1. The assumption that the adaptive filter and unknown channel have equal tap lengths is made, basically, for ease of analysis. The results of the analyses can be easily extended to the case in which the adaptive filter is 'longer' than the unknown channel. A 'shorter' adaptive filter, however, leads to biased asymptotic estimation of the unknown channel. This undermodelling bias is in addition to any bias introduced by the correlation characteristics of the input and disturbance signals. Under some circumstances, such as in the closed loop configuration of telephone networks, the existence of this additional source of bias

increases the possibility of instability. As long as such instability does not occur, then the results of the transient performance analyses obtained in this chapter can be extended also to this 'shorter' adaptive filter case.

2. Zero initial conditions for the adaptive filter typically are chosen in practice.

By introducing the n-tuple input signal vector:

$$U(k) = (u(k),\ u(k-1),\ \ldots,\ u(k-n+1))^T$$

we obtain the following expression for the output signal:

$$\begin{aligned} y(k) &= (\theta - \hat{\theta}(k))^T U(k) + s(k) \\ &= \tilde{\theta}(k)^T U(k) + s(k) \end{aligned} \tag{3.3}$$

where $\tilde{\theta}(k) = \theta - \theta(k)$ is the residual filter parameter vector. Good estimation (or in the case of echo cancellation, good echo cancellation) results when $\tilde{\theta}(k) \approx 0$. Using the LMS algorithm for adaptation leads to the following update equation for the residual filter parameter vector:

$$\begin{aligned} \tilde{\theta}(k+1) &= \tilde{\theta}(k) - \mu y(k) U(k) \\ &= (I - \mu U(k) U(k)^T)\tilde{\theta}(k) - \mu U(k) s(k), \quad \tilde{\theta}(0) = \theta \end{aligned} \tag{3.4}$$

where μ is the update stepsize. Equation (3.4) describes the dynamics of LMS adaptive FIR filter system.

Remark:

3. For stability and asymptotic performance reasons, μ is chosen typically to be very 'small' (see Assumption 6 and Remark 7).

The following assumption makes the application of (deterministic) Averaging Theory feasible.

Assumption 2 *The input, $u(k)$, and disturbance, $s(k)$, signals are zero mean, bounded and stationary so that the limits:*

$$R_m \stackrel{\Delta}{=} E(U(k)U(k)^T) = \lim_{N\to\infty} 1/N \sum_{k=k_0}^{N-1+k_0} U(k)U(k)^T,\ \forall k_0$$

$$P_m \triangleq E(U(k)s(k)) = \lim_{N\to\infty} 1/N \sum_{k=k_0}^{N-1+k_0} U(k)s(k), \ \forall k_0$$

$$\sigma_u^2 \triangleq \lim_{N\to\infty} 1/N \sum_{k=k_0}^{N-1+k_0} u(k)^2, \ \forall k_0$$

$$\sigma_s^2 \triangleq \lim_{N\to\infty} 1/N \sum_{k=k_0}^{N-1+k_0} s(k)^2, \ \forall k_0$$

exist for all m, where m =length of $U(k) = (u(k)\ u(k-1)\ \ldots\ u(k-m+1))^T$.

We also make the following assumption to simplify the analysis.

Assumption 3 *The input and disturbance signals are uncorrelated with each other over time:*

$$\lim_{N\to\infty} 1/N \sum_{k=0}^{N-1} u(k-l)s(k) = 0, \quad \forall l$$

that is, $P_m = 0$.

Remark:

4. Recall that in the application of echo cancellation, $u(k)$ contains the far end speech signal, $s_{FE}(k)$, and, possibly, a delayed and attenuated version of the near end speech signal, $s_{NE}(k)$, resulting from incomplete far end acoustic echo cancellation. Assumption 3 will be valid if $s_{NE}(k)$ and $s_{FE}(k)$ are uncorrelated (which is typical of speech) and if the autocorrelation length of $s_{NE}(k)$ is less than the delay associated with travelling the full network loop (NE $\rightarrow$ FE $\rightarrow$ NE). This remark is addressed in more detail in Chapter 4.

Assumption 4 *The input signal,* $u(k)$*, is such that the autocorrelation sequence* $\{r_l\}$*:*

$$r_j = \lim_{N\to\infty} 1/N \sum_{k=0}^{N-1} u(k)u(k-j), \ j = \ \ldots -2, -1, 0, 1, 2, \ \ldots$$

is absolutely summable:

$$\sum_{j=-\infty}^{\infty} |r_j| < \infty$$

Remark:

5. Assumption 4 guarantees the existence of the power spectrum $\Phi_{uu}(\omega)$ of the input signal.

Assumption 5 *The power spectrum $\Phi_{uu}(\omega)$ of the input signal is positive definite:*

$$\Phi_{uu}(\omega) > 0, \quad 0 \leq \omega \leq 2\pi$$

Remark:

6. Assumption 5 implies that the input signal covariance matrix R_n is positive definite for all dimensions, n.

Assumption 6 *The update stepsize is such that:*

$$\mu \ll \frac{1}{n\sigma_u^2} \triangleq \frac{1}{Tr(R_n)} < \frac{1}{\lambda_{max}(R_n)} \tag{3.5}$$

where $Tr(.) = Trace(.)$.

Remark:

7. Assumption 6 is not at variance with typical choices of μ in practice. In particular, the validity of Assumption 6 guarantees stability of the averaged system and, generally, guarantees that the dynamics of the second order moments of $\tilde{\theta}_k$ also remain stable. These stability issues are addressed in more detail in the next section.

3.3 Review of Transient Performance Analyses

The dynamics of the LMS adaptive FIR filter in the parallel configuration of Figure 3.1 have been studied by numerous authors (*e.g.* [69], [30], [70], [71], [72], [28], [8], [73]). A common feature of these studies is that they indicate that the transient performance or convergence rate is linked to the eigenvalues, λ_i, of the $n \times n$ input signal autocorrelation matrix, R_n, where n is the tap length or parameter dimension

of the FIR estimator. This indicates a strong dependence of convergence rate on the second order moments or, equivalently, on the autocorrelation function of the input signal. It also suggests a dependence on the parameter dimension, n, of the FIR filter. The existence of this dimension dependence is supported in very general cases by Vapnik and Chervonenkis theory [74] which leads us to expect a penalty on convergence rate with an increase in filter parameter dimension.

Despite the apparent existence of these input signal and dimension dependencies of convergence rate, few studies have attempted to quantify them. Many instead attempt to determine bounds on the LMS adaptation constant μ to ensure stability and/or the value of μ which optimizes the convergence rate. The results vary depending on the assumptions made and approach taken. Earlier analyses considered convergence of $E[||\tilde{\theta}(k)||_2]$ or convergence in the mean (where $||.||_2$ denotes the Euclidean norm), while more recent analyses consider convergence of $E[\tilde{\theta}(k)\tilde{\theta}(k)^T]$ or convergence in second order moments. The latter approach, because it takes into account variations around the mean, necessarily leads to tougher restrictions on μ to ensure stability. A summary of some of these results is given in Table 3.1.

Table 3.1: Optimal Convergence Value of and Upper Stable Value of μ.

Approach	Input Conditions	Optimal, μ^*	Upper Stable
Type 1 [30],[28]			$\mu < 1/\lambda_{max}(R_n)$
Type 2 [7],[28],[70]	Gaussian white	$1/[(n+2)\sigma_u^2]$	$\mu < 2\mu^*$
Type 2 [28]	white,kurtosis= ν_u	$1/[\sigma_u^2(n-1+\nu_u)]$	$\mu < 2\mu^*$
Type 2* [8]	Gaussian	$n\sigma_u^2/E[\|\|U(k)\|\|_2^4]$ $n\sigma_u^2/[(n\sigma_u^2)^2+2Tr(R^2)]$	$\mu < 2\mu^*$

Type 1: Convergence in Mean, $E[||\tilde{\theta}(k)||_2]$;
Type 2: Convergence in 2^{nd} order moments, $E[\tilde{\theta}(k)\tilde{\theta}(k)^T]$. This approach assumes that the input signal vectors $U(k)$ are *i.i.d.*.
* Disturbance assumed to be white.

It must be emphasized that in analysing the convergence in second order moments, all studies to date make the assumption that the input signal vectors $U(k)$ are independent and identically distributed. This assumption, which is not valid since $U(k)$ and $U(k-1)$ have $n-1$ elements in common, is reported in [71], [75], [76], [77] to cause only relatively small errors if the stepsize is sufficiently small. However, this leads one to question the validity of the results in Table 3.1 for this second order moments approach. In general, stability requires μ to be sufficiently smaller than that indicated in Table 3.1.

Remark:

8. Since for Gaussian input signals

$$Tr(R_n^2) < [Tr(R_n)]^2$$

then, according to Table 3.1, convergence in second order moments for such signals is guaranteed when

$$\mu < \frac{1}{3Tr(R_n)}$$

Taking into account the error due to the use of the i.i.d. assumption in deriving the results, a 'safe' choice would be that indicated in Assumption 6:

$$\mu \ll 1/(Tr(R_n))$$

.

The few authors who have attempted to quantify transient performance have considered one of two transient performance measures. The measure proposed in [70] is a convergence cost function similar to that of (3.15) in Section 5. However, it is noted that this cost function depends on the unknown channel vector θ. Consequently, besides determining an optimal convergence rate value for μ, analyses of this cost function are not carried out.

In [28], [30] quantitative analysis of the initial rate of convergence (in the mean) is attempted by proposing as a measure, the effective initial time constant τ defined as the time constant of the 'learning curve' $\epsilon(k) = E[(\tilde{\theta}(k)^T U(k) + s(k))^2]$:

$$[\epsilon(0) - \epsilon(\infty)]e^{-1/\tau} = [\epsilon(1) - \epsilon(\infty)]$$

As in other analyses, n different modes of convergence are shown, in general, to exist. In [30] the analysis is simplified by considering only the case in which the input signal is white for which case the time constant is given by

$$\tau = \frac{1}{\mu\sigma_u^2} \tag{3.6}$$

This indicates no dependence of convergence rate on dimension.

To enable quantitative analysis, [28] chooses to examine the case in which the unknown channel vector θ is uniformly distributed among the eigenvectors of the input signal autocorrelation matrix R_n so that:

$$||R_n\theta||_2^2 = \frac{Tr(R_n^2)}{n}||\theta||_2^2 \tag{3.7}$$

The result is that, if μ is sufficiently small such that $\mu \ll 1/\lambda_{max}(R)$, then:

$$\begin{aligned} \frac{1}{\tau} &\approx 2\mu \frac{Tr(R_n^2)}{Tr(R_n)} \\ &= 2\mu\sigma_u^2(1+\rho) \qquad (3.8) \\ \text{where } \rho &= \sigma_\lambda^2/\lambda_{ave}^2 \\ \sigma_\lambda^2 &= \text{variance of eigenvalues of } R_n \\ \lambda_{ave} &= \text{mean of eigenvalues of } R_n \end{aligned}$$

The result of (3.8) indicates that the convergence rate is input signal dependent and, in particular, increases with the spread of the eigenvalues of R_n as measured by ρ. Since $0 \leq \rho \leq n-1$, then the result of (3.8) also suggests that convergence rate is dependent on dimension. This dimension dependence vanishes when $\rho = 0$, that is, when the input signal is white.

The above results indicate that, for fixed μ, a greater dimension may penalize the convergence rate. Furthermore, this penalty may be input signal dependent. In the following sections we carry out analyses to quantify these suggested dimension and input signal dependencies of convergence rate.

3.4 Averaged System Equations

In this section we apply the Averaging Method, as presented in [78] and as discussed in Appendix B, to (3.4) to obtain an equation which approximates its dynamics and is considerably simpler to analyse.

Application of averaging to (3.4) yields:

$$\begin{aligned} \tilde{\theta}^{av}(k+1) &= (I-\mu R_n)\tilde{\theta}^{av}(k), \quad \tilde{\theta}^{av}(0) = \theta \qquad (3.9) \\ &= (I-\mu R_n)^{k+1}\theta \qquad (3.10) \end{aligned}$$

Remark:

9. The application of averaging relies on μ being sufficiently small such that the residual parameter vector $\tilde{\theta}(k)$ is slowly time varying in comparison to the input $u(k)$ and disturbance $s(k)$ signals. As a result of this separation in time scales, we can approximate (3.4) with that in which the variations in $\tilde{\theta}(k)$, due to $u(k)$ and $s(k)$, are averaged out.

Assumptions 5 and 6 guarantee that:

$$\tilde{\theta}^{av}(k) \to 0, \quad \text{or, equivalently,} \quad \hat{\theta}^{av}(k) \to \theta, \quad \text{exponentially} \tag{3.11}$$

that is, good asymptotic performance of the averaged adaptive filter system.

The solution, $\tilde{\theta}^{av}(k)$, of the averaged equation of (3.9) or (3.10) approximates the solution, $\tilde{\theta}(k)$, of the original equation within the averaging error:

$$||\tilde{\theta}(k) - \tilde{\theta}^{av}(k)|| = O(\delta(\mu)) \tag{3.12}$$

where:

$$\delta(\mu) = \sup_{k_0} \sup_{\tilde{\theta} \in D} \sup_{k \in [0, L/\mu)} \mu || \sum_{i=k_0}^{k+k_0} (U(i)U(i)^T - R)\tilde{\theta} - s(i)U(i)|| \tag{3.13}$$

with L independent of μ and $D = \{\tilde{\theta} : ||\tilde{\theta}|| < ||\theta||\}$.

According to (3.13), the averaging error increases with μ, or alternatively, approaches zero as $\mu \to 0$. Thus, for sufficiently small μ, the averaged system approximates the original system sufficiently well. This claim is quantified in the appendix at the end of this chapter.

3.5 Convergence Cost Function

In this section we propose a cost function, which provides a measure of the 'expected' convergence rate of the averaged system. The meaning of 'expected' will be elaborated upon.

A measure of performance at sampling interval k is given by the squared Euclidean norm of the parameter error:

$$C_k \triangleq ||\tilde{\theta}^{av}(k)||_2^2 / ||\theta||_2^2$$

To enable analysis of overall performance we need to define a cost function which provides a suitable measure of the convergence of C_k to zero (assuming (3.11) is true). Consider the function

$$C_{0,N} \triangleq \sum_{k=0}^{N} ||\tilde{\theta}^{av}(k)||_2^2 / ||\theta||_2^2. \tag{3.14}$$

$C_{0,N}$ is nondecreasing in N and, provided (3.11) is true, as N approaches infinity, $C_{0,N}$ will approach a constant, the value of which is dependent on the convergence

rate. Therefore, a suitable convergence cost function is:

$$C \triangleq \sum_{k=0}^{\infty} ||\tilde{\theta}^{av}(k)||_2^2 / ||\theta||_2^2. \tag{3.15}$$

Similar cost functions have been proposed by other authors such as [70]. It can be interpreted as a (normalized) measure of the total power, over time, of the residual filter parameter vector, $\tilde{\theta}^{av}(k)$.

By combining (3.10) and (3.15) we obtain:

$$C = \sum_{k=0}^{\infty} ||(I - \mu R)^k \theta||_2^2 / ||\theta||_2^2. \tag{3.16}$$

$$= \sum_{k=0}^{\infty} \sum_{i=1}^{n} [(1 - \mu\lambda_i)^{2k} \frac{\gamma_i^2(\theta)}{||\theta||_2^2}] \tag{3.17}$$

where λ_i and $\gamma_i(\theta)$, $i = 1, 2, \ldots, n$, are the eigenvalues of R_n and the components of θ in the directions of the eigenvectors of R_n, respectively. Note that the eigenvectors of the symmetric matrix R_n, because of their orthonormal property, do not appear in (3.17).

Clearly, for a given input signal autocorrelation matrix R_n the value of C depends on the orientation of the unknown channel vector, θ, with respect to the eigenvectors of R_n. For example, for a given input signal, if θ lies along the direction of the eigenvector corresponding to a large/small eigenvalue of R_n then C will be large/small, that is, the convergence rate will be fast/slow. Our primary aim is to determine how convergence rate is affected, in general, by the filter dimension and input signal autocorrelation. To this end, we consider the 'expected' convergence rate over an ensemble of systems, each having the same dimension and input/disturbance signals, but differing in the coefficients of their unknown channel vector, θ. Considering (3.16), (3.17) a measure of the expected convergence rate is therefore given by:

$$\bar{C}_e \triangleq E_\theta[\sum_{k=0}^{\infty} \{\frac{||(I - \mu R)^k \theta||_2^2}{||\theta||_2^2}\}] = \sum_{k=0}^{\infty} \sum_{i=1}^{n} \{(1 - \mu\lambda_i)^{2k} E_\theta[\frac{\gamma_i^2(\theta)}{||\theta||_2^2}]\} \tag{3.18}$$

If we assume that θ has a flat distribution over the ensemble then:

$$E_\theta(\frac{\gamma_i^2(\theta)}{||\theta||_2^2}) = E_\theta(\frac{\gamma_i^2(\theta)}{\sum_{i=1}^{n} \gamma_i^2(\theta)}) = \frac{1}{n} \tag{3.19}$$

Combining (3.18) and (3.19) leads to:

$$\bar{C}_e = \frac{1}{n} \sum_{k=0}^{\infty} \sum_{i=1}^{n} (1 - \mu\lambda_i)^{2k} \tag{3.20}$$

$$= \frac{1}{2n} \sum_{i=1}^{n} \frac{1}{\mu\lambda_i} + \frac{1}{2n} \sum_{i=1}^{n} \frac{1}{2 - \mu\lambda_i} \tag{3.21}$$

Since, from Assumptions 5 and 6, $1 < 2 - \mu\lambda_i < 2,\ i = 1, 2, ..., n$, then

$$\frac{1}{2n\mu}Tr(R_n^{-1}) + \frac{1}{4} < \bar{C}_e < \frac{1}{2n\mu}Tr(R_n^{-1}) + \frac{1}{2} \tag{3.22}$$

In general, for sufficiently small μ, the additional terms $1/4$ and $1/2$ on the LHS and RHS, respectively, of (3.22) are negligible in comparison to:

$$C_e \triangleq \frac{1}{2n\mu}Tr(R_n^{-1}) \tag{3.23}$$

Assuming sufficiently small μ, we propose C_e of (3.23) as a transient performance cost function, the analysis of which follows.

3.6 Cost Function Analysis

In this section we analyse the cost function C_e of (3.23) under the condition that μ is fixed. This determines how the 'expected' convergence rate of the averaged LMS adaptive filter system is affected by the characteristics of $u(k)$ and the dimension, n, of the adaptive filter. In the next section we consider the case in which μ is adjusted to maintain asymptotic performance.

We begin by analysing the convergence cost function for signals which satisfy Assumptions 2-6. This is followed by analyses for a smaller class of input signals - those described by autoregressive processes.

Theorem 1 *Consider an averaged LMS adaptive FIR filter system, the dynamics of which are described by (3.10). Let n be the dimension of the FIR filter and C_e be the expected convergence cost functions defined in (3.23). Suppose Assumptions 2-6 of Section 3.2 are valid. Then,*

C_e is a nondecreasing function of n and, furthermore, is a nonincreasing function of n only when the input signal is discrete white or uncorrelated over time.

Proof: See Appendix F.1.

Theorem 1 leads to the following important result.

Result 1 *Unless the input signal, $u(k)$, is discrete white, the expected convergence rate will deteriorate with increasing dimension, n.*

Theorem 1 also leads to the conclusion that, for a given input signal and dimension of our adaptive filter system, the convergence cost function of (3.23) is bounded from above by:

$$C_e \leq \lim_{n\to\infty} \frac{Tr(R_n^{-1})}{2\mu n} \triangleq C_e(\infty) \tag{3.24}$$

The following Theorem extends the work by Gray [59] to obtain this bound.

Theorem 2 *Consider an averaged LMS adaptive FIR filter system, the dynamics of which are described by (3.10). Let n be the dimension of the FIR filter and C_e be the expected convergence cost functions defined in (3.23). Suppose Assumptions 2-6 of Section 3.2 are valid. Then for a given input signal and dimension, C_e is bounded from above by:*

$$C_e \leq \frac{1}{2\mu}\frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_{uu}^{-1}(\omega)d\omega \;=\; C_e(\infty) \tag{3.25}$$

where $\Phi_{uu}^{-1}(\omega)$ is the power spectrum of the input signal.

Proof: See Appendix F.2.

Any input signal dependence of $C_e(\infty)$ arises from the integral:

$$L_u \triangleq \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_{uu}^{-1}(\omega)d\omega \tag{3.26}$$

Note that, as can be seen by comparing (3.24) and (3.25), this integral is actually $\lim_{n\to\infty}[Tr(R_n^{-1})/n]$. The following Lemma determines, for a given signal power, the signal condition which minimizes this integral.

Lemma 1 *Consider a function $\Phi_{uu}(\omega)$, $-\pi \leq \omega \leq \pi$. Suppose the function is constrained by:*
(a) $\Phi_{uu}(\omega) > 0 \;\; \forall\omega$
(b)

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_{uu}(\omega)d\omega = \sigma^2. \tag{3.27}$$

Then, the integral:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_{uu}^{-1}(\omega)d\omega \tag{3.28}$$

is minimized by $\Phi_{uu}(\omega) = \sigma^2, \;\; \forall\omega$.

Proof: See Appendix F.3.

Lemma 1 leads to the important result.

Result 2 *For a given input signal power, the asymptotic (in n) expected convergence cost function, $C_e(\infty)$, is minimized when the input signal has a flat power spectrum, or, equivalently, is discrete white.*

Remark:

10. This result is clearly in agreement with Result 1.

The term L_u of (3.26), in fact, can be interpreted as a measure of the autocorrelation of $u(k)$. To see this let $A(\omega)$ be a filter which transforms $u(k)$ into a discrete white signal of variance σ^2. It follows that:

$$L_u = [1/2\pi \int_{-\pi}^{\pi} |A(\omega)|^2 d\omega]/\sigma^2$$

So L_u is a measure of the filter power required to whiten or decorrelate the signal $u(k)$.

Result 3 *For sufficiently large dimension, n, (or, more formally, in the limit as $n \rightarrow \infty$) the expected convergence cost function C_e increases with input signal autocorrelation.*

Summary:

(i) The expected convergence cost function, C_e,

(a) is independent of dimension, n, when the input signal is discrete white;

(b) increases with dimension when the input signal is autocorrelated (*i.e.* not discrete white).

(ii) As dimension increases, C_e increases towards a finite value $C_e(\infty) \triangleq \lim_{n\rightarrow\infty} C_e$, which

(a) is minimum for discrete white input signals;

(b) increases the more the input signal is autocorrelated.

3.6.1 Analysis for Autoregressive Input Signals

In this section we analyse the influence of dimension and input signal characteristics on the cost function C_e for the particular case in which the input signal is described

by an autoregressive (AR) model. Such models are typically used for voiced speech and, therefore, are of particular interest to the area of adaptive echo cancellation. We will use the following model:

$$u(k) = \frac{1}{A(q^{-1})} w(k) \tag{3.29}$$

where:

$A(q^{-1}) = 1 + a_1 q^{-1} + a_2 q^{-2} + ... + a_p q^{-p}$,

$w(k)$ is a discrete white zero mean signal of variance:

$$\sigma^2 = [1/2\pi \int_{-\pi}^{\pi} \frac{1}{|A(\omega)|^2} d\omega]^{-1} \tag{3.30}$$

The variance of $w(k)$ as given by (3.30) leads to $u(k)$ having unit variance.

Remark:

11. As is implied by (3.30), σ^2 is dependent on the AR parameters of the input signal.

As commented in the previous subsection, a measure of the correlation within such a signal is given by:

$$L_u = [1/2\pi \int_{-\pi}^{\pi} |A(\omega)|^2 d\omega]/\sigma^2 = (1 + a_1^2 + a_2^2 + ... + a_p^2)/\sigma^2 \tag{3.31}$$

The following Theorem quantitatively relates the expected convergence cost function, C_e, with dimension and AR parameters of the input signal.

Theorem 3 *Consider an averaged n^{th} dimension LMS adaptive FIR filter system, the dynamics of which are described by (3.10). Suppose the signal, $u(k)$, being input to the filter system is described by the p^{th} order AR model of (3.29). Then, for $n \geq p$, the cost function, C_e, of (3.23) is given by:*

$$C_e = \frac{1}{2\mu}[1 + (1 - \frac{2}{n})a_1^2 + (1 - \frac{4}{n}a_2^2) + ... + (1 - \frac{2p}{n})a_p^2]/\sigma^2, \quad \forall n \geq p \tag{3.32}$$

where σ^2 is as defined in (3.30).

Proof: See Appendix F.4.

Remark:

12. Using an extension of the Levinson algorithm, an iterative relationship can be developed to obtain C_e as an explicit function of a_i, n and μ also for $1 \leq n \leq p-1$. However, such a relationship is relatively complicated and provides little insight into the dependence of C_e on dimension and input signal autocorrelation.

13. Typically, the order of the AR model for voiced speech is relatively small ($p = 10-20$). So, in the application of adaptive echo cancellation, in which the filter dimension, n, is considerably larger ($n = 100-2000$), the condition in (3.32) that $n \geq p$ is not particularly limiting.

14. The cost function C_e, as given in (3.32), depends on the AR coefficients not only through the bracketted '[ ]' term but also through σ^2. By applying the iterative procedure indicated in [79], an explicit expression relating σ^2 to the AR coefficients can be obtained. For example, for $p = 2$:

$$\sigma^2 = (1 - a_2^2)\left[1 - \frac{a_1}{(1+a_2)^2}\right]$$

The expression for σ^2, however, becomes increasingly more complicated as the AR model order increases beyond $p = 2$. Consequently, as p increases it becomes more difficult, if not impossible, to gain insight into the general effect of the AR coefficients on σ^2. It is worth noting, however, that for AR modelled voiced speech, σ^2 is, typically, considerably smaller than unity.

The result of (3.32) indicates that:

$$\lim_{n\to\infty} C_e = \frac{1}{2\mu}[1 + a_1^2 + a_2^2 + ... + a_p^2]/\sigma^2 \tag{3.33}$$

So for sufficiently large dimension, C_e is determined by, and increases with the autocorrelation measure $L_u = [1 + a_1^2 + a_2^2 + ... + a_p^2]/\sigma^2$ of the input signal. Clearly, (3.33) is in agreement with the result of Theorem 2.

The variation of C_e with dimension, n, is easily verified from (3.32) to be:

$$C_e(n+1) - C_e(n) \triangleq \Delta_n C_e = \frac{1}{\mu n(n+1)}[\frac{a_1^2 + 2a_2^2 + ... + pa_p^2}{\sigma^2}] \tag{3.34}$$

As expected, $\Delta_n C_e$ is positive, that is, C_e increases with dimension, n. A loose link between the effect of n on C_e, as measured by $\Delta_n C_e$, and the autocorrelation level of the input signal, as measured by L_u, is indicated by (3.34):

$$\frac{L_u}{\mu n(n+1)} \leq \Delta_n C_e + \frac{1}{\mu n(n+1)\sigma^2} \leq \frac{pL_u}{\mu n(n+1)} \tag{3.35}$$

This suggests that the adverse effect of n on C_e may increase with autocorrelation level, L_u, and AR model order, p, of the input signal. A direct link can be achieved by considering input signals described by the first order AR (AR1(a)) model:

$$u(k) = \frac{1}{1 - aq^{-1}} w(k) \tag{3.36}$$

where:

$a \in (-1, 1)$;

$w(k)$ is a discrete white zero mean signal of variance $\sigma^2 = 1 - a^2$.

Note that the variance of $u(k)$ is unity. The parameter a provides a measure of autocorrelation of $u(k)$ not only through the relationship of (3.31):

$$L_u = (1 + a^2)/(1 - a^2),$$

but also through the expression for the autocorrelation function of $u(k)$:

$$r_l = \lim_{N \to \infty} \frac{1}{N} \sum_{k=0}^{N-1} u(k)u(k - l) = a^{|l|}.$$

That is, a larger value of $|a|$, or a^2, leads to a broader autocorrelation function/greater autocorrelation length.

Applying (3.32) we obtain:

$$C_e = \frac{1}{2\mu}[\frac{1 + (1 - \frac{2}{n})a^2}{1 - a^2}], \quad n \geq 1 \tag{3.37}$$

According to (3.37), the cost function C_e for an AR1(a) input signal increases with dimension, n, and with the input signal autocorrelation factor a^2. In fact, it can be easily verified that the variation of C_e with n and a^2 is given by:

$$\Delta_n C_e = \frac{1}{\mu n(n + 1)}[\frac{a^2}{1 - a^2}] \tag{3.38}$$

$$d_a C_e \triangleq \frac{dC_e(a)}{da^2} = \frac{1}{\mu}[\frac{1 - 1/n}{(1 - a^2)^2}] \tag{3.39}$$

Both $d_a C_e$ and $\Delta_n C_e$ are positive. Furthermore, $\Delta_n C_e$ increases with a^2 and $d_a C_e$ increases with n. This leads to the following result.

Result 4 *For AR1(a) input signals, the expected convergence cost function, C_e, increases with dimension, n, and with a^2, or input signal autocorrelation. The adverse effect of dimension on C_e is accentuated by higher input signal autocorrelation and vice versa.*

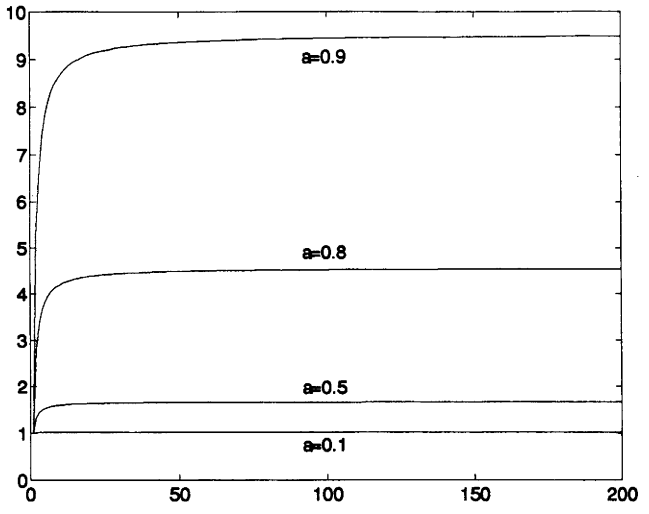


Figure 3.2: Plot of $Tr(R_n^{-1})/n$ vs n for AR1(a) input signals.

This result is supported by the plots in Figure 3.2 of

$$\frac{Tr(R_n^{-1})}{n} = \frac{1+(1-2/n)a^2}{1-a^2}$$

over n for various values of a.

Figure 3.2 indicates that for AR1(a) input signals the adverse effect of dimension on C_e diminishes rapidly with n. The expression for $\Delta_n C_e$ given in (3.34) indicates that the effect of n will also diminish for higher order AR input signals (more typical of voiced speech). A comparison between the expressions of $\Delta_n C_e$ for AR1(a) inputs, (3.38), and higher order inputs, (3.34), however, indicates that the diminishing effect for higher order AR inputs may not be as great as for AR1(a) inputs. This is also suggested by Figure 3.3 which includes plots of:

$$\frac{Tr(R_n^{-1})}{n} = [1+(1-\frac{2}{n})a_1^2+(1-\frac{4}{n}a_2^2)+...+(1-\frac{2p}{n})a_p^2]/\sigma^2 \qquad (3.40)$$

for the AR vectors:

$$AR1 = [1,\ 0.9];\ AR2 = [1,\ 0.9,\ 0.9^2,\ ...,\ 0.9^6];\ AR3 = [1,\ 0.9,\ ...,\ 0.9^{20}]. \qquad (3.41)$$

Through the application of the Yule-Walker method [80], the AR parameters for three different unit variance voiced speech segments were obtained. These are included in Table 3.2 together with the corresponding values for L_u, the autocorrelation measure

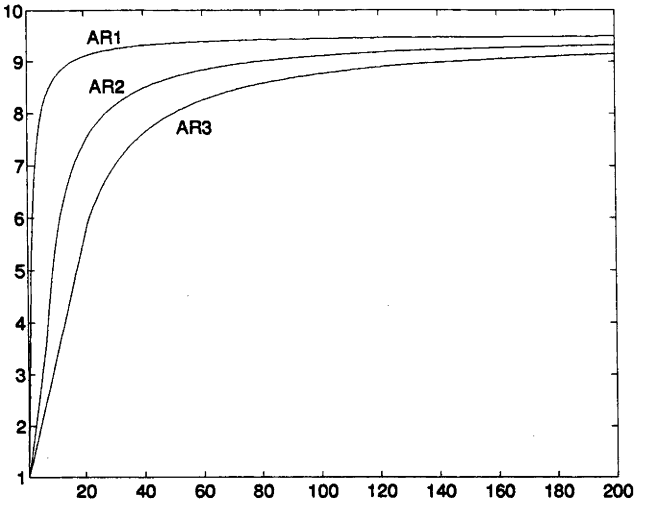


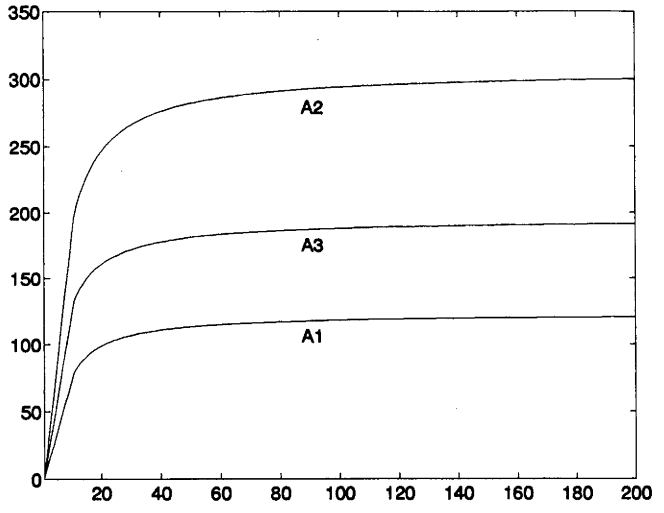
Figure 3.3: Plot of $Tr(R_n^{-1})/n$ vs n for AR inputs generated by the AR vectors of (3.41).

of (3.31). Figure 3.4 includes plots of $Tr(R_n^{-1})/n$ as given in (3.40) for these AR parameter sets. The plots indicate that as the dimension increases the expected convergence rate for the voiced speech inputs becomes considerably poorer than for discrete white inputs of the same variance - $Tr(R_n^{-1})/n = 1$, $\forall n$. This is due directly to the greater value of the autocorrelation measure

$$L_u = \lim_{n\to\infty} Tr(R_n^{-1})/n$$

for the voiced speech.

Figure 3.4 also indicates that the effect of n diminishes as n increases and, in particular, suggests that the effect of increasing n will become insignificant for $n > n_0 \approx 100$. However, it should be remembered that Figure 3.4 shows plots of $Tr(R_n^{-1})/n \approx 2\mu C_e$ and, therefore, only a small change in $Tr(R_n^{-1})/n$ may still lead to a significant change in C_e, particularly for sufficiently small μ.



Dimension, n

Figure 3.4: Plot of $Tr(R_n^{-1})/n$ vs n for AR inputs of Table 3.2.

Table 3.2: Voiced Speech AR parameter vectors - used for Figure 3.4.

AR filter	AR coefficients (1, a_1, a_2, ..., a_p)	L_u
A1	1.0, -1.6559, 1.6588, -1.0009, 0.7491, -0.4902, 0.4537, -0.1718, 0.2451, -0.1751, 0.0673	122.9
A2	1.0, -1.7480, 1.2405, -0.5314, 0.6742, -0.3161, -0.2506, 0.0494, 0.5712, -0.3844, 0.0743	306.4
A3	1.0, -1.7466, 1.3329, -0.5290, 0.4849, -0.4553, 0.0704, 0.4252, -0.0569, -0.3278, 0.2077	195.1

Summary of AR Signal Analysis:

(i) For first order AR (AR1(a)) input signals:

(a) The convergence cost function C_e increases with the filter dimension and the input signal autocorrelation;

(b) The adverse effect of dimension on C_e is accentuated by higher input signal autocorrelation and vice versa.

(ii) For higher order AR input signals:

(a) The convergence cost function C_e increases with dimension, n;

(b) As $n \rightarrow \infty$, C_e increases towards $C_e(\infty)$, the value of which increases with input signal autocorrelation;

(c) The rate at which C_e increases with n is dependent on the AR coefficients and, in particular, will tend to increase with input signal autocorrelation level and

AR model order.

(iii) In general for AR input signals, the adverse effect of dimension, n, on C_e diminishes with increasing n.

(iv) For AR modelled voiced speech inputs, the values of C_e indicate considerably poorer expected convergence rates than for discrete white inputs. This is due directly to the higher autocorrelation levels of voiced speech. The adverse effect of n on C_e diminishes with increasing n, although the adverse effect may remain significant for relatively large n.

3.7 Asymptotic Performance Analysis

A number of studies have essentially quantified asymptotic performance of the LMS adaptive FIR filter in the configuration of Figure 3.1. In this section, rather than attempt to carry out our own asymptotic analysis, we review the results of some of these previously conducted analyses.

Note that we are considering the asymptotic performance of the original LMS adaptive FIR filter system, as opposed to the averaged system. Although the averaged parameter vector $\tilde{\theta}^{av}(k) \rightarrow 0$, estimation noise causes the original parameter vector $\tilde{\theta}(k)$ to converge to a nonzero vector.

A measure of asymptotic performance is given by

$$E[||\tilde{\theta}(\infty)||_2^2] \triangleq \lim_{k\rightarrow\infty} E[||\tilde{\theta}(k)||_2^2] \tag{3.42}$$

where $||.||_2$ is the Euclidean norm. This measure is obtained by studying the evolution of $E[\tilde{\theta}(k)\tilde{\theta}(k)^T]$, the second order moments of $\tilde{\theta}(k)$. In order to do this quantitatively, however, it is necessary to invoke the following input signal independence assumption.

Assumption 7 *The input signal vectors $U(k)$ are independent and identically distributed.*

As commented previously, this assumption is not satisfied in the case of an LMS adaptive FIR filter system since U_k and U_{k-1} have $n-1$ elements in common. However, the errors due to the use of this assumption are relatively small if μ is sufficiently small [71], [75], [76], [77].

A summary of the quantitative results obtained to date for the performance measure of (3.42) is as follows. Note that we do not call upon the validity of Assumption 6 immediately.

Summary of Asymptotic Analysis Results

Consider the LMS adaptive FIR filter system introduced in Section 3.2. Let σ_u^2 and σ_s^2 be the variance of the input signal and disturbance signal, respectively.

(a) Suppose Assumptions 1-5, 7 are valid and that μ is sufficiently small so that the system is convergent.

(b) If ,in addition, to Supposition (a)

(i) the input signal is Gaussian;

(ii)the system satisfies the stationarity condition;

$$E[\tilde{\theta}(k+1)\tilde{\theta}(k+1)^T] = E[\tilde{\theta}(k)\tilde{\theta}(k)^T] \ \forall k \tag{3.43}$$

(iii) the LMS update stepsize is sufficiently small such that Assumption 6 is valid:

$$\mu \ll 1/(n\sigma_u^2) \tag{3.44}$$

then as derived in [30]

$$E[||\tilde{\theta}(\infty)||_2^2] = \mu n \sigma_s^2 \tag{3.45}$$

(c) If, in addition to Supposition (a)

(i) the input signal is Gaussian

then as derived in [70], [28]

$$E[||\tilde{\theta}(\infty)||_2^2] = \mu\sigma_s^2 \left[\sum_{i=1}^{n} \frac{1}{1-\mu\lambda_i}\right] \left[2 - \sum_{i=1}^{n} \frac{\mu\lambda_i}{1-\mu\lambda_i}\right]^{-1} \tag{3.46}$$

(d) If, in addition to Supposition (a)

(i) the disturbance is i.i.d

then as shown in [8]

$$E[||\tilde{\theta}(\infty)||_2^2] = \mu n \sigma_s^2 \left[2 - \frac{\mu E[||U(k)||_2^4]}{n\sigma_u^2}\right]^{-1} \tag{3.47}$$

or (ii) the disturbance is i.i.d and the input signal is Gaussian

then [8]

$$E[||\tilde{\theta}(\infty)||_2^2] = \mu n \sigma_s^2 \left[2 - \mu n \sigma_u^2 [1 + \frac{2Tr(R^2)}{(Tr(R))^2}]\right]^{-1} \tag{3.48}$$

Under Assumption 6, the RHS of (3.46) and (3.48) can be considerably simplified. In particular, for (3.46), application of the approximations:

$$\sum_{i=1}^{n} \frac{1}{1-\mu\lambda_i} = \sum_{i=1}^{n}[1+\mu\lambda_i+O(\mu^2\lambda_i^2)] \approx n(1+\mu\sigma_u^2) \tag{3.49}$$

$$\frac{\mu\lambda_i}{1-\mu\lambda_i} = \sum_{i=1}^{n}[\mu\lambda_i+O(\mu^2\lambda_i^2)] \approx \mu n\sigma_u^2 \tag{3.50}$$

leads to:

$$E[(||\tilde{\theta}_\infty||_2)^2] \approx \mu n\sigma_s^2 \frac{1+\mu\sigma_u^2}{2-\mu n\sigma_u^2} \tag{3.51}$$

$$\approx \frac{\mu n\sigma_s^2}{2} \tag{3.52}$$

Note the slight difference between (3.52) and (3.45).

Using the fact that:

$$0 < Tr(R^2) \leq (n\sigma_u^2)^2$$

with the result of (3.48) indicates that for Gaussian input signals

$$\mu n\sigma_s^2\left[2-\mu n\sigma_u^2\right]^{-1} \leq E[(||\tilde{\theta}_\infty||_2)^2] \leq \mu n\sigma_s^2\left[2-3\mu n\sigma_u^2\right]^{-1} \tag{3.53}$$

This result simplifies to (3.52) when the condition of (3.44) is imposed or Assumption 6 is valid. Finally, (3.47) suggests that when Assumption 6 is valid, the result of (3.52) should hold for most input signals which are not Gaussian.

It is important to note that the asymptotic results reported above do not depend on the autocorrelation levels of the input or disturbance signals, but only on the variance of each. There is, however, a strong link between asymptotic performance and filter dimension. In particular, asymptotic performance deteriorates with increasing dimension, n. For sufficiently small μ such that Assumption 6 is valid, the deterioration is linear in n. When Assumption 6 is not valid, the asymptotic performance deteriorates at a considerably faster rate with increasing n.

The above discussion leads to the following result.

Result 5 *Subject to the validity of Assumptions 1-7, in order to maintain asymptotic performance of the LMS adaptive FIR filter as dimension, n, increases, the update stepsize μ must be reduced linearly with n:*

$$\mu = \frac{\mu_0}{n\sigma_s^2}, \quad \textit{where } \mu_0 \textit{ is a constant.}$$

Remark:

15. A simplified interpretation of the NLMS algorithm is that it is the LMS algorithm with μ normalized via

$$\mu_{NLMS} = \frac{\mu_{LMS}}{n\sigma_u^2}$$

This suggests that the NLMS algorithm shows a similar dimension dependence to the LMS algorithm in which $\mu = \mu_{LMS}$ is adjusted to maintain asymptotic performance.

Adjusting μ according to Result 5 leads to the convergence cost function C_e taking the form:

$$\begin{aligned} C_e^{ass} \triangleq C_e(\mu = \mu_0/(n\sigma_s^2)) &= \frac{\sigma_s^2 n}{2\mu_0}\left[\frac{Tr(R_n^{-1})}{n}\right] \\ &\approx \frac{\sigma_s^2 n}{2\mu_0} L_u \quad \text{for large } n \end{aligned} \tag{3.54}$$

where L_u is the measure of input signal autocorrelation as given in (3.26), (3.31). Thus, for sufficiently large n, C_e^{ass} increases linearly with n. For the case in which the input signal is an autoregressive process, then, as indicated by (3.32), C_e^{ass} increases linearly with n, for all (positive) n:

$$\Delta_n C_e^{ass} = \left[\frac{\sigma_s^2}{2\mu_0}\right] L_u, \quad \text{for all } n \tag{3.55}$$

It is clear from (3.54), (3.55) that the adverse effect of increasing n depends strongly on the autocorrelation level of the input signal. This is demonstrated in Figure 3.5 which shows plots of

$$n\frac{Tr(R_n^{-1})}{n} = Tr(R_n^{-1})$$

for unit variance signals generated by the three autoregressive filters of Table 3.2. The plot for a white signal of unit variance can not be discerned from the horizontal axis, since for such a signal $Tr(R^{-1}) = n$. In particular, when μ is adjusted to maintain asymptotic performance, the cost function for these AR input signals increases with n at a rate of 100-300 times as fast as that for a white input signal of equal variance.

The above discussion is summarized by the following.

Result 6 *(i) When μ is adjusted to maintain asymptotic performance, the convergence cost function increases (the expected convergence rate decreases) essentially at*

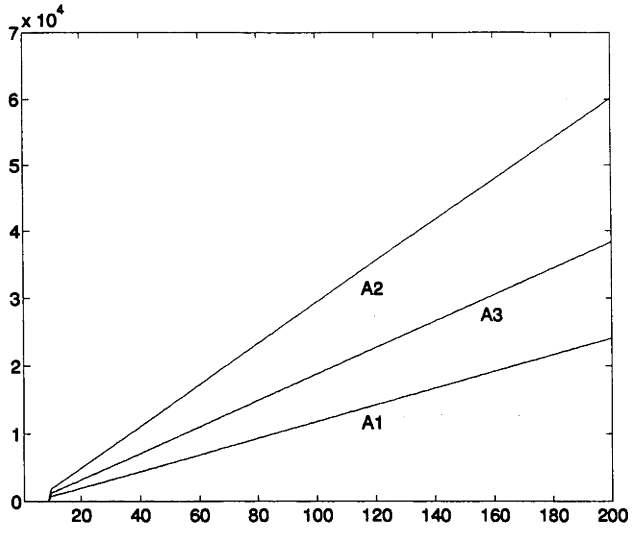


Figure 3.5: Plot of $Tr(R_n^{-1})$ vs n for AR inputs of Table 3.2.

a linear rate with increasing n.
(ii) The rate of increase is proportional to the autocorrelation level of the input signal.
(iii) The rate of increase is of the order of 100 times greater for voiced speech input signals than for white input signals of the same signal variance.

3.8 Conclusion

This chapter analysed the influence of filter parameter dimension and input signal autocorrelation on the convergence rate of an LMS adaptive FIR filter to a time invariant FIR modelled unknown channel. Averaging Theory was used to simplify the analyses. The convergence rate of the averaged system was shown to be dependent on the initial conditions of the adaptive filter system. Supposing that all initial conditions have equal probability, we proposed that a suitable measure of the expected convergence rate over an ensemble of initial conditions is provided by the cost function

$$C_e = \frac{1}{2\mu}\left[\frac{Trace(R_n^{-1})}{n}\right]$$

where μ is the LMS update constant, n is the filter parameter dimension and R_n is the $n \times n$ input signal autocorrelation matrix.

Analyses for the case in which μ is fixed indicated that

(a) when the input signal is discrete white, the convergence cost function C_e is independent of n;

(b) when the input signal is non-white, the cost function C_e increases (the expected convergence rate decreases) monotonically with dimension n towards a finite value, which is determined by and increases with the input signal autocorrelation.

When the input signal is described by an autoregressive (AR) model (typically used for voiced speech) the cost function, C_e, depends explicitly on the AR parameters, as well as on dimension and μ. Restricting the analyses to μ being fixed, it was shown that for a first order AR modelled input signal, the expected convergence rate deteriorates with input signal autocorrelation as well as with dimension. Furthermore, an increase in autocorrelation accentuates the adverse effect of dimension and vice versa. For higher order AR modelled input signals, an increase in input signal autocorrelation may also accentuate the adverse effect of dimension. Plots of C_e for AR modelled voiced speech inputs indicate considerably poorer expected convergence rates than for discrete white inputs. This is due to the high autocorrelation levels of voiced speech.

In general, for fixed μ, the adverse effect of n diminishes as n increases, although this will be less pronounced for higher order AR inputs. This observation is important because it suggests that for sufficiently large n, an increase in n may not cause a deterioration in the expected convergence rate. For AR modelled voiced speech inputs, however, the adverse effect may remain significant for relatively large n.

The cost function was also analysed for the case in which μ is adjusted to maintain asymptotic performance. These analyses indicated that under such circumstances the expected convergence cost function increases linearly with n, at a rate which is proportional to the autocorrelation level of the input signal. In particular, for highly autocorrelated input signals such as voiced speech, when asymptotic performance is maintained, an increase in n will cause a significant deterioration in the expected convergence rate, even for large n.

Chapter 3 Appendix: Averaging Error Analysis

In Section 6 we carried out transient performance analyses on an averaging approximation of the original system. The closeness of the approximation to the original system, or the averaging error, improves with a reduction of μ. The averaging error, however, may also depend on signal characteristics and/or filter dimension. In order to generate a confidence level in the relevance of the transient performance analyses of this chapter, we carry out in this appendix quantitative analyses of the averaging error.

We begin with a brief heuristic analysis. This is followed with a formal analysis in which we quantify the results suggested heuristically. To enable quantification - of the effects of dimension and signal characteristics on the averaging error - we introduce (i) an ensemble, having given properties, and (ii) an expectation operation of the averaging error over the ensemble. The use of such a combination leads to an explicit relationship between the expected averaging error and n-dimensional matrices involving signal moments. Examination of this relationship follows.

Heuristic Analysis

The averaging error vector $\tilde{\theta}^{av}(k+1) - \tilde{\theta}(k+1)$ is an n-dimensional vector. Therefore, depending on the norm used, the averaging error $||\tilde{\theta}^{av}(k+1) - \tilde{\theta}(k+1)||$ will tend to grow with n. For example,

$\lim_{k\to\infty} ||\tilde{\theta}^{av}(k+1) - \tilde{\theta}(k+1)||_1$ grows like n;

$\lim_{k\to\infty} ||\tilde{\theta}^{av}(k+1) - \tilde{\theta}(k+1)||_2$ grows like $n^{1/2}$;

where $||.||_1$ and $||.||_2$ denote the 1-norm and the Euclidean norm, respectively. The rate of growth will depend on the signal characteristics.

To determine the signal dependence of the averaging error we begin by combining the original and averaged system equations of (3.4) and (3.10) to obtain:

$$||\tilde{\theta}^{av}(k+1) - \tilde{\theta}(k+1)|| = ||A_{k+1}\theta + B_{k+1}|| \tag{3.56}$$

$$\leq ||A_{k+1}\theta|| + ||B_{k+1}|| \tag{3.57}$$

where:

$$A_{k+1} \triangleq (I - \mu R(k))(I - \mu R(k-1))...(I - \mu R(0)) - (I - \mu R)^{k+1} \tag{3.58}$$

$$B_{k+1} \triangleq [(I - \mu R(k))(I - \mu R(k-1))...(I - \mu R(1))\mu P(0) \tag{3.59}$$

$$
\begin{aligned}
&+ \quad (I-\mu R(k))(I-\mu R(k-1))...(I-\mu R(2))\mu P(1) \\
&+ \quad ..... \\
&+ \quad (I-\mu R(k))\mu P(k-1) \\
&+ \quad \mu P(k)] \\
R(k) \;&\triangleq\; U(k)U(k)^T \text{ and } P(k) \triangleq U(k)s(k).
\end{aligned}
$$

We restrict the heuristic analysis of signal characteristics to the simplified case in which there is no disturbance, that is, $B_k = 0$, $\forall k$. Expanding the RHS of (3.58) leads to:

$$
\begin{aligned}
A_k \;&=\; \{(-\mu k)[1/k \sum_{i=0}^{k-1}(R(i)-R)]\} \qquad (3.60) \\
&+ \quad \{(\mu k)^2[1/k^2 \sum_{i=0}^{k-1}\sum_{j=i+1}^{k-1}(R(j)R(i)-R^2)]\} \\
&+ \quad .....+\{(-\mu k)^k[1/k^k\,(R(k-1)R(k-2)...R(0)-R^k)]\}
\end{aligned}
$$

Assuming μ is sufficiently small, then over a $1/\mu$ time scale interval:

$$k \in [0, L/\mu), \quad L \text{ independent of } \mu$$

the averaging error term $||A_k\theta||$ will be dominated by:

$$(\mu k)||1/k \sum_{i=0}^{k-1}(R(i)-R)||, \quad \text{on a time scale } 1/\mu$$

that is, by the second order moments of $r_{k,l} = u(k)u(k-l)$. Similarly, it can be shown that over a $1/\mu$ time scale, $||B_k||$ is dominated by the second order moments of $p_{k,l} = s(k)u(k-l)$.

The averaging error term beyond a $1/\mu$ time scale:

$$A_j\theta, \quad j \in [mL/\mu, (m+1)L/\mu), \; m \geq 1$$

can be written in terms of the $1/\mu$ time scale error terms as follows:

$$
\begin{aligned}
A_{k+mL/\mu}\theta \;&=\; A_k(mL/\mu)\tilde{\theta}_{mL/\mu} \qquad (3.61) \\
&+ \quad \sum_{i=0}^{m-1}(I-\mu R)^k(I-\mu R)^{(m-i)L/\mu}A_{L/\mu}(iL/\mu)\tilde{\theta}_{iL/\mu}, \quad m \geq 1, \; k \in [0, L/\mu)
\end{aligned}
$$

where

$$A_k(iL/\mu) = (I-\mu R(k-1+iL/\mu))(I-\mu R(k-2+iL/\mu))...(I-\mu R(iL/\mu))-(I-\mu R)^k.$$

Thus, beyond a $1/\mu$ time scale, the averaging error term $||A_j\theta||$ is determined by $1/\mu$ time scale error terms $||A_{L/\mu}(iL/\mu)\tilde{\theta}(iL/\mu)||$ and by the rate of decay of $||(I-\mu R)^k||$. Equivalently, $||A_j\theta||_{j>L/\mu}$ is determined by the second order moments of $r_{k,l}$ and by the autocorrelation characteristics of $u(k)$.

Similarly, it can be shown that beyond a $1/\mu$ time scale, $||B_j||$ is determined by the second order moments of $r_{k,l}$ and of $p_{k,l}$ and by the autocorrelation function of $u(k)$.

In summary, these heuristic considerations suggest that, for sufficiently small μ, the averaging error:

decreases with a reduction in μ;

increases with dimension n;

depends on the second order moments of $r_{k,l} = u(k)u(k-l)$ and $p_{k,l} = s(k)u(k-l)$ as well as on the autocorrelation characteristics of the input $u(k)$.

Formal Analysis

In this section we quantify the effects of dimension and signal moments on the averaging error, or, more specifically, on the bound given in (3.57). To do this we consider the expectation of the bound over an ensemble of systems. The ensemble and expectation operation are described below.

Ensemble

(i) Each system is n-dimensional.

(ii) The input signal in all systems is a stochastic process defined on the same ergodic probability space $(\Omega_u,\ \Sigma_u,\ P_u)$.

(iii) The disturbance signal in all systems is a stochastic process defined on the same ergodic probability space $(\Omega_s,\ \Sigma_s,\ P_s)$.

(iv) The Euclidean length of the unknown channel vector $||\theta||_2$ is the same in each system.

(v) The unknown channel vector, θ, is uniformly distributed around the n-dimensional hypersphere of radius $||\theta||_2$.

Expectation Operation

$$E[E_\theta[||\tilde{\theta}(k+1) - \tilde{\theta}^{av}(k+1)||]] \leq E[E_\theta[||A_k\theta||_2]] + E[||B_k||_2] \qquad (3.62)$$

where:

$E_\theta[.]$ is the expectation with respect to θ;

$E[.]$ is the expectation with respect to the input and disturbance signals.

Remark:

16. Without a-priori knowledge of the unknown channel vector θ, we should assume that θ has equal probability of lying in any direction, hence the inclusion of property (v).

17. The bound on $||\tilde{\theta}(k+1) - \tilde{\theta}^{av}(k+1)||$, as given in (3.57), is dependent on the orientation of θ with respect to the eigenvectors of A_k. This dependence is not of interest to us. Instead, we are interested in how dimension and signal characteristics, in general, affect the averaging error. The approach we take to enable analysis of this general effect is to take the expectation of the averaging error with respect to θ.

The following Theorem provides a bound on each of the error terms

$$E[E_\theta[||A_k\theta||_2]], \quad E[||B_k||_2]$$

The result involves the Hilbert-Schmidt norm $||.||_{HS}$ defined as follows.

Let $A \in \mathcal{R}^{n\times n}$ and let a_{ij} be the ij^{th} element of A. Then

$$||A||_{HS} = \left[\frac{Tr(A^TA)}{n}\right]^{1/2} = \left[\frac{\sum_{j=1}^{n}\sum_{i=1}^{n} a_{ij}^2}{n}\right]^{1/2}$$

Theorem 4 *Consider the original LMS adaptive FIR system and the averaged LMS adapted FIR system described by (3.4) and (3.10), respectively and Assumptions 1-6. Let $||.||_2$, $||.||_{i2}$ and $||.||_{HS}$ be the Euclidean norm, the induced Euclidean norm and the Hilbert Schmidt norm, respectively. Let A_k and B_k be as defined in (3.58) and (3.59), respectively and $A_k(k_0)$, $B_k(k_0)$ be defined as:*

$$\begin{aligned}
A_k(k_0) &\triangleq (I-\mu R(k-1+k_0))(I-\mu R(k-2+k_0))...(I-\mu R(k_0)) - (I-\mu R)^k \\
B_k(k_0) &\triangleq [(I-\mu R(k-1+k_0))(I-\mu R(k-2+k_0))...(I-\mu R(1+k_0))\mu P(k_0) \\
&+ (I-\mu R(k-1+k_0))(I-\mu R(k-2+k_0))...(I-\mu R(2+k_0))\mu P(1+k_0) \\
&+ ..... \\
&+ (I-\mu R(k-1+k_0))\mu P(k-2+k_0) \\
&+ \mu P(k-1)]
\end{aligned}$$

Suppose

$$\sup_{k\in[0,L/\mu)} ||A_k||_{i2} + (1-\mu\lambda_{min}(R))^{L/\mu} < 1 \tag{3.63}$$

Then over the ensemble described by ensemble properties (i)-(v):

$$E[E_\theta[||A_k\theta||_2]] < \sup_{k\in[0,L/\mu)} \mu E[||\sum_{l=0}^{k} R(l) - R||_{HS}]||\theta||_2 \times \left\{(1+2\beta\mu)^{L/\mu}\frac{2-(1-\mu\lambda_{min}(R))^{L/\mu}}{1-(1-\mu\lambda_{min}(R))^{L/\mu}}\right\} \quad (3.64)$$

$$E[||B_k||_2] < \sup_{k\in[0,L/\mu)} \mu E[||\sum_{l=0}^{k} P(l)||_2] \times \left\{(1+(1+G)\beta\mu)^{L/\mu}\frac{2-(1-\mu\lambda_{min}(R))^{L/\mu}}{1-\sup_{k\in[0,L/\mu)}||A_k||_{i2}-(1-\mu\lambda_{min}(R))^{L/\mu}}\right\} \quad (3.65)$$

where:

$$\beta = \sup_k ||R(k)||_{i2} = \sup_k Trace(R(k))$$

$$G = \frac{E^{1/2}[||P(k)||_2^2]\sup_{k\in[0,L/\mu)} E^{1/2}[||\sum_{l=0}^{k} R(l) - R||_{i2}^2]}{\beta \sup_{k\in[0,L/\mu)} E[||\sum_{l=0}^{k} P(l)||_2]}$$

Therefore

$$\sup_k E[E_\theta[||\tilde{\theta}(k) - \tilde{\theta}^{av}(k)||_2] = O(\delta_1(\mu)) + O(\delta_3(\mu)) \quad (3.66)$$

where

$$\delta_1(\mu) \triangleq \sup_{k\in[0,L/\mu)} \mu E[||\sum_{l=0}^{k} R(l) - R||_{HS}]||\theta||_2, \quad (3.67)$$

$$\delta_3(\mu) \triangleq \sup_{k\in[0,L/\mu)} \mu E[||\sum_{l=0}^{k} P(l)||_2] \quad (3.68)$$

and $E[E_\theta[||\tilde{\theta}(k) - \tilde{\theta}^{av}(k)||_2]$ is the averaging error measure defined in (3.62).

Proof: See Appendix F.5.

Remark:

18. For sufficiently small μ, the supposition of (3.63) will be valid.

19. When μ satisfies the stronger constraint (than that imposed by Assumption 6):

$$\mu \ll \beta \quad (3.69)$$

we have

$$(1+2\beta\mu)^{L/\mu} \approx e^{2L\beta}, \quad (1-\mu\lambda_{min}(R_n))^{L/\mu} \approx e^{-L\lambda_{min}(R_n)}$$

The bracketted { } term of (3.64) is then

$$\left\{(1+2\beta\mu)^{L/\mu}\frac{2-(1-\mu\lambda_{min}(R))^{L/\mu}}{1-(1-\mu\lambda_{min}(R))^{L/\mu}}\right\} \approx \frac{e^{2L\beta}}{e^{L\lambda_{min}(R_n)}-1}$$

which increases with $\beta/[\lambda_{min}(R_n)]$, or, essentially, the eigenvalue spread of R_n. A similar result holds for the bracketted { } term of (3.65).

The result of (3.66) indicates that, for sufficiently small μ, the averaging error depends on the second order moments of

$$\sum_{j=0}^{k} u(j)u(j-l), \quad \sum_{j=0}^{k} s(j)u(j-l), \quad l=0,1,...,n-1, \quad k\in[0,L/\mu).$$

This is in agreement with the heuristic analysis results of the previous subsection.

Rather than analyze $\delta_1(\mu)$ and $\delta_3(\mu)$ we will analyze the closely related terms

$$\delta_2(\mu) \geq \delta_1(\mu), \quad \delta_4(\mu) \geq \delta_3(\mu)$$

defined by:

$$\delta_2(\mu) = \sup_{k\in[0,L/\mu)} \mu E^{1/2}\left[||\sum_{l=0}^{k} R(l) - R||_{HS}^2\right] \tag{3.70}$$

$$\delta_4(\mu) = \sup_{k\in[0,L/\mu)} \mu E^{1/2}\left[||\sum_{l=0}^{k} P(l)||_2^2\right] \tag{3.71}$$

The decision to do this is primarily for ease of analysis, but is validated by the following argument. The bound $\delta_2(\mu) \geq \delta_1(\mu)$ will become tighter as the variation over the ensemble of

$$\frac{||\sum_{l=0}^{k} R(l) - R||_{HS}}{E\left[||\sum_{l=0}^{k} R(l) - R||_{HS}\right]}$$

reduces. It is expected that, because of the smoothing effect of the summation within the numerator of this term, the variation will reduce as k increases. Now, as we shall see, $\delta_2(\mu)$ is essentially determined by

$$\mu E^{1/2}\left[||\sum_{l=0}^{k} R(l) - R||_{HS}^2\right] \text{ with } k = L/\mu.$$

Consequently, if μ is sufficiently small, then the bound $\delta_2(\mu) \geq \delta_1(\mu)$ should become sufficiently tight. Similar comments apply to $\delta_3(\mu), \delta_4(\mu)$.

The following theorem, which extends results of [69], [81], quantifies: $\delta_2(\mu)$ and $\delta_4(\mu)$.

Theorem 5 *Let $\delta_2(\mu)$ and $\delta_4(\mu)$ be as defined in (3.70) and (3.71). Let $r_{k,l} = u(k)u(k-l)$, $r_l = E[r_{k,l}]$, $p_{k,l} = s(k)u(k-l)$, $p_l = E[p_{k,l}] = 0$. Suppose:*

(i) $f_l(\omega)$, the power spectrum of $r_{k,l}$, exists and is twice differentiable at $\omega = 0$;

(ii) $g_l(\omega)$, the power spectrum of $p_{k,l}$, exists and is twice differentiable at $\omega = 0$.

Then

$$\delta_2(\mu) = (\mu L)^{1/2}\left\{\sum_{l=-n+1}^{n-1}(1-\frac{|l|}{n})[2\pi f_l(0)+O(\mu/L)^{1/2}]\right\}^{1/2} \tag{3.72}$$

$$\delta_4(\mu) = (\mu L)^{1/2}\left\{\sum_{l=0}^{n-1}[2\pi g_l(0)+O(\mu/L)^{1/2}]\right\}^{1/2} \tag{3.73}$$

where:

$$2\pi f_l(0) = \sum_{m=-\infty}^{\infty} E[(r_{k,l}-r_l)(r_{k+|m|,l}-r_l)],\quad 2\pi g_l(0) = \sum_{m=-\infty}^{\infty} E[(p_{k,l})(p_{k+|m|,l})] \tag{3.74}$$

Furthermore, if

(iii) the limits

$$\lim_{N\to\infty} 1/N \sum_{k=k_0}^{N-1+k_0}(r_{k,l}-r_l)(r_{k+|m|,l}-r_l) \triangleq E[(r_{k,l}-r_l)(r_{k+|m|,l}-r_l)]$$

$$\lim_{N\to\infty} 1/N \sum_{k=k_0}^{N-1+k_0}(p_{k,l})(p_{k+|m|,l}) \triangleq E[(p_{k,l})(p_{k+|m|,l})]$$

exist uniformly in k_0, then:

$$f_l(0) > 0,\quad g_l(0) > 0,\ \forall l \tag{3.75}$$

and, consequently, $\delta_2(\mu)$ and $\delta_4(\mu)$ increase with dimension, n.

Proof: See Appendix F.6.

Remark:

20. The twice differentiability condition on $f(\omega)|_{\omega=0}$ requires the existence of:

$$\frac{d^2 f(\omega)}{d\omega^2}|_{\omega=0} = -1/2\pi \sum_{m=-\infty}^{\infty}\{E[(r_{k,l}-r_l)(r_{k+|m|,l}-r_l)]\}m^2,$$

A sufficient condition for this is:

$$\sum_{m=-\infty}^{\infty}|E[(r_{k,l}-r_l)(r_{k+|m|,l})-r_l)]|m^2 < \infty, \tag{3.76}$$

As commented in [69], this imposes a covariance decay requirement on $r_{t,l}$. Similar comments can be made concerning $g(\omega)$ and $p_{k,l}$.

As indicated in (3.74), $2\pi f_l(0)$ and $2\pi g_l(0)$ are determined by summations of the covariance sequences of $r_{k,l}$ and $p_{k,l}$, respectively. Thus, $2\pi f_l(0)$ and $2\pi g_l(0)$ will tend to decrease as the covariance sequence/second order moments of $r_{k,l}$ and $p_{k,l}$, respectively, decay more quickly to zero.

This comment together with the results of Theorem 5 suggest the following:

Result 7 *1. The expected averaging error as defined in (3.62) will increase as the second order moments of $r_{k,l} = u(k)u(k-l)$ and $p_{k,l} = s(k)u(k-l)$ decay more slowly to zero.*
2. The expected averaging error will increase with dimension. This effect of dimension will tend to be accentuated by a slower rate of decay of the second order moments of $r_{k,l}$ and $p_{k,l}$.
3. The expected averaging error will increase like $\mu^{1/2}$.

The following Theorem considers the particular case of $u(k)$ and $s(k)$ each having a Gaussian probability distribution.

Theorem 6 *Let $\delta_2(\mu)$ and $\delta_4(\mu)$ be as defined in (3.70) and (3.71). Let the sequences $u(k)$ and $s(k)$ associated with $\delta_2(\mu)$ and $\delta_4(\mu)$ have autocorrelation functions given by $B_q = E[u(k)u(k-q)]$ and $D_q = E[s(k)s(k-q)]$, respectively.*
Suppose:
(i) $u(k)$ and $s(k)$ each have a Gaussian probability distribution;
(ii) the autocorrelation functions of $u(k)$ and $s(k)$ are each bounded above and below by an exponentially decaying function:

$$|B_q| \leq Ba_u^{|q|}, \ |D_q| \leq Da_s^{|q|}$$

where $B = B_q|_{q=0}$, $D = D_q|_{q=0}$ $0 \leq a_u, a_s < 1$.

Then

$$\begin{aligned} \delta_2(\mu) &\leq (\mu nBL)^{1/2}\{\left[\sum_{l=-n+1}^{n-1} \frac{1}{n}(1-\frac{|l|}{n})\left(2|l|a_u^{2|l|} + \frac{(1+a_u^2)(1+a_u^{2|l|})}{1-a_u^2}\right)\right] \\ &+ O(\mu/L)^{1/2}\}^{1/2} \qquad (3.77) \\ \delta_4(\mu) &\leq (\mu nBDL)^{1/2}\left\{\left[\frac{1+a_u a_s}{1-a_u a_s} + O(\mu/L)^{1/2}\right]\right\}^{1/2} \qquad (3.78) \end{aligned}$$

with equality when $u(k)$ and $s(k)$ are each described by first order AR (AR1(a)) processes:

$$B_q = Ba_u^{|q|}, \ D_q = Da_s^{|q|}, \ -1 < a_u, a_s < 1.$$

L is a constant independent of μ.

Proof: See Appendix F.7.

Plots of the bracketed term '[ ]' on the RHS of (3.77) over n for various values of a_u are shown in Figure 3.6. These indicate, together with (3.78) the following:

Result 8 *When the input and disturbance signals are Gaussian the effect of dimension on the averaging error factors $\delta_2(\mu)$ and $\delta_4(\mu)$ is accentuated by a slower decay rate of the autocorrelation functions of these signals. In particular, when both autocorrelation functions decay exponentially, $\delta_2(\mu)$ and $\delta_4(\mu)$ grow like $n^{1/2}$, the rate of growth increasing considerably with greater autocorrelation of each signal.*

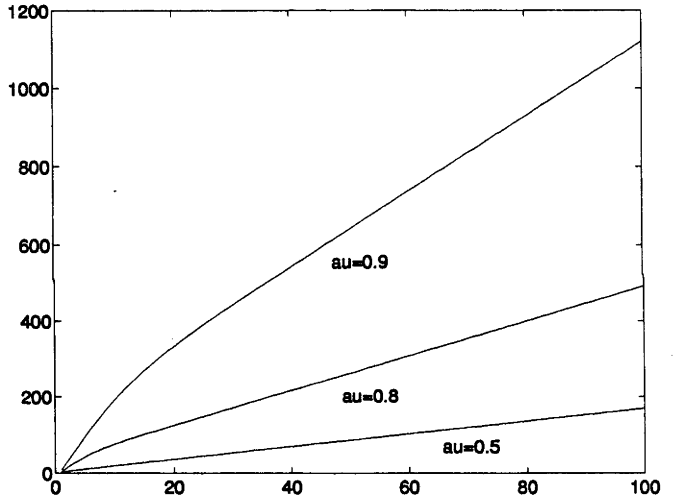


Figure 3.6: Plot of bracketed '[ ]' of (3.77) over n for various a_u - autocorrelation decay rate.

Summary

In this appendix we analyzed the error $||\tilde{\theta}(k) - \tilde{\theta}^{av}(k)||_2$ due to the averaging approximation made in Section 4. To enable quantification of the effects of dimension and signal characteristics on the averaging error, we considered the expectation of the averaging error over a defined ensemble of systems. It was shown that, for sufficiently small μ, the expectation value was bounded by a function which depended largely on

the second order moments of $\sum_{j=0}^{k} u(j)u(j-l)$, $\sum_{j=0}^{k} s(j)u(j-l)$ over a $1/\mu$ time scale. An explicit relationship was derived between the bound and these second order moments. It then followed that the bound increases:

(i) when the second order moments of $r_{k,l} = u(k)u(k-l)$ and $p_{k,l} = s(k)u(k-l)$ decay more slowly to zero;

(ii) with dimension, n;

(iii) like $\mu^{1/2}$.

For the particular case of $u(k)$ and $s(k)$ each having a Gaussian probability distribution, it was shown that the adverse effect of dimension on the averaging error bound is accentuated by the autocorrelation functions of $u(k)$ and $s(k)$ decaying more slowly to zero. More specifically, when the signal autocorrelation functions decay exponentially to zero, the bound increases like $n^{1/2}$, the rate of increase growing considerably as the autocorrelation functions decay more slowly.

The results above suggest that, in order to maintain a sufficiently small averaging error when moving from a relatively low dimension LMS/FIR system to a relatively high dimension LMS/FIR system, the size of μ should be reduced - particularly when the input and disturbance signals are such that the second order moments of $r_{k,l}$ and $p_{k,l}$ decay slowly to zero. In the particular case of Gaussian signals, the size of μ should be reduced according to

$$\mu = \frac{1}{Cn}$$

where C(constant) is dependent on and dramatically increases with the autocorrelation of the signals.

These results indicate that in order that the transient performance analyses of Section 6 remain relevant to the original system, the size of μ should be reduced as the dimension n increases. It may be felt that this requirement detracts from the usefulness of the transient performance results. However, as indicated in the asymptotic performance analyses, a similar reduction in μ is necessary to maintain asymptotic performance as dimension increases. It should be added, though, that the rate at which μ needs to be reduced, so as to maintain the same averaging error, increases with more slowly decaying fourth order signal statistics. The maintenance of asymptotic performance, in contrast, is dependent only on the power of the disturbance signal.

Chapter 4

Analysis of the LMS/FIR Filter in Closed Loop Echo Cancellation

4.1 Introduction

In the previous chapter we analyzed the dynamics of the LMS adaptive FIR filter in an open loop configuration. In effect, we considered only one end of an echo cancelling network and ignored the existence of the other end. In particular, we ignored the possibility that, due to the echo path at the other end of the network, the input signal $u(k)$ may contain a filtered (and delayed) version of the output signal $y(k)$. As might be expected, the existence of this feedback may lead to instability. It may also lead to the transient performance and asymptotic performance of the adaptive filter being more adversely affected by increasing signal correlation levels and/or filter dimension.

In this chapter, we carry out analyses which take into account this feedback. More specifically, we analyze the dynamics of the closed loop configuration of Figure 1.6 in which an unknown channel and a neighbouring LMS adaptive FIR filter are located at each end of the loop. Such a configuration is representative of the echo cancellation network of Figure 4.1, where the unknown channels are the echo paths located within the 2 to 4-wire hybrid or acoustic enclosure and the driving signals, $s_1(k), s_2(k)$, are the subscriber signals (speech and/or noise). The closed loop system we consider also allows for additive noise, $n_1(k), n_2(k)$ in the linking channels of the loop.

To reduce verbosity and simplify notation we will assume that the closed loop is an

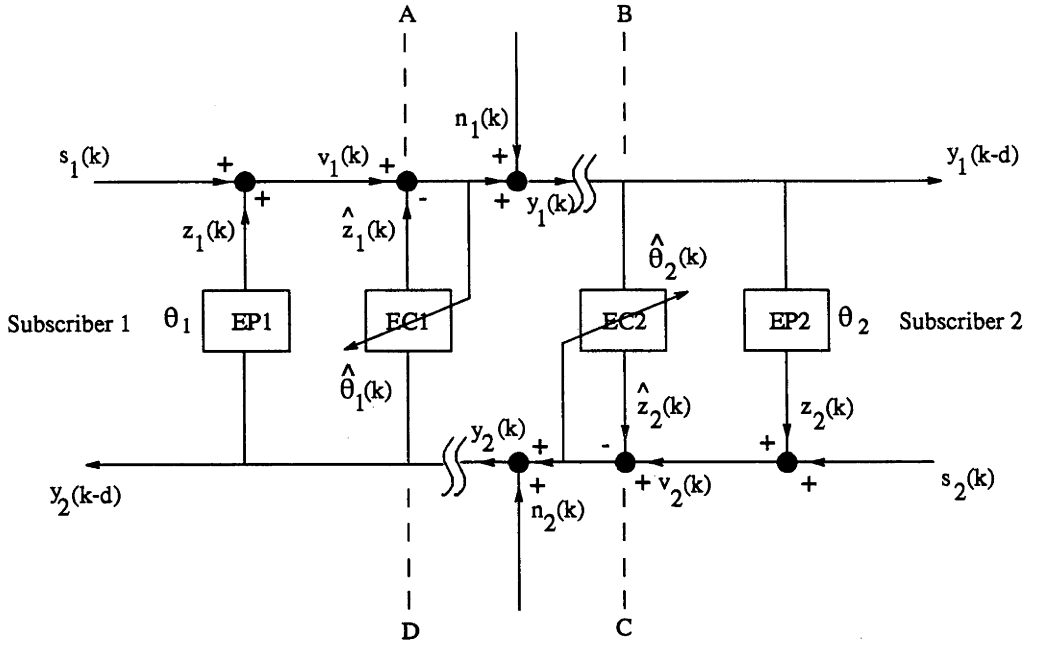


Figure 4.1: Double echo canceller network.

echo cancelling network which is being driven by subscriber signals. Of particular interest is the effect of the subscriber signal correlation levels and the dimension of the adaptive echo cancellers on the performance of the double echo canceller (**DEC**) system. By performance we mean the ability of the echo cancellers to suppress echoes generated by both echo paths.

The dynamics of the DEC system are nonlinear in nature and, in general, considerably more difficult to analyze quantitatively than the dynamics of the open loop system of Chapter 3. In order to carry out dynamical analyses we either invoke a number of simplifying assumptions or attempt only semi-formal analyses. Nevertheless, the results provide a very useful understanding of the effects of subscriber signal correlation levels and, to a lesser extent, echo canceller dimension on the performance of the DEC system.

The outline for the chapter is as follows. We begin in Section 2 by heuristically analysing the dynamics of the DEC system to highlight the effects of the feedback structure on performance. This is followed by a brief review of the results of studies conducted on closed loop systems similar to the DEC system being considered. After developing the system equations in Section 3, we carry out a semi-formal analysis in Section 4 on the DEC system to gain a semi-quantitative understanding of the effects of subscriber signal correlation levels and echo path/canceller dimension on

performance. To enable quantitative analyses to be carried out, we then consider a simplified DEC system in which the time invariant echo paths and adaptive echo cancellers are single tap FIR filters. The analyses carried out lead to the determination of bounds on the subscriber signal correlation levels within which good performance is obtained. In Section 6, with the aim of obtaining an explicit equality relationship between DEC performance and subscriber signal correlation levels, we conduct rigorous dynamical analyses on the single tap DEC system in which the transmission channels impose a single sample delay and the subscriber signals are assumed to be first order autoregressive processes.

Note: Appendix I lists the assumptions made in this chapter and other chapters.

4.2 Heuristics and Literature Review

In this section we heuristically analyse the DEC system to gain some insight into the effects of the feedback structure on performance. We follow this with a brief review of analyses conducted on the DEC system or similar closed loop adaptive systems.

Consider Figure 4.1 and, in particular, the end of the DEC system near subscriber 1. (We will ignore the channel noise signals, n_1 and n_2.) In its attempt to minimize the residual echo in y_1, echo canceller EC1 interprets any part of the transmitting channel signal, v_1, which is cross correlated with the receiving channel signal, y_2, as being part of the echo. However, v_1 contains not only the echo, z_1, but also the subscriber signal s_1. Thus if s_1 and y_2 are cross correlated, the echo canceller will adapt incorrectly. Signal y_2 largely contains s_2. Consequently, if s_1 and s_2 are cross correlated, we would expect biased adaptation. Also, y_2 contains an attenuated version of y_1, as a result of leakage via the far end echo path (EP2), and therefore contains an attenuated version of s_1. Thus biased adaptation of EC1 may also occur if s_1 shows a sufficiently broad autocorrelation function. A broad autocorrelation function of s_2 and significant cross correlation levels between s_1 and s_2 may, similarly, cause biased adaptation of EC2. In summary, the feedback configuration of the DEC system has the potential for causing poor asymptotic behaviour not only when the subscriber signals are cross correlated, but also when the such signals show sufficiently broad autocorrelation functions.

The transient performance of either echo canceller in a DEC system, as quantified in Chapter 3, is dependent largely on the autocorrelation level of the input signal. With the feedback characteristic of the DEC system, the input signal is composed not only

of the far end signal just transmitted, but also possibly of one or more delayed and filtered versions of the far end signal and/or near end signal. This suggests that poorer transient performance may occur with an increase in the autocorrelation level of one or both subscriber signals and an increase in cross correlation of these signals. As indicated by the analyses of Chapter 3, an increase in dimension of the echo cancellers should accentuate this signal dependence.

Various authors have carried out analyses to quantify these effects, particularly that pertaining to asymptotic performance. Earlier analyses [24], [21] considered closed loop adaptive echo cancelling systems in which only one of the closed loop ends was adaptive, the other end being modelled as a fixed attenuation. These analyses showed that unstable/bursting/poor asymptotic behaviour may be caused simply by large differences between the power of the subscriber signals. Later analyses [21], [22], [82] considered closed loop systems in which both ends were adaptive - that is, the DEC system we consider. The analyses indicated that stable/good asymptotic behaviour was achievable in this system under a greater range of circumstances. In particular, when the transmission channels impose a delay of at least one sampling interval it is shown that, for sufficiently small μ, good asymptotic performance is obtained if subscriber signals are zero mean white noise [82] or sufficiently rich and uncorrelated with each other [22]. It is also indicated in [22] that the convergence rate is exponentially fast at a rate bounded below by

$$(1 - \min\{\mu\lambda_{min}(R_n^{(i)})\})^k, \quad i = \{1, 2\}$$

where $R_n^{(i)}$ is the $n \times n$ autocorrelation matrix of the i^{th} subscriber signal.

In the analyses we carry out in the following sections, we strengthen and extend these results. We examine the case in which the subscriber signals are cross correlated as well as uncorrelated with each other. We quantify asymptotic and transient performance in terms of cross correlation and autocorrelation levels of the subscriber signals and, to a less extent, dimension of the echo paths/cancellers.

4.3 System Equations

In this section we develop the equations which describe the dynamics of the DEC system of Figure 4.1. We assume that both echo path/echo canceller pairs satisfy Assumption 1 of Chapter 3. That is, the echo paths are time invariant FIR filters with a tap length of n and the echo cancellers are LMS adaptive FIR filters also with

a tap length of n. Furthermore, the echo cancellers have zero initial conditions. We assume all signals are sampled. At sampling instant k subscriber 1 (2) sends signal $s_1(k)$ ($s_2(k)$) while receiving a delayed version of the signal $y_2(k)$ ($y_1(k)$); additive noise $n_1(k)$ ($n_2(k)$) occurs within the transmitting channel of subscriber 1 (2). The echo paths $EP1, EP2$ and echo cancellers $EC1, EC2$ are parametrized by the n-tap coefficient vectors

$$\begin{aligned} EP1: \quad \theta_1 &= (\theta_{1,0}\, \theta_{1,1}\, \theta_{1,2} ... \theta_{1,n-1})^T \\ EP2: \quad \theta_2 &= (\theta_{2,0}\, \theta_{2,1}\, \theta_{2,2} ... \theta_{2,n-1})^T \\ EC1: \quad \hat{\theta}_1(k) &= (\hat{\theta}_{1,0}(k)\, \hat{\theta}_{1,1}(k)\, \hat{\theta}_{1,2}(k) ... \hat{\theta}_{1,n-1}(k))^T \\ EC2: \quad \hat{\theta}_2(k) &= (\hat{\theta}_{2,0}(k)\, \hat{\theta}_{2,1}(k)\, \hat{\theta}_{2,2}(k) ... \hat{\theta}_{2,n-1}(k))^T. \end{aligned}$$

We make the following additional assumptions regarding the DEC network.

Assumption 8 *Each echo path is attenuating, that is:*

$$||\theta_1||_1 < 1, \quad ||\theta_2||_1 < 1 \tag{4.1}$$

where $||\ ||_1$ *is the 1-norm.*

Assumption 9 *Both transmission channels of the DEC system loop (i.e. channel* $A \rightarrow B$ *and channel* $C \rightarrow D$ *of Figure 4.1) impose a transmission delay of* $d \geq 1$ *sample intervals.*

Remark

1. Any difference Δd between the sample delays imposed by the transmission channels can be allowed for by assuming that the first Δd coefficients of the appropriate echo path vector θ_i are zero.

2. Under the highly unlikely condition of zero transmission delay, $d = 0$, the DEC systems forms an algebraic loop, which, in turn, increases the tendency of the system to show unstable/chaotic behaviour. See, for example, [22] for a more detailed discussion.

The signals received by subscribers 1,2 are, respectively:

$$y_1(k) = s_1(k) + n_1(k) + \tilde{\theta}_1(k)^T \phi_{2,k} \tag{4.2}$$

$$
\begin{aligned}
y_2(k) &= s_2(k) + n_2(k) + \tilde{\theta}_2(k)^T \phi_{1,k} \qquad (4.3)\\
\text{where } \tilde{\theta}_1(k) &= \theta_1 - \hat{\theta}_1(k), \quad \tilde{\theta}_2(k) = \theta_2 - \hat{\theta}_2(k)\\
\phi_1(k) &= (y_1(k-d)\ y_1(k-d-1)\ ...\ y_1(k-d-n+1))^T\\
\phi_2(k) &= (y_2(k-d)\ y_2(k-d-1)\ ...\ y_2(k-d-n+1))^T
\end{aligned}
$$

Applying the LMS algorithm to the echo cancellers leads to the following update equations for the residual echo parameters, $\tilde{\theta}_i(k)$:

$$\tilde{\theta}_1(k+1) = \tilde{\theta}_1(k) - \mu(y_1(k) - n_1(k))\phi_2(k) \qquad (4.4)$$

$$\tilde{\theta}_2(k+1) = \tilde{\theta}_2(k) - \mu(y_2(k) - n_2(k))\phi_1(k) \qquad (4.5)$$

where we have assumed the same LMS adaptation constant is used in both echo cancellers. As remarked in Chapter 3, μ should be chosen such that

$$\mu \ll min\{\frac{1}{n\hat{\sigma}_1^2(k)}, \frac{1}{n\hat{\sigma}_2^2(k)}\} \qquad (4.6)$$

where $\hat{\sigma}_1^2(k), \hat{\sigma}_2^2(k)$ is an estimate of the variance of $y_1(k), y_2(k)$, respectively.

Equations (4.2)-(4.5), together with the initial conditions:

$$\phi_1(0), \phi_2(0), \tilde{\theta}_1(0) = \theta_1, \tilde{\theta}_2(0) = \theta_2 \qquad (4.7)$$

describe the dynamics of the DEC system driven by external signals:

$$s_1(k), s_2(k), n_1(k), n_2(k)$$

The dynamics are clearly nonlinear in nature.

4.3.1 Signal Assumptions

Assumption 10 *The subscriber and channel noise signals are bounded and stationary so that the limits:*

$$Av_{sisj(l)} \triangleq \lim_{N\to\infty} 1/N \sum_{k=k_0}^{N-1+k_0} s_i(k)s_j(k-l), \quad \forall k_0, \forall l \quad i,j = \{1,2\} \qquad (4.8)$$

$$Av_{ninj(l)} \triangleq \lim_{N\to\infty} 1/N \sum_{k=k_0}^{N-1+k_0} n_i(k)n_j(k-l), \quad \forall k_0, \forall l \quad i,j = \{1,2\} \qquad (4.9)$$

$$Av_{sinj(l)} \triangleq \lim_{N\to\infty} 1/N \sum_{k=k_0}^{N-1+k_0} s_i(k)n_j(k-l), \quad \forall k_0, \forall l \quad i,j = \{1,2\} \qquad (4.10)$$

exist.

We also make the following assumption to simplify the analysis.

Assumption 11

$$\begin{array}{lrll} (a)\ Av_{ninj(l)} & = & \delta_{ij}Q_i & l=0 \\ & = & 0 & otherwise \\ (b)\ Av_{nisj(l)} & = & 0 & \forall l,\ i,j=\{1,2\} \end{array}$$

Remark:

3. Considering the typical characteristics of noise, Assumption 11(b) is most plausible. The validity of Assumption 11(a) requires the channel noise to be 'white'.

4.4 Analysis for FIR Echo Paths/Echo Cancellers of Arbitrary Dimension

In this section we present a semi-formal analysis of the dynamics of a DEC network in which the echo paths and echo cancellers are n-tap FIR filters, as opposed to the single tap FIR filters to be considered in the subsequent sections. The aim is to develop an understanding of the requirements of the subscriber signals for such high order DEC systems to perform well. The system equations which describe the dynamics of these higher order DEC systems are those of (4.2)-(4.5). To simplify the analysis, we will assume that channel noise is absent. Initially, we specialize to subscriber 1.

Combining (4.2)-(4.5) leads to:

$$\begin{aligned} \tilde{\theta}_1(k+1) &= \tilde{\theta}_1(k) - \mu[s_1(k)\phi_2(k) + (\tilde{\theta}_1(k)^T\phi_2(k))\phi_2(k)] \\ &= [I - \mu\phi_2(k)\phi_2(k)^T]\tilde{\theta}_1(k) \\ &\quad -\mu s_1(k)[\phi_1(k-d)^T\tilde{\theta}_2(k-d)\ \ \phi_1(k-1-d)^T\tilde{\theta}_2(k-1-d) \\ &\quad \ldots \phi_1(k-n+1-d)^T\tilde{\theta}_2(k-n+1-d)]^T \\ &\quad -\mu s_1(k)[s_2(k-d)\ \ s_2(k-1-d) \ldots s_2(k-n+1-d)]^T \end{aligned} \quad (4.11)$$

If μ is sufficiently small then $\tilde{\theta}_2$ (and $\tilde{\theta}_1$) will vary slowly with time and we may approximate $\tilde{\theta}_2(k-j)$ by $\tilde{\theta}_2(k)$, where $d \leq j \leq n-1-d$. As $\tilde{\theta}(k) = (\tilde{\theta}_1(k)\ \tilde{\theta}_2(k))^T$ is slowly varying, and $y(k) = (y_1(k)\ y_2(k))^T$ is essentially signal driven, we may approximate the $\tilde{\theta}$ update equation by averaging the signal dynamics. (This is

essentially what is done in the application of Averaging theory in Sections 5,6.) The resulting update equation is:

$$\begin{pmatrix} \tilde{\theta}_1(k+1) \\ \tilde{\theta}_2(k+1) \end{pmatrix} = \begin{pmatrix} I - \mu Av(\Phi_2(k)) & -\mu Av(s_1(k)\Psi_1(k-d)) \\ -\mu Av(s_2(k)\Psi_2(k-d)) & I - \mu Av(\Phi_1(k)) \end{pmatrix} \begin{pmatrix} \tilde{\theta}_1(k) \\ \tilde{\theta}_2(k) \end{pmatrix} + \begin{pmatrix} -\mu Av(s_1(k)\Upsilon_2(k)) \\ -\mu Av(s_2(k)\Upsilon_1(k)) \end{pmatrix} \tag{4.12}$$

where:

$\Phi_i(k)$ is the $n \times n$ correlation matrix $\phi_i(k)\phi_i(k)^T$;

$\Psi_i(k)$ is the $n \times n$ matrix $(\phi_i(k)\ \ \phi_i(k-1) \ldots \phi_i(k-n+1))$;

$\Upsilon_i(k)$ is the $n \times 1$ vector $(s_i(k-d)\ \ s_i(k-1-d) \ldots s_i(k-n+1-d))^T$

with $i = 1, 2$;

$Av(x(k)) = 1/M \sum_{k=0}^{M-1} x(k), \quad M \gg 1.$

(Note that the averaged factors involving y_1 or y_2 are still $\tilde{\theta}_1$ and $\tilde{\theta}_2$ dependent.)

Consider (4.12) rewritten as:

$$\tilde{\theta}(k+1) = (I - A(\tilde{\theta}(k)))\tilde{\theta}(k) + b, \quad \tilde{\theta}(k) = (\tilde{\theta}_1(k)\ \tilde{\theta}_2(k))^T$$

where the form of $A(\tilde{\theta}(k))$ and b can be easily ascertained from (4.12). Assuming that A is invertible, the asymptotic value of $\tilde{\theta}(k)$ is given by:

$$\lim_{k \to \infty} \tilde{\theta}(k) \triangleq \tilde{\theta}_\infty \approx (A(\tilde{\theta}_\infty))^{-1} b$$

The presence of cross correlation between the subscriber signals results in the bias term, b, being nonzero and, thus, causes the residual echo parameter vector to converge to a nonzero asymptotic value. As a consequence, echoes within the DEC system will never be completely cancelled.

Broader autocorrelation functions of the subscriber signals result in an increase in the off diagonal terms of the matrix A, that is:

$$Av(y_i(k-d)y_i(k-l-d)),\ 1 \le l \le n-1; \quad Av(s_i(k)y_i(k-j-2d)),\ 0 \le j \le 2(d+n-1)$$

This tends to lead to:

(a) an increase in the gain of the matrix A^{-1} and, thus, in the presence of cross correlation, results in $\tilde{\theta}_\infty$ lying further from the origin;

(b) slower convergence of the echo parameter vector, $\tilde{\theta}(k)$, to its asymptotic value.

In order to quantify more strongly these conclusions we apply the analysis techniques outlined in Appendix A. This leads to the following result.

Result 9 *Under the assumption of no undermodelling (Assumption 1), the residual echo parameter vector $\tilde{\theta}(k)$ will uniformly contract to within an l_1 ball $B : ||\tilde{\theta}||_1 \leq B$ at an exponential rate:*

$$(||\tilde{\theta}(k)||_1 - B) \leq (1 - \mu\nu)^k \tilde{\theta}(0)$$

if

$$\begin{aligned} 0 < \nu &< \min_{i=1,2} \min_{j \in [0,n-1]} \{Av(y_i(k-j-d)y_i(k-j-d)) \qquad (4.13) \\ &- \left[\sum_{m=0}^{n-1} |Av(y_i(k-m-d)y_i(k-j-d))| + |Av(s_i(k)y_i(k-j-m-2d))|\right]\} \\ \text{and } 0 < \nu &< 1/\mu \qquad (4.14) \end{aligned}$$

where

$$B \triangleq \frac{\sum_{j=0}^{n-1} \{|Av(s_1(k)s_2(k-j-d))| + |Av(s_2(k)s_1(k-j-d))|\}}{\nu} \qquad (4.15)$$

Remark:

4. Since the averaged factors involving y_1 and y_2 are dependent on $\tilde{\theta}_1$ and $\tilde{\theta}_2$, the conditions of (4.13) and (4.14) implicitly restrict the results to some domain D in residual echo parameter vector space:

   $$D : \tilde{\theta}(k), \tilde{\theta}(k) \in D, \ \forall k.$$

   This domain may be, in fact, of zero size.

The signals y_1 and y_2 largely contain the subscriber signals s_1 and s_2, respectively. Combining this relationship with the result above leads to the following remarks.

Remark:

5. Subscriber signals showing broader autocorrelation functions tend to lead to the upperbound on ν, as given in (4.13), becoming lower and even becoming negative. This suggests that more highly autocorrelated subscriber signals will show, for a given value of μ, slower convergence or may even show nonconvergent behaviour.

6. Assuming condition (4.13) is satisfied so that $\nu > 0$, then (4.15) suggests that when the subscriber signals are uncorrelated with each other, the residual parameter vector $\tilde{\theta}(k)$ will converge to the origin thus resulting in unbiased or 'optimal' asymptotic echo suppression within the DEC system. On the other hand, cross correlated subscriber signals tend to lead to incomplete or biased asymptotic echo suppression. The degree of bias increases not only with cross correlation but also with increasing autocorrelation which acts to decrease the denominator ν in (4.15).

7. For a given level of autocorrelation and cross correlation of the subscriber signals, an increase in the dimension n of the echo paths/echo cancellers leads to the bound on ν decreasing and B increasing. That is, poorer transient and asymptotic performance. This adverse effect of dimension is accentuated by greater correlation within and between the subscriber signals.

8. The autocorrelation and cross correlation functions of zero mean signals decay towards zero as the lag index increases. Combining this with the comments above and Result 9 suggests that the adverse signal correlation effects on performance decrease as the transmission channel delay d increases. In particular, Result 9 suggests that the asymptotic bias vanishes when the cross correlation function of the subscriber signals decays to zero before lag $l = d + n$. Furthermore, when the subscriber signals' autocorrelation functions are sufficiently narrow and the delay d sufficiently large, then the term

$$\sum_{m=0}^{n-1} |Av(s_i(k)y_i(k-d-j-m))|$$

of (4.13) also vanishes. In this case, the conditions of (4.13) and (4.14) resemble the equivalent l_1 norm conditions of a pair of decoupled open loop systems of Chapter 3.

4.5 Analysis of Single Tap Echo Paths/Cancellers

With the aim of obtaining a more quantitative understanding of the effects of subscriber signal correlation than that provided in the previous section, in this section we carry out formal analyses of a simplified DEC network in which the echo paths and echo cancellers are single tap FIR filters:

$$\theta_i \;=\; (\theta_i\, 0...0)^T; \quad i = \{1, 2\}$$

$$\hat{\theta}_i(k) \;=\; (\hat{\theta}_i(k)\; 0\; ...\; 0)^T; \quad i = \{1,2\}$$

We begin by considering the situation in which there is no cross correlation between the subscriber signals. This is typical of speech. To complete the analyses, we then consider the case of non zero cross correlation between the subscriber signals. In both cases we focus on the case in which the transmission channels impose single sample delay, $d = 1$. The results are then extended to include longer sample delays.

To enable application of standard analysis techniques (described in Appendix A), we use the Averaging Method, discussed in Appendix B, to develop a set of relatively simple equations whose solution approximates that of the system equations. Prior to developing these equations, we identify a domain in residual echo parameter, $\tilde{\theta}$, space to which the averaging approximation must be restricted. Our ultimate aim is to identify conditions under which the residual echo parameters, $\tilde{\theta}_1(k), \tilde{\theta}_2(k)$, converge as closely and as quickly as possible to $\tilde{\theta} = 0$ and hence yield good echo cancellation.

4.5.1 System Equations

The system equations of (4.2)-(4.5) for the single tap DEC system with transmission channel delay $d = 1$ take the simplified form:

$$\tilde{\theta}_1(k+1) \;=\; \tilde{\theta}_1(k) - \mu(y_1(k) - n_1(k))y_2(k-1) \tag{4.16}$$

$$\tilde{\theta}_2(k+1) \;=\; \tilde{\theta}_2(k) - \mu(y_2(k) - n_2(k))y_1(k-1) \tag{4.17}$$

$$y_1(k+1) \;=\; s_1(k+1) + n_1(k+1) + \tilde{\theta}_1(k+1)y_2(k) \tag{4.18}$$

$$y_2(k+1) \;=\; s_2(k+1) + n_2(k+1) + \tilde{\theta}_2(k+1)y_1(k) \tag{4.19}$$

while the initial conditions of (4.7) become:

$$y_1(0), y_2(0), \theta_1, \theta_2 \tag{4.20}$$

4.5.2 Domain of Analysis

To determine the domain D in residual echo parameter space over which the Averaging Method may be applied, consider (4.16),(4.17) with (4.20) rewritten as the matrix equation:

$$\tilde{\theta}(k+1) = \tilde{\theta}(k) - \mu f(\tilde{\theta}(k), y(k)); \quad \tilde{\theta}(k=0) = \theta, \;\; y(k=0) = y(0) \tag{4.21}$$

where

$$\tilde{\theta}(k) = \begin{pmatrix} \tilde{\theta}_1(k) \\ \tilde{\theta}_2(k) \end{pmatrix}; \quad \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}; \quad y(k) = \begin{pmatrix} y_1(k) \\ y_2(k) \end{pmatrix}; \tag{4.22}$$

$$f(\tilde{\theta}(k), y(k)) = \begin{pmatrix} f_1(\tilde{\theta}(k), y(k)) \\ f_2(\tilde{\theta}(k), y(k)) \end{pmatrix} = \begin{pmatrix} (y_1(k) - n_1(k))\, y_2(k-1) \\ (y_2(k) - n_2(k))\, y_1(k-1) \end{pmatrix} \tag{4.23}$$

and (4.18),(4.19) with (4.20) rewritten as the matrix equation:

$$y(k+1) = g(k+1) + G(\tilde{\theta}(k+1))y(k); \quad y(k=0) = y(0), \quad \tilde{\theta}(k=0) = \theta \tag{4.24}$$

where

$$g(k) = \begin{pmatrix} s_1(k) + n_1(k) \\ s_{2(k)} + n_2(k) \end{pmatrix}; \quad G(\tilde{\theta}(k)) = \begin{pmatrix} 0 & \tilde{\theta}_1(k) \\ \tilde{\theta}_2(k) & 0 \end{pmatrix} \tag{4.25}$$

The averaging method deals with equations of the form:

$$\tilde{\theta}(k+1) = \tilde{\theta}(k) - \mu f(\tilde{\theta}(k), k)$$

We massage our problem into this form by approximating $y(k)$ in (4.24)) by a signal vector, $\bar{y}(k, \tilde{\theta}(k))$ where, $\bar{y}(k)$ is defined by:

$$\bar{y}(k+1, z) = g(k+1) + G(z)\bar{y}(k, z); \quad \bar{y}(k=0) = y(0) \tag{4.26}$$

where $z = (z_1\ z_2)^T$ is constant. $\bar{y}(k, z)$ is well defined provided $G(z)$ is a stable matrix, that is: $|z_1 z_2| < 1$. Comparing (4.26) and (4.24) it can be seen that $\bar{y}(k, \tilde{\theta}(k))$ is a good approximation for $y(k)$. Indeed, as shown in Appendix G.1:

$$|\bar{y}(k, \tilde{\theta}(k)) - y(k)| = O(\mu), \quad \text{on a time scale} \quad 1/\mu \tag{4.27}$$

provided:

(a) μ is sufficiently small, and

(b) $G(\tilde{\theta}(k))$ is a stable matrix for every $\tilde{\theta}(k)$, on a time scale $1/\mu$.

Condition (b) requires $\tilde{\theta}$ to be restricted to the interior of the domain,

$$D: \ |\tilde{\theta}_1(k)\tilde{\theta}_2(k)| < \Theta < 1, \quad \text{on a time scale } 1/\mu$$

Assuming μ is sufficiently small, this is guaranteed by the validity of Assumption 8.

Replacing $y(k)$ in (4.21) by $\bar{y}(\tilde{\theta}(k))$ then yields:

$$\tilde{\theta}(k+1) = \tilde{\theta}(k) - \mu f(\tilde{\theta}(k), \bar{y}(k, \tilde{\theta}(k))) + O(\mu^2), \quad \text{on a time scale } 1/\mu \tag{4.28}$$

For the first order approximation on μ, the second term is irrelevant and can be ignored. Equation (4.28) is in the form suitable for the application of averaging.

4.5.3 Averaged Residual Echo Parameter Update Equations

We will now develop the averaged parameter update equations.

Combining (4.26) and (4.28), and specializing to $\tilde{\theta}_1$ results in:

$$\tilde{\theta}_1(k+1) = \tilde{\theta}_1(k) - \mu f_1(\tilde{\theta}(k), k), \quad \tilde{\theta}_1(k=0) = \theta_1 \tag{4.29}$$

$$f_1(\tilde{\theta}(k), k) = \tilde{\theta}_1(k)\bar{y}_2(k-1, \tilde{\theta}(k))\bar{y}_2(k-1, \tilde{\theta}(k)) + s_1(k)\bar{y}_2(k-1, \tilde{\theta}(k)) \tag{4.30}$$

Applying the Averaging Method yields:

$$\tilde{\theta}_1^{av}(k+1) = \tilde{\theta}_1^{av}(k) - \mu f_1^{av}(\tilde{\theta}^{av}(k)), \quad \tilde{\theta}_1^{av}(k=0) = \theta_1 \tag{4.31}$$

$$f_1^{av}(\tilde{\theta}^{av}) = \tilde{\theta}_1^{av} Av_{\bar{y}2\bar{y}2(0)}(\tilde{\theta}^{av}) + Av_{s1\bar{y}2(1)}(\tilde{\theta}^{av}) \tag{4.32}$$

where

$$Av_{\bar{y}2\bar{y}2(l)}(z) = \lim_{M\to\infty} 1/M \sum_{k=0}^{M-1} \bar{y}_2(k, z)\bar{y}_2(k-l, z), \tag{4.33}$$

$$Av_{s1\bar{y}2(l)}(z) = \lim_{M\to\infty} 1/M \sum_{k=0}^{M-1} s_1(k)\bar{y}_2(k-l, z) \tag{4.34}$$

with z constant in the summations.

A similar set of equations can be obtained for $\tilde{\theta}_2^{av}(k+1)$. According to Averaging Theory (see Appendix B), the solution, $\tilde{\theta}^{av}(k) = (\tilde{\theta}_1^{av}(k)\ \tilde{\theta}_2^{av}(k))^T$, to the averaged equations approximates $\tilde{\theta}(k) = (\tilde{\theta}_1(k)\tilde{\theta}_2(k))^T$ to within an error given by:

$$|\tilde{\theta}(k) - \tilde{\theta}^{av}(k)| = O(\delta(\mu)), \quad \text{on a time scale } 1/\mu \tag{4.35}$$

where

$$\delta(\mu) = \sup_{\tilde{\theta}\in D} \sup_{k\in[0,1/\mu)} \mu \Big| \sum_{m=0}^{k} f(\tilde{\theta}, m) - f^{av}(\tilde{\theta}) \Big| \tag{4.36}$$

with $f = (f_1\ f_2)^T, f^{av} = (f_1^{av}\ f_2^{av})^T$ as given in (4.30), (4.32), respectively, and the $\tilde{\theta}_2, \tilde{\theta}_2^{av}$ equivalents. Furthermore [78], if the averaged system has an asymptotically stable equilibrium within the domain D then the approximations hold on an infinite time scale.

Remark:

9. $\delta(\mu)$ is related to the stationarity of the signals and will be smaller when the signals show greater stationarity over the convergence window of the LMS algorithm.

4.5.4 Zero Cross Correlation Analysis

In this section we make the additional assumption:

Assumption 12 $Av_{s1s2(l)}, Av_{s2s1(l)} \equiv 0, \quad \forall l$

Remark:

10. Assumption 12 is typically valid for the case in which $s_1(k)$ and $s_2(k)$ are different speech signals.

By employing (4.26), and specializing initially to $\tilde{\theta}_1$, it can be shown that:

$$\begin{aligned} Av_{\bar{y}2\bar{y}2(0)}(z_1,z_2) &= \frac{1}{1-(z_1z_2)}[2z_2^2z_1Av_{s1\bar{y}2(1)}(z_1,z_2) \\ &\quad +2z_2z_1Av_{s2\bar{y}2(2)}(z_1,z_2) \\ &\quad +z_2^2(Av_{s1s1(0)}+Av_{n1n1(0)}) \\ &\quad +Av_{s2s2(0)}+Av_{n2n2(0)}] \qquad (4.37) \\ Av_{s1\bar{y}2(l)}(z_1,z_2) &= \sum_{n=0}^{\infty}(z_1z_2)^n[z_2Av_{s1s1(l+1+2n)}] \qquad (4.38) \\ Av_{s2\bar{y}2(l)}(z_1,z_2) &= \sum_{n=0}^{\infty}(z_1z_2)^n[Av_{s2s2(l+2n)}] \qquad (4.39) \end{aligned}$$

Combining (4.37)-(4.39) with (4.31),(4.32), and doing the same for the equivalent equations involved in the update of $\tilde{\theta}_2^{av}(k)$, we find that the adaptation of the averaged residual echo parameters is given by the following matrix equation:

$$\begin{pmatrix} \tilde{\theta}_1^{av}(k+1) \\ \tilde{\theta}_2^{av}(k+1) \end{pmatrix} = \begin{pmatrix} 1-(\alpha_{22}(\vartheta_k)+\gamma_{22}(\vartheta_k)) & -(\beta_{11}(\vartheta_k)+\eta_{11}(\vartheta_k)) \\ -(\beta_{22}(\vartheta_k)+\eta_{22}(\vartheta_k)) & 1-(\alpha_{11}(\vartheta_k)+\gamma_{11}(\vartheta_k)) \end{pmatrix} \begin{pmatrix} \tilde{\theta}_1^{av}(k) \\ \tilde{\theta}_2^{av}(k) \end{pmatrix} \qquad (4.40)$$

where

$$\begin{aligned} \vartheta_k &= \tilde{\theta}_1^{av}(k)\tilde{\theta}_2^{av}(k); \\ \alpha_{ii}(\vartheta_k) &= \frac{\mu}{1-(\vartheta_k)^2}[Av_{sisi(0)}+\sum_{n=0}^{\infty}(\vartheta_k)^n 2\vartheta_k Av_{sisi(2+2n)}] \qquad (4.41) \\ \beta_{ii}(\vartheta_k) &= \frac{\mu}{1-(\vartheta_k)^2}[\vartheta_k Av_{sisi(0)}+\sum_{n=0}^{\infty}(\vartheta_k)^n(1+(\vartheta_k)^2)Av_{sisi(2+2n)}] \qquad (4.42) \\ \gamma_{ii}(\vartheta_k) &= \frac{\mu}{1-(\vartheta_k)^2}Av_{nini(0)} \qquad (4.43) \\ \eta_{ii}(\vartheta_k) &= \frac{\mu}{1-(\vartheta_k)^2}\vartheta_k Av_{nini(0)} \qquad (4.44) \end{aligned}$$

with $i = \{1, 2\}$.

Standard analysis of (4.40)-(4.44) leads to the following theorem.

Theorem 7 *(1) Consider the single tap, single delay DEC system described by (4.16)-4.20) and Assumptions 1,8-11. (See Appendix I for a listing of Assumptions). Let* $||.||_1$ *denote the 1-norm. Suppose:*

(a) Assumption 12 is valid;

then, $\exists\ \mu^* > 0$ *such that* $\forall \mu \in [0, \mu^*]$ *the averaged update equations (4.40)-(4.44) are valid.*

(2) Consider the "averaged" system described by (4.40)-(4.44). Let Q_i *be as defined in Assumption 11. Suppose, in addition to Supposition (a):*

(b) $||(\theta_1, \theta_2)||_1 \leq 2(\Theta)^{1/2}$, *for some* $\Theta \in [0, 1)$;

(c) $\mu < \mu^*$;

(d)

$$|Av_{sisi(l)}| \leq R_i r_i^l \ \ i = \{1, 2\}, \ \ where \ \ 0 < r_i < 1, \ \ R_i = Av_{sisi(0)} \tag{4.45}$$

If, in addition,

$$\mu \ < \ \min_{i=1,2}(1 - \Theta^2)/[\frac{R_i(1 + \Theta r_i^2)}{1 - \Theta r_i^2} + Q_i] \tag{4.46}$$

$$\nu \ < \ \min_{i=1,2} \frac{1}{1 + \Theta}[\frac{R_i(1 - r_i^2(1 + 2\Theta))}{1 - \Theta r_i^2} + Q_i] \tag{4.47}$$

for some $0 < \nu < 1/\mu$

then,

(i) the "averaged" system will remain stable and the "averaged" parameter vector $(\tilde{\theta}_1^{av}(k)\ \tilde{\theta}_2^{av}(k))^T$, *will converge to the origin at a rate bounded above by*

$$||(\tilde{\theta}_1^{av}(k), \tilde{\theta}_2^{av}(k))||_1 \leq (1 - \mu\nu)^k ||(\theta_1, \theta_2)||_1.$$

(ii) the original system will remain stable, and its parameter vector, $(\tilde{\theta}_1(k),\ \tilde{\theta}_2(k))^T$, *will remain in the domain defined by*

$$D_0 : |\tilde{\theta}_1(k)\tilde{\theta}_2(k)| < \Theta, \ \ \forall k$$

and converge to within a neighbourhood, $O(\delta(\mu))$, *of the origin, where* $\delta(\mu)$ *is the averaging error as defined in (4.36).*

Proof: See Appendix G.2.

Theorem 7 refers to a DEC system in which there is a single sample delay along both transmission channels of the DEC system. The following corollary allows for an equal but arbitrary sample delay along each transmission channel, *i.e.* $A \rightarrow B$ and $C \rightarrow D$ of Figure 4.1.

Corollary 1 *Consider the DEC system of Theorem 7, but with the relaxed assumption that the transmission channels of the DEC system cause a delay of d sample intervals. Assume suppositions (a)-(d) of Theorem 7 are valid.*

If

$$\mu < \min_{i=1,2}(1-\Theta^2)/[\frac{R_i(1+\Theta r_i^{2d})}{1-\Theta r_i^{2d}} + Q_i] \tag{4.48}$$

$$\nu < \min_{i=1,2} \frac{1}{1+\Theta}[\frac{R_i(1-r_i^{2d}(1+2\Theta))}{1-\Theta r_i^{2d}} + Q_i] \tag{4.49}$$

for some $0 < \nu < 1/\mu$

then (i)the averaged system and (ii)the original system will exhibit the behaviour described in Theorem 7, (i) and (ii), respectively.

Proof: Trivial extension of the proof of Theorem 7.

Remarks

11. Supposition (d) of Theorem 7 requires only that the autocorrelation functions of the subscriber signals $s_1(k)$ and $s_2(k)$ are bounded above and below by exponentially decaying functions. Clearly, broader autocorrelation functions lead to larger values of r_1 and r_2.

12. For smaller values of r_i and Θ, the upperbounds on μ and ν in (4.46) and (4.47) are larger. This suggests that a DEC system is more likely to remain stable and the residual echo parameter vector, $(\tilde{\theta}_1(k),\ \tilde{\theta}_2(k))^T$, converge more rapidly towards the origin when:

    (a) the autocorrelation functions of both $s_1(k)$ and $s_2(k)$ are narrower;

    (b) the tap coefficient vectors of the FIR echo paths lie closer to the origin.

    Similar comments hold for (4.48) and (4.49).

13. The presence of noise causes the upper bound on μ to decrease. This indicates that as noise is introduced, smaller μ may be required if the condition of

either (4.46) or (4.48) is to be satisfied. The presence of noise, on the other hand, causes the upper bound on ν to increase which improves the likelihood of stability and the transient performance.

14. $\delta(\mu)$, because of its dependence on signal stationarity, is affected by the presence of noise. In general, $\delta(\mu)$ (and, therefore, the size of the neighbourhood around the origin to which the original echo parameter vector, $\tilde{\theta}(k)$, converges) increases with the introduction of noise.

15. The condition in (4.47) or (4.49) is not satisfied if:

$$\frac{Q_i}{R_i} < \frac{r_i^2(1+2\Theta)-1}{1-\Theta r_i^2}, \quad \frac{Q_i}{R_i} < \frac{r_i^{2(d+1)}(1+2\Theta)-1}{1-\Theta r_i^{2(d+1)}}$$

or, in the noiseless case, if:

$$r_i^2 > 1/(1+2\Theta) \text{ or } r_i^{2(d+1)} > 1/(1+2\Theta)$$

respectively. Although (4.47), (4.49) may be conservative, this suggests that not all combinations of Θ and r_i satisfying

$$0 < \Theta < 1; \quad 0 < r_i < 1$$

will yield exponentially stable behaviour of the "averaged" system. In particular, performance problems of the averaged system, and probably of the original system, are likely to occur when Θ is close to (but less than) unity (i.e. when the echo paths impose little attenuation) and the subscriber signals show relatively broad autocorrelation functions.

16. In general, because the subscriber signal autocorrelation levels $|Av_{sisi(l)}|$ tend to decrease as $|l|$ increases, the presence of longer transmission delays along the channels of the DEC system will assist in reducing the adverse effects of subscriber signal autocorrelation on echo canceller performance. This is indicated by the fact that, for given values of r_i, conditions (4.48) and (4.49) are more easily satisfied than (4.46) and (4.47). Furthermore, (4.48) and (4.49) are more easily satisfied as the delay d increases. In particular, if the autocorrelation functions of both subscriber signals decay to zero before lag $l = 2d$, then (4.48), (4.49) take the form:

$$\begin{aligned} \mu &< \min_{i=1,2}(1-\Theta^2)/[R_i+Q_i] \\ \nu &< \min_{i=1,2}[R_i+Q_i]/(1+\Theta) \end{aligned}$$

In essence, we see a vanishing of the adverse effects of subscriber signal autocorrelation, which arose due to the feedback nature of the DEC system.

4.5.5 Nonzero Cross Correlation (Noise Absent)

In this section, we no longer assume that there is zero cross correlation between the subscriber signals. However to simplify the mathematics we make the additional assumption that channel noise is absent:

Assumption 13 $n_1(k), n_2(k) = 0, \quad \forall k$

By employing (4.26) and specializing to $\tilde{\theta}_1$, it can be shown that:

$$\begin{aligned} Av_{\bar{y}2\bar{y}2(0)}(z_1,z_2) &= \frac{1}{1-(z_1z_2)^2}[2z_2^2z_1Av_{\bar{y}2s1(1)} + 2z_1z_2Av_{\bar{y}2s2(2)} \\ &\quad +z_2^2Av_{s1s1(0)} + Av_{s2s2(0)} \\ &\quad +2z_2Av_{s1s2(1)}] \qquad (4.50) \\ Av_{s1\bar{y}2(l)}(z_1,z_2) &= \sum_{n=0}^{\infty}(z_1z_2)^n[z_2Av_{s1s1(l+1+2n)} + Av_{s1s2(l+2n)}] \qquad (4.51) \\ Av_{s2\bar{y}2(l)}(z_1z_2) &= \sum_{n=0}^{\infty}(z_1z_2)^n[Av_{s2s2(l+2n)} + z_2Av_{s2s1(l+1+2n)}] \qquad (4.52) \end{aligned}$$

The matrix equation, equivalent to (4.40), when cross correlation between the subscriber signals is present, becomes:

$$\begin{aligned} \begin{pmatrix} \tilde{\theta}_1^{av}(k+1) \\ \tilde{\theta}_2^{av}(k+1) \end{pmatrix} &= \begin{pmatrix} 1-(\alpha_{22}(\vartheta_k)+\rho_{21}(\vartheta_k)) & -(\beta_{11}(\vartheta_k)+\xi_{12}(\vartheta_k)) \\ -(\beta_{22}(\vartheta_k)+\xi_{21}(\vartheta_k)) & 1-(\alpha_{11}(\vartheta_k)+\rho_{12}(\vartheta_k)) \end{pmatrix} \begin{pmatrix} \tilde{\theta}_1^{av}(k) \\ \tilde{\theta}_2^{av}(k) \end{pmatrix} \\ &\quad + \frac{1}{1-\vartheta^2}\begin{pmatrix} -\mu Av_{s2s1(1)} \\ -\mu Av_{s1s2(1)} \end{pmatrix} \qquad (4.53) \end{aligned}$$

where $\vartheta_k = \tilde{\theta}_1^{av}(k)\tilde{\theta}_2^{av}(k)$, $\alpha_{ii}(\vartheta_k)$, and $\beta_{ii}(\vartheta_k)$ are as given in (4.41), (4.42), and $\rho_{ij}(\vartheta_k)$ and $\xi_{ij}(\vartheta_k)$ are:

$$\begin{aligned} \rho_{ij}(\vartheta_k) &= \frac{\mu}{1-(\vartheta_k)^2}\tilde{\theta}_j^{av}[2Av_{sisj(1)} + \sum_{n=0}^{\infty}2\vartheta_k(\vartheta_k)^nAv_{sisj(3+2n)}] \qquad (4.54) \\ \xi_{ij}(\vartheta_k) &= \frac{\mu}{1-(\vartheta_k)^2}\tilde{\theta}_j^{av}[\vartheta_kAv_{sisj(1)} + \sum_{n=0}^{\infty}(1+(\vartheta_k)^2)(\vartheta_k)^nAv_{sisj(3+2n)}] \qquad (4.55) \end{aligned}$$

with $i,j = \{1,2\}; i \neq j$.

Standard analysis of (4.53)-(4.55),(4.41) and (4.42) leads to the following theorem.

Theorem 8 *(1) Consider the single tap, single delay DEC system described by (4.16)-(4.20) and Assumptions 1, 8-10. (See Appendix I for an easy reference to these assumptions.) Let* $||.||_1$ *be the 1-norm. Suppose:*

(a) Assumption 13 is valid,

then, $\exists\ \mu^* > 0$ *such that* $\forall \mu \in [0, \mu^*]$ *the averaged update equations (4.53)-(4.55),(4.41),(4.42) are valid.*

(2) Consider the "averaged" system described by these averaged equations. Suppose, in addition to Supposition (a):

(b) $||(\theta_1, \theta_2)||_1 \leq 2(\Theta)^{1/2}$, *for some* $\Theta \in [0, 1)$;

(c) $\mu < \mu^*$;

(d) $|Av_{sisi(l)}| \leq R_i r_i^l \ \ i = \{1, 2\}$, *where* $0 < r_i < 1$, $R_i = Av_{sisi(0)}$;

(e) $|Av_{s1s2(l)}|, |Av_{s2s1(l)}| \leq G g^l$ *where* $0 < g < 1$, $G = Av_{s1s2(0)} = Av_{s2s1(0)}$.

If, in addition

$$\mu < \min_{i=1,2}(1-\Theta^2)1/[\frac{R_i(1+\Theta r_i^2)}{1-\Theta r_i^2} + \frac{4(\Theta)^{1/2}Gg}{1-\Theta g^2}] \tag{4.56}$$

$$\nu < \min_{i=1,2} \frac{1}{1-(\Theta)^2}[\frac{R_i(1-\Theta)(1-r_i^2(1+2\Theta))}{1-\Theta r_i^2} - \frac{2(\Theta)^{1/2}Gg(2+\Theta+g^2)}{1-\Theta g^2}] \tag{4.57}$$

$$B \triangleq \frac{|Av_{s1s2(1)}| + |Av_{s2s1(1)}|}{(1-\Theta^2)\nu} \leq 2(\Theta)^{1/2} \tag{4.58}$$

for some $0 < \nu < 1/\mu$

then,

(i) the "averaged" system will remain stable and the averaged parameter vector, $\tilde{\theta}^{av}(k)$, *will converge to within the* l_1 *ball* B : $||(\tilde{\theta}_1^{av}, \tilde{\theta}_2^{av})||_1 < B$ *at a rate bounded above by*

$$||(\tilde{\theta}_1^{av}(k), \tilde{\theta}_2^{av}(k))||_1 - B \leq (1-\mu\nu)^k ||(\theta_1, \theta_2)||_1.$$

(ii) the original system will remain stable, and its parameter vector, $(\tilde{\theta}_1(k),\ \tilde{\theta}_2(k))^T$, *will remain in the domain defined by*

$$D_0 : |\tilde{\theta}_1(k)\tilde{\theta}_2(k)| < \Theta,\ \ \forall k$$

and converge to within the l_1 *ball:*

$$||(\tilde{\theta}_1, \tilde{\theta}_2)||_1 < B + O(\delta(\mu))$$

where $\delta(\mu)$ *is the averaging error as defined in (4.36).*

Proof: See Appendix G.3.

Corollary 2 *Consider the DEC system of Theorem 8. but with the relaxed assumption that the channels of the DEC system cause a delay of d sampling intervals in each transmission direction. Assume suppositions (a)-(e) of Theorem 8 are valid.*

If, in addition

$$\mu < \min_{i=1,2}(1-\Theta^2)1/[\frac{R_i(1+\Theta r_i^{2d})}{1-\Theta r_i^{2d}} + \frac{4(\Theta)^{1/2}Gg^d}{1-\Theta g^{2d}}] \tag{4.59}$$

$$\nu < \min_{i=1,2}\frac{1}{1-(\Theta)^2}[\frac{R_i(1-\Theta)(1-r_i^{2d}(1+2\Theta))}{1-\Theta r_i^{2d}} - \frac{2(\Theta)^{1/2}Gg^d(2+\Theta+g^{2d})}{1-\Theta g^{2d}}] \tag{4.60}$$

$$B \triangleq \frac{|Av_{s1s2(d)}| + |Av_{s2s1(d)}|}{(1-\Theta^2)\nu} \leq 2(\Theta)^{1/2} \tag{4.61}$$

for some $0 < \nu < 1/\mu$

then, (i)the averaged system and (ii)the original system will exhibit the behaviour described in Theorem 8, (i) and (ii), respectively.

Proof: Trivial extension of Theorem 8 proof.

Remarks:

17. The upperbounds on μ and ν in (4.56) and (4.57) increase as r_i, g, G and Θ decrease. Assuming condition (4.58) is satisfied, this suggests that the DEC system is more likely to remain stable and the averaged residual echo parameter vector, $(\tilde{\theta}_1^{av}(k)\tilde{\theta}_2^{av}(k))$, converge more rapidly to within the l_1 ball B when:
    (a) the autocorrelation functions of the subscriber signals are narrower;
    (b) the cross correlation levels between the subscriber signals are lower;
    (c) the echo path tap coefficient vector (θ_1, θ_2) lies closer to the origin (that is, the echo paths cause greater attenuation).
    Similar comments hold for (4.59)-(4.61). The dependence of transient performance on cross correlation levels, in addition to the autocorrelation levels, of the subscriber signals is due to the feedback structure of the DEC system. The dependence of transient performance on the echo path attenuation is intuitive since it governs how much the feedback structure of the DEC system accentuates the adverse signal effects.

18. Condition (4.58) is included to ensure that, assuming (4.56) and (4.57) are satisfied, the average parameter vector will remain inside the l_1 ball:

$$||(\tilde{\theta}_1^{av}\ \tilde{\theta}_2^{av})||_1 \leq 2(\Theta)^{1/2}$$

This simplifies the analysis as well as ensures that the averaged parameter vector remains within the averaging domain. Similar comments hold for (4.59)-(4.61).

19. A smaller value of B, which suggests better asymptotic performance, results from:

    (a) lower levels of subscriber signal cross correlation which helps not only by reducing the numerator but also by increasing the denominator in the left hand term of (4.58) (or (4.61));

    (b) narrower autocorrelation functions of the subscriber signals, which act to increase the size of the denominator in the left hand term of (4.58) (or (4.61)).

20. In the absence of cross correlation, that is,

$$Av_{s1s2(l)} = 0, \ \forall \ l \ \text{and} \ G = 0$$

    the conditions and results of Theorem 8 simplify to those of Theorem 7 (in the case of zero channel noise). Thus Theorem 7 (without noise) is a special case of Theorem 8.

21. When the subscriber signals are 'white' then, assuming the subscriber signals are not delayed versions of each other, then

$$Av_{s1s2(l)} = Av_{s2s1(l)} = 0, \ l \neq 0$$

    and, consequently, 'optimal' asymptotic performance should be obtained.

22. A comparison between Theorem 7 and Theorem 8 shows that the presence of cross correlation reduces both the likelihood of stability and the transient performance by reducing the upperbounds on both μ and ν. To more clearly illustrate the adverse effects of cross correlation on stability and transient performance, let us assume:

$$|Av_{s1s2(l)}|, |Av_{s2s1(l)}| \leq p \min_{i=1,2} R_i r_i^l, \quad \forall l \ \ p \in [0,1].$$

    A greater value of p implies a greater level of cross correlation between the subscriber signals. The condition of (4.57) then becomes:

$$\nu < \min_{i=1,2} \frac{R_i}{(1-\Theta^2)(1-\Theta r_i^2)}[(1-r_i^2(1+2\Theta))(1-\Theta)-2\Theta^{1/2}pr_i(2+\Theta+r_i^2)] \quad (4.62)$$

    Stability is guaranteed when ν, and therefore the bracketted term, [ ], of the RHS of (4.62), is larger than zero. Figure 4.2 shows r *vs* Θ plots of [ ] = 0 for various values of p. For each value of p, all (r, Θ) points lying below the

plotted line satisfy [ ] > 0. As indicated in Figure 4.2, an increase in p, resulting from an increase in cross correlation levels, leads to a reduction in size of the region in (r, Θ) space which satisfies [ ] > 0. Consequently, an increase in cross correlation levels leads to a decrease in the likelihood of stability and good transient behaviour.

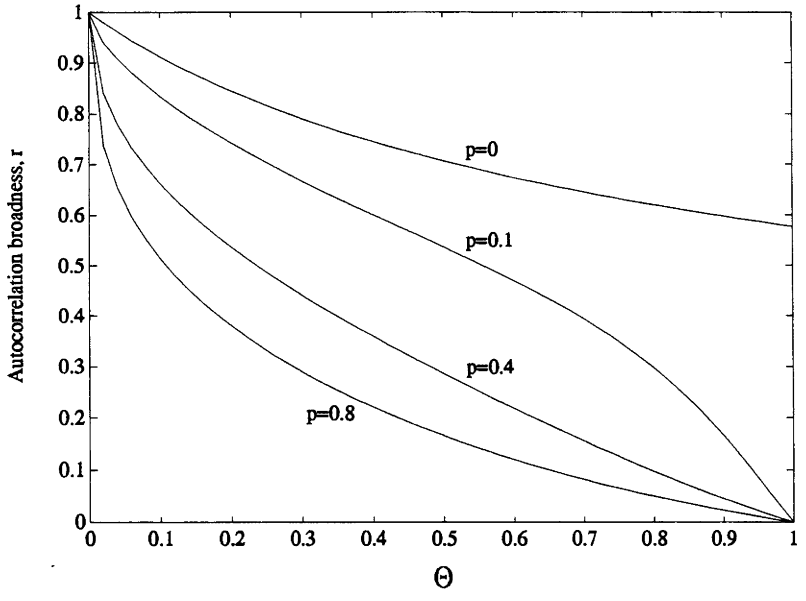


Figure 4.2: Upper edges of uniform contraction regions for varying amounts of cross correlation (given by p) - eqn (4.62)

23. In general, because subscriber signal cross correlation levels $|Av_{s1s2(l)}|$ and autocorrelation levels $|Av_{sisi(l)}|$ tend to decrease as $|l|$ increases, the presence of longer transmission delays will assist in reducing the adverse effects of subscriber signal cross correlation and autocorrelation on echo canceller performance. This is indicated by the fact that, for a given value of r_i and g, conditions (4.59), (4.60) and (4.61) are more easily satisfied as the delay, d, increases and are more easily satisfied than (4.56),(4.57) and (4.58). In particular, if the cross correlation function $Av_{s1s2(l)}$ and the autocorrelation functions $Av_{s1s1(l)}, Av_{s2s2(l)}$ of the subscriber signals decay to zero before lag $l = d$ and $l = 2d$, respectively, then the (4.59)-(4.61) take the form:

$$\begin{aligned} \mu &< \min_{i=1,2}(1-\Theta^2)/R_i \\ \nu &< \min_{i=1,2} R_i/(1+\Theta) \\ B = 0 &\leq 2(\Theta)^{1/2} \end{aligned}$$

Thus, unbiased asymptotic echo suppression is obtained and the dynamics of the closed loop DEC system resemble the dynamics of a pair of decoupled open loop systems.

4.5.6 Simulations

We conclude the examination of single tap DEC systems by presenting a number of 'simple' simulations which illustrate the effects of broad autocorrelation functions and the presence of cross correlation on the performance of the system. The simulation conditions are:

Adaptive system	Single tap, single delay DEC system
Adaptation stepsize	$\mu = 0.002$
Initial conditions	$\tilde{\theta}_1(0) = \theta_1 = 0.8,\ \tilde{\theta}_2(0) = \theta_2 = -0.7,\ y_1(0) = 0,\ y_2(0) = 0$
Subscriber signals	AR1(a) of (3.36) with a given in Table 4.1

In simulation (a), the subscriber signals are non-cross correlated AR1(0) or discrete white signals. In simulation (b), both subscriber signals show broader autocorrelation functions than in (a), but again have zero cross correlation. In simulation (c), the subscriber signals show not only relatively broad autocorrelation functions but also a relatively broad cross correlation function.

The simulation results are given in Figure 4.3 which shows plots of the variation in residual echo level, $z_1(k) - \hat{z}_1(k) = \tilde{\theta}_1(k)y_2(k-1)$, over time at one end of the DEC loop. The residual echo level over time at the other end of the loop was observed, in all cases, to be similar to that shown. As indicated by a comparison between Figure 4.3a and 4.3b, broader subscriber signal autocorrelation functions may lead to slower convergence rates, while Figure 4.3c suggests that the presence of subscriber signal cross correlation may lead to incomplete echo cancellation.

Table 4.1: Single tap DEC Simulations - Description of subscriber signals.

Sim.	Subscriber Signals*		Autocorre lation fn.s$^+$		Cross Correlation fn.
no.	s1	s2	As1(j)	As2(j)	C(j)
(a)	AR1(0)	AR1(0)	$\sim \delta_{0,j}$	$\sim \delta_{0,j}$	~ 0
(b)	AR1(0.8)	AR1(0.5)	$\sim 0.8^j$	$\sim 0.5^j$	~ 0
(c)	AR1(0.7)	s1	$\sim 0.7^j$	$\sim 0.7^j$	$\sim 0.7^j$

* AR1(a) = autoregressive sequence as in eqn. (3.36)

$^+$ $\delta_{0,0} = 1,\ \delta_{0,j} = 0$ when $j \neq 0$.

4.5.7 Summary

The examination of the single tap DEC system conducted in this section indicates the following.

(i) Broader autocorrelation functions and greater cross correlation levels of the subscriber signals reduce the convergence rate of the DEC system. Sufficiently broad autocorrelations functions and/or high cross correlation levels can lead to nonconvergent behaviour, irrespective of the update stepsize μ.

(ii) An increase in cross correlation levels of the subscriber signals causes a deterioration in asymptotic performance. This is accentuated by high autocorrelation levels of the subscriber signals.

(iii) The adverse signal effects on transient and asymptotic performance are accentuated when the echo paths impose little attenuation.

(iv) In general, when the subscriber signals are 'white' then good asymptotic performance is obtained.

(v) For a given update stepsize μ, which is sufficiently small, an increase in DEC channel noise may lead to better transient, but poorer asymptotic performance.

(vi) An increase in the delay imposed by the DEC channels tends to reduce the adverse effects of broad autocorrelation functions and high cross correlation levels of the subscriber signals. Furthermore, when the subscriber signals' correlation functions are sufficiently narrow and the delay is sufficiently large, the dynamics of the DEC system simplify to a pair of decoupled open loop systems of Chapter 3.

4.6 Single Tap Single Delay DEC System with AR Subscriber Signals

The analyses conducted in the previous sections of this chapter have provided, at best, bounds on the effects of subscriber signal correlation levels on the performance of the DEC system. In this section we aim to obtain an explicit equality expression relating performance to correlation levels. In order to achieve this, we reconsider the single tap, single delay DEC system examined in Section 4.5 under the additional assumption that the subscriber signals are first order autoregressive processes as described by (3.36) of Chapter 3.

The analyses are organized as follows. We begin by introducing a number of quantitative assumptions on the subscriber signals in addition to the autoregressive assump-

tion. Initially, the subscriber signals are assumed to be of equal power. Averaging Theory is again applied to obtain an approximate simpler system, analysis of which follows. Finally we briefly explore the case in which the subscriber signals are of unequal power.

4.6.1 Assumptions

In addition to Assumptions 1, 8, 9 and 13 we impose the following assumptions on the subscriber signals.

Assumption 14 *The subscriber signals are described by the first order autoregressive processes:*

$$\begin{array}{rcll} s_1(k+1) & = & as_1(k) + (1-a^2)^{1/2} w_1(k) & s_1(0) = 0 \\ s_2(k+1) & = & as_2(k) + (1-a^2)^{1/2} w_2(k) & s_2(0) = 0 \end{array} \tag{4.63}$$

where: $|a| < 1$ *and* $w_1(k)$ *and* $w_2(k)$ *are wide sense stationary discrete white signals.*

Furthermore, the signals $w_1(k)$ and $w_2(k)$ are such that the following assumptions hold for $i, j = \{1, 2\}$, $i \neq j$:

Assumption 15 *Bounded signals:*

$$|w_i(k)| < W, \ \forall k$$

Assumption 16 *Zero mean property:*

$$|\frac{1}{M} \sum_{k=m}^{M+m-1} w_i(k)| \leq \frac{C}{\sqrt{M}} \ \forall m, M$$

Assumption 17 *Well defined autocorrelation properties:*

$$|\frac{1}{M} \sum_{k=m}^{M+m-1} w_i(k) w_i(k-l) - V^2 \delta_{0,l}| \leq \frac{C}{\sqrt{M}} \ \forall m, M, l$$

Assumption 18 *Well defined cross correlation properties:*

$$|\frac{1}{M}\sum_{k=m}^{M+m-1} w_i(k)w_j(k-l) - pV^2\delta_{0,l}| \leq \frac{C}{\sqrt{M}} \ \forall m, l, M \ \textit{where} \ |p| \leq 1$$

where the constant C is positive and independent of the integers m, l and M in the above inequalities.

Remarks:

24. The numerator $(1-a^2)^{1/2}$ in (4.63) is included so that the signals $s_1(k)$, $s_2(k)$, $w_1(k)$, $w_2(k)$ have the same power (l_2 norm).

25. The signals $w_i(k)$ are deterministic in the sense that Assumptions 16-18 do not refer to a probability distribution or to convergence in a probabilistic sense.

The wide sense stationarity of $w_1(k)$ and $w_2(k)$ implies that $s_1(k)$ and $s_2(k)$ are also wide sense stationary signals. Because of the stability of the AR filter involved the subscriber signals $s_i(k)$ enjoy properties similar to the signals $w_i(k)$. More specifically, for some positive constant S only depending on a, W and V :

$$|s_i(k)| \leq \sqrt{\frac{(1+|a|)}{(1-|a|)}}W, \ \forall k \tag{4.64}$$

$$|\frac{1}{M}\sum_{k=m}^{M+m-1} s_i(k)| \leq \frac{S}{\sqrt{M}} \ \forall m, M \tag{4.65}$$

$$|\frac{1}{M}\sum_{k=m}^{M+m-1} s_i(k)s_i(k-l) - a^lV^2| \leq \frac{S}{\sqrt{M}} \ \forall m, M, l \tag{4.66}$$

$$|\frac{1}{M}\sum_{k=m}^{M+m-1} s_i(k)s_j(k-l) - pa^lV^2| \leq \frac{S}{\sqrt{M}} \ \forall m, l, M \tag{4.67}$$

The above inequalities embody all the assumptions on the subscriber signals we need to derive our results.

(4.66) and (4.67) indicate that the signal parameter, a, is a measure of the broadness of the autocorrelation and crosscorrelation functions of the subscriber signals, while the signal parameter, p, is a measure of the degree of cross correlation between the subscriber signals.

4.6.2 Analysis

Following the same procedure of Section 4.5.2, we define our domain of analysis as

$$D : |\tilde{\theta}_1(k)\tilde{\theta}_2(k)| < \Theta < 1$$

On this domain D and for sufficiently small μ, the dynamics of the DEC system, at least on a time scale $1/\mu$, are described by the averaged system equations:

$$\begin{array}{rcll} \tilde{\theta}_1^{av}(k+1) & = & \tilde{\theta}_1^{av}(k) - \mu A v_{\bar{y}1\bar{y}2(1)}(\bar{\theta}(k)) & \tilde{\theta}_1^{av}(0) = \theta_1 \\ \tilde{\theta}_2^{av}(k+1) & = & \tilde{\theta}_2^{av}(k) - \mu A v_{\bar{y}2\bar{y}1(1)}(\bar{\theta}(k)) & \tilde{\theta}_2^{av}(0) = \theta_2 \end{array} \tag{4.68}$$

where $\tilde{\theta}_k^{av} = (\tilde{\theta}_1^{av}(k), \tilde{\theta}_2^{av}(k))$ and

$$Av_{\bar{y}i\bar{y}j(l)}(z) = \lim_{M\to\infty} 1/M \sum_{k=m}^{M-1+m} \bar{y}_i(k,z)\bar{y}_j(k-l,z), \ \forall l \ \ i,j \in \{1,2\} \tag{4.69}$$

and $\bar{y}_i(k,z)$ is as defined by:

$$\begin{array}{rcll} \bar{y}_1(k+1,z) & = & s_1(k+1) + z_1\bar{y}_2(k,z) & \bar{y}_1(0) = y_1(0) \\ \bar{y}_2(k+1,z) & = & s_2(k+1) + z_2\bar{y}_1(k,z) & \bar{y}_2(0) = y_2(0) \end{array} \tag{4.70}$$

Remark:

26. The limit of (4.69) exists (on the domain $|z_1 z_2| < \Theta < 1$), is independent of m and, moreover, the limiting value is approached uniformly in m, by virtue of Assumptions 16-18.

The validity of Assumptions 16-18 leads to the averaging error being given by:

$$|\tilde{\theta}(k) - \bar{\theta}(k)| = 0(\delta(\mu)) = O(\sqrt{\mu}), \ \text{ on a time scale } 1/\mu \tag{4.71}$$

Notice that we also have the approximation (because $\bar{y}(k,z)$ is analytic in z in D):

$$|\bar{y}(k,\bar{\theta}(k)) - y(k)| = O(\sqrt{\mu}) \ \text{ on a time scale } 1/\mu \tag{4.72}$$

Furthermore [78], if the averaged system has an asymptotically stable equilibrium within the domain D then the approximations can be extended to hold on an infinite time scale:

$$\begin{aligned} |\tilde{\theta}(k) - \bar{\theta}(k)| &= O(\sqrt{\mu}) \quad \forall k \geq 0 \\ |\bar{y}(k, \bar{\theta}(k)) - y(k)| &= O(\sqrt{\mu}) \quad \forall k \geq 0 \end{aligned} \tag{4.73}$$

Through the application of the averaging approximation results above and analysis of the averaged system equations (4.68) we obtain the following result.

Theorem 9 *Consider the single tap, single delay DEC system described by (4.16)-(4.20) and Assumptions 1, 8,9,13-18. (See Appendix I for an easy reference to these assumptions.) Let* $\Theta \in (max(a^2, \theta_1^2, \theta_2^2), 1)$. *Then there exists a positive constant* $\mu*(\Theta, a, p)$ *such that for all positive* $\mu < \mu*$:

$$\begin{aligned} |\tilde{\theta}(k) - \tilde{\theta}^{av}(k)| &= O(\sqrt{\mu}) \quad \forall k \geq 0 \\ |\bar{y}(k, \tilde{\theta}^{av}(k)) - y(k)| &= O(\sqrt{\mu}) \quad \forall k \geq 0 \end{aligned} \tag{4.74}$$

Here $\tilde{\theta}^{av}(k)$ *is defined in (4.68) and* $\bar{y}(k, z)$ *is defined in (4.70) whilst* $\tilde{\theta}(k)$ *and* $y(k)$ *are the variables of the DEC system.*

Furthermore, if, in addition,

$$\mu < \frac{(1-\Theta)(1-a^2\Theta)}{V^2(1+a^2)} \tag{4.75}$$

then the averaged parameters $\tilde{\theta}_1^{av}(k)$, $\tilde{\theta}_2^{av}(k)$ *converge exponentially fast (with rate* λ*) to:*

$$\tilde{\theta}_1^{av}(k),\ \tilde{\theta}_2^{av}(k) \rightarrow \bar{\theta}_{12,s} \triangleq \frac{-(1+a^2) + \sqrt{(1+a^2)^2 - 4a^2p^2}}{2ap} as\ \lambda^k \rightarrow 0 \tag{4.76}$$

The convergence rate λ *(uniform estimate over the domain of attraction) can be estimated as:*

$$\lambda < 1 - \mu V^2 min\left(\frac{[(1+a^2) + \sqrt{(1+a^2)^2 - 4a^2p^2}]/2 - |ap|\sqrt{\Theta}}{(1+\Theta)(1+a^2\Theta)}, \frac{1-a^2}{(1+\Theta)(1+a^2\Theta)}\right)$$

Proof: See Appendix G.4.

Equations (4.74) and (4.76) indicate that the parameter estimate, $\hat{\theta}(k)$, converges exponentially fast (with rate λ) to an $O(\sqrt{\mu})$ neighbourhood of $(\theta_1 + \bar{\theta}_{12,s}, \theta_2 + \bar{\theta}_{12,s})$.

Remark:

27. When the subscriber signals are 'white', i.e. $a = 0$, and regardless of the cross correlation, even when the external subscriber signals are identical (perfectly correlated), the averaged DEC system shows optimal asymptotic behaviour in that the averaged parameter vector converges to the origin. Consequently, the original parameter error $(\tilde{\theta}_1(k), \tilde{\theta}_2(k))$ converges to a $\sqrt{\mu}$ small neighbourhood of the origin exponentially fast $(\bar{\theta}_s = (\bar{\theta}_{12,s}, \bar{\theta}_{12,s}) = (0,0))$. The transients are governed by a convergence rate overbounded by $\lambda < 1 - \mu V^2 \frac{1}{1+\Theta}$.

28. In the case in which the external subscriber signals are uncorrelated i.e. $p = 0$ the DEC system shows similar optimal asymptotic behaviour to that of 'white' subscriber signals with the parameter error converging to a $\sqrt{\mu}$ small neighbourhood of the origin exponentially fast. The transients are governed by a convergence rate overbounded by $\lambda < 1 - \mu V^2 \frac{1-a^2}{(1+\Theta)(1+a^2\Theta)}$. Notice that the convergence rate (or at least its bound) is adversely affected by the presence of the auto-correlation coefficient a.

29. In the case in which the subscriber signals are correlated and not 'white' i.e. $ap \neq 0$ the DEC system no longer can achieve full asymptotic echo suppression. The parameter error is biased. The bias grows with both the autocorrelation and crosscorrelation of the subscriber signals. This is the subject of Figure 4.4. At the same time the transients become longer.

30. Many simulations have been carried out on the single tap, single delay DEC system with correlated first order autoregressive subscriber signals. The results agree with Theorem 9 and the above discussion. Figures 4.5 and 4.6 provide an illustration by showing the time evolution of the residual echo parameters $(\tilde{\theta}_{1,k}, \tilde{\theta}_{2,k})$ and residual echoes $(\tilde{\theta}_{1,k}y_{2,k-1}, \tilde{\theta}_{2,k}y_{1,k-1})$, respectively, for the case in which the subscriber signals (4.63) have autocorrelation factor $a = 0.8$ and crosscorrelation factor $p = 0.665$. $\mu = 0.002$ and $V = 1$ for this simulation. As indicated in Figure 4.5, both residual echo parameters converge approximately to $\bar{\theta}_{12,s} = -0.368$, while Figure 4.6 indicates incomplete asymptotic echo suppression.

4.6.3 Subscriber Signals of Unequal Power

The previous section considered subscriber signals which were of equal power and described by the first order autoregressive (AR1) processes of (4.63). In this subsection we briefly explore the case in which subscriber signals are AR1 processes of unequal power. In particular, we consider when Assumptions 17 and 18 are replaced by:

Assumption 19

$$|\frac{1}{M}\sum_{k=m}^{M+m-1} w_i(k)w_i(k-l) - V_i^2\delta_{0,l}| \leq \frac{C}{\sqrt{M}} \; \forall m, M, l, \; \textit{where } V_i > 0$$

Assumption 20

$$|\frac{1}{M}\sum_{k=m}^{M+m-1} w_i(k)w_j(k-l) - pV_1V_2\delta_{0,l}| \leq \frac{C}{\sqrt{M}} \; \forall m, l, M \; \textit{where } |p| \leq 1$$

where $i, j = \{1, 2\}, i \neq j$.

Consequently, the subscriber signal inequalities of (4.66), (4.67) are replaced by:

$$|\frac{1}{M}\sum_{k=m}^{M+m-1} s_i(k)s_i(k-l) - a^l V_i^2| \leq \frac{S}{\sqrt{M}} \; \forall m, M, l \tag{4.77}$$

$$|\frac{1}{M}\sum_{k=m}^{M+m-1} s_i(k)s_j(k-l) - pa^l V_1V_2| \leq \frac{S}{\sqrt{M}} \; \forall m, l, M \tag{4.78}$$

The following Theorem quantifies the effect, on echo canceller performance, of subscriber signals having unequal power.

Theorem 10 *Consider the single tap, single delay DEC system described by (4.18)-(4.20), Assumptions 1 8,9,13-16,19,20 and*

$$\begin{aligned} \tilde{\theta}_1(k+1) &= \tilde{\theta}_1(k) - \frac{\mu}{V_2^2}y_1(k+1)y_2(k) \quad \tilde{\theta}_1(0) = \theta_1 \\ \tilde{\theta}_2(k+1) &= \tilde{\theta}_2(k) - \frac{\mu}{V_1^2}y_2(k+1)y_1(k) \quad \tilde{\theta}_2(0) = \theta_2 \end{aligned} \tag{4.79}$$

Let $\Theta \in (a^2, 1)$. *Suppose:*

$$\frac{V_2}{V_1}|\theta_1| + \frac{V_1}{V_2}|\theta_2| < 2\sqrt{\Theta} < 2 \tag{4.80}$$

Then there exists a positive constant $\mu*(\Theta, a, p)$ such that for all positive $\mu < \mu*$:

$$\begin{array}{rcll} |\tilde{\theta}(k) - \tilde{\theta}^{av}(k)| & = & O(\sqrt{\mu}) & \forall k \geq 0 \\ |\bar{y}(k, \tilde{\theta}^{av}(k)) - y(k)| & = & O(\sqrt{\mu}) & \forall k \geq 0 \end{array} \tag{4.81}$$

Here $\bar{y}(k,z)$ is defined in (4.70) and $\tilde{\theta}^{av}(k)$ is defined by:

$$\begin{array}{rcll} \tilde{\theta}_1^{av}(k+1) & = & \tilde{\theta}_1^{av}(k) - \dfrac{\mu}{V_2^2} Av_{\bar{y}1\bar{y}2(1)}(\tilde{\theta}^{av}(k)) & \tilde{\theta}_1^{av}(0) = \theta_1 \\ \tilde{\theta}_2^{av}(k+1) & = & \tilde{\theta}_2^{av}(k) - \dfrac{\mu}{V_1^2} Av_{\bar{y}2\bar{y}1(1)}(\tilde{\theta}^{av}(k)) & \tilde{\theta}_2^{av}(0) = \theta_2 \end{array} \tag{4.82}$$

whilst $\tilde{\theta}(k)$ and $y(k)$ are the variables of the DEC system.

Furthermore, if, in addition,

$$\mu < \frac{(1-\Theta)(1-a^2\Theta)}{1+a^2} \tag{4.83}$$

then the averaged parameters $\tilde{\theta}_1^{av}(k)$, $\tilde{\theta}_2^{av}(k)$ converge exponentially fast (with rate λ) to:

$$\begin{array}{rcl} \tilde{\theta}_{1,k}^{av} & \rightarrow & \bar{\theta}_{1,s} \triangleq \dfrac{V_2}{V_1} \dfrac{-(1+a^2) + \sqrt{(1+a^2)^2 - 4a^2p^2}}{2ap} \quad as\ \lambda^k \rightarrow 0 \\ \tilde{\theta}_{2,k}^{av} & \rightarrow & \bar{\theta}_{2,s} \triangleq \dfrac{V_1}{V_2} \dfrac{-(1+a^2) + \sqrt{(1+a^2)^2 - 4a^2p^2}}{2ap} \quad as\ \lambda^k \rightarrow 0 \end{array} \tag{4.84}$$

The convergence rate λ (uniform estimate over the domain of attraction) can be estimated as:

$$\lambda < 1 - \mu\ min\left(\frac{[(1+a^2) + \sqrt{(1+a^2)^2 - 4a^2p^2}]/2 - |ap|\sqrt{\Theta}}{(1+\Theta)(1+a^2\Theta)}, \frac{1-a^2}{(1+\Theta)(1+a^2\Theta)}\right)$$

Proof: The proof follows along the same lines as the proof of Theorem 9.

Remark:

31. The condition of (4.80) imposes a constraint on the ratio of the power of the subscriber signals. This condition guarantees that $(\tilde{\theta}_1^{av}(0), \tilde{\theta}_2^{av}(0)) = (\theta_1, \theta_2)$ lies in an invariant subset of the domain of attraction of the equilibrium $(\bar{\theta}_{1,s}, \bar{\theta}_{2,s})$ of (4.84). However, because (4.80) is only a sufficient condition, its violation does not necessarily mean $(\tilde{\theta}_1^{av}(k), \tilde{\theta}_2^{av}(k))$ will fail to converge to $(\bar{\theta}_{1,s}, \bar{\theta}_{2,s})$. On the

other hand, violation of (4.80) will increase the chances of unstable behaviour, such as bursting, in the original system. Consider, for example, the case:

$$V_2/V_1 \gg 1 \text{ and } a, p \neq 0, \text{ such that } \bar{\theta}_{1,s} = M \gg 1$$

The width of the stability domain

$$D = \{(\tilde{\theta}_1^{av}(k), \tilde{\theta}_2^{av}(k)) : |\tilde{\theta}_1^{av}(k)\tilde{\theta}_2^{av}(k)| < \Theta < 1\}$$

in the region of the equilibrium $(\bar{\theta}_{1,s}, \bar{\theta}_{2,s}) = (M, 1/M)$ is $O(1/M)$. Recall that the averaged system only approximates the original system within an error $O(\sqrt{\mu})$ Thus, if M is sufficiently large, stability problems become a real possibility.

32. As indicated by (4.79), the DEC system considered in Theorem 10 is different to that of Theorem 9, not only because the subscriber signals have unequal power, but also because the adaptation constant of each echo canceller has been normalized with respect to the power of the far end subscriber signal. This normalization was included to enable quantitative estimation of the domain of attraction of the equilibrium of (4.84).

    It is important to note, however, that this normalization does not affect the location of the equilibrium of the averaged system or, equivalently, the location of the asymptotic $O(\sqrt{\mu})$ ball to which the original system parameter vector $\tilde{\theta}(k)$ converges. This is because the equation:

$$Av_{\bar{y}1\bar{y}2(1)}(\tilde{\theta}^{av}(k)) = Av_{\bar{y}2\bar{y}1(1)}(\tilde{\theta}^{av}(k)) = 0$$

    the solution of which identifies the equilibria of the averaged system (4.82), is independent of any normalization of μ. In particular, assume μ is sufficiently small and $\tilde{\theta}^{av}(0) = (\theta_1, \theta_2)$ lies sufficiently close to $\bar{\theta}_s = (\bar{\theta}_{1,s}, \bar{\theta}_{2,s})$ so that $\tilde{\theta}^{av}(0)$ is within the domain of attraction of $\bar{\theta}_s$. Then the averaged parameter vector $\tilde{\theta}^{av}(k)$ of the DEC system with Assumptions 1,8,9,13-16,19,20 and in which μ is not normalized, will also converge to the equilibrium $\bar{\theta}_s$ of Theorem 10.

The result of Theorem 10 indicates that when the subscriber signals are of unequal power, the adverse effect of correlation within and between the subscriber signals on asymptotic echo canceller performance will be accentuated at one end of the DEC system and diminished at the other end. In particular, in order to achieve sufficiently good asymptotic performance of both echo cancellers within the DEC system, the

autocorrelation and cross correlation of the subscriber signals should be reduced to smaller levels as the ratio of the power of the subscriber signals increases(decreases) above(below) unity.

Summary

The rigorous analysis conducted in this section on the single tap, single delay DEC system with first order AR subscriber signals indicates the following.

(i) Good asymptotic performance is achieved when the subscriber signals show zero cross correlation and/or are 'white'.

(ii) The convergence rate is maximum when the subscriber signals are 'white'.

(iii) When the subscriber signals are not 'white', the transient and asymptotic performance deteriorates with increasing autocorrelation and/or cross correlation levels.

(iv) The adverse signal effects on asymptotic performance are accentuated when the subscriber signals are of unequal power.

(v) As the power ratio of the subscriber signals increases (decrease) above (below) unity, the size of μ needs to be reduced to ensure stability.

4.7 Conclusion

In this chapter we analysed the dynamics of a closed loop system representative of the double echo canceller (DEC) system in which an echo path and a neighbouring LMS adaptive FIR echo canceller are located at each end of the loop. The system is driven by subscriber signals entering either end of the loop. A number of analyses were conducted ranging from semi-formal to rigorous depending on the assumptions made. In particular, because of the nonlinear dynamics of the DEC system, a number of simplifying assumptions were needed to enable quantitative/rigorous analysis to be carried out. The analyses indicated the following.

(i) When the subscriber signals are cross correlated, the adaptive filtering is biased. This bias grows and, subsequently, the asymptotic performance deteriorates with greater cross correlation levels and broader autocorrelation functions of the subscriber signals. Zero bias and (apart from estimation noise, which is of $O(\mu)$) complete asymptotic echo suppression occurs when the subscriber signals show zero cross correlation.

(ii) The possibility of nonconvergent/unstable behaviour increases with increasing cross correlation and/or autocorrelation levels of the subscriber signals.

(iii) The transient performance deteriorates with greater cross correlation levels and

broader autocorrelation functions of the subscriber signals. The convergence rate is maximum when the subscriber signals are non cross correlated and are 'white', that is, each is uncorrelated over time.

(iv) Importantly, when both subscriber signals are 'white' the cross correlation function of the subscriber signals will be, in general, zero for all sample lags. [The exception is the unlikely case of the subscriber signals being replicas or time shifted versions of each other, in which case the cross correlation function is nonzero for the lag corresponding to the time shift.] Consequently, 'white' subscriber signals generally lead to optimal asymptotic and transient performance.

(v) The adverse effects of broad autocorrelation function and high cross correlation levels of the subscriber signals are accentuated by an increase in dimension of either echo canceller.

(vi) The adverse subscriber signal effects reduce as the delay, imposed by the loop channels, increases. In particular, if the delay of the loop channels is sufficiently long and the subscriber signals show sufficiently narrow autocorrelation and cross correlation functions, then the dynamics of the closed loop DEC system simplify to the dynamics of a pair of decoupled open loop LMS adaptive FIR systems. In this case, unbiased asymptotic echo suppression is achieved. Furthermore, the transient performance can be enhanced/optimised by whitening the input signal to each echo canceller.

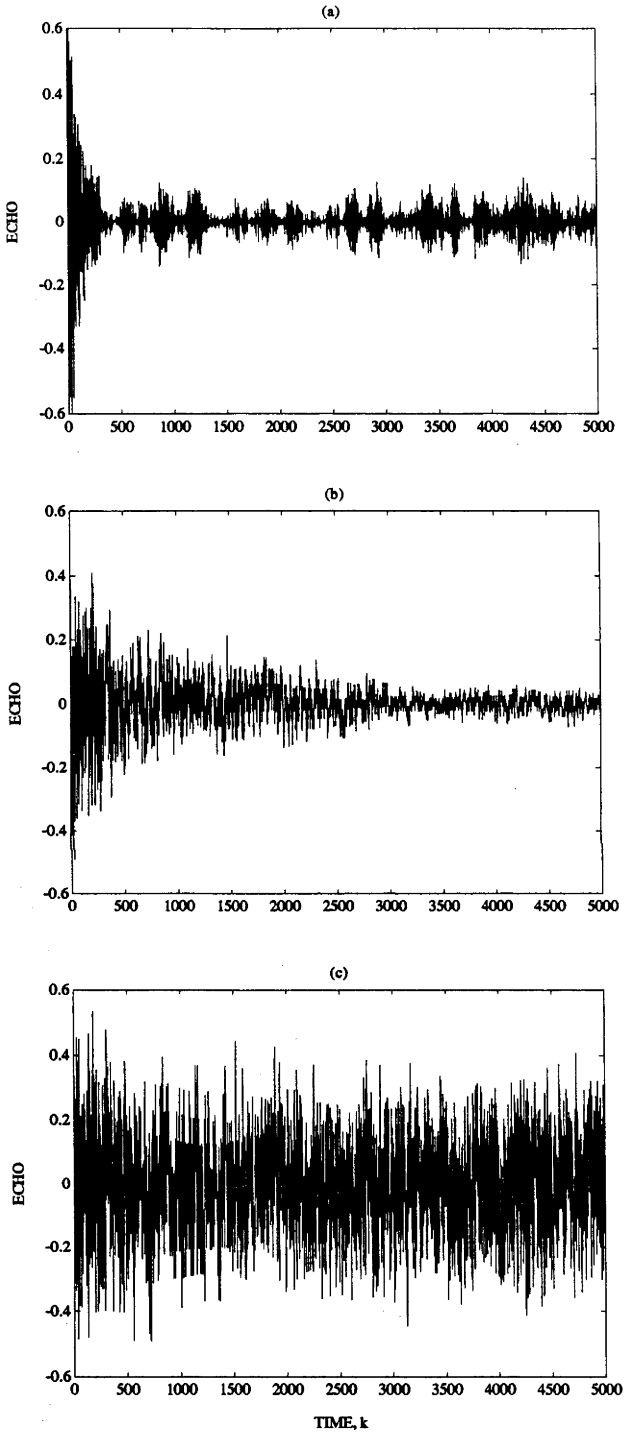


Figure 4.3: Plot of residual echoes over time with $\tilde{\theta}_1(0) = 0.8$, $\tilde{\theta}_2(0) = -0.7$, $\mu = 0.002$ and subscriber signals (a) of Table 4.1(a), (b) of Table 4.1(b), (c) of Table 4.1(c).

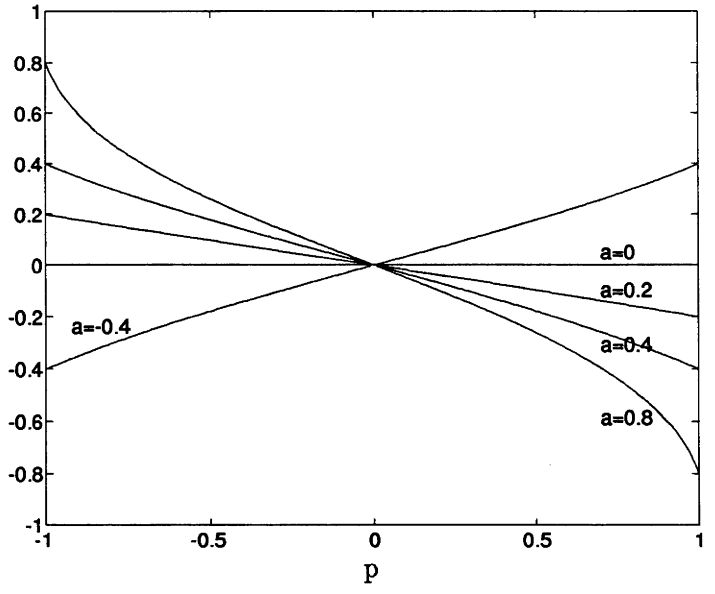


Figure 4.4: Plot of $\bar{\theta}_{12,s}$ as given in (4.76) for correlated subscriber signals (4.63) with cross correlation factor p and autocorrelation factor a.

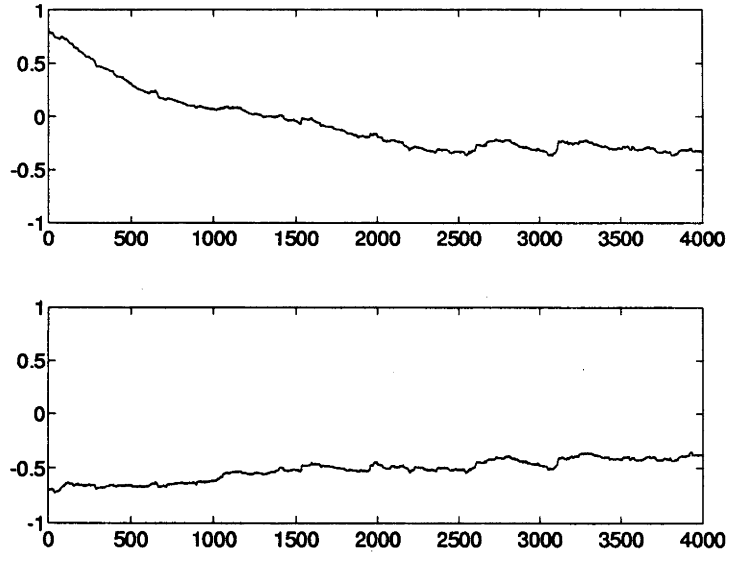


Figure 4.5: Plot of $\tilde{\theta}_{1,k}$, $\tilde{\theta}_{2,k}$ over time with $\tilde{\theta}_{1,0} = 0.8$, $\tilde{\theta}_{2,0} = -0.7$, $\mu = 0.002$ and signal parameters $a = 0.8$, $p = 0.665$.

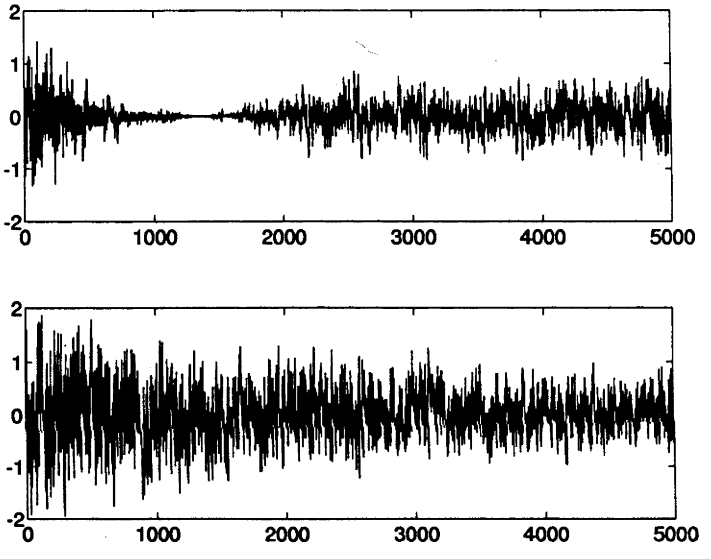


Figure 4.6: Plot of residual echoes over time with $\tilde{\theta}_{1,0} = 0.8$, $\tilde{\theta}_{2,0} = -0.7$, $\mu = 0.002$ and signal parameters $a = 0.8$, $p = 0.665$.

Chapter 5

Signal Conditioning for Echo Cancellers

5.1 Introduction

The analyses carried out in Chapters 3 and 4 highlighted the dependence of the performance of the LMS adaptive FIR echo canceller on the correlation characteristics of the input signal to the echo canceller or of the subscriber signals of the telecommunication network. In particular, the analytical results imply that for a given echo canceller dimension, the transient and asymptotic performance of the double echo canceller (DEC) system can be improved by whitening the subscriber signals. The results also suggest that, assuming the delay imposed by the transmission channels of the DEC system is sufficiently long, transient performance improves by whitening the input signal of each echo canceller. For sufficiently long transmission channel delays, unbiased asymptotic echo suppression is obtained, particularly for speech transmission echo cancellation systems, in which the subscriber signals are typically not cross correlated.

In this chapter we present two performance enhancing schemes based on these signal conditioning ideas. The first scheme, examined in Section 2, focusses on subscriber signal whitening. It is only applicable to 4-wire loop circuit echo cancellation, since access to the subscriber signals is required. The scheme involves filtering the subscriber signals with digital scramblers. The pseudo-random property of digital scramblers leads to the filtered subscriber signals being essentially 'white'. Although, generally, this is sufficient for good asymptotic performance, a simple extension of the scheme also is presented which ensures zero cross correlation between the scram-

bled subscriber signals even in the event that the subscriber signals are replicas or time shifted versions of each other. Simulations demonstrate the ability of this signal conditioning scheme to provide enhanced echo cancellation.

The second scheme, which is presented in Section 3, is based on the input signal whitening approach. This scheme may be used for enhancing either acoustic echo cancellation or 4-wire loop circuit echo cancellation. Its success relies on the assumption that the input signal to each echo canceller within the DEC system is well modelled as an autoregressive (AR) process. Such AR modelling is typically employed for speech. The scheme involves applying standard linear prediction techniques to obtain an estimate of the autoregressive filter and using this filter to whiten the input. In acoustic echo cancellation, the input signal is that which is received by the audience in the acoustic enclosure. Consequently, indirect whitening methods, using the autoregressive filter estimate, are required. Several indirect whitening methods have been proposed previously. We review and suggest extensions to these methods.

A couple of points concerning the two signal conditioning schemes need to be highlighted.

(i) The nature of the scrambler whitening scheme restricts its applicability. Firstly, the transmission efficiency of digital signals, such as the scrambled subscriber signals, over analogue channels is typically less than that over digital channels. For example, the bit rate achievable for digital/PCM encoded voice frequency channels is about 20kb/s [84], far less than the desired rate of 64kb/s - 8 bit words, 8kHz sampling rate - which is achievable with digital channels. Secondly, signal compression techniques which are used for reducing the bit rate over digital systems rely on the signals having redundant 'information', or being autocorrelated. These comments indicate that the scrambling scheme should be restricted to digital networks (including digital local 2-wire subscriber lines) which do not require signal compression. Fibre optic networks are suitable.

(ii) The AR whitening scheme involves two estimators, the echo canceller and the linear predictor, rather than just the one estimator, the echo canceller, of the standard approach. In general, such a scheme may not provide transient performance improvements since it involves a greater overall filter parameter dimension, $n + p$ where p and n are the dimension of the AR filter and echo canceller, respectively. Recall, however, that the adverse effect of dimension on transient performance grows with input signal autocorrelation. This was shown for the LMS estimator, but it may also apply to the linear predictor. With this in mind, performance improvements should

arise with the the AR scheme because the highly autocorrelated speech based signal is feeding only the linear predictor, the parameter dimension of which is typically p =10 to 20. This is much less than the parameter dimension of the typical echo canceller, $n = 100 \rightarrow 4000$, which, in the standard approach, is that being fed by the speech based signal.

5.2 Signal Conditioning with Digital Scramblers

We begin by introducing the scrambler and two common ways in which such devices can be used to whiten signals as well as decorrelate two identical signals. A discussion then follows on the use of these scrambling techniques for improving echo canceller performance in 4-wire loop telephony.

5.2.1 Scramblers

A scrambler, as shown in Figure 5.1, is basically a binary feedback shift register. Clearly, this requires the summation operator $\oplus$ to be modulo-2 and the tap delay coefficients, $h_1, h_2, ..., h_m$, to be binary. The tap coefficients are chosen such that the binary polynomial

$$h(z) = 1 + h_1 z + h_2 z^2 + \ ... \ + h_m z^m$$

is primitive, that is, has no binary factors other than itself and unity. (Tables of primitive binary polynomials are given by various authors, e.g. [83], [84].) This leads to the binary sequence $\{x_k\}$ output by the scrambler showing zero autocorrelation or randomlike properties over a sampling length $P_x = 2^m - 1$. This length P_x of zero autocorrelation is, in fact, the maximum possible for any m-tap binary feedback shift register. Furthermore, the sequences output by two scramblers having different tap coefficient vectors will be orthogonal, that is, show zero cross correlation.

These pseudorandom and orthogonality properties of scrambler sequences are illustrated in Figure 5.2. Figure 5.2a shows the autocorrelation function of a sequence output by a 7-tap scrambler, (the sequence is binary antipodal, $\{-1, 1\}$) while Figure 5.2b shows the cross correlation function of antipodal sequences output by a 5-tap and a 7-tap scrambler.

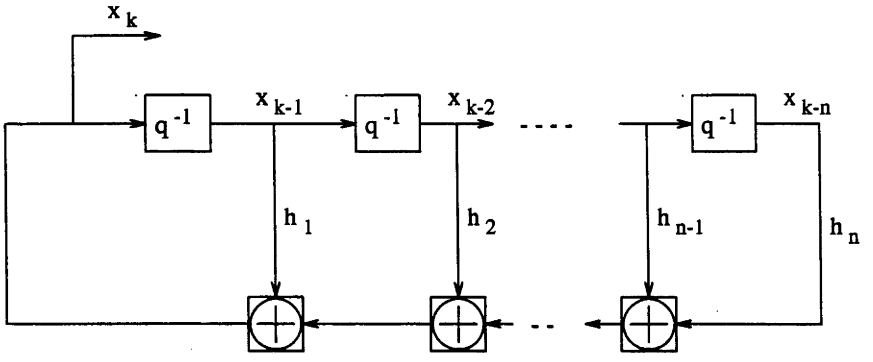


Figure 5.1: Scrambler (isolated) - modulo 2 feedback shift register, with suitably chosen binary tap coefficients, h_i

5.2.2 Signal Scrambling

A scrambler is commonly employed to interact with a given binary signal, such as a binary subscriber signal, to produce a scrambled (randomlike) signal. Scrambling is an attractive process for whitening signals because it is deterministic and, consequently, through the use of a suitable descrambling device, the original signal can be completely recovered. The scrambling-descrambling process can be carried out by either of two methods:

1. Frame Synchronized Scrambling/Descrambling
2. Self Synchronized Scrambling/Descrambling

Frame Synchronized Scrambling/Descrambling

Frame synchronized scrambling of an input sequence $\{b_k\}$, shown in Figure 5.3a, involves modulo-2-summing the output sample, x_k, of the scrambler with the input sample, b_k to produce $c_k = x_k \oplus b_k$. The output or scrambled sequence, $\{c_k\}$, is pseudorandom with a period, P_c, equal to the lowest common multiple (LCM) of P_x, the period of the scrambler sequence and, P_b, the period of the input sequence [84].

The descrambling process as shown in Figure 5.3b, used to recover the original sequence $\{b_k\}$, involves modulo-2 summing the scrambled sequence $\{c_k\}$ with the output of a scrambler of the same form as the first scrambler. The correct operation of the frame synchronized scrambler depends on the alignment in time of the states of the scrambler and descrambler. This is achieved by a mechanism known as frame synchronization. See [84] for a discussion on this topic.

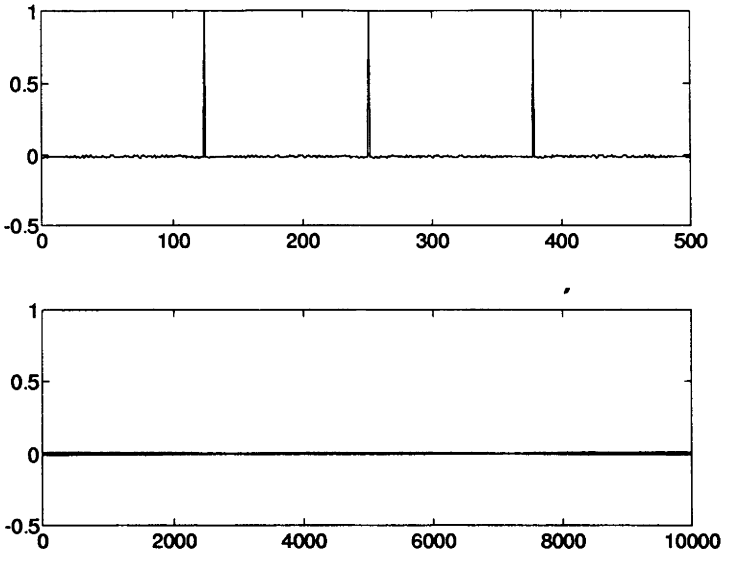


Figure 5.2: (a) Autocorrelation function of an antipodal sequence output by a 7-tap scrambler (isolated), (b) Cross correlation function of antipodal sequences output by a 5-tap and a 7-tap scrambler (isolated)

Self Synchronized Scrambling/Descrambling

Self synchronized scrambling of an input sequence $\{b_k\}$, shown in Figure 5.4a, differs from frame synchronized scrambling essentially by having the scrambled sequence, $\{c_k\}$, fed back into the scrambler. This results in the state of the scrambler depending not only on its initial state (as in frame synchronized scrambling), but also on all of the past input samples, b_0, b_1, ..., b_k. The scrambled samples, c_k, are given by:

$$c_k = b_k \oplus h_1 c_{k-1} \oplus ... \oplus h_m c_{k-m} \tag{5.1}$$

The period, P_c, of the self synchronized scrambled sequence, like that of frame synchronized scrambled sequences, is given by $P_c = LCM(P_x, P_b)$, where $P_x = 2^m - 1$ and P_b is the period of the input sequence. However, there is one exception to this rule. One of the 2^m possible states of the m-tap scrambler, the determination of which is dependent on the input sequence, will lead to the output sequences having a period, $P_x = P_b$, the period of the input sequence, which is a serious limitation if P_b is small. To illustrate this characteristic of self synchronized scramblers, consider the following example, given in [84]. Suppose we have a self synchronized scrambler with a tap coefficient vector $h = [1, 1]$. That is, the output sample is given by $c_k = b_k \oplus c_{k-1} \oplus c_{k-2}$. Now, assume that at some initial time i the input sequence is

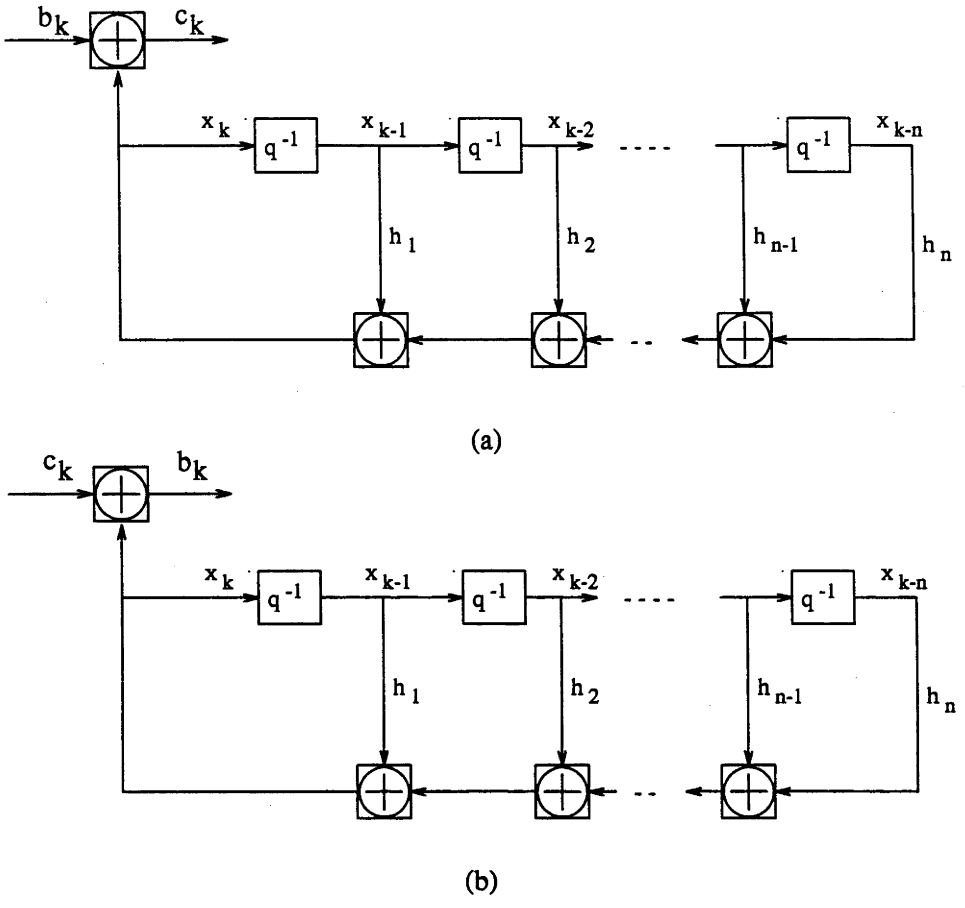
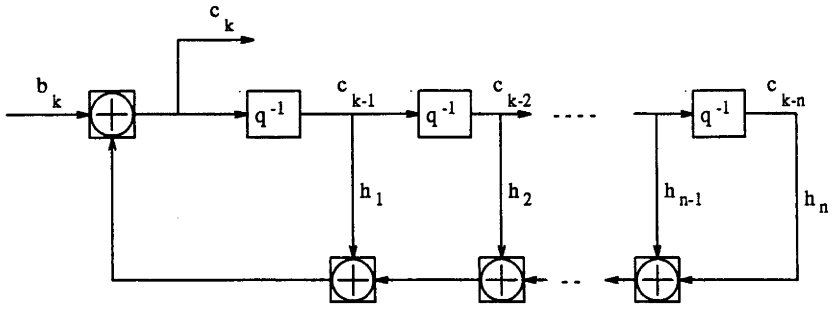


Figure 5.3: (a) Frame synchronized scrambler with input sequence $\{b_k\}$, (b) Frame synchronized descrambler

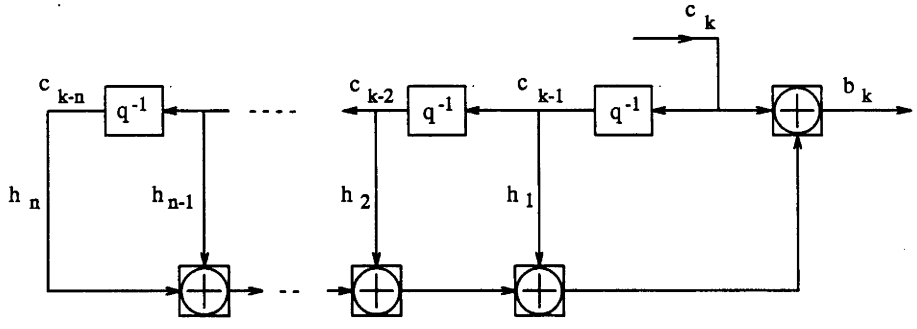
alternating zeroes and ones, period two: $(b_i,\ b_{i+1},\ b_{i+2},\ b_{i+3},\ ...) = (0,\ 1,\ 0,\ 1,\ ...)$. If the state of the scrambler at time i happens to be $(c_{i-1}, c_{i-2}) = (0, 1)$ then the output sequence is also alternating zeroes and ones, period two. However, if at time i the scrambler is in any of the other three $(= 2^2 - 1)$ possible states then the output sequence will have period 2x3=6.

Clearly, the probability, 2^{-m}, of a self synchronized scrambler being in the undesirable state for a given input sequence becomes negligible for sufficiently large m.

Descrambling, as shown in Figure 5.4b, to regain the original sequence, $\{b_k\}$, involves inputting the scrambled sequence, $\{c_k\}$, into a 'reversely' (or inversely) structured device to that of the scrambler. As the name suggests, self synchronized scrambling has the advantage, over frame synchronized scrambling, of not requiring synchronization between scrambler and descrambler [84]. Of course, there is still the need for clock synchronization between the scrambling and descrambling devices.



(a)



(b)

Figure 5.4: (a) Self synchronized scrambler with input sequence $\{b_k\}$, (b) Self synchronized descrambler

Self synchronized scrambling has an important disadvantage. It is unprotected against error propagation [84]. The sensitivity to error propagation increases with the number of nonzero taps within the scrambler. In particular, error multiplication is by a factor equal to the number of nonzero taps plus one [84]. For this reason, self synchronized scramblers used in practice involve a minimum number of nonzero taps. For many scrambler tap lengths, this minimum number is three.

Scrambled Signal Characteristics

Consider a sequence which is a constant sequence of zeroes or ones. This sequence is fully autocorrelated - the correlation length is equal to the sequence length. The autocorrelation function of the output of an m-tap length self/frame synchronized scrambler fed by such a sequence is just that of the isolated scrambler, a series of impulses at lags $l = (2^m - 1)^i$, $i = 0, 1, 2, \ldots$ - see, for example, Figure 5.2. Intuitively, as the correlation length of the input decreases, the autocorrelation function impulses

at the large lag values should disappear. Furthermore, when the autocorrelation length of the input drops below $P_x = 2^m - 1$, all but the zero lag impulse should disappear. In particular, it is expected that, if the autocorrelation length of the input sequence is significantly less than P_x then the self/frame synchronized scrambled sequence should be essentially discrete white. This intuition is supported by a large number of simulations.

As suggested by the orthogonalizing properties of isolated scramblers and as indicated by many simulations, the scrambled sequences output by two scramblers (either frame synchronized or self synchronized) having different 'primitive' tap coefficient vectors are approximately orthogonal or show essentially zero cross correlation, irrespective of the input signal characteristics. In the case of self synchronized scramblers (because the state of a self synchronized scrambler is influenced by all of the past input samples as well as by its initial state), approximately orthogonal sequences can be obtained from the same scrambler by simply changing the initial state vector.

Figures 5.5 and 5.6 illustrate the whitening and orthogonalizing capabilities of self synchronized scramblers. Figure 5.5a shows the autocorrelation function of a signal $u(k)$ described by the first order autoregressive (AR1) model of (3.36) with a=0.9. Prior to scrambling, this signal was 1-bit quantized:

$$u(k) < 0 \rightarrow u(k) = 0, \quad u(k) \geq 0 \rightarrow u(k) = 1.$$

Importantly, the autocorrelation function of the antipodal version of this quantized signal is the same as that shown in Figure 5.5a. The autocorrelation function of the antipodal signal after being self synchronized scrambled, with a 7-tap scrambler, is shown in Figure 5.5b. Scrambling has transformed a signal having a relatively broad autocorrelation function into a signal having an autocorrelation function resembling that of a zero mean discrete white signal. Frame synchronized scramblers show similar whitening capabilities.

Figure 5.6a shows the cross correlation function of two equal AR1 signals (with $a = 0.7$). The cross correlation function of the signals, after one of the AR1 signals was self synchronized scrambled by a 5-tap scrambler and the other by a 7-tap scrambler, is shown in Figure 5.6b. [Note: the signals were 1-bit quantized prior to scrambling.] Scrambling has reduced the cross correlation levels approximately to zero. Similar orthogonalizing characteristics are exhibited by pairs of frame synchronized scramblers having different 'primitive' tap coefficient vectors and by pairs of self synchronized scramblers which differ only by their initial state vectors. Of course,

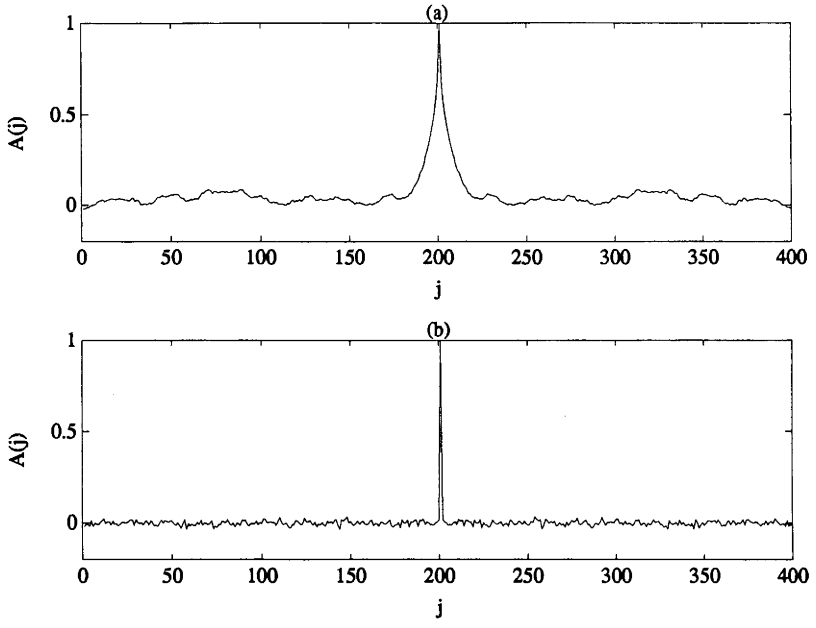


Figure 5.5: (a) Autocorrelation function of a first order autoregressive signal of (3.36) with a=0.9, (b) Autocorrelation function of signal in (a) after being self synchronized scrambled with a 7-tap scrambler

input signals for each tap coefficient vector/initial state vector configuration can be concocted to show that there are exceptions, but the probability of the occurrence of such signals is likely to be small.

5.2.3 Scrambler Scheme for Echo Cancellation

The scrambler schemes examined in the previous subsection provide a means of obtaining a 'white' or decorrelated sequence of bits. Since each sample of a digital signal is a linear combination of a block of bits and the blocks corresponding to each sample do not overlap, then a white sequence of bits implies a white digital signal. This then suggests that a scheme based on the use of scramblers to scramble the bits of the digital subscriber signals in a DEC 4-wire loop system should provide enhanced/optimal echo canceller performance. Such a scheme is shown in Figure 5.7a. It involves placing frame synchronized or self synchronized scramblers having the same configuration (tap coefficient/initial state vectors) with the subscriber digital transmitters and corresponding descramblers with the digital receivers. Use of such a scheme, in general, should lead to the digital signals entering the 4-wire loop showing approximately delta shaped or 'white' autocorrelation functions. In general, the signals will also show zero cross correlation.

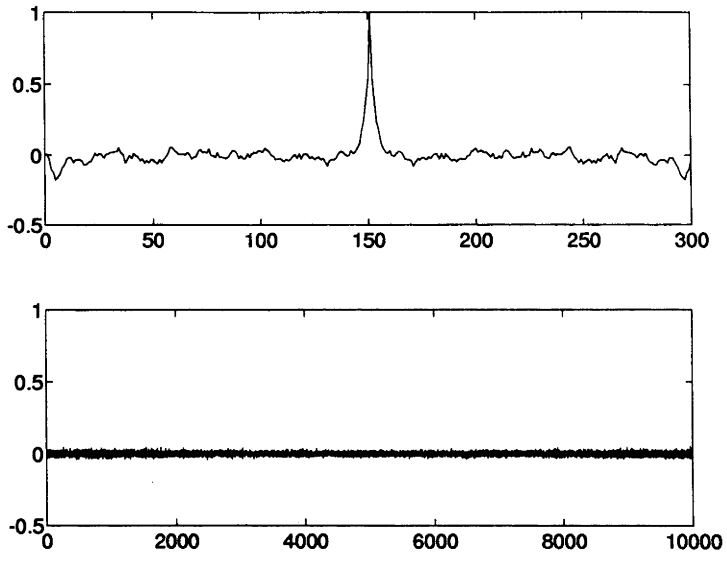
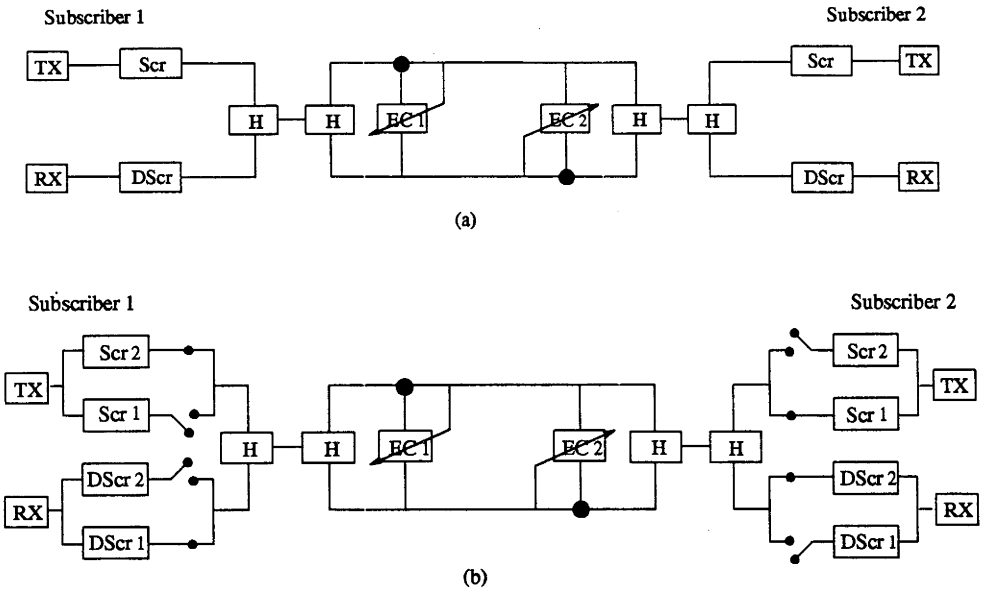


Figure 5.6: (a) Cross correlation function of two equal first order autoregressive signals of (3.36) with a=0.7, (b) Cross correlation function of the signals in (a) after one was self synchronized scrambled by a 5-tap scrambler and the other by a 7-tap scrambler

An exception to the occurrence of 'white' scrambled subscriber signals is that in which the autocorrelation length of the subscriber signal approaches or extends past $P_x = 2^m - 1$, where m is the tap length of the scrambler. However, the likelihood of this occurrence is negligible when scramblers, having sufficiently long tap coefficient vectors e.g. $m = 23$, are selected. In the case of self synchronized scrambling, the tap coefficient vectors should each have a minimal number of nonzero taps so as to minimize the error multiplication factor (discussed in section 5.2.2.).

An extension of the basic scheme involves providing each subscriber with an orthogonal or differently configured scrambler. Such a scheme ensures zero cross correlation between the scrambled subscriber signals even in the event that the subscriber signals are identical or time shifted versions of each other. A possible problem might be foreseen in providing each subscriber with a differently configured scrambler. This may be a serious limitation with frame synchronized scramblers. However, it should not be a significant problem with self synchronized scramblers for which a different configuration can be obtained by not only changing the tap coefficient vector, but also by simply changing the initial state.

An alternative approach to that of providing each subscriber with a differently configured scrambler is that in which each subscriber is provided with the same pair of



Scr=Scrambler, DScr=Descrambler, H=Hybrid, EC=Echo Canceller

Figure 5.7: Scrambler line coding schemes proposed for improving echo cancellation in DEC 4-wire loop networks, (a) Basic scheme - each subscriber uses the same scrambler configuration, (b) Paired scheme - all subscribers have the same pair of differently configured scramblers

differently configured scramblers/descramblers - as illustrated in Figure 5.7b. This approach relies on the assumption that each subscriber link-up involves no more than two subscribers (although it could easily be extended to allow for more subscribers in a link-up.) Zero cross correlation of the scrambled subscriber signals can be obtained by ensuring that both subscribers do not select the same configuration. The selection of different configurations would be easily carried out during the establishment of the subscriber link-up.

5.2.4 Simulations

We conclude this section by comparing the results of simulations of LMS/FIR echo canceller performance in DEC systems with and without the scrambler scheme employed. The simulation conditions were the same as those described in Chapter 4 Section 4.5 and Table 4.1. In all of the simulations the subscriber signals were converted into 1-bit quantized antipodal sequences. In each of the simulations involving scrambling, the following self synchronized scramblers were used:

Scrambler 1	Tap coefficient vector:	[01001]	Initial state vector:	[00001]
Scrambler 2		[0010001]		[0000001]

The simulation results are given in Figure 5.8, which shows plots of the variation in residual echo level over time at one end of the DEC loop. The residual echo level over time at the other end of the DEC loop was observed, in all cases, to be similar to that shown. In simulation (a) the subscriber signals are uncorrelated AR(0) or discrete white signals. As expected, good performance is achieved with or without the scrambling scheme employed. In both simulations (b) and (c), each subscriber signal has a relatively broad autocorrelation function. In simulation (c) the subscriber signals also show a relatively broad cross correlation function. In both of these latter simulations, the use of the scrambler scheme has led to a greatly improved (transient and asymptotic) performance.

5.3 Signal Conditioning via Autoregressive Filtering

In this section we examine the second signal conditioning method, the objective of which is to enhance transient performance by whitening the input signal to each echo canceller. A basic assumption of the method investigated is that the input signal, like speech, is well modelled as an autoregressive (AR) process:

$$u(k) = \frac{w(k)}{A(q^{-1})} \tag{5.2}$$

where $w(k)$ is a zero mean discrete white signal and $A(q^{-1})$ is an FIR filter of tap length p =10 to 50. Under this assumption, approximate whitening of the input signal should be achieved simply by filtering the input signal with an estimate $\hat{A}(q^{-1})$ of the filter $A(q^{-1})$. In acoustic echo cancellation, however, such a scheme is not permitted since the whitened input signal is that which is received by the audience in the acoustic enclosure. In this section we examine and suggest extensions to a number of proposed approaches which address this problem.

In this section we do not consider AR filter estimation. There are standard techniques/algorithms for such estimation. One popular technique is presented in Appendix D.

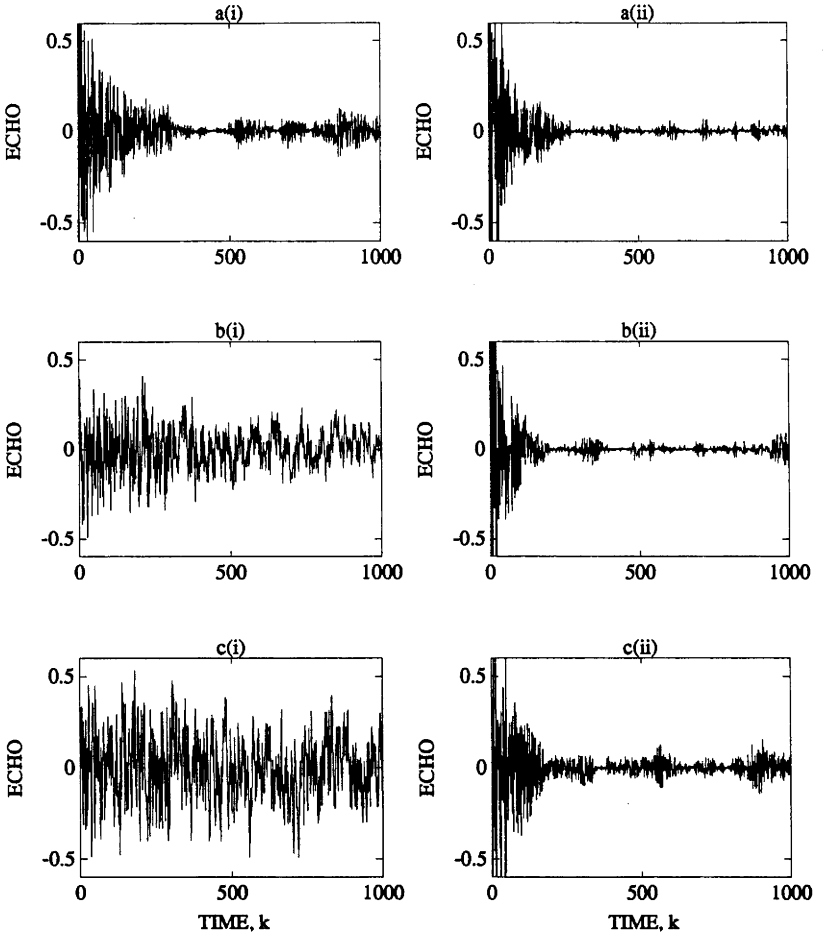


Figure 5.8: Plot of residual echo over time (i) without and (ii) with the scrambling scheme of Figure 5.7b employed for $\tilde{\theta}_1(0) = 0.8$, $\tilde{\theta}_2(0) = -0.7$, $\mu = 0.002$ and subscriber signals (a) of Table 4.1(a), (b) of Table 4.1(b), (c) of Table 4.1(c)

5.3.1 AR Based Whitening Schemes

Consider the standard LMS adaptive open loop system of Figure 5.9. Again, we assume that the echo path and echo canceller are n-tap FIR filters parametrized by the filter vectors θ and $\hat{\theta}(k)$, respectively. We assume that the input $u(k)$ to the echo path/canceller is well modelled by a p^{th} order autoregressive (AR) process, which is possibly non-stationary:

$$\begin{aligned} u(k) &= w(k) - [a_1(k)u(k-1) + a_2(k)u(k-2) + ... + a_p(k)u(k-p)] \\ \text{or } w(k) &= A(k)^T U_a(k) \\ \text{where } A(k) &= (1\ a_1(k)\ a_2(k)\ ...\ a_p(k))^T \end{aligned}$$

$$U_a(k) = (u(k)\ u(k-1)\ u(k-2)\ ...\ u(k-p))^T$$

and $w(k)$ is a zero mean 'white' signal of variance $\sigma_w^2(k)$. Let

$$\hat{A}(k) = (1\ \hat{a}_1(k)\ \hat{a}_2(k)\ ...\ \hat{a}_p(k))^T$$

be an estimate of the AR filter vector $A(k)$. Furthermore, we define:

$$\begin{aligned}
\hat{w}(k) &= \hat{A}(k)^T U_a(k) \\
\hat{W}(k) &= (\hat{w}(k)\ \hat{w}(k-1)\ \hat{w}(k-2)\ ...\ \hat{w}(k-n+1))^T \\
S_a(k) &= (s(k)\ s(k-1)\ s(k-2)\ ...\ s(k-p))^T
\end{aligned}$$

When $\hat{A}(k) = A(k)$ then

$$E[\hat{W}(k)\hat{W}(k)^T] = I\sigma_w^2(k)$$

where I is the $n \times n$ identity matrix.

Using the above notation and assumptions, we now examine a number of AR filtering schemes which have been proposed, particularly for acoustic echo cancellation.

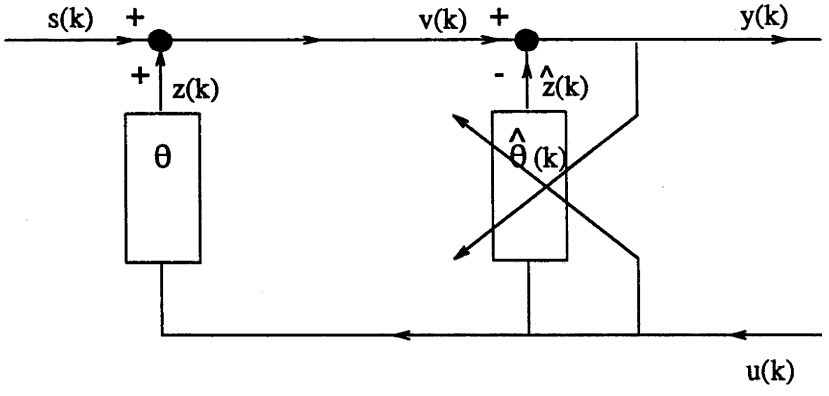


Figure 5.9: The standard LMS adaptive open loop system

Scheme A

The first scheme, which we call **Scheme A**, and which was proposed in [37] is illustrated in Figure 5.10. This scheme involves filtering/whitening of the input to the echo canceller via the estimate $\hat{A}(q^{-1})$. Refiltering of this filtered signal with $1/\hat{A}(q^{-1})$ just prior to the echo path ensures that the original input signal is received by the near end subscriber/acoustic enclosure. Filtering with $\hat{A}(q^{-1})$ is again employed after

the echo path in an attempt to decorrelate the input signal within the echo. LMS adaptation of the echo canceller uses this filtered echo+disturbance signal as well as the filtered input signal $\hat{A}(q^{-1})u(k)$. The adaptation is summarized by the following set of equations.

$$\begin{aligned}
\hat{\theta}(k+1) &= \hat{\theta}(k) + \mu\hat{W}(k)x(k) && (5.3)\\
x(k) &= \hat{A}(k)^T S_a(k) \\
&\quad +\hat{A}(k)^T[\theta^T U(k)\ \theta^T U(k-1)\ ...\ \theta^T U(k-p)]^T \\
&\quad -\hat{\theta}(k)^T[\hat{A}(k)^T U_a(k)\ \hat{A}(k-1)^T U_a(k-1)\ ...\ \hat{A}(k-n+1)^T U_a(k-n+1)]^T \\
&= \hat{A}(k)^T S_a(k) && (5.4)\\
&\quad +\theta^T[\hat{A}(k)^T U_a(k)\ \hat{A}(k)^T U_a(k-1)\ ...\ \hat{A}(k)^T U_a(k-n+1)]^T \\
&\quad -\hat{\theta}(k)^T\hat{W}(k)
\end{aligned}$$

where $U(k) = (u(k)\ u(k-1)\ u(k-2)\ ...u(k-n+1))^T$.

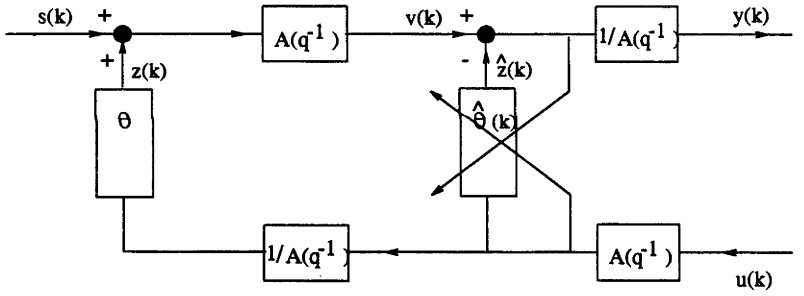


Figure 5.10: The LMS adaptive open loop system with AR whitening Scheme A

When the autoregressive estimate is time invariant $\hat{A}(k) = \hat{A}$ (which requires the input to be stationary), then $x(k)$ of (5.4) can be rewritten as

$$\begin{aligned}
x(k) &= \hat{A}(k)^T S_a(k) + (\theta - \hat{\theta}(k))^T[\hat{A}^T(U_a(k)\ U_a(k-1)\ ...\ U_a(k-n+1)]^T && (5.5)\\
&= \hat{A}(k)^T S_a(k) + (\theta - \hat{\theta}(k))^T\hat{W}(k) && (5.6)
\end{aligned}$$

and the update equation for the echo canceller becomes

$$\tilde{\theta}(k+1) = (\theta - \hat{\theta}(k+1)) = [I - \mu\hat{W}(k)\hat{W}(k)^T]\tilde{\theta}(k) + \mu\hat{A}(k)^T S_a(k)\hat{W}(k) \qquad (5.7)$$

Assuming $\hat{A}$ is a good estimate of the time invariant AR filter vector A then

$$E[\hat{W}(k)\hat{W}(k)^T] \approx I\sigma_w^2$$

Consequently, the echo canceller converges at a rate similar to that when the input signal to the echo canceller/path is white.

Speech, however, is nonstationary and, therefore, the autoregressive estimate is time varying. In this case $x(k)$ does not, in general, take the form of (5.5),(5.6). In particular, if the nonstationarity of the input extends only over $M < n$ sample intervals then

$$\hat{A}(k)^T U_a(k-i) = \hat{w}(k-i)$$

only holds for $0 \leq i \leq M$. One approach proposed by [37] and [87], involves retaining the same AR estimate for n sample intervals, where n is the tap length of the echo canceller. We will call this approach in combination with (5.3) **Scheme A1.** Under Scheme A1, (5.5) holds. However, clearly this scheme does not address the actual problem, since (5.6) and (5.7) only follow if the stationarity of the input signal extends past n sample intervals.

The stationarity of speech is typically about 160 sampling intervals at an 8kHz sampling rate, while at the same rate, the dimension n of acoustic echo paths ranges from 200 to 4000. Thus, in many acoustic echo cancellation cases, Scheme A1 (like scheme A) may lead to poor AR estimation and provide far from white input convergence rates.

Scheme B

A more recent approach proposed [36], [88], [89] is illustrated in Figure 5.11. In this scheme only the signals used in the LMS adaptation algorithm are filtered. The update equation for this approach, which we label **Scheme B**, is

$$\hat{\theta}(k+1) = \hat{\theta}(k) + \mu \hat{W}(k)\hat{A}(k)^T Y_a(k) \tag{5.8}$$

$$\begin{aligned} Y_a(k) &= S_a(k) + [\theta^T U(k)\ \theta^T U(k-1)\ ...\ \theta^T U(k-p)]^T \\ &- [\hat{\theta}^T(k)U(k)\ \hat{\theta}^T(k-1)U(k-1)\ ...\ \hat{\theta}^T(k-p)U(k-p)]^T \end{aligned} \tag{5.9}$$

A modified version of Scheme B is proposed in [88]. It involves replacing

$$\hat{z}(k-i) = \hat{\theta}(k-i)U(k-i), \quad i = 1, 2, ..., p$$

in (5.9) by

$$\hat{z}(k, k-i) = \hat{\theta}(k)U(k-i), \quad i = 1, 2, ..., p.$$

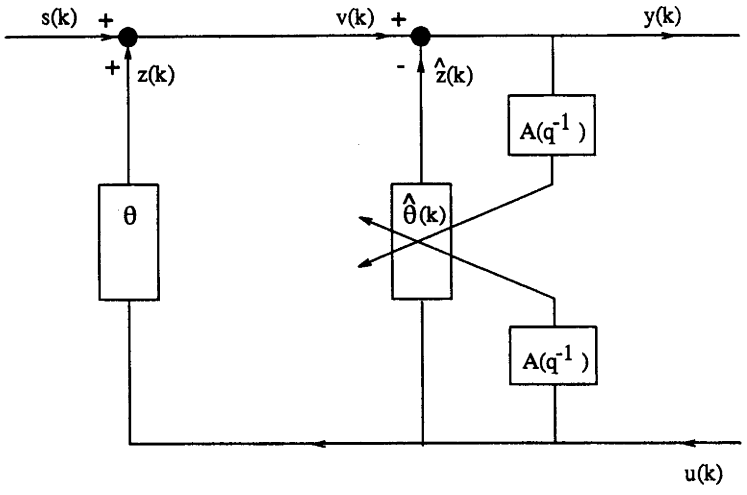


Figure 5.11: The LMS adaptive open loop system with AR whitening Scheme B

That is, the most recent estimate $\hat{\theta}(k)$ is used to obtain 'improved' estimates of the last p echo samples. This modification, in general, requires an extra pn multiplications per sampling interval. The update equation for this modified version, which we call **Scheme B1** is:

$$\hat{\theta}(k+1) = \hat{\theta}(k) + \mu\hat{W}(k)\hat{A}(k)^T\bar{Y}_a(k) \tag{5.10}$$

$$\bar{Y}_a(k) = [U_a(k)\ U_a(k-1)\ ...\ U_a(k-n+1)](\theta - \hat{\theta}(k)) + S_a(k) \tag{5.11}$$

which implies

$$\begin{aligned}(\theta - \hat{\theta}(k+1)) = \tilde{\theta}(k+1) &= (I - \mu\hat{W}(k)\hat{A}(k)^T[U_a(k)\ U_a(k-1)\ ...\ U_a(k-n+1)])\tilde{\theta}(k) \\ &\quad -\mu\hat{W}(k)\hat{A}(k)^T S_a(k)\end{aligned} \tag{5.12}$$

This scheme, like Schemes A, A1 and B, experiences problems with non-stationary input signals - the update equation of (5.12) resembles the LMS equation for a white input, only when the input signal is stationary over a period of at least n sample intervals. The configuration of Scheme B, B1 however, enables the nonstationary input problem to be reduced. In particular, consider the update equation in Scheme B1 for the i^{th} tap coefficient of the echo canceller:

$$\hat{\theta}_i(k+1) = \hat{\theta}_i(k) + \mu\hat{w}(k-i)\hat{A}(k)^T\bar{Y}_a(k) \tag{5.13}$$

The same whitening filter $A_k(q^{-1})$ is used in the updating of all tap coefficients. Equation (5.12) implies that this is the root of the problem. An alternative, which is hinted at, but otherwise ignored in [88], involves using different, more appropriate whitening filters for each tap:

$$\hat{\theta}_i(k+1) = \hat{\theta}_i(k) + \mu\hat{w}(k-i)\hat{A}(k-i)^T\bar{Y}_a(k) \tag{5.14}$$

which implies

$$\begin{aligned}\tilde{\theta}_i(k+1) &= \tilde{\theta}_i(k) - \mu\hat{w}(k-i)\hat{A}(k-i)^T S_a(k) \\ &\quad -\mu\hat{w}(k-i)\hat{A}(k-i)^T[U_a(k)\ U_a(k-1)\ ...\ U_a(k-n+1)](\theta - \hat{\theta}(k))\end{aligned} \tag{5.15}$$

We call this approach **Scheme B2.** Assuming the input signal is stationary over a period of M sample intervals, this scheme leads to the adaptation of the i^{th} tap being decoupled from its M nearest taps: $\hat{\theta}_{i-M/2}, ..., \hat{\theta}_{i+M/2}$. The coupling with the taps outside of the M tap neighbourhood will grow with the nonstationarity of the input signal, but, in many cases, will be 'small'.

The increase in complexity per sample interval of the scheme described over Scheme B1 is np. However, by using the same whitening filter vector $\hat{A}(k-i)$ in the update equations for each of the taps in the M tap neighbourhood of tap i, the increase in complexity is only np/M.

A third version of Scheme B, which we label **Scheme B3** and appears to be novel, uses different AR estimates, as in Scheme B2, but does not use the modification of Scheme B1. The success of this scheme relies on the update stepsize μ being sufficiently small so that $\hat{\theta}(k)$ changes slowly with time in comparison to the signals. Under such conditions, the dynamics of Scheme B3 approximate those of Scheme B2, at least on the short/fast time scale of the signals.

Remark:

1. The error in $\hat{\theta}(k)$ arising from the slowly time varying approximation used in Scheme B3 is similar to that due to the averaging approximation used in Chapters 3 and 4 and, consequently, is typically $O(\mu^\beta)$, where $0 < \beta \leq 1$.

5.3.2 Discussion

The brief analyses of the previous subsection indicate that of the schemes presented, Scheme B2, in general, should provide the best convergence rate improvements over the standard LMS adaptive FIR echo canceller. Furthermore, in many cases, the transient performance of this scheme should closely resemble that achieved with 'white' input signals. The improvements obtained with the remaining schemes should approach that of Scheme B2 when:
(i) μ is sufficiently small - Scheme B3;
(ii) the stationarity of the input signal extends past n sample intervals - Schemes A,

A1 and B1;

(iii) μ is sufficiently small and the input stationarity extends past n sample intervals - Scheme B.

In [36] and [88] the results of various simulations involving Schemes A1, B, B1 and the direct AR filtering scheme of Figure 5.12 are reported. The simulations involve an 'unknown' channel, the impulse response of which is based on measurements of an acoustic echo path. The impulse response length was 512 samples with a sampling rate of 8kHz. The input was speech. A new AR estimate was obtained every $M = 160$ samples. The results reported were an average over either 7 speakers [36] or 16 speakers [88]. The value(s) of the stepsize μ used was not given. It is indicated that each of the three indirect AR filtering schemes provide transient performance improvements over the standard LMS adaptive FIR echo canceller. Furthermore, the transient performance of the these schemes was 'close' to that obtained by the direct AR filtering/whitening scheme. In fact, the transient performance of either Scheme A1 or Scheme B1 was indistinguishable from the direct scheme.

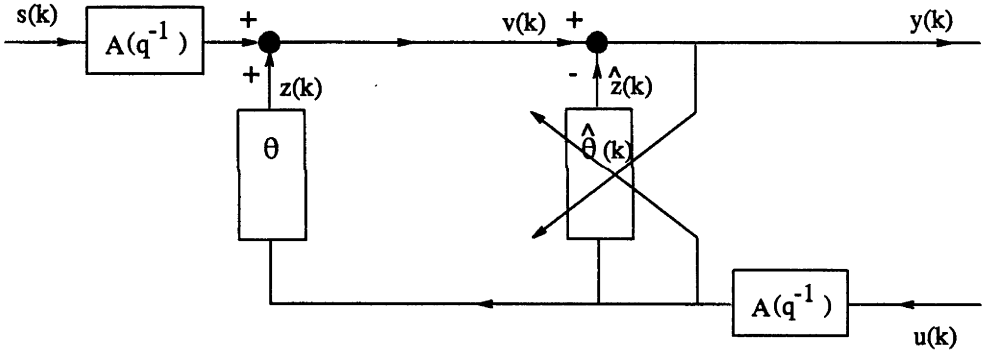


Figure 5.12: The LMS adaptive open loop system with direct AR filtering scheme

The ability of the indirect AR filtering schemes to provide transient performance improvements similar to that obtained with direct AR whitening is further supported by simulations carried out for the thesis. The simulation conditions, however, were less realistic than those reported in [36] and [88]. In particular, the signals used were not speech but synthesized AR signals derived from measured speech signals. Furthermore, the impulse response of each of the unknown channels used was arbitrarily chosen. The tap lengths ranged from 50 to 300. The simulations indicated that all indirect schemes, presented in the previous subsection, performed as well as the direct AR filtering scheme. These results were obtained for stationary as well as non-stationary input AR signals. It should be added, however, that each of the non-stationary signals used was derived from a periodic sequence of AR processes.

Remark:

2. It was noted in [36] and [88] that the performance of the echo canceller in Schemes B and B1 was improved by introducing an appropriate delay ν into the upper AR filter - that used to filter the feedback signal $y(k)$. This delay, effectively, compensates for any delay within the echo path and echo canceller. Use of such delays in the upper AR filter(s) of Schemes A,A1,B2 and B3 should also improve their performance.

A comparison of the complexity of the AR whitening schemes presented is given in Table 5.1. In all but Scheme A1 it is assumed that the AR estimate is obtained only once every $M \leq n$ samples, where the choice of M depends on the stationarity of the input signal. In Scheme A1 the AR estimate is obtained once every n sample intervals. The table includes the number of multiplications required to obtain an AR estimate once every M samples via the Levinson-Durbin algorithm , which as indicated in Appendix D, is approximately $2p + (4p + p^2)/M$.

Table 5.1: Comparison of Complexity

Scheme	Multiplications per sample	Memory
Standard	$2n$	n
A	$2n + 6p + (4p + p^2)/M$	$3n + 3p$
A1	$2n + 6p + (4p + p^2)/n$	$3n + 3p$
B	$2n + 4p + (4p + p^2)/M$	$3n + 2p$
B1	$(2 + p)n + 4p + (4p + p^2)/M$	$3n + 2p$
B2	$(2 + p + p/M)n + 4p + (4p + p^2)/M$	$3n + 2p$
B3	$(2 + p/M)n + 4p + (4p + p^2)/M$	$3n + 2p$

Typically, the order of the AR filter for speech is $p = 10$ to 50. Studies carried out in [36], [88] indicate that good transient performance is achievable with AR filters of lower order $p < 10$. Combining this with the fact that for acoustic echo cancellation $n = 400 : 4000$ suggests that the complexity of the AR whitening schemes is determined largely by the size of n. Consequently, Schemes A, A1 and B typically introduce very little extra computational cost over the standard approach. For speech input signals in which the stationarity length $M \approx 160$ is considerably larger than the AR model order $p \approx 10$, Scheme B3 also introduces little extra computation. On the other hand, the cost of Schemes B1 and B2 is about $(2 + p)/2$ times, that of the standard approach.

It is suggested in [88] that the performance benefits gained in Schemes B and B1 by increasing the AR order, drops quickly after $p = 1$. In fact, for Scheme B1, an

order of $p = 2$ is suggested to be optimal for the lower (input signal, $u(k)$) AR filter, while $p = 1$ is suggested for the upper (residual echo signal, $y(k)$) AR filter. This reduction in AR filter orders is taken further in [89], where it is suggested that the top AR filter is not needed in Scheme B. It should be added that the algorithm used in [89], although similar to Scheme B, employed modifications such as variable stepsize. Taking these comments into account, the computational cost of Schemes B1 and B2, in practice, may be only 2 to 3 times that of Schemes A, A1 and B.

As usual, the most appropriate choice of the indirect AR schemes presented depends on the importance of low computational cost relative to performance. This choice is influenced by the system properties. When the input signal is stationary, the low cost Schemes A, A1 and B should be chosen. As the signals become more non-stationary then, as long as μ is sufficiently small, Scheme B3 should be chosen - it provides the best compromise between cost and performance. It should be added that a disadvantage of Schemes A and A1 is that of poorer tracking capabilities of the echo path. This is due to the delay introduced by the inverse AR filter and the AR filter located just before and after the echo path, respectively.

5.4 Conclusion

In this chapter we presented two signal conditioning schemes, the objective of each being to improve the performance of LMS adaptive FIR echo cancellers in speech transmission telecommunication networks. The first scheme focussed on whitening the subscriber signals of the telecommunication network and, consequently, is not applicable to acoustic echo cancellation. It involved placing a low computational digital scrambling/descrambling device with the transmitter/receiver of each subscriber. The scrambled subscriber signals typically show 'white' autocorrelation functions and zero cross correlation. Under some circumstances, cross correlation may occur. A simple extension of the scheme, which basically has each subscriber using a different scrambling device, removes this adverse possibility. The scheme is expected to enhance the performance of the LMS/FIR echo cancelling system to that achieved with white uncorrelated subscriber signals. This is supported by simulations.

The second scheme focussed on whitening the input of each echo canceller within the network. As indicated by the analyses of Chapter 4, the success of this approach relies on the transmission delay, imposed by the central 4-wire loop channels, being sufficiently large. The scheme, which may be used for both 4-wire loop circuit echo

cancellation and acoustic echo cancellation, assumes that the input signal, like speech, is well modelled as an autoregressive (AR) process. In acoustic echo cancellation, use of an estimate of the AR filter to whiten the input directly is not permitted - the input is that received by the acoustic enclosure. Alternative schemes, using additional filtering or based on AR filtering only of the signals used in the LMS algorithm, were examined. The computational complexity of such schemes is generally comparable to that of the standard LMS/FIR filter. The schemes should improve the performance of the LMS/FIR filter and, in many cases, the performance should match or closely approach that achieved with white input signals.

Chapter 6

Dimension Reduced LMS/FIR Estimation

6.1 Introduction

The analyses of Chapters 3 and 4 indicated that an increase in the parameter dimension of the LMS adaptive FIR echo canceller adversely affects the transient and asymptotic performance of the echo canceller (Chapter 3) and of the double echo canceller telecommunication network (Chapter 4). This adverse dependence is accentuated by high signal correlation levels and can be reduced significantly by applying the signal conditioning schemes proposed in Chapter 5. However, even when the network's subscriber signals and/or the echo cancellers' input signals are white and non cross correlated, an increase in dimension n leads to a deterioration (linearly in n) in asymptotic performance - when the update stepsize μ is fixed. On the other hand, if μ is reduced so as to maintain asymptotic performance, then an increase in dimension leads to a linear reduction in transient performance.

These comments imply that significant improvements in asymptotic and/or transient performance might be achieved by basing the echo canceller on more efficiently parametrized models (of echo paths) than the standard FIR model. This approach has been followed before and lead to the investigation of IIR parametrized models, particularly for acoustic echo cancellation. However, as indicated in Chapter 2, despite the reduction in parameter dimension, IIR based echo cancellers are generally inferior to the FIR echo canceller. A less ambitious, more flexible parametrized model is that in which a number of the FIR taps are fixed to zero or, effectively, are made 'inactive'. Such a parametrization may be suitable for echo cancellation since, as

discussed in more detail in Section 6.2, the impulse response of an echo path typically shows inactive or zero regions. An appealing feature about this 'active' parametrization is that, unlike the IIR parametrization, it is FIR based and, consequently, the analytical results of Chapters 3 and 4 apply, at least qualitatively.

The 'active' parametrization has been used by various authors such as [40], [39] in the development of algorithms for channel estimation applications such as echo cancellation. The approach taken in [40] to estimate the active taps involves: (i) detecting the position of the most active (largest) tap by comparing cross correlation estimates:

$$\text{position of most active tap} = \arg_i \max \left\{ \frac{\sum_{k=i+1}^{N} v(k)u(k-i)}{\sum_{k=i+1}^{N} u^2(k-i)} \right\} \tag{6.1}$$

where $v(k)$ is the observed output of the echo path; (ii) LMS adaptively estimating the coefficient of this tap; (iii) by effectively removing this estimated tap, detecting the position of the second most active tap; (iv) repeating the procedure until all m active taps are estimated. As reported, such an approach leads to improved LMS convergence properties when m is sufficiently small.

The alternate approach taken in [39] aims at jointly (position) detecting and (coefficient) estimating the m most active taps by comparing Normalized Least Mean Square (NLMS) estimates of each tap coefficient. At any one time only $\bar{m} > m$ taps out of a total of n taps are estimated. After a given number of iterations, the m most active taps are chosen while the remaining $\bar{m} - m$ (least active current) taps are replaced by other taps not previously considered. The procedure is repeated until all n taps have been considered. Reported simulation results indicate convergence rate improvements over simultaneous NLMS estimation of all n taps.

A possible modification to the approach of [39], so as to achieve further convergence rate improvements, would be to use estimation algorithms, such as the Least Squares (LS) algorithm, which converge faster, at least initially, than the LMS algorithm. This, however, would lead to greater computational requirements. An alternative approach is to use a suitable low computational approximation of the LS algorithm to detect the position of the active taps prior to LMS estimation of their coefficients. This is the basis of our approach. In particular, we show through 'concrete' analysis, that for white input signals a more suitable measure of tap activity than that of (6.1) is:

$$X_j = \frac{[\sum_{k=i+1}^{N} v(k)u(k-i)]^2}{\sum_{k=i+1}^{N} u^2(k-i)}. \tag{6.2}$$

A number of issues concerning this approach need to be addressed. In particular,

having detected the position of the active regions, and copied these into the echo canceller prior to estimation, under what conditions can we expect improved estimation performance. Problems are likely to occur if the input signal is autocorrelated, since this would lead to coupling of the taps (both active and inactive) within the echo canceller. Biased estimation may result. Such coupling may also affect the accuracy of the detection procedure. Another issue is that of the additional computational complexity introduced by the 'active' tap detector. As we shall see, the coupling and computational difficulties are avoided or minimized by ensuring that the input signal to the echo path/canceller is 'white'.

The chapter is organized as follows. We begin in Section 2 by examining typical circuit and acoustic echo path impulse response structures. This suggests that, for an echo path having an n-tap long impulse response, a suitable parametrization for the echo canceller is that in which $n - m$ of the taps are set to zero. The number m of nonzero or 'active' taps maybe chosen a-priori or, preferably, determined on-line. In Section 3, we formalize our system with a number of assumptions. In Section 4 we conduct a brief analysis to indicate that when the input signal is 'white', this active' tap parametrization provides unbiased estimation as well as improved performance. In Section 5, we develop low complexity procedures, based on the Least Squares (LS) method, for detecting the number and positions of the active taps of the echo path. Using the results of Section 5, we propose in Section 6 a low complexity algorithm for on-line detection (of position) and LMS estimation (of coefficients) of the active taps of the echo path. Simulations demonstrate the performance advantages of this algorithm over the standard LMS algorithm.

6.2 Echo Path Structures

As mentioned in Chapter 1, circuit echo paths within 4-wire loop telephony networks typically have impulse responses that consist of an initial inactive (or zero) region followed by a dispersive (or nonzero) region [1]. Furthermore, the inactive region maybe considerably longer than the dispersive region. This suggests that a suitable parametrization for circuit echo cancellers is that in which the coefficients of the first M FIR taps is set to zero, where M is an unknown number to be determined.

Like circuit echo paths, the impulse responses of acoustic echo paths may also show an initial flat region. They may also show a sparse structure - that is, nonzero taps or groups of nonzero taps sparsely separated by taps having zero or insignificant

coefficients. This sparse structure description is suggested by the fact that only certain acoustic tracks between the loudspeaker (receiver) and microphone (transmitter) of the acoustic enclosure will involve dominant reflection points, such as walls. Such tracks result in dominant nonzero taps (or small groups of dominant nonzero taps) within the impulse response of the acoustic echo path. Furthermore, these nonzero taps will tend to be sparsely separated and their number, considerably less than the total tap length of the echo path.

This sparse structure description for acoustic echo paths also has been suggested by other authors, [90], [91]. In particular, [91] proposes as a model for acoustic echo paths, a sequence of two filters such that the overall impulse is formed by the response of an FIR filter driven by a sparse delta sequence. An examination of acoustic echo path impulse responses lends support to this notion that acoustic echo paths are sparsely structured. An example is shown in Figure 6.1a. This impulse response was derived from a measured acoustic echo path impulse response, Figure 6.1b, by applying the technique presented in Appendix E, which essentially removes the effects of estimation/measurement noise. The measured impulse response of Figure 6.1b was obtained [1] from a room approximately $5m \times 10m \times 3m$.

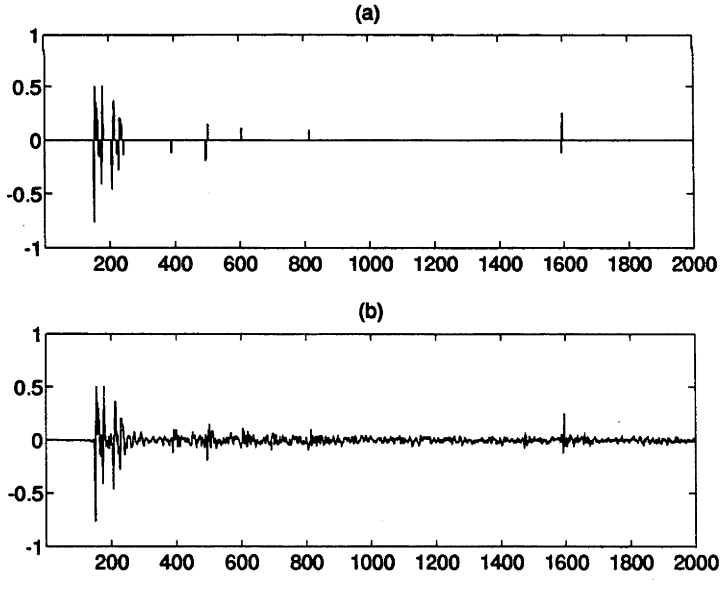


Figure 6.1: Acoustic echo path impulse response showing sparse structure: (a) is derived from the measured impulse response shown in (b) via the technique of Appendix E.

[1] The measurements were made by CSIRO Radiophysics, Sydney, Australia.

The above comments suggest that a suitable parametrization model for both circuit and acoustic echo cancellers is that of an FIR filter in which the coefficients of $n-m$ of the n FIR taps are set to zero. The remaining m taps are allowed to be nonzero or 'active'. The position or lag of each of these active taps is to be determined. The number m should also be determined, although, a simpler approach might involve selecting a value for m, one which over-estimates the number of active taps.

6.3 System Description

Throughout this chapter we consider the open loop LMS adaptive estimation set-up of Figure 1.5 of Chapter 3. We assume that Assumptions 1-6 of Chapter 3 are valid - see Appendix I for an easy reference to these assumptions. At sampling instant k, the input to the echo path/unknown channel is $u(k)$, an additive disturbance $s(k)$ occurs within the unknown channel, $v(k)$ is the observed output of the unknown channel and $\hat{z}(k)$ is the output of the unknown channel estimator. As will become evident in the following sections, to avoid problems with the use of the active tap parametrization as well as to enable the development of our fast, low computational active tap detection schemes, we require that the input signal satisfies the following assumption.

Assumption 21 *The input signal is uncorrelated over time (white) such that the autocorrelation matrix of the input signal $u(k)$:*

$$R_n = \sigma_u^2 I$$

where I is the $n \times n$ identity matrix and σ_u^2 is the variance of $u(k)$.

The use of the active tap parametrization is based on the validity of the following assumption.

Assumption 22 *The time invariant n-tuple FIR modelled echo , $\Theta(q^{-1})$, has only $m < n$ nonzero taps:*

$$\Theta(q^{-1}) = b_{j_1} q^{-j_1} + b_{j_2} q^{-j_2} + ... + b_{j_m} q^{-j_m} \tag{6.3}$$

where $0 \leq j_i < j_{i+1} \leq n, \quad i = 1, 2, ..., m.$

Remark:

1. Assumption 22 is rather general. Clearly it includes the case in which the echo path has a sparse impulse response structure, typical of acoustic echo paths, and the case in which the impulse response has an initial flat or 'inactive' region, as in circuit echo paths.

6.4 Requirements and Benefits of Active Tap Parametrization

In this section we argue that the active tap parametrization tends to cause difficulties, unless the input signal is white. After invoking Assumption 21, we then carry out a brief analysis to indicate the benefits such a parametrization can yield when the echo path satisfies Assumption 22.

Consider the case in which the input is filtered white noise, $w(k)$:

$$u(k) = F(q^{-1})w(k), \quad E[w^2(k)] = \sigma_u^2$$

The observed output of the echo path of Assumption 22 is:

$$v(k) = \Theta(q^{-1})F(q^{-1})w(k) + s(k) = \sum_{i=1}^{m} b_{j_i} F(q^{-1})w(k - j_i) + s(k) \qquad (6.4)$$

and the output of the estimator $\hat{\Theta}(q^{-1})$ with all n taps active is:

$$\hat{z}(k) = \hat{\Theta}(q^{-1})F(q^{-1})w(k) = \sum_{i=i}^{n} \hat{\theta}_i F(q^{-1})w(k - i) \qquad (6.5)$$

The quality/performance of the estimated model can be measured by (for fixed $\hat{\Theta}(q^{-1})$), $E[(v(k) - \hat{z}(k))^2]$. Observe that:

$$E[(v(k) - \hat{z}(k))^2] = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\Theta(e^{i\omega}) - \hat{\Theta}(e^{i\omega})|^2 |F(e^{i\omega})|^2 \sigma_u^2 d\omega + \sigma_s^2 \qquad (6.6)$$

The input signal filter $F(q^{-1})$ results in a frequency weighting on the estimation error, that is, a biased estimation error. In the time domain, this may be interpretted as coupling between the taps of the residual filter $\Theta(q^{-1}) - \hat{\Theta}(q^{-1})$. In particular, coupling leads to nonactive taps appearing to be active and active taps appearing less active. Under such coupling, the detection of active taps and the unbiased estimation of their coefficients is likely to be difficult.

The coupling increases as the impulse response of $F(q^{-1})$ lengthens, or equivalently, as the input becomes more autocorrelated. On the other hand, coupling can be reduced by using approximate input signal whitening schemes. If the active taps are sufficiently sparsely separated, the whitening schemes may remove the coupling completely. Zero coupling, however, is only guaranteed when the input signal is white or satisfies Assumption 21, so that

$$F(q^{-1}) = 1 = |F(e^{i\omega})|$$

Under the validity of Assumption 21, the following discussion motivates the use of the active tap parametrization. When the input is white (6.6) becomes:

$$E[v(k) - \hat{z}(k))^2] = [\sum_{i=1}^{m}(b_{j_i} - \hat{\theta}_{j_i})^2 + \sum_{i \neq j_i} \hat{\theta}_i^2]\sigma_u^2 + \sigma_s^2 \tag{6.7}$$

Note that the taps of the estimates are decoupled. Define

$$\Delta\theta_i(N) \triangleq \theta_i - \hat{\theta}_i(N)$$

where N is the number of data points used to obtain the estimate. Since the input is white, then the following property will hold asymptotically in N for different estimates of θ_i.

$$\lim_{N\to\infty} \{\Delta\theta_i(N)\} \sim AsN(0, \sigma_i),$$
$$\lim_{N\to\infty} E(\Delta\theta_i(N)\Delta\theta_j(N)) = \left\{ \begin{array}{ll} \sigma_i^2 & i = j \\ 0 & i \neq j \end{array} \right\}$$

For example, the cross correlation estimator

$$\hat{\theta}_i(N) = \left[\frac{\sum_{k=i+1}^{N} v(k)u(k-i)}{\sum_{k=i+1}^{N} u(k-i)^2} \right]$$

has this property with $\sigma_i = \sqrt{\frac{\sigma_s^2}{\sigma_u^2(N-i)}}$. The same holds for the least squares (LS) estimator. The equivalent result for the LMS estimator is:

$$\sigma_i = \sqrt{\mu\sigma_s^2/2} \tag{6.8}$$

Consequently, with expectation now with respect to $\Delta\theta_i(N)$:

$$E\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} |\Theta(e^{i\omega}) - \hat{\Theta}(e^{i\omega})|^2 \sigma_u^2 d\omega\right)$$
$$= \left[\sum_{i=1}^{n} \sigma_i^2\right] \sigma_u^2 \tag{6.9}$$

This implies that the asymptotic error increases linearly with the number of estimated coefficients. On the other hand, as indicated by (6.7), neglecting to estimate a tap b_{j_i} which contributes to $v(k)$ results in a bias error. The corresponding error is proportional to:

$$b_{j_i}^2 \sigma_u^2$$

The above discussion indicates that improved asymptotic performance of the LMS estimator (and of the LS estimator) can be expected by estimating only those taps which satisfy:

$$\theta_i^2 > \sigma_i^2$$

that is, only those taps which can be distinguished from the estimation noise, or, are "active". This clearly motivates the use of active tap detection schemes.

Remark:

2. The above discussion does not consider transient performance. However, as quantified in Chapter 3, when the input signal is white (that is Assumption 21 is valid), then for fixed μ, the convergence rate of the LMS estimator is essentially independent of the number of estimated coefficients, n. On the other hand, if an improvement in asymptotic performance is not as important as transient performance, then μ could be increased linearly as the number of estimated coefficients n decreases. According to the analyses of Chapter 3, this leads to the asymptotic performance being maintained and the transient performance increasing linearly.

   In short, a reduction in the number of estimated parameters (but not below the number of active taps within the echo path), improves transient performance (if μ is increased with decreasing n) or asymptotic performance (if μ is fixed). The LMS estimation algorithm presented in this chapter considers only the case of fixed μ - which results in enhanced asymptotic performance.

6.5 Active Tap Detection

In this section we develop procedures, based on the Least Squares (LS) method, for detecting the position of the active taps of the FIR modelled echo path. We begin by considering the case in which the number of active taps, m, is known, or, alternatively, the case in which we want to detect only the m most active taps. By modifying the LS method appropriately, we then develop a simple scheme for determining the

number of unknown active taps, in addition to detecting their positions. Simulations demonstrate the ability of the scheme to detect the correct number of active taps even with an input to disturbance ratio of $0dB$.

Consider the echo path as described by Assumption 22, which we parametrize by the n long parameter vector θ:

$$\theta = (0(j_1 - 1), b_{j_1}, 0(j_2 - j_1 - 1), b_{j_2}, ..., 0(j_m - j_{m-1}), b_{j_m}, 0(n - j_m))^T \qquad (6.10)$$

where $0(j)$ is the zero matrix of size $1 \times j$. Our aim is to determine the positions of the m nonzero elements of θ. We can achieve this by obtaining an estimate $\hat{\theta}$ of θ which has only m nonzero elements. An approach to obtaining this estimate is to consider minimizing the Least Squares cost function:

$$V_N(\hat{\theta}(N)) = \sum_{k=1}^{N}(v(k) - U(k)^T\hat{\theta}(N))^2 \qquad (6.11)$$

under the restriction that all but m elements of $\hat{\theta}$ are zero. As before, $U(k)$ is the $n \times 1$ input signal vector at time k.

To write this more formally, we introduce the following notation.

$$\begin{aligned} t^m &\triangleq \text{the set}\{t_1, t_2, ..., t_m\} \\ \hat{\theta}(N, t^m) &\triangleq (\hat{b}_{t_1}(N), \hat{b}_{t_2}(N), ..., \hat{b}_{t_m}(N))^T \\ J(t^m) &\triangleq n \times m \text{ matrix with} \\ &\qquad l^{th}\text{row} = 0(m) \text{ if } l \neq t_i \\ &\qquad l^{th}\text{row} = (0(t_i - 1), 1, 0(m - t_i)) \text{ if } l = t_i \end{aligned}$$

The restricted LS cost function of (6.11) can be then rewritten as:

$$\begin{aligned} V_N(\hat{\theta}(N, t^m)) &= N[J(t^m)\hat{\theta}(N, t^m) - \hat{\theta}^{LS}(N)]^T R_N[J(t^m)\hat{\theta}(N, t^m) - \hat{\theta}^{LS}(N)] \\ &+ V_N(\hat{\theta}^{LS}(N)) \end{aligned} \qquad (6.12)$$

where $\hat{\theta}^{LS}(N)$ is the unrestricted LS estimate (that is, all n elements of $\hat{\theta}^{LS}(N)$ may be nonzero) and is given by:

$$\hat{\theta}^{LS}(N) = R(N)^{-1}f(N) \qquad (6.13)$$

$$R(N) = \frac{1}{N}\sum_{k=1}^{N} U(k)U(k)^T \qquad (6.14)$$

$$f(N) = \frac{1}{N}\sum_{k=1}^{N} v(k)U(k) \qquad (6.15)$$

Remark:

3. Since $v(k) = \theta^T U(k) + s(k)$ then

$$\begin{aligned} f(N) &= R(N)\theta + \frac{1}{N}\sum_{k=1}^{N} U(k)s(k) \\ &\rightarrow R\theta \text{ as } N \rightarrow \infty \text{ when Assumption 3 is valid} \end{aligned}$$

This leads to $\hat{\theta}^{LS}(N) \rightarrow \theta$ as $N \rightarrow \infty$, that is, consistency.

4. In general, even if θ has only m nonzero elements, the unrestricted LS estimate $\hat{\theta}^{LS}(N)$ will be optimal (in a LS sense):

$$V_N(\hat{\theta}^{LS}(N)) \leq V_N(\hat{\theta}(N, t^m))$$

The solution to our problem is given by the $arg\hat{\theta}(N, t^m)$ which minimizes the multivariable LS cost function

$$W_N(\hat{\theta}(N, t^m)) \triangleq N[J(t^m)\hat{\theta}(N, t^m) - \hat{\theta}^{LS}(N)]^T R(N)[J(t^m)\hat{\theta}(N, t^m) - \hat{\theta}^{LS}(N)] \quad (6.16)$$

and is given by:

$$\begin{aligned} \hat{\theta}^{LS}(N, t^m) &= [J(t^m)^T R(N) J(t^m)]^{-1} J(t^m)^T R(N)\hat{\theta}^{LS}(N) \\ &= [J(t^m)^T R(N) J(t^m)]^{-1} J(t^m)^T f(N) \end{aligned} \quad (6.17)$$

The corresponding cost is

$$\begin{aligned} W_N(\hat{\theta}^{LS}(N, t^m)) &= N(\hat{\theta}^{LS}(N))^T [R(N) \\ &\quad - R(N)J(t^m)[J(t^m)^T R(N) J(t^m)]^{-1} J(t^m)^T R(N)]\hat{\theta}^{LS}(N) \\ &= N f(N)^T [R(N)^{-1} \\ &\quad - J(t^m)[J(t^m)^T R(N) J(t^m)]^{-1} J(t^m)^T] f(N) \end{aligned} \quad (6.18)$$

The cost function of (6.18) depends only on the input and disturbance signals, through $R(N)$ and $f(N)$, and on the set of tap/element positions t^m. We can achieve our aim of determining the positions of the m active taps of the unknown channel by determining the set t^m which minimizes the RHS of (6.18). This involves calculating and comparing the value of the RHS of (6.18) for $(n)!/[(m)!(n-m)!]$ different combinations. This is likely to be a mammoth task. The following Lemma, which extends the work of [92], indicates that, for sufficiently large N, the task can be considerably simplified if the input signal is white.

Lemma 2 *Consider signals $u(k)$ and $s(k)$ which satisfy the assumptions detailed in Section 3. Let $U(k)$ be the n-tap input signal vector, and $R(N), f(N), \hat{\theta}^{LS}(N,t^m)$, $W_N(\hat{\theta}^{LS}(N,t^m))$ be as defined in (6.14),(6.15),(6.13),(6.18), respectively.*

Then:

$$||\hat{\theta}^{LS}(N,t^m) - \bar{\theta}(N,t^m)||_2 \rightarrow 0 \quad w.p.1 \text{ as } N \rightarrow \infty \tag{6.19}$$

$$1/N||W_N(\hat{\theta}^{LS}(N,t^m)) - \bar{W}_N(t^m)||_2 \rightarrow 0 \quad w.p.1 \text{ as } N \rightarrow \infty \tag{6.20}$$

where:

$$\bar{\theta}(N,t^m) = [J(t^m)^T \bar{R}(N) J(t^m)]^{-1} J(t^m)^T f(N) \tag{6.21}$$

$$\bar{W}_N(t^m) = N f(N)^T [\bar{R}(N)^{-1} - J(t^m)[J(t^m)^T \bar{R}(N) J(t^m)]^{-1} J(t^m)^T] f(N) \tag{6.22}$$

$$\bar{R}(N) = \frac{1}{N} \sum_{k=1}^{N} diag(u(k)^2, u(k-1)^2, ..., u(k-n+1)^2) \tag{6.23}$$

Proof: See Appendix H.

The cost function $\bar{W}_N(t^m)$ has the important property that the contribution of each tap is decoupled from the rest. Consequently, the set of m tap positions which minimizes $\bar{W}_N(t^m)$ is given by the indices $t_i = j$ corresponding to the m greatest values of:

$$X_N(j) \triangleq \frac{\left[\sum_{k=j+1}^{N} v(k)u(k-j)\right]^2}{\sum_{k=j+1}^{N} u^2(k-j)} \tag{6.24}$$

This leads to the following important result.

Result 10 *Subject to the conditions of Lemma 2, then, for sufficiently large N, the positions of the m most active taps of the FIR modelled channel are given by the indices corresponding to the m greatest values of $X_N(j)$ of (6.24).*

An equivalent way of stating Result 10, which will prove more useful in the subsequent discussion, is that the positions of the m most active taps are given by the indices corresponding to the m smallest values of the single tap LS cost functions:

$$\bar{V}_N(j) \triangleq \sum_{k=1}^{N} v^2(k) - X_N(j) \tag{6.25}$$

It is important to note that, subject to the conditions of Lemma 2,

$$\sum_{i=0}^{m} \bar{V}_N(t_i) \rightarrow V_N(\hat{\theta}(N, t^m)) + m \sum_{k=1}^{N} v^2(k) \; as \; N \rightarrow \infty \tag{6.26}$$

A more sophisticated approach than that above enables the number of active taps to be determined as well as their positions. Such an approach requires the cost function considered to have the property of structural consistency (that is the correct number and position of active taps is determined as $N \rightarrow \infty$). The LS cost function does not enjoy this property. In particular, the LS cost function $V_N(\hat{\theta})$, irrespective of the true order of the FIR modelled unknown channel, decreases as the order of the LS estimate increases. A standard approach to counteract against this bias towards higher order estimates, so as to enable estimation of the true model order, is to introduce some term which penalizes order or dimension. Typically, one uses

$$V_N(\hat{\theta})[1 + m\frac{C(N)}{N}] \tag{6.27}$$

or

$$\log V_N(\hat{\theta}) + m\frac{C(N)}{N} \tag{6.28}$$

where m is the dimension or number of estimated taps. To guarantee structural consistency [93] , the following condition must hold.

$$C(N) \rightarrow \infty, \; \frac{C(N)}{N} \rightarrow 0 \;\; \text{as } N \rightarrow \infty. \tag{6.29}$$

Some of the well known dimension penalizing cost functions [93] which satisfy this consistency condition are:

$$\begin{aligned} V_N^{BIC} &= \log V_N(\hat{\theta}) + m\frac{\log N}{N} \;\; \text{Akaike's B-Information Criterium} \\ V_N^{\phi} &= \log V_N(\hat{\theta}) + m\frac{\log\log N}{N} \;\; \phi\text{-Criterium} \end{aligned}$$

Because $V_N(\hat{\theta}) = O(N)$, the following alternate cost function also enjoys the property of structural consistency:

$$V_N^{alt} \triangleq V_N(\hat{\theta}) + m \log N \tag{6.30}$$

$$\sim V_N(\hat{\theta}) \left(1 + (\frac{N}{V_N(\hat{\theta})})\frac{m \log N}{N}\right) \tag{6.31}$$

Remark:

5. A more general form of (6.30) is:

$$V_N^{alt2} \triangleq V_N(\hat{\theta}) + Km \log N \tag{6.32}$$

where K is a constant independent of m and N, but may be dependent on, for example, the variance, σ_s^2 and σ_u^2, of the disturbance and input signals, respectively.

Subject to the conditions of Lemma 2, for sufficiently large N, the result of (6.26) indicates that we can approximate

$$V_N^{alt}(N, t^m) \triangleq V_N(\hat{\theta}(N, t^m)) + m \log N$$

by the decoupled cost function:

$$\bar{V}_N^{alt}(t^m) \triangleq \sum_{i=1}^{m} \bar{V}_N^{alt}(t_i) - m \sum_{k=1}^{N} v^2(k) \tag{6.33}$$

$$\text{where } \bar{V}_N^{alt}(t_i) \triangleq \bar{V}_N(t_i) + \log N \tag{6.34}$$

and $\bar{V}_N(t_i)$ is as defined in (6.25). Alternatively,

$$\bar{V}_N^{alt}(t^m) = \sum_{k=1}^{N} v^2(k) - \sum_{i=1}^{m} [X_N(t_i) - \log N] \tag{6.35}$$

Thus, for sufficiently large N, an estimate of the number m and the positions t_i, $i = 1, 2, ..., m$ of active taps is given by that set of indices which minimizes the RHS of (6.35) or, equivalently, which maximizes

$$X_N^{alt}(t^m) \triangleq \sum_{i=1}^{m} [X_N(t_i) - \log N] \tag{6.36}$$

It is clear that the RHS of (6.36) is a monotonically increasing function of m so long as $X_N^{alt}(t_i) > \log N$. This leads to the following important result.

Result 11 *Subject to the conditions of Lemma 2, then, for sufficiently large N, the positions of the active taps of an FIR modelled channel are given by those indices j for which*

$$X_N(j) > \log N \tag{6.37}$$

Remark:

6. It should be emphasized that Result 11 provides a criterion for estimating the number m and positions t_i of the active taps of the echo path. In comparison, Result 10 only provides a criterion for determining the positions of the m most active taps, or alternatively, the positions of an a-priori known number of active taps.

Result 11 suggests the following simple procedure for detecting the positions of the active taps of the echo path.

Algorithm 1 *1. Set* $f_j(0) = 0,\ r_j(0) = r_0,\ x_j(0) = 0,\ j = 1, 2, ..., n.$

2. Update $f_{j\leq k}(k),\ r_{j\leq k}(k),\ x_{j\leq k}(k)$ *at time* k *via:*

$$\begin{aligned} f_j(k) &= f_j(k-1) + v(k)u(k-j) \\ r_j(k) &= r_j(k-1) + u(k-j)^2 \\ x_j(k) &= \frac{f_j^2(k)}{r_j(k)} \end{aligned}$$

3. At time k*, an estimate of the positions of the active taps is given by the set of indices which satisfy:*

$$x_j(k) > \log(k) \tag{6.38}$$

Remark:

7. The initial condition r_0 of $r_j,\ j = 1, 2, ..., n$ must be nonzero but should be sufficiently small so that its effect on $r_j(k)$ rapidly decays with k.

8. The condition of (6.37) does not require a comparison of all n taps (which is the requirement suggested by Result 10). That is, there is no need for simultaneous computation of n values of $X_N(j)$. In particular, this means that the detection of the active taps can be carried out in consecutive blocks of length $\bar{n} < n$ rather than in one block of length n - as is the method of Algorithm 1. Clearly this is beneficial when there are computational constraints.

The results of two sets of simulations based on Algorithm 1 are shown in Figure 6.2 - plotted is the number of active taps detected over time. The mean result of ten similar simulations is shown. The impulse response (in discrete time domain) of the channel being estimated is shown in Figure 6.3 - 11 out of 300 taps are nonzero. This impulse response was derived (via the approach of Appendix E) from a truncated version of a room acoustic impulse response. The input and disturbance signals used in the simulations were both zero mean white Gaussian signals with variance:
(a) $\sigma_u^2 = 1.0$ and $\sigma_s^2 = 0.01$, (b) $\sigma_u^2 = 1.0$ and $\sigma_s^2 = 1.0$.
Similar results were achieved for nonGaussian (zero mean white) signals and for the case in which the disturbance signal is nonwhite (but still zero mean and uncorrelated with the input).

It is interesting to note that in the case of relatively low level disturbance, the algorithm, as indicated by Figure 6.2a, converges relatively slowly but smoothly to provide an unbiased estimate of the number of active taps. In the relatively high level disturbance case of Figure 6.2b, the algorithm provides an estimate which increases quickly towards the active tap number rapidly, but overshoots. The estimate then decays, with some 'ringing', towards the true number of active taps. The apparent asymptotic overestimation is due to estimation noise - an examination of the results of individual simulations indicates the estimation is unbiased. Many simulations support these comments and suggest that the algorithm experiences significant overshooting/overestimation problems when the input to disturbance ratio drops below $0dB$.

6.6 LMS Estimation via Detection

In this section we use Result 11 to propose an algorithm for LMS estimating the coefficients of only the active taps of the FIR modelled unknown channel/echo path. Simulations demonstrate the performance advantages of this algorithm over that of the standard LMS algorithm.

One approach to LMS estimating the active taps of the echo path, which makes use of the detection results derived in Section 5, would involve first running Algorithm 1 for a given number $N \gg n$ of sample intervals and then LMS estimating the coefficients only of those taps detected as being active (at time $k = N$). The approximate LS

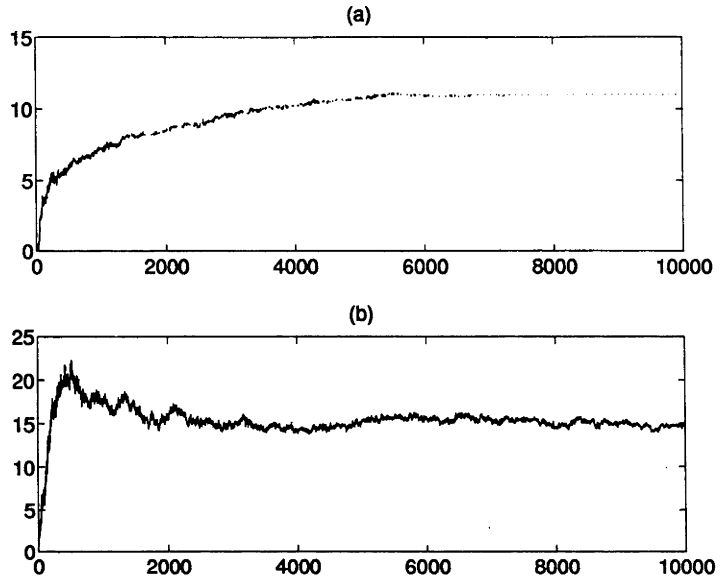


Figure 6.2: Number of active FIR taps detected over time of the Channel of Figure 6.3 - detection scheme based on that of Result 11

estimate of each of the detected active taps, as given by (6.21):

$$\bar{\theta}_j(N) = \frac{\sum_{k=j+1}^{N} v(k)u(k-j)}{\sum_{k=j+1}^{N} u(k-j)^2}$$

could be used as the initial LMS estimate.

An alternative LMS estimation approach to that suggested above involves determining at each sample interval k the indices which satisfy the active condition $x_j > \log(k)$. The corresponding taps in the LMS estimator are then LMS adapted for that sample interval. The LMS coefficients of those taps which are not detected as being active may be frozen for that sample interval or, alternatively, may have a forgetting function applied to them. The application of a forgetting function is actually preferred because it ensures that the estimate is structurally consistent. This approach is summarized by the following algorithm.

Algorithm 2 *1. Choose the LMS forgetting factor* $\alpha \in [0, 1)$.
Set $f_j(0) = 0,\ r_j(0) = r_0,\ x_j(0) = 0,\ \hat{\theta}_j(0) = 0,\ j = 1, 2, ..., n.$

2. Update $f_j(j \leq k),\ r_j(j \leq k),\ x_j(j \leq k)$ *at time* k *via:*

$$\begin{aligned} f_j(k) &= f_j(k-1) + v(k)u(k-j) \\ r_j(k) &= r_j(k-1) + u(k-j)^2 \end{aligned}$$

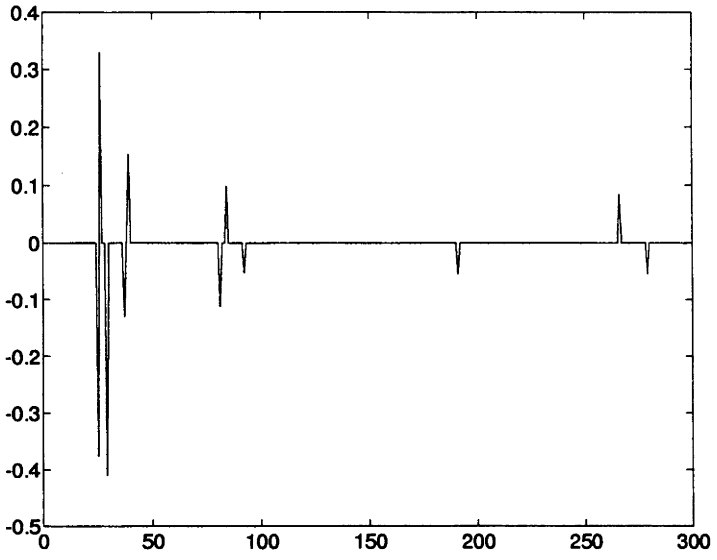


Figure 6.3: The impulse response of the Channel used in simulations.

$$x_j(k) = \frac{f_j^2(k)}{r_j(k)}$$

3. Determine the set of indices $\{t^m\}$ *which satisfy* $x_j(k) > \log(k)$. *Construct an* $n \times 1$ *vector* $g(k)$ *with ones in the positions corresponding to the set of indices* $\{t^m\}$ *and zeros in the remaining positions.*

4. Update $\hat{\theta}_j(k)$ *at time* k *via:*

$$\begin{aligned} \epsilon(k) &= v(k) - U(k)^T\hat{\theta}(k), \quad \textit{where } \hat{\theta}(k) = (\hat{\theta}_0(k), \hat{\theta}_1(k), ..., \hat{\theta}_{n-1}(k))^T \\ \hat{\theta}_j(k+1) &= \alpha^{1-g_j(k)}\hat{\theta}_j(k) + \mu\epsilon(k)g_j(k)u(k-j) \end{aligned}$$

5. Return to step 2.

Remark:

9. Choice of $\alpha = 1$ corresponds to freezing of the LMS estimates of those taps not detected during each sample interval. Because this leads to structural inconsistency, such a choice is not permitted in Algorithm 2.

10. One approach to modifying Algorithm 2 to enable estimation of a slowly time varying echo path involves:

(i) periodically reinitializing $f_j(k), r_j(k), x_j(k)$:

$f_j(k) = 0,\ r_j(k) = r_0,\ x_j(k) = 0,\ k = cT,\ c = 0, 1, ...,$ and T sufficiently large

(ii) LMS updating the j^{th} tap coefficient at time $k = cT + \bar{k},\ 1 \leq \bar{k} < T$ only if

$$x_j(k) > \log(\bar{k})$$

Figure 6.4 shows the results of simulations based on Algorithm 2 - plotted is the squared Euclidean norm of the residual parameter vector $(||\theta - \hat{\theta}(k)||_2)^2$ over number of sample intervals. The mean results of ten similar simulations are shown in both Figure 6.4a and 6.4b. The impulse response, θ, of the channel being estimated is that shown in Figure 6.3. In both Figure 6.4a and 6.4b, the input and disturbance signals are zero mean Gaussian signals of unity variance. In Figure 6.4a the disturbance signal is white while in Figure 6.4b the disturbance signal is generated by a first order autoregressive filter $H(z) = 1/(1 - 0.8z^{-1})$ and, consequently, is autocorrelated. Similar results were achieved with nonGaussian signals. For comparison purposes, Figure 6.5 shows the results of similar simulations based on the standard LMS algorithm in which all $n = 300$ taps are estimated. The simulation results indicate a considerable improvement in asymptotic performance of the LMS-active tap estimator over the standard LMS estimator. Furthermore, this improvement does not depend on the disturbance signal being white (or the input and disturbance signals being Gaussian).

Remark:

11. As remarked in Section 4, the LMS estimation algorithm presented above considers only the case of fixed μ - which results in enhanced asymptotic performance. To enhance transient performance instead, the proposed algorithm could be modified so that μ is made proportional to the number of active taps detected.

6.7 Conclusion

In this chapter we tackled the adverse effects of large filter parameter dimension on the performance of the LMS adaptive FIR filter. This problem is particularly relevant to echo cancellation since the impulse response length n of echo paths is typically moderately long to very long. We began by proposing, as a means of reducing the parameter dimension, an 'active' tap parametrization in which only $m < n$ taps in

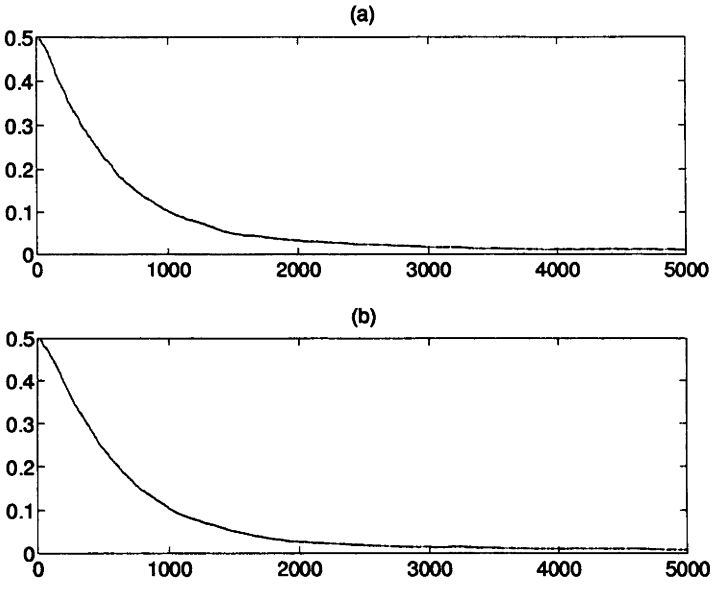


Figure 6.4: Performance over time of the LMS-active tap estimator of Algorithm 2. Plotted is $(||\theta - \hat{\theta}(k)||_2)^2$ over time, where θ is the impulse response vector of the channel of Figure 6.3. The input signal in both (a) and (b) is a zero mean Gaussian white process, while the disturbance signal is either (a) a zero mean Gaussian white process or (b) a zero mean Gaussian first order AR process with pole $p = 0.8$. The forgetting factor $\alpha = 0.9$ and update constant $\mu = 0.001$.

the adaptive FIR echo canceller were allowed to be nonzero. This parametrization was chosen due to the observation that echo path impulse responses typically show zero or 'inactive' regions.

Analyses carried out showed that, under the condition that the input signal is white, performance improvements in LMS estimation are achievable with this parametrization. A simple procedure, based on the Least Squares method, was then developed which, for sufficiently large N (the number of sample intervals), detects the correct number and position of active taps. Using this active tap detection procedure we proposed an LMS based estimation algorithm. Simulations indicated that such an "LMS-active tap detection" algorithm provides considerable performance improvement over the standard LMS algorithm.

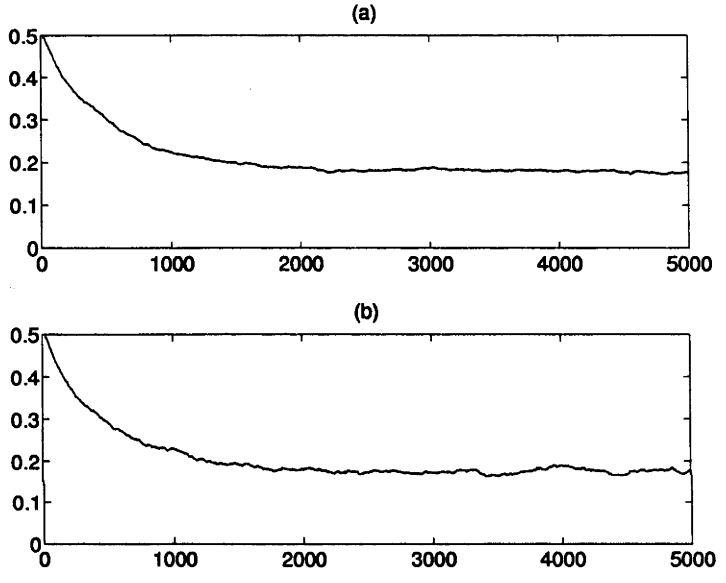


Figure 6.5: Performance over time of the standard LMS estimator. Plotted is $(||\theta - \hat{\theta}(k)||_2)^2$ over time, where θ is the impulse response vector of the channel of Figure 6.3. The input signal in both (a) and (b) is a zero mean Gaussian white process, while the disturbance signal is either (a) a zero mean Gaussian white process or (b) a zero mean Gaussian first order AR process with pole $p = 0.8$. The update constant $\mu = 0.001$.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

In this thesis we have focussed on echo cancellation in speech transmission telecommunication systems via the popular LMS adaptive FIR filter. This technique typically provides adequate suppression of circuit echoes, generated by impedance mismatching within central 4-wire loop telephone networks. However, poor performance is occasionally observed. The LMS adaptive FIR echo canceller is also used to suppress acoustic echoes which occur during hands free telephony and teleconferencing. High computational cost and poor transient performance, however, are common problems.

The performance problems in both of these applications has been linked to the correlation characteristics of the transmitted/driving signals. The large FIR filter parameter dimensions required in the latter application (to adequately model the 'long' acoustic echo paths) has been also linked to the observed poor transient performance. Motivated by this, the objectives of the work leading to this thesis were as follows.

1. Quantify the effects of signal correlation characteristics and filter dimension on performance of the LMS adaptive FIR filter.

2. Based on these results and the typical characteristics of the echo cancellation application, develop performance enhancing schemes for the LMS adaptive FIR echo canceller.

The first objective was tackled by conducting dynamical analyses on an open loop and a closed loop LMS adaptive FIR filter system, the former being representative of one end of an echo cancellation network, while the latter is representative of the full

network. In particular, the open loop system consisted of:

- the adaptive filter connected next to and in parallel with time invariant FIR modelled unknown channel of equal tap length;
- additive disturbance within the unknown channel;
- stationary, zero mean input and disturbance signals which were not cross correlated;
- positive definite input power spectrum;
- 'sufficiently small' LMS update stepsize, μ.

This setup ensured unbiased estimation of the unknown channel.

We began by developing a quantitative measure of the expected convergence rate of the LMS adaptive FIR filter to the unknown channel. We proposed the cost function:

$$C_e = \frac{Trace(R_n^{-1})}{2\mu n}$$

where R_n is the $n \times n$ autocorrelation matrix of the input signal and n is the tap length of the FIR filter. An increase in the cost function implied a decrease in the expected convergence rate. Analysis of the cost function C_e followed, with particular attention being given to input signals, such as speech, which are well modelled as autoregressive (AR) processes. The results are summarized below.

- In general, for a given μ,
  - the cost function increases monotonically with dimension for 'non-white' or coloured inputs and is independent of dimension when the input is 'white';
  - as dimension n increases, the cost function increases towards a finite value which is determined by and increases with input autocorrelation level.

- For autoregressive (AR) input signals, an explicit relationship is obtained between the cost function, the AR coefficients, filter dimension, n, and μ. The adverse effect of dimension tends to increase with input autocorrelation level.

- For the filter dimensions typical in echo cancellation, the cost function is of the order of 100 times greater with AR modelled speech input signals than with 'white' inputs of the same variance. This is due to the high autocorrelation level of speech.

- When μ is adjusted to maintain asymptotic performance, the cost function increases approximately linearly with dimension. The linear rate increases with increasing input autocorrelation level.

The closed loop adaptive filter system consisted of:

- an LMS adaptive FIR filter and FIR modelled unknown channel at each end of the loop;
- the tap lengths of the adaptive filters and unknown channels were equal;
- the loop driven by signals entering from within the unknown channels;
- driving signals stationary and zero mean;
- sufficiently small μ.

With reference to echo cancellation, the driving signals represented the subscriber signals. Semi-formal to rigorous analyses were conducted. The results are summarized below.

• For given filter dimension/tap length and μ, the transient and asymptotic performance of both adaptive filters deteriorates with increasing autocorrelation and cross correlation levels of the subscriber signals. An increase in dimension accentuates this effect.

• For given dimension and μ, asymptotic and transient performance improves as the subscriber signals become 'whiter'.

• When the channels connecting the two adaptive channel pairs impose a sufficiently long delay and/or the subscriber signals are sufficiently 'white', the dynamics simplify to those of a pair of open loop systems.

These results together with the properties of speech transmission echo cancellation telecommunication systems - highly autocorrelated input/subscriber signals and large filter dimensions - indicate convincingly that considerable performance improvements of such systems can be achieved by employing schemes which:

• whiten the subscriber signals, or, if the telecommunication channels impose a sufficiently large delay, whiten the input signals to the echo cancellers;

• reduce the filter dimension, *i.e.* the number of adaptive parameters of the FIR echo canceller.

Following the first approach we proposed a scheme which uses digital scramblers at each end of the network to ensure the subscriber signals are 'white'. Under the assumption of a fully digital network, the scheme is shown to provide greatly enhanced performance. A second scheme considered makes use of the autoregressive (AR) nature of speech signals. It involves using estimates of the AR filter of the echo canceller input signal to reduce input signal autocorrelation. We heuristically analysed exist-

ing versions of this AR based scheme as well as proposed and analysed modifications. The analyses indicated that the schemes, under a variety of conditions, essentially 'whiten' the input signal. Simulations supported the analytical results.

The second approach, that of reducing filter dimension, was tackled by proposing a parametrization for the n-tap FIR based echo canceller in which only $m < n$ taps are allowed to be nonzero or 'active'. The remaining $n - m$ are set to zero. This parametrization was based on the observation that the (discrete time domain) impulse responses of echo paths typically contain regions of inactive or zero taps. We showed that when the input signal is white, such a parametrization can lead to improved transient and/or asymptotic performance of the adaptive filter. Assuming a 'white' input signal we then:

- developed, via a least squares approach, a measure of the activity/inactivity of each tap of an FIR modelled echo path;
- developed, using this activity measure, a low computational cost algorithm for determining the lag position of the 'active' or nonzero taps of an echo path;
- proposed a modified LMS algorithm which uses the 'active' tap detection algorithm to reduce the echo canceller parameter dimension and, hence provide improved asymptotic and/or transient performance - importantly, the modification introduces only a minor increase in computation.

7.2 Future Directions of Research

7.2.1 Dynamical Analyses

Open Loop Analyses

- Extend the open loop analytical results of the thesis, at least qualitatively, to allow for some degree of nonstationarity within the input/disturbance signals and the unknown channel. This might involve first developing a suitable signal model and possibly a new performance cost function.
- Determine the performance of the LMS adaptive FIR filter with nonstationary speech driving signals, using as a reference, the performance with 'white' driving signals. This may be based solely on simulation results, or, more ambitiously, on the analysis of a performance cost function.

Closed Loop Analyses

- Apply the Lyapunov technique, which enabled rigorous analysis of a simplified closed loop system consisting of single tap echo paths/cancellers, single sample delay transmission channels and first order autoregressive (AR) subscriber signals, to the cases of:
  larger delay systems;
  higher order AR subscriber signals.

- Strengthen the semi-quantitative results obtained for closed loop systems of arbitrary echo path/canceller dimension (tap length). In particular, obtain a quantitative relationship relating performance, dimension and subscriber signal correlation levels. This will require a performance cost function to be developed.

- Carry out a simulation study of closed loop systems having arbitrary dimensions and channel delays and driven by speech subscriber signals. Compare to 'white' subscriber signals.

- Show more conclusively, through simulation and/or theoretical analysis, that the dynamics of the closed loop system simplify to a pair of decoupled open loop systems when the channel delay is sufficiently longer than the cross and autocorrelation lengths of the subscriber signals.

7.2.2 Signal Conditioning Schemes

Scrambler Scheme
Carry out simulation studies to determine the performance improvements achievable with the scrambler scheme when the subscriber signals are speech.

AR Whitening Schemes Compare the performance benefits obtained with the six AR whitening schemes presented using more realistic simulation conditions, such as (i) speech input and disturbance signals and (ii) measured echo path impulse responses.

7.2.3 Dimension Reduced LMS/FIR Estimation

- The proposed structurally consistent cost function - which lead to the active tap criterion - involved replacing the LS cost function $V_N = O(N)$ by N. The use of other possible $O(N)$ replacements such as those depending on the variance of the disturbance and/or input signals (*e.g.* $\sigma_s^2 N$, $\sigma_s^2 N/\sigma_u^2$) should be explored.

- Determine the deterioration in performance of the active tap detection algorithm when the 'white' input is replaced by an approximately 'white' input. This will require the development of a suitable signal model as well as a performance cost function.
- Explore the effect of the disturbance level on the transient and asymptotic performance of the active tap detection algorithm.
- Explore, more fully, extensions to the proposed active tap LMS algorithm which enable tracking of time varying echo paths.

7.2.4 Nonlinear Effects in Channels

Determine sources of nonlinearity in speech transmission echo cancellation networks. Examples are A/D (analogue to digital) and D/A conversion, the loudspeaker and/or microphone in acoustic systems. Quantify the effects of these nonlinearities and, if detrimental, develop methods which reduce the effects.

Appendix A

Preliminary Concepts

A.1 Order Function and O(.) Function

A function $\delta(\mu)$ is called an order function if $\delta(\mu)$ is continuous and positive in an interval $(0, \mu^+]$ and if $\lim_{\mu \to 0} \delta(\mu)$ exists [78].

The $O(.)$ function is defined as follows [78]. Consider two order functions δ_1 and δ_2.

If there exists a constant, C, such that:

$$\frac{|\delta_2(\mu)|}{|\delta_1(\mu)|} \leq C, \quad \text{for } \mu \to 0$$

then

$$\delta_2(\mu) = O(\delta_1(\mu)) \text{ for } \mu \to 0$$

A.2 Time Scale

Consider a function $\phi_\mu(k)$, parametrized by μ, and an order function $\delta(\mu)$. If the equality:

$$\phi_\mu(k) = O(\delta(\mu)) \text{ as } \mu \to 0$$

is valid for $0 \leq k < L/\mu$, with L being a constant independent of μ, then the equality is said to hold on the timescale $1/\mu$ [78].

A.3 Lipschitz Continuity

Consider the vectorfield $f(k,x)$ with:

$$k \geq 0, \quad x \in D \subset R^n, \quad f : N \times R^n \rightarrow R^n$$

If there exists some λ, which is independent of k , x, such that for all $x_1, x_2 \in D$

$$||f(k,x_1) - f(k,x_2)|| \leq \lambda ||x_1 - x_2||$$

Then, $f(k,x)$ is Lipschitz continuous in $x \in D$, uniformly in k [78].

A.4 Uniform Contraction

Consider the difference equation [94]:

$$x_{k+1} = A(x_k)x_k + b(x_k), \qquad x_{k=0} = x_0 \tag{A.1}$$

where $x, x_0 \in D \subset R^n$, A is an $n \times n$ matrix, b is an $n \times 1$ matrix and D is an open set containing the origin.

Suppose the induced l_p norm of $A(x)$ is bounded by:

$$||A(x)||_{ip} \leq 1 - \epsilon \tag{A.2}$$

$$\text{for all } x \in D_0 = \{x \in D : \ ||x||_p \leq \chi(\text{constant})\} \tag{A.3}$$

$$\text{for some } 0 < \epsilon < 1 \text{ which is independent of } k, x,$$

(a) If

$$b(x_k) = 0 \ \ \forall \ x_k \tag{A.4}$$

then (A.1) is a uniform contraction to zero on $x \in D_0$.

This implies that:

(i) there exists a unique equilibrium at $x = 0$ for all $x_0 \in D_0$

(ii) the solution $x \equiv 0$ is exponentially stable with D_0 being a subset of the domain of exponential attraction.

Note that $||x_k - 0||_p$ is bounded by:

$$||x_k - 0||_p \leq (1 - \epsilon)^k ||x_0||_p \tag{A.5}$$

and that this bound will converge more quickly to zero as ϵ increases towards unity.

(b) If

$$||b(x_k)||_p \leq b(\text{constant}) \ \forall \ x_k \tag{A.6}$$

$$b/\epsilon \leq \chi(\text{constant}) \tag{A.7}$$

then (A.1) is a uniform contraction on $x \in D_0$ to the l_p ball, B:

$$||x||_p = b/\epsilon \tag{A.8}$$

This implies that a unique solution does not necessarily exist, but that for all $x_0 \in D_0$ the solution x_k will remain stable.

Note that as ϵ increases towards unity:

(i) the l_p ball B will reduce in size

(ii) if $||x_0||_p \geq b/\epsilon$, then the solution x_k will converge more quickly to within the l_p ball B since $||x_k||_p - b/\epsilon \leq (1-\epsilon)^k ||x_0||_p$.

Appendix B

The Averaging Method

Consider the difference equation [78]:

$$x_{k+1} = x_k - \mu f(k, x_k), \quad x_{k=0} = x_0 \tag{B.1}$$

where $x, x_0 \in D \subset R^n$, $k \geq 0$, $f : N \times R^n \to R^n$.

Suppose the limit:

$$\bar{f}(z) = \lim_{M \to \infty} 1/M \sum_{k=k_0}^{M-1+k_0} f(k, z) \tag{B.2}$$

where $x = z$ is held constant, is well defined and is independent of k_0.

Consider the difference equation:

$$\bar{x}_{k+1} = \bar{x}_k - \mu \bar{f}(\bar{x}_k), \quad \bar{x}_{k=0} = x_0 \tag{B.3}$$

where $\bar{x} \in D \subset R^n$, and $\bar{f}$ is as defined in (B.2).

(a) If $f(k, x_k)$ is Lipschitz continuous in $x \in D$, uniformly in k, then, $\bar{f}(\bar{x}_k)$ is Lipschitz continuous in $\bar{x} \in D$ and

$$|x_k - \bar{x}_k| = O(\delta^{1/2}(\mu)), \quad \text{on the time scale } 1/\mu \tag{B.4}$$

$$\text{where } \delta(\mu) = \sup_{k_0} \sup_{x \in D} \sup_{k \in [0, L/\mu)} \mu \left| \sum_{m=k_0}^{k+k_0} f(m, x) - \bar{f}(x) \right| \tag{B.5}$$

(b) If the condition in (a) is satisfied and $f(k, x_k)$ has Lipschitz continuous first order partial derivatives with respect to x in D then:

$$|x_k - \bar{x}_k| = O(\delta(\mu)), \quad \text{on the time scale } 1/\mu \tag{B.6}$$

(c) If the condition in (a) or (b) is satisfied and $\bar{x} \equiv 0$ (with $x_0 = 0$) is an exponentially stable solution of (B.3) with domain of exponential attraction $D_0 \subset D$ then, for

$x_0 \in D_0$, the bound on the approximation in (B.4) or (B.6), respectively, holds for all time, that is, the time scale of the approximation is extended to infinity.

Appendix C

Recursive Least Squares

The Recursive Least Squares (RLS) method is a popular approach for obtaining an estimate $\hat{\theta}$ of an unknown parameter vector θ of a linear regression:

$$v(t) = \theta^T \phi(t) + e(t) \tag{C.1}$$

where

$v(t)$ is an observed variable;

$\phi(t)$ is an observed vector and is called the regression vector; and

$e(t)$ is unobserved noise resulting from additive disturbance and/or measurement errors.

As an example, the output of an FIR modelled system is often described by a linear regression:

$$v(t) = \theta_0 u(t) + \theta_1 u(t-1) + ... + \theta_{n-1} u(t-n+1) + e(t) \tag{C.2}$$

for which

$$\theta = (\theta_0 \ \theta_1 \ ... \ \theta_{n-1})^T, \quad \phi(t) = (u(t) \ u(t-1) \ ... \ u(t-n+1))^T \tag{C.3}$$

Other examples are outputs from AR modelled and ARX modelled systems.

The Least Squares approach is based on the idea of choosing as an estimate, that $\hat{\theta}$ which minimizes the cost function:

$$V_N(\hat{\theta}) = \frac{1}{N} \sum_{t=1}^{N} \gamma(N,t)[v(t) - \hat{\theta}^T \phi(t)]^2 \tag{C.4}$$

where $\gamma(N,t)$ is a weighting function.

We will consider here the exponential weighting function (see *e.g* [95] for a more general approach):

$$\gamma(N,t) = (1-\alpha)^{N-t}, \quad 0 \leq \alpha < 1 \tag{C.5}$$

The criterion $V_N(\hat{\theta})$ is quadratic in $\hat{\theta}$. We can obtain the Least Squares estimate by differentiating (C.4) with respect to $\hat{\theta}$,

$$\frac{dV_N(\hat{\theta})}{d\hat{\theta}} = 2\sum_{t=1}^{N}(1-\alpha)^{N-t}\left[-\phi(t)v(t) + \phi(t)\phi(t)^T\hat{\theta}\right], \tag{C.6}$$

equating the result to zero and then solving for $\hat{\theta}$:

$$\hat{\theta} \triangleq \hat{\theta}(N) = R(N)^{-1}f(N) \tag{C.7}$$

$$R(N) = \sum_{t=1}^{N}(1-\alpha)^{N-t}\phi(t)\phi(t)^T \tag{C.8}$$

$$f(N) = \sum_{t=1}^{N}(1-\alpha)^{N-t}\phi(t)v(t) \tag{C.9}$$

From (C.7)-(C.9) we have:

$$(1-\alpha)R(t-1) = R(t) - \phi(t)\phi^T(t) \tag{C.10}$$

$$\begin{aligned}(1-\alpha)f(t-1) &= (1-\alpha)R(t-1)\hat{\theta}(t-1) \\ &= R(t)\hat{\theta}(t-1) - \phi(t)\phi^T(t)\hat{\theta}(t-1)\end{aligned} \tag{C.11}$$

Thus

$$\begin{aligned}\hat{\theta}(N) &= R(N)^{-1}\left[(1-\alpha)f(N-1) + \phi(N)v(N)\right] \\ &= R(N)^{-1}\left[R(N)\hat{\theta}(N-1) + \phi(N)\{-\phi^T(N)\hat{\theta}(N-1) + v(N)\}\right] \\ &= \hat{\theta}(N-1) + R(N)^{-1}\phi(N)\{v(N) - \phi^T(N)\hat{\theta}(N-1)\}\end{aligned} \tag{C.12}$$

Equation (C.12) together with

$$R(N) = (1-\alpha)R(N-1) + \phi(N)\phi^T(N) \tag{C.13}$$

is a version of the exponentially weighted RLS algorithm.

The computational complexity, resulting from the need to obtain the inverse of the matrix $R(N)$ every sampling interval, can be avoided by using the matrix inversion lemma [95]:

Lemma 3 *Let A, B, C and D be matrices of compatible dimensions, so that the product BCD and the sum $A + BCD$ exist. Then*

$$[A + BCD]^{-1} = A^{-1} - A^{-1}B[DA^{-1}B + C^{-1}]^{-1}DA^{-1} \tag{C.14}$$

Following common practice we define $P(N) \triangleq R^{-1}(N)$. Applying (C.14) to (C.13) with

$$A = (1-\alpha)R(N-1) = (1-\alpha)P(N-1)^{-1},\ B = \phi(N),\ C = 1,\ D = \phi^T(N)$$

leads to

$$\begin{aligned} P(N) &= \frac{P(N-1)}{1-\alpha} \\ &\quad -\frac{P(N-1)}{1-\alpha}\phi(N)\left[\phi^T(N)\frac{P(N-1)}{1-\alpha}\phi(N)+1\right]^{-1}\phi^T(N)\frac{P(N-1)}{1-\alpha} \\ &= \frac{1}{1-\alpha}\left[P(N-1) - \frac{P(N-1)\phi(N)\phi^T(N)P(N-1)}{\phi^T(N)P(N-1)\phi(N)+(1-\alpha)}\right] \end{aligned} \tag{C.15}$$

Using (C.15) we also obtain:

$$\begin{aligned} P(N)\phi(N) &= \frac{1}{1-\alpha}\left[P(N-1)\phi(N) - \frac{P(N-1)\phi(N)\phi^T(N)P(N-1)\phi(N)}{\phi^T(N)P(N-1)\phi(N)+(1-\alpha)}\right] \\ &= \frac{P(N-1)\phi(N)}{\phi^T(N)P(N-1)\phi(N)+(1-\alpha)} \end{aligned} \tag{C.16}$$

Combining (C.12), (C.16) and (C.15) leads to the standard RLS algorithm.

Algorithm 3 Initialization: *Choose $P(0)$ and $\hat{\theta}(0)$.*

Recursion:

$$\hat{\theta}(N) = \hat{\theta}(N-1) + L(N)\{v(N) - \phi^T(N)\hat{\theta}(N-1)\} \tag{C.17}$$

$$L(N) = \frac{P(N-1)\phi(N)}{\phi^T(N)P(N-1)\phi(N)+(1-\alpha)} \tag{C.18}$$

$$P(N) = \frac{1}{1-\alpha}\left[P(N-1) - \frac{P(N-1)\phi(N)\phi^T(N)P(N-1)}{\phi^T(N)P(N-1)\phi(N)+(1-\alpha)}\right] \tag{C.19}$$

Remark:

A common choice [95] of initial values is:

$$P(0) = K\ I,\ \ \hat{\theta}(0) = 0$$

where I is the identity matrix and K is a large constant.

Appendix D

Autoregressive Filter Estimation - the Levinson-Durbin Algorithm

In this appendix we present the Levinson-Durbin Algorithm which is a recursive procedure, based on the Yule-Walker Method, for obtaining an estimate of the autoregressive (AR) filter of an autoregressive signal [85]. The technique uses the autocorrelation values of the signal and, therefore, assumes that the signal is stationary. For nonstationary AR signals such as speech, the signal is assumed to be stationary over a given period of time. For speech, a period of 20ms is often suggested [20].

Algorithm 4 Levinson-Durbin Algorithm:
Let r_j, $j = 0, 1, ...$ be the autocorrelation function of an AR modelled signal. Let p be the order of the AR filter to be estimated.

Initialization: $a_0^{(0)} = 1$ *and* $D_0 = r_0$.

Recursion: For $m = 0, 1, ..., p-1$, *do:*

$$\beta_m = \sum_{t=0}^{m} a_t^{(m)} r_{m+1-t} \tag{D.1}$$

$$K_{m+1} = -\frac{\beta_m}{D_m} \tag{D.2}$$

$$a_0^{(m+1)} = 1 \tag{D.3}$$

$$a_t^{(m+1)} = a_t^{(m)} + K_{m+1} a_{m+1-t}^{(m)}, \quad \text{for } t = 1, 2, ..., m+1 \tag{D.4}$$

$$D_{m+1} = (1 - K_{m+1}^2) D_m \tag{D.5}$$

The coefficients of the AR estimate are given by:

$$a_0^{(p)},\ a_1^{(p)},\ ...,\ a_p^{(p)}$$

In the case of an AR signal which is non-stationary but is assumed to be stationary over L sample intervals, the autocorrelation values r_j are replaced by appropriate short term autocorrelation estimates, such as:

$$r_j(k_0, L) = \frac{1}{L} \sum_{k=k_0+j+1}^{k_0+L} u(k)u(k-j)w(k), \quad j = 0, 1, 2, ..., p$$

where $w(k)$ is associated with a windowing function. The relatively high computational complexity of this approach could be avoided by using, instead, an iterative procedure such as [86],

$$\begin{aligned} \hat{\sigma}_u^2(k) &= \alpha\hat{\sigma}_u^2(k-1) + (1-\alpha)u(k)^2 \\ r_j(k+1) &= \alpha r_j(k) + \frac{u(k)u(k-j)}{\hat{\sigma}_u^2(k)}, \quad j = 0, 1, 2, ...p \end{aligned} \tag{D.6}$$

The number of multiplications per sampling interval required in this autocorrelation scheme is only $2(p+1)$.

The complexity of the Levinson-Durbin recursion is approximately $4p + p^2$. If a new AR estimate is obtained every M sample intervals and we assume the complexity to be spread evenly over the M sample block then:

The total complexity of the Levinson-Durbin per sampling interval $\approx 2p+(4p+p^2)/M$

where we have assumed the autocorrelation estimate is obtained via (D.6).

Appendix E

Removing Measurement Noise Effects - Sparse Channel Impulse Response Estimation

In this appendix we present a procedure for removing the effects of measurement noise from estimated time domain channel impulse responses. The procedure relies on the true impulse response having a sufficiently large number of zero taps and, consequently, its applicability is restricted to channels having, for example, a sparse structure. This procedure may be viewed as an off-line scheme for active tap detection of sparse channels.

Consider an unbiased estimate of the discrete time domain impulse response of a sparse channel. Generally, the presence of measurement noise or disturbance causes the tap coefficient estimate of each of the zero taps of the sparse channel to be nonzero. If we assume the input is white, then the discussion of Section 6.5 suggests that asymptotically (at least for LS,LMS estimates) the zero tap estimates form a zero mean i.i.d. Gaussian distribution:

$$\{\hat{\theta}_i\} \sim N(0, \sigma^2),\ i.i.d.,\ \text{where}\ \ \theta_i = 0 \tag{E.1}$$

Under the validity of (E.1), we use the following result, which is inspired by the work of Donoho cited in [99], to develop a procedure for removing the effects of the noise, or, equivalently, for determining which taps are zero.

Result 12 *Let* $\{z_i\} \sim N(0, \sigma^2), i.i.d.$. *Define the event* $A_M \triangleq \{\sup_{i \leq M} |z_i| \leq \sigma\sqrt{2logM}\}$ *Then,* $Prob(A_M) \rightarrow 1$ *as* $M \rightarrow \infty$.

A first glance suggests that, in order to use the threshold $\sigma\sqrt{2logM}$ to determine which taps are zero, a-priori knowledge of the indices i of the zero taps is required. This requirement is avoided for sparse channels by applying the following iterative procedure.

Algorithm 5 *1. Include initially the indices of all n tap estimates $\{\hat{\theta}_i\}$ in the set S of zero taps. Set $M = n$.*

2. Determine rms value σ_S of the estimates of the taps in set S.

3. Determine the indices i of those taps, the estimated coefficients of which satisfy:

$$|\hat{\theta}_i| \leq \sigma_S\sqrt{2\log M}. \tag{E.2}$$

Include only these indices in the set S. Set M =number of elements of set S.

4. Repeat steps 2 to 3 a given number of times or, alternatively, until the difference in σ_S from one iteration to the next has dropped below a given value.

Appendix F

Chapter 3 Proofs

F.1 Proof of Theorem 1

The proof of Theorem 1 follows from proof that $Tr(R_n^{-1})/n$ is a nondecreasing function of n and, furthermore, is a nonincreasing function of n only when the input is white, or, equivalently:

$$R_n = I\sigma_u^2, \text{ where I is the } n \times n \text{ identity matrix}$$

To begin, we note that R_n is an n^{th} order positive definite symmetric Toeplitz matrix. Let R_{n+1} be a $n+1^{th}$ order symmetric Toeplitz matrix, with first row:

$$r_0,\ r_1,\ r_2,\ \ldots,\ r_n.$$

Denote the inverse of R_{n+1} by Q_{n+1} the $(i,j)^{th}$ element of which is $q_{i,j}$. Following the procedure outlined in [80] we let

$$b_n(k,1) = q_{k,1}/q_{1,1} \text{ for } k = 1,\ 2,\ \ldots,\ n+1 \tag{F.1}$$

$$\rho_{n+1} = 1/q_{1,1} \tag{F.2}$$

After an extension of the Levinson algorithm, as given in [80], for the inverse of a symmetric Toeplitz matrix, we have:

$$Tr(Q_{n+1}) = \frac{1}{\rho_{n+1}}[(n+1)+(n-1)b_{n+1}^2(2,1)+(n-3)b_{n+1}^2(3,1)+\ldots+(1-n)b_{n+1}^2(n+1,1)] \tag{F.3}$$

An extension of the results given in [80] for the Cholesky decomposition, via the Levinson algorithm, of the inverse of a symmetric Toeplitz matrix leads to

$$Tr(Q_{n+1}) = \frac{1}{\rho_n}(1 + b_n^2(1) + b_n^2(2) + \ldots + b_n^2(n))$$

$$+ \quad \frac{1}{\rho_{n-1}}(1 + b_{n-1}^2(1) + \ \ldots \ + b_{n-1}^2(n-1))$$
$$+ \quad .....$$
$$+ \quad \frac{1}{\rho_1}(1 + b_1^2(1)) \ + \ \frac{1}{\rho_0} \tag{F.4}$$

Combining (F.3) and (F.4) yields:

$$\frac{Tr(Q_{n+1})}{n+1} - \frac{Tr(Q_n)}{n} = \frac{2}{n(n+1)}[\frac{b_{n+1}^2(2,1) + 2b_{n+1}^2(3,1) + ... + nb_{n+1}^2(n+1,1)}{\rho_{n+1}}] \tag{F.5}$$

As indicated by the Levinson algorithm given in [80], the following properties hold for symmetric Toeplitz matrices:

$$\begin{aligned} \rho_{n+1} &= \rho_n(1 - b_{n+1}^2(n+1,1)) \\ det(R_{n+1}) &= \Pi_{i=1}^{n+1}\rho_i \end{aligned}$$

where $det(.)$ denotes determinant. Combining these properties with the fact that for positive definite matrices, $det(R_n) > 0$ leads to $\rho_n > 0$. This, together with (F.5), leads to:

$$\frac{Tr(Q_{n+1}}{n+1} - \frac{Tr(Q_n)}{n} \geq 0 \tag{F.6}$$

with equality only if:

$$b_{n+1}^2(k,1) = 0 \quad \text{for } 2 \leq k \leq n+1 \tag{F.7}$$

Combining (F.7) with the following Levinson algorithm equation (from [80]) for symmetric Toeplitz matrices:

$$\begin{aligned} q_{j+1,k+1} &= [b_{n+1}(j+1,1)b_{n+1}(k+1,1) - b_{n+1}(n+2-j)b_{n+1}(n+2-k)]/\rho_{n+1} \\ &+ \ q_{j,k}, \quad 1 \leq j,k \leq n \end{aligned} \tag{F.8}$$

indicates that:

$q_{j,k} = q_{1,1} \quad j = k$
$= 0 \quad j \neq k$

Thus, (F.7) implies $Q_{n+1} = Iq_{1,1}$ (where $I = (n+1) \times (n+1)$ identity matrix) or equivalently $R_{n+1} = I/q_{1,1} = Ir_0$.

For the LMS/FIR system in question, $r_0 = \sigma_u^2$ and the proof is complete.

F.2 Proof of Theorem 2

According to Gray [59] we have the following result.

Lemma 4 *Consider a real sequence* $\{r_k\}$, $|k| = 0, 1, 2, \ldots$ *which is absolutely summable:*

$$\sum_{k=-\infty}^{\infty} |r_k| < \infty$$

and is symmetric with respesct to $k = 0 : r_k = r_{-k}$. *Let*

(a) $f(\omega)$ *be the Fourier transform of the sequence,* $\{r_k\}$*:*

$$f(\omega) = \lim_{n\to\infty} \sum_{k=-n}^{n} r_k e^{ik\omega}$$

(b) R_n *be the symmetric Toeplitz matrix of order n constructed from* $\{r_k\}$ *such that the first row is*

$$r_0, r_1, r_2, \cdots, r_{n-1}$$

If $f(\omega) > 0, \ \forall\, \omega$, *then*

$$\lim_{n\to\infty} \frac{1}{n} Tr(R_n^{-1}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f^{-1}(\omega) d\omega$$

Combining this result with the expression for $C_e(\infty)$ of (3.24) leads to (3.25) of Theorem 2.

F.3 Proof of Lemma 1

Let

$$\sigma^2 = 1/2\pi \int_{-\pi}^{\pi} \Phi_{uu}(\omega) d\omega \tag{F.9}$$

be the given signal power. Consider the power spectrum:

$$\Phi_{uu}(\omega) = \sigma^2 + \alpha\delta(\omega) \tag{F.10}$$

where α is a positive constant. According to (F.9) $\delta(\omega)$ must satisfy:

$$1/2\pi \int_{-\pi}^{\pi} \delta(\omega) d\omega = 0 \tag{F.11}$$

The extrema of

$$F(\alpha) \triangleq 1/2\pi \int_{-\pi}^{\pi} \Phi_{uu}^{-1}(\omega) d\omega$$

are given by $dF/d\alpha = 0$, that is:

$$1/2\pi \int_{-\pi}^{\pi} \frac{d}{d\alpha} [\frac{1}{\sigma^2 + \alpha\delta(\omega)}] d\omega = 1/2\pi \int_{-\pi}^{\pi} \frac{-\delta(\omega)}{[\sigma^2 + \alpha\delta(\omega)]^2} d\omega = 0 \tag{F.12}$$

Now let:

$$\begin{aligned}\Omega_1 &= \{\omega|\delta(\omega) > 0\} \\ \Omega_2 &= \{\omega|\delta(\omega) < 0\}\end{aligned}$$

So, the extrema of F satisfy:

$$\begin{aligned}0 &= 1/2\pi \int_{-\pi}^{\pi} \frac{-\delta(\omega)}{[\sigma^2 + \alpha\delta(\omega)]^2} d\omega \\ &= \frac{1}{2\pi\sigma^2}[\int_{\Omega_1} \frac{-\delta(\omega)}{[1 + \frac{\alpha}{\sigma^2}\delta(\omega)]^2} d\omega + \int_{\Omega_2} \frac{-\delta(\omega)}{[1 + \frac{\alpha}{\sigma^2}\delta(\omega)]^2} d\omega] \\ &\geq \int_{\Omega_1} -\delta(\omega)d\omega + \int_{\Omega_1} -\delta(\omega)d\omega \quad \text{with equality when } \alpha\delta(\omega) = 0 \ \forall\omega \quad \text{(F.13)}\end{aligned}$$

However, from (F.11), we have

$$\int_{\Omega_1} -\delta(\omega)d\omega + \int_{\Omega_1} -\delta(\omega)d\omega = 0$$

Thus, the only noncontradictory solution to (F.12) occurs when

$$\begin{aligned}\alpha\delta(\omega) &= 0, \ \forall\omega \\ \text{or } \Phi_{uu}(\omega) &= \sigma^2, \ \forall\omega\end{aligned}$$

F.4 Proof of Theorem 3

Consider first when $n \geq p+1$. Let

R_n be the $n \times n$ covariance matrix constructed from the signal u_k;

A be the $n \times 1$ AR vector $A = (1, a_1, a_2, ..., a_p, 0, 0, ..., 0)^T$;

S be the $n \times 1$ vector $S = (\sigma^2, 0, 0, ..., 0)^T$.

As shown in [80] for $n \geq p+1$:

$$R_n A = S$$

Thus, assuming R_n is invertible then

$$A = R_n^{-1} S$$

It follows that, if $q_{k,1}(n)$ is the $(k,1)^{th}$ element of R_n^{-1}, then

$$\begin{aligned}q_{k,1}(n) = A(k)/\sigma^2 &= 1/\sigma^2 \quad k = 1 \\ &= a_k/\sigma^2 \quad 2 \leq k \leq p+1 \\ &= 0 \quad p+2 \leq k \leq n\end{aligned}$$

By combining the above set of equalities with (F.3), and noting that $\rho_n = \sigma^2$, $b_n(k,1) = A(k)$, then we obtain, for $n \geq p+1$:

$$\frac{Tr(R_n^{-1})}{n} = \frac{1}{\sigma^2}[1 + (1 - \frac{2}{n})a_1^2 + (1 - \frac{4}{n})a_2^2 + ... + (1 - \frac{2p}{n})a_p^2] \qquad \text{(F.14)}$$

By using the result of (F.14) together with (F.5) and (F.1) then it can be easily verified that the result of (F.14) also holds for $n = p$. The result of Theorem 3 follows from (F.14) and (3.23).

F.5 Proof of Theorem 4

Let $0 \leq k < L/\mu$ and

$$\begin{aligned} A_{k+1}(k_0) &\triangleq (I-\mu R(k+k_0))(I-\mu R(k-1+k_0))...(I-\mu R(k_0)) \\ &\quad -(I-\mu R)^{k+1} \qquad \text{(F.15)} \\ B_{k+1}(k_0) &\triangleq [(I-\mu R(k+k_0))(I-\mu R(k-1+k_0))...(I-\mu R(1+k_0))\mu P(k_0) \\ &+ (I-\mu R(k+k_0))(I-\mu R(k-1+k_0))...(I-\mu R(2+k_0))\mu P(1+k_0) \\ &+ ..... \\ &+ (I-\mu R(k+k_0))\mu P(k-1+k_0) \\ &+ \mu P(k+k_0)] \qquad \text{(F.16)} \end{aligned}$$

We begin by showing that

$$A_{k+1+mL/\mu} \triangleq A_{k+1+mL/\mu}(0),\ B_{k+1+mL/\mu} \triangleq B_{k+1+mL/\mu}(0),\ m \geq 0$$

can be written in terms of the $1/\mu$ time scale terms $A_k(iL/\mu)$, $B_k(iL/\mu)$, $i \leq m$.

Consider the $1/\mu$ scale time interval $mL/\mu \leq mL/\mu + k \leq (m+1)L/\mu$. The difference between the original and averaged parameter vectors in this time interval is given by:

$$\begin{aligned} \Delta\tilde{\theta}_{mL/\mu+k+1} &\triangleq \tilde{\theta}_{mL/\mu+k+1} - \tilde{\theta}^{av}_{mL/\mu+k+1} \\ &= A_{k+1}(mL/\mu)\tilde{\theta}_{mL/\mu} + (I-\mu R)^{k+1}(\tilde{\theta}_{mL/\mu} - \tilde{\theta}^{av}_{mL/\mu}) + B_{k+1}(mL/\mu) \\ &= A_{k+1}(mL/\mu)\tilde{\theta}^{av}_{mL/\mu} \qquad \text{(F.17)} \\ &+ [A_{k+1}(mL/\mu) + (I-\mu R)^{k+1}](\tilde{\theta}_{mL/\mu} - \tilde{\theta}^{av}_{mL/\mu}) + B_{k+1}(mL/\mu) \end{aligned}$$

It follows from (F.17) that:

$$\begin{aligned} \Delta\tilde{\theta}_{mL/\mu} &= A_{L/\mu}((m-1)L/\mu)(I-\mu R)^{(m-1)L/\mu}\tilde{\theta}_0 + B_{L/\mu}((m-1)L/\mu) \\ &+ [(I-\mu R)^{L/\mu} + A_{L/\mu}((m-1)L/\mu)](\tilde{\theta}_{(m-1)L/\mu} - \tilde{\theta}^{av}_{(m-1)L/\mu}) \\ &= \sum_{j=0}^{m-1} \left[\Pi_{i=j+1}^{m-1}\left((I-\mu R)^{L/\mu} + A_{L/\mu}(iL/\mu)\right)\right] \\ &\times \left[A_{L/\mu}(jL/\mu)(I-\mu R)^{jL/\mu}\theta + B_{L/\mu}(jL/\mu)\right] \qquad \text{(F.18)} \end{aligned}$$

Substitution of (F.18) into (F.17) then leads to:

$$\begin{aligned}
\Delta\tilde{\theta}_{mL/\mu+k+1} &= \{A_{k+1}(mL/\mu)(I-\mu R)^{mL/\mu} + \left[(I-\mu R)^{k+1} + A_{k+1}(mL/\mu)\right] \\
&\times \left[\sum_{j=0}^{m-1}\left[\Pi_{i=j+1}^{m-1}\left((I-\mu R)^{L/\mu} + A_{L/\mu}(iL/\mu)\right)\right]\right. \\
&\times \left.\left[A_{L/\mu}(jL/\mu)(I-\mu R)^{jL/\mu}\right]\right]\}\theta \qquad \text{(F.19)} \\
&+ \{B_{k+1}(mL/\mu) + \left[(I-\mu R)^{k+1} + A_{k+1}(mL/\mu)\right] \\
&\times \left[\sum_{j=0}^{m-1}\left[\Pi_{i=j+1}^{m-1}\left((I-\mu R)^{L/\mu} + A_{L/\mu}(iL/\mu)\right)\right] B_{L/\mu}(jL/\mu)\right]\}
\end{aligned}$$

Comparing (F.19) with (3.56) implies that the first bracketted '{ }' term of (F.19) is

$$A_{mL/\mu+k+1}(0) = A_{mL/\mu+k+1}$$

and that the second bracketted '{ }' term is

$$B_{mL/\mu+k+1}(0) = B_{mL/\mu+k+1}.$$

Consider now $||A_{mL/\mu+k+1}\theta||_2$. Combining properties (iv) and (v) of the ensemble leads to

$$E_\theta[||A_{mL/\mu+k+1}\theta||_2] = ||A_{mL/\mu+k+1}||_{HS}||\theta||_2 \qquad \text{(F.20)}$$

By applying the triangular inequality together with the following inequalities:

$$||AB||_{i2} \leq ||A||_{i2}||B||_{i2}, \quad \text{where } A, B \in \mathcal{R}^{n\times n} \qquad \text{(F.21)}$$

$$||AB||_{HS} \leq ||A||_{i2}||B||_{HS}, \; ||A||_{HS}||B||_{i2}, \quad \text{where } A, B \in \mathcal{R}^{n\times n} \qquad \text{(F.22)}$$

we obtain:

$$\begin{aligned}
||A_{mL/\mu+k+1}||_{HS} &\leq ||A_{k+1}(mL/\mu)||_{HS}||(I-\mu R)^{mL/\mu}||_{i2} \\
&+ [1 + ||A_{k+1}(mL/\mu)||_{i2}] \\
&\times [\sum_{j=0}^{m-1}\left[\Pi_{i=j+1}^{m-1}||(I-\mu R)^{L/\mu}||_{i2} + ||A_{L/\mu}(iL/\mu)||_{i2}\right] \\
&\times ||(I-\mu R)^{jL/\mu}||_{i2}||A_{L/\mu}(jL/\mu)||_{HS}||] \qquad \text{(F.23)} \\
&\leq ||A_{k+1}(mL/\mu)||_{HS}||(I-\mu R)^{mL/\mu}||_{i2} \\
&+ [1 + \sup_{k_0}\sup_{k\in[0,L/\mu)}||A_k(k_0)||_{i2}] \\
&\times [\sum_{j=0}^{m-1}\left[||(I-\mu R)^{L/\mu}||_{i2} + \sup_{k_0}\sup_{k\in[0,L/\mu)}||A_k(k_0)||_{i2}\right]^{m-1-j}
\end{aligned}$$

$$
\begin{aligned}
&\times \quad ||(I-\mu R)^{jL/\mu}||_{i2}||A_{L/\mu}(jL/\mu)||_{HS}||] \\
&\leq \quad ||A_{k+1}(mL/\mu)||_{HS}||(I-\mu R)^{mL/\mu}||_{i2} \qquad \text{(F.24)} \\
&+ \quad \left[2-||(I-\mu R)^{L/\mu}||_{i2}\right] \\
&\times \quad \left\{\sum_{j=0}^{m-1}||(I-\mu R)^{jL/\mu}||_{i2}||A_{L/\mu}(jL/\mu)||_{HS}||\right\}
\end{aligned}
$$

where in the last step we applied the condition of (3.63).

Using the fact that,

$$||(I-\mu R)^{jL/\mu}||_{i2} = (1-\mu\lambda_{min}(R))^{jL/\mu}$$

and the ergodic property of the ensemble, then it follows that $\forall m \geq 0$

$$
\begin{aligned}
E[||A_{mL/\mu+k+1}||_{HS}] \quad &\leq \quad \frac{2-(1-\mu\lambda_{min}(R))^{L/\mu}-(1-\mu\lambda_{min}(R))^{mL/\mu}}{1-(1-\mu\lambda_{min}(R))^{L/\mu}} \\
&\times \quad \sup_{k\in[0,L/\mu)} E[||A_k||_{HS}] \\
&\leq \quad \frac{2-(1-\mu\lambda_{min}(R))^{L/\mu}}{1-(1-\mu\lambda_{min}(R))^{L/\mu}} \sup_{k\in[0,L/\mu)} E[||A_k||_{HS}||] \qquad \text{(F.25)}
\end{aligned}
$$

The proof of (3.64) is completed by examining the $1/\mu$ time scale term

$$\sup_{k\in[0,L/\mu)} E[||A_k||_{HS}]$$

The RHS of (3.58) can be expanded as in (3.60). We make the following denotations to simplify notation:

$$A_{1,k} \triangleq \sum_{i=0}^{k-1}(R(i)-R) \qquad \text{(F.26)}$$

$$A_{2,k} \triangleq \sum_{i=0}^{k-1}\sum_{j=i+1}^{k-2}(R(j)R(i)-R^2)$$

$$A_{j,k} \triangleq \text{the } j^{th} \text{ bracketed}'\{\ \}' \text{ term on the RHS of (3.60)} \qquad \text{(F.27)}$$

Using the triangular inequality of the Hilbert Schmidt norm yields:

$$E[||A_k||_{HS}] \leq \mu E[||A_{1,k}||_{HS}] + \mu^2 E[||A_{2,k}||_{HS}] + ... + \mu^k E[||A_{k,k}||_{HS}] \qquad \text{(F.28)}$$

We will show that, on a $1/\mu$ time scale, each of the terms $E[||A_{j,k}||]$, $2 \leq j \leq k$ is overbounded by an expression involving $\sup_{k\in[0,L/\mu)} E[||A_{1,k}||]$.

From the definition of $A_{2,k}$ above, it follows that:

$$\begin{aligned} A_{2,k} &= \sum_{i=0}^{k-1}[\sum_{j=i+1}^{k-2}(R(j)-R)]R(i) + R\sum_{i=0}^{k-1}[\sum_{j=0}^{k-2-i}(R(j)-R)] \\ &= \sum_{i=0}^{k-1} A_{1,k-i-1}(i+1)R(i) + R\sum_{i=0}^{k-1} A_{1,k-i-1} \end{aligned} \tag{F.29}$$

where $A_{1,k}(k_0) \triangleq \sum_{i=k_0}^{k-1+k_0}(R_i - R)$.

Combining the inequality of (F.22) with $\beta = \sup_k ||R_k||_{i2}$ and the ergodic property of the ensemble leads to

$$E[||A_{2,k}||_{HS}] \leq 2\beta \sum_{i=0}^{k-1} E[||A_{1,k-i-1}||_{HS}] \tag{F.30}$$

$$\leq 2\beta k \sup_{k\in[0,L/\mu)} E[||A_{1,k}||_{HS}] \tag{F.31}$$

From the definition of $A_{3,k}$ above, it follows that

$$\begin{aligned} A_{3,k} &= \sum_{i=0}^{k-2}[\sum_{j=i+1}^{k-3}\sum_{l=j+1}^{k-4}(R(l)R(j)-R^2)]R(i) + R^2\sum_{i=0}^{k-2}[\sum_{j=0}^{k-3-i}\sum_{l=0}^{k-4-i-j}(R(l)-R)] \\ &= \sum_{i=0}^{k-2} A_{2,k-i-2}(i+1)R(i) + R^2\sum_{i=0}^{k-2}[\sum_{j=0}^{k-3-i} A_{1,k-2-i-j}] \end{aligned} \tag{F.32}$$

Applying (F.22), (F.31) and the ergodic property of the ensemble leads to

$$\begin{aligned} E[||A_{3,k}||_{HS}] &\leq 2\beta^2\sum_{i=0}^{k-2}(k-i-2)\sup_{k\in[0,L/\mu)} E[||A_{1,k}||_{HS}] \\ &+ \beta^2\sum_{i=0}^{k-2}[\sum_{j=0}^{k-3-i}\sup_{k\in[0,L/\mu)} E[||A_{1,k}||_{HS}] \\ &= 3/4\left[(2\beta)^2\frac{k(k-1)}{2}\sup_{k\in[0,L/\mu)} E[||A_{1,k}||_{HS}]\right] \end{aligned} \tag{F.33}$$

Similarly, it can be shown that:

$$\begin{aligned} E[||A_{j,k}||_{HS}] &\leq \frac{j}{2^{j-1}}\left[(2\beta)^{j-1}C_{j-1}^k\sup_{k\in[0,L/\mu)} E[||A_{1,k}||_{HS}]\right] \\ &< \left[(2\beta)^{j-1}C_{j-1}^k\sup_{k\in[0,L/\mu)} E[||A_{1,k}||_{HS}]\right] \end{aligned} \tag{F.34}$$

where $C_j^k = k!/((k-j)!j!)$. Combining (F.28) and (F.34) with the fact that

$$(1+x)^k = \sum_{j=0}^{k} C_j^k x^j$$

leads to

$$\sup_{k\in[0,L/\mu)} E[||A_k||_{HS}] < \sup_{k\in[0,L/\mu)} E[||\mu\sum_{i=0}^{k}(R(i)-R)||_{HS}](1+2\beta\mu)^{L/\mu} \qquad \text{(F.35)}$$

By following a similar procedure to that above it can be shown that

$$\begin{aligned} E[||B_{mL/\mu+k}(k_0)] &\leq \left\{1+\frac{1+\sup_{k\in[0,L/\mu)}||A_k||_{i2}}{1-[\sup\sup_{k\in[0,L/\mu)}||A_k||_{i2}+(I-\mu R)^{L/\mu}]}\right\} \\ &\times \sup_{k\in[0,L/\mu)} E[||B_k||_2] \\ &\leq \frac{2-(I-\mu R)^{L/\mu}}{1-[\sup\sup_{k\in[0,L/\mu)}||A_k||_{i2}+(I-\mu R)^{L/\mu}]} \\ &\times \sup_{k\in[0,L/\mu)} E[||B_k||_2] \qquad \text{(F.36)} \end{aligned}$$

$$\sup_{k\in[0,L/\mu)} E[||B_k||_2] < \sup_{k\in[0,L/\mu)} E[||\mu\sum_{i=0}^{k-1}P(i)||_2](1+\mu(1+G)\beta)^{L/\mu} \qquad \text{(F.37)}$$

where G is as defined in Theorem 4.

F.6 Proof of Theorem 5

The results of (3.72) and (3.73) of Theorem 5 follow directly from applying the result of the following Lemma to:

$$\begin{aligned} E[\mu^2||\sum_{l=0}^{k-1}(R-R(l))||_{HS}^2] &= \mu^2 k\{\sum_{l=-n+1}^{n-1}(1-\frac{|l|}{n})E[1/k(\sum_{i=0}^{k-1}r_{i,l}-r_l)^2]\} \\ E[\mu^2||\sum_{l=0}^{k-1}(P(l))||_2^2] &= \mu^2 k\{\sum_{l=-n+1}^{n-1}(1-\frac{|l|}{n})E[1/k(\sum_{i=0}^{k-1}p_{i,l})^2]\} \end{aligned}$$

The Lemma below was obtained by Bitmead [69] through an extension of the work of [81].

Lemma 5 *Consider a stationary sequence r_k with mean r. Suppose the power spectrum $f(\omega)$ of r_k exists and is twice differentiable at $\omega = 0$. Then*

$$\begin{aligned} E[1/k(\sum_{m=t}^{k-1+t}(r_m-r))^2] &= \{\sum_{m=-\infty}^{\infty}E[(r_i-r)(r_{i+m}-r)] \qquad \text{(F.38)} \\ &+ O(k^{-1/2}\frac{d^2f(\omega)}{d\omega^2}|_{\omega=0})+O(k^{-3/4})+O(k^{-3/2})]\} \end{aligned}$$

We prove now (3.75). Assuming the limit

$$\lim_{N\to\infty} 1/N \sum_{t=t_0}^{N-1+t_0}(r_{t,l}-r_l)(r_{t+|m|,l}-r_l) \triangleq E[(r_{t,l}-r_l)(r_{t+|m|,l}-r_l)]$$

exists uniformly in t_0, then it can be easily verified that:

$$\sum_{m=-k+1}^{k-1}(1-\frac{|m|}{k})E[(r_{t,l}-r_l)(r_{t+|m|,l}-r_l)] = E[1/k(\sum_{i=t}^{k-1+t} r_{i,l}-r_l)^2] \qquad \text{(F.39)}$$

Therefore,

$$\begin{aligned} 2\pi f_l(0) &= \lim_{k\to\infty}\sum_{m=-k}^{k} E[(r_{t,l}-r_l)(r_{t+|m|,l}-r_l)] \\ &= \lim_{k\to\infty}\sum_{m=-k}^{k}\frac{|m|}{k}E[(r_{t,l}-r_l)(r_{t+|m|,l}-r_l)] \\ &+ \lim_{k\to\infty} E[1/k(\sum_{i=t}^{k-1+t} r_{i,l}-r_l)^2] \end{aligned} \qquad \text{(F.40)}$$

The twice differentiability condition of Theorem 5 requires the existence of

$$\frac{d^2 f(\omega)}{d\omega^2}|_{\omega=0} = -1/2\pi \sum_{m=-\infty}^{\infty}\{E[(r_{t,l}-r_l)(r_{t+|m|,l}-r_l)]\}m^2.$$

This implies

$$\lim_{k\to\infty}\frac{1}{k}\sum_{m=-k}^{k}|m||E[(r_{t,l}-r_l)(r_{t+|m|,l}-r_l)]| = 0 \qquad \text{(F.41)}$$

Thus,

$$2\pi f_l(0) = \lim_{k\to\infty} E[1/k(\sum_{i=t}^{k-1+t} r_{i,l}-r_l)^2] > 0 \qquad \text{(F.42)}$$

The proof of $g_l(0) > 0$ follows along similar lines.

F.7 Proof of Theorem 6

From (3.72) we have

$$\begin{aligned} 2\pi f_l(0) &= \sum_{m=-\infty}^{\infty} E[(r_{t,l}-r_l)(r_{t+|m|,l}-r_l)] \\ &= \sum_{m=-\infty}^{\infty}[E[r_{t,l}r_{t+|m|,l}]-r_l^2] \\ &= \sum_{m=-\infty}^{\infty}[E[u(t)u(t-l)u(t+|m|)u(t+|m|-l)]-E^2[u(t)u(t-l)]] \end{aligned}$$

Since $u(k)$ is assumed to be Gaussian, then

$$\begin{aligned} E[u(t)u(t-l)u(t+m)u(t+m-l)] &= E[u(t)u(t-l)]E[u(t+m)u(t+m-l)] \\ &+ E[u(t)u(t+m)]E[u(t-l)u(t+m-l)] \\ &+ E[u(t)u(t+m-l)]E[u(t-l)u(t+m)] \end{aligned}$$

and, therefore,

$$E[u(t)u(t-l)u(t+m)u(t+m-l)] - E^2[u(t)u(t-l)] = B_m^2 + B_{l-m}B_{l+m} \quad \text{(F.43)}$$

When $|B_q| \leq Ba_u^{|q|}$, with $B = B_q|_{q=0}$, $0 \leq a_u < 1$, then

$$\begin{aligned}
2\pi f_l(0) &= \sum_{m=-\infty}^{\infty} (B_{|m|}^2 + B_{l-|m|}B_{l+|m|}) \\
&= B_0^2 + B_l^2 + 2\sum_{m=1}^{\infty} B_m^2 + 2\sum_{m=1}^{|l|-1} B_{l-m}B_{l+m} + 2\sum_{m=|l|}^{\infty} B_{l-m}B_{l+m} \\
&\leq B^2[1 + a_u^{2|l|} + \frac{2a_u^2}{1-a_u^2} + 2(|l|-1)a_u^{2|l|} + \frac{2a_u^{2|l|}}{1-a_u^2}] \\
&= B^2[\frac{(1+a_u^2)(1+a_u^{2|l|})}{1-a_u^2} + 2|l|a_u^{2|l|}] \qquad \text{(F.44)}
\end{aligned}$$

Combining (F.44) with (3.72) leads to (3.77). The proof to (3.78) follows along similar lines to that above.

Appendix G

Chapter 4 Proofs

G.1 Proof of (4.27)

Consider the update equations of (4.21) and (4.24) in the more general form:

$$x(k+1) = A(\theta(k))x(k) + Bu(k) \tag{G.1}$$

$$\theta(k+1) = \theta(k) + \mu h(\theta(k), x(k)) \tag{G.2}$$

where we assume:

(i) $A(\theta)$ has Lipschitz continuous first order derivatives in θ;
(ii) $h(\theta, x)$ is Lipschitz continuous in (θ, x);
(iii) $A(\theta)$ is stable for $\theta \in D$.

Suppose $\theta(0) \in D^o \subset D$, then $\theta(k)$ remains inside D over a $O(1/\mu)$ time interval.

Consider:

$$x^s(k+1, \bar{\theta}) = A(\bar{\theta})x^s(k, \bar{\theta}) + Bu(k), \quad \bar{\theta} \in D \tag{G.3}$$

Assumption (iii) implies that $x^s(k)$ is well defined on $k \in [0, \infty)$. Moreover, Assumption (i) and (iii) imply that $x^s(k)$ is continuously differentiable in $\bar{\theta}$, since

$$\mathcal{D}x^s(k+1, \bar{\theta}) = (\mathcal{D}A(\bar{\theta}))x^s(k, \bar{\theta}) + A(\bar{\theta})[\mathcal{D}x^s(k, \bar{\theta})] \tag{G.4}$$

where $\mathcal{D}$ is the differentiable operator in $\bar{\theta}$.

Define: $z(k) \triangleq x(k) - x^s(k, \theta(k))$. Then

$$
\begin{aligned}
z(k+1) &= x(k+1) - x^s(k, \theta(k+1)) \\
&= A(\theta(k))x(k) + Bu(k) - [A(\theta(k+1))x^s(k, \theta(k+1)) + Bu(k)] \\
&= A(\theta(k))[x(k) - x^s(k, \theta(k))] \\
&+ A(\theta(k))x^s(k, \theta(k)) - A(\theta(k+1))x^s(k+1, \theta(k+1)) \\
&= A(\theta(k))z(k) + [A(\theta(k)) - A(\theta(k+1))]x^s(k, \theta(k)) \\
&+ A(\theta(k+1))[x^s(k, \theta(k)) - x^s(k, \theta(k+1))] \qquad \text{(G.5)}
\end{aligned}
$$

Assumption (i) implies that

$$
||[A(\theta(k)) - A(\theta(k+1))]x^s(k, \theta(k))|| \leq \gamma_1 ||\theta(k) - \theta(k+1)|| ||x^s(k, \theta(k))|| \qquad \text{(G.6)}
$$

where γ_1 is the Lipschitz constant of $A(\theta)$. Since $x^s(k)$ is well defined then the RHS of (G.6) is $O(\mu)$. Also, since $x^s(k)$ is continuously differentiable in $\bar{\theta}$, then

$$
||A(\theta(k+1))[x^s(k, \theta(k)) - x^s(k, \theta(k+1))]|| \leq ||A(\theta(k+1))|| \gamma_2 ||\theta(k) - \theta(k+1)|| \qquad \text{(G.7)}
$$

the RHS of which is $O(\mu)$.

Thus, with $A(\theta)$ stable, we have

$$
||z(k+1)|| = O(\mu)
$$

for as long as $\theta(k)$ remains in D.

G.2 Proof of Theorem 7

The aim is to determine conditions which will ensure (4.40) is a uniform contraction to zero over some domain $D_0 \subset D$, where $D : |\bar{\theta}_1 \bar{\theta}_2| < 1$. (See Appendix A for a discussion on Uniform Contraction.)

Using the l_1 induced norm in the uniform contraction condition of (A.2) and applying this to (4.40), we find the required condition is:

$$
\max_{i=1,2} [|1 - \alpha_{ii} - \gamma_{ii}| + |\beta_{ii} + \eta_{ii}|] < 1 - \epsilon \quad \text{for some } 0 < \epsilon < 1 \qquad \text{(G.8)}
$$

Equation (A.3) specifies a condition on the domain D_0 over which this may hold. Taking this into account and the restriction on D, we define D_0 by:

$$
D_0 : \; ||(\bar{\theta}_1, \bar{\theta}_2)||_1 \leq 2(\Theta)^{1/2} \qquad \text{(G.9)}
$$

where $\Theta \in [0,1)$ is given by:

$$|\bar{\theta}_1\bar{\theta}_2| \leq \Theta < 1.$$

Considering (G.8), we choose to restrict $\alpha_{ii} + \gamma_{ii}$ by:

$$0 < \alpha_{ii} + \gamma_{ii} < 1, \quad i = \{1,2\} \tag{G.10}$$

The lower bound is necessary, while the upper bound inhibits the oscillatory behaviour which is likely to occur if the upper bound is increased to its allowable limit of $2 - \epsilon$.

With this restriction, (G.8) will be satisfied if:

$$\min_{i=1,2}[\alpha_{ii} + \gamma_{ii} - |\beta_{ii} + \eta_{ii}|] > \epsilon \tag{G.11}$$

Let $|Av_{sisi(l)}|$ be overbounded by:

$$|Av_{sisi(l)}| < R_i r_i^l \text{ where } 0 < r_i < 1, \quad R_i = Av_{sisi(0)} \text{ is finite} \tag{G.12}$$

A sufficient condition for the upper bound of (G.10) is:

$$\max_{i=1,2} \frac{\mu}{1-\Theta^2}[R + \sum_{n=0}^{\infty}(\Theta)^n 2\Theta R_i r_i^{2+2n} + Q] < 1$$

that is,

$$\mu < \min_{i=1,2}(1-\Theta^2)/[\frac{R_i(1+\Theta r_i^2)}{(1-\Theta r_i^2)} + Q_i] \tag{G.13}$$

A sufficient condition for the lower bound of (G.10) is (G.11). A sufficient condition for (G.11) is:

$$\epsilon/\mu < \min_{i=1,2} \frac{1}{1+\Theta}[\frac{R_i(1-r_i^2(1+2\Theta))}{(1-\Theta r_i^2)} + Q_i] \tag{G.14}$$

With ν substituted for ϵ/μ, the above results lead to Theorem 7.

G.3 Proof of Theorem 8

Following a similar procedure to the proof of Theorem 7, we find the required conditions for (4.53) to be a uniform contraction to the l_1 ball B:

$$||(\bar{\theta}_1\ \bar{\theta}_2)||_1 \leq B, \text{ for all } (\bar{\theta}_1\ \bar{\theta}_2)^T \in D_0,$$

where D_0 is as defined in (G.9), are given by:

$$\max_{i,j=1,2;i\neq j}[|1-\alpha_{ii}-\rho_{ij}| + |\beta_{ii}+\xi_{ij}|] < 1-\epsilon \text{ for some } 0 < \epsilon < 1 \tag{G.15}$$

$$\frac{\mu(|Av_{s1s2(1)}| + |Av_{s2s1(1)}|)}{(1-\Theta^2)\epsilon} = B < 2(\Theta)^{1/2} \tag{G.16}$$

As in Theorem 7 proof, we satisfy (G.15) by applying the restrictions:

$$0 < \alpha_{ii} + \rho_{ij} < 1, \quad i,j = \{1,2\} \ i \neq j \tag{G.17}$$

$$min_{i,j=1,2;i\neq j}[\alpha_{ii} + \rho_{ij} - |\beta_{ii} + \xi_{ij}|] > \epsilon \tag{G.18}$$

Let $|Av_{sisi(l)}|$ be overbounded as in (G.12) and $|Av_{s1s2(l)}|, |Av_{s2s1(l)}|$ be overbounded by:

$$|Av_{s1s2(l)}|, |Av_{s2s1(l)}| < Gg^l, \text{ where } 0 < g < 1, \ G = Av_{s1s2(0)}, Av_{s2s1(0)} \tag{G.19}$$

Following the same procedure as in Theorem 7 proof, we find that with ν substituted for ϵ/μ, equations (4.56),(4.57) are sufficient conditions for (G.17),(G.18) respectively. Finally, (G.16) is satisfied by (4.58).

G.4 Proof of Theorem 9

The averaged system (4.68) has equilibria given by the solutions to:

$$Av_{\bar{y}1\bar{y}2(1)}(z) = Av_{\bar{y}1\bar{y}2(1)}(z) = 0$$

We, therefore, begin by determining an explicit expression for these averaged terms of (4.69). This is achieved by combining the equations (4.70) and (4.63) with Assumptions 16-18 to construct a Lyapunov equation, the solution of which depends explicitly on the averaged terms (4.69) - see next section. The constructed Lyapunov equation was solved by using MAPLE V (see Section G.6) to obtain:

$$Av_{\bar{y}1\bar{y}2(1)}(z) = V^2 \frac{apz_1^2z_2^2 + a^2z_1^2z_2 + z_1z_2^2 + 2apz_1z_2 + z_1 + a^2z_2 + ap}{(1 - z_1^2z_2^2)(1 - a^2z_1z_2)} \tag{G.20}$$

$$Av_{\bar{y}1\bar{y}2(1)}(z) = V^2 \frac{apz_1^2z_2^2 + a^2z_2^2z_1 + z_2z_1^2 + 2apz_1z_2 + z_2 + a^2z_1 + ap}{(1 - z_1^2z_2^2)(1 - a^2z_1z_2)} \tag{G.21}$$

From the expressions (G.20) and (G.21) we conclude that the averaged system (4.68) has two finite equilibria $\tilde{\theta}_s^{av} = (\bar{\theta}_{12,s}, \bar{\theta}_{12,s})$ and $\bar{\theta}_u = (\bar{\theta}_{12,u}, \bar{\theta}_{12,u})$ given by (via MAPLE V):

$$\begin{aligned} \bar{\theta}_{12,s} &= -(1+a^2) + \sqrt{(1+a^2)^2 - 4a^2p^2}/2ap \\ \bar{\theta}_{12,u} &= -(1+a^2) - \sqrt{(1+a^2)^2 - 4a^2p^2}/2ap \end{aligned} \tag{G.22}$$

It is easily verified that for all $|a| < 1$ and all $|p| < 1$, $\bar{\theta}_u = (\bar{\theta}_{12,u}, \bar{\theta}_{12,u})$ lies outside the domain D of interest, indeed $|\bar{\theta}_{12,u}| > 1$. Consequently, this equilibrium will

be ignored. On the other hand, $\bar{\theta}_s = (\bar{\theta}_{12,s}, \bar{\theta}_{12,s}) \in D$. This is because, from the definition of Θ in Theorem 9,

$$|a| < \sqrt{\Theta} < 1 \tag{G.23}$$

which guarantees $|\bar{\theta}_{12,s}| < \sqrt{\Theta} < 1$. We, therefore, consider only $\bar{\theta}_s$.

From the Jacobian evaluated at $\bar{\theta}_s$ we obtain that the equilibrium $\bar{\theta}_s$ is (locally) asymptotically stable. A domain of attraction may be estimated as follows.

First we rewrite the system as follows:

$$\begin{aligned}
[\tilde{\theta}_1^{av}(k+1) + \tilde{\theta}_2^{av}(k+1)] &= [\tilde{\theta}_1^{av}(k) + \tilde{\theta}_2^{av}(k)] \\
&\times \left(1 - \mu V^2 \frac{1+a^2}{(1-\tilde{\theta}_1^{av}(k)\tilde{\theta}_2^{av}(k))(1-a^2\tilde{\theta}_1^{av}(k)\tilde{\theta}_2^{av}(k))}\right) \\
&- \mu V^2 \frac{2ap(1+\tilde{\theta}_1^{av}(k)\tilde{\theta}_2^{av}(k))}{(1-\tilde{\theta}_1^{av}(k)\tilde{\theta}_2^{av}(k))(1-a^2\tilde{\theta}_1^{av}(k)\tilde{\theta}_2^{av}(k))}
\end{aligned}$$

$$\begin{aligned}
[\tilde{\theta}_1^{av}(k+1) - \tilde{\theta}_2^{av}(k+1)] &= [\tilde{\theta}_1^{av}(k) - \tilde{\theta}_2^{av}(k)] \\
&\times \left(1 - \mu V^2 \frac{1-a^2}{(1+\tilde{\theta}_1^{av}(k)\tilde{\theta}_2^{av}(k))(1-a^2\tilde{\theta}_1^{av}(k)\tilde{\theta}_2^{av}(k))}\right)
\end{aligned}$$

Let us select both $|\theta_1| < \sqrt{\Theta} < 1$ and $|\theta_2| < \sqrt{\Theta} < 1$. Consider the domain:

$$E = \{(\tilde{\theta}_1^{av}(k), \tilde{\theta}_2^{av}(k)) \ : \ |\tilde{\theta}_1^{av}(k)| + |\tilde{\theta}_2^{av}(k)| < 2\sqrt{\Theta}\} \subset D \tag{G.24}$$

On this domain E we have the estimates:

$$\begin{aligned}
|\tilde{\theta}_1^{av}(k+1) + \tilde{\theta}_2^{av}(k+1)| &\leq |\tilde{\theta}_1^{av}(k) + \tilde{\theta}_2^{av}(k)|(1-\mu\alpha_k) + \mu\alpha_k \frac{2|ap|(1+\Theta)}{1+a^2} \\
|\tilde{\theta}_1^{av}(k+1) - \tilde{\theta}_2^{av}(k+1)| &\leq |\tilde{\theta}_1^{av}(k) - \tilde{\theta}_2^{av}(k)|(1-\mu\frac{V^2(1-a^2)}{(1+\Theta)(1+a^2\Theta)})
\end{aligned}$$

Here

$$\alpha_k = \frac{V^2(1+a^2)}{(1-\tilde{\theta}_1^{av}(k)\tilde{\theta}_2^{av}(k))(1-a^2\tilde{\theta}_1^{av}(k)\tilde{\theta}_2^{av}(k))}$$

On the domain $E \subset D$, α_k is positive and bounded above by $\dfrac{V^2(1+a^2)}{(1-\Theta)(1-a^2\Theta)}$. Let μ be sufficiently small (and positive), such that $\mu\dfrac{V^2(1+a^2)}{(1-\Theta)(1-a^2\Theta)} < 1$.

As the initial conditions are such that they belong to E, we conclude by induction that the domain E is invariant and that the trajectories converge to the invariant subset $F = \{\tilde{\theta}_1^{av}(.) = \tilde{\theta}_2^{av}(.)\} \cap E$. Indeed:

$$|\tilde{\theta}_1^{av}(k+1)| + |\tilde{\theta}_2^{av}(k+1)| \ < \ 2\sqrt{\Theta}, \ \forall k$$

since:

$$\begin{aligned} |\tilde{\theta}_1^{av}(k+1) + \tilde{\theta}_2^{av}(k+1)| &\leq (1-\mu\alpha_k)(2\sqrt{\Theta}) + \mu\alpha_k \frac{2|ap|(1+\Theta)}{1+a^2} \leq 2\sqrt{\Theta} \ \ \forall k \\ |\tilde{\theta}_1^{av}(k) - \tilde{\theta}_2^{av}(k)| &\leq |\theta_1 - \theta_2|(1 - \mu\frac{V^2(1-a^2)}{(1+\Theta)(1+a^2\Theta)})^k \leq 2\sqrt{\Theta} \\ &\rightarrow \ 0 \ as \ k \rightarrow \infty \end{aligned} \tag{G.25}$$

We now consider the dynamics restricted to the invariant set F. In F the dynamics are governed by the recursion:

$$\varphi_{k+1} = \varphi_k - \mu V^2 \frac{ap\varphi_k^2 + (a^2+1)\varphi_k + ap}{(1-\varphi_k^2)(1-a^2\varphi_k^2)} \tag{G.26}$$

which can be rewritten as:

$$[\varphi_{k+1} - \bar{\theta}_{12,s}] = (1 - \mu V^2 \frac{ap(\varphi_k - \bar{\theta}_{12,u})}{(1-\varphi_k^2)(1-a^2\varphi_k^2)})[\varphi_k - \bar{\theta}_{12,s}] \tag{G.27}$$

On the domain $|\varphi_k| < \sqrt{\Theta}$ we have the estimate

$$ap(\varphi_k - \bar{\theta}_{12,u}) > \frac{(1+a^2) + \sqrt{(1+a^2)^2 - 4a^2p^2}}{2} - |ap|\sqrt{\Theta} > 0$$

Thus with $|\varphi_0| < \sqrt{\Theta}$, φ_k converges exponentially to $\bar{\theta}_{12,s}$.

G.4.1 Constructing the Lyapunov Equation

The Lyapunov equation relevant to our system can be constructed by combining Equations (4.70) and (4.63) in conjunction with the Assumptions 16-18 as follows.

Let: $x(k) = (y_1(k,z)\ y_2(k,z)\ s_1(k)\ s_2(k))^T$, $w_k = (w_1(k)\ w_2(k))^T$, $z = (z_1, z_2)^T$.

Then,

$$x(k+1)x(k+1)^T = Ax(k)x(k)^TA^T + Ax(k)w(k)^TB^T + Bw(k)x(k)^TA^T + Bw(k)w(k)^TB^T$$

where:

$$A = \begin{pmatrix} 0 & z_1 & a & 0 \\ z_2 & 0 & 0 & a \\ 0 & 0 & a & 0 \\ 0 & 0 & 0 & a \end{pmatrix} \quad B = \sqrt{(1-a^2)} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Define:

$$P_{m+1} \triangleq \lim_{M \to \infty} 1/M \sum_{k=m}^{M+m-1} x(k+1)x(k+1)^T$$

Assumption 18 implies that

$$\lim_{M \to \infty} 1/M \sum_{k=m}^{M+m-1} x(k)w(k)^T = 0 \text{ (4 × 2 zero matrix)}$$

Combining this with Assumption 17 leads to

$$P_{m+1} = AP_mA^T + BWB^T$$

where

$$W = V^2 \begin{pmatrix} 1 & p \\ p & 1 \end{pmatrix}$$

For $a^2, z_1^2, z_2^2 < 1$ and $w_1(k), w_2(k)$ bounded (Assumption 16) we have $x(k)$ bounded. This leads to the existence of the following limit and, consequently, the Lyapunov equation:

$$\lim_{m \to \infty} P_{m+1} = P = APA^T + BWB^T$$

The averaged term $Av_{\bar{y}1\bar{y}2(1)}(z)$ is obtained from the solution P of this Lyapunov equation via:

$$Av_{\bar{y}1\bar{y}2(1)}(z) = \lim_{M \to \infty} 1/M \sum_{k=m}^{M-1+m} \bar{y}_1(k)(z)\bar{y}_2(k-1)(z) = c_1Pc_2^T$$

with

$$c_1 = \begin{pmatrix} 0 & z_1 & a & 0 \end{pmatrix} \quad c_2 = \begin{pmatrix} 0 & 1 & 0 & 0 \end{pmatrix}$$

The term $Av_{\bar{y}2\bar{y}1(1)}(z)$ is similarly obtained.

G.4.2 Maple procedures to solve a Lyapunov equation

The following Maple procedures were used to find the expressions for the averaged terms $Av_{\bar{y}1\bar{y}2(1)}(z)$ and $Av_{\bar{y}2\bar{y}1(1)}(z)$.

These procedures allow one to solve a Lyapunov equation of the type $P = APA^T + BWB^T$ where $W = W^T \geq 0$.

```
with(linalg);

tenp:=proc(a) local  i,j,k,l,n,aa;
n:=coldim(a);
aa:=matrix(n^2,n^2);
for i to n do
   for j to n do
     for k to n do
       for l to n do
aa[(i-1)*n+k,(j-1)*n+l]:=a[i,j]*a[k,l]
od;od;od;od;
aa
end;

colstack:=proc(b) local i,j,n;
n:=coldim(b);
aa:=matrix(n^2,1);
for i to n do
  for j to n do
aa[(i-1)*n+j,1]:=b[i,j]
od;od;
aa
end;

makemat:=proc(p) local n,i,j,a;
n:=sqrt(rowdim(p));
a:=matrix(n,n);
for i to n do
  for j to n do
a[i,j]:=p[(i-1)*n+j,1]
od;od;
a
end;
```

```
lyap:=proc(a,b,v) local d,c,f;
d:=add(scalarmul(tenp(a),-1),array(identity,1..coldim(a)^2,1..coldim(a)^2));
c:=multiply(multiply(b,v),transpose(b));
f:=colstack(c);
makemat(linsolve(d,f))
end;
```

Appendix H

Proof of Lemma 2

The results of [92] indicate that for an input signal which satisfies Assumptions 2 and 21 the following holds w.p.1. as $N \to \infty$:

$$\|R_N - \sigma_u^2 I\|_2 \leq 2n\,(\log N/N)^{1/2} + C\frac{n^2}{N}. \tag{H.1}$$

where n is the dimension of R_N, I is the $n \times n$ identity matrix and C is a constant. This implies that

$$\|R_N - \bar{R}_N\|_2 \to 0, \quad w.p.1.\ N \to \infty. \tag{H.2}$$

Since

$$\|R_N^{-1} - \bar{R}_N^{-1}\| \leq \|R_N^{-1}\|_2 \|R_N - \bar{R}_N\|_2 \|\bar{R}_N^{-1}\|_2, \tag{H.3}$$

$$\|R_N^{-1}\|_2 < C_1\ (constant) \quad \text{and} \quad \|R_N^{-1}\|_2 < C_2\ (constant) \quad \text{for large } N, \tag{H.4}$$

we have

$$\|R_N^{-1} - \bar{R}_N^{-1}\|_2 \to 0, \quad w.p.1.\ N \to \infty. \tag{H.5}$$

Since f_N is bounded then

$$\|R_N^{-1} f_N - \bar{R}_N^{-1} f_N\|_2 \to 0, \quad w.p.1.\ N \to \infty \tag{H.6}$$

$$\|f_N^T R_N^{-1} f_N - f_N^T \bar{R}_N^{-1} f_N\|_2 \to 0, \quad w.p.1.\ N \to \infty. \tag{H.7}$$

The results of Lemma 2 follow.

Appendix I

List of Assumptions

Assumption 1

(i) The unknown channel is time invariant and is adequately modelled by an n-tap digital FIR filter with tap coefficient vector

$$\hat{\theta} = (\hat{\theta}_0 \ \hat{\theta}_1 \ \hat{\theta}_2 \ ... \ \hat{\theta}_{n-1})^T. \tag{I.1}$$

(ii) The LMS adaptive FIR filter has a tap length of n and at sampling instant k has the tap coefficient vector

$$\hat{\theta}(k) = (\hat{\theta}_0(k) \ \hat{\theta}_1(k) \ \hat{\theta}_2(k) \ ... \ \hat{\theta}_{n-1}(k))^T. \tag{I.2}$$

(iii) The tap coefficients of the LMS adaptive FIR filter are initially set to zero:

$$\hat{\theta}_i(0) = 0, \ i = 0, 1, 2, ..., n-1.$$

Assumption 2

The input, $u(k)$, and disturbance, $s(k)$, signals are zero mean, bounded and stationary so that the limits:

$$\begin{aligned}
R_m \triangleq E(U(k)U(k)^T) &= \lim_{N\to\infty} 1/N \sum_{k=k_0}^{N-1+k_0} U(k)U(k)^T, \ \forall k_0 \\
P_m \triangleq E(U(k)s(k)) &= \lim_{N\to\infty} 1/N \sum_{k=k_0}^{N-1+k_0} U(k)s(k), \ \forall k_0 \\
\sigma_u^2 &\triangleq \lim_{N\to\infty} 1/N \sum_{k=k_0}^{N-1+k_0} u(k)^2, \ \forall k_0 \\
\sigma_s^2 &\triangleq \lim_{N\to\infty} 1/N \sum_{k=k_0}^{N-1+k_0} s(k)^2, \ \forall k_0
\end{aligned}$$

exist for all m, where m =length of $U(k) = (u(k)\ u(k-1)\ ...\ u(k-m+1))^T$.

Assumption 3

The input and disturbance signals are uncorrelated with each other over time:

$$\lim_{N\to\infty} 1/N \sum_{k=0}^{N-1} u(k-l)s(k) = 0, \quad \forall l$$

that is, $P_m = 0$.

Assumption 4

The input signal, $u(k)$, is such that the autocorrelation sequence $\{r_l\}$:

$$r_j = \lim_{N\to\infty} 1/N \sum_{k=0}^{N-1} u(k)u(k-j), \quad j = \ ...-2,\ -1,\ 0,\ 1,\ 2,\ ...$$

is absolutely summable:

$$\sum_{j=-\infty}^{\infty} |r_j| < \infty.$$

Assumption 5

The power spectrum $\Phi_{uu}(\omega)$ of the input signal is positive definite:

$$\Phi_{uu}(\omega) > 0 \quad 0 \leq \omega \leq 2\pi.$$

Assumption 6

The update stepsize is such that:

$$\mu \ll \frac{1}{n\sigma_u^2} \triangleq \frac{1}{Tr(R_n)} < \frac{1}{\lambda_{max}(R_n)} \tag{I.3}$$

where $Tr(.) = Trace(.)$.

Assumption 7

The input signal vectors $U(k)$ are independent and identically distributed.

Assumption 8

Each echo path is attenuating, that is:

$$||\hat{\theta}_1||_1 < 1, \quad ||\hat{\theta}_2||_1 < 1 \tag{I.4}$$

where $||.||_1$ is the 1-norm.

Assumption 9

Both transmission channels of the DEC system loop (*i.e.* channel $A \rightarrow B$ and channel $C \rightarrow D$ of Figure 4.1) impose a transmission delay of $d \geq 1$ sample intervals.

Assumption 10

The subscriber and channel noise signals are bounded and stationary so that the limits:

$$Av_{sisj(l)} \triangleq \lim_{N\to\infty} 1/N \sum_{k=k_0}^{N-1+k_0} s_i(k)s_j(k-l), \quad \forall k_0, \forall l \quad i,j=\{1,2\} \tag{I.5}$$

$$Av_{ninj(l)} \triangleq \lim_{N\to\infty} 1/N \sum_{k=k_0}^{N-1+k_0} n_i(k)n_j(k-l), \quad \forall k_0, \forall l \quad i,j=\{1,2\} \tag{I.6}$$

$$Av_{sinj(l)} \triangleq \lim_{N\to\infty} 1/N \sum_{k=k_0}^{N-1+k_0} s_i(k)n_j(k-l), \quad \forall k_0, \forall l \quad i,j=\{1,2\} \tag{I.7}$$

exist.

Assumption 11

$$\begin{array}{lcll} (a)\ Av_{ninj(l)} & = & \delta_{ij}Q_i & l=0 \\ & = & 0 & \text{otherwise} \\ (b)\ Av_{nisj(l)} & = & 0 & \forall l,\ i,j=\{1,2\} \end{array}$$

Assumption 12

$Av_{s1s2(l)}, Av_{s2s1(l)} \equiv 0, \quad \forall l.$

Assumption 13

$n_1(k), n_2(k) = 0, \quad \forall k.$

Assumption 14

The subscriber signals are described by the first order autoregressive processes:

$$\begin{array}{rcll} s_1(k+1) & = & as_1(k) + (1-a^2)^{1/2}w_1(k) & s_1(0)=0 \\ s_2(k+1) & = & as_2(k) + (1-a^2)^{1/2}w_2(k) & s_2(0)=0 \end{array} \tag{I.8}$$

where: $|a| < 1$ and $w_1(k)$ and $w_2(k)$ are wide sense stationary discrete white signals.

Assumption 15

Bounded signals:

$$|w_i(k)| < W, \ \forall k$$

Assumption 16

Zero mean property:

$$|\frac{1}{M} \sum_{k=m}^{M+m-1} w_i(k)| \leq \frac{C}{\sqrt{M}} \ \forall m, M$$

Assumption 17

Well defined autocorrelation properties:

$$|\frac{1}{M}\sum_{k=m}^{M+m-1} w_i(k)w_i(k-l) - V^2\delta_{0,l}| \leq \frac{C}{\sqrt{M}} \; \forall m, M, l$$

Assumption 18

Well defined cross correlation properties:

$$|\frac{1}{M}\sum_{k=m}^{M+m-1} w_i(k)w_j(k-l) - pV^2\delta_{0,l}| \leq \frac{C}{\sqrt{M}} \; \forall m, l, M \text{ where } |p| \leq 1$$

where the constant C is positive and independent of the integers m, l and M in the above inequalities.

Assumption 19

$$|\frac{1}{M}\sum_{k=m}^{M+m-1} w_i(k)w_i(k-l) - V_i^2\delta_{0,l}| \leq \frac{C}{\sqrt{M}} \; \forall m, M, l, \text{ where } V_i > 0$$

Assumption 20

$$|\frac{1}{M}\sum_{k=m}^{M+m-1} w_i(k)w_j(k-l) - pV_1V_2\delta_{0,l}| \leq \frac{C}{\sqrt{M}} \; \forall m, l, M \text{ where } |p| \leq 1$$

where $i, j = \{1, 2\}, i \neq j$.

Assumption 21

The input signal $u(k)$ is uncorrelated over time (white) so that its $n \times n$ autocorrelation matrix:

$$R_n = \sigma_u^2 I$$

where I is the $n \times n$ identity matrix and σ_u^2 is the variance of $u(k)$.

Assumption 22

The time invariant n-tuple FIR modelled echo , $\Theta(q^{-1})$, has only $m < n$ nonzero taps:

$$\Theta(q^{-1}) = b_{j_1}q^{-j_1} + b_{j_2}q^{-j_2} + ... + b_{j_m}q^{-j_m} \tag{I.9}$$

where $0 \leq j_i < j_{i+1} \leq n, \quad i = 1, 2, ..., m.$

Bibliography

[1] M.M. Sondhi and D.A. Berkley, "Silencing Echoes on the Telephone Network", *Proceedings of the IEEE,* Vol. 68, No. 8, pp948-963, 1980.

[2] R.H. Moffet, "Echo and Delay Problems in Some Digital Communication Systems", *IEEE Communications Magazine*, Vol. 25, No. 8, pp41-47, 1987.

[3] Personal communication with M.A. Gilloire, France Telecom, Centre National d'Etudes des Telecommunications, Lannion Cedex.

[4] R. Wehrmann, "Acoustic Echo Control - A Perpetual Challenge", *Signal Processing*, Vol. 27, p253, 1992.

[5] S.L. Gay and R.J. Mammone, "Fast Converging Subband Acoustic Echo Cancellation using RAP on the WE DSP16A", *Proceedings International Conference on Acoustic, Speech, Signal Processing*, ICASSP'90, pp1141-1144, Albuquerque, NM, 1990.

[6] C.W.K Gritton and D.W. Lin, "Echo Cancellation Algorithms", *IEEE ASSP Magazine*, pp. 30-37, April 1984.

[7] D.G. Messerschimtt, "Echo Cancellation in Speech and Data Transmission", *IEEE Journal on Selected Areas in Communications*, vol. SAC-2, pp283-297, 1984.

[8] D.T.M Slock, "On the Convergence Behaviour of the LMS and the Normalized LMS Algorithms", *IEEE Trans. on Signal Processing*, Vol. 41, No. 9, pp2811-2825, 1993.

[9] K. Yamazaki, S. Aly and D. Falconer, "Convergence behaviour of a jointly adaptive transversal and memory-based echo canceller", *IEE Proceedings-F*, vol. 138, No. 4 pp. 361-370, 1991.

[10] R.G. Gould and G.K. Helder, "Transmission Delay and Echo Suppression", *IEEE Spectrum*, pp47-54, April 1970.

[11] M. Rhee and C. Kang, "Effects of Line Code on the Echo Cancellation Scheme in the ISDN U-Interface Tranceiver", *Proceedings of Trends in Electronics Conference*, TENCON'87, pp172-176, Seoul, Korea, 1987.

[12] S. Yamamoto *et. al*, "The Echo Canceller using the Fast Kalman Filter Algorithm", *Proceedings 8th World Congress of the International Federation of Automatic Control*, IFAC'81, pp593-598, Kyoto, Japan, 1981.

[13] S. Makino and S. Shimada, "Echo Control in Telecommunications", *Journal of the Acoustical Society of Japan*, Vol. 11 (6), pp309-316, 1990.

[14] G.S. Fang, "Voice Channel Echo Cancellation", *IEEE Communications Magazine*, Vol. 21, No. 9, pp11-14, 1983.

[15] A. Gilloire and J. Zurcher, "Achieving the Control of the Acoustic Echo in Audio Terminals", *Proceedings EUSIPCO'88*, pp491-494, Grenoble, France, 1988.

[16] W. Hsu, F. Chui and D.A. Hodges, "An Acoustic Echo Canceller", *IEEE Journal of Solid State Circuits*, Vol. 24, No. 6, pp1639-1646, 1989.

[17] H. Perez and F. Amano, "A new Subband Echo Canceller Structure", *Trans. Inst. Electron. Inf. Commun. Eng. (IEICE)*, Vol. E73, No. 10, pp1625-1631, 1990.

[18] S. Makino and Y. Kaneda, "Acoustic Echo Canceller Algorithms Based on the Variation Characteristics of a Room Impulse Response", *Proceedings International Conference on Acoustic, Speech, Signal Processing*, ICASSP'90, pp1133-1136, Albuquerque, NM, 1990.

[19] H. Yasukawa, I. Furukawa and Y. Ishiyama, "Characteristics of Acoustic Echo Cancellers using Sub-band Sampling and Decorrelation Methods", *Electronics Letters*, Vol. 24, No. 16, pp1039-1040, 1988.

[20] E. Hansler, "The Hands Free Telephone Problem - An Annotated Bibligraphy", *Signal Processing*, Vol. 27, pp259-271, 1992.

[21] W.A. Sethares, C.R. Johnson, Jr. and C.E. Rohrs, "Bursting in Adaptive Hybrids", *IEEE Trans. on Communications*, Vol. 37, No. 8, pp791-799, 1989.

[22] I.M.Y. Mareels and R.K. Boel, "A Performance Oriented Analysis of a Double Hybrid Adaptive Echo Cancelling System", *Journal of Mathematics in Estimation and Control*, vol.2, no.1 pp. 71-94, 1992.

[23] Z. Ding, C.R. Johnson, Jr. and W.A. Sethares, "Frequency Dependent Bursting in Adaptive Echo Cancellation and its Prevention using Double-Talk Detectors", *International Journal of Adaptive Control and Signal Processing*", Vol. 4, pp219-236, 1990.

[24] W.A. Sethares and I.M.Y. Mareels, "Dynamics of an Adaptive Hybrid", *IEEE Trans. on Circuits and Systems*, Vol, 38, No. 1, pp1-11, 1991.

[25] G.J. Rey, R.R. Bitmead and C.R. Johnson, "The Dynamics of Bursting in Simple Feedback Systems with Leakage", *IEEE Trans. on Circuits and Systems*, Vol, 38, No. 5, pp476-488, 1991.

[26] K. Murano, S. Unagami and F. Amano, "Echo Cancellation and Applications", *IEEE Communications Magazine*, pp49-55, Jan 1990.

[27] K. Wesolowski, C.M. Zhao and W. Rupprecht, "Adaptive LMS Transversal Filters with Controlled Length", *IEE Proceedings-F*, Vol. 139, No. 3, pp233-238, 1992.

[28] W.A. Gardner, "Learning Characteristics of Stochastic-Gradient-Descent Algorithms: A General Study, Analysis and Critique", *Signal Processing*, Vol. 6, pp113-133, 1984.

[29] N.J. Bershad, "Analysis of the Normalized LMS Algorithm with Gaussian Inputs", *IEEE Trans. on Acoustic, Speech, Signal Processing*, Vol. ASSP-34, pp793-806, 1986.

[30] B. Widrow, J.M. McCool, M.G. Larimore, C.R. Johnson, "Stationary and Nonstationary Learning Characteristics of the LMS Adaptive Filter", *Proceedings of the IEEE*, Vol. 64, No. 8, pp1151-1162, 1976.

[31] M. Tarrab and A. Feuer, "Convergence and Performance Analysis of the Normalized LMS Algorithm with Uncorrelated Gaussian Data", *IEEE Trans. on Information Theory*, Vol. IT-34, No. 4, pp680-691, 1988.

[32] N. Furuya *et. al*, "Audio Conference Equipment with Acoustic Echo Canceller", *NEC Res. and Dev. Journal*, Vol. 76, pp18-23, 1985.

[33] O. Horna, "Correction Algorithms for Extended Echo Cancellers", *Proceedings International Conference on Acoustic, Speech, Signal Processing*, ICASSP'81, pp560-563, Atlanta, GA, 1981.

[34] K. Ozeki and T. Umeda, "An Adaptive Filtering Algorithm using an Orthonormal Projection to an Affine Subspace and its Properties", *Electronics and Communications in Japan*, vol.67-A, pp19-27, 1984.

[35] H. Yasukawa, S. Shimada and I. Furukawa, "Acoustic Echo Canceller with High Speech Quality", *Proceedings International Conference on Acoustic, Speech, Signal Processing*, ICASSP'87, pp2125-2128, Dallas. TX, 1987.

[36] C. Acker, P. vary and H. Ostendarp, "Acoustic Echo Cancellation using Prediction Residual Signals", *Proceedings Second European Conference on Speech Communication and Technology*, pp1297-1300, Genova, Italy, 1991.

[37] S. Yamamoto *et al*, "An adaptive Echo Canceller with linear Predictor", *Trans. IECE Japan*, Vol.E62, pp851-857, 1979.

[38] J. Sikorav, "Experiments to Identify and Track Non Stationarities in Audio Conference Rooms", iProceedings International Conference on Acoustic, Speech, Signal Processing, ICASSP'88, pp2566-2569, New York, NY, 1988.

[39] H. Sato *et al*, "A Fast Adaptive FIR Filter Algorithm with Reduced Computation for Adaptive Tap Position", *Proceedings 1993 IEICEJ*, pp20-25, Sendai, Japan, 1993.

[40] Y. Cheng and D.M. Etters, "Analysis of an Adaptive Technique for Modeling Sparse Systems", *IEEE Trans. on Acoustic, Speech, Signal Processing*, Vol. 37, No. 2, pp254-263, 1989.

[41] H. Schutze, "Comparison of LMS- and Fast-Kalman-Algorithm for Acoustic Echo Control", *International Workshop on Acoustic Echo Control*, Berlin, Germany, 1989.

[42] A. Gilloire and T. Petillon, "A Comparison of NLMS and Fast RLS Algorithms for the Identification of Time Varying Systems with Noisy Inputs - Application to acoustic Echo Cancellation", *Proceedings EUSIPCO-90*, pp417-420, Barcelona, Spain, 1990.

[43] H. Schutze and Z. Ren, "Numerical Characteristics of Fast Recursive Least Squares Transversal Adaptation Algorithm - A Comparative Study", *Signal Processing*, Vol. 27, No. 3, pp317-332, 1992.

[44] D.T.M. Slock and T. Kailath, "Numerically Stable Fast Transversal Filters for Recursive Least Squares Adaptive Filtering", *IEEE Trans. on Signal Processing*, Vol. sp-39, pp92-114, 1991.

[45] A. Benallal and A. Gilloire, "A New Method to Stabilize Fast RLS Algorithms Based on a First-Order Model of the Propagation of Numerical Errors", *Proceedings International Conference on Acoustic, Speech, Signal Processing*, ICASSP'88, pp1373-1376, New York, NY, 1988.

[46] B. Friedlander, "Lattice Filters for Adaptive Processing", *Proceedings of IEEE*, Vol. 70, No. 8, pp. 829-867, 1982.

[47] J. Chao and S. Tsujii, "A Stable and Distortion-Free IIR Echo and Howling Canceler", *IEEE Trans. on Signal Processing*, Vol. 39, No. 8, pp1812-1821, 1991.

[48] M.L. Honig, "Convergence models for Lattice Joint Process Estimators and Least Squares Algorithms", *IEEE Trans. on Acoustic, Speech, Signal Processing*, Vol. ASSP-31, pp415-425, 1983.

[49] E.H. Satorius and S.T. Alexander, "Channel Equalization using Adaptive Lattice Algorithms", *IEEE Trans. Communications*, Vol. COM-27, pp899-905, 1979.

[50] H. Fan and W.K. Jenkins, "An Investigation of an Adaptive IIR Echo Canceller: Advantages and Problems", *IEEE Trans. on Acoustics, Speech, Signal Processing*, Vol. 36, No. 12, pp1819-1833, 1988.

[51] G. Long, D. Shwed and D. Falconer, "Study of a Pole-Zero Adaptive Echo Canceller", *IEEE Trans. on Circuits and Systems*, Vol. CAS-34, No. 7, pp765-769, 1987.

[52] A.V. Zitzewitz, "Considerations on Acoustic Echo Cancellation Based on Real Time Experiments", *Proceedings EUSIPCO-90*, pp1987-1990, Barcelona, Spain, 1990.

[53] J. Chao, P. Huang and S. Tsujii, "Adaptive Echo Cancellers for Acoustic conference systems in Closed Loop", *Proceedings Communications Systems: Towards Global Integration*, ICCS'90, pp760-764, Singapore, 1990.

[54] O. Muron and J.Sikorav, "Modelling of Reverberators and Audioconference Rooms", *Proceedings International Conference on Acoustic, Speech, Signal Processing*, ICASSP'86, pp921-924, Tokyo, Japan, 1986.

[55] G.W. Davidson and D.D. Falconer, "Reduced Complexity Echo Cancellation Using Orthonormal Functions", *IEEE Trans. on Circuits and Systems*, Vol. 38, No. 1, pp20-28, 1991.

[56] J.J. Shynk, "Frequency Domain and Multirate Adaptive Filtering", *IEEE Signal Processing Magazine*, pp14-37, Jan. 1992.

[57] D.F. Elliot and K.R. Rao, *Fast Transforms: Algorithms, Analyses, Applications*, Orlando, FL:Academic, 1982.

[58] M.J. Shense, "The Spectral Dynamics of Evolving LMS Adaptive Filters", *Proceedings International Conference on Acoustic, Speech, Signal Processing*, ICASSP'79, pp950-953, Washington, DC, 1979.

[59] R.M. Gray, "Toeplitz and Circulant Matrices: II", Technical Report No. 6504-1, April 1977, Information Systems Laboratory, Stanford University, California.

[60] W. Kellermann, "Analysis and Design of Multirate Systems for Cancellation of Acoustical Echoes", *Proceedings International Conference on Acoustic, Speech, Signal Processing*, ICASSP'88, pp2570-2573, New York, NY, 1988.

[61] P.P. Vaidyanathan, "Multirate Digital Filters, Filterbanks, Polyphase Networks, and Applications: A Tutorial", *Proceedings IEEE*, Vol. 78, pp56-93, 1990.

[62] P.L. Chu, "Quadrature Mirror Filter Design for an Arbitrary Number of Equal Bandwidth Channels", *IEEE Trans. on Acoustic, Speech, Signal Processing*, Vol. ASSP-33, pp203-218, 1985.

[63] J. Chen and J. Vandewalle, "Design of a Zero Delay Sub-Band Acoustic Echo Canceller", *Proceedings ISCAS 1988*, pp1325-1328, Espoo, Finland, 1988.

[64] J. Chen *et al*, "A New Structure for Sub-band Acoustic Echo Canceller", *Proceedings International Conference on Acoustic, Speech, Signal Processing*, ICASSP'88, pp2574-2577, New York, NY, 1988.

[65] A. Gilloire and M .Vitterli, "Adaptive Filtering in Sub-bands", *Proceedings International Conference on Acoustic, Speech, Signal Processing*, ICASSP'88, pp1572-1575, New York, NY, 1988.

[66] S. Yamamoto and S. Kitayama, "An Adaptive Echo Canceller and Variable Step Gain Method", *Trans. IECE Japan*, Vol. E65, pp1-8, 1982.

[67] R. Wehrmann, J. van der List and P. Meissner, "A Noise Insensitive Compromise Gradient Method for the Adjustment of Adaptive Echo Cancellers", *IEEE Trans. Communications*, Vol. COM-28, pp753-759, 1980.

[68] R. Frenzel and M.E. Hennecke, "A Robust Echo Compensator: Implementation and Realtime Measurements", *Proceedings 1991 Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, 1991.

[69] R.R. Bitmead, B.D.O. Anderson and T.S. Ng, "Convergence Rate Determination for Gradient-based Adaptive Estimators", *Automatica*, Vol. 22, No. 2, pp185-191, 1986.

[70] A. Feuer and E. Weinstein, "Convergence Analysis of LMS Filters with Uncorrelated Gaussian Data", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 1, pp222-229, 1985.

[71] J.E. Mazo, "On the Independence Theory of Equalizer Convergence", *Bell System Technology Journal*, Vol. 58, pp963-993, 1979.

[72] B. Widrow *et al*, "Adaptive Noise Cancellation: Principles and Applications", *Proceedings of the IEEE*, Vol. 63, pp1692-1716, 1975.

[73] L.I. Horowitz, K.D. Senne, "Performance Advantage of Complex LMS for Controlling Narrow-Bank Adaptive Arrays", *IEEE Trans. Acoustic, Speech, Signal Processing*, Vol. ASSP-29, pp722-735, 1991.

[74] V. Vapnik, *Estimation of Dependences Based on Empirical Data*, Springer-Verlag, New York, 1982.

[75] D.C. Farden, "Stochastic Approximation with Correlated Data", *IEEE Trans. on Information Theory*, Vol. IT-27, pp105-113, 1981.

[76] D.C. Farden, "Tracking Properties of Adaptive Signal Processing Algorithms", *IEEE Trans. on Acoustic, Speech, Signal Processing*, Vol. ASSP-29, pp439-446, 1981.

[77] S.K. Jones, R.K. Cavin, III and W.M. Reed, "Analysis of Error-gradient Adaptive Linear Estimators for a Class of Stationary Dependent Processors", *IEEE Trans. on Information Theory*, Vol. IT-28, pp318-329, 1982.

[78] J.A. Sanders and F. Verhulst, '*Averaging Methods in Nonlinear Dynamical Systems*, Springer-Verlag, New York, Applied Mathematical Sciences **59**, 1985.

[79] K.J. Astrom, *Introduction to Stochastic Control Theory*, Academic Press, New York, 1970.

[80] S.L. Marple, Jr., *Digital Spectral Analysis with Applications*, Prentice-Hall, New Jersey, 1987.

[81] I.A. Ibragimov, Y.V. Linnik, *Independent and Stationary Sequences of Random Variables*, Wolters-Noordhoff, Groningen, 1971.

[82] H. Schultze, "Convergence of Acoustic Echo Cancellers for Hands-Free Telephones Operating Under Feedback Conditions", *IEEE Trans. on Speech and Audio Processing*, Vol. 1, No. 2, pp257-260, 1993.

[83] P.H. Bardell, "Design Considerations for Parallel Pseudorandom Pattern Generators", *Journal of Electronic Testing: Theory and Applications*, Vol. 1, pp73-87, 1990.

[84] E.A. Lee, D.G. Messerschmitt, *Digital Communications*, Kluwer Academic Publishers, Boston, 1988.

[85] A. Gersho and R.M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, Boston, 1992.

[86] R. Pettigrew and V. Cuperman, "Backward Pitch Prediction for Low-delay Speech Coding", *Proceedings Global Telecommunications Conference*, Globecom'89, pp147-1252, Dallas, TX, 1989.

[87] U. Schultheiss, "Uber die Adaption eines Kompensators fur akustische Echos", *VDI-Fort-schritt-Berichte*, Reihe 10, Nr. 90, 1988.

[88] C. Acker and P. Vary, "Combined Implementation of Predictive Speech Coding and Acoustic Echo Cancellation", Signal Processing VI, EUSIPCO-92, Vol.III, pp. 1641-1644, Brussels, 1992.

[89] M. Mboup, M. Bonnet and O. Macchi, "A New Adaptive Prewhitening Filter for Acoustic Echo Cancellation", *2nd Workshop on Acoustic Echo Cancellation*, Italy, 1991.

[90] J.B. Allen and D.A. Berkley, "Image Method for Efficiently Simulating Small-room Acoustics", *Journal Acoust. Soc. Amer.*, Vol. 65, pp943-950, 1979.

[91] J.R. Casar-Corredera and J.A. Alcazar-Fernandez, "An Acoustic Echo Canceller for Teleconference Systems", *Proceedings International Conference on Acoustic, Speech, Signal Processing*, ICASSP'86, pp1317-1320, Tokyo, Japan, 1986.

[92] L. Ljung and B. Wahlberg, "Asymptotic Properties of the Least Squares Method for Estimating Transfer Functions and Disturbance Spectra" *Advances in Applied Probability*, Vol. 24, pp412-440, 1992.

[93] F. Gustafsson, "Estimation of Discrete Parameters in Linear Systems", Ph.D. Thesis, Linkoping Studies in Science and Technology, Dissertation No. 271, 1992.

[94] J.K. Hale, *Ordinary Differential Equations*, Kreiger, Florida, 1980.

[95] L. Ljung and T. Soderstrom, *Theory and Practice of Recursive Identification*, MIT Press, Cambridge, 1983.

[96] L. Ljung, *System Identification: Theory for the User*, Prentice-Hall, New Jersey, 1987.

[97] G.A. Clark, S.R. Parker and S.K. Mitra, "A Unified Approach to Time- and Frequency-domain Realization of FIR Adaptive Digital Filters", *IEEE Trans. Acoustic, Speech, Signal Processing*, Vol. ASSP-31, pp1073-1083, 1983.

[98] R.R. Bitmead and B.D.O. Anderson, "Adaptive Frequency Sampling Filters", *IEEE Trans. Circuits and Systems*, Vol. CAS-28, pp524-533, 1981.

[99] H.Cramer and M.R.Leadbetter, *Stationary and Related Stochastic Srocesses: Sample Function Properties and their Applications*, Wiley, New York, 1967.