

ADAPTIVE ALGORITHMS FOR CHANNEL  
ESTIMATION: USING APRIORI  
INFORMATION FOR OPTIMAL DESIGN

BY

**MUHAMMAD SAQIB SOHAIL**

A Thesis Presented to the  
DEANSHIP OF GRADUATE STUDIES

**KING FAHD UNIVERSITY OF PETROLEUM & MINERALS**

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the  
Requirements for the Degree of

**MASTER OF SCIENCE**

In

**ELECTRICAL ENGINEERING**

JUNE 2008

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS  
DHAHRAN 31261, SAUDI ARABIA

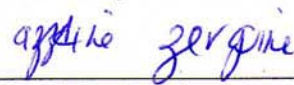
DEANSHIP OF GRADUATE STUDIES

This thesis, written by **MUHAMMAD SAQIB SOHAIL** under the direction of his thesis adviser and approved by his thesis committee, has been presented to and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN ELECTRICAL ENGINEERING**.

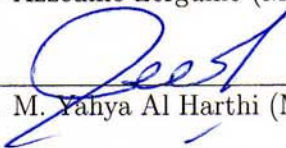
Thesis Committee

for 

Dr. Tareq Y. Al-Naffouri (Adviser)



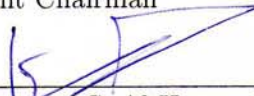
Dr. Azzedine Zerguine (Member)



Dr. M. Yahya Al Harthi (Member)

for 

Dr. Ibrahim O. Habiballah  
Department Chairman

  
Dr. Mohammad S. Al-Homoud  
Dean of Graduate Studies

Date

23/8/01



# ACKNOWLEDGMENTS

I am truly grateful to ALLAH, the most merciful, the most gracious for his countless blessings on me. It is He who gave me strength to complete this work. I pray that He continues to shower his blessings and gives me the wisdom to use whatever I have learned in His cause.

I am grateful to King Fahd University of Petroleum & Minerals for providing me the opportunity to study and carry out my research at such a prestigious university under the guidance of highly qualified faculty.

I am truly indebted to my thesis advisor Dr. Tareq Al-Naffouri for the endless hours of help, suggestions, ideas and advice. I was fortunate to take two courses with him and learn from his immense knowledge. Halfway through the thesis he had to leave for University of Southern California, USA, but in spite of the geographical distance and loads of work, he was in constant touch, advising, reviewing and helping me out in every problem I faced during the course of the thesis. I would also like to thank him for flying back to Saudi Arabia on a very tight schedule to attend my thesis defense despite his various engagements.

I am also grateful to my thesis committee members Dr. Azzedine Zerguine and Dr. Yahya Al-Harhi for their detailed review and excellent advice during

the preparation of this thesis.

I would also like to express my deep gratitude to Dr. Mohammad Moinuddin for his support and guidance. From the time I first came to KFUPM, to the final days when I was working on the thesis, I could always seek his advise and guidance whenever needed.

I am grateful to my friends Ahmed, Zeeshan and Salman for always being there for me. I cannot forget the time we spent together working long hours, supporting each other through the frustrating dull patches during the course of the thesis, playing games, exploring the city and its various restaurants, searching for jobs and having long informative discussions on a vast spectrum of topics. I would also like to show appreciation to my mat'am (kitchen) members Mazhar bhai, Ahmed, Zeeshan, Salman, Babar, Asif, Akhlaq bhai and Ayaz. They have really made my stay at KFUPM a truly memorable experience of my life. Special thanks is due to my friend Tayeb for translating the abstract of this work into arabic.

This work would not have been possible without the support and encouragement of my Mother and Father. Their prayers really carried me through all the tough times. I owe it to my mother for being my first teacher and taking a keen interest in my study throughout my life. Whatever I am today is because of my mother. I would also like to thank my sisters Zainab and Maymoona and brother Fahad for their continued friendship, love, support and countless indirect contributions to my life. Lastly, I would like to thank my wife Madiha for her constant support and encouragement.

TO MY MOTHER AND FATHER

# TABLE OF CONTENTS

<b>LIST OF TABLES</b>	<b>viii</b>
<b>LIST OF FIGURES</b>	<b>ix</b>
<b>ABSTRACT (ENGLISH)</b>	<b>xii</b>
<b>ABSTRACT (ARABIC)</b>	<b>xiv</b>
<b>1 INTRODUCTION</b>	<b>1</b>
<b>I Frequency Domain Channel Estimation in OFDM</b>	<b>4</b>
<b>2 INTRODUCTION TO CHANNEL ESTIMATION</b>	<b>5</b>
2.1 System Model . . . . .	6
2.2 Literature Review . . . . .	9
2.2.1 Channel Estimation using Pilots . . . . .	10
2.2.2 Blind Channel Estimation . . . . .	10
2.2.3 Semi Blind Channel Estimation . . . . .	11
2.2.4 Data Aided Channel Estimation . . . . .	12
2.2.5 Constraints Used in Channel Estimation/Data Detection .	12
2.2.6 Time Domain Channel Estimation . . . . .	14
2.2.7 Frequency Domain Channel Estimation . . . . .	15
2.3 Disadvantage of Performing Channel Estimation in Time Domain	16
2.4 Can We Perform Channel Estimation Reliably in Frequency Domain?	18

2.5	Input/Output Relationship in the Frequency Domain . . . . .	19
2.6	Pilot/Output Relationship in Frequency Domain . . . . .	20
2.7	A Parameter Reduction Approach . . . . .	20
2.8	Conclusion . . . . .	21
<b>3</b>	<b>INTERPOLATION BASED FREQUENCY DOMAIN ESTIMATION</b>	<b>22</b>
3.1	Least Squares . . . . .	25
3.2	Simulation Parameters . . . . .	26
3.3	Effect of Number of Pilots . . . . .	26
3.4	Effect of Section Length . . . . .	27
3.5	Effect of Varying the Number of Parameters . . . . .	29
3.6	A Scheme to Improve the Channel Estimate . . . . .	31
3.7	Least Squares with Regularization . . . . .	35
3.8	Conclusion . . . . .	36
<b>4</b>	<b>EIGENVALUE APPROACH TO FREQUENCY DOMAIN CHANNEL ESTIMATION</b>	<b>38</b>
4.1	Iterative Channel Estimation using the Expectation Maximization Approach . . . . .	42
4.1.1	The Maximization Step . . . . .	42
4.1.2	The Expectation Step . . . . .	46
4.1.3	Summary of the EM Algorithm . . . . .	48
4.2	Using Time-Correlation to Improve the Channel Estimate . . . . .	49
4.2.1	Developing a Frequency Domain Time-Variant Model . . . . .	49
4.2.2	Initial (Pilot-Based) Channel Estimation . . . . .	51
4.2.3	Iterative (Data-Aided) Channel Estimation . . . . .	52
4.2.4	Cyclic FB Kalman . . . . .	54
4.2.5	Helix based FB Kalman . . . . .	55
4.2.6	Using Code to Enhance the Estimate . . . . .	55
4.2.7	Forward Kalman Filter . . . . .	55

4.3	Time Domain Multiple Access Channel Estimation . . . . .	56
4.4	Simulation Results . . . . .	57
4.4.1	Effect of Modeling Noise . . . . .	58
4.4.2	EM based Least Squares . . . . .	60
4.4.3	Kalman Filter based Receivers . . . . .	61
4.4.4	Pilot Design . . . . .	65
4.5	Conclusion . . . . .	67
	<b>APPENDIX A: Moment Calculation</b>	<b>69</b>
 <b>II Least Mean Square Adaptive Filters with Optimum Error Nonlinearity</b>		<b>70</b>
<b>5</b>	<b>INTRODUCTION TO ADAPTIVE FILTERS</b>	<b>71</b>
5.1	System Model . . . . .	72
5.2	Evaluating Adaptive Filters . . . . .	73
5.2.1	Error Measures . . . . .	73
5.2.2	Various Assumptions used for Evaluating Performance of an Adaptive Filter . . . . .	74
5.2.3	Performance Criteria . . . . .	74
5.3	Fundamental Energy Relation . . . . .	76
5.4	A Brief Overview of Previous Work . . . . .	77
<b>6</b>	<b>OPTIMUM ERROR NON LINEARITY–AT STEADY STATE</b>	<b>80</b>
6.1	The MSE for General Error Nonlinearity . . . . .	81
6.2	Excess Mean Square Error . . . . .	84
6.3	Optimum Choice of the Nonlinearity . . . . .	85
6.4	Optimum Error Nonlinearity for Some Special Cases . . . . .	87
6.4.1	Gaussian Noise . . . . .	88
6.4.2	Laplacian Noise . . . . .	89
6.4.3	Binary Noise . . . . .	90



6.4.4	Gaussian Mixture . . . . .	91
6.5	Optimum Error Nonlinearity with Conditional Analysis . . . . .	92
6.6	The Conditional Error Nonlinearities for Special Noise Cases . . . . .	95
6.7	Role of the Variance of $e_a(i)$ . . . . .	96
6.8	Calculating the Variance of $e_a(i)$ . . . . .	97
6.9	Simulation Results . . . . .	99
6.9.1	The Effect of Varying $\sigma_{ea}^2$ . . . . .	99
6.9.2	Switched Mode Case . . . . .	102
6.9.3	Using Averaged Value of $\sigma_{ea}^2$ Over a Window Size . . . . .	108
6.9.4	The Effect of Window Size . . . . .	110
6.9.5	Time Variant Channels . . . . .	112
6.10	Conclusion . . . . .	117
<b>APPENDIX A: Error Nonlinearity for Laplacian Noise</b>		<b>118</b>
<b>APPENDIX B: Error Nonlinearity for Binary Noise</b>		<b>121</b>
<b>7</b>	<b>CONCLUSION AND FUTURE WORK</b>	<b>123</b>
7.1	Conclusion . . . . .	123
7.2	Future Work . . . . .	124
<b>REFERENCES</b>		<b>126</b>
<b>VITAE</b>		<b>139</b>

# LIST OF TABLES

5.1	<i>Examples for <math>f[e(i)]</math>.</i>	73
-----	---	----

# LIST OF FIGURES

2.1	OFDM transmitter block diagram. . . . .	7
2.2	CIR in the frequency domain partitioned in four subchannels. . . . .	19
3.1	CIR in the frequency domain divided into 4, 8 and 16 parts. . . . .	24
3.2	BER plot showing effect of number of pilots. . . . .	27
3.3	BER plot showing effect of section length. . . . .	28
3.4	BER plot showing effect of paramters . . . . .	29
3.5	Example of averaging scheme. . . . .	32
3.6	Err and BER plots for the averaging scheme. . . . .	33
3.7	Err and BER plots for the averaging scheme. . . . .	34
3.8	BER plots for regularized LS . . . . .	37
4.1	Block diagram of the system. . . . .	39
4.2	MSE and BER comparison. . . . .	59
4.3	MSE comparison EM based least squares (16 pilots) . . . . .	60
4.4	BER comparison EM based least squares (16 pilots) . . . . .	61
4.5	MSE comparison EM based least squares (24 pilots) . . . . .	62
4.6	BER comparison EM based least squares (24 pilots) . . . . .	62
4.7	BER comparison for various uncoded freq. domain methods( 16 and 24 pilots). . . . .	63
4.8	BER comparison for frequency domain coded methods (16 pilots). . . . .	64
4.9	BER comparison for the coded case (16 pilots) . . . . .	65
4.10	BER comparison of time and frequency domain uncoded techniques (24 pilots) . . . . .	66

4.11	MSE for all pilot patterns. . . . .	67
4.12	MSE for all pilot patterns. . . . .	68
6.1	Error nonlinearity for the Gaussian case . . . . .	89
6.2	Error nonlinearity for the Laplacian case . . . . .	91
6.3	Error nonlinearity for the Binary case . . . . .	92
6.4	Error nonlinearity for the Gaussian mixture case . . . . .	93
6.5	MSD of Optimum filter at various values of $\sigma_{ea}^2$ in Gaussian Noise	100
6.6	MSD Optimum filter at various values of $\sigma_{ea}^2$ in Binary Noise . . .	101
6.7	MSD Optimum filter at various values of $\sigma_{ea}^2$ in Laplacian Noise .	101
6.8	Effect of value of $\sigma_{ea}^2$ onMSD (Gaussian noise). . . . .	102
6.9	Effect of value of $\sigma_{ea}^2$ on convergence speed (Gaussian noise). . . .	103
6.10	Effect of value of $\sigma_{ea}^2$ on MSD (Binary noise). . . . .	103
6.11	Effect of value of $\sigma_{ea}^2$ on convergence speed (Binary noise). . . . .	104
6.12	Effect of value of $\sigma_{ea}^2$ on MSD (Laplacian noise). . . . .	104
6.13	Effect of value of $\sigma_{ea}^2$ on convergence speed (Laplacian noise). . . .	105
6.14	MSD learning curves for Gaussian noise (switched mode case) . .	106
6.15	MSD learning curves for Laplacian noise (switched mode case) . .	106
6.16	MSD learning curves for Binary noise (switched mode case) . . . .	107
6.17	MSD learning curves for Gaussian mixture noise with $m=0.3$ , $\sigma_{v_1}^2=100$ and $\sigma_{v_2}^2 = 1$ ( <i>switchedmodecase</i> ) . . . . .	107
6.18	MSD learning curves for Gaussian noise windowing method . . . .	108
6.19	MSD learning curves for Laplacian windowing method . . . . .	109
6.20	MSD learning curves for Binary noise windowing method . . . . .	109
6.21	MSD learning curves for Gaussian mixture noise windowing method	110
6.22	Steady state value of MSD vs window size (Binary Noise) . . . . .	111
6.23	Steady state value of MSD vs window size (Laplacian noise) . . . . .	111
6.24	MSD for RW time varying channel in Gaussian noise . . . . .	113
6.25	MSD for RW time varying channel in Binary noise . . . . .	113
6.26	MSD for AR time varying channel in Gaussian noise . . . . .	114

6.27 MSD for AR time varying channel in Binary noise . . . . .	115
6.28 MSD for probabilistic AR time varying channel in Gaussian noise . . .	116
6.29 MSD for probabilistic AR time varying channel in Binary noise . . . .	116

# THESIS ABSTRACT

**NAME:** Muhammad Saqib Sohail  
**TITLE OF STUDY:** Adaptive Algorithms for Channel Estimation: Using  
Apriori Information for Optimal Design  
**MAJOR FIELD:** Electrical Engineering  
**DATE OF DEGREE:** JUNE 2008

This thesis is concerned with using a priori information on the communication problem to enhance the performance of adaptive algorithms. We demonstrate this for two scenarios.

In the first part of the thesis, we consider the problem of channel estimation for multiuser OFDM transmission over block time varying channels. We show that frequency domain based channel estimation can be more practical and less computationally complex than the time domain based estimation. We show how a priori information (time and frequency correlation) allows us to design Kalman filters and obtain a very good performance in environments with high Doppler.

In the second part we use our a priori knowledge about the output noise distribution to design adaptive filters with optimum error nonlinearities. Specifically,

we use the energy relation that is usually used for mean square analysis to derive the mean square error for general error nonlinearities and subsequently use that to design optimum error nonlinearities given our knowledge about the noise distribution. We use a similar technique to design adaptive filters with optimum error nonlinearities and data normalization that is usually used when the input data is correlated.

## خلاصة الرسالة

الاسم الكامل	: محمد ثاقب سهيل
عنوان الرسالة	: تصميم خوارزميات متكيفة لتقدير القناة باستخدام المعلومات مسبقه لتصميم أمثل
التخصص	: الهندسة الكهربائية
تاريخ الشهادة	: يونيو 2008

تستقدم هذه الرسالة معلومات مسبقة عن مشكلة الاتصال. لتحسين اداء الخوارزميات المتكيفة. تحديداً تم عرض سيناريوهين لتحسين الاداء. في الجزء الاول من هذه الرسالة ، تمت دراسة مشكلة تقدير القناة للبحث لعدد من المستخدمين (multiuser) عبر قناة متغيرة زمنياً. تبين الرسالة ان تقدير القناة علي اساس المجال الترددي يمكن ان يكون اكثر عملياً و اقل تعقيداً من التقدير علي اساس المجال الزمني. تبين الرسالة ايضاً ان المعلومات المسبقه ( التردد ، الزمن و الارتباط) تسمح لنا بتصميم مرشحات كلمان (Kalman Filters) و بالحصول على اداء جيد جداً في محيط ذي دوبلر (Doppler) عال.

في الجزء الثاني من هذه الرسالة، تمت دراسة المرشحات المتكيفة مع الزمن و ال تستعمل علاقات خطأ لا خطية تحديداً تم استعمال المعلومات المسبقة عن توزيع الضجيج في الخرج لتصميم علاقات لا خطية مثلي. استخدمت الرسالة علاقة الطاقة التي تستخدم عادة في تحليل المربع المتوسط (mean square analysis) للوصول الى علاقة عامة المتوسط مربع الخطأ (mean square analysis) عندما تستخدم المرشحات لعلاقات للخطأ لا خطية عامة. من ثم تم تصميم مرشحات متغير مع الزمن رات لا خطيات مثلي. كذلك تم استخدام طريقة مماثلة لتصميم مرشحات متكيفة مع الزمن ذات لا خطيات مثلي و ذات معطيات تسوية (data normalization) والتي عادة ما تستخدم عندما تكون معطيات الدخل مترابطة.



# CHAPTER 1

## INTRODUCTION

In a wireless system, data is sent over a time variant fading channel. At the receiver, we get the received signal convolved and corrupted with noise. Naturally we are interested in recovering the transmitted data. Suppose we have information about the channel over which the data is being transmitted. In this case, we can faithfully obtain the transmitted data by making use of the received signal and the channel information (through equalization). In reality, we do not have the prior knowledge of the transmission channel, and hence we have to settle for an estimate of the channel obtained at the receiver using some estimation technique.

Channel estimation is thus an important step in receiver design. In a communication system, the sole purpose of the channel estimation is to recover the transmitted data. The aim of this thesis is to design optimum channel estimators that make use of a priori information about the communication problem. Specifically, the thesis will consider two problems in this regard

- 1) *Frequency domain based channel estimation for multiuser OFDM system.*

In this part, we will show how a priori information about channel correlation in frequency and time can help reduce pilot overhead and complexity by considering channel estimation in frequency domain.

2) *An Adaptive Filter with Optimum Error Nonlinearity.* The channel estimator will be based on least mean squares filter that employs an error nonlinearity. We will show that a priori knowledge about noise statistics will allow us to design adaptive filters with optimum error nonlinearities.

Accordingly, the thesis is divided into two parts, each dealing with one of these two problems. Part 1 consists Chapters 2 to 4 while part 2 is composed of Chapters 5 and 6. Chapter 2 is an introduction to channel estimation in OFDM and provides a brief survey of some approaches to channel estimation given in the literature like pilot based channel estimation, semi blind channel estimation, blind channel estimation and data aided channel estimation. It also surveys techniques on the basis whether the channel estimation was performed in time domain or frequency domain and builds the case for frequency domain channel estimation by comparing the merits and demerits of the two methods. Lastly it outlines the input/output relationship in the frequency domain.

Chapter 3 explores a simple interpolation based frequency domain channel estimation technique. Here we assume the interpolation matrix to be either linear or quadratic. We develop the Least Square solution to this problem and introduce a scheme to improve the channel estimate.

Chapter 4 details the eigenvalue based approach to frequency domain channel

estimation. Next it discusses the use of the Expectation Maximization (EM) algorithm to make the channel estimation process iterative in order to improve the estimate. It also discusses how time correlation can be incorporated in the algorithm to further improve the estimate and discusses various forms of the Kalman filter that we have used.

Chapter 5 provides a brief introduction to adaptive filters and surveys the literature for the design of optimum nonlinearity. We take a look at the commonly used measures for evaluating the performance of adaptive filters and give a brief introduction to the fundamental energy conservation relation.

Chapter 6 details the design of adaptive filters with optimum error nonlinearity at steady state. We start with deriving the expression of excess mean square error. Next we derive the optimum choice of the nonlinearity and solve it for some special cases. We also derive the optimum error nonlinearity with conditional analysis and show the results for some special cases.

Chapter 7 provides some concluding remarks and some insight about future work.

# Part I

## Frequency Domain Channel

### Estimation in OFDM

## CHAPTER 2

# INTRODUCTION TO CHANNEL ESTIMATION

With the advent of the modern digital communication age, demands on the data transmission rates have exceeded several Mbps and will continue to grow in the foreseeable future as the telecommunication industry continues to offer more sophisticated and advanced services. Orthogonal frequency division multiplexing (OFDM) is a technology that promises to meet these transmission demands. Since the last decade, OFDM has attracted considerable attention. The main reason for this interest is the substantial advantage it offers in high rate transmissions over frequency selective fading channels like robustness to multi-path fading and capability to control the data rate according to the transmission channel [2]. OFDM effectively divides a wide band frequency selective fading channel into a large number of narrow band flat fading channels over which parallel data streams are transmitted thereby increasing the symbol duration. The insertion of a cyclic pre-

fix (CP), of adequate length, in the transmission symbol reduces the inter symbol interference (ISI). The CP, which is a cyclic extension of the IFFT output, has to be at least as long as the channel impulse response (CIR) in order to avoid ISI[1]. This also enables the OFDM system to have simple receiver structure utilizing a frequency-domain equalizer (FEQ) with only one complex multiplication per subcarrier to mitigate frequency selectivity. As such, OFDM has found wide acceptance and application. It is already a part of many digital communication standards and is being used the world over. OFDM has been selected as the physical layer of choice for broadband wireless communications systems ([2], [3], [4], [5], [6]).

## 2.1 System Model

Consider a sequence of  $T + 1$  data symbols  $\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_T$ , each of length  $N$ , to be transmitted in an OFDM system. Every symbol  $\mathbf{x}_i$ , undergoes an IFFT operation to produce the time domain symbol

$$\mathbf{x}_i = \sqrt{N}Q^* \mathbf{x}_i \quad (2.1)$$

where  $Q$  is the  $N \times N$  FFT matrix. In order to counter the effect of ISI, a length  $P$  CP  $\underline{\mathbf{x}}_i$  is appended to the symbol  $\mathbf{x}_i$ , which results in the super symbol  $\bar{\mathbf{x}}_i$ , each of length  $N + P$ . The CP serves to mitigate the multi-path effect but the estimation of channel characteristics of fading channels require densely spaced pilot tones

specially for those channels with a small coherence bandwidth [32]. Figure 2.1 shows the basic elements of an OFDM transmitter.

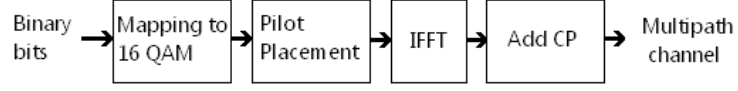


Figure 2.1: OFDM Transmitter Block Diagram.

Let  $\underline{h}_i$  be the channel of maximum length  $P + 1$ . We consider a block fading model and assume that the channel remains unchanged for each super-symbol but varies from one super-symbol to the next according to the following state space model.

$$\underline{h}_{i+1} = \mathbf{F}\underline{h}_i + \mathbf{G}\underline{u}_i \quad (2.2)$$

where  $\underline{h}_o \sim \mathcal{N}(0, \Pi_o)$  and  $\underline{u}_o \sim \mathcal{N}(0, \sigma_u^2)$ . the matrices  $\mathbf{F}$  and  $\mathbf{G}$  are a function of the doppler spread, the power delay profile (frequency correlation), and the transmit filter. The matrices are given as

$$\mathbf{F} = \begin{bmatrix} \alpha(0) & & \\ & \ddots & \\ & & \alpha(P) \end{bmatrix} \quad \text{and} \quad \mathbf{G} = \begin{bmatrix} \sqrt{1 - \alpha^2(0)} & & \\ & \ddots & \\ & & \sqrt{(1 - \alpha^2(P))e^{-\beta P}} \end{bmatrix}$$

$\alpha(p)$  is related to the Doppler frequency  $f_D(p)$  by  $\alpha(p) = J_0(2\pi f_D T(p))$ . The variable  $\beta$  corresponds to the exponent of the channel decay profile while the

factor  $\sqrt{(1 - \alpha^2(p))e^{-\beta p}}$  ensures that each link maintains the exponential decay profile ( $e^{-\beta p}$ ) for all time. We assume this information is known at the receiver. The model thus captures both frequency and time correlation.

The passage of  $\bar{\mathbf{x}}_k$  symbols through the channel  $\underline{\mathbf{h}}$ , produces the received sequence  $\bar{\mathbf{y}}_k$  at the receiver. The received packet (of length  $N + P$ ) is split into a length  $N$  packet  $\mathbf{y}_k$  and a length  $P$  prefix  $\underline{\mathbf{y}}_k$ . The prefix absorbs all the ISI present between the  $\bar{\mathbf{x}}_{k-1}$  and  $\bar{\mathbf{x}}_k$  packets and is hence discarded. The time domain relation of the input and the output can be expressed as

$$\mathbf{y}_i = \mathbf{x}_i \otimes \mathbf{h}_i + \mathbf{n}_i \quad (2.3)$$

where  $\otimes$  denotes convolution. Equation (2.3) takes a more transparent form in the frequency domain as

$$\mathcal{Y}_i = \text{diag}(\mathcal{X}_i)\mathcal{H}_i + \mathcal{N}_i \quad (2.4)$$

or

$$\mathcal{Y}_i = \text{diag}(\mathcal{X}_i)\mathbf{Q}_{P+1}\underline{\mathbf{h}}_i + \mathcal{N}_i \quad (2.5)$$

The relationship in (2.5) follows from the FFT relationship

$$\mathcal{H}_i = \mathbf{Q} \begin{bmatrix} \underline{\mathbf{h}}_i \\ 0 \end{bmatrix} = \mathbf{Q}_{P+1}\underline{\mathbf{h}}_i \quad (2.6)$$



where  $\mathbf{Q}_{P+1}$  consists of the first  $P + 1$  columns of  $\mathbf{Q}$ . Alternatively, with

$$\mathbf{X}_i \triangleq \text{diag}(\boldsymbol{\alpha}_i)\mathbf{Q}_{P+1} \quad (2.7)$$

we can write

$$\mathbf{y}_i = \mathbf{X}_i\mathbf{h}_i + \mathcal{N}_i \quad (2.8)$$

which is no longer diagonal. We will discuss the disadvantage of this decoupled relationship in this part.

## 2.2 Literature Review

As mentioned in the introduction, our aim is to design an algorithm for channel estimation in OFDM. In this section, we will take a look at the literature relating to channel estimation in OFDM systems. We will provide an overview of the various approaches to channel estimation and the different constraints assumed on channel and data.

The availability of an accurate channel transfer function estimate is one of the prerequisites for coherent symbol detection in an OFDM receiver. Numerous research contributions have appeared in literature on the topic of channel estimation, in recent years. One way to classify these works is as according to the method used for channel estimation (training based, semi blind, blind and data aided). Another approach to classify these algorithms is based on the constraints

used for channel and data recovery.

### **2.2.1 Channel Estimation using Pilots**

One technique for channel estimation is to use pilots. As equalization requires channel state information (CSI), pilots on predetermined subcarriers are sent as training signals in OFDM systems, and the channels for pilot subcarriers are directly estimated, while those for non pilot subcarriers need to be estimated through interpolation with the channel estimates from adjacent pilot subcarriers [7], [8], [9], [10], [11], [12]. This in turn, is achieved at the cost of a reduction in the number of useful subcarriers available for data transmission. In [13], the authors have developed a channel estimator by introducing an extended channel and its finite impulse response approximation.

### **2.2.2 Blind Channel Estimation**

Since the number of pilots must be greater than the number of channel taps, the use of cyclic prefix (CP) and pilot symbols entails a significant bandwidth loss, motivating blind methods. Several works have attempted to perform blind channel estimation in OFDM. The authors in [14] explored transmitter redundancy for blind channel estimation while in [15], a blind identification exploiting receiver diversity which can get CSI during one OFDM symbol was investigated. In [17] the authors present a fast converging blind channel estimator for OFDM-systems based on the Maximum Likelihood principle. A non redundant precoding along

with cyclic prefix was explored in [18]. In [19], second-order cyclostationary statistics and antenna precoding are used while [20] employs finite-alphabet constraint for blind channel estimation. The authors in [21] suggest an approach which relies on the i.i.d. assumption of the data sequence and uses the cyclic prefix redundancy present in OFDM systems and [23] developed an a posteriori probability based two dimensional channel estimation algorithm.

### **2.2.3 Semi Blind Channel Estimation**

In semi blind methods, both the pilots and natural constraints are used for channel estimation([24], [25]). In [26] a semi-blind channel estimation using receiver diversity is proposed for OFDM systems in the presence of virtual carriers. The authors in [27] employed a semiblind channel estimation method using selected channel parameter estimation and error reduction algorithms. The work presented in [46] proposes a pilot aided algorithm for frequency domain channel estimation for a single-user and multiple receiving antennas system in the presence of synchronous interference while the authors in [28] used delay sub-space based approach for channel estimation. In [29], coding along with pilots was used for channel estimation. Similarly other works have explored various other semi-blind techniques for channel estimation. Coding and cyclic prefix were investigated for channel estimation in [31]. Authors of [32] used interpolated LS by applying phase shifted samples while [33] proposed to include a phase rotation term in the frequency domain interpolation.

## 2.2.4 Data Aided Channel Estimation

The motivation behind estimating the channel response is to recover the data being transmitted. The detected data can be, in turn, used to improve the channel estimate, thus giving rise to an iterative method for channel and data recovery. Several works have explored this idea of joint data and channel estimation ([30], [48], [49], [50], [52], [53], [39], [43], [54], [45], [58], [59]). A data aided approach seems most appropriate for channel estimation as it makes a collective use of data and channel constraints for estimation.

## 2.2.5 Constraints Used in Channel Estimation/Data Detection

All the works mentioned earlier, use a subset of the following constraints on the channel estimate or data, regardless of the estimation technique used. Following is a survey of these constraints and the work that employs them.

### Data Constraints:

**Finite alphabet constraint:** Data is usually drawn from a finite alphabet set. The authors in [20], [36] and [39] make use of this constraint.

**Code:**Data usually exhibits some form of redundancy like a code that helps reduce the row probability or err [23], [31], [50].

**Transmit precoding:** The data might also contain some form of precoding (to facilitate equalization at the receiver) such as a cyclic prefix, silent guard bands [47], [51] and known symbol precoding [61].

**Pilots:** Pilots represent the most primitive form of redundancy and are usually inserted to perform channel estimation or simply to initialize the estimation process [12], [8], [9], [29], [45], [13], [28].

**Channel Constraints:**

**Finite delay spread:** The channel is usually of finite impulse response with a maximum delay spread that is assumed available to the receiver.

**Sparsity:** the sparsity of a multipath fading channel is defined as the ratio of the time duration spanned by the multipaths to their number [16], [35], [40]. the number of paths and their delays are usually stationary. However, their amplitudes and relative phases usually vary much more rapidly with time. this essentially reduces the number of parameters to be estimated to that of the number of multipaths in the channel.

**Frequency correlation:** In addition to information about which of the channel taps are inactive, we usually have additional statistical information about the active ones. Thus, it is usually assumed that the taps are Gaussian ( zero mean or not depending on whether the channel exhibits Rayleigh or Rician fading) with a certain covariance matrix. this matrix is a measure of the frequency correlation among the taps [38], [55].

**Time correlation:** As channels vary with time, they exhibit some form of time correlation. time-variant behavior could also be more structured, e.g., following a state-space model [48], [49], [57].

**Uncertainty information:** Channel also suffers from non ideal effects such as nonlinearities and rapid time-variations that are difficult to model. The aggregate effect of this non ideal behavior could be represented as uncertainty information that can be used to build robust receivers [22].

Regardless of the approach used for channel estimation or the constraints employed, estimation can be carried out in any of the two domains (time and frequency). Below, we classify the approaches that are used in either of these two domains. We also discuss the advantages and disadvantages of estimation in these domains. All these methods for channel estimation are either in the frequency domain or in the time domain. Below is a survey of various works in the two domains.

## 2.2.6 Time Domain Channel Estimation

A lot of researchers have opted for channel estimation in the time domain. A joint carrier frequency synchronization and channel estimation scheme using the expectation-maximization (EM) approach is proposed in [41]. A time domain minimum mean square error (MMSE) channel estimation technique based on subspace tracking for OFDM system is put forward in [42]. In [43], a joint channel and data estimation algorithm is presented which makes a collective use of data and channel constraints. A simplified joint frequency-offset and channel estimation technique for Multi-Symbol Encapsulated MSE OFDM system is proposed in [24], while authors in [27] present a sequential method for channel response estimation based on

Carrier Frequency Offset and symbol timing estimation by exploiting the structure of the packet preamble of IEEE 802.11a standard. The authors in [44] take a statistical approach and estimate the channel based on Power Spectral Density (PSD) and LS estimation for OFDM systems with timing offsets. An iterative receiver structure with joint detection and channel estimation based on implicit pilots is proposed in [45] and [46] presents a pilot aided channel estimation algorithm in the presence of synchronous noise by exploiting the a priori available information about the interference structure.

### **2.2.7 Frequency Domain Channel Estimation**

In the past years, various techniques for channel estimation in the frequency domain have also been explored. Researchers in [32] apply phase shifted samples in the frequency-domain to an interpolated LS to estimate the channel while in [33], the authors propose to include a phase rotation term in the frequency domain interpolation for better CIR window location. Channel estimation using polynomial cancelation coding (PCC) training symbols and frequency domain windowing is proposed in [34]. A sub-band approach to channel estimation and channel equalization is proposed in [37]. A low-complexity iterative channel estimator is proposed in [30]. The minimum mean square error (MMSE) channel estimation in the frequency domain is considered in [38] while researchers in [28] present delay subspace-based channel estimation techniques for OFDM systems over fast-fading channels.

## 2.3 Disadvantage of Performing Channel Estimation in Time Domain

Most channel estimation algorithms for OFDM presented in literature perform estimation in the time domain (instead of the frequency domain) [13], [41], [43], [24], [27]. By performing estimation in the time domain, one can decrease the degrees of freedom from  $N$ , the number of frequency bins, to  $P + 1$ , the number of (time domain) channel taps. This is a drastic reduction since the number of channel taps is usually less than the cyclic prefix which is usually designed to be less than  $\frac{N}{4}$ . The reduction in the parameter estimation space in turn results in improved estimation accuracy.

There is a certain price that we have to pay, however, for this gain. We lose the diagonal structure of the channel by performing the estimation in the time domain. Thus, instead of frequency domain relationship (2.4) in which the various equations are decoupled, we employ the time-frequency relationship (2.8) which is no more diagonal (decoupled). This loss in transparency in return complicates channel estimation and makes it more computationally complex. For example, while the estimation in (2.4) is performed on a bin by bin basis according to

$$\hat{\mathcal{H}}(l) = \frac{\mathcal{Y}(l)}{\mathcal{X}(l)} \quad l = 1, 2, \dots, N \quad (2.9)$$



channel estimation in (2.8) requires size  $P + 1$  matrix inversion

$$\hat{\underline{\mathbf{h}}} = (\mathbf{X}^* \mathbf{X})^{-1} \mathbf{X}^* \mathbf{y} \quad (2.10)$$

where  $(\mathbf{X}^* \mathbf{X})$  is invertible. Moreover, since data detection is best performed in the frequency domain, estimating the channel in the time domain makes it necessary to perform an extra IFFT operation (to obtain the frequency domain estimate  $\mathcal{H}$  from the time domain estimate  $\hat{\underline{\mathbf{h}}}$  and use it for data detection). Thus, for data-aided channel estimation techniques, each channel estimation step would require one such IFFT operation.

Apart from the computational complexity, performing channel estimation in the time domain might be over solving a problem. For example, in multiple access OFDM systems, like WiMAX, users are not interested in the whole frequency spectrum, but only in that part of the spectrum in which they are operating. In fact these users don't have access to the whole spectrum but only a part of it is available to them. Moreover, even if some users were interested in estimating the whole spectrum, many standards would not be able to support that as there are not enough pilots to do so.

## 2.4 Can We Perform Channel Estimation Reliably in Frequency Domain?

Channel estimation in the frequency domain avoids the previously mentioned disadvantages. Moreover, the structure that characterizes the estimation problem in the time domain continues to characterize the estimation problem in the frequency domain. Specifically, the time and frequency correlation exhibited by the time domain channel maps in to corresponding correlation of the channel frequency response.

The only problem with channel estimation in the frequency domain is the increase in the number of parameter to be estimated [38]. If we can reduce the parameter estimation space, then we can avoid the one disadvantage of frequency domain estimation as compared to time domain estimation. The frequency response of the channel is inherently limited by the degrees of freedom of the time domain impulse response. How does this limited degree of freedom manifests itself in the frequency domain? Figure 2.2 demonstrates the length 64 frequency response resulting from a 16 tap channel with exponential decay profile similar to the one we employ in our simulations. Note that within a narrow enough band of spectrum, the spectrum looks linear or quadratic. As such, we employ model reduction in this work to estimate the spectrum, thereby reducing the number of parameters to be estimated.

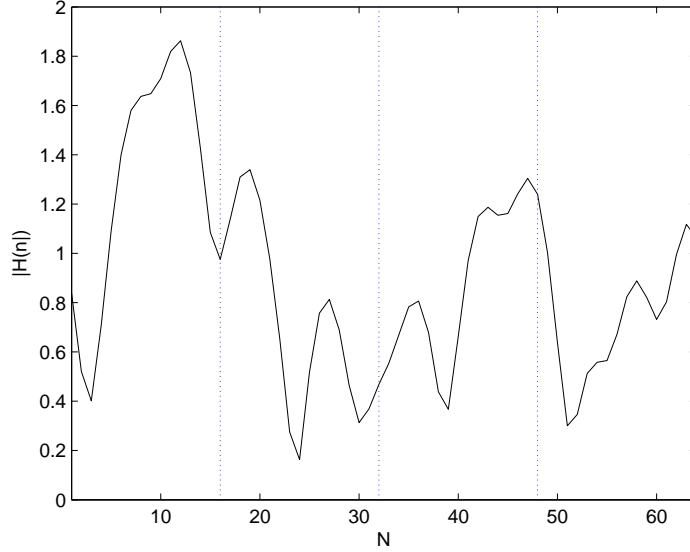


Figure 2.2: CIR in the frequency domain partitioned in four subchannels.

## 2.5 Input/Output Relationship in the Frequency Domain

The input/output relationship of the OFDM system is best described in the frequency domain. A frequency domain channel response of length  $N$  is shown in figure 2.2. We start by partitioning the channel response into a number of sections each of length  $L_f$  producing a total of  $\lceil \frac{N}{L_f} \rceil$  sections<sup>1</sup>. Let us denote the  $j^{\text{th}}$  section of the frequency response by  $\underline{\mathcal{H}}^{(j)}$ . The input/output equation for this section is given by

$$\underline{\mathbf{y}}_i^{(j)} = \text{diag}(\underline{\mathbf{x}}_i^{(j)}) \underline{\mathcal{H}}_i^{(j)} + \underline{\mathcal{N}}_i^{(j)} \quad (2.11)$$

---

<sup>1</sup>In a multi-access OFDM system, we can choose the section length to be the number of carriers allocated to each user. However, the sections need not have equal length over the frequency response.

Where  $\underline{\mathbf{y}}^{(j)}$ ,  $\underline{\mathbf{x}}^{(j)}$ ,  $\underline{\mathbf{h}}^{(j)}$  and  $\underline{\mathcal{N}}^{(j)}$  are the  $j^{\text{th}}$  sections of  $\mathbf{y}$ ,  $\mathbf{x}$ ,  $\mathbf{h}$  and  $\mathcal{N}$  respectively. From now onwards, we will drop the dependence on  $j$  for notational convenience. Equation (2.11) can now be written as

$$\underline{\mathbf{y}}_i = \text{diag}(\underline{\mathbf{x}}_i)\underline{\mathbf{h}}_i + \underline{\mathcal{N}}_i \quad (2.12)$$

where  $\underline{\mathcal{N}}_i \sim \mathcal{N}(0, \sigma_i^2 I)$  is the additive white gaussian noise.

## 2.6 Pilot/Output Relationship in Frequency Domain

In general, the receiver needs pilots to obtain a channel estimate. The pilot locations within the OFDM symbol are denoted by the index set  $I_p = i_1, i_2, \dots, i_{L_p}$ . Also, let  $\text{diag}(\underline{\mathbf{x}}_{I_p})$  denote the matrix  $\text{diag}(\underline{\mathbf{x}})$  pruned of the rows that do not belong to  $I_p$ . Then, the pilot/output equation can be derived from the input/output relation (2.12) as

$$\underline{\mathbf{y}}_{I_p} = \text{diag}(\underline{\mathbf{x}}_{I_p})\underline{\mathbf{h}} + \underline{\mathcal{N}}_{I_p} \quad (2.13)$$

## 2.7 A Parameter Reduction Approach

The main hindrance in performing channel estimation in the frequency domain, as opposed to the time domain estimation, is the increased number of parameters

to be estimated. Our goal here is to apply some parameter reduction technique to reduce the number of frequency domain parameters to be estimated. Dropping the dependence on time index  $i$  for notational convenience, we consider that  $\underline{\mathcal{H}}$  can be expressed as

$$\underline{\mathcal{H}} = \mathbf{V}_p \boldsymbol{\alpha}_d \quad (2.14)$$

where  $\mathbf{V}_p$  is a known matrix and  $\boldsymbol{\alpha}_d$  is the vector of parameters to be determined. We will consider two different approaches for estimating  $\underline{\mathcal{H}}$ . One approach is to consider a linear or quadratic approximation which we will discuss in chapter 3. Another way to go about solving for  $\underline{\mathcal{H}}$  is based on Eigenvalue decomposition and is discussed in chapter 4.

## 2.8 Conclusion

In this chapter, we have given a brief literature survey of the channel estimation problem. The frequency domain system model is given and advantages and disadvantages of performing channel estimation in time and frequency domains were also discussed. We introduced the parameter reduction model in order to model the channel in the frequency domain. In the next chapter we will take a detailed look at the parameter reduction model.

## CHAPTER 3

# INTERPOLATION BASED FREQUENCY DOMAIN ESTIMATION

The problem that we encounter when performing channel estimation in frequency domain is the increased number of parameters to be estimated. For frequency domain estimation, we require to estimate  $N$  parameters while in the case of time domain estimate, we only need to estimate  $P + 1$  parameters. We can eliminate this disadvantage if we can find a way to decrease the parameter estimation space for the frequency domain estimation, such that it is comparable to the number of parameters needed for time domain estimation. The frequency response of the channel is inherently limited by the degrees of freedom of the time domain impulse response. Figure 3.1 shows a length 64 frequency response of a 16 tap channel with an exponential delay profile similar to the one that will be use in

simulations. We can see that with in a narrow enough band width, the spectrum can be approximated as linear or quadratic. Mathematically speaking, let  $\mathcal{H}(k)$  be a sub band of the frequency spectrum of width  $L_f$ (where  $k = 1, 2, \dots, \lfloor \frac{N}{L_f} \rfloor$ ). If the frequency spectrum is linear in this sub band, then we can write

$$\underline{\mathcal{H}}(k) = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & L_f \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \quad (3.1)$$

If the spectrum is quadratic, we can write

$$\underline{\mathcal{H}}(k) = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2^2 \\ \vdots & \vdots & \vdots \\ 1 & L_f & L_f^2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} \quad (3.2)$$

In general, we can write

$$\underline{\mathcal{H}}(k) = \mathbf{V}_p \boldsymbol{\alpha}_d \quad (3.3)$$

where  $\mathbf{V}_p$  is the interpolation matrix and  $\boldsymbol{\alpha}_d$  is the vector of interpolation parameters.

The input/output relation is given by equation(2.12). Replacing  $\underline{\mathcal{H}}$  from equa-

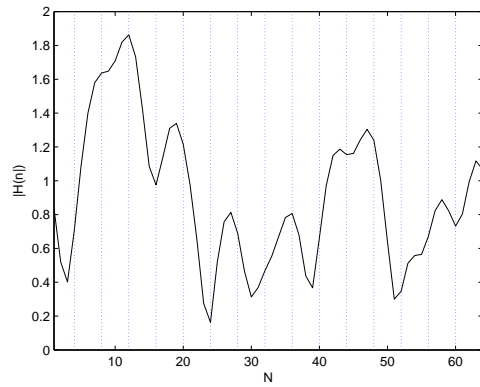
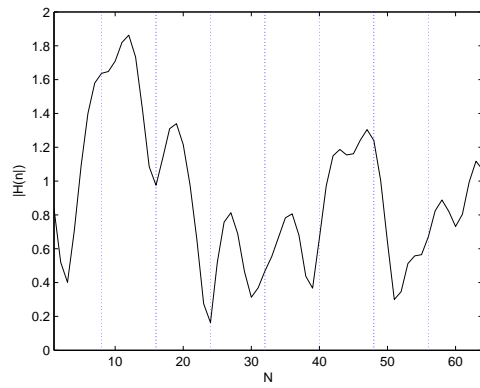
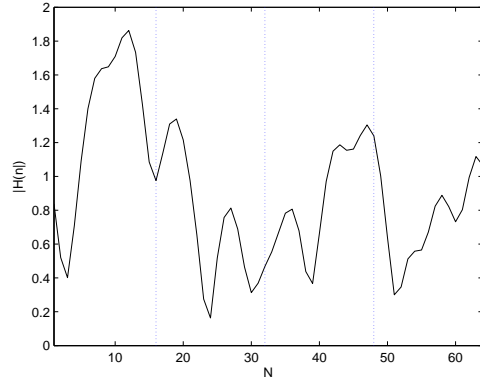


Figure 3.1: CIR in the frequency domain divided into 4, 8 and 16 parts.



tion (3.3) results in

$$\begin{aligned}\underline{\mathbf{y}} &= \text{diag}(\underline{\mathbf{x}})\mathbf{V}_p\boldsymbol{\alpha}_d + \underline{\mathcal{N}} \\ &= \underline{\mathbf{X}}\boldsymbol{\alpha}_d + \underline{\mathcal{N}}\end{aligned}\tag{3.4}$$

where  $\underline{\mathbf{X}} = \text{diag}(\underline{\mathbf{x}})\mathbf{V}_p$  and  $\underline{\mathcal{N}}$  is zero mean white gaussian noise. Pruning the above equation yields

$$\underline{\mathbf{y}}_{I_p} = \underline{\mathbf{X}}_{I_p}\boldsymbol{\alpha}_d + \underline{\mathcal{N}}_{I_p}\tag{3.5}$$

### 3.1 Least Squares

The solution of equation (3.5) is based on minimizing

$$\hat{\boldsymbol{\alpha}}_d = \arg \min_{\boldsymbol{\alpha}_d} \{ \|\underline{\mathbf{y}}_{I_p} - \underline{\mathbf{X}}_{I_p}\boldsymbol{\alpha}_d\|^2 \}\tag{3.6}$$

Solving it as a least square problem [62], yields

$$\hat{\boldsymbol{\alpha}}_d = (\underline{\mathbf{X}}_{I_p}^* \underline{\mathbf{X}}_{I_p})^{-1} \underline{\mathbf{X}}_{I_p}^* \underline{\mathbf{y}}_{I_p}\tag{3.7}$$

The estimate of the  $j^{\text{th}}$  section of the channel is thus given by

$$\underline{\mathbf{h}} = \mathbf{V}_p \hat{\boldsymbol{\alpha}}_d\tag{3.8}$$

Concatenation of all  $M$  such section, will give us the complete channel response  $\mathcal{H}$ .

## 3.2 Simulation Parameters

Consider an OFDM system where an iid sequence of  $T + 1$  data symbols  $\mathbf{x}_o^T$  are to be transmitted. The length of each symbol,  $N$ , is 64. We use a CP of length 15. The modulation scheme used is 16 QAM with grey coding. The channel impulse response(CIR) is considered to consist of 16 complex taps(maximum length allowable for the channel with a CP length of 15). The exponential decay profile  $E[|h_0(k)|^2]$  of the channel remains fixed over any OFDM symbol and is taken to be  $e^{-0.2k}$ . These parameters are used throughout the simulations.

## 3.3 Effect of Number of Pilots

By intuition, we know that increasing the number of pilots should yield a better channel estimate and hence better BER performance. Figure 3.2 is plotted for 16 and 32 pilots. In both cases, we use 2 interpolation parameters and 8 sections. As evident from the figure, increasing the number of pilots will lead to a better channel estimate.

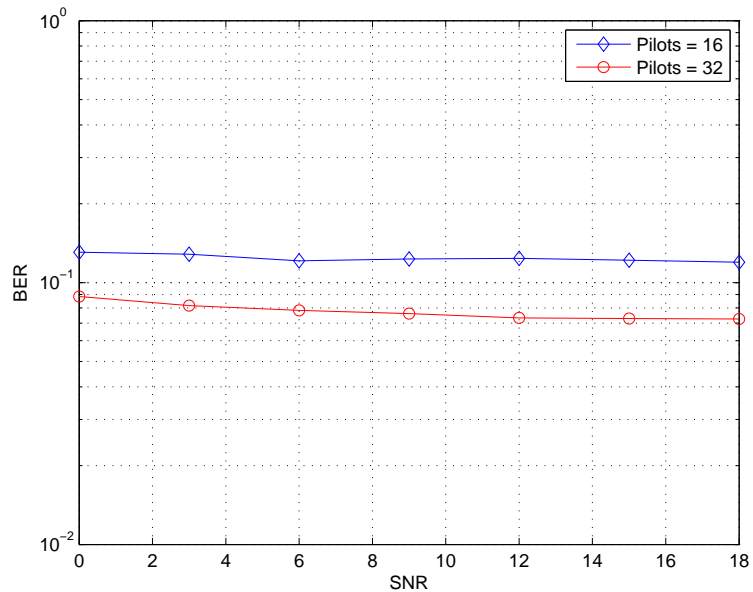
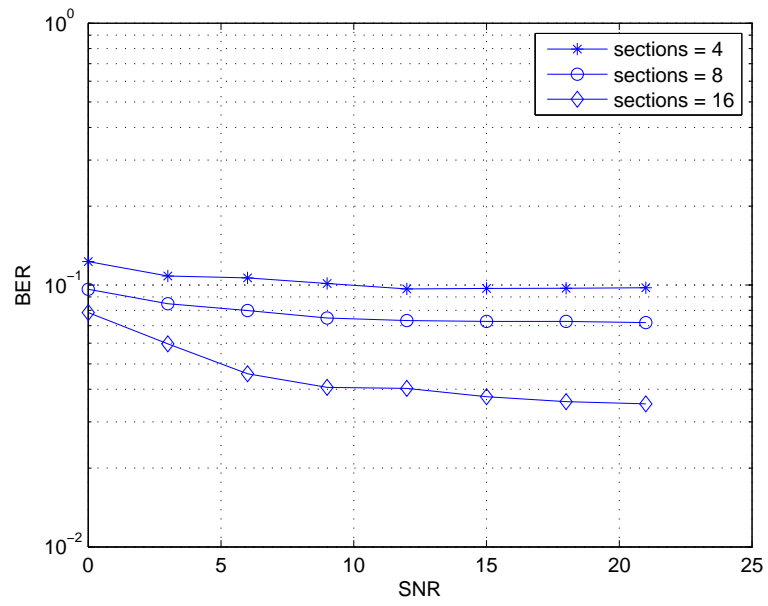


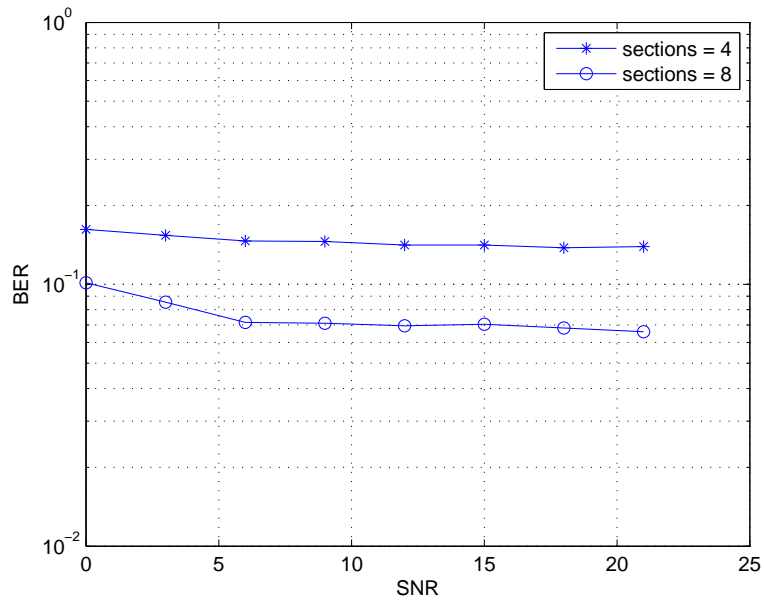
Figure 3.2: BER curve for 8 sections 2 parameters

### 3.4 Effect of Section Length

Another way to improve the channel estimate would be to divide the channel into a larger number of sections. We employ 32 pilots, 2 interpolation parameters and divide the channel into 4, 8 and 16 sections respectively. The BER curves for these three cases are shown in figure 3.3(a). We can see that decreasing the section length, i.e. increasing the number of section per channel, results in a better BER performance. Figure 3.3(b) shows BER performance for 32 pilots and 3 parameters and shows the same trend. So for a better channel estimate, we should use a larger number of sections.



(a)



(b)

Figure 3.3: (a) BER curve for 32 pilots and 2 parameters (b) BER curve for 32 pilots and 3 parameters.

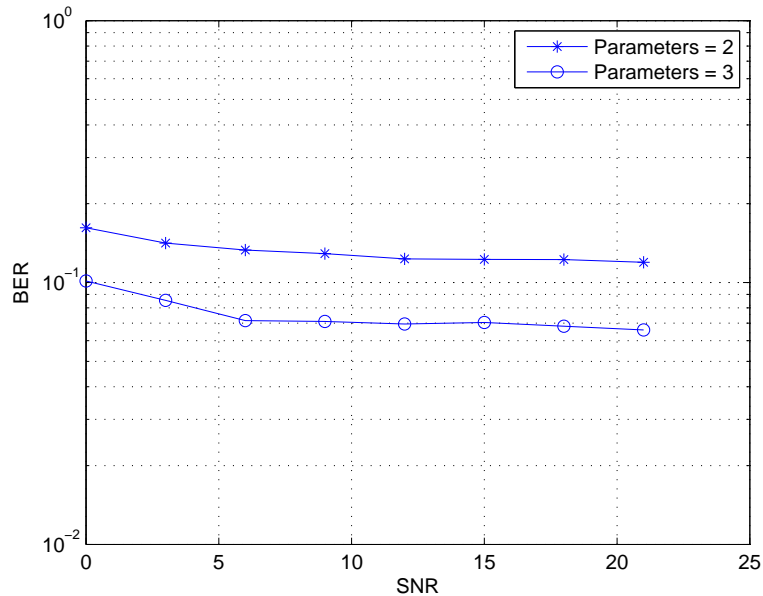


Figure 3.4: BER curve using 32 pilots and 8 sections

### 3.5 Effect of Varying the Number of Parameters

The channel estimate is also affected by varying the number of estimation parameters. We plot the BER of the system using 2 and 3 interpolation parameters. In both cases, we use 32 pilots and divided the channel into 8 sections. Figure 3.4 shows the effect of changing the number of parameter on the BER performance. As we can see, increasing the number of parameters from 2 to 3, results in improved BER performance specially at high SNR. So increasing the number of interpolation parameters improves the channel estimate. For figures 3.2-3.4 above, we conclude that

- Increasing the number of pilots improves the channel estimate.
- Increasing the number of sections in which the frequency domain channel response is divided, improves the channel estimate.

- Increasing the number of interpolation parameters improves the channel estimate.

However, there is a limit to the extent we can increase these parameters. Increasing the number of pilots means fewer carriers are available for data transmission purpose. The number of sections and the number of interpolation parameters are in turn both limited by the number of pilots we use. For the Least Square solution of equation (3.5), requires the following condition to be fulfilled

$$\text{number of pilots in each section} \geq \text{number of interpolation parameters} \quad (3.9)$$

So if use 32 pilots and 2 interpolation parameters, then every section of the channel response must have at least 2 pilots. That means that at most we can divide the channel response into 16 sections. If we increase the number of interpolation parameters to 3, than each section must have at least 3 pilots. In this case channel response can be divided into a maximum of 8 sections.

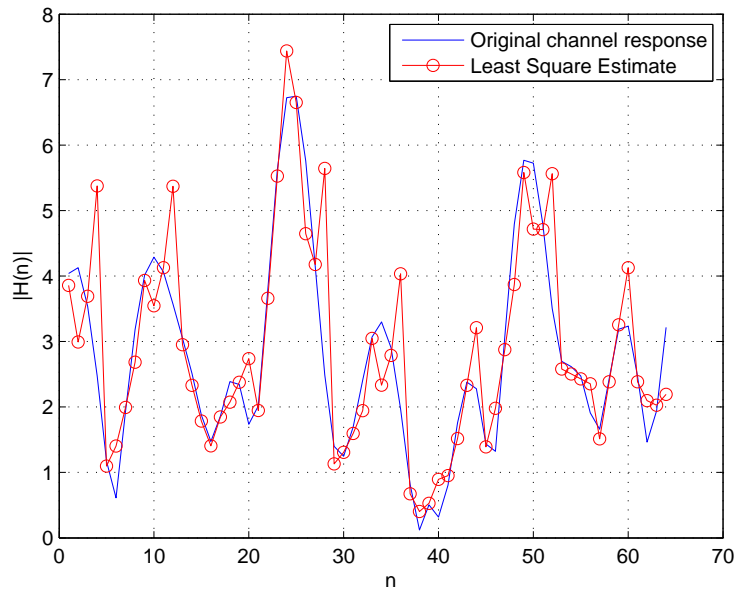
This limitation can be avoided if we use a regularized Least Square solution for equation (3.5). In that case we can have as many sections and interpolation parameters as we want as long as there is at least one pilot per section. i.e.

$$\text{number of pilots in each section} \geq 1 \quad (3.10)$$

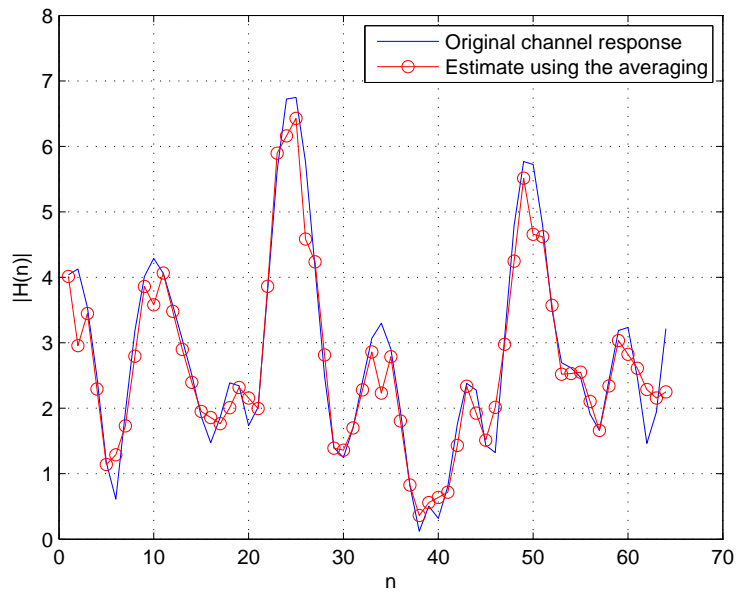
## 3.6 A Scheme to Improve the Channel Estimate

The interpolation method we use in this chapter is polynomial-based. As such, we expect the point at the edge of each section to be inflated as shown in figure 3.5(a). The first figure is for 2 interpolation parameters and the second is for 3 parameters. both have 32 pilots and 16 sections. If we can somehow correct these inflated points, our estimate is bound to improve. In order to reduce this error, we use an **Averaging Scheme**. This scheme sets the estimate of the edge point of each section to be the average of the second last point of the current section and the first point of the next section.

Figures 3.5(a) and 3.5(b) show the original LS based channel estimate( 32 pilots/2 parameters/16 sections) with inflated points and compares it with the averaging scheme. We can see that using the averaging scheme improves the channel estimate. Let us compare the performance of the two methods to get a better idea of the advantage offered by the averaging scheme. We consider the case of 32 pilots. Figures 3.6(a) and 3.6(b) show the Error and BER plots for the two schemes using 3 interpolation parameters, 8 sections per channel and figures 3.7(a) and 3.7(b) show the same plots for 2 interpolation parameters, 16 sections per channel. It is evident that the averaging scheme outperforms the LS solution but it does so at the cost of added computational complexity. Also an interesting observation is that the averaging scheme performs better in the case of 16 sections. This is logical as the later case has more number of sections and more edge points will be corrected by the averaging scheme than in the former case.



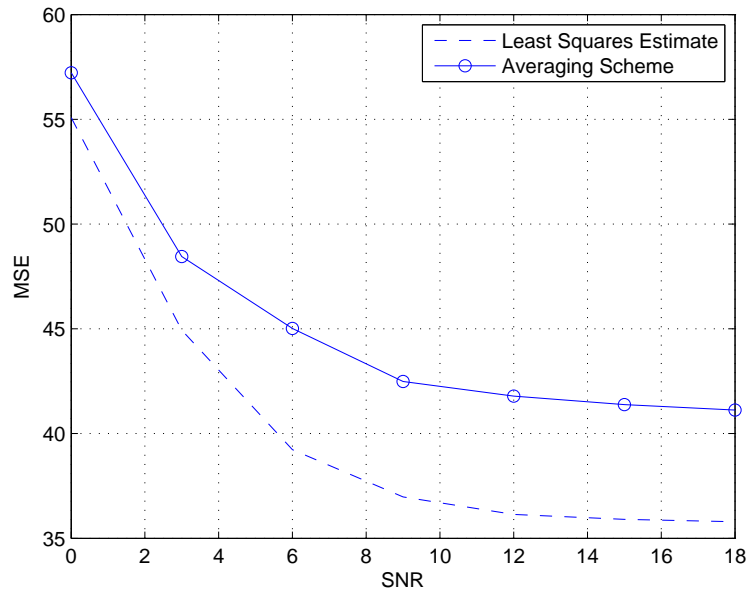
(a)



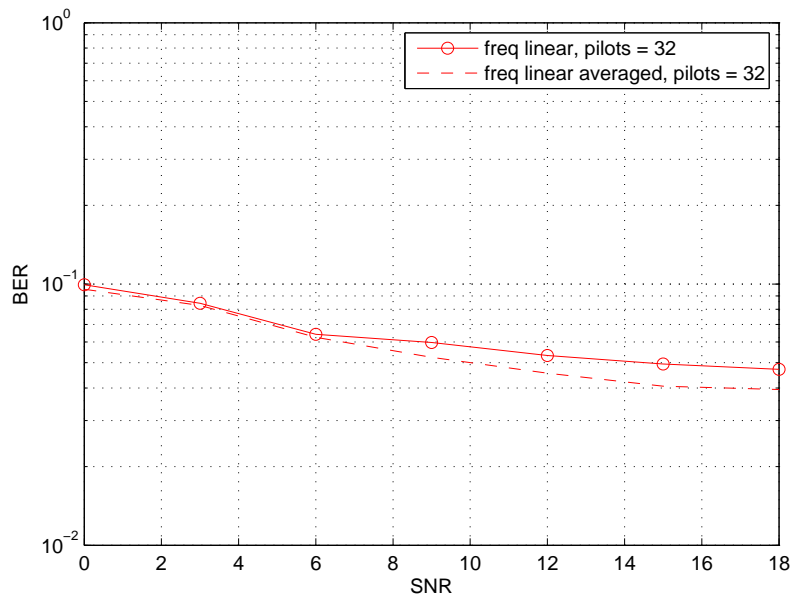
(b)

Figure 3.5: (a) CIR inflated at the edges(32/2/16) (b) Removing inflation using averaging scheme.



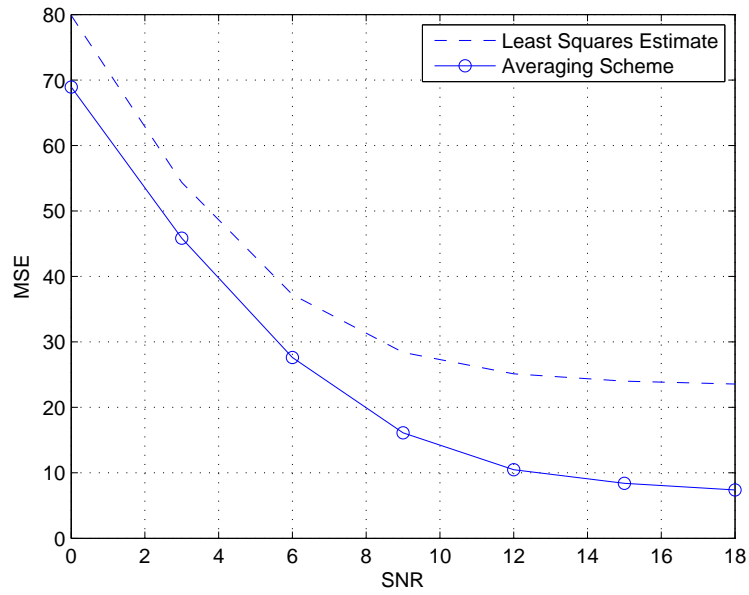


(a)

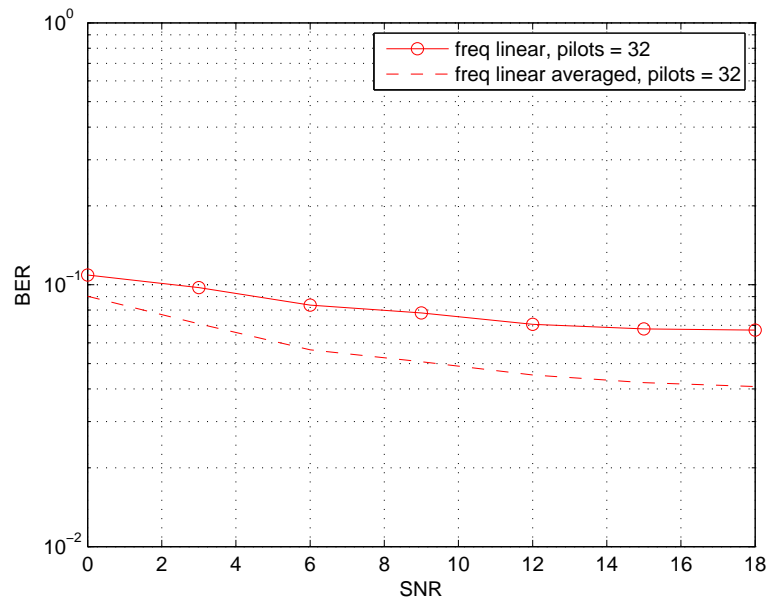


(b)

Figure 3.6: (a) Error plot for LS and averaging scheme for (32/3/8) (b) BER plots for LS and averaged scheme for (32/3/8)



(a)



(b)

Figure 3.7: (a) Error plot for LS and averaging scheme for (32/2/16) (b) BER plots for LS and averaged scheme for (32/2/16).

### 3.7 Least Squares with Regularization

Ideally, we estimate  $\boldsymbol{\alpha}_d$  using some I/O relationship by maximizing the corresponding log-likelihood function

$$\hat{\boldsymbol{\alpha}}_d = \arg \max_{\boldsymbol{\alpha}_d} \{ \ln p(\underline{\mathbf{y}} | \underline{\mathbf{X}}, \boldsymbol{\alpha}_d) + \ln p(\boldsymbol{\alpha}_d) \}$$

When the channel obeys the I/O relationship (3.4) (so that  $\ln p(\underline{\mathbf{y}} | \underline{\mathbf{X}}, \boldsymbol{\alpha}_d) = -\|\underline{\mathbf{y}} - \underline{\mathbf{X}}\boldsymbol{\alpha}_d\|_{\sigma_n^{-2}}^2$  up to some additive constant  $\ln p(\boldsymbol{\alpha}_d) = -\|\boldsymbol{\alpha}_d\|_{R_{\alpha_d}^{-1}}^2$  up to some additive constant), then the LS estimate is given by

$$\hat{\boldsymbol{\alpha}}_d = \arg \max_{\boldsymbol{\alpha}_d} \left\{ \|\underline{\mathbf{y}}_{I_p} - \underline{\mathbf{X}}_{I_p} \boldsymbol{\alpha}_d\|_{\sigma_n^{-2} \mathbf{I}}^2 + \|\boldsymbol{\alpha}_d\|_{R_{\alpha_d}^{-1}}^2 \right\} \quad (3.11)$$

where  $\sigma_n^2$  is the noise variance. The estimate of  $\boldsymbol{\alpha}_d$  that minimizes the MSE is given by

$$\hat{\boldsymbol{\alpha}}_d = \mathbf{R}_{\alpha_d} \underline{\mathbf{X}}_{I_p}^* [\sigma_n^{-2} \mathbf{I} + \underline{\mathbf{X}}_{I_p} \mathbf{R}_{\alpha_d} \underline{\mathbf{X}}_{I_p}^*]^{-1} \underline{\mathbf{y}}_{I_p} \quad (3.12)$$

We will assume that  $\mathbf{R}_{\alpha_d}$  is the identity matrix, that is we assume no correlation between parameters. The advantage of regularized LS solution is that there is no restriction on the number of pilots per section and they can be less than the number of parameters. This allows us to try those combinations of interpolation parameters, number of pilots and number of sections which are not possible in the

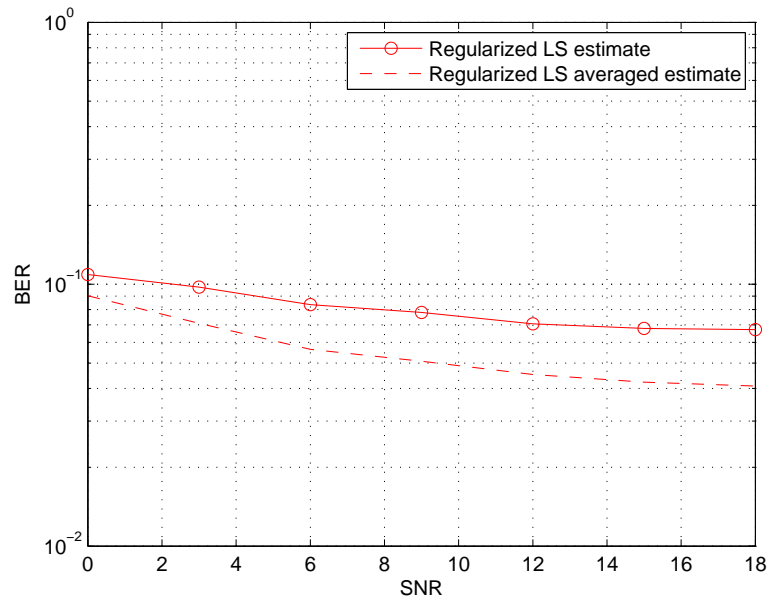
---

<sup>1</sup>where  $\boldsymbol{\alpha}_d$  is a column vector and the notation  $\|\boldsymbol{\alpha}_d\|_{R_{\alpha_d}^{-1}}^2$  refers to the weighted Euclidean norm of  $\boldsymbol{\alpha}_d$ . i.e,  $\text{vb} \|\boldsymbol{\alpha}_d\|_{R_{\alpha_d}^{-1}}^2 = \boldsymbol{\alpha}_d^* R_{\alpha_d}^{-1} \boldsymbol{\alpha}_d$

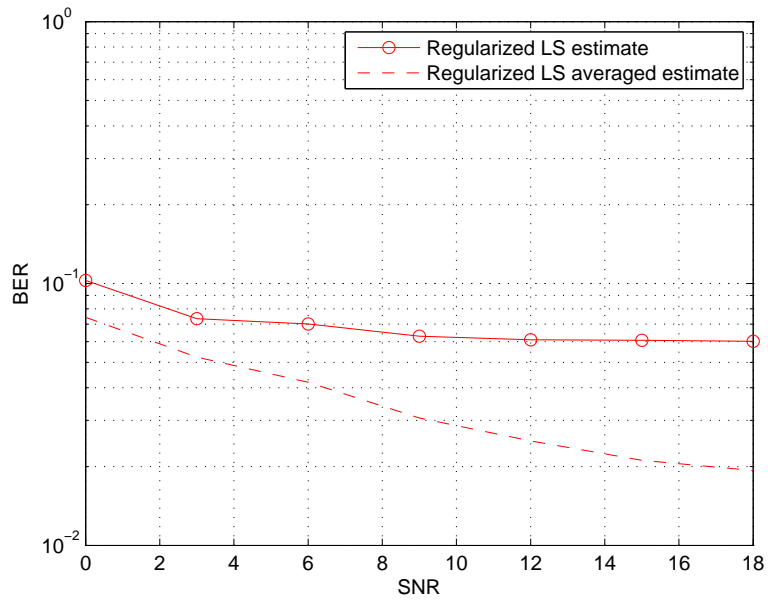
non-regularized case. Following are the BER curves for the regularized case with and without averaging. Figure 3.8(a) and 3.8(b) are both plotted for 32 pilots and 16 sections. The number of interpolation parameters used is 2 and 3 respectively.

## 3.8 Conclusion

The interpolation techniques based on simple linearization and quadratic approximation investigated in this chapter require very dense pilot placement and this increases the number of frequency domain parameters to be estimated. The inherent limit on the number of interpolation parameters per section can be removed by considering a regularized LS solution for the estimation problem. Still, we find that this method requires a high number of estimation parameter for channel estimation. As such, this method proves to be too expansive. Hence the need to explore some other method to represent the channel in frequency domain. In the next chapter, we will consider an alternate method for this purpose.



(a)



(b)

Figure 3.8: (a) BER curve for regularized solution with 32 pilots/2 parameters/16 sections (b) BER curve for regularized solution with 32 pilots/3 parameters/16 sections.

## CHAPTER 4

# EIGENVALUE APPROACH TO FREQUENCY DOMAIN CHANNEL ESTIMATION

Another approach for reducing the parameters in frequency domain channel estimation is the eigenvalue approach. Assuming that the second order statistics of the channel is available at the receiver, we can find its Eigenvalue decomposition. Using model reduction, we can represent  $\mathcal{H}$  using dominant eigenvalue and treat the rest as modeling noise. The block diagram of the system is given in figure 4.1.

The input/output equation that involves the  $j^{th}$  section is given by equation (2.12) while equation (2.13) gives its pruned form. We reproduce them here for

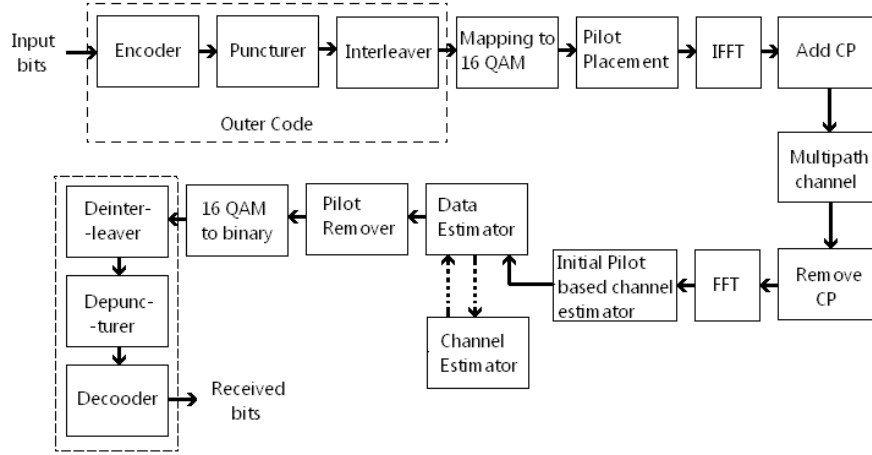


Figure 4.1: Block diagram of the system.

easy reference

$$\underline{\mathbf{y}}_i^{(j)} = \text{diag}(\underline{\mathbf{x}}_i^{(j)}) \underline{\mathbf{H}}_i^{(j)} + \underline{\mathcal{N}}_i^{(j)} \quad (4.1)$$

Dropping the dependence on  $j$  and  $i$  for notational convenience and pruning

$$\underline{\mathbf{y}}_{I_p} = \text{diag}(\underline{\mathbf{x}}_{I_p}) \underline{\mathbf{H}} + \underline{\mathcal{N}}_{I_p} \quad (4.2)$$

Obviously, the pilots are not enough to estimate the elements of  $\underline{\mathbf{H}}$ . So we resort to model reduction starting from the autocorrelation function of  $\underline{\mathbf{H}}$ , that is  $\mathbf{R}_{\underline{\mathbf{H}}}$ . To this end, consider the eigenvalue decomposition of  $\mathbf{R}_{\underline{\mathbf{H}}}$ ,

$$\mathbf{R}_{\underline{\mathbf{H}}} = \sum_{l=1}^{L_f} \lambda_l \mathbf{v}_l \mathbf{v}_l^T$$

where  $\lambda_1 \geq \lambda_2 \dots \geq \lambda_{L_f}$  are the (ordered) eigenvalues of  $\mathbf{R}_{\underline{\mathcal{H}}}$  and  $\mathbf{v}_1, \dots, \mathbf{v}_{L_f}$  are the corresponding eigenvectors. We can use this decomposition to represent  $\underline{\mathcal{H}}$  as

$$\underline{\mathcal{H}} = \sum_{l=1}^{L_f} \alpha_l \mathbf{v}_l$$

where  $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_{L_f}]^T$  is a parameter vector, to be estimated, with zero mean and autocorrelation matrix  $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{L_f})$ . We now represent  $\underline{\mathcal{H}}$  using the dominant eigenvalues and treat the rest as modeling noise <sup>1</sup>, i.e.

$$\underline{\mathcal{H}} = \mathbf{V}_d \boldsymbol{\alpha}_d + \mathbf{V}_n \boldsymbol{\alpha}_n \quad (4.3)$$

Upon substituting (4.3) in (4.1), we obtain

$$\underline{\mathcal{Y}} = \text{diag}(\underline{\mathcal{X}}) \mathbf{V}_d \boldsymbol{\alpha}_d + \underline{\mathcal{N}} + \text{diag}(\underline{\mathcal{X}}) \mathbf{V}_n \boldsymbol{\alpha}_n \quad (4.4)$$

$$= \underline{\mathbf{X}}_d \boldsymbol{\alpha}_d + \underline{\mathcal{N}}' \quad (4.5)$$

where  $\underline{\mathbf{X}}_d = \text{diag}(\underline{\mathcal{X}}) \mathbf{V}_d$  and  $\underline{\mathcal{N}}' = \underline{\mathcal{N}} + \underline{\mathbf{X}}_n \boldsymbol{\alpha}_n$  with  $\underline{\mathbf{X}}_n = \text{diag}(\underline{\mathcal{X}}) \mathbf{V}_n$ . The noise  $\underline{\mathcal{N}}'$  includes both the additive and *modeling* noise. We consider it to be zero mean white gaussian noise with autocorrelation

$$\mathbf{R}_{\underline{\mathcal{N}}'} = \mathbf{R}_{\underline{\mathcal{N}}} + \text{diag}(\underline{\mathcal{X}}) \mathbf{V}_n \text{diag}(\lambda_n) \mathbf{V}_n^* \text{diag}(\underline{\mathcal{X}})^* \quad (4.6)$$

---

<sup>1</sup>The cutoff between the parameters that are considered dominant and the ones that are considered as part of the modeling noise depends on the relative values of the  $\lambda_j$ 's. In our simulations, we use the condition  $\frac{\lambda_{j+1}}{\lambda_j} > 5$  to place our cutoff.



Now equation (4.5) can be used to construct a pilot/output equation, similar to (4.2), as

$$\underline{\mathbf{y}}_{I_p} = \underline{\mathbf{X}}_{d,I_p} \boldsymbol{\alpha}_d + \underline{\mathcal{N}}'_{I_p} \quad (4.7)$$

Which can be used to estimate  $\boldsymbol{\alpha}_d$  by maximizing the log likelihood function

$$\hat{\boldsymbol{\alpha}}_d^{\text{MAP}} = \arg \max_{\boldsymbol{\alpha}_d} \left\{ \ln p(\underline{\mathbf{y}}_{I_p} | \underline{\mathbf{X}}_{d,I_p}, \boldsymbol{\alpha}_d) + \ln p(\boldsymbol{\alpha}_d) \right\} \quad (4.8)$$

The maximum a posterior MAP estimate of parameter  $\boldsymbol{\alpha}$  is thus given by

$$\hat{\boldsymbol{\alpha}}_d^{\text{MAP}} = \arg \min_{\boldsymbol{\alpha}_d} \left\{ \|\underline{\mathbf{y}}_{I_p} - \underline{\mathbf{X}}_{d,I_p} \boldsymbol{\alpha}_d\|_{\underline{\mathbf{R}}_{\mathcal{N}'}}^{-1} + \|\boldsymbol{\alpha}_d\|_{\underline{\boldsymbol{\Lambda}}_d}^{-1} \right\} \quad (4.9)$$

which simplifies to

$$\hat{\boldsymbol{\alpha}}_d = \underline{\boldsymbol{\Lambda}}_d \underline{\mathbf{X}}_{d,I_p}^* \left[ \underline{\mathbf{R}}_{\mathcal{N}'} + \underline{\mathbf{X}}_{d,I_p} \underline{\boldsymbol{\Lambda}}_d \underline{\mathbf{X}}_{d,I_p}^* \right]^{-1} \underline{\mathbf{y}}_{I_p} \quad (4.10)$$

The resulting mean square error is given by

$$\mathbf{R}_e = \left[ \underline{\boldsymbol{\Lambda}}_d^{-1} + \underline{\mathbf{X}}_{d,I_p}^* \underline{\mathbf{R}}_{\mathcal{N}'}^{-1} \underline{\mathbf{X}}_{d,I_p} \right]^{-1} \quad (4.11)$$

The estimate of the  $j^{\text{th}}$  section of the spectrum is then given by  $\hat{\underline{\mathcal{H}}} = \mathbf{V}_d \hat{\boldsymbol{\alpha}}_d$ . The concatenation of all  $\lceil \frac{N}{L_f} \rceil$  sections produces the frequency domain based estimate of the frequency response  $\hat{\mathcal{H}}$ .

## 4.1 Iterative Channel Estimation using the Expectation Maximization Approach

Pilot based channel estimation, whether in the time domain or frequency domain, does not make full use of the constraints on the data. One can thus implement iterative (data-aided) techniques for channel estimation [39], [43]. Using the data aided approach, we can improve the channel estimate [30], [43]. Thus providing the motivation to use the expectation-maximization (EM) algorithm. The EM algorithm is used to estimate a parameter in the case when some of the data required for estimation is unobserved. The algorithm first performs an initial estimate of the unobserved data and uses this estimate to compute the maximum-likelihood (ML) estimate of the parameter to be estimated. This is the maximization step. Next, the algorithm uses the estimated parameter to update the estimate of the unobserved data. This is the expectation step. These steps are repeated iteratively until a convergent solution is reached [60]. Next, we will discuss the EM algorithm in detail.

### 4.1.1 The Maximization Step

In the previous subsection we find  $\hat{\boldsymbol{\alpha}}_d$  by maximizing the log likelihood function given by equation (4.8). Since the input  $\boldsymbol{\mathcal{X}}$  (and hence  $\underline{\boldsymbol{X}}_d$ ) is not observable, we can employ the EM algorithm and instead of maximizing (4.8) we can maximize an averaged form of the log likelihood function. Specifically, starting from an initial estimate  $\hat{\boldsymbol{\alpha}}_d^{(0)}$ , calculated say using pilots, the estimate  $\hat{\boldsymbol{\alpha}}_d$  is calculated

iteratively with the estimate at the  $k^{th}$  iteration given by

$$\hat{\boldsymbol{\alpha}}_d^{(k)} = \arg \max_{\boldsymbol{\alpha}_d} \left\{ E_{\mathcal{X}_i | \mathcal{Y}_i, \hat{\boldsymbol{\alpha}}_d^{(k-1)}} \ln p(\underline{\mathbf{y}}_{I_p} | \underline{\mathbf{X}}_{d, I_p}, \boldsymbol{\alpha}_d) + \ln p(\boldsymbol{\alpha}_d) \right\} \quad (4.12)$$

which simplifies to <sup>2</sup>

$$\hat{\boldsymbol{\alpha}}_d^{\text{MAP}} = \arg \min_{\boldsymbol{\alpha}_d} \left\{ E \left\| \underline{\mathbf{y}}_{I_p} - \underline{\mathbf{X}}_{d, I_p} \boldsymbol{\alpha}_d \right\|_{\underline{\mathbf{R}}_{\mathcal{N}'}}^2 + \|\boldsymbol{\alpha}_d\|_{\underline{\boldsymbol{\Lambda}}_d}^2 \right\} \quad (4.13)$$

Strictly speaking, the noise correlation  $\underline{\mathbf{R}}_{\mathcal{N}'}$  is itself dependent on the input due to the modeling noise (see equation (4.6)). Hence in performing the expectation in (4.13), we need to take this into account. Treating the general case is difficult, so we consider the following three cases for  $\underline{\mathbf{R}}_{\mathcal{N}'}$ :

Case 1:  $\underline{\mathbf{R}}_{\mathcal{N}'}$  is a constant:

This happens when we ignore the modeling noise so that

$$\underline{\mathbf{R}}_{\mathcal{N}'} = \sigma^2 \mathbf{I}$$

where the expectation in (4.13) is taken with respect to  $\underline{\mathbf{X}}_d$  given  $\underline{\mathbf{y}}$  and the most recent estimate  $\boldsymbol{\alpha}_d$ . In this case  $\underline{\mathbf{R}}_{\mathcal{N}'}$  becomes independent of  $\underline{\mathbf{X}}_d$  and it would then be straight forward to carry the expectation in (4.13). Specifically, upon

---

<sup>2</sup>the expectation is taken with respect to the input given the output and the most recent estimate  $\hat{\boldsymbol{\alpha}}_d^{k-1}$ . This information is understood & dropped for notational convenience.

completing the squares, (4.13) can be equivalently written as

$$\begin{aligned} \min_{\alpha_d} & \mathbf{y}_i^* \mathbf{R}_{\mathcal{N}'}^{-1} \mathbf{y}_i - \alpha_d^* E[\underline{\mathbf{X}}_d^*] \mathbf{R}_{\mathcal{N}'}^{-1} \mathbf{y}_i - \mathbf{y}_i^* \mathbf{R}_{\mathcal{N}'}^{-1} E[\underline{\mathbf{X}}_d] \alpha_d \\ & + \alpha_d^* E[\underline{\mathbf{X}}_d^*] \mathbf{R}_{\mathcal{N}'}^{-1} E[\underline{\mathbf{X}}_d] \alpha_d - \alpha_d^* E[\underline{\mathbf{X}}_d^*] \mathbf{R}_{\mathcal{N}'}^{-1} E[\underline{\mathbf{X}}_d] \alpha_d \\ & + \alpha_d^* E[\underline{\mathbf{X}}_d^* \mathbf{R}_{\mathcal{N}'}^{-1} \underline{\mathbf{X}}_d] \alpha_d + \alpha_d^* \Lambda_d^{-1} \alpha_d \end{aligned}$$

which can be simplified to

$$\hat{\alpha}_d^{\text{MAP}} = \arg \min_{\alpha_d} \|\mathbf{y} - E[\underline{\mathbf{X}}_d] \alpha_d\|_{\frac{1}{\sigma_n^2} \mathbf{I}}^2 + \|\alpha_d\|_{\frac{1}{\sigma_n^2} \text{Cov}[\underline{\mathbf{X}}_d]}^2 + \|\alpha_d\|_{\Lambda_d^{-1}}^2 \quad (4.14)$$

Case 2: Taking Expectation of  $\mathbf{R}_{\mathcal{N}'}$ :

Instead of ignoring the modeling noise, we can split the expectation in (4.13) into an expectation over  $\mathbf{R}_{\mathcal{N}'}$  and an *independent* expectation taken over the rest of the terms i.e., we can approximate (4.13) as

$$\hat{\alpha}_d^{\text{MAP}} = \arg \min_{\alpha_d} \left\{ E\|\mathbf{y}_{I_p} - \underline{\mathbf{X}}_{d, I_p} \alpha_d\|_{E[\mathbf{R}_{\mathcal{N}'}]^{-1}}^2 + \|\alpha_d\|_{\Lambda_d^{-1}}^2 \right\} \quad (4.15)$$

Now the expectation of  $\mathbf{R}_{\mathcal{N}'}$  is given by

$$E[\mathbf{R}_{\mathcal{N}'}] = \sigma^2 \mathbf{I} + E[\text{diag}(\underline{\mathbf{X}}) \mathbf{V}_n \Lambda_n \mathbf{V}_n^* \text{diag}(\underline{\mathbf{X}}^*)] \quad (4.16)$$

We show in Appendix A that this expectation can be expressed as

$$E[\mathbf{R}_{\mathcal{N}'}] = \sigma^2 \mathbf{I} + E[\underline{\mathbf{D}}] \mathbf{V}_n \Lambda_n \mathbf{V}_n^* E[\underline{\mathbf{D}}^*] + \text{Cov}[\underline{\mathbf{D}}] \text{diag}(\mathbf{V}_n \Lambda_n \mathbf{V}_n^*)$$

where  $\underline{\mathbf{D}} = \text{diag}(\underline{\mathbf{X}})$  and where  $\text{diag}(\mathbf{V}_n \boldsymbol{\Lambda}_n \mathbf{V}_n^*)$  is a diagonal matrix whose diagonal coincides with the diagonal of the matrix  $\mathbf{V}_n \boldsymbol{\Lambda}_n \mathbf{V}_n^*$ . The now averaged  $\mathbf{R}_{\underline{\mathcal{N}}'}$  does not depend on  $\underline{\mathbf{X}}$  any more. Replacing  $\mathbf{R}_{\underline{\mathcal{N}}'}$  by its expectation, it is then straight forward to carry the expectation in (4.15) which comes out to be

$$\hat{\boldsymbol{\alpha}}_d^{\text{MAP}} = \arg \min_{\boldsymbol{\alpha}_d} \|\underline{\mathbf{y}} - E[\underline{\mathbf{X}}_d] \boldsymbol{\alpha}_d\|_{E[\mathbf{R}_{\underline{\mathcal{N}}'}]^{-1}}^2 + \|\boldsymbol{\alpha}_d\|_{\text{Cov}[\underline{\mathbf{D}}] \text{diag}(\mathbf{V}_n \boldsymbol{\Lambda}_n \mathbf{V}_n^*)}^2 + \|\boldsymbol{\alpha}_d\|_{\Lambda_d}^2 \quad (4.17)$$

Case 3:  $\mathbf{X}$  is constant modulus:

In the constant modulus case, it is possible to evaluate (4.13) exactly. Specifically, and starting from the expression for the autocorrelation  $\mathbf{R}_{\underline{\mathcal{N}}'}$

$$\mathbf{R}_{\underline{\mathcal{N}}'} = \sigma^2 \mathbf{I} + \underline{\mathbf{D}} \mathbf{V}_n \boldsymbol{\Lambda}_n \mathbf{V}_n^* \underline{\mathbf{D}}^*$$

we can write

$$\begin{aligned} \mathbf{R}_{\underline{\mathcal{N}}'}^{-1} &= (\sigma^2 \mathbf{I} + \underline{\mathbf{D}} \mathbf{V}_n \boldsymbol{\Lambda}_n \mathbf{V}_n^* \underline{\mathbf{D}}^*)^{-1} \\ &= \underline{\mathbf{D}}^{-*} \left( \frac{\sigma^2}{\mathcal{E}} \mathbf{I} + \mathbf{V}_n \boldsymbol{\Lambda}_n \mathbf{V}_n^* \right)^{-1} \underline{\mathbf{D}}^{-1} \\ &= \underline{\mathbf{D}}^{-*} \mathbf{R}_{\underline{\mathcal{N}}''}^{-1} \underline{\mathbf{D}}^{-1} \end{aligned}$$

where  $\mathbf{R}_{\underline{\mathcal{N}}''} \triangleq \frac{\sigma^2}{\mathcal{E}} \mathbf{I} + \mathbf{V}_n \boldsymbol{\Lambda}_n \mathbf{V}_n^*$  and where we used the fact that  $\underline{\mathbf{D}} \underline{\mathbf{D}}^* = \mathcal{E} \mathbf{I}$  since the input is constant modulus. With this in mind, we conclude that

$$\underline{\mathbf{X}}_d^* \mathbf{R}_{\underline{\mathcal{N}}'}^{-1} = \mathbf{V}_d^* \underline{\mathbf{D}}^* \mathbf{R}_{\underline{\mathcal{N}}'}^{-1} = \mathbf{V}_d^* \mathbf{R}_{\underline{\mathcal{N}}''}^{-1} \underline{\mathbf{D}}^{-1}$$

$$\underline{\mathbf{R}}_{\underline{\mathcal{N}}'}^{-1} \underline{\mathbf{X}}_d = \mathbf{D}^{-1*} \underline{\mathbf{R}}_{\underline{\mathcal{N}}''}^{-1} \mathbf{V}_d$$

and

$$\underline{\mathbf{X}}_d^* \underline{\mathbf{R}}_{\underline{\mathcal{N}}'}^{-1} \underline{\mathbf{X}}_d = \mathbf{V}_d^* \underline{\mathbf{R}}_{\underline{\mathcal{N}}''}^{-1} \mathbf{V}_d$$

Thus, in the constant modulus case, (4.13) can be equivalently written as

$$\begin{aligned} \hat{\boldsymbol{\alpha}}_d^{(j)} = \arg \min_{\boldsymbol{\alpha}_d} & \boldsymbol{\mathcal{Y}}^* E[\mathbf{D}^{-1*}] \underline{\mathbf{R}}_{\underline{\mathcal{N}}''}^{-1} E[\mathbf{D}^{-1}] \boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{Y}}^* E[\mathbf{D}^{-1*}] \underline{\mathbf{R}}_{\underline{\mathcal{N}}''}^{-1} \mathbf{V}_d \boldsymbol{\alpha}_d \\ & - \boldsymbol{\alpha}_d^* \mathbf{V}_d^* \underline{\mathbf{R}}_{\underline{\mathcal{N}}''}^{-1} E[\mathbf{D}^{-1}] \boldsymbol{\mathcal{Y}} + \boldsymbol{\alpha}_d^* \mathbf{V}_d^* \underline{\mathbf{R}}_{\underline{\mathcal{N}}''}^{-1} \mathbf{V}_d \boldsymbol{\alpha}_d + \boldsymbol{\alpha}_d^* \boldsymbol{\Lambda}_d^{-1} \boldsymbol{\alpha}_d \end{aligned}$$

which upon simplification becomes

$$\hat{\boldsymbol{\alpha}}_d^{\text{MAP}} = \arg \min_{\boldsymbol{\alpha}_d} \|E[\mathbf{D}^{-1}] \boldsymbol{\mathcal{Y}} - \mathbf{V}_d \boldsymbol{\alpha}_d\|_{\underline{\mathbf{R}}_{\underline{\mathcal{N}}''}^{-1}}^2 + \|\boldsymbol{\alpha}_d\|_{\boldsymbol{\Lambda}_d^{-1}}^2 \quad (4.18)$$

In the simulations further ahead, we compare the approximate solutions (4.14) & (4.17) with the exact EM solution (4.18) for a constant modulus input. Simulations show that replacing  $\underline{\mathbf{R}}_{\underline{\mathcal{N}}'}$  with its expectation is almost as good as calculating the expectation exactly.

## 4.1.2 The Expectation Step

As we have seen above, the maximization step assumes the presence of some expectations. By inspecting subsection 4.1.1, we see we need to calculate the

following moments.

$$E[\underline{\mathbf{X}}_d], \text{Cov}[\underline{\mathbf{X}}_d^*], E[\underline{\mathbf{D}}], E[\underline{\mathbf{D}}\underline{\mathbf{B}}\underline{\mathbf{D}}^*], \text{ and } E[\underline{\mathbf{D}}^{-1}] \quad (4.19)$$

Now as  $\underline{\mathbf{X}}_d = \text{diag}(\underline{\boldsymbol{\mathcal{X}}})\mathbf{V}_d = \underline{\mathbf{D}}\mathbf{V}_d$  we can see that we can express the moments of  $\mathbf{X}_d$  in terms of moments of  $\mathbf{D}$ . Specifically we have that

$$E[\underline{\mathbf{X}}_d] = E[\underline{\mathbf{D}}]\mathbf{V}_d$$

and

$$\begin{aligned} \text{Cov}[\underline{\mathbf{X}}_d^*] &= E[\underline{\mathbf{X}}_d\underline{\mathbf{X}}_d^*] - E[\underline{\mathbf{X}}_d]E[\underline{\mathbf{X}}_d^*] \\ &= E[\underline{\mathbf{D}}]\mathbf{V}_d\mathbf{V}_d^*E[\underline{\mathbf{D}}^*] + \text{Cov}[\underline{\mathbf{D}}]\text{diag}(\mathbf{V}_d\mathbf{V}_d^*) - E[\underline{\mathbf{D}}]\mathbf{V}_d\mathbf{V}_d^*E[\underline{\mathbf{D}}^*] \\ &= \text{Cov}[\underline{\mathbf{D}}]\text{diag}(\mathbf{V}_d\mathbf{V}_d^*) \end{aligned}$$

Moreover, we show in appendix A that

$$E[\underline{\mathbf{D}}\underline{\mathbf{B}}\underline{\mathbf{D}}^*] = E[\underline{\mathbf{D}}]\mathbf{B}E[\underline{\mathbf{D}}^*] + \text{Cov}[\underline{\mathbf{D}}]\text{diag}(\mathbf{B}) \quad (4.20)$$

From above it follows that in order to calculate the expectations in (4.19), it is enough to calculate the following three moments

$$E[\text{diag}(\underline{\boldsymbol{\mathcal{X}}})], \text{Cov}[\text{diag}(\underline{\boldsymbol{\mathcal{X}}})] \ \& \ E[\text{diag}(\underline{\boldsymbol{\mathcal{X}}})^{-1}] \quad (4.21)$$

where the expectation is performed given the output  $\underline{\mathcal{Y}}$  and the most recent channel estimate  $\hat{\underline{\mathcal{H}}}$ . In carrying out these expectations, we will assume that the elements of  $\underline{\mathcal{X}}$  are independent.<sup>3</sup> With this in mind, it is easy to see that we can evaluate the moments in (4.21) and hence in (4.19) by calculating

$$E[\underline{\mathcal{X}}(l)], \text{Cov}[\underline{\mathcal{X}}(l)] = E[|\underline{\mathcal{X}}(l)|^2] - |E[\underline{\mathcal{X}}(l)]|^2, \quad E\left[\frac{1}{\underline{\mathcal{X}}(l)}\right]$$

Now assuming that  $\underline{\mathcal{X}}(l)$  is drawn from the alphabet  $A = \{A_1, \dots, A_M\}$  with equal probability, it is can be shown that [43]

$$E[\underline{\mathcal{X}}(l)|\underline{\mathcal{Y}}(l), \underline{\mathcal{H}}(l)] = \frac{\sum_{j=1}^M A_j e^{-\frac{|\underline{\mathcal{Y}}(l) - \underline{\mathcal{H}}(l)A_j|^2}{\sigma^2}}}{\sum_{j=1}^M e^{-\frac{|\underline{\mathcal{Y}}(l) - \underline{\mathcal{H}}(l)A_j|^2}{\sigma^2}}} \quad (4.22)$$

$$E[|\underline{\mathcal{X}}(l)|^2|\underline{\mathcal{Y}}(l), \underline{\mathcal{H}}(l)] = \frac{\sum_{j=1}^M |A_j|^2 e^{-\frac{|\underline{\mathcal{Y}}(l) - \underline{\mathcal{H}}(l)A_j|^2}{\sigma^2}}}{\sum_{j=1}^M e^{-\frac{|\underline{\mathcal{Y}}(l) - \underline{\mathcal{H}}(l)A_j|^2}{\sigma^2}}} \quad (4.23)$$

$$E\left[\frac{1}{\underline{\mathcal{X}}(l)}|\underline{\mathcal{Y}}(l), \underline{\mathcal{H}}(l)\right] = \frac{\sum_{j=1}^M \frac{1}{A_j} e^{-\frac{|\underline{\mathcal{Y}}(l) - \underline{\mathcal{H}}(l)A_j|^2}{\sigma^2}}}{\sum_{j=1}^M e^{-\frac{|\underline{\mathcal{Y}}(l) - \underline{\mathcal{H}}(l)A_j|^2}{\sigma^2}}} \quad (4.24)$$

### 4.1.3 Summary of the EM Algorithm

Now let us summarize the EM based estimation algorithm developed so far.

1. Calculate the initial channel estimate  $\hat{\underline{\mathcal{H}}}_0$  using pilots (4.9).
2. Calculate the moments of the input given the current channel estimate  $\hat{\underline{\mathcal{H}}}_i$  and the output  $\underline{\mathcal{Y}}$  using equations (4.22)-(4.24).

---

<sup>3</sup>This is in general not true because the elements of  $\underline{\mathcal{H}}$  are not independent (as the elements of  $\underline{\mathcal{H}}$  are the fourier transform of the impulse response  $\underline{\mathbf{h}}$ ). However, we continue to use this approximation as this maintains the transparency of element-by-element equalization in OFDM.



3. Calculate the channel estimate using either one of the methods (4.14), (4.17) or (4.18) outlined in Section 4.1.
4. Iterate between step 2 and 3.

We can run the algorithm for a specific number of times or until some predefined minimum error threshold is reached.

## 4.2 Using Time-Correlation to Improve the Channel Estimate

The receiver developed in the previous section performs channel estimation symbol by symbol. In other words, the channel is block fading & hence is totally independent from symbol to symbol. In a practical scenario the channel impulse responses are correlated over time. In this section, we will show how to use time correlation to enhance the estimate of  $\alpha_d$ . To this end, let's first develop a model for the time variation of the parameter  $\alpha_d$ .

### 4.2.1 Developing a Frequency Domain Time-Variant Model

Consider the block fading model in (2.2) and let's assume for simplicity that the diagonal matrices  $\mathbf{F}$  and  $\mathbf{G}$  are actually scalar multiples of the identity, i.e.

$$\mathbf{F} = f\mathbf{I} \quad \mathbf{G} = \sqrt{1 - f^2}\mathbf{I}$$

where  $f$  is a function of Doppler frequency (see [43]). We will use the time domain model in (2.2) to derive a similar model for  $\alpha$ . To this end, recall that

$$\mathcal{H}_i = \mathbf{Q}_{P+1} \mathbf{h}_i$$

Thus, the  $j^{\text{th}}$  section of  $\mathcal{H}_i$ ,  $\underline{\mathcal{H}}_i^{(j)}$ , is related to  $\mathbf{h}_i$  by

$$\underline{\mathcal{H}}_i^{(j)} = \mathbf{Q}_{P+1}^{(j)} \mathbf{h}_i \quad (4.25)$$

where  $\mathbf{Q}_{P+1}^{(j)}$  corresponds to the  $j^{\text{th}}$  section of  $\mathbf{Q}_{P+1}$ , i.e.,  $\mathbf{Q}_{P+1}$  pruned of all its rows except those of the  $j^{\text{th}}$  section. Now, we can replace  $\underline{\mathcal{H}}_i^{(j)}$  by its representation using the dominant parameters  $\alpha_d$ , to get

$$\mathbf{V}_d \alpha_{d,i} = \mathbf{Q}_{P+1}^{(j)} \mathbf{h}_i$$

or

$$\alpha_{d,i} = \mathbf{V}_d^+ \mathbf{Q}_{P+1}^{(j)} \mathbf{h}_i$$

where  $\mathbf{V}_d^+$  is the pseudo inverse of  $\mathbf{V}_d$ . Multiplying both sides of (2.2) by  $\mathbf{V}_d^+ \mathbf{Q}_{P+1}^{(j)}$  yields a dynamical recursion for  $\alpha_d$

$$\alpha_{d,i+1} = \mathbf{F}_\alpha \alpha_{d,i} + \mathbf{G}_\alpha \mathbf{u}_i \quad (4.26)$$

where  $\mathbf{F}_\alpha = f\mathbf{I}$  and  $\mathbf{G}_\alpha = \sqrt{1-f^2}\mathbf{V}_d^+\mathbf{Q}_{P+1}^{(j)}$  and where

$$E[\boldsymbol{\alpha}_{d,0}\boldsymbol{\alpha}_{d,0}^*] = \boldsymbol{\Lambda}_d$$

Note that the dependence of  $\mathbf{G}_\alpha$  and  $\boldsymbol{\alpha}_d$  on  $j$  has been suppressed for notational convenience. We are now ready to implement the EM algorithm to the frequency domain system governed by the dynamical equation (4.26). As we have seen in section 4.1, the algorithm will consist of an initial estimation step, a maximization step, and an expectation step.

## 4.2.2 Initial (Pilot-Based) Channel Estimation

In the initial channel estimation step, the frequency domain system is described by equations (4.7) and (4.26), reproduced here for convenience.

$$\underline{\mathbf{y}}_{I_p,i} = \underline{\mathbf{X}}_{d,I_p,i}\boldsymbol{\alpha}_{d,i} + \underline{\mathcal{N}}'_{I_p,i} \quad (4.27)$$

$$\boldsymbol{\alpha}_{d,i+1} = \mathbf{F}_\alpha\boldsymbol{\alpha}_{d,i} + \mathbf{G}_\alpha\mathbf{u}_i \quad (4.28)$$

Now given a sequence  $i = 0, 1, \dots, T$  of pilot bearing symbols, we can obtain the optimum estimate of  $\{\boldsymbol{\alpha}_{i,d}\}_{i=0}^T$  by applying a forward-backward Kalman to (4.27)-(4.28)(see [56]), i.e., by implementing the following equations

**Forward run:** Starting from the initial conditions  $\mathbf{P}_{0|-1} = \mathbf{\Pi}_0$  and  $\boldsymbol{\alpha}_{0|-1} = \mathbf{0}$  and for  $i = 1, \dots, T$ , calculate

$$\mathbf{R}_{e,i} = \mathbf{R}_{\mathcal{N}'} + \underline{\mathbf{X}}_{d,I_p,i} \mathbf{P}_{i|i-1} \underline{\mathbf{X}}_{d,I_p,i}^* \quad (4.29)$$

$$\mathbf{K}_{f,i} = \mathbf{P}_{i|i-1} \underline{\mathbf{X}}_{d,I_p,i}^* \mathbf{R}_{e,i}^{-1} \quad (4.30)$$

$$\hat{\boldsymbol{\alpha}}_{i|i} = \left( \mathbf{I} - \mathbf{K}_{f,i} \underline{\mathbf{X}}_{d,I_p,i} \right) \hat{\boldsymbol{\alpha}}_{i|i-1} + \mathbf{K}_{f,i} \boldsymbol{\mathcal{Y}}_i \quad (4.31)$$

$$\hat{\boldsymbol{\alpha}}_{i+1|i} = \mathbf{F}_\alpha \hat{\boldsymbol{\alpha}}_{i|i} \quad (4.32)$$

$$\mathbf{P}_{i+1|i} = \mathbf{F}_\alpha \left( \mathbf{P}_{i|i-1} - \mathbf{K}_{f,i} \mathbf{R}_{e,i} \mathbf{K}_{f,i}^* \right) \mathbf{F}_\alpha^* + \frac{1}{\sigma_n^2} \mathbf{G}_\alpha \mathbf{G}_\alpha^* \quad (4.33)$$

**Backward run:** Starting from  $\boldsymbol{\lambda}_{T+1|T} = \mathbf{0}$  and for  $i = T, T-1, \dots, 0$ , calculate

$$\boldsymbol{\lambda}_{i|T} = \left( \mathbf{I}_{P+N} - \underline{\mathbf{X}}_{d,I_p,i}^* \mathbf{K}_{f,i} \right) \mathbf{F}_i^* \boldsymbol{\lambda}_{i+1|T} + \underline{\mathbf{X}}_{d,I_p,i} \mathbf{R}_{e,i}^{-1} \left( \boldsymbol{\mathcal{Y}}_i - \underline{\mathbf{X}}_{d,I_p,i} \hat{\boldsymbol{\alpha}}_{i|i} \right) \quad (4.34)$$

$$\hat{\boldsymbol{\alpha}}_{i|T} = \hat{\boldsymbol{\alpha}}_{i|i-1} + \mathbf{P}_{i|i-1} \boldsymbol{\lambda}_{i|T} \quad (4.35)$$

The desired estimate is  $\hat{\boldsymbol{\alpha}}_{i|T}$ . This gives us an initial estimate to run the data-aided part of the algorithm with.

### 4.2.3 Iterative (Data-Aided) Channel Estimation

For this part, we use the whole data symbol and not just the pilot part. Thus, in this case our system is described by equations (4.5) and (4.26) also reproduced

here for convenience

$$\underline{\mathbf{y}}_i = \underline{\mathbf{X}}_{d,i} \boldsymbol{\alpha}_{i,d} + \underline{\mathcal{N}}'_i \quad (4.36)$$

$$\boldsymbol{\alpha}_{d,i+1} = \mathbf{F}_\alpha \boldsymbol{\alpha}_{d,i} + \mathbf{G}_\alpha \mathbf{u}_i \quad (4.37)$$

If the data symbols  $\underline{\mathbf{X}}_{d,i}$  were known, we would have employed the forward-backward Kalman-Filter (4.29)-(4.35) on the above state-space model. Since the input is not available, we replace it by its estimate along an expectation maximization algorithm. Specifically, along the lines developed in [43] we can show that the FB Kalman filter needs to be applied to the following state space model

$$\underline{\mathbf{y}}_i = \begin{bmatrix} E[\underline{\mathbf{X}}_{d,i}] \\ Cov[\underline{\mathbf{X}}_{d,i}^*]^{\frac{1}{2}} \end{bmatrix} \boldsymbol{\alpha}_{i,d} + \begin{bmatrix} \underline{\mathcal{N}}'_i \\ \mathbf{0} \end{bmatrix} \quad (4.38)$$

$$\boldsymbol{\alpha}_{d,i+1} = \mathbf{F}_\alpha \boldsymbol{\alpha}_{d,i} + \mathbf{G}_\alpha \mathbf{u}_i \quad (4.39)$$

where the expectations in (4.38) are taken given the output  $\underline{\mathbf{y}}_i$  and most recent channel estimate  $\boldsymbol{\alpha}_{d,i}$ . The expectations that appears in (4.38) are calculated as we did in Section 4.1.2. In contrast to the symbol by symbol EM algorithm of section 4.1, there are several ways of implementing the EM iterations in the time-correlated multi-symbol case. In the symbol by symbol algorithm of Section 4, there was one dimension to iterate against (channel estimation versus data detection). When the channels are time correlated over several OFDM symbols

as is the case here, there are two dimensions we can iterate against:

1. We can iterate between channel estimation & data detection.
2. We could also iterate against time using the Kalman filter where the previous channel estimate informs the subsequent channel estimate.

Depending on how we schedule iterations across these two dimensions, we get different receivers. We discuss two such receivers here, the Cyclic and the Helix Kalman based receivers.

#### 4.2.4 Cyclic FB Kalman

In the cyclic based Kalman, we initialize the algorithm using the FB Kalman implemented over the pilot symbols. This is then used to initialize the data aided version, where the channel estimate is used to obtain the data estimate, and that allows us to propagate the estimate to the next symbol. The process is continued until the forward steps are completed followed by the backward run. The EM steps are repeated again (  $2^{nd}$  forward run followed by  $2^{nd}$  backward run and so on). In other words, we iterate only *once* between channel estimation & data detection before invoking the Kalman to move to the next symbol and so on. The iterations thus trace circles over the OFDM symbols which motivates the name Cyclic Kalman.

### 4.2.5 Helix based FB Kalman

The Helix based FB Kalman is a more general version of the Cyclic Kalman. The two filters are initialized in the same way. However at each symbol, we iterate several times between channel estimation and data detection before moving on the next symbol (whereas the cyclic Kalman iterates once between the channel estimate and data estimate at each step). This allows to refine the channel estimate as much as possible before propagating it using the Kalman to the next OFDM symbol. The iterations in this case draw a helix shape, hence the name.

### 4.2.6 Using Code to Enhance the Estimate

In any practical system, an outer code is usually implemented that extends over several OFDM symbols. The outer code can be used to enhance the data aided channel estimate. Specifically, following data detection, the code can be invoked to enhance the data estimate (through error correction). Now the (hard) data obtained is more refined and hence can be used to enhance the channel estimate by employing the FB Kalman again. Our simulation shows that invoking the code can have a profound effect on performance.

### 4.2.7 Forward Kalman Filter

One drawback of the FB Kalman implementation is the latency and memory involved as one needs to store *all* symbols to perform the backward run. One way around that is to implement the forward only Kalman which avoids the latency

problem. The forward only Kalman thus suffers as a result in performance and is not able to make use of the code to enhance the data estimate.

### 4.3 Time Domain Multiple Access Channel Estimation

For fair comparison, we need to compare the frequency domain (LS and Kalman) receiver, with the time domain counterpart. How do users estimate the channel in the time domain given their limited share of the spectrum? To describe this, we just need to write the input/output equations seen by each user. The input/output equation for the  $j^{\text{th}}$  user is given by (see (4.1))

$$\underline{\mathbf{y}}_i^{(j)} = \text{diag}(\underline{\mathbf{x}}_i^{(j)})\underline{\mathbf{H}}_i^{(j)} + \underline{\mathcal{N}}_i^{(j)}$$

Now  $\underline{\mathbf{H}}_i^{(j)}$  is related to the impulse response by (see (4.25))

$$\underline{\mathbf{H}}_i^{(j)} = \mathbf{Q}_{P+1}^{(j)} \mathbf{h}_i$$

where as described in Section 4.2.1,  $\mathbf{Q}_{P+1}^{(j)}$  is  $\mathbf{Q}_{P+1}$  pruned of all rows that don't belong to the  $j^{\text{th}}$  section. So, we can write

$$\underline{\mathbf{y}}_i^{(j)} = \text{diag}(\underline{\mathbf{x}}_i^{(j)})\mathbf{Q}_{P+1}^{(j)}\mathbf{h}_i + \underline{\mathcal{N}}_i^{(j)} \tag{4.40}$$



Equation (4.40) can be used for initial *time-domain* estimate using pilots and for symbol-by-symbol EM-based estimation. If we use in addition the dynamic recursion of (2.2)

$$\mathbf{h}_{i+1} = \mathbf{F}\mathbf{h}_i + \mathbf{G}\mathbf{u}_i$$

we can implement the various kind of Kalman filters discussed in the previous section for *time-domain* channel estimation. It is important to note that the computational complexity involved in the time domain case is much higher than in the Eigen estimate as the significant eigenvalues  $\alpha_d$  are less than the channel length.

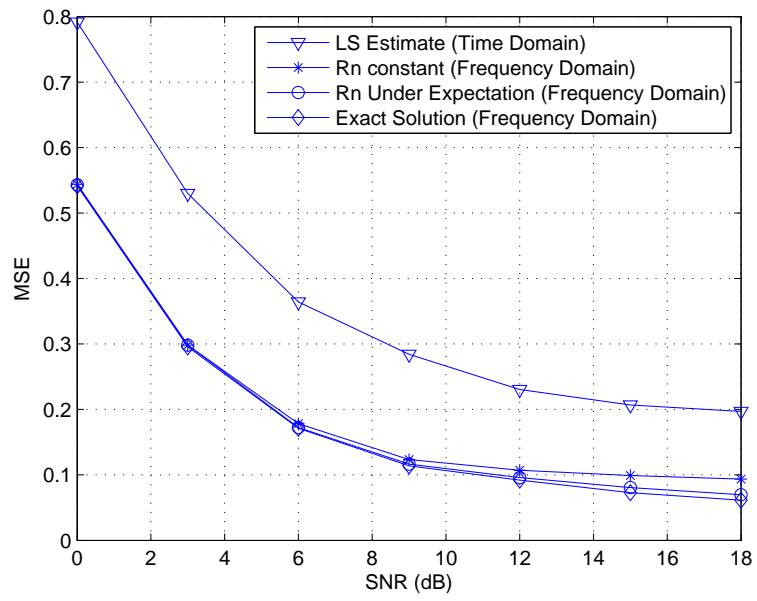
## 4.4 Simulation Results

We consider an OFDM system that transmits 6 symbols with 64 carriers and a cyclic prefix of length  $P = 15$  each with a time variation of  $f = 0.9$ . The data bits are mapped to 16 QAM through Gray coding (except for figures 4.2(a) and 4.2(b) which use a 4 QAM). The OFDM symbol serves 4 users each occupying 16 frequency bins. In addition, the OFDM symbol carries 16 or 24 pilots equally divided between the users. The channel impulse response consists of 15 complex taps (the maximum length possible). It has an exponential delay profile  $E[|h_0(k)|^2] = e^{-0.2k}$  and remains fixed over any OFDM symbol. Where specified, an outer code is used to provide robustness. The outer code is 1/2 rate convolutional code. In what follows, we compare the performance of frequency domain based channel estimation using various techniques for both the coded and uncoded cases. We also

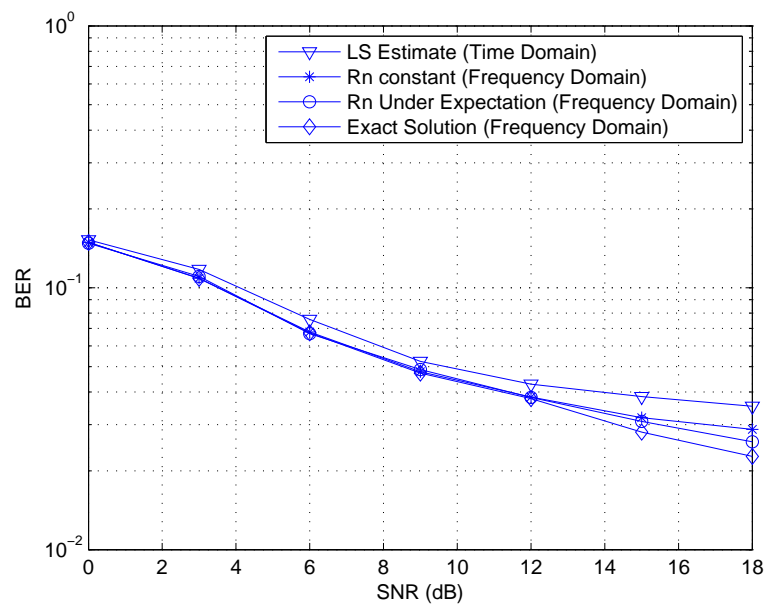
benchmark our method with the time domain method briefly described in Section 4.3 (see [43] also).

#### **4.4.1 Effect of Modeling Noise**

Figures 4.2(a) and 4.2(b) show the MSE and BER curves for the three cases considered in section 4.1 comparing the various treatment of the noise. We plot the figures 4.2(a) and 4.2(b) for constant modulus using 16 pilots. As evident from the graphs, the inclusion of the modeling noise improves the result. We also note that the expectation of the noise and the exact solution have almost comparable results.



(a)



(b)

Figure 4.2: (a) MSE (b) BER.

## 4.4.2 EM based Least Squares

For a fair comparison between the time domain and the frequency domain techniques for a multiple access system, we compare the time domain LS estimate with the frequency domain LS and LS with EM estimate. Figures 4.3-4.6 show the MSE of the channel estimate and the BER performance for these methods for the uncoded case. Figures 4.4 and 4.6 show the BER performance while figures 4.3 and 4.5 show the MSE at 16 and 24 pilots, respectively. Comparing them, we see that increasing the number of pilots improves the LS estimate, with the frequency domain method wading better than the time domain method.

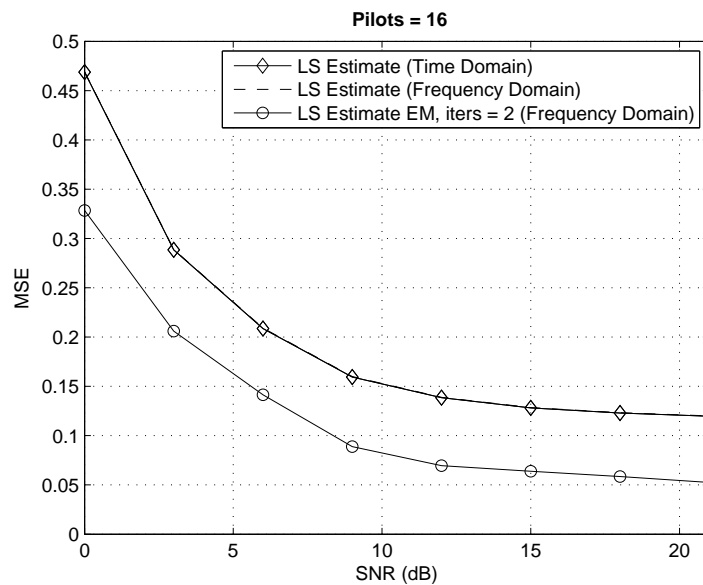


Figure 4.3: MSE comparison EM based least squares (16 pilots).

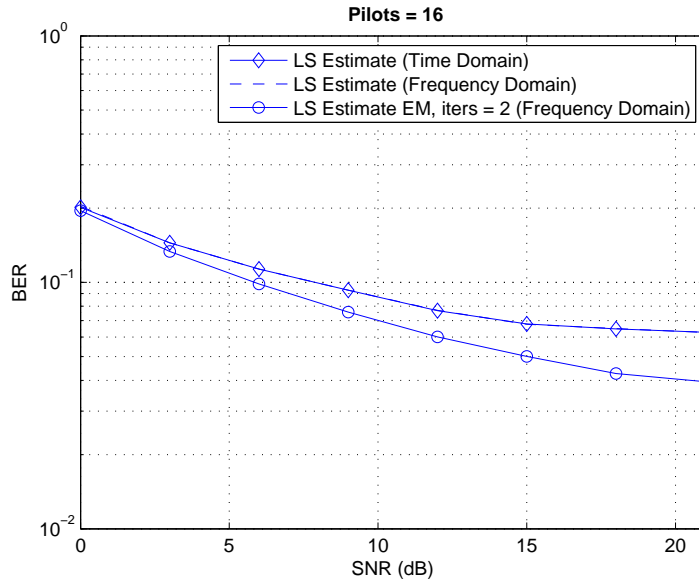


Figure 4.4: BER comparison EM based least squares (16 pilots).

### 4.4.3 Kalman Filter based Receivers

Figure 4.7(a) compares the BER performance of frequency domain Forward Kalman, Cyclic and Helical Kalman filters with the time domain LS method and Helix Kalman for the uncoded case at 16 pilots. As expected, we see that using Kalman filter improves the EM based estimate in the frequency domain. We also see that Helix based Kalman performs better than other frequency domain based techniques and that for the uncoded 16 pilot case, the frequency domain methods fairs better than the time domain methods.

Figure 4.7(b) shows the same comparison for 24 pilots uncoded case. For the case of 24 pilots, we note that though the time domain estimate methods perform better than frequency domain methods, the performance of the frequency domain Helix Kalman is comparable to the time domain Helix Kalman. Figure 4.8 compares the BER performances of frequency domain channel estimation of

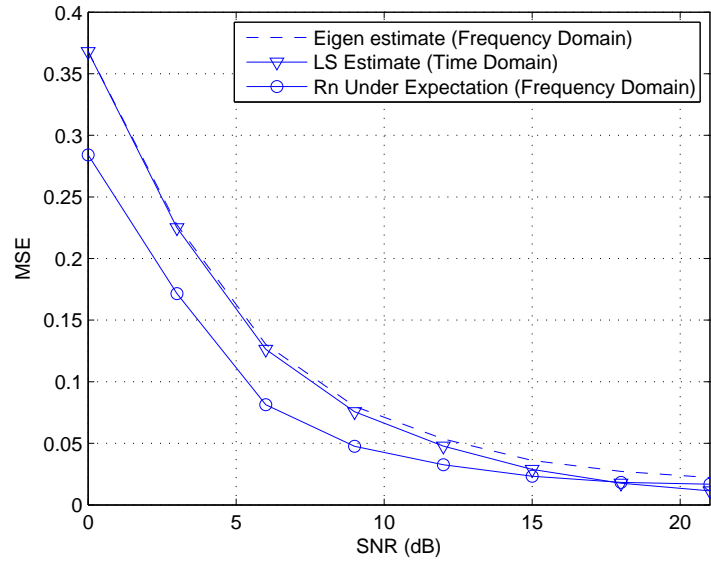


Figure 4.5: MSE comparison EM based least squares (24 pilots).

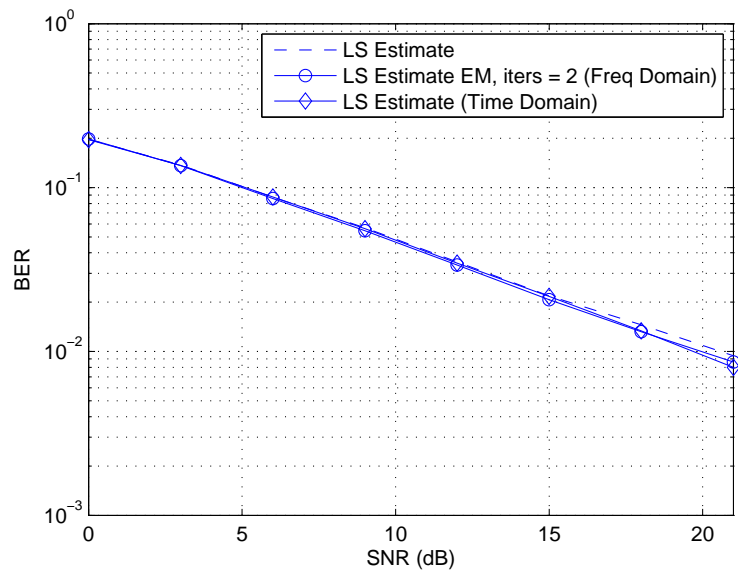
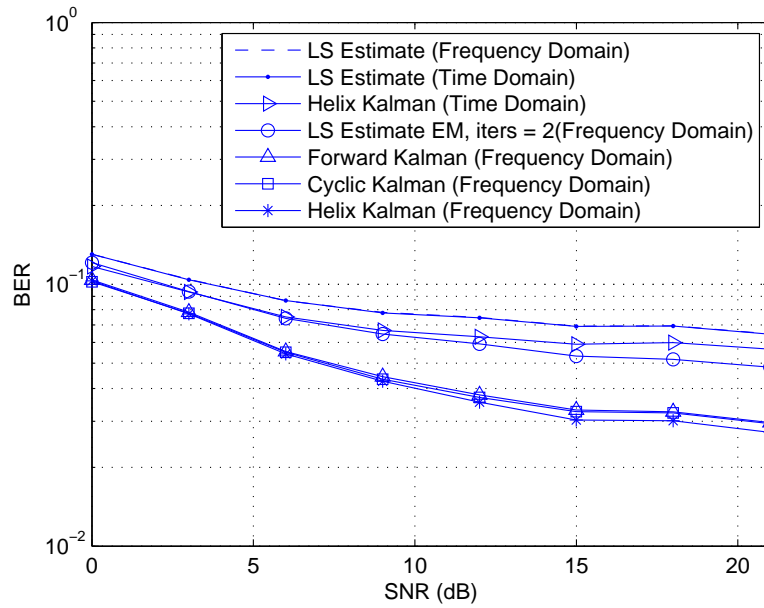
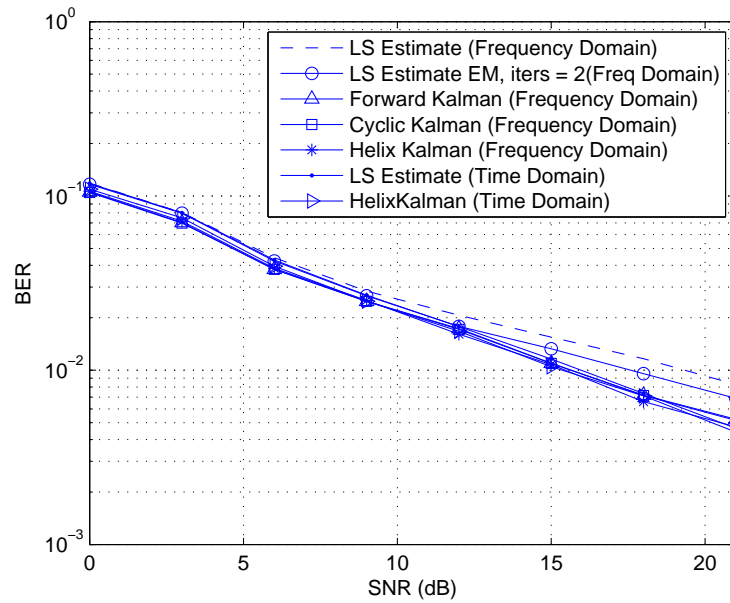


Figure 4.6: BER comparison EM based least squares (24 pilots).



(a)



(b)

Figure 4.7: BER comparison for various uncoded freq. domain methods (a) using 16 pilots (b) using 24 pilots.

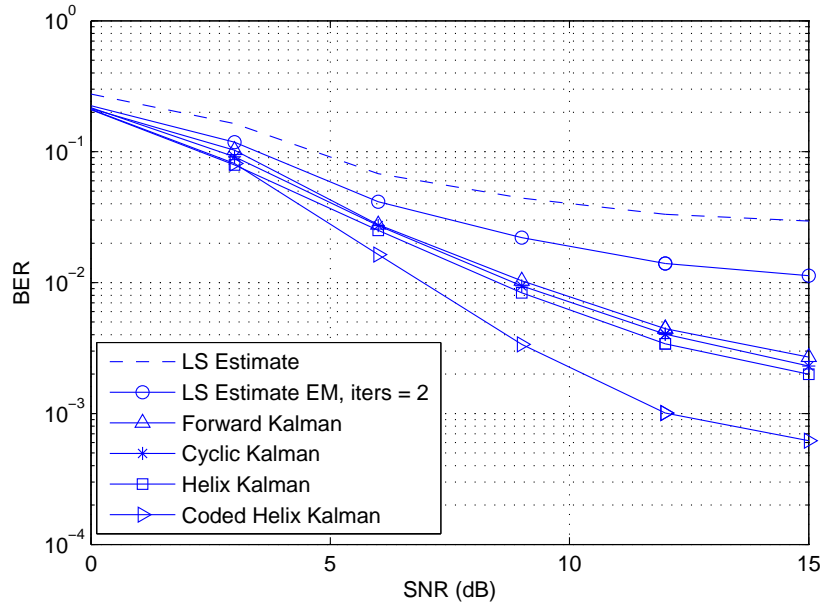


Figure 4.8: BER comparison for frequency domain coded methods (16 pilots).

various Kalman filters with the LS and LS EM estimate for the 16 pilot case. Here we utilize the outercode to enhance the estimate. We see that the code enhancement technique is superior to the rest of the techniques. Figure 4.9 shows the result of the comparison of frequency domain Helix Kalman and coded Kalman with the time domain Helix Kalman (16 pilots). We can see that for the multiple access case, the frequency domain technique fairs better than the time domain estimation method, while the coded Kalman outperforms all other techniques.



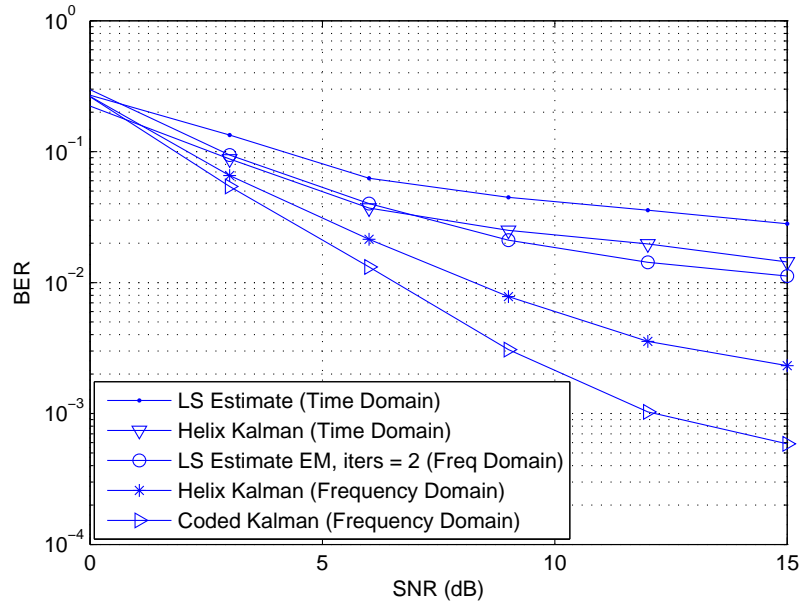


Figure 4.9: BER comparison for the coded case (16 pilots).

In order to see a fair comparison between the time domain and the frequency domain techniques for a multiple access system, we compare the frequency domain Helix Kalman with the time domain Helix Kalman obtained from the procedure outlined in Section 4.3. We plot figure 4.10 for 24 pilots with using the outercode and employing 6 eigenvalues per section to estimate the channel in the frequency domain. As we can see from the figure, the frequency domain Helix Kalman outperforms the time domain Helix Kalman.

#### 4.4.4 Pilot Design

From figures 4.4 and 4.6, it can be established that pilot density has a profound effect on the channel estimation algorithm. It will be worthwhile to investigate the effect of pilot pattern on the channel estimation algorithm as well. Here we

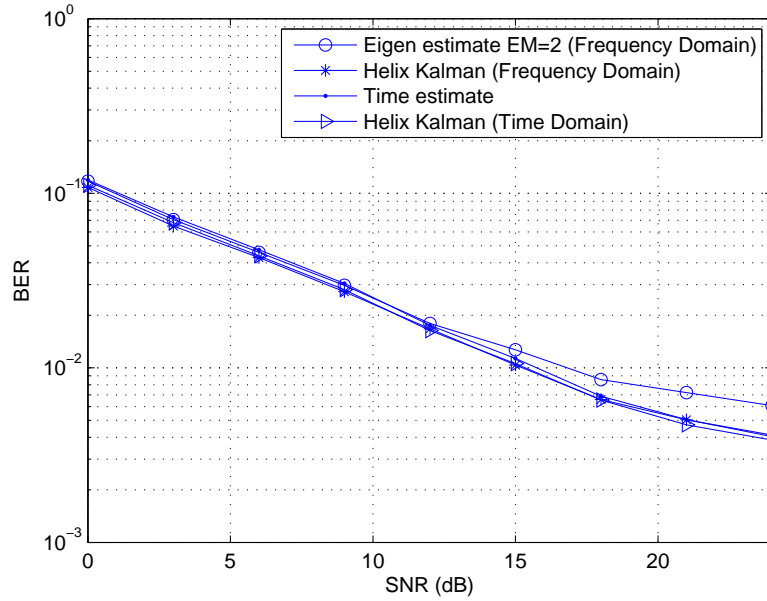


Figure 4.10: BER comparison of time and frequency domain uncoded techniques (24 pilots).

find the optimal pilot pattern that minimizes the MSE of the estimate for the pilot placement, given by (4.11), in the frequency domain. Consider the case of 16 pilots, with an OFDM symbol of length 64. Considering 4 users, each user will have access to 16 frequency bins. Assuming the pilots be equally divided among all the users, the spectrum available to every user will have 4 pilots each. This means there are 1820 different combination of pilot patterns that are possible. We perform an exhaustive search for the minimum MSE for all 1820 pilot patterns. Figure 4.11 shows that the minimum MSE occurs at combinations 682 and 1010, both of which are equispaced combinations. Since we are consider a block OFDM system, we have another option of varying the pilot pattern over time, i.e., change the pilot positions in subsequent OFDM symbols. The equidistant pilot placement has already been established as better to the other pilot placement patterns. Now

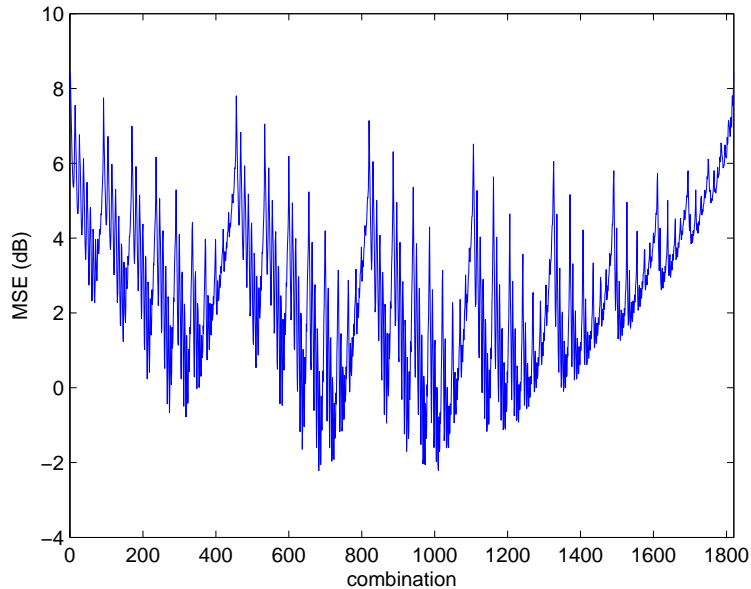


Figure 4.11: MSE for all pilot patterns.

we compare the BER performance of OFDM blocks which have the same equidistance patterns with that of those which have alternating equidistance pilots. We find that alternating pilot pattern provides a slightly better BER performance than the non alternating scheme. Figure 4.12 shows the BER comparison for equidistance non alternating and alternating schemes.

## 4.5 Conclusion

We present an OFDM receiver design based on a semi-blind low complexity frequency domain channel estimation algorithm for multi-access OFDM system. Opposed to the time domain case which estimates the whole spectrum, we propose a frequency domain approach in which the user estimates the part of the spectrum in which he operates. The advantage of this is reduction in computational cost

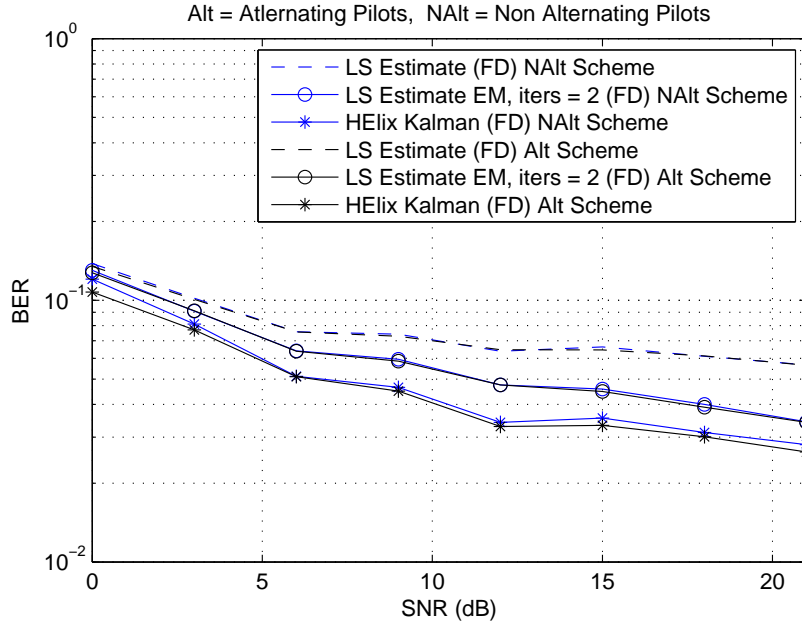


Figure 4.12: BER comparison for alternating and non alternating equidistant schemes (16 pilots) .

incurred by each user. Also, the user might not have access to the entire spectrum. We estimate the channel parameters based on the eigenvalue technique, greatly reducing the number of parameters to be estimated. The receiver uses the pilots to kick start the estimation process and then iterates between channel and data recovery. Our receiver utilizes data (finite alphabet set, code, transmit precoding, pilots) and channel (finite delay spread, frequency correlation, time correlation) constraints. Thanks to the decoupled relation in the frequency domain, data recovery is done on an element by element basis while the channel estimation boils down to solving a regularized least squares problem. We propose to improve the estimate making use of the time correlation information of the channel by relaxing the latency requirement. For this purpose, we employ Cyclic and Helix based FB Kalman filters and use the outer code to enhance the channel estimate. We make

use of both the frequency and time correlation which results in a relatively low training overhead. The simulation results show the performance of our algorithm. Our results maybe extended to multiple antenna OFDM systems.

## APPENDIX A: Moment Calculation

Now to calculate an expectation of the form  $E[\underline{\mathbf{D}}\underline{\mathbf{B}}\underline{\mathbf{D}}]$ , which appears in (4.16), we note that by our assumption different elements of  $\underline{\mathbf{D}}$  are independent making the expectation that involves them in  $E[\underline{\mathbf{D}}\underline{\mathbf{B}}\underline{\mathbf{D}}]$  separable, i.e. for these terms, we have

$$E[\underline{\mathbf{D}}\underline{\mathbf{B}}\underline{\mathbf{D}}^*] = E[\underline{\mathbf{D}}]\underline{\mathbf{B}}E[\underline{\mathbf{D}}^*] \quad (\text{A-1})$$

The identical forms, however, interact according to

$$E[\underline{\mathbf{D}}\underline{\mathbf{B}}\underline{\mathbf{D}}] = E[\underline{\mathbf{D}}\text{diag}(\underline{\mathbf{B}})\underline{\mathbf{D}}] \quad (\text{A-2})$$

$$= E[\underline{\mathbf{D}}\underline{\mathbf{D}}^*]E[\text{diag}(\underline{\mathbf{B}})] \quad (\text{A-3})$$

By combining (A-1) and (A-3), we see that

$$E[\underline{\mathbf{D}}\underline{\mathbf{B}}\underline{\mathbf{D}}^*] = E[\underline{\mathbf{D}}]\underline{\mathbf{B}}E[\underline{\mathbf{D}}^*] + \text{Cov}[\underline{\mathbf{D}}]\text{diag}(\underline{\mathbf{B}})$$

## Part II

Least Mean Square Adaptive

Filters with Optimum Error

Nonlinearity

## CHAPTER 5

# INTRODUCTION TO ADAPTIVE FILTERS

Adaptive filters are time variant systems that learn from their environment and adapt to the variations in the signal statistics. Adaptive filters are useful whenever we need to process a signal arising from unknown statistics. In the realm of modern communication, the role of adaptive filters is of vital importance. For instance, adaptive filters have been extremely important in achieving the high efficiency and high reliability on ubiquitous telecommunication services.

In this second part of the thesis, we show how to design adaptive filters with optimum nonlinearities from a priori information about the noise statistics. This part is in turn divided into two chapters. The first chapter provides a quick overview of adaptive filters with error nonlinearities. It explains how the performance of adaptive filters is evaluated (and the assumptions needed for that) and introduces the energy relation as a convenient tool for accessing performance.

## 5.1 System Model

Consider the following system identification scenario

$$d(i) = \mathbf{u}_i \mathbf{w}^o + v(i) \quad (5.1)$$

where  $d(i)$  is the desired signal response,  $\mathbf{u}_i$  is an input row regression vector,  $\mathbf{w}^o$  denotes an unknown column vector that we wish to estimate and  $v(i)$  is the measurement noise. Various adaptive schemes ([62], [63]) have been proposed in literature for the estimation for  $\mathbf{w}^o$ . The general form of the recursive update equation used by these schemes can be represented as [62]

$$\mathbf{w}_{i+1} = \mathbf{w}_i + \mu \mathbf{u}_i^T f[e(i)] \quad (5.2)$$

where  $\mathbf{w}_i$  is the estimate of  $\mathbf{w}^o$  at time  $i$ ,  $\mu$  is the step-size and

$$e(i) = d(i) - \mathbf{u}_i \mathbf{w}_i \quad (5.3)$$

is the estimation error. The nonlinearity  $f[e(i)]$  controls the correction term in (5.2) and is known as the scalar error nonlinearity.

This class of algorithms is general enough to include the special cases listed in Table 5.1. Several of these algorithms were already considered in the literature (see, e.g., [64, 66, 67, 68] and the many references therein).



Table 5.1: *Examples for  $f[e(i)]$ .*

ALGORITHM	ERROR NONLINEARITIES $f[e(i)]$
LMS	$e(i)$
LMF	$e^3(i)$
LMF family	$e^{2k+1}(i)$
LMMN	$ae(i) + be^3(i)$
Sign error	$\text{sgn}[e(i)]$
Sat. nonlin.	$\int_0^{e(i)} \exp\left(-\frac{z^2}{2\sigma_{\text{sat}}^2}\right) dz$

## 5.2 Evaluating Adaptive Filters

As we shall see in the literature review section, many adaptive algorithms have been suggested in literature. To evaluate the performance of these algorithms, various error measures and performance criteria are used which are summarized in this section. In this section, we also summarize the various assumptions that have been used to evaluate the performance of adaptive filters.

### 5.2.1 Error Measures

Evaluating the performance of an adaptive filter deals with a study of the time-evolution and the steady-state values of  $E[|e(i)|^2]$  and  $E[||\mathbf{v}_i||^2]$ <sup>1</sup>, where  $\mathbf{v}_i$  is the weight error vector defined as

$$\mathbf{v}_i = \mathbf{w}^o - \mathbf{w}_i \quad (5.4)$$

The steady-state values of  $E[|e(i)|^2]$  and  $E[||\mathbf{v}_i||^2]$  represent the mean-square-error and the mean-square-deviation (MSD) performances of an adaptive filter, respectively, whereas their time-evolution relate to the learning or the transient

---

<sup>1</sup>The notation  $||\mathbf{v}_i||^2$  denotes the squared Euclidean norm of a vector, i.e.,  $||\mathbf{v}_i||^2 = \mathbf{v}_i \mathbf{v}_i^T$ .

behavior of the filter. In carrying out the performance evaluation of adaptive filters, it is convenient to define the a-priori and a-posteriori estimation errors [62]

$$e_a(i) = \mathbf{u}_i \mathbf{v}_i, \quad \text{and} \quad e_p(i) = \mathbf{u}_i \mathbf{v}_{i+1} \quad (5.5)$$

The estimation error  $e(i)$  and the a-priori error  $e_a(i)$  are related by

$$e(i) = e_a(i) + v(i) \quad (5.6)$$

## 5.2.2 Various Assumptions used for Evaluating Performance of an Adaptive Filter

Various assumptions and techniques have been employed in the literature to characterize the performance measure of adaptive filters such as linearization [66, 67, 69], restricted class of nonlinearities [70, 71, 72, 73], assumptions on the statistics of the errors [66, 70, 74, 75], restricted class of inputs [66, 70, 71], independence assumption [63], Gaussian noise [66, 70, 74] (see [76] for more details). Most techniques in the literature use a combination of these assumptions.

## 5.2.3 Performance Criteria

There are different ways of evaluating the performance of an adaptive filter. Most common performance measures of an adaptive filter are the following:

*a) Convergence Speed.* Convergence speed of an adaptive algorithm is an important performance measure. It shows how fast an adaptive algorithm converges to its steady-state. Different adaptive algorithms have been designed in order to improve the convergence speed of adaptive filters. For example, the LMF algorithm of [67] employs a nonlinearity of third error norm to give a faster convergence compared to the LMS algorithm, while the NCLMS algorithm of [77] enhances the convergence performance of the LMS algorithm by applying a constraint based on noise variance. The NLMS [78] and the NLMF [79] algorithms enhance the speed of the LMS in the presence of correlated input.

*b) Mean and Mean-square-error Stability.* Stability of an adaptive algorithm is another important issue. The stability of an adaptive algorithm is analyzed both in the mean and mean-square-error sense [62, 63] where we require the various error measure to remain bounded. For thi to happen, we usually choose the step size of the adaptive filter to be small enough.

*c) Steady-state Behavior.* Steady-state behavior is yet another important performance measure of adaptive algorithms. Usually two measures of performance indices are of interest: steady-state excess mean-square error (EMSE) and steady-state mean-square-deviation (MSD) defined respectively as

$$\text{EMSE} = \lim_{i \rightarrow \infty} E[|e(i)|^2] - \sigma_v^2 \quad (5.7)$$

$$\text{MSD} = \lim_{i \rightarrow \infty} E[|\mathbf{v}_i|^2] \quad (5.8)$$

This is true assuming noise is iid and independent of input.

### 5.3 Fundamental Energy Relation

The fundamental energy relation is a relation that makes it possible to analyze adaptive filters under weak assumptions. The relation was originally developed in [80, 81, 82] in the context of robustness analysis of adaptive filters within a deterministic framework, it has since been used in [64, 83, 84] to study both the transient and the steady state performance of adaptive filters. The energy conservation relation enables us to avoid many of the assumptions mentioned in Section 5.2.2. We can thus study adaptive filters under the most general conditions.

The energy relation is easy to develop. Using error measures defined in (5.5), we rewrite the adaptive algorithm (5.2) as follows

$$\mathbf{w}^o - \mathbf{w}_{i+1} = \mathbf{w}^o - (\mathbf{w}_i + \mu \mathbf{u}_i^T f[e(i)]) \quad (5.9)$$

$$\mathbf{v}_{i+1} = \mathbf{v}_i - \mu \mathbf{u}_i^T f[e(i)] \quad (5.10)$$

multiplying both sides of equation (5.10) by  $\mathbf{u}_i$  and keeping in view the definitions of (5.5), we obtain the relation between  $e_p(i)$ ,  $e_a(i)$ ,  $e(i)$  as

$$\mathbf{u}_i \mathbf{v}_{i+1} = \mathbf{u}_i \mathbf{v}_i - \mu \mathbf{u}_i \mathbf{u}_i^T f[e(i)] \quad (5.11)$$

$$e_p(i) = e_a(i) - \mu f[e(i)] \|\mathbf{u}_i\|^2 \quad (5.12)$$

Solving the above equation for  $\mu f[e(i)]$  and substituting it in (5.10) leads to the energy conservation relation

$$\|\mathbf{v}_{i+1}\|^2 + \frac{|e_a(i)|^2}{\|\mathbf{u}_i\|^2} = \|\mathbf{v}_i\|^2 + \frac{|e_p(i)|^2}{\|\mathbf{u}_i\|^2} \quad (5.13)$$

The above relation is called the fundamental energy conservation relation [83]. It shows the time evolution of the energies of the error quantities and enables us to perform steady state analysis in a transparent manner. Replacing the a-posteriori error  $e_p(i)$  in (5.13) by its equivalent expression in (5.12) results in the following energy relation

$$\|\mathbf{v}_{i+1}\|^2 = \|\mathbf{v}_i\|^2 - 2\mu e_a(i)f[e(i)] + \mu^2\|\mathbf{u}_i\|^2 f^2[e(i)] \quad (5.14)$$

The above equation applies to the class of adaptive filters given by (5.2)-(5.3). It is an exact relationship, as we we have not used any assumption or approximation in deriving it. Consequently, the analysis based on the fundamental energy relation will be more general and rigorous. This relation will serve as the starting point for our discussion on the design of optimum error nonlinearities in the next chapter.

## 5.4 A Brief Overview of Previous Work

The LMS algorithm is a simple algorithm used to update adaptive filter coefficients. Due to its success, many variants of this algorithms have been suggested in literature to improve its steady-state error, speed of convergence or to reduce the

computational complexity. These variations can be classified into adaptive filters that employ an error nonlinearity and those that employ a data nonlinearity in the update term. Here we will discuss those adaptive filters which are nonlinear in error only.

In the past, researchers have designed optimum error nonlinearities by first analyzing the mean and the mean-square behavior of adaptive filter for general error nonlinearities and then by optimizing the result based on the analysis. However, as the analysis is based on restrictive assumptions (e.g. Gaussian/white input, Gaussian noise, input regressor independence, linearization of nonlinearity), the optimum error nonlinearity obtained is limited by the assumptions used for the analysis. Most of these techniques can be divided into two categories:

1. Techniques with intuitively suggested error nonlinearity function [67, 78, 74, 85, 87] based on intuitive arguments.
2. Techniques which derive the optimum error nonlinearity functions under certain assumptions [68, 86, 88].

In this part, our aim is to design a more general optimum error nonlinearity in the steady-state by relaxing some of the assumptions previously used like the assumption on the distribution of input regressor elements and assumption on the distribution of noise by avoiding any linearization arguments. The optimum nonlinearities obtained thus will not only be more general but will also encompass more diverse scenarios of adaptive filtering. Thus, our approach first performs a steady state mean-square analysis for a general error nonlinearity and then

optimizes the choice of the nonlinearity by minimizing the steady state error. In analyzing the optimum error nonlinearity, we rely on the fundamental energy relation which is discussed next.

## CHAPTER 6

# OPTIMUM ERROR NON LINEARITY–AT STEADY STATE

As we pointed out in the previous chapter, many adaptive algorithms have been proposed in literature employing various kind of error nonlinearities. These nonlinearities were obtained under various restrictions and assumptions or were even simply intuitively motivated. Here, we use the energy relation to derive the EMSE from adaptive filters employing a general error nonlinearity under quite general assumptions. We use that in Section 6.3 to derive the optimum error nonlinearity and in Section 6.5 to derive the conditional error nonlinearity (both of which turn to be a function of the additive noise). In Section 6.4 we demonstrate how the nonlinearity manifests itself for various noise distributions. Finally, Section 6.9 demonstrates the performance of the optimum nonlinearity in various scenarios.



## 6.1 The MSE for General Error Nonlinearity

In this section, we use the energy relation to derive the mean square error for the general error nonlinearity. Specifically, starting from the energy relation (5.14) and taking expectation of both sides we obtain

$$E[||\mathbf{v}_{i+1}||^2] = E[||\mathbf{v}_i||^2] - 2\mu E[e_a(i)f[e(i)]] + \mu^2 E[||\mathbf{u}_i||^2 f^2[e(i)]] \quad (6.1)$$

We assume the filter to be stable, so it will eventually reach steady-state. Such that at  $i \rightarrow \infty$ ,  $E[||\mathbf{v}_{i+1}||^2] = E[||\mathbf{v}_i||^2]$ . Then at steady state, (6.1) becomes

$$\lim_{i \rightarrow \infty} E[e_a(i)f[e(i)]] = \frac{\mu}{2} \lim_{i \rightarrow \infty} E[||\mathbf{u}_i||^2 f^2[e(i)]] \quad (6.2)$$

In order to evaluate the two expectations in (6.2), we will introduce the following assumptions:

**AN.** The noise sequence  $\{v(i)\}$  is independent, identically distributed, and independent of the input sequence  $\{\mathbf{u}_i\}$ .

**AG.** The filter is long enough such that  $e_a(i)$  is Gaussian.

**AU.** The random variables  $||\mathbf{u}_i||^2$  and  $f^2[e(i)]$  are asymptotically uncorrelated, i.e.,

$$\lim_{i \rightarrow \infty} E[||\mathbf{u}_i||^2 f^2[e(i)]] = E[||\mathbf{u}_i||^2] \lim_{i \rightarrow \infty} E[f^2[e(i)]] \quad (6.3)$$

Assumptions **AG** and **AU** get more realistic with the increase in the length of the filter. We can use the central limit theorem to justify assumption **AG**. Assumption

**AU** is a weaker assumption than the independence assumption usually employed in literature. We can justify it for long filters using an ergodic argument on  $\|\mathbf{u}_i\|^2$ .

With these assumptions let's first evaluate the left hand side of (6.2). TO do so, we will use Price's Theorem [65]

$$E[xf[y+z]] = \frac{E[xy]}{E[y^2]}E[f[y+z]]$$

where  $x$  and  $y$  are jointly Gaussian and independent of  $z$ . Using this result, together with assumptions **AN** and **AG**, we can rewrite the left hand side of (6.2) as

$$\begin{aligned} E[e_a(i)f[e(i)]] &= E[e_a(i)f[e_a(i) + v(i)]] \\ &= E[e_a^2(i)] \frac{E[e_a(i)f[e_a(i) + v(i)]]}{E[e_a^2(i)]} \end{aligned} \quad (6.4)$$

where the expectation  $E[e_a(i)f[e(i)]]$  depends on  $e_a(i)$  through the second moment  $E[e_a^2(i)]$  *only* and therefore the ratio in (6.4) is a function of  $E[e_a^2(i)]$ . This leads us to define

$$h_G[E[e_a^2(i)]] \triangleq \frac{E[e_a(i)f[e(i)]]}{E[e_a^2(i)]} \quad (6.5)$$

Combining (6.4) and (6.5), we get

$$E[e_a(i)f[e(i)]] = E[e_a^2(i)]h_G[E[e_a^2(i)]] \quad (6.6)$$

Now employing the assumption **AU**, we see that the expectation on the right hand side of (6.2) can be split as

$$E[||\mathbf{u}_i||^2 f^2[e(i)]] = E[||\mathbf{u}_i||^2] E[f^2[e(i)]] \quad (6.7)$$

Also as  $e_a(i)$  is zero mean Gaussian by assumption **AG** and independent of the noise, so  $E[f^2[e(i)]]$  will also depend on  $e_a(i)$  through its *second moment only*. Thus, we define

$$h_U[E[e_a^2(i)]] \triangleq E[f^2[e(i)]] \quad (6.8)$$

Combining (6.7) and (6.8) yields

$$E[||\mathbf{u}_i||^2 f^2[e(i)]] = E[||\mathbf{u}_i||^2] h_U[E[e_a^2(i)]] \quad (6.9)$$

Replacing the expectations of (6.2) by those in (6.6) and (6.9), the energy relation takes on the form

$$\lim_{i \rightarrow \infty} E[e_a^2(i)] \lim_{i \rightarrow \infty} h_G[E[e_a^2(i)]] = \frac{\mu}{2} E[||\mathbf{u}_i||^2] \lim_{i \rightarrow \infty} h_U[E[e_a^2(i)]] \quad (6.10)$$

The above relation was derived for a general memoryless error nonlinearity. Next, we calculate the excess mean square error using these results.

## 6.2 Excess Mean Square Error

To calculate the excess mean square error, we rearrange (6.10) as

$$\begin{aligned}\lim_{i \rightarrow \infty} E[e_a^2(i)] &= \frac{\mu}{2} E[\|\mathbf{u}_i\|^2] \frac{\lim_{i \rightarrow \infty} h_U[E[e_a^2(i)]]}{\lim_{i \rightarrow \infty} h_G[E[e_a^2(i)]]} \\ &= \frac{\mu}{2} \text{Tr}(\mathbf{R}) \frac{\lim_{i \rightarrow \infty} h_U[E[e_a^2(i)]]}{\lim_{i \rightarrow \infty} h_G[E[e_a^2(i)]]}\end{aligned}\quad (6.11)$$

as both  $h_U$  and  $h_G$  are analytic in their arguments and denoting the excess mean-square error by  $S = \lim_{i \rightarrow \infty} E[e_a^2(i)]$ , we can write

$$\begin{aligned}\lim_{i \rightarrow \infty} h_U[E[e_a^2(i)]] &= h_U[S] \\ \lim_{i \rightarrow \infty} h_G[E[e_a^2(i)]] &= h_G[S]\end{aligned}\quad (6.12)$$

This means that the EMSE satisfies the nonlinear relationship

$$S = \frac{\mu}{2} \text{Tr}(\mathbf{R}) \frac{h_U[S]}{h_G[S]}\quad (6.13)$$

The EMSE can then be calculated by evaluating  $h_U$  and  $h_G$  for a given error nonlinearity  $f$  and solving for the fixed point equation (6.13).

### 6.3 Optimum Choice of the Nonlinearity

Consider the expression of the excess mean-square error (6.13) rewritten as

$$S = \frac{\mu}{2} \text{Tr}(\mathbf{R}) \frac{E[f^2[e(i)]]}{E[f'[e(i)]]} \quad (6.14)$$

Now the excess mean-square error cannot be less than the Cramer-Rao bound of the underlying estimation process (i.e., estimating  $\mathbf{u}_i \mathbf{w}^o$  from  $\mathbf{u}_i \mathbf{w}_i$ ). Thus we can write

$$\frac{E[f^2[e(i)]]}{E[f'[e(i)]]} \geq \frac{2}{\mu \text{Tr}(\mathbf{R})} \alpha = \alpha' \quad (6.15)$$

where  $\alpha'$  is non zero because the adaptive filters does not have infinite memory as we are using an adaptive filter with non-vanishing step size. The  $\alpha$  used in equation (6.16) is nothing but the EMSE at steady state. Now denoting the pdf of  $e(i)$  with  $p_e$  and assuming it is differentiable, we claim that the nonlinearity

$$f[e(i)] = -\alpha' \frac{p'_e[e(i)]}{p_e[e(i)]} \quad (6.16)$$

is optimum in the sense that it attains the lower bound on the EMSE. To prove this claim, let us evaluate the numerator and denominator of (6.15) for the optimum

choice of  $f$ . The numerator will be

$$\begin{aligned}
E[f^2[e(i)]] &= \int_{-\infty}^{\infty} f^2[e(i)]p_e[e(i)]de(i) \\
&= (\alpha')^2 \int_{-\infty}^{\infty} \left( \frac{p'_e[e(i)]}{p_e[e(i)]} \right)^2 p_e[e(i)]de(i) \\
&= (\alpha')^2 \int_{-\infty}^{\infty} \frac{(p'_e[e(i)])^2}{p_e[e(i)]} de(i)
\end{aligned} \tag{6.17}$$

and the denominator will be

$$\begin{aligned}
E[f'[e(i)]] &= \int_{-\infty}^{\infty} f'[e(i)]p_e[e(i)]de(i) \\
&= f[e(i)]p_e[e(i)]|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f[e(i)]p'_e[e(i)]de(i)
\end{aligned} \tag{6.18}$$

Now for the same choice of  $f$ , we have

$$\begin{aligned}
E[f'[e(i)]] &= -\alpha' p'_e[e(i)]|_{-\infty}^{\infty} + \alpha' \int_{-\infty}^{\infty} \frac{(p'_e[e(i)])^2}{p_e[e(i)]} de(i) \\
&= \alpha' \int_{-\infty}^{\infty} \frac{(p'_e[e(i)])^2}{p_e[e(i)]} de(i)
\end{aligned} \tag{6.19}$$

assuming  $p'_e$  decays to zero as  $e(i)$  reaches  $\pm\infty$ . Using (6.17) and (6.19) we attain the desired lower bound

$$\frac{E[f^2[e(i)]]}{E[f'[e(i)]]} = \alpha' \tag{6.20}$$

which validates our claim of (6.16). As  $\alpha$  is just EMSE at steady-state, therefore the value of  $\alpha'$  in equation (6.16) can be set as

$$\alpha' = \frac{2\sigma_{ea}^2}{\mu \text{Tr}(\mathbf{R})} \quad (6.21)$$

The optimum error nonlinearity is thus given by

$$f[e(i)] = -\frac{2\sigma_{ea}^2}{\mu \text{Tr}(\mathbf{R})} \frac{p'_e[e(i)]}{p_e[e(i)]} \quad (6.22)$$

## 6.4 Optimum Error Nonlinearity for Some Special Cases

In the previous section, we derived the non linearity in (6.22) under weaker assumption as compared to what is already available in literature. For example, we have not assumed any restriction on the statistics of the noise or the symmetry of its pdf. The nonlinearity (6.22) also holds true irrespective of the input statistics and color. We have not used any linearization argument in deriving (6.22) which makes our result accurate over all stages of adaptation. Note that assumption **AU** that assumes  $\|\mathbf{u}_i\|^2$  and  $f^2[e(i)]$  asymptotically uncorrelated which is a weaker condition than the independence assumption usually taken in literature. Also this expression is generic as we have not placed any restriction on the class of the error nonlinearities as is sometimes done in the literature.

Now  $e(i)$  is the sum of two independent variables  $e_a(i)$  and  $v(i)$  and its pdf is

given by the convolution of the two pdf's. By assumption **AG**, we know the pdf of  $e_a(i)$  is zero mean Gaussian. Thus we just need to model the noise statistics in order to complete the modeling of  $p_e$

$$\begin{aligned} p_e[e(i)] &= p_{e_a}[e(i)] * p_v[e(i)] \\ &= \frac{1}{\sqrt{2\pi\sigma_{e_a}^2}} e^{-\frac{e^2}{2\sigma_{e_a}^2}} * p_v[e(i)] \end{aligned} \quad (6.23)$$

where  $*$  is the convolutional operator. Let us see how the error nonlinearity will manifest itself for different noise statistics.

#### 6.4.1 Gaussian Noise

For the case when noise  $v(i)$  is Gaussian with mean zero and variance  $\sigma_v^2$ , and since  $e_a(i)$  is also Gaussian with mean zero and variance  $\sigma_{e_a}^2$  the error  $e(i)$  will be Gaussian with zero mean and variance  $\sigma_e^2 = \sigma_v^2 + \sigma_{e_a}^2$ . The pdf of  $e(i)$  is thus given as

$$p_e[e(i)] = \frac{1}{\sqrt{2\pi(\sigma_{e_a}^2 + \sigma_v^2)}} \exp \left[ \frac{-e^2(i)}{2(\sigma_v^2 + \sigma_{e_a}^2)} \right] \quad (6.24)$$

and its derivative is given as

$$p'_e[e(i)] = -\frac{e(i)}{(\sigma_v^2 + \sigma_{e_a}^2)\sqrt{2\pi(\sigma_{e_a}^2 + \sigma_v^2)}} \exp \left[ \frac{-e^2(i)}{2(\sigma_v^2 + \sigma_{e_a}^2)} \right] \quad (6.25)$$



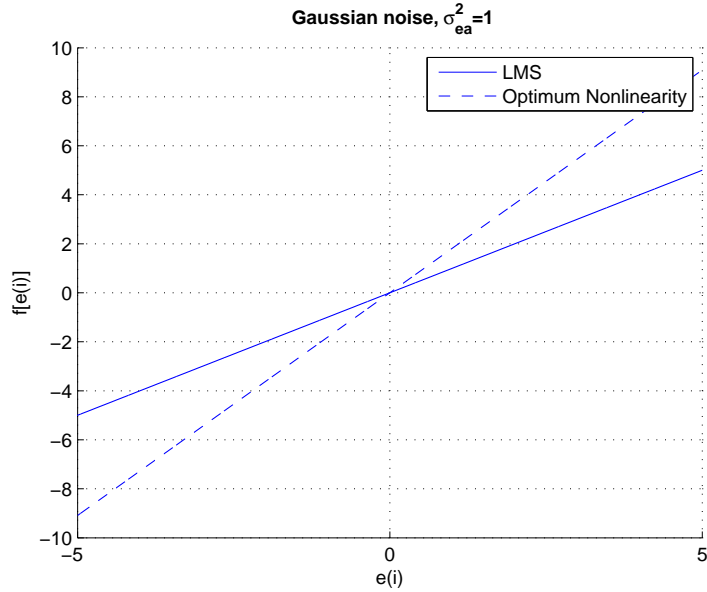


Figure 6.1: Error nonlinearity for the Gaussian case

Replacing (6.24) and (6.25) in (6.22), we get the optimum nonlinearity for the *Gaussian* noise

$$f_{opt}[e(i)] = \frac{2\sigma_{ea}^2}{\mu\text{Tr}(\mathbf{R})(\sigma_{ea}^2 + \sigma_v^2)}e(i) \quad (6.26)$$

The shape of the nonlinearity is shown in Figure 6.1

## 6.4.2 Laplacian Noise

For the case of Laplacian noise, the pdf of the noise is given by

$$p_v[e(i)] = \frac{1}{2b}e^{-\frac{|v|}{b}} \quad (6.27)$$

The constant  $b$  is related to the noise variance as

$$\sigma_v^2 = 2b^2 \quad \text{or} \quad b = \sqrt{\frac{\sigma_v^2}{2}} \quad (6.28)$$

Substituting (6.27) in (6.23) yields

$$p_e[e(i)] = \frac{1}{\sqrt{2\pi\sigma_{ea}^2}} e^{-\frac{e^2}{2\sigma_{ea}^2}} * \frac{1}{2b} e^{-\frac{|v|}{b}} \quad (6.29)$$

The convolution is evaluated in Appendix A. After evaluating  $p'_e[e(i)]$  we arrive at the following expression for the nonlinearity

$$f_{opt}[e(i)] = -\frac{2\sigma_{ea}^2}{\mu\text{Tr}(\mathbf{R})} \left\{ \frac{e^{e(i)/b}\gamma_1[e(i)] - e^{-e(i)}g_-[e(i)]}{e^{e(i)/b}\left(1 - \text{erf}\left[\frac{b e(i) + \sigma_{ea}^2}{b\sqrt{2\sigma_{ea}^2}}\right]\right) + e^{-e(i)/b}\left(1 + \text{erf}\left[\frac{b e(i) - \sigma_{ea}^2}{b\sqrt{2\sigma_{ea}^2}}\right]\right)} \right\} \quad (6.30)$$

where

$$\gamma_1[e(i)] = \frac{1}{b} - \frac{1}{b}\text{erf}\left[\frac{b e(i) + \sigma_{ea}^2}{b\sqrt{2\sigma_{ea}^2}}\right] - \sqrt{\frac{2}{\pi\sigma_{ea}^2}} e^{(b e(i) + \sigma_{ea}^2)^2 / 2b^2\sigma_{ea}^2} \quad (6.31)$$

$$g_-[e(i)] = \frac{1}{b} + \frac{1}{b}\text{erf}\left[\frac{b e(i) - \sigma_{ea}^2}{b\sqrt{2\sigma_{ea}^2}}\right] - \sqrt{\frac{2}{\pi\sigma_{ea}^2}} e^{(b e(i) - \sigma_{ea}^2)^2 / 2b^2\sigma_{ea}^2} \quad (6.32)$$

The general shape of the nonlinearity is demonstrated in Figure 6.2

### 6.4.3 Binary Noise

In the Binary noise case, we have

$$v(i) = \begin{cases} b & \text{with probability } 1/2 \\ -b & \text{with probability } 1/2 \end{cases} \quad (6.33)$$

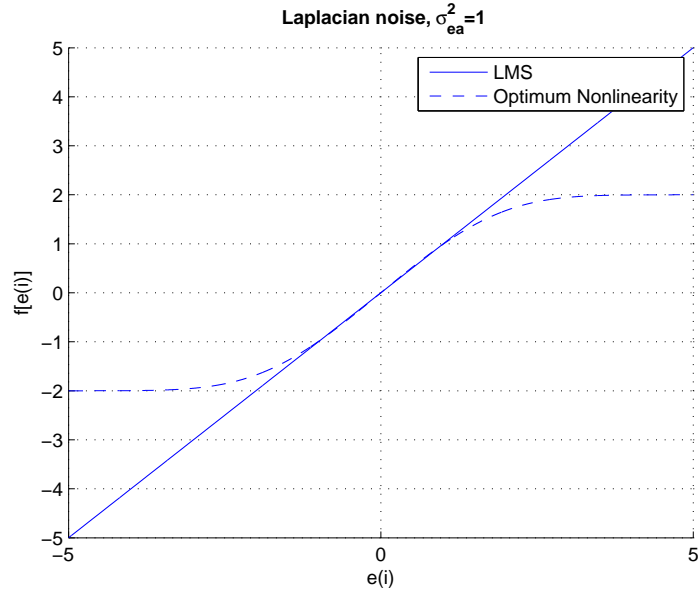


Figure 6.2: Error nonlinearity for the Laplacian case

We can show (see Appendix B for derivation) that the optimal nonlinearity takes the following form

$$f_{opt}[e(i)] = \frac{2}{\mu \text{Tr}(\mathbf{R})} \left[ e(i) - b \tanh\left(\frac{b e(i)}{\sigma_{ea}^2}\right) \right] \quad (6.34)$$

The general shape of the nonlinearity is shown in Figure 6.3

#### 6.4.4 Gaussian Mixture

In the *Gaussian Mixture* noise case, we have a mixture of two Gaussian noises of variances  $\sigma_{v_1}^2$  &  $\sigma_{v_2}^2$ . The two noises are weighted by coefficients  $m$  and  $1 - m$

$$v = mv_1 + (1 - m)v_2 \quad (6.35)$$

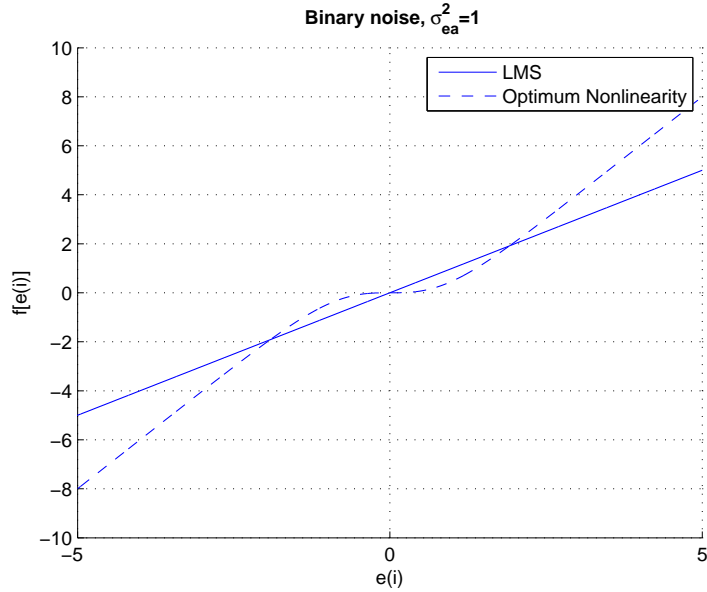


Figure 6.3: Error nonlinearity for the Binary case

This makes the noise Gaussian with variance given by  $(\sigma_v^2 = m^2\sigma_{v_1}^2 + (1-m)^2\sigma_{v_2}^2)$ .

Hence the nonlinearity in this case is similar as that in the Gaussian case and is given by

$$f_{opt}[e(i)] = \frac{2\sigma_{ea}^2}{\mu\text{Tr}(\mathbf{R})(\sigma_{ea}^2 + m^2\sigma_{v_1}^2 + (1-m)^2\sigma_{v_2}^2)}e(i) \quad (6.36)$$

The general shape of the nonlinearity is show in Figure 6.4

## 6.5 Optimum Error Nonlinearity with Conditional Analysis

Is it possible to derive an adaptive filter that combines error nonlinearity with a simple data nonlinearity? Specifically how can we obtain a normalized LMS

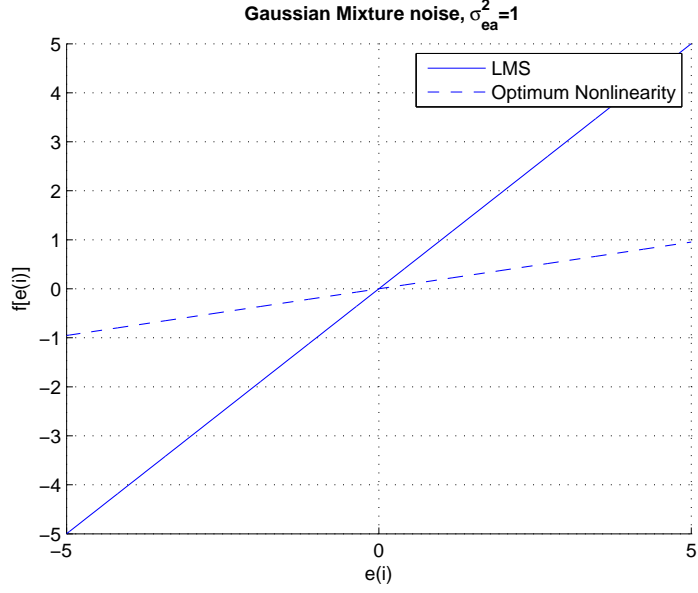


Figure 6.4: Error nonlinearity for the Gaussian mixture case

algorithm with an error nonlinearity. i.e., an algorithm having the update

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu f[e(i)] \frac{\mathbf{u}_i}{\|\mathbf{u}_i\|^2} \quad (6.37)$$

One motivation for this normalization is that it allows us to deal better with correlated input. It turns out that we can do so by repeating the same analysis we performed in the Section 6.3 conditioned on the input. Specifically starting again from the energy relation (5.14) and taking the expectation on both sides conditioned on  $\mathbf{u}_i$  yields

$$\begin{aligned} E [\|\mathbf{v}_{i+1}\|^2 | \mathbf{u}_i] &= E [\|\mathbf{v}_i\|^2 | \mathbf{u}_i] - 2\mu E [e_a(i) f[e(i)] | \mathbf{u}_i] + \mu^2 \|\mathbf{u}_i\|^2 E [f^2[e(i)] | \mathbf{u}_i] \\ &= E [\|\mathbf{v}_i\|^2 | \mathbf{u}_i] - 2\mu E [e_a(i) | \mathbf{u}_i] f[e(i) | \mathbf{u}_i] + \mu^2 \|\mathbf{u}_i\|^2 E [f^2[e(i)] | \mathbf{u}_i] \end{aligned} \quad (6.38)$$

Assuming that the filter is stable, it should eventually reach its steady-state wherein  $E [||\mathbf{v}_{i+1}||^2|\mathbf{u}_i] = E [||\mathbf{v}_i||^2|\mathbf{u}_i]$  as  $i \rightarrow \infty$ . By assuming the stationarity of  $\mathbf{u}_i$ , it can be shown that

$$\lim_{i \rightarrow \infty} E [e_a(i)|\mathbf{u}_i f[e(i)|\mathbf{u}_i]] = \frac{\mu}{2} ||\mathbf{u}_i||^2 \lim_{i \rightarrow \infty} E [f^2[e(i)|\mathbf{u}_i]] \quad (6.39)$$

Next, the expectation  $E [e_a(i)|\mathbf{u}_i f[e(i)|\mathbf{u}_i]]$  in the above equation is evaluated using Price's Theorem and can be setup to the following

$$E [e_a(i)|\mathbf{u}_i f[e(i)|\mathbf{u}_i]] = E [e_a^2(i)|\mathbf{u}_i] h_G [E [e_a^2(i)|\mathbf{u}_i]] \quad (6.40)$$

where

$$h_G [E [e_a^2(i)|\mathbf{u}_i]] = \frac{E [e_a(i)|\mathbf{u}_i f[e(i)|\mathbf{u}_i]]}{E [e_a^2(i)|\mathbf{u}_i]}. \quad (6.41)$$

Knowing that  $h_G [E [e_a^2(i)|\mathbf{u}_i]]$  can take the alternative form of  $E [f'[e(i)|\mathbf{u}_i]]$ , we can write the above equation as follows

$$E [e_a(i)|\mathbf{u}_i f[e(i)|\mathbf{u}_i]] = E [e_a^2(i)|\mathbf{u}_i] E [f'[e(i)|\mathbf{u}_i]] \quad (6.42)$$

Finally, by substituting the value of the above expectation in equation (6.39), the following relation is obtained

$$S = \frac{\mu}{2} ||\mathbf{u}_i||^2 \frac{\lim_{i \rightarrow \infty} E [f^2[e(i)|\mathbf{u}_i]]}{\lim_{i \rightarrow \infty} E [f'[e(i)|\mathbf{u}_i]]}, \quad (6.43)$$

where  $S = \lim_{i \rightarrow \infty} E [e_a^2(i) | \mathbf{u}_i]$ . Therefore, by using the same approach developed for the case optimum error nonlinearity, we get the following optimum error nonlinearity with conditional analysis

$$f_{opt}[e(i) | \mathbf{u}_i] = -\alpha' \frac{p'_e[e(i) | \mathbf{u}_i]}{p_e[e(i) | \mathbf{u}_i]} \quad (6.44)$$

where

$$\alpha' = \frac{2\sigma_{ea}^2}{\mu \|\mathbf{u}_i\|^2}. \quad (6.45)$$

This is the same nonlinearity (6.22) obtained earlier except that  $\text{Tr}(R)$  is now replaced by  $\|\mathbf{u}_i\|^2$ .

## 6.6 The Conditional Error Nonlinearities for Special Noise Cases

Given the similarity between the optimum nonlinearity and the conditional nonlinearity, we can easily deduce the conditional nonlinearity for the four noise cases studied earlier. Specifically, the optimum conditional nonlinearity for the Gaussian noise is given by

$$f_{opt}[e(i)] = \frac{2\sigma_{ea}^2}{\mu \|\mathbf{u}_i\|^2 (\sigma_{ea}^2 + \sigma_v^2)} e(i) \quad (6.46)$$

For the Laplacian Noise case, we have

$$f_{opt}[e(i)] = -\frac{2\sigma_{ea}^2}{\mu\|\mathbf{u}_i\|^2} \left\{ \frac{e^{e(i)/b}\gamma_1[e(i)] - e^{-e(i)}g_-[e(i)]}{e^{e(i)/b}\left(1 - \operatorname{erf}\left[\frac{b e(i) + \sigma_{ea}^2}{b\sqrt{2\sigma_{ea}^2}}\right]\right) + e^{-e(i)/b}\left(1 + \operatorname{erf}\left[\frac{b e(i) - \sigma_{ea}^2}{b\sqrt{2\sigma_{ea}^2}}\right]\right)} \right\} \quad (6.47)$$

where  $\gamma_1[e(i)]$  and  $g_-[e(i)]$  are defined in equation (6.31) and (6.32). For Binary Noise case, the optimum nonlinearity in the conditional analysis case comes out to be

$$f_{opt}[e(i)] = \frac{2\sigma_{ea}^2}{\mu\|\mathbf{u}_i\|^2(\sigma_{ea}^2 + \sigma_v^2)} \left( e(i) - e^{-\left(\frac{e^2(i)+1}{2\sigma_{ea}^2}\right)} \tanh\left[\frac{e(i)}{\sigma_{ea}^2}\right] \right) \quad (6.48)$$

and the nonlinearity for the Gaussian mixture case is given by

$$f_{opt}[e(i)] = \frac{2\sigma_{ea}^2}{\mu\|\mathbf{u}_i\|^2(\sigma_{ea}^2 + m^2\sigma_{v1}^2 + (1-m)^2\sigma_{v2}^2)} e(i) \quad (6.49)$$

## 6.7 Role of the Variance of $e_a(i)$

By investigating the expressions of the optimum error nonlinearities derived above, we see that they all depend on the variance of  $e_a(i)$ . Thus the choice of the variance of  $e_a(i)$  plays a vital role in the performance of the adaptive filter with optimum nonlinearity in any noise environment. Consider the case of Gaussian noise, for example, if we substitute the value of  $f_{opt}[e(i)]$  in equation (5.2), the weight update



rule of the LMS algorithm is modified to the following

$$\mathbf{w}_{i+1} = \mathbf{w}_i + \frac{2\sigma_{ea}^2}{\text{Tr}(\mathbf{R})(\sigma_{ea}^2 + \sigma_v^2)} e(i) \mathbf{u}_i^T \quad (6.50)$$

It can be observed that the weight update rule for the optimum nonlinearity is independent of the step-size value  $\mu$ . Moreover since, the values of  $\text{Tr}(\mathbf{R})$  and  $\sigma_v^2$  are fixed for a given system, the variance of  $e_a(i)$  ( $\sigma_{ea}^2$ ) is the only parameter which can control the step-size of the adaptive filter. Therefore, the optimum nonlinearity for Gaussian noise can be considered as the LMS algorithm with a new and actually variable step-size  $\mu' = \frac{2\sigma_{ea}^2}{\text{Tr}(\mathbf{R})(\sigma_{ea}^2 + \sigma_v^2)}$ . In the case of laplacian and binary noise, we have totally different nonlinearities as compared to the LMS algorithm. There  $\sigma_{ea}^2$  plays the same role of controlling the step-size of the adaptive rule as well as controlling the general shape of the nonlinearity.

## 6.8 Calculating the Variance of $e_a(i)$

It has been established in the previous section that the value of  $e_a(i)$  plays an important role in the performance of optimum nonlinearity. Now we will consider how to estimate this variance. Note that  $e_a(i)$  is time variant and therefore must be estimated online. The problem is that  $e_a(i)$  is not accessible, so we calculate  $\sigma_{ea}^2$  from  $e(i)$  using the fact that

$$\sigma_e^2 = \sigma_{ea}^2 + \sigma_v^2 \quad (6.51)$$

In practice, we estimate  $\sigma_{ea}^2$  by estimating the variance of  $e(i)$  over a window of samples of  $e(i)$  and subtracting the noise variance i.e.,

$$\widehat{\sigma}_{ea}^2 = \widehat{\sigma}_e^2 - \sigma_v^2 \quad (6.52)$$

To assure the stability of the algorithm, we enforce that the value of  $\widehat{\sigma}_{ea}^2$  stays within an interval  $[a, b]$  where the limits  $a$  and  $b$  are experimentally derived.

At the early stages of adaptation however, there are not enough samples of  $e(i)$  and so we can not estimate  $\widehat{\sigma}_e^2$  reliably. In this case, we simply employ the LMS update equation. When enough samples of  $e(i)$  are available, we can estimate  $\sigma_e^2(i)$  reliably so we switch to the optimal non linearity. So, the actual nonlinearity implemented is

$$f[e(i)] = \begin{cases} \mu e(i) & \text{for early stages} \\ f_{opt}[e(i)] & \text{when enough samples of } e(i) \text{ are available} \end{cases} \quad (6.53)$$

There is another justification for this switching approach. The derivation of the optimum nonlinearity suggests that the optimum rule should work well when the steady-state has been reached. Therefore, it is best to implement the nonlinearity in two modes. In the transient mode, the LMS algorithm is implemented. In the second (steady-state) mode, the algorithm switches to the optimum nonlinearity. The switching is activated on the basis of the energy of the estimation error  $e(i)$ . If this energy is less than a certain threshold  $\gamma$ , the algorithm switches from mode

1 to mode 2, that is

$$f[e(i)] = \begin{cases} e(i) & \text{if } |e(i)|^2 \geq \gamma \\ f_{opt}[e(i)] & \text{if } |e(i)|^2 < \gamma \end{cases}$$

In the 2nd mode, i.e. when  $|e(i)|^2 \geq \gamma$ ,  $\sigma_{ea}^2$  is either set to a fixed value or calculated using the windowing method (i.e. by averaging  $e^2(i)$  over a window, estimating  $\sigma_e^2$  and using that to estimate  $\sigma_{ea}^2$ ). The advantage of doing so is that in the steady state,  $e_a(i)$  is much more stable and hence calculating  $\sigma_{ea}^2(i)$  should be easy.

## 6.9 Simulation Results

In this section, some simulations are carried out to validate the theoretical findings in system identification scenario. The unknown system to be identified is an FIR system with impulse response  $[0.0351, -0.0688, 0.1205, -0.258, 0.9054, -0.2561, 0.1018, -0.0731, 0.0673, -0.0673]^T$ . The performance measure is the mean square deviation (MSD). We investigate our results for four different noise environments, Gaussian, Laplacian, Binary and Gaussian mixture. The signal to noise ratio used is always set to 10 dB. The results are averaged over 100 runs.

### 6.9.1 The Effect of Varying $\sigma_{ea}^2$

Here we investigate the effect of varying  $\sigma_{ea}^2$  on the MSD of the optimum non linearity algorithm as discussed in Section 6.7. The convergence rate of our al-

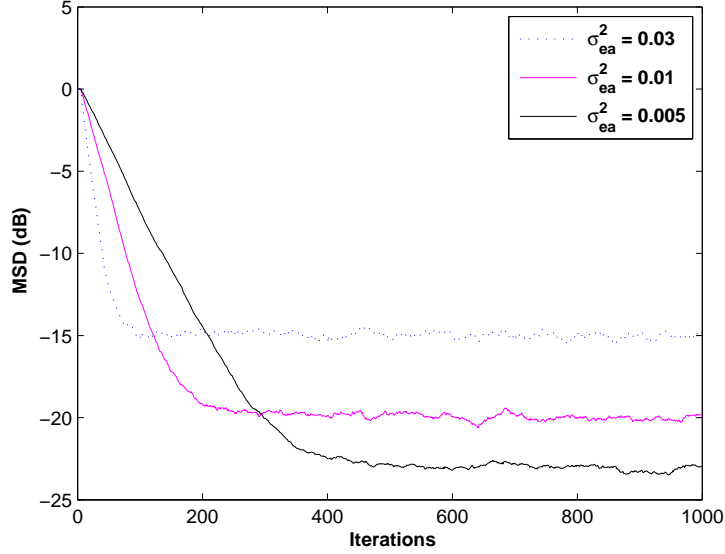


Figure 6.5: MSD of Optimum filter at various values of  $\sigma_{ea}^2$  in Gaussian Noise algorithms is not dependent on the step size  $\mu$ , but rather on the value of  $\sigma_{ea}^2$

Here, three different values of  $\sigma_{ea}^2$  (0.03, 0.01, 0.005) are used. Figure 6.5 shows the MSD curves for various values of  $\sigma_{ea}^2$  in Gaussian Noise. It can be seen from the figure that the convergence speed of the optimum nonlinearity decreases with decreasing  $\sigma_{ea}^2$ . However the steady state value of the MSD decreases with decreasing  $\sigma_{ea}^2$ . Figure 6.7 plot the same result for the Laplacian noise case and shows the same trend. Figure 6.6 shows the same plot for the binary noise case which is plotted for  $\sigma_{ea}^2 = 2.5, 1.3, 1$ .

From above, we can see that the variance  $\sigma_{ea}^2$  plays a role similar to the step size in an adaptive filter. It can be used to decrease the steady-state error but that results in slower convergence speed (and vice versa). It will be thus worthwhile to study the effect of changing the variance of  $e_a(i)$  on the convergence speed and

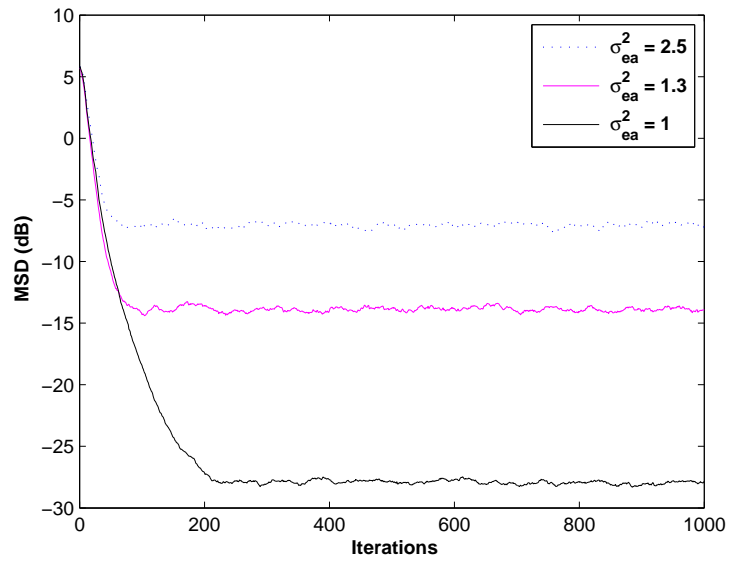


Figure 6.6: MSD Optimum filter at various values of  $\sigma_{ea}^2$  in Binary Noise

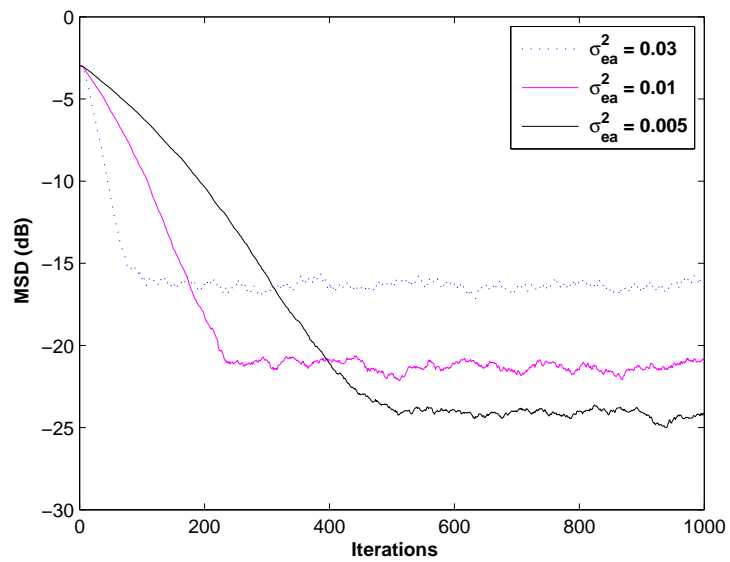


Figure 6.7: MSD Optimum filter at various values of  $\sigma_{ea}^2$  in Laplacian Noise

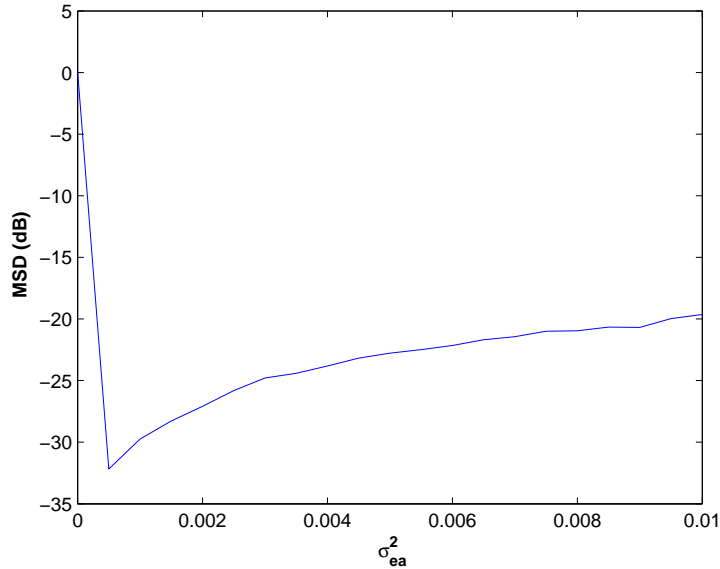


Figure 6.8: Effect of value of  $\sigma_{ea}^2$  on MSD (Gaussian noise).

the steady state MSD achieved. Figure 6.8 plots the steady state MSD versus  $\sigma_{ea}^2$  and Figure 6.9 plots the number of iterations vs  $\sigma_{ea}^2$  required to reach steady state (convergence). In both cases, we vary the variance  $\sigma_{ea}^2$  in steps of 0.0005. Figures 6.10 and 6.11 compare the same for the Binary Noise case. All these figures show essentially the same trend: decreasing  $\sigma_{ea}^2$  speed will result in a lower steady-state MSD, slower convergence speed & vice versa. It is up to the designer to decide what matters to him.

### 6.9.2 Switched Mode Case

The derivation of the optimum nonlinearity suggests that it will perform better if invoked at the steady state. As we explained in Section 6.8, in the switched mode case, the optimum nonlinearity works in two modes. In mode 1, it follows that same update rule as LMS, while in mode 2 it switches to the optimum

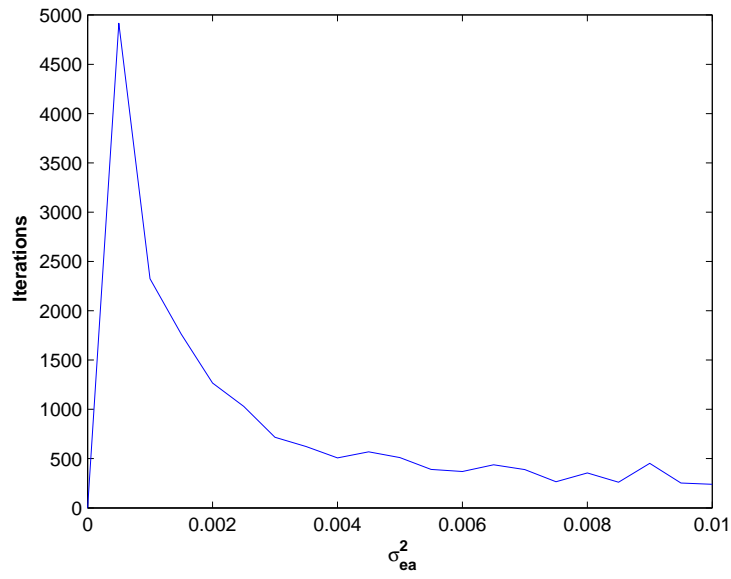


Figure 6.9: Effect of value of  $\sigma_{ea}^2$  on convergence speed (Gaussian noise).

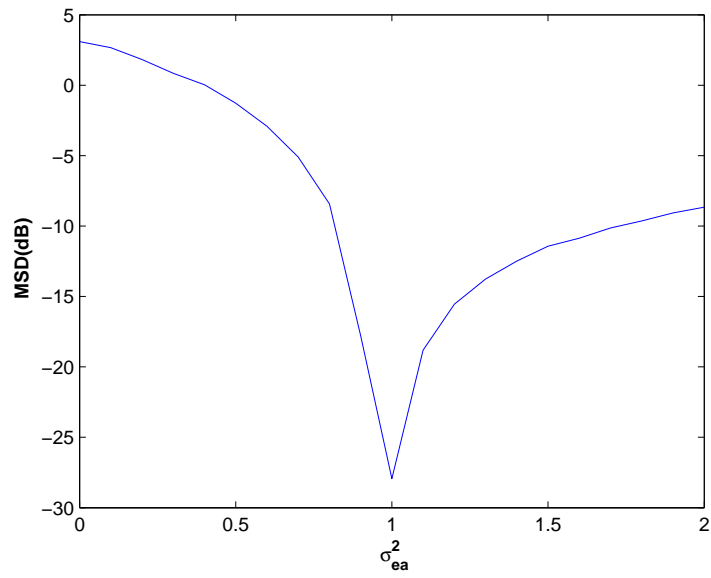


Figure 6.10: Effect of value of  $\sigma_{ea}^2$  on MSD (Binary noise).

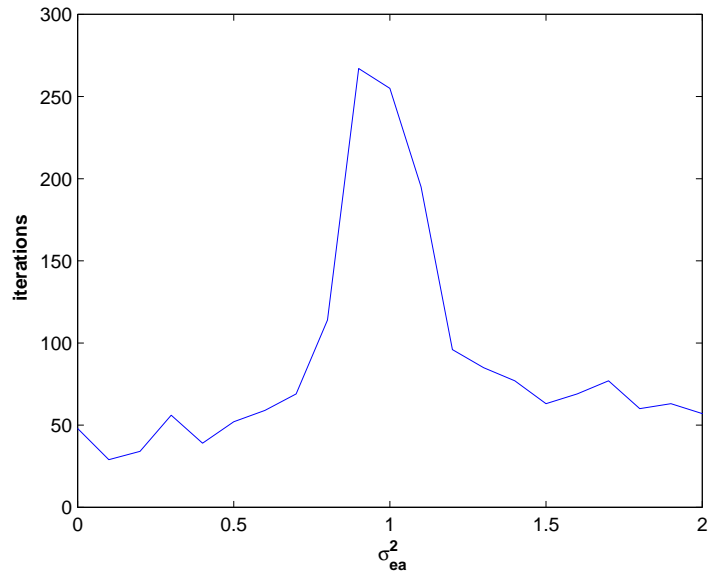


Figure 6.11: Effect of value of  $\sigma_{ea}^2$  on Comparison of convergence speed (Binary noise).

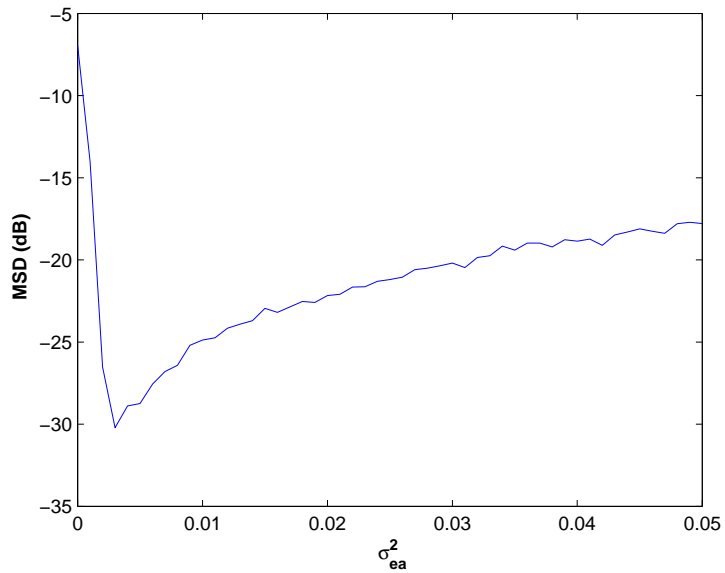


Figure 6.12: Effect of value of  $\sigma_{ea}^2$  on MSD (Laplacian noise).



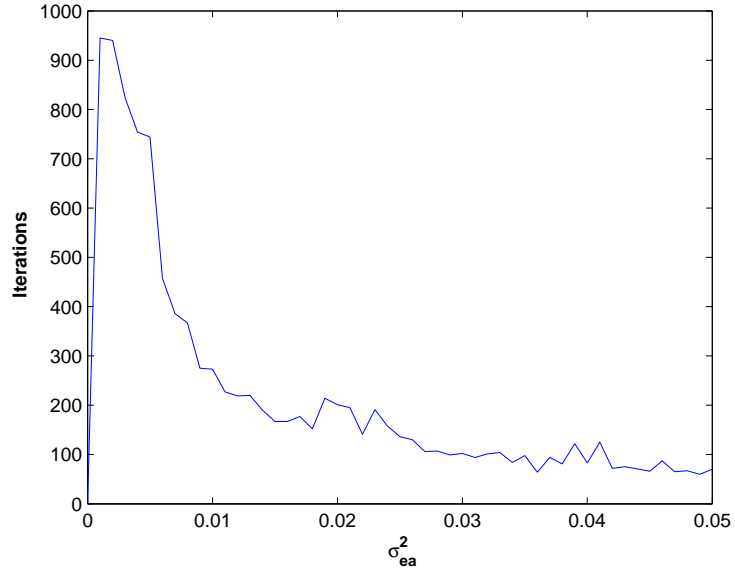


Figure 6.13: Effect of value of  $\sigma_{ea}^2$  on convergence speed (Laplacian noise).

update equation in which the optimum nonlinearity uses a fixed value of  $\sigma_{ea}^2(i)$ . The algorithm switches from mode 1 to mode 2 when the error energy reaches a predefined threshold. We run these simulations by fixing the variance  $\sigma_{ea}^2$ . Simulation are carried out for Gaussian noise, Laplacian noise, Binary noise and Gaussian mixture noise. (see Figures 6.14, 6.15, 6.16 and 6.17). The Figures compare the learning curves of the LMS, conditional nonlinearity and the optimum nonlinearity in the four noise environments. The value of  $\sigma_{ea}^2$  for the Gaussian and Laplacian noise environments is set to 0.001 and is set to 1 for the Binary noise case. It can be depicted form the figures that there is a great improvement in the steady-state MSD for the conditional and optimum nonlinearity as compared to the LMS (Note also that all three algorithms maintain the same convergence speed).

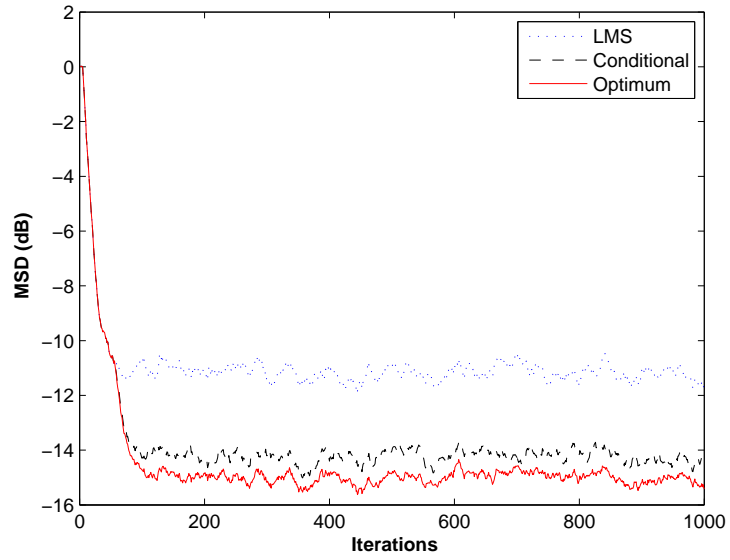


Figure 6.14: MSD learning curves for Gaussian noise (switched mode case)

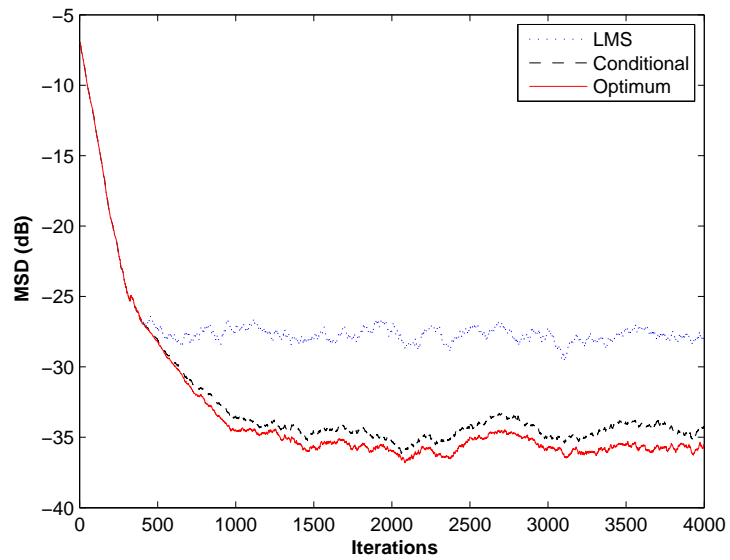


Figure 6.15: MSD learning curves for Laplacian noise (switched mode case)

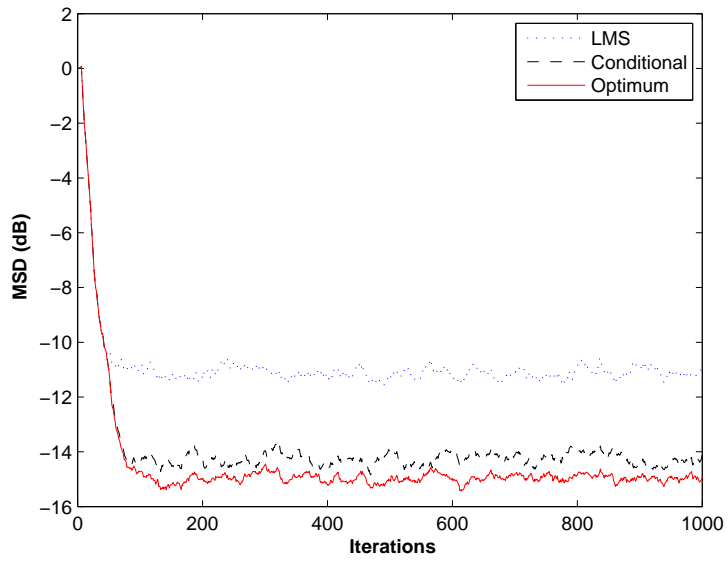


Figure 6.16: MSD learning curves for Binary noise (switched mode case)

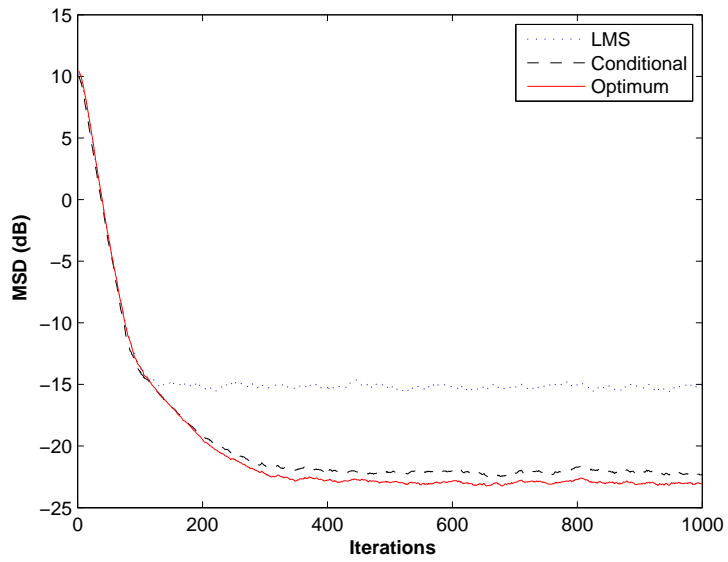


Figure 6.17: MSD learning curves for Gaussian mixture noise with  $m=0.3$ ,  $\sigma_{v_1}^2=100$  and  $\sigma_{v_2}^2=1$  (switched mode case)

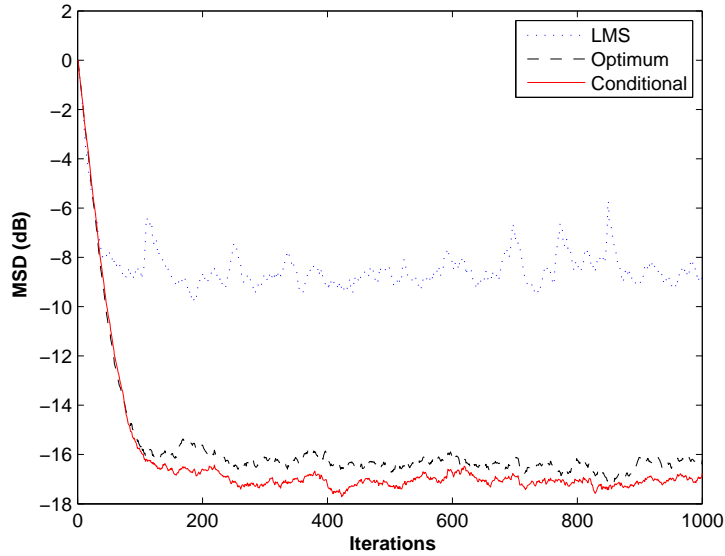


Figure 6.18: MSD learning curves for Gaussian noise windowing method

### 6.9.3 Using Averaged Value of $\sigma_{ea}^2$ Over a Window Size

In the previous simulation, we fixed the value of  $\sigma_{ea}^2$  throughout the simulation. Here, as opposed to fixing  $\sigma_{ea}^2$ , we run the adaptive filters here by estimating  $\sigma_{ea}^2$  and averaging it over a window size of 10 (as explained in Section 6.8). Simulation experiments are carried out for all the four noise environments. In the first 10 iterations, where there are not enough samples of  $e(i)$  to estimate  $\sigma_e^2(i)$ , we use the LMS update rule (as explained in Section 6.8). Figures 6.18 - 6.21 compare the performance of the LMS and the adaptive filter with optimum and conditional nonlinearities. It can be seen from these figures that there is substantial improvement in the steady-state MSD of the optimum nonlinearity as compared to that of the LMS algorithm.

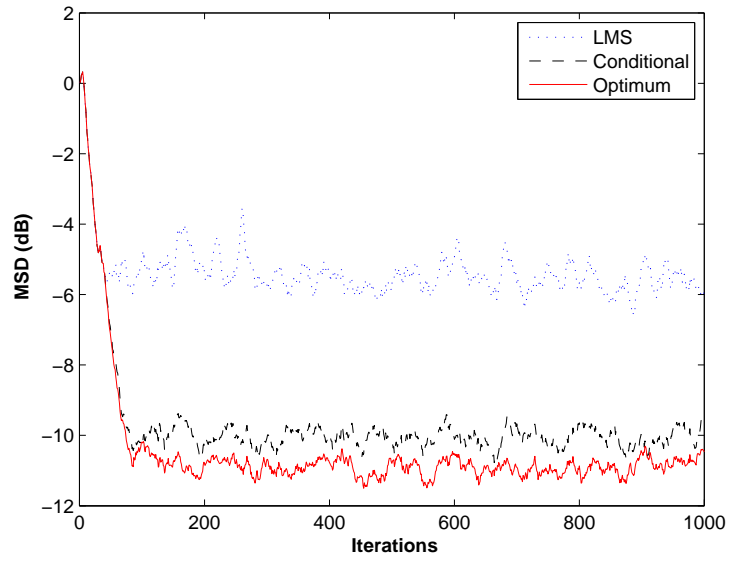


Figure 6.19: MSD learning curves for Laplacian windowing method

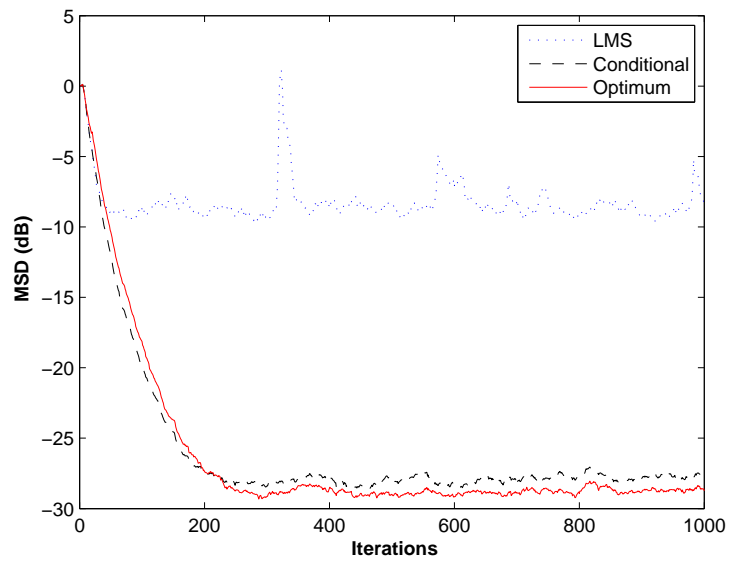


Figure 6.20: MSD learning curves for Binary noise windowing method

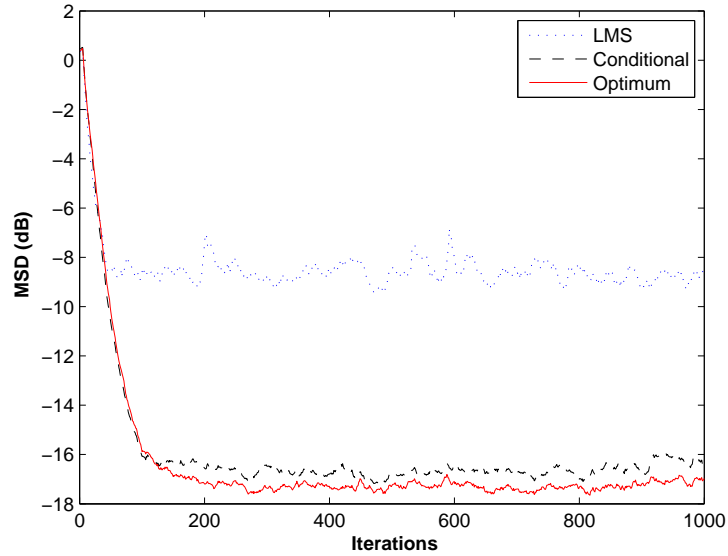


Figure 6.21: MSD learning curves for Gaussian mixture noise windowing method

### 6.9.4 The Effect of Window Size

How does the window size used in the window size used in the estimation of  $\sigma_{ea}^2$  affect the performance of the adaptive filter? In subsection 6.9.3, we used a window size of 10. Here we demonstrate how the window size is varied. We demonstrate that for binary noise (6.22) and Laplacian noise (6.23) when the window size is varied over the range 1 – 40. We see that there is a marginal improvement in the MSD with increasing window size.

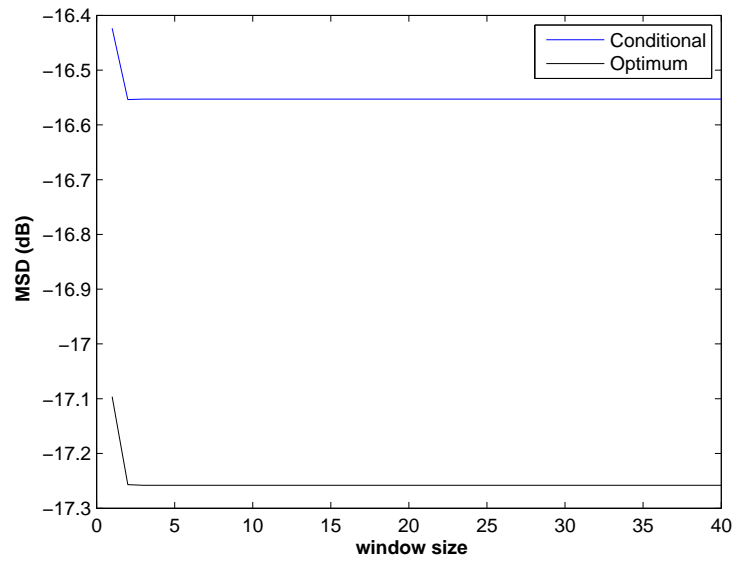


Figure 6.22: Steady state value of MSD vs window size (Binary Noise)

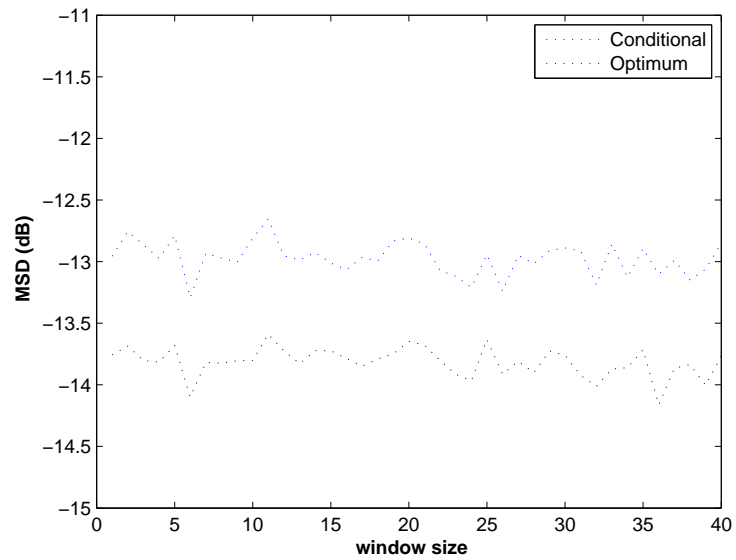


Figure 6.23: Steady state value of MSD vs window size (Laplacian noise)

## 6.9.5 Time Variant Channels

Adaptive algorithms are most valuable when the environment is time variant. In the ensuing we compare the performance of the LMS to the two adaptive algorithm derived in this thesis for time variant channels.

Here we will compare LMS with the optimum error nonlinearity and the optimum error nonlinearity with conditional analysis algorithms. We will study the behavior of these algorithms for three channel models [77]: 1) random walk 2) auto-regressive (AR) model and 3) probabilistic AR model. We assume the channel to be of length 10. The input vector is assumed to be a white Gaussian sequence, as normally considered in literature. We use the windowing technique to estimate the value of  $\sigma_{ea}^2$  with a window length of 5. The LMS is set for fast convergence time and uses a step size of 0.1. The signal to noise ratio is set to 10 dB and the results are averages over 100 runs. Next we present the channel models and their respective results.

1. Random Walk: The random walk model for channel coefficient is given by the following equation

$$c_{k+1} = c_k + \delta_k \tag{6.54}$$

where  $\delta_k$  is a white Gaussian sequence whose elements are uncorrelated with a mean of zero and variance  $\sigma_\delta^2 = 10^{-5}$ . Figures 6.24 and 6.25 compare the three algorithm (optimal, conditional and LMS) in the the random walk channel model for Gaussian and Binary cases. We see that the optimum and conditional nonlinearities perform much better than the LMS.



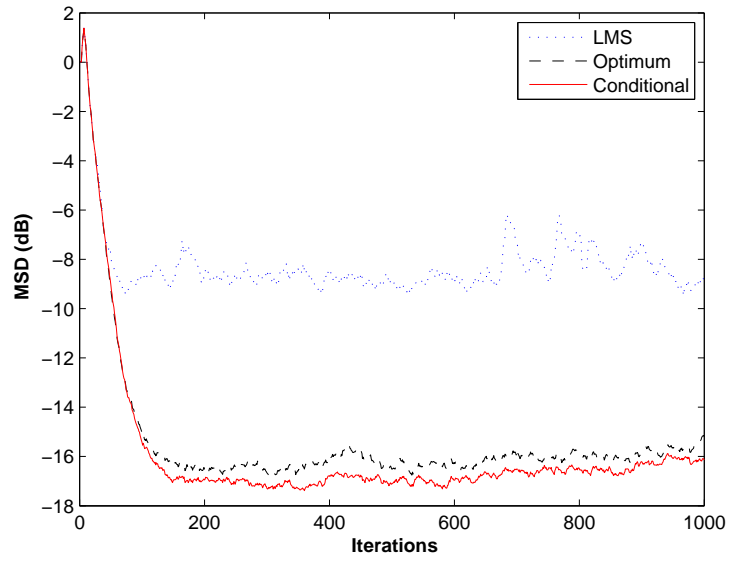


Figure 6.24: MSD for RW time varying channel in Gaussian noise

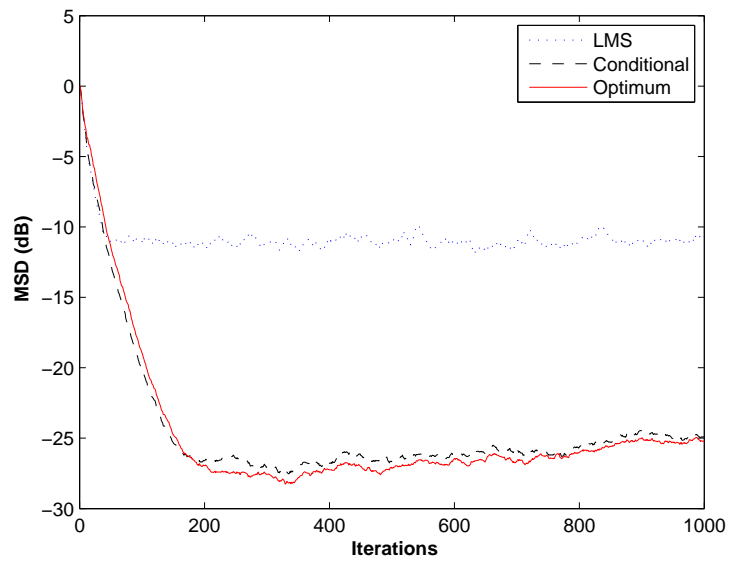


Figure 6.25: MSD for RW time varying channel in Binary noise

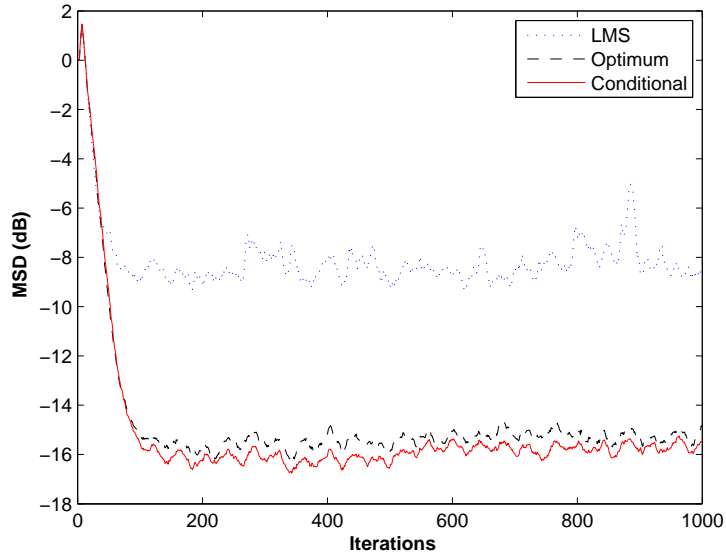


Figure 6.26: MSD for AR time varying channel in Gaussian noise

2. AR Model: The AR model for the variation of channel coefficient is given as

$$c_{k+1} = ac_k + \delta_k \quad (6.55)$$

where  $0 < a < 1$  (we use  $a = 0.90$ ) and  $\delta$  is the same as in the random walk model. Figures (6.26) and (6.27) compare the MSD learning curves of the LMS, the optimum error non linearity and the optimum error with conditional analysis for Gaussian and Binary noise cases.

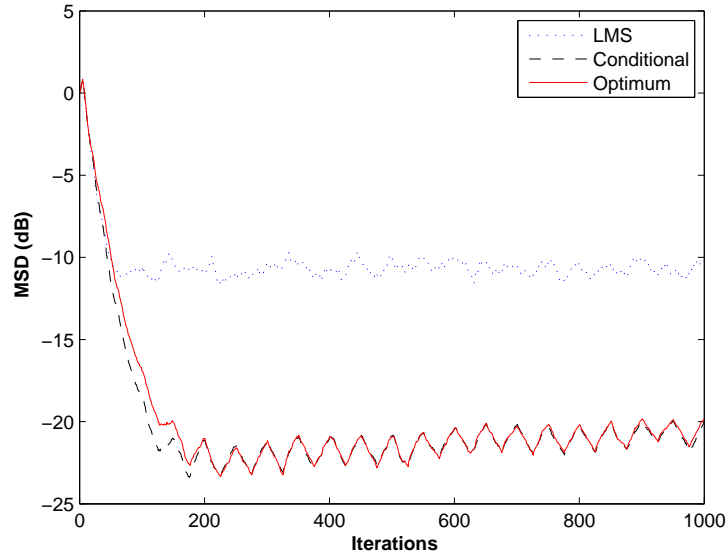


Figure 6.27: MSD for AR time varying channel in Binary noise

3. Probabilistic AR Model: In the probabilistic AR model, the channel coefficients still vary according to (6.55) but with a probability  $p$ . In our simulations, we vary the channel taps with probability  $p = 0.05$ . The results of the probabilistic channel for Gaussian and Binary noise cases are shown in figures (6.28) and (6.29).

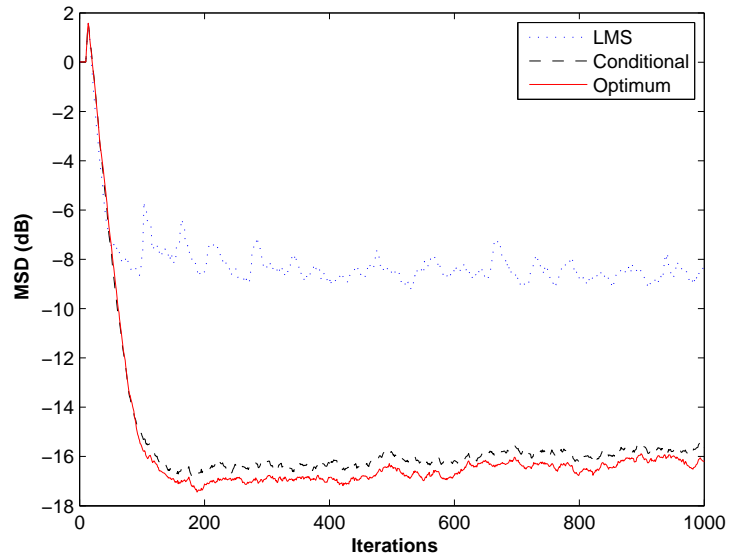


Figure 6.28: MSD for probabilistic AR time varying channel in Gaussian noise

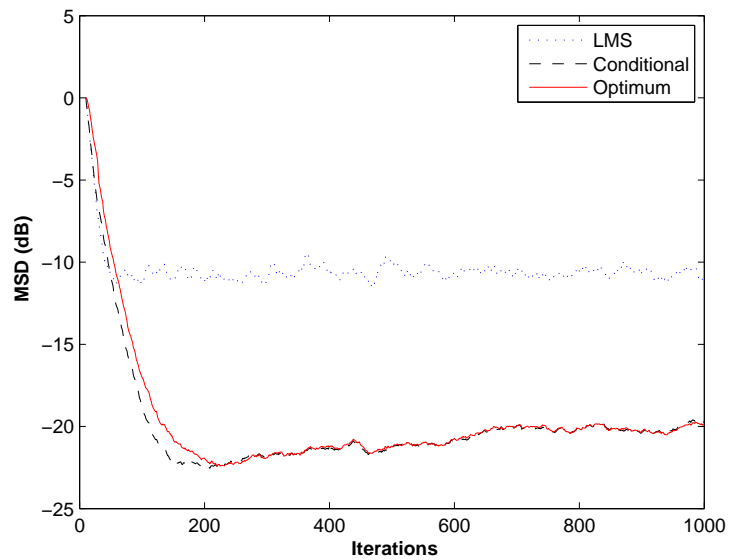


Figure 6.29: MSD for probabilistic AR time varying channel in Binary noise

## 6.10 Conclusion

In this second part of the thesis we have designed steady-state optimum error nonlinearity for adaptive filters. We used the energy relation to derive the steady state behavior of adaptive filters with general error nonlinearities. The nonlinearity was derived under weaker conditions as compared to what is normally employed in literature hence it is more general. The nonlinearity turns out to be a function of the pdf of additive noise and variance of estimation error. A conditional optimum error nonlinearity is also derived along the same lines. A major issue was the evaluation of variance of estimation error which is time variant. Extensive simulations were carried out which have demonstrated that optimum nonlinearity outperforms the LMS. This was also demonstrated in time variant case. The proper online estimation of the variance of estimation error, which is needed in the implementation of the optimum nonlinearity, is still in need of further investigation.

# APPENDIX A: Error Nonlinearity for Laplacian

## Noise

The solution of equation 6.29 is given as

$$\begin{aligned} p_e[e(i)] &= \int_{-\infty}^{\infty} p_{ea}[e(i) - t]p_v[t]dt \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_{ea}^2}} \exp\left[-\frac{(e(i) - t)^2}{2\sigma_{ea}^2}\right] \frac{1}{2b} \exp\left[\frac{-|t|}{b}\right] dt \\ &= \frac{1}{\sqrt{2\pi\sigma_{ea}^2}} \frac{1}{2b} \left( \int_{-\infty}^0 \exp\left[-\frac{(e(i) - t)^2}{2\sigma_{ea}^2}\right] \exp\left[\frac{-t}{b}\right] dt + \right. \\ &\quad \left. \int_0^{\infty} \exp\left[-\frac{(e(i) - t)^2}{2\sigma_{ea}^2}\right] \exp\left[\frac{t}{b}\right] dt \right) \end{aligned}$$

Now performing the substitution

$$\begin{aligned} \frac{e(i) - t}{\sqrt{2\sigma_{ea}^2}} = u &\Rightarrow e(i) - \sqrt{2\sigma_{ea}^2} u = t \\ \frac{-dt}{\sqrt{2\sigma_{ea}^2}} = du &\Rightarrow dt = -\sqrt{2\sigma_{ea}^2} du \end{aligned}$$

So the above equation becomes

$$\begin{aligned}
p_e[e(i)] &= -\frac{1}{\sqrt{2\pi\sigma_{ea}^2}} \frac{1}{2b} \left( \int_{-\infty}^{\frac{e(i)}{\sqrt{2\sigma_{ea}^2}}} \exp[-u^2] \exp\left[\frac{-e(i) + \sqrt{2\sigma_{ea}^2}u}{b}\right] \sqrt{2\sigma_{ea}^2} du \right. \\
&\quad \left. + \int_{\frac{e(i)}{\sqrt{2\sigma_{ea}^2}}}^{\infty} \exp[-u^2] \exp\left[\frac{e(i) - \sqrt{2\sigma_{ea}^2}u}{b}\right] \sqrt{2\sigma_{ea}^2} du \right) \\
&= -\frac{1}{\sqrt{\pi}} \frac{1}{2b} \left( \int_{-\infty}^{\frac{e(i)}{\sqrt{2\sigma_{ea}^2}}} \exp[-u^2] \exp\left[\frac{-e(i)}{b}\right] \exp\left[\frac{\sqrt{2\sigma_{ea}^2}u}{b}\right] du \right. \\
&\quad \left. + \int_{\frac{e(i)}{\sqrt{2\sigma_{ea}^2}}^{\infty} \exp[-u^2] \exp\left[\frac{e(i)}{b}\right] \exp\left[\frac{-\sqrt{2\sigma_{ea}^2}u}{b}\right] du \right) \\
&= -\frac{1}{\sqrt{\pi}} \frac{1}{2b} \cdot \\
&\quad \left( e^{-\frac{e(i)}{b}} \int_{-\infty}^{\frac{e(i)}{\sqrt{2\sigma_{ea}^2}}} \exp\left[-(u^2 - 2(u) \left(\frac{\sqrt{2\sigma_{ea}^2}}{2b}\right) + \left(\frac{\sqrt{2\sigma_{ea}^2}}{2b}\right)^2 - \left(\frac{\sqrt{2\sigma_{ea}^2}}{2b}\right)^2)\right] du \right. \\
&\quad \left. + e^{\frac{e(i)}{b}} \int_{\frac{e(i)}{\sqrt{2\sigma_{ea}^2}}^{\infty} \exp\left[-(u^2 + 2(u) \left(\frac{\sqrt{2\sigma_{ea}^2}}{2b}\right) + \left(\frac{\sqrt{2\sigma_{ea}^2}}{2b}\right)^2 - \left(\frac{\sqrt{2\sigma_{ea}^2}}{2b}\right)^2)\right] du \right) \\
&= -\frac{1}{\sqrt{\pi}} \frac{1}{2b} \cdot \\
&\quad \left( e^{-\frac{e(i)}{b}} \int_{-\infty}^{\frac{e(i)}{\sqrt{2\sigma_{ea}^2}}} \exp\left[-\left(u - \frac{\sqrt{2\sigma_{ea}^2}}{2b}\right)^2\right] e^{\left(\frac{\sqrt{2\sigma_{ea}^2}}{2b}\right)^2} du \right. \\
&\quad \left. + e^{\frac{e(i)}{b}} \int_{\frac{e(i)}{\sqrt{2\sigma_{ea}^2}}^{\infty} \exp\left[-\left(u + \frac{\sqrt{2\sigma_{ea}^2}}{2b}\right)^2\right] e^{\left(\frac{\sqrt{2\sigma_{ea}^2}}{2b}\right)^2} du \right) \tag{A-1}
\end{aligned}$$

Performing another substitution

$$u - \frac{\sqrt{2\sigma_{ea}^2}}{2b} = w \quad \text{and} \quad u + \frac{\sqrt{2\sigma_{ea}^2}}{2b} = v$$

in (A-1), we get

$$p_e[e(i)] = -\frac{1}{2b\sqrt{\pi}} \exp\left[\frac{\sigma_{ea}^2}{2b^2}\right] \left( e^{-\frac{e(i)}{b}} \int_{-\infty}^y e^{-w^2} dw + e^{\frac{e(i)}{b}} \int_z^{\infty} e^{-v^2} dv \right) \quad (\text{A-2})$$

where

$$\begin{aligned} \frac{e(i)}{\sqrt{2\sigma_{ea}^2}} - \frac{\sqrt{2\sigma_{ea}^2}}{2b} = y &\quad \Rightarrow \quad \frac{b e(i) - \sigma_{ea}^2}{b\sqrt{2\sigma_{ea}^2}} = y \\ \frac{e(i)}{\sqrt{2\sigma_{ea}^2}} + \frac{\sqrt{2\sigma_{ea}^2}}{2b} = z &\quad \Rightarrow \quad \frac{b e(i) + \sigma_{ea}^2}{b\sqrt{2\sigma_{ea}^2}} = z \\ \text{also } \operatorname{erf}(x) &= \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \end{aligned}$$

which simplifies to

$$\begin{aligned} p_e[e(i)] &= -\frac{1}{2b\sqrt{\pi}} \exp\left[\frac{\sigma_{ea}^2}{2b^2}\right] \frac{\sqrt{\pi}}{2} \left( e^{-\frac{e(i)}{b}} (1 + \operatorname{erf}(y)) + e^{\frac{e(i)}{b}} (1 - \operatorname{erf}(z)) \right) \\ &= -\frac{1}{4b} e^{\sigma_{ea}^2/2b^2} \left\{ e^{\frac{e(i)}{b}} \left( 1 - \operatorname{erf}\left[\frac{b e(i) + \sigma_{ea}^2}{b\sqrt{2\sigma_{ea}^2}}\right] \right) + e^{-\frac{e(i)}{b}} \left( 1 + \operatorname{erf}\left[\frac{b e(i) - \sigma_{ea}^2}{b\sqrt{2\sigma_{ea}^2}}\right] \right) \right\} \quad (\text{A-3}) \end{aligned}$$

and

$$\begin{aligned} p'_e[e(i)] &= -\frac{1}{4} e^{\sigma_{ea}^2/2b^2} \left\{ e^{e(i)/b} \left( \frac{1}{b} - \frac{1}{b} \operatorname{erf}\left[\frac{b e(i) + \sigma_{ea}^2}{b\sqrt{2\sigma_{ea}^2}}\right] - \sqrt{\frac{2}{\pi\sigma_{ea}^2}} e^{(b e(i) + \sigma_{ea}^2)^2/2b^2\sigma_{ea}^2} \right) \right. \\ &\quad \left. + e^{-e(i)/b} \left( \frac{1}{b} + \frac{1}{b} \operatorname{erf}\left[\frac{b e(i) - \sigma_{ea}^2}{b\sqrt{2\sigma_{ea}^2}}\right] - \sqrt{\frac{2}{\pi\sigma_{ea}^2}} e^{(b e(i) - \sigma_{ea}^2)^2/2b^2\sigma_{ea}^2} \right) \right\} \quad (\text{A-4}) \end{aligned}$$



Replacing equations (A-3) and (A-4) in (6.22), the optimum nonlinearity for the case of *Laplacian* noise will become

$$f_{opt}[e(i)] = -\frac{2\sigma_{ea}^2}{\mu \text{Tr}(\mathbf{R})} \left\{ \frac{e^{e(i)/b} \gamma_1[e(i)] - e^{-e(i)} \gamma_2[e(i)]}{e^{e(i)/b} \left(1 - \text{erf} \left[ \frac{b e(i) + \sigma_{ea}^2}{b \sqrt{2\sigma_{ea}^2}} \right] \right) + e^{-e(i)/b} \left(1 + \text{erf} \left[ \frac{b e(i) - \sigma_{ea}^2}{b \sqrt{2\sigma_{ea}^2}} \right] \right)} \right\} \quad (\text{A-5})$$

where

$$\gamma_1[e(i)] = \frac{1}{b} - \frac{1}{b} \text{erf} \left[ \frac{b e(i) + \sigma_{ea}^2}{b \sqrt{2\sigma_{ea}^2}} \right] - \sqrt{\frac{2}{\pi \sigma_{ea}^2}} e^{(b e(i) + \sigma_{ea}^2)^2 / 2b^2 \sigma_{ea}^2} \quad (\text{A-6})$$

$$\gamma_2[e(i)] = \frac{1}{b} + \frac{1}{b} \text{erf} \left[ \frac{b e(i) - \sigma_{ea}^2}{b \sqrt{2\sigma_{ea}^2}} \right] - \sqrt{\frac{2}{\pi \sigma_{ea}^2}} e^{(b e(i) - \sigma_{ea}^2)^2 / 2b^2 \sigma_{ea}^2} \quad (\text{A-7})$$

## APPENDIX B: Error Nonlinearity for Binary Noise

In the binary noise case, the noise pdf is given as

$$p_v[e(i)] = \begin{cases} b & \text{with probability 0.5} \\ -b & \text{with probability 0.5} \end{cases} \quad (\text{B-1})$$

substituting it in (6.23), we get

$$\begin{aligned} p_e[e(i)] &= \frac{1}{\sqrt{2\pi\sigma_{ea}^2}} e^{-\frac{e^2(i)}{2\sigma_{ea}^2}} * p_v[e(i)] \\ &= \frac{1}{2} \frac{1}{\sqrt{2\pi\sigma_{ea}^2}} e^{-\frac{(e(i)-b)^2}{2\sigma_{ea}^2}} + \frac{1}{2} \frac{1}{\sqrt{2\pi\sigma_{ea}^2}} e^{-\frac{(e(i)+b)^2}{2\sigma_{ea}^2}} \\ &= \frac{1}{2} \frac{1}{\sqrt{2\pi\sigma_{ea}^2}} e^{-\frac{e(i)^2+b^2}{2\sigma_{ea}^2}} \left[ e^{\frac{b e(i)}{\sigma_{ea}^2}} + e^{-\frac{b e(i)}{\sigma_{ea}^2}} \right] \end{aligned} \quad (\text{B-2})$$

and its differentiation yields

$$\begin{aligned}
p'_e[e(i)] &= -\frac{1}{4\sigma_{ea}^2} \frac{2e(i)}{\sqrt{2\pi\sigma_{ea}^2}} e^{-\frac{e(i)^2+b^2}{2\sigma_{ea}^2}} \left[ e^{\frac{b e(i)}{\sigma_{ea}^2}} + e^{-\frac{b e(i)}{\sigma_{ea}^2}} \right] \\
&+ \frac{1}{2} \frac{1}{\sqrt{2\pi\sigma_{ea}^2}} e^{-\frac{e(i)^2+b^2}{2\sigma_{ea}^2}} \left[ \frac{b}{\sigma_{ea}^2} e^{\frac{b e(i)}{\sigma_{ea}^2}} - \frac{b}{\sigma_{ea}^2} e^{-\frac{b e(i)}{\sigma_{ea}^2}} \right] \quad (\text{B-3})
\end{aligned}$$

Replacing (B-2) and (B-3) in (6.22), we get the optimum error nonlinearity for the binary noise case as

$$f_{opt}[e(i)] = \frac{2}{\mu \text{Tr}(\mathbf{R})} \left[ e(i) - b \tanh\left(\frac{b e(i)}{\sigma_{ea}^2}\right) \right] \quad (\text{B-4})$$

## CHAPTER 7

# CONCLUSION AND FUTURE WORK

### 7.1 Conclusion

In part 1 of the thesis, we investigate two parameter reduction approaches for channel estimation in the frequency domain for channel estimation in multiuser systems, where each user has access to a limited portion of the spectrum. First we investigated a simple linearization and quadratic approach. As the results show, this technique requires dense pilot placement and a large number of interpolation parameters, stressing the system. As such this limits the scope of this approach. Thus motivating us to use a more innovative approach. Next we investigated the eigenvalue based interpolation technique for channel estimation. The channel is initially estimated using pilots. This initial pilot based estimate is improved using a data aided approach. For this purpose we use the EM algorithm. We first

estimate the data and then use it to estimate the channel. By iterating between these two steps, the receiver improves the channel estimate. By incorporating time correlation information, we can further improve the estimate. To this end we employ the Kalman Filter. If we relax the latency constraint, we can further improve the channel estimate by employing Cyclic and Helix based Kalman Filters. When an outer code is used, the code will reduce the number of errors in the received data and this corrected data can in turn be used to further enhance the channel estimate.

In part 2, we have designed the steady state optimum error nonlinearity for long adaptive filters. Starting from an averaged form of energy conservation relation, we have derived the relation of optimum error nonlinearity and optimum nonlinearity with conditional analysis. Closed form expressions for some special cases of intrust have also been derived. A windowed approach is suggested to make a real time estimate of the variance of  $e$ . This provides a practical implementation option and also shows an improvement in the performance of the optimum and conditional case as compared to the LMS.

## 7.2 Future Work

It has been shown in Section 4.4.4 that the performance of the channel estimation algorithm is affected by the pilot pattern used. The pilot design in Section 4.4.4 was 1 dimensional, meaning that we consider pilot placement only in the frequency domain. One direction of further research could be investigating 2 dimensional

optimal pilot design for this technique, considering pilot placement in both the time and the frequency domains.

The Eigenvalue approach here has been developed for the SISO case. We have demonstrated the viability of this technique. Future research on this work can focus on adapting this technique for the MIMO case. The adaptation will follow the same design procedure as in the SISO case and most of the design parameters would also be unchanged.

In part 2, we have designed the optimum error nonlinearity and the conditional error nonlinearity. Future work on this part can derive the optimum Data nonlinearity for long adaptive filters along similar lines. Also the effect of variance of  $\sigma_{ea}^2$  on the window size can be further explored.

# REFERENCES

- [1] L. Hanzo and T. Keller, *OFDM and MC-CDMA*, John Wiley and Sons, 2006.
- [2] W. Zhendao and G. B. Giannakis, "Wireless multicarrier communications," *IEEE Signal Processing Magazine*, vol. 17, no. 3, pp. 2948, May 2000.
- [3] M. Speth, S. A. Fechtel, G. Fock, and H. Meyr, "Optimum receiver design for wireless broad-band systems using OFDM-Part 1," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1668-1677, Nov. 1999.
- [4] European Telecommunications Standards Institute, ETS 300 744: Digital Video Broadcasting (DVB-T).
- [5] IEEE Std 802.11a/D7.0-1999, Part11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: High Speed Physical Layer in the 5GHz Band.
- [6] ETSI, "Broadband Radio Access Networks (BRAN); HIPERLAN type 2 technical specification; Physical (PHY) layer," August 1999. DTS/BRAN-0023003 V0.k.

- [7] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM system," *IEEE Trans. Consumer Electr.*, vol. 44, no. 3, pp. 1122-1128, Aug. 1998.
- [8] Y. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," *Proc. IEEE Vehicular Tech. Conf.*, 1999, vol. 2, pp. 1131-1135.
- [9] J.K. Cavers, "An analysis of pilot symbol assisted modulation for rayleigh fading channels (mobile radio)," *IEEE Trans. Vehicular Tech.*, vol. 40, no. 4, pp. 686-693, Nov. 1991.
- [10] F. Tufvesson and T. Maseng, "Pilot assisted channel estimation for OFDM in mobile cellular systems," *Proc. IEEE Vehicular Tech. Conf.*, May 1997, vol. 3, pp. 1693-1643.
- [11] S. Ohno and G. B. Giannakis, "Optimal training and redundant precoding for block transmissions with application to wireless OFDM," *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, 2001, pp. 2389-2392.
- [12] J. W. Choi and Y. H. Lee, "Optimum Pilot Pattern for Channel Estimation in OFDM Systems", *IEEE Transactions on Wireless Communications*, Vol. 4, No. 5, September 2005.
- [13] Seongwook Song and Andrew C. Singer, "Pilot-Aided OFDM Channel Estimation in the Presence of the Guard Band", *IEEE Transactions on Communications*, Vol. 55, No. 8, August 2007

- [14] C. Tepedelenlioglu and G. B. Giannakis, "Transmitter redundancy for blind estimation and equalization of time- and frequency-selective channels," *IEEE Trans. Signal Processing*, vol. 48, pp. 2029-2043, July 2000.
- [15] Chengyang Li, S. Roy, "Subspace-Based Blind Channel Estimation for OFDM by Exploiting Virtual Carriers," *IEEE Trans. on Wireless Comm.*, vol. 2, pp. 141-150, Jan. 2003.
- [16] B. Yang, K. Ben Letaief, R. Cheng, and Z. Cao, "Channel estimation for OFDM transmission in multipath fading channels based on parametric channel modeling," *IEEE Trans. Commun.*, vol. 49, no. 3, pp. 467-479, Mar. 2001.
- [17] M. C. Necker and G. L. Stüber, "Totally blind channel estimation for OFDM over fast varying mobile channels," *Proc. IEEE Int. Conf. on Communications (ICC)*, New York, Apr. 2002, pp. 421-425.
- [18] C. Shin, R. W. Heath, Jr. and E. J. Powers, "Blind Channel Estimation for MIMO OFDM Systems", *IEEE Transactions On Vehicular Technology*, Vol. 56, No. 2, March 2007
- [19] H. Bölcskei, R. W. Heath, Jr., and A. J. Paulraj, "Blind channel identification and equalization in OFDM-based multi-antenna systems," *IEEE Trans. Signal Processing*, vol. 50, pp. 96-109, Jan. 2002.
- [20] S. Zhou and G. B. Giannakis, "Finite-alphabet based channel estimation for OFDM and related multicarrier systems," *IEEE Trans. Commun.*, vol. 49, pp. 1402-1414, Aug. 2001.



- [21] X. Zhuang, Z. Ding, and A. L. Swindlehurst, "A statistical subspace method for blind channel identification in OFDM communications," *Proc. IEEE ICASSP*, Istanbul, Turkey, 2000.
- [22] Y. Li, L. J. Cimini, and N. R. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels", *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 902-915, Jul. 1998.
- [23] F. Sanzi and M. C. Necker, "Totally Blind APP Channel Estimation for Mobile OFDM Systems", *IEEE Communications Letters*, Vol. 7, No. 11, November 2003
- [24] Xianbin Wang, Wu, Y., Chouinard, J.-Y., Hsiao-Chun Wu, "On the design and performance analysis of multisymbol encapsulated OFDM systems", *IEEE Transactions on Vehicular Technology*, Volume 55, Issue 3, May 2006 Page(s):990 - 1002.
- [25] Cirpan, H.A., Panayirci, E., Dogan, H., "Nondata-aided channel estimation for OFDM systems with space-frequency transmit diversity", *IEEE Transactions on Vehicular Technology*, Volume 55, Issue 2, March 2006 Page(s):449 - 457
- [26] Wang Kunji, Zhang Jianhua, Li Chaojun, Huang Chen, "Semi-Blind OFDM Channel Estimation Using Receiver Diversity in the presence of virtual carriers", *IEEE First International Conference on Communications and Networking in China, ChinaCom '06*. Oct. 2006. Pages(s):1-4.

- [27] Jianhua Liu, Jian Li, “Parameter estimation and error reduction for OFDM-based WLANs ”, *IEEE Transactions on Mobile Computing*, Volume 3, Issue 2, April-June 2004 Page(s):152 - 163
- [28] Simeone, O., Bar-Ness, Y., Spagnolini, U., “Pilot-based channel estimation for OFDM systems by tracking the delay-subspace”, *IEEE Transactions on Wireless Communications*, Volume 3, Issue 1, Jan. 2004 Page(s):315 - 325
- [29] B. Song, L. Gui, and W. Zhang, “Comb Type Pilot Aided Channel Estimation in OFDM Systems with transmit diversity”, *IEEE Transactions on Broadcasting*, Vol. 52, No. 1, March 2006
- [30] Dowler, A., Nix, A., McGeehan, J., “Data-derived iterative channel estimation with channel tracking for a mobile fourth generation wide area OFDM system”, *IEEE GLOBECOM 03, Global Telecommunications Conference, 2003.* , Volume 2, 1-5 Dec. 2003 Page(s):804 - 808 Vol.2
- [31] G. Alrawi, T. Y. Al-Naffouri, A. Bahai, and J. Cioffi, “Exploiting error control coding and cyclic prefix in channel estimation for coded OFDM systems,” *Proc. IEEE GlobeCom*, Nov. 2002, pp. 1152-1156.
- [32] Chae-Hyun Lim, Dong Seog Han, “Robust LS channel estimation with phase rotation for single frequency network in OFDM” , *IEEE Transactions on Consumer Electronics*, Volume 52, Issue 4, Nov. 2006 Page(s):1173 - 1178

- [33] P.Yun Tsai, T. Dar Chiueh, “Frequency-Domain Interpolation-Based Channel Estimation in Pilot Aided OFDM Systems”, *IEEE 59th VTC*, vol 1, pp. 420-424, 2004.
- [34] J. Jiang, Tian Tang, Yongjin Zhang, Ping Zhang, “A channel Estimation Algorithm for OFDM based on PCC training Symbols and Frequency Domian Windowing”, *IEEE ISCIT 2006*, pp. 629-632, 2006.
- [35] I. Kang, M. P. Fitz, and S. B. Gelfand, “Blind estimation of multipath channel parameters: a modal analysis approach,” *IEEE Trans. Commun.*, vol. 47, no. 8, pp. 1140-1150, Aug. 1999.
- [36] Z. Shengli and G. B. Giannakis, ”Finite-alphabet based channel estimation for OFDM and related multicarrier systems,” *IEEE Trans. Commun.*, vol. 49, no. 8, pp. 1402-1414, Aug. 2001.
- [37] Damián Marelli and Minyue Fu, “A subband approach to channel estimation and equalization for DMT and OFDM systems”, *IEEE Transactions on Communications*, Vol. 53, No. 11, November 2005.
- [38] O. Edfors, M. Sandell, J. van de Beek, K. S. Wilson, and P. O. Brjesson, “OFDM channel estimation by singular value decomposition,” *IEEE Trans. Signal Proc.*, vol. 46, no. 7, pp. 931-939, Jul. 1998.
- [39] C. Aldana, E. de Carvalho, and J. M. Cioffi, “Channel estimation for multi-carrier multiple input single output systems using the EM algorithm,” *IEEE Trans. Signal Proc.*, vol. 51, no. 12, pp. 3280-3292, Dec. 2003.

- [40] M.C. Vanderveen, A.-J. Van der Veen, and A. Paulraj, "Estimation of multipath parameters in wireless communications," *IEEE Trans. Signal Proc.*, vol. 46, no. 3, pp. 682-690, Mar. 1998.
- [41] Jong-Ho Lee, Jae Choong Han, Seong-Cheol Kim, "Joint carrier frequency synchronization and channel estimation for OFDM systems via the EM algorithm", *IEEE Transactions on Vehicular Technology*, Volume 55, Issue 1, Jan. 2006 Page(s):167 - 172
- [42] Xiaolin Hou, Zhan Zhang, Kayama, H., "Time-Domain MMSE Channel Estimation Based on Subspace Tracking for OFDM Systems", *Vehicular Technology Conference*, 2006. VTC 2006-Spring. IEEE 63rd Volume 4, 7-10 May 2006 Page(s):1565 - 1569
- [43] T. Y. Al-Naffouri, "An EM-Based Forward-Backward Kalman Filter for the Estimation of Time-Variant Channels in OFDM", *IEEE Trans. Signal Processing*, vol. 55, Jul. 2007.
- [44] Shengping Qin, Peide Liu, Xin Zhang, Laibo Zheng, Deqiang Wang, "Channel estimation with timing offset based on PSD & LS estimation for wireless OFDM systems", *International Symposium on Intelligent Signal Processing and Communication Systems, 2007*. page(s): 248-251 ISPACS 2007.
- [45] R. Dinis, N. Souto, J. Silva, A. Kumar and A. Correia, "Joint Detection and Channel Estimation for OFDM Signals with Implicit Pilots", *IEEE Mobile and Wireless Communications Summit*, Hungary, Vol. 1 pp. 1-5, July 2007.

- [46] Jeremic, A., Thomas, T.A., Nehorai, A., “OFDM channel estimation in the presence of interference”, *IEEE Transactions on Signal Processing*, Volume 52, Issue 12, Dec. 2004 Page(s):3429 - 3439
- [47] G.B. Giannakis, “Filter banks for blind channel identification and equalization,” *IEEE Signal Proc. Lett.*, vol. 4, no. 6, pp. 184-187, Jun. 1997.
- [48] C. Komninakis, C. Fragouli, A. Sayed, and R. Wesel, “Multi-input multi-output fading channel tracking and equalization using Kalman estimation,” *IEEE Trans. Signal Proc.*, vol. 50, no. 5, pp. 1065-1076, May 2002.
- [49] R.A. Iltis, “Joint estimation of PN code delay and multipath using the extended Kalman filter,” *IEEE Trans. Commun.*, vol. 38, no. 10, pp. 1677-1685, Oct. 1990.
- [50] B. Lu, X. Wang, and Y. Li, “Iterative receivers for space-time blockcoded OFDM systems in dispersive fading channels,” *IEEE Trans. Wireless Commun.*, vol. 1, no. 2, pp. 213-225, Apr. 2002.
- [51] B. Muquet, M. de Courville, G. B. Giannakis, Z. Wang, and P. Duhamel, “Reduced complexity equalizers for zero-padded OFDM transmissions,” *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, Istanbul, Turkey, Jun. 2000, pp. 2973-2976
- [52] A. Ghorokhov and J. Linnartz, “Robust OFDM receivers for dispersive time varying channels: equalization and channel acquisition,” *Proc. IEEE Int. Conf. Commun.*, 2002, pp. 470-474.

- [53] Y. Xie and C. N. Georghiades, "Two EM-type channel estimation algorithms for OFDM with transmitter diversity," *IEEE Trans. Commun.*, vol. 51, no. 1, pp. 106-115, Jan. 2003.
- [54] M. Jiang, J. Akhtman, F. Guo, and L. Hanzo. "Iterative joint channel estimation and multi-user detection for high-throughput multiple-antenna OFDM systems". In *Proceedings of the Spring'06 IEEE Vehicular Technology Conference*, Melbourne, Australia, May 2006.
- [55] G. Sütber, *Principles of mobile communication*, Kluwer Academic, 2001.
- [56] T. Kailath, A. H. Sayed and B. Hassibi, *Linear estimation*, Prentice Hall, 2000.
- [57] M. K. Tsatsanis, G. B. Giannakis, and G. Zhou, "Estimation and equalization of fading channels with random coefficients," *Signal Proc.*, vol. 53, no. 2-3, pp. 211-229, Sept. 1996.
- [58] M. D. Estarki and E. Karami, "Joint Blind and Semi-Blind MIMO Channel Tracking and Decoding Using CMA Algorithm", *IEEE Vehicular Technology Conference*, Dublin, vol. 1, pp 2223-2227, April 2007.
- [59] T. Cui and C. Tellambura, "Joint data detection and channel estimation for OFDM systems". *IEEE Transactions on Communications*, 54 (4). pp. 670-679, 2006.
- [60] T. K. Moon, "The expectation-maximization algorithm", *IEEE Signal Proc. Mag.*, 13(6):47-60, Nov. 1996.

- [61] G. Leus and M. Moonen, “Semi-blind channel estimation for block transmissions with non-zero padding”, *Proc. Asilomar Conf. on Signals, Syst. and Computers*, Nov. 2001, pp. 762-766.
- [62] Ali H. Syed, *Fundamentals of Adaptive Filtering*, John Wiley & Sons, Inc. 2003.
- [63] S. Haykin, *Adaptive Filter Theory*, Englewood Cliffs, NJ: Prentice Hall, 1996.
- [64] N. R. Yousef and A. H. Sayed, “A Unified Approach to the Steady-State and Tracking Analysis of Adaptive Filters,” *IEEE Transactions on Signal Processing*, vol. 49, pp. 314–324, February 2001.
- [65] Price, R., “A Useful Theorem for Non-Linear Devices Having Gaussian Inputs,” *IEEE Trans. Inform.* , Th. 4, 69-72, 1958.
- [66] D. L. Duttweiler, “Adaptive filter performance with nonlinearities in the correlation multiplier,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 30, no. 4, pp. 578–586, Aug 1982.
- [67] E. Walach and B. Widrow, “The Least Mean Fourth(LMF) adaptive algorithm and its family,” *IEEE transactions, Information Theory*, vol. 30, no. 2, pp. 275–283, 1984.
- [68] N.J. Bershad, “Analysis of the normalized LMS algorithm with gaussian inputs.,” *IEEE transactions on Acoustic, Speech and Signal Processing*, vol. ASSP-34, pp. 793–806, 1986.

- [69] W. A. Sethares, “Adaptive algorithms with nonlinear data and error functions,” *IEEE transactions on Signal Processing*, vol. 40, no. 9, pp. 2199–2206, 1992.
- [70] V. Mathews and S. Cho, “Improved convergence analysis of stochastic gradient adaptive filters using the sign algorithm,” *IEEE transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 4, pp. 450–454, Apr 1987.
- [71] N. J. Bershad, “On error-saturation nonlinearities in LMS adaptation,” *IEEE transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 4, pp. 440–452, 1988.
- [72] N. J. Bershad and M. Bonnet, “Saturation effects in LMS adaptive echo cancellation for binary data,” *IEEE transactions on Acoustics, Speech, and Signal Processing*, vol. 38, pp. 1687–1696, Oct 1990.
- [73] E. Eweda, “Comparison of RLS, LMS, and sign algorithms for tracking randomly time-varying channels,” *IEEE transactions on Signal Processing*, vol. 42, pp. 2937–2944, Nov 1994.
- [74] S. Koike, “Convergence analysis of a data echo canceler with a stochastic gradient adaptive FIR filter using the sign algorithm,” *IEEE transactions on Signal Processing*, vol. 43, no. 12, pp. 2852–2861, Dec 1995.
- [75] V.H. Nascimento and A. H. Sayed, “Stability of the LMS adaptive filter by means of a state equation”, *Proc. 36th Annual Allerton Conference on Communication, Control and Computing*, pp. 245-251, 1998.



- [76] T. Y. Al-Naffouri and A. H. Sayed, “Adaptive filters with error nonlinearities: Mean-square analysis and optimum design” *EURASIP Journal on Applied Signal Processing, Special issue on nonlinear signal processing and applications*, vol. 2001, no. 4.
- [77] Y. Wei, S. B. Gefand, and J. V. Krogmeier, “Noise constrained least mean squares algorithm,” *IEEE Transactions on Signal Processing*, vol. 49, pp. 1961–1970, Sept 2001.
- [78] J.I., Nagumo and A. Noda, “A learning method for system identification,” *IEEE Transactions, Automatic control*, vol. AC-12, pp. 282–287, 1967.
- [79] A. Zerguine and M. Bettayeb, “Adaptive System Identification Using the Normalized Least Mean Fourth Algorithm,” *Proceedings of EUSIPCO-98, Greece*, pp. 737–740, September 1998.
- [80] M. Rupp and A. H. Sayed, “A Time-Domain Feedback Analysis Of Adaptive Algorithms via the Small Gain Theorem ,” *Proceedings. SPIE*, vol. 2563, pp. 458–469, 1995.
- [81] M. Rupp and A. H. Sayed, “A Time-Domain Feedback Analysis Of Filtered-Error Adaptive Gradient Algorithms,” *IEEE transactions on Signal Processing*, vol. 44, pp. 1428–1439, 1996.
- [82] A. H. Sayed and M. Rupp, *Robustness Issues in Adaptive Filtering, DSP Handbook*, CRC Press, 1998.

- [83] T. Y. Al-Naffouri and A. H. Sayed, “Transient Analysis of Adaptive Filters With Error Nonlinearities,” *IEEE Transactions on Signal Processing*, vol. 51, pp. 653–663, March 2003.
- [84] T. Y. Al-Naffouri and A. H. Sayed, “Transient Analysis of Data-Normalized Adaptive Filters,” *IEEE Transactions on Signal Processing*, vol. 51, pp. 639–652, March 2003.
- [85] J. A. Chambers, O. Tanrikulu and A. G. Constantinides, “Least mean mixed-norm adaptive filtering,” *Electronic Letters*, vol. 30, no. 19, pp. 1574–1575, 1994.
- [86] R. H. Kwong and E. W. Johnston, “A Variable Stepsize LMS algorithm,” *IEEE transactions on signal processing*, vol. 40, pp. 1633–1642, 1992.
- [87] R. W. Harris, D.M. Chabries, and F. A. Bishop, “A Variable Stepsize (VS) algorithm,” *IEEE transactions on Acoustic, Speech and Signal Processing*, vol. 34, pp. 499–510, 1986.
- [88] T. Y. Al-Naffouri, A. H. Sayed and T. Kailath, “On the selection of optimal nonlinearities for stochastic gradient adaptive algorithms,” *in Proc. ICASSP*, vol. 1, pp. 464-467, Istambul, Turkey, June 2000.

## VITAE

- Muhammad Saqib Sohail got his bachelors degree in Electrical Engineering (major electronics and communications) with honors from University of Engineering and Technology, Lahore, Pakistan in December 2004.
- Worked as lecturer at UET, Lahore from January 2005 to February 2006.
- Worked as project engineer at National Advanced Systems Company (Motorola), Riyadh from March 2006 to August 2006.
- Joined King Fahd University of Petroleum and Minerals, Dhahran in September 2006.
- Completed masters degree requirement in June 2008 with a CGPA of 3.72/4.0.
- email: saqib@kfupm.edu.sa