Density estimation communication systems

Wang, Peng

2007

Wang, P. (2007). Density estimation communication systems. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/3472>

<https://doi.org/10.32657/10356/3472>

Nanyang Technological University

Downloaded on 26 Apr 2023 07:36:43 SGT

Density Estimation for Communication Systems

by
Wang Peng

A Dissertation Submitted in Partial Fulfilment of the
Requirements for the Degree of
Doctor of Philosophy

in the
School of Electrical & Electronic Engineering
of the

NANYANG TECHNOLOGICAL UNIVERSITY

2007

Acknowledgements

I would like to express my sincere appreciation to Associate Professor Ser Wee. Ever since 1998, when I was a Master's candidate under his supervision in Nanyang Technological University, he has been a source of invaluable advice, guidance and encouragement.

I also want to thank my colleagues in the Center for Signal Processing, especially Mr. Lu Dawei, Dr. Yu Zhuliang, Mr. Yang Fan and Ms. Hoon Torshiang, for always being there whenever help is needed.

I am deeply grateful to my friends in Singapore. Their trust and friendship have made the past four years a memorable time in my life. Special thanks go to Mr. Timmy Tan, Dr. Li Yuan, Mr. Xu Yong and Mr. Ge Qi.

Furthermore, I want to thank my parents, siblings and friends in China. They have shaped me into the person that I am today.

Finally, I owe a debt of gratitude to my wife, Furong. Without her love, patience and understanding, this thesis would not have been possible.

Abstract

Density estimation refers to a family of techniques that attempt to model the underlying statistics (e.g. the probability density function) of a signal based on a sequence of observations. Today, it has become one of the most important tools in data analysis. While parametric density estimation has already gained a wide recognition in communications, the non- and semi-parametric density estimation techniques are relatively less attended. The objective of this study is to fill this gap by investigating two problems that cannot be easily handled with parametric density estimation approaches.

The first problem is on the use of nonparametric density estimate to improve the data efficiency of *a posteriori* ER (error rate) estimator. The idea is not new. However, the optimization of the density estimator for ER estimation has never been addressed in the literature. This thesis presents an optimization approach to minimizing the mean square error of the ER estimation. In general, an ER estimator based on density estimation requires the transmitted signal to be known *a priori*. In this study, we develop a new blind ER estimation approach that combines the strengths of the nonparametric density estimation and the ER extrapolation techniques. The effectiveness of the proposed approach is verified by computer simulations.

The second problem concerns ISI (Inter-Symbol Interference) mitigation. Conventional equalization approaches rely on accurate estimation of the CIR (Channel Impulse

Response). However, accurate CIR estimation is often associated with a costly waste of bandwidth. This thesis proposes a blind channel estimation method, where the amplitude of the CIR is derived from a non- or semi-parametric density estimate of the underlying statistics. Based on the findings obtained, we develop a new channel equalization algorithm. The proposed equalizer greatly reduces the difficulties encountered in conventional channel estimation techniques without degrading the BER performance. For a symbol-based detector, its performance is comparable to the decision-feedback equalizer in some cases. If a sequence detector is implemented, its performance can approach that of the maximum-likelihood sequence estimator.

Contents

Acknowledgments	i
Abstract	iii
List of Figures	viii
List of Tables	xii
1 Introduction	1
1.1 Motivation	1
1.2 Objectives	2
1.3 Major Contributions	3
1.4 Organization	5
2 Density Estimation	7
2.1 Introduction	7
2.2 Overview of Density Estimation	8
2.3 Nonparametric Density Estimation	9
2.3.1 Histogram and Naive Estimation	9
2.3.2 Kernel Estimation	10
2.3.3 Nearest Neighbor Estimation	12
2.3.4 Variable Kernel Estimation	13
2.3.5 Orthogonal Series Estimation	13
2.4 Finite Mixture Analysis	15

3	Kernel Real-Error Monitor	17
3.1	Introduction	17
3.2	Overview of Error Rate Estimation	18
3.3	Performance Criteria	21
3.4	Real-Error Counting Monitor	23
3.5	Kernel Real-Error Monitor	24
3.5.1	Principle of Kernel Real-Error Monitoring	25
3.5.2	Optimal Smoothing Effect	27
3.5.3	Kernel Real-Error Monitoring for M -ary Modulation	35
3.5.4	Blind Estimation Inability	38
3.5.5	Computer Simulations	39
3.6	Summary	46
4	Kernel Pseudo-Error Monitor	47
4.1	Introduction	47
4.2	Traditional Pseudo-Error Monitor	48
4.2.1	Principle of Pseudo-Error Monitoring	48
4.2.2	Statistical Properties	50
4.2.3	Traditional Pseudo-Error Monitors	53
4.3	Kernel Pseudo-Error Monitor	54
4.3.1	Principle of Kernel Pseudo-Error Monitoring	54
4.3.2	Properties and Optimization	59
4.3.3	Performance Advantages	59
4.3.4	Computer Simulations	61
4.4	Summary	66

5	Origin Analysis Equalizer	67
5.1	Introduction	67
5.2	Overview of Channel Equalization	68
5.3	Principle of Origin Analysis Equalization	70
5.4	Origin Localization	74
5.4.1	Nonparametric PDF Estimation	74
5.4.2	Peak Extraction	75
5.4.3	Peak Verification	76
5.4.4	L and σ_n^2 Estimation	77
5.4.5	Localization Examples	78
5.4.6	Kernel Localization	81
5.4.7	Limitations	82
5.5	Origin Identification	83
5.5.1	Non-Blind Identification	83
5.5.2	Blind Identification	84
5.5.3	Blind Identification for M -ary Modulation	97
5.6	Decision Logic	98
5.7	Computer Simulations	100
5.8	Summary	104
6	Semi-Parametric Origin Analysis Equalizer	106
6.1	Introduction	106
6.2	Finite Gaussian Mixture Model	107
6.3	Maximum Likelihood Localization	107
6.3.1	Unconstrained EM Localization	108

6.3.2	Constrained EM Localization	111
6.3.3	Rate of Convergence	118
6.3.4	Spurious Convergence Control	121
6.3.5	EM Localization for M -ary Modulation	125
6.4	EM-Based Origin Identification	126
6.5	Sequence Detection	127
6.6	Adaptive Origin Analysis Equalization	130
6.7	Error Performance Analysis	131
6.7.1	Error Performance of Symbol Detection	131
6.7.2	Error Performance of Sequence Detection	136
6.7.3	Effect of Channel Length Misestimation	140
6.8	Computer Simulations	143
6.9	Summary	145
7	Conclusions and Recommendations	152
7.1	Conclusions	152
7.2	Recommendations for Future Research	154
	Author's Publications	157
	Bibliography	159

List of Figures

3.1	A unified framework for error monitoring	19
3.2	Inability of kernel real-error monitor in blind estimation	39
3.3	Normalized rms errors of the optimum kernel real-error monitor and the real-error counting monitor	40
3.4	ER estimates produced by the optimum kernel real-error monitor and the real-error counting monitor	42
3.5	Percent coverage of the optimum kernel real-error monitor and the real- error counting monitor	43
3.6	Normalized rms errors of the optimum kernel real-error monitor and the monitor with triple smoothing effect	43
3.7	ER estimates produced by the optimum kernel real-error monitor and the monitor with triple smoothing effect	44
3.8	ER estimates produced by the optimum kernel real-error monitor for 16-QAM	45
4.1	Relationship between the logarithmic ER and the SNR for BPSK trans- mission over an AWGN channel	49
4.2	Normalized rms error of the traditional pseudo-error monitor	52
4.3	Block diagram of the kernel pseudo-error monitor based on threshold mod- ification and linear extrapolation	56

4.4	Block diagram of the kernel pseudo-error monitor based on signal degradation and linear extrapolation	58
4.5	Normalized rms errors of the threshold modification monitor and the kernel real- and pseudo-error monitors	62
4.6	Percent coverage of the threshold modification monitor and the kernel real- and pseudo-error monitors	63
4.7	ER estimates produced by the optimum kernel pseudo-error monitor and the threshold modification monitor	64
4.8	ER estimates produced by the optimum kernel pseudo-error monitor in non-Gaussian environment	65
4.9	ER estimates produced by the optimum kernel pseudo-error monitor in the presence of ISI	65
5.1	Notions of <i>concurrent origin pair</i> and <i>concurrent state pair</i>	72
5.2	Block diagram of <i>origin analysis equalizer</i>	73
5.3	An example of nonparametric <i>origin localization</i>	79
5.4	An example of nonparametric <i>origin localization</i>	80
5.5	Accuracy of nonparametric <i>origin localizers</i>	82
5.6	Identification of the <i>state-origin</i> correspondences	96
5.7	Origin identification for 4-QAM	99
5.8	BER performance of nonparametric <i>origin analysis equalizer</i> for a channel with impulse response $[1, 0.5]^T$	102
5.9	BER performance of nonparametric <i>origin analysis equalizer</i> for a channel with impulse response $[1, -1]^T$	102
5.10	BER performance of nonparametric <i>origin analysis equalizer</i> for a channel with impulse response $[1, -0.5, 0.7]^T$	103

5.11	BER performance of nonparametric <i>origin analysis equalizer</i> for non-Gaussian noise.	103
6.1	Histogram of the <i>origin localization</i> error of unconstrained EM localizer for a channel with impulse response $[1, 0.5, 0.25]^T$	112
6.2	Error performance of unconstrained EM <i>origin localizer</i>	112
6.3	Histogram of the <i>origin localization</i> error of constrained EM localizer Ia for a channel with impulse response $[1, 0.5, 0.25]^T$	114
6.4	Histogram of the <i>origin localization</i> error of constrained EM localizer IIa for a channel with impulse response $[1, 0.5, 0.25]^T$	114
6.5	Histogram of the <i>origin localization</i> error of constrained EM localizer Ia for a channel with impulse response $[1, 0.5, 0.7]^T$	116
6.6	Histogram of the <i>origin localization</i> error for constrained EM localizer IIa for a channel with impulse response $[1, 0.5, 0.7]^T$	116
6.7	Convergence of constrained EM <i>origin localizers</i> for a channel with impulse response $[1, 0.5, 0.7]^T$	119
6.8	Correspondence between the <i>origin localization</i> error and the variance estimation error in constrained EM localization Ia	122
6.9	Histogram of the <i>origin localization</i> error of constrained EM localizer Ia ⁺ with SCC I for a channel with impulse response $[1, 0.5, 0.7]^T$	124
6.10	Histogram of the <i>origin localization</i> error of constrained EM localizer Ia ⁺ with SCC II for a channel with impulse response $[1, 0.5, 0.7]^T$	124
6.11	Illustrations of error propagation in OAE	127
6.12	Possible state transitions in a 3-tap binary channel	129
6.13	Probability of consecutive errors	135

6.14	Probability of independent error in symbol and sequence detection	141
6.15	Effect of underestimation of channel order	142
6.16	BER performance of semi-parametric <i>origin analysis equalizer</i> for a channel with impulse response $[1, -0.5, 0.7]^T$	146
6.17	BER performance of semi-parametric <i>origin analysis equalizer</i> for a COST 207 TU channel with impulse response $[1, -0.66]^T$	146
6.18	BER performance of semi-parametric <i>origin analysis equalizer</i> for a COST 207 BU channel with impulse response $[1, 0.21, 0.78]^T$	147
6.19	BER performance of semi-parametric <i>origin analysis equalizer</i> for a COST 207 HT channel with impulse response $[1, 0, 0, 0, -0.52, 0.17]^T$	147
6.20	BER performance of semi-parametric <i>origin analysis equalizer</i> for a COST 207 TU channel with impulse response $[-0.73, 1]^T$	148
6.21	BER performance of semi-parametric <i>origin analysis equalizer</i> for a 7-tap channel	148
6.22	Effect of underestimation of channel length	149
6.23	Performance of semi-parametric <i>origin analysis equalizer</i> for 4-QAM	150

List of Tables

5.1	<i>Origin</i> transition record obtained for a communication channel with impulse response $[1, 0.25, 0.5]^T$ and at SNR 13 dB (2000 symbols)	93
5.2	<i>Origin</i> transition record obtained for a communication channel with impulse response $[1, 0.5, 0.25]^T$ and at SNR 13 dB (2000 symbols)	93
5.3	<i>Origin</i> transition record obtained for a communication channel with impulse response $[-0.5, 0.25, 1]^T$ and at SNR 13 dB (2000 symbols)	94
5.4	<i>Origin</i> transition record obtained for a communication channel with impulse response $[-1, -0.25, -0.5]^T$ and at SNR 13 dB (2000 symbols)	94
6.1	Performance indices of various EM <i>origin localizers</i> for a channel with impulse response $[1, 0.5, 0.25]^T$ and at SNR 13 dB (100 samples)	117
6.2	Performance indices of various EM <i>origin localizers</i> for a channel with impulse response $[1, 0.5, 0.7]^T$ and at SNR 13 dB (100 samples)	117
6.3	Performance indices of constrained EM <i>origin localizer</i> Ia^+ for a channel with impulse response $[1, 0.5, 0.7]^T$ and at SNR 13 dB (100 samples)	120
6.4	Rates of occurrence of spurious convergence with and without spurious convergence control (100 samples)	125

Chapter 1

Introduction

1.1 Motivation

As an old Chinese proverb says, all mutations follow some immutable principles. However, in most circumstances, these principles are hidden in their dynamic presentations and cannot be observed directly. Density estimation¹ provides us a method to see through the illusive appearances and understand the nature of things better.

In general, density estimation refers to a family of techniques that attempt to accurately and concisely model a set of random observations. According to the specific demand on the accuracy and simplicity of the model, three approaches of density estimation can be performed, parametric, semi-parametric and nonparametric estimations. While parametric density estimation has been widely used in communications, e.g. the prevalent Gaussian model for a channel suffering from additive noise only, the other two approaches are relatively less attended to. Parametric models have distinct advantages in clear presentation and ease of use, but their suitability in more complicated environments

---

¹In this thesis, the term ‘density’ refers specifically to probability density.

is often doubtful. For example, the Gaussian model cannot accurately reflect the channel characteristics in the presence of ISI (Inter-Symbol Interference). It is our belief that the problems associated with parametric modeling can be, at least, alleviated by using semi- or non-parametric density estimate.

1.2 Objectives

The main objective of this thesis work is to investigate the properties and possible applications of the semi- and non-parametric density estimation techniques to communications for problems that cannot be handled easily by the parametric approaches. In particular, two problems have been investigated: the estimation of the *a posteriori* Error Rate (ER) and the mitigation of the ISI.

The *a priori* ER estimation methods involve the use of some parametric statistical models to predict the behavior of a communication system. Due to the generality and simplicity consideration, a parametric model can only attend to ensemble features. Consequently, performance prediction made by a parametric model is not accurate enough for the purpose of system monitoring. In contrast, an *a posteriori* ER estimate is derived from observations made at the receiver and it takes into account additional information of the system that are neglected in parametric modeling. One major concern in *a posteriori* ER estimation is the observation time required for generating consistently reliable estimates. The conventional method of counting the transmission errors needs an extremely long observation time. Moreover, it requires the transmitted signal to be known *a priori* (non-blind ER estimation). By extrapolating the ER from a set of PERs (Pseudo-Error Rates), the pseudo-error monitoring technique overcomes the problem of needing the transmitted signal to be known and makes blind estimation possible. In addition, as a result of more

efficient exploitation of the observations (compared to real-error counting approach), the observation cost can be reduced greatly. However, its efficiency is still inadequate when the system needs to work with high fidelity. In this study, we will illustrate how nonparametric density estimation can be used, solely or together with the pseudo-error monitoring technique, to enhance the data efficiency of *a posteriori* ER estimator.

The presence of ISI is of critical concern in high-speed data communications. During the past decades, a considerable amount of effort has been devoted to the exploration for effective and efficient algorithms to ISI mitigation. However, nearly all these algorithms rely on the processing of the CIR (Channel Impulse Response). While CIR has been used as the standard representation of channel characteristics, it is not the only possible representation. From the point of view of statistics, the effect of ISI is a change in the distribution of the decision variable. Hence, there is a possibility to derive the desired information from the underlying density instead of from the CIR. The other objective of this study is to investigate the role that density estimation can play in channel equalization.

1.3 Major Contributions

The major contributions of this thesis are summarized as follows.

- As mentioned in the previous section, the *a posteriori* ER estimation method based on real-error counting is extremely inefficient in the sense that only the transmission errors are involved in the construction of the ER estimate. Several researchers have explored the feasibility of using nonparametric density estimate in error monitoring [1–4]. However, the optimization of the density estimator for ER estimation has not been addressed. This study develops an optimization algorithm that minimizes the mean square error of the ER estimate.

CHAPTER 1. INTRODUCTION

- Error monitoring based solely on nonparametric density estimation still relies on the knowledge of the transmitted signal for reliable ER estimation. In this study, the problem is overcome by integrating ER extrapolation with nonparametric density estimation. The resultant error monitoring approach inherits the advantages of both techniques and is able to provide reliable and efficient blind ER estimation.
- A number of ISI mitigation or equalization algorithms have been studied in the literature, e.g., linear equalization, decision-feedback equalization, maximum likelihood sequence estimation, etc. Most of these algorithms require accurate CIR estimation, which is, however, often costly to obtain. This study presents a novel approach to channel modeling, termed “*origin analysis*”, with the objective of avoiding associated with CIR estimation. The *origin analysis* consists of two processes: the first process, *origin classification*, constructs a model for the underlying density, and the second process, *origin identification*, aims to retrieve the phase information that has been discarded in the statistical modeling process. On the basis of *origin analysis*, a new equalization algorithm has been developed in this study. As verified by Monte Carlo simulations, the proposed method can provide good performance in mitigating the effect of ISI without the need to perform the difficult task of channel estimation.
- The *origin analysis* produces a set of parameters, from which the CIR can be derived easily. Therefore, the *origin analysis* can also be considered as a blind approach to CIR estimation. In this sense, it has the potential to be integrated with CIR-based equalization algorithms. A detailed discussion on this issue, however, is beyond the scope of the study.

1.4 Organization

This thesis consists of seven chapters, and the content of each chapter is as follows.

- Chapter 1, the present chapter, outlines the motivation, objectives, major contributions and scope of this PhD study.
- Chapter 2 reviews the literature on density estimation, and briefly introduces the most important semi- and non-parametric density estimation algorithms. The advantages and disadvantages of these algorithms are also described.
- Chapter 3 and 4 are dedicated to *a posteriori* ER estimation. An overview of ER estimation is given first in Chapter 3. This is then followed by a description of two popular error monitors: one based on error counting and the other based on nonparametric density estimation. A novel algorithm that optimizes the density estimator for ER estimation is proposed in this chapter with simulation results demonstrating the advantage of using the proposed optimal density estimate.
- Error monitoring based on the estimation of the underlying density is essentially a nonblind approach. Chapter 4 addresses this problem. Chapter 4 begins with a discussion on pseudo-error monitoring, the most well-known blind estimation approach, from which, it can be clearly seen that pseudo-error monitoring has complementary characteristics to the approach based on density estimation. A novel error monitoring technique is then developed, which is an integration of the above two approaches. Computer simulations show that the proposed error monitor, as expected, successfully combines the strengths of both approaches and has the capability to generate reliable and efficient blind ER estimate.

CHAPTER 1. INTRODUCTION

- Chapter 5 and 6 investigate the application of density estimation techniques in channel ISI mitigation. After a short review of the existing equalization algorithms, Chapter 5 introduces the principle of *origin analysis equalization*. As the simplest realization of this principle, an *origin analysis equalizer* based on nonparametric density estimation is studied extensively. Monte Carlo simulations are performed to demonstrate the strengths and limitations of this equalizer.
- Equalization based on nonparametric *origin analysis* has serious limitations. This is due to the fact that an *origin analysis equalizer* still relies on a set of parameters, which cannot be directly derived from a nonparametric density estimate (in another word, parameterization of the nonparametric model is necessary). Semi-parametric modeling seems to be more suitable in the context of channel equalization. Chapter 6 studies *origin analysis equalization* based on semi-parametric density estimation and verifies this conjecture.
- Chapter 7 lists the problems that should be included in future investigations and concludes this thesis work.

Chapter 2

Density Estimation

2.1 Introduction

Non- and semi-parametric density estimation techniques have already received a wide recognition in various areas, such as economics, medicine, science, and engineering, as a powerful tool for performing explorative data analysis. However, relatively little effort has been devoted to use these techniques in communications to solve the problems caused by the use of some simplistic statistical models. The objective of this thesis is to help fill this gap.

This chapter briefly reviews several important density estimation algorithms. Of special interest are the naive and kernel nonparametric approaches and the EM (Expectation-Maximization) semi-parametric approach. The kernel estimator will be exploited in Chapter 3 and 4 to achieve higher data efficiency in *a posteriori* error rate estimation, while the naive and the EM estimators will be used in Chapter 5 and 6, respectively, to overcome the difficulties of estimating the channel impulse response for channel equalization.

2.2 Overview of Density Estimation

In probability and statistics, density estimation refers to a family of techniques that are used to construct an estimate for an unobservable underlying probability density function (PDF) from the observed data. There exist three different types of density estimation algorithms: parametric, semi-parametric, and nonparametric.

Parametric algorithms assume that the observed data are drawn from one of a known family of distributions [5]. The problem of density estimation is accordingly translated into the problem of estimation of a set of parameters that defines the assumed distribution. For example, in communications, the Gaussian statistical model is often assumed to describe the transmission channel. Parametric estimation is attractive for its simplicity. However, it is unrealistic to expect that a generic model such as the Gaussian would be able to represent the concerned statistic in all circumstances. This explains why so many models other than the Gaussian have been developed in the literature for communication channels.

In contrast to parametric methods, a nonparametric approach makes no *a priori* assumptions about the functional form of the underlying density, i.e., it allows the data to speak for themselves. Nonparametric approaches are more flexible in forming the density estimates. Consequently, they can provide more accurate descriptions for the concerned statistics than parametric approaches. However, this flexibility is not always an advantage. In the cases where the density estimates are used for further analysis and investigation, a mathematically tractable density estimate is usually more appreciated.

Semi-parametric methods seek to achieve a balance between the tractability and the accuracy of the result. In semi-parametric estimation, the underlying density is regarded as a convex linear combination¹ of component densities from a known family, Gaussian

---

¹A convex combination is a linear combination of data points (which can be vectors or scalars) where all coefficients are positive and sum up to 1.

densities for instance. Thus, semi-parametric approaches have more freedom in representing the densities than parametric approaches and can provide more tractable results than nonparametric approaches. Semi-parametric methods are frequently grouped into parametric estimation in the literature, due to the fact that they are also based on some parametric models.

2.3 Nonparametric Density Estimation

Nonparametric density estimation, which aims to reveal the underlying distribution based on one or several random samples from a population, has become one of the most active research areas in statistics since the publication of the classical work of Parzen in 1962 [6]. In this section, prevalent nonparametric algorithms are introduced briefly.

2.3.1 Histogram and Naive Estimation

The histogram was developed in the 19th century, and it is still in common use today as a visual clue to the general shape of the underlying distribution [5]. The histogram is simple to calculate; however, as a density estimator, it leaves too much to be desired. One main drawback of this method is the discontinuity of the resultant estimate at the bin boundaries, which may cause serious problem if the derivative of the density estimate is required. Besides, since the histogram is zero outside a certain range, it is inapplicable when the tails of the underlying distribution are highly concerned, e.g., in error rate estimation. A more detailed description of the deficiency of the histogram estimator can be found in [7].

A histogram density estimator for data sequence $\{X_i\}$ can be constructed as follows

$$\hat{f}(x) = \frac{1}{n} \times \frac{n_x}{w_x} \quad (2.1)$$

CHAPTER 2. DENSITY ESTIMATION

---

where n is the sample size, n_x is the number of observations (data) falling into the bin that covers x , and w_x denotes the corresponding bin width. The bin specifications are of critical importance for histogram estimation. The number of bins and the bin widths affect the overall and local smoothness of the density estimate [8,9]. Sometimes, the choice of bin locations can also have a significant impact on the accuracy of the estimate [10]. Thus, the bin locations should be determined specifically rather than arbitrarily.

The naive approach [11] has been developed to overcome the difficulty of designing data-specific bin locations. A naive density estimator can be written as

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n W\left(\frac{x - X_i}{h}\right) \quad (2.2)$$

where h specifies the bin width and W is a weight function given by

$$W(x) = \begin{cases} 0.5 & \text{if } |x| < 1, \\ 0 & \text{otherwise.} \end{cases} \quad (2.3)$$

As is shown in Eq. (2.2), the naive estimate is nothing more than a histogram obtained by setting the bin width to $2h$ and arranging the bins such that each point x is right at the center of a bin. The naive method frees the histogram from a particular choice of bin locations, but it is still unable to provide continuous density estimates.

2.3.2 Kernel Estimation

The development of kernel density estimation is normally credited to Parzen [6], although its principle was first introduced by Rosenblatt in 1956 [11]. The kernel density estimator can be interpreted as a generalization of the naive estimator. By replacing weight function W with kernel function K in Eq. (2.2), we can obtain the expression of kernel density estimator

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (2.4)$$

CHAPTER 2. DENSITY ESTIMATION

---

where h is often referred to as the smoothing parameter, window width or bandwidth in this context. To assure that the resultant $\hat{f}(x)$ is a *bona fide* probability density, the kernel function K must satisfy

$$\int_{-\infty}^{+\infty} K(x) dx = 1. \quad (2.5)$$

A common practice is to choose a density function as the kernel, the standard Gaussian density for instance. It can be seen from Eq. (2.4) that $\hat{f}(x)$ composed of density kernels is also a density. The smoothing parameter h determines how many details exhibited by the sample distribution will be maintained in the density estimate. If h is set too small, the spurious fine structure can become visible, and if h is set too large, some important features can be obscured. In both cases, $\hat{f}(x)$ will display a large deviation from the actual density. The optimum value of h relies on the choice of the kernel, the actual density, and the criterion used to evaluate the quality of the density estimate.

The first [11] and the most widely used measure of the global accuracy of $\hat{f}(x)$ is the mean integrated square error (MISE), which is defined as

$$\text{MISE } \hat{f} = \text{E} \int_{-\infty}^{+\infty} [\hat{f}(x) - f(x)]^2 dx. \quad (2.6)$$

The optimum smoothing parameter in the sense of minimizing the MISE, denoted as h_0 , has been found to be [6]

$$h_0 = \left[\int_{-\infty}^{+\infty} x^2 K(x) dx \right]^{-2/5} \left[\int_{-\infty}^{+\infty} K(x)^2 dx \right]^2 \left[\int_{-\infty}^{+\infty} f''(x)^2 dx \right]^{-1/5} n^{-1/5}. \quad (2.7)$$

If the concerned statistic is a Gaussian distribution with variance σ^2 , for the choice of the standard Gaussian kernel h_0 can be calculated simply as follows

$$h_0 \approx 1.06\sigma n^{-1/5}. \quad (2.8)$$

A quick way of choosing the smoothing parameter, therefore, would be to estimate σ from the collected data and then to substitute the estimate into the above equation. For many purposes this h_0 is certainly adequate, while for others it can still serve as a good starting point for subsequent fine tuning.

2.3.3 Nearest Neighbor Estimation

The nearest neighbor estimation represents an attempt to adapt the amount of smoothing to the local density of data [5, 12–14]. The degree of smoothing is controlled by an integer k , chosen to be considerably smaller than the sample size, typically $k = \sqrt{n}$.

Define the distance $d(x_1, x_2)$ between two points on the line to be $|x_1 - x_2|$. Given an observation sequence of x , $\{X_i\}$, we have n distances for an arbitrary x , from x to each of the observations. Denote the distances as $\{d_i(x)\}$ and arrange them in ascending order, i.e.,

$$d_1(x) \leq d_2(x) \leq \cdots \leq d_n(x).$$

The k^{th} nearest neighbor density estimate can then be formed as

$$\hat{f}(x) = \frac{k-1}{2nd_k(x)}. \quad (2.9)$$

As can be seen, the nearest neighbor estimator can be interpreted as a kernel estimator using $d_k(x)$ as the smoothing parameter and the rectangular function with width $k-1$ and height 0.5 as the kernel.

It follows from Eq. (2.9) that the nearest neighbor estimate is inversely proportional to the size of the box that covers $k-1$ observations. Note that in the tails of the distribution, the distance $d_k(x)$ is larger than in the main part. Accordingly, the problem of inadequate smoothing in the tails is alleviated. However, as the tails of the distribution can easily be overemphasized, the nearest neighbor estimation method is inappropriate when an estimate of the entire density is required [5], e.g., in error rate calculation. It should also be noted that the nearest neighbor estimate has discontinuous derivative at points $x = X_i$ and its integral is unbounded.

2.3.4 Variable Kernel Estimation

The variable kernel estimation is another effort to adapt the smoothing to the local density of data [5, 12, 15, 16]. It is defined by

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{hd_k(X_j)} K \left[\frac{x - X_i}{hd_k(X_j)} \right] \quad (2.10)$$

where $d_k(X_j)$ is the distance from X_j to the k^{th} nearest point in the data set consisting of the other $n - 1$ data points. The window width of the kernel placed on the point X_j is proportional to $d_k(X_j)$, so that data points in sparse regions will be weighted by flatter kernels. While parameter k determines the local variation of the smoothing effect, h decides the amount of the overall smoothing.

In contrast with the nearest neighbor estimate, a variable kernel estimate is itself a PDF provided the kernel K is. Furthermore, as in static kernel estimation, all the local smoothness properties of the kernel will be inherited by the estimate.

2.3.5 Orthogonal Series Estimation

The orthogonal series estimation solves the problem of PDF approximation from a different point of view [12, 17, 18]. The basic principle underlying this method is to estimate the PDF by estimating the coefficients of its orthogonal series expansion.

Let $\{\varphi_i(x)\}$ denote a complete orthonormal system of functions, and assume that the desired density $f(x)$ is square-integrable and converges to its series expansion on $\{\varphi_i(x)\}$

$$f(x) = \sum_{i=0}^{\infty} c_i \varphi_i(x) \quad (2.11)$$

where c_i is the projection of $f(x)$ onto $\varphi_i(x)$, i.e.,

$$c_i = \int_{-\infty}^{\infty} f(x) \varphi_i^*(x) dx \quad (2.12)$$

CHAPTER 2. DENSITY ESTIMATION

where $\varphi_i^*(x)$ represents the complex conjugate of $\varphi_i(x)$. The coefficient c_i can be estimated unbiasedly from the observation sequence $\{X_i\}$ as follows

$$\hat{c}_i = \frac{1}{n} \sum_{j=1}^n \varphi_i^*(X_j). \quad (2.13)$$

Various orthogonal function systems can be used as the basis of the expansion. The most popular orthogonal series density estimator, the Gram-Charlier's estimator, chooses the Hermite polynomials and retains only the first M expansion terms to estimate $f(x)$, i.e.,

$$\hat{f}(x) = h(x) \sum_{i=0}^M c_i H_i(x) \quad (2.14)$$

where $h(x)$ is the standard Gaussian function and

$$H_i(x) = (-1)^i h^{-1}(x) \frac{d^i h(x)}{dx^i}. \quad (2.15)$$

Let μ_x and σ_x^2 be respectively the mean and variance of x . The coefficient c_i appeared in this series expansion can be expressed in terms of the moments of the normalized random variable $(x - \mu_x)\sigma_x^{-1}$

$$c_i = \frac{1}{i!} \left[\mu_i - \frac{i^{[2]}}{2^1 \times 1!} \mu_{i-2} + \frac{i^{[4]}}{2^2 \times 2!} \mu_{i-4} - \cdots \right] \quad (2.16)$$

where $i^{[k]} = i!/(i-k)!$ and μ_k denotes the k^{th} order moment of $(x - \mu)\sigma^{-1}$. Thereby, the problem of PDF approximation is translated into moment estimation. A simple moment estimate can be derived from the observations of x , $\{X_i\}$, as follows

$$\hat{\mu}_k = \frac{1}{n} \sum_{i=1}^n \left(\frac{X_i - \hat{\mu}_x}{\hat{\sigma}_x} \right)^k \quad (2.17)$$

where $\hat{\mu}_x$ and $\hat{\sigma}_x$ are respectively the sample mean and standard deviation. In pursuit of better PDF approximation, the moment estimators that are more robust against spurious data should be employed [19].

The convergence speed of orthogonal series estimator is independent of the dimension of the data, which gives it an advantage over the kernel estimator for multivariate density estimation. However, except for rather special choices of the expansion basis, the resultant PDF approximate is not guaranteed to be a density.

2.4 Finite Mixture Analysis

Semi-parametric density estimation often refers specifically to the technique of finite mixture analysis. A finite mixture is a distribution that can be expressed as the superposition of a finite number of component distributions, i.e.,

$$f(x) = \sum_{i=1}^c P_i g_i(x; \boldsymbol{\theta}_i) \quad (2.18)$$

where c denotes the number of mixing components, $g_i(x; \boldsymbol{\theta})$ is the i^{th} component distribution depending on the parameter vector $\boldsymbol{\theta}_i$, and $\{P_i\}$ represent the mixing proportions. Let us use $\boldsymbol{\Psi}$ to represent the union of $\{\boldsymbol{\theta}_i\}$ and $\{P_i\}$. Eq. (2.18) becomes

$$f(x) = \sum_{i=1}^c g_i(x; \boldsymbol{\Psi}). \quad (2.19)$$

The likelihood function for $\boldsymbol{\Psi}$ can be formed from the observed data $\{X_i\}$ as follows

$$\mathcal{L}(\boldsymbol{\Psi}) = \prod_{j=1}^n f\{X_j\} = \prod_{j=1}^n \sum_{i=1}^c g_i(X_j; \boldsymbol{\Psi}). \quad (2.20)$$

In many cases, it is more convenient to use the log-likelihood function given by

$$\log \mathcal{L}(\boldsymbol{\Psi}) = \sum_{j=1}^n \log \left\{ \sum_{i=1}^c g_i(X_j; \boldsymbol{\Psi}) \right\}. \quad (2.21)$$

The maximum-likelihood (ML) estimate of $\boldsymbol{\Psi}$ can then be obtained by as a solution to the likelihood equation

$$\partial \log \mathcal{L}(\boldsymbol{\Psi}) / \partial \boldsymbol{\Psi} = 0. \quad (2.22)$$

There are numerous techniques for finding the solution to the above equation, e.g., the Newton-Raphson method, quasi-Newton methods, Fisher's scoring method, and the EM (expectation-maximization) method [20–23]. Among these techniques, the EM algorithm has received the most attention in the literature because of the guaranteed convergence [20–25]. The EM algorithm is an iterative procedure. Each iteration consists of an E-step

CHAPTER 2. DENSITY ESTIMATION

---

and an M-step. The E-step computes the conditional expectation of the log-likelihood of the complete data given the observations and the current parameter estimate; and the subsequent M-step updates the parameter estimate to maximize this likelihood function. A detailed discussion on the EM algorithm for Gaussian mixture analysis will be given in Chapter 6.

Chapter 3

Kernel Real-Error Monitor

3.1 Introduction

During the past decades, a considerable amount of effort has been devoted to theoretical analysis of the error rate (ER) performances of data communications systems in various environments. Unfortunately, due to the lack of either effectiveness or efficiency, the *a priori* ER estimation methods, in the sense of presuming some specific channel models, can hardly satisfy the demand for a practical system monitoring solution. The *a posteriori* ER estimation techniques, which attempt to derive the ER estimate from the observations at the receiver have thus been developed.

The simplest *a posteriori* ER estimation scheme is to train the system and count the transmission errors. This real-error counting approach is, however, extremely inefficient in exploiting the available data. As a consequence, a formidably long observation time is needed for producing a reliable estimate for a low ER. This chapter investigates how and to what extent the nonparametric density estimation technique can help to reduce the observation cost.

3.2 Overview of Error Rate Estimation

The ultimate performance criterion for telecommunications systems is the degree of user satisfaction. However, due to the high cost of deploying a real-time feedback mechanism, the predictable reluctance of users to give responses, the suspicious credibility of the feedback information, and most importantly, the great inconvenience caused to users, this criterion is inapplicable in practical system monitoring. A common strategy of simplifying the problem is to select a set of gross features that are fairly easily measurable as the indicator of system performance. A well designed feature set, such as those suggested in [26–28], covers all aspects of service that have clear impact on users’ satisfaction. Usually, one or a few numerical measures, referred to as performance parameters, are defined for each selected feature. The full set of performance parameters thus form a numerical description of quality of service.

The error rate is the most direct measure of the accuracy of data transmission, which is a primary concern in telecommunications. It is defined to be the ratio of the number of data units received in error to the total number of data units transmitted. Within the scope of the general concept of error rate, several more specific terms have been defined, e.g., the bit error rate (BER), symbol error rate (SER), block error rate, residual error rate, and etc. This dissertation deals with the raw SER, i.e., the rate of erroneous symbols at the input to the detector (before performing error correction). Nevertheless, the approaches presented in the thesis can be readily extended to the estimation of other ER parameters.

The ER estimation can be based on theoretical calculation or field data analysis. The former methodology, termed *a priori* ER estimation, plays an important role in algorithm development and system design. It allows different algorithms or systems to be compared

Fig. 3.1. A unified framework for error monitoring

in laboratory by their potential of providing accurate data transmissions. During the past decades, a considerable amount of effort has been devoted to the theoretical prediction of the ER behaviors of telecommunication systems under various conditions [29–37].

Relying on specific channel models, where the name “*a priori* ER estimation” comes from, the theoretical analysis techniques are, unfortunately, inapplicable in system monitoring and maintenance. On the one hand, the communication channels are unpredictable in nature. As a result, the assumed channel models and thus the *a priori* solutions can lose their legitimacy easily. On the other hand, a relatively reliable channel model usually involves a number of parameters that are practically difficult to measure. The high cost of estimating these parameters make *a priori* solutions undesirable even in the cases that the channel models are adequate.

The *a posteriori* ER estimation methods are developed to fulfill the uprising demand for a timely alert for the time variation of the performance of a system in providing reliable data transmission. Unlike those *a priori* approaches, the *a posteriori* methods care little about the operation principles underlying the system. Instead, they attempt to estimate the ER based on an inductive analysis of the observed data.

A device that is used in system maintenance to provide *a posteriori* ER estimate is called an error monitor. Fig. 3.1 illustrates the typical error monitoring process. As is

CHAPTER 3. KERNEL REAL-ERROR MONITOR

---

shown, an error monitor extracts the effective information from the observations, with or without recurring to the source, from which an ER estimate is then derived.

The error monitors can be categorized into blind and non-blind monitors according to whether the source information input is needed or not; and according to how the effective information is exploited in calculating the ER, they can be grouped into real- and pseudo-error monitors. In real-error monitoring, which is the focus of this chapter, the effective information is used to calculate the ER directly. In pseudo-error monitoring, which will be addressed in the next chapter, a set of intermediate parameters, i.e. the pseudo error rates, are firstly estimated, which are then synthesized to give the desired ER estimate.

When inter-symbol interference (ISI) is present, a process known as channel equalization is often incorporated into the receiver, to suppress the interference and thus improve the ER performance (see Chapter 5 for a detailed discussion on channel equalization). In that case, the error monitor can be placed either before or after the ISI suppression, used to evaluate the necessity or effectiveness of implementing a channel equalizer. Note that the main challenge in error monitoring is to determine if the system performance is good enough to be acceptable, that is, the estimation of a low ER. The estimation of a high ER, e.g. in the presence of ISI or strong noise, is comparatively simple—if an error monitor can deal with low ERs, it should be able to deal with high ERs as well. Therefore, this study will not consider the ISI specifically. In other words, we assume that the ISI, if any, has been removed through channel equalization. In Chapter 4, simulation results will be given to show the effectiveness of the proposed approaches in the presence of ISI.

Overall, *a priori* ER estimators aim to predict the ensemble ER performance of the system in the environment consistent with the assumed channel model, while *a posteriori* estimators aim to measure the ER performance of the system at a specific moment.

3.3 Performance Criteria

Before going further into the discussion of specific error monitoring techniques, it is appropriate to clarify what makes a good error monitor. According to [38], an ideal error monitor should possess the following features:

1. Sensitivity to the factors that affect the error rate and robustness against those that have no or little influence.
2. Accuracy and reproducibility to a level adequate for the use being made of the information provided by the error monitor.
3. Fast response to change in the error rate to allow rapid detection of faults.
4. Ability to estimate the error rate with little or no interruption in the flow of useful information.
5. Simple implementation, so that monitoring does not substantially increase the cost or complexity of the system.

Feature 1 and 2 are related to the reliability of the ER estimator; feature 3 essentially highlights the efficiency of the estimation algorithm in exploiting the useful information buried in the observations; feature 4 addresses the online operability, or as is often referred to as, the blind estimation capability¹; and finally, feature 5 attends to the implementation cost. In this study we use all these features in the evaluation of an *a posteriori* approach,

---

¹The two concepts are slightly different indeed. Blind estimation capability emphasizes the independence on the *a priori* knowledge of the transmitted signal and is more stringent than online operability. A blind ER estimator can be used online or offline, while the online performance of a non-blind estimator is usually unacceptable. In this study, we do not consider their difference.

CHAPTER 3. KERNEL REAL-ERROR MONITOR

---

except feature 5, which, although a major concern in practice, has no direct relation with the monitoring capability.

While the blind estimation capability is generally a simple “yes” or “no” question, the reliability and efficiency of an ER estimator have to be presented numerically. A prevalent approach in point estimation is to measure the reliability of an estimator by its estimation bias, variance, and mean square error (MSE) [39], i.e.,

$$\text{bias } \hat{P}_0 = E \hat{P}_0 - P_0 \quad (3.1a)$$

$$\text{var } \hat{P}_0 = E \hat{P}_0^2 - E^2 \hat{P}_0 \quad (3.1b)$$

$$\text{MSE } \hat{P}_0 = E(\hat{P}_0 - P_0)^2 = \text{var } \hat{P}_0 + \text{bias}^2 \hat{P}_0 \quad (3.1c)$$

where, and in the rest of the thesis, the hat symbol over an variable denotes an estimate of the true variable, and P_0 represents the error rate to estimate. The MSE reflects the overall reliability, and the bias and variance quantify the accuracy (constant error) and consistency (variable error) respectively.

Another approach is to grade the error monitoring capability by the probability that the ER estimate falls within a given interval, i.e.,

$$\mathcal{C}(\hat{P}_0) = P\{\varepsilon_l \leq \hat{P}_0 \leq \varepsilon_u\} \quad (3.2)$$

where ε_l and ε_u are respectively the lower and upper tolerance limits, $[\varepsilon_l, \varepsilon_u]$ is called a tolerance interval for $\hat{P}_0$, and $\mathcal{C}$ is known as the percent coverage [40].

In light of the criteria of estimation reliability, there are different definitions for the term efficiency. In general, a point estimator is said to be more efficient than another if it can achieve higher reliability with a fixed observation time or use fewer observations to satisfy a given reliability requirement.

3.4 Real-Error Counting Monitor

The simplest and most accurate *a posteriori* method of estimating the ER is to compare the received signal to the original one and then enumerate the differences. Such a comparison is generally impossible since the original signal is unknown to the receiver. To allow the comparison to be made, it is necessary to perform system training or, when a return path exists, to send the received signal back to the source [38, 41].

In real-error counting monitoring, two types of events are defined for the observations: event ‘0’, corresponding to the transmission errors, is assumed to occur with constant probability P_0 ; event ‘1’, corresponding to the properly decoded symbols, is assumed to occur with constant probability $1 - P_0$. The number of event ‘0’ out of n observations, denoted as n_0 , is a binomial random variable with parameters n and P_0 , i.e., $n_0 \sim \mathcal{B}(n, P_0)$. The probability of detecting k event ‘0’ is given by

$$P(n_0 = k) = \binom{n}{k} P_0^k (1 - P_0)^{n-k} \quad (3.3)$$

and the first two moments of n_0 are respectively

$$\mathbb{E} n_0 = nP_0 \quad (3.4a)$$

$$\text{var } n_0 = nP_0(1 - P_0). \quad (3.4b)$$

The ER estimate, $\hat{P}_0$, is approximated as the rate of occurrence of event ‘0’ within the observation time, i.e.,

$$\hat{P}_0 = \frac{n_0}{n} \quad (3.5)$$

It follows directly from Eq. (3.4) that

$$\text{bias } \hat{P}_0 = 0 \quad (3.6a)$$

$$\text{var } \hat{P}_0 = \frac{1}{n} P_0 (1 - P_0) \approx \frac{1}{n} P_0 \quad (3.6b)$$

$$\text{MSE } \hat{P}_0 = \text{var } \hat{P}_0 \approx \frac{1}{n} P_0. \quad (3.6c)$$

As can be seen from Eq. (3.6a), $\hat{P}_0$ obtained in this way is asymptotically unbiased.

When the ER is very small, the real-error counting monitor (RECM) requires a prohibitively long observation time to fulfill the requirement of making reliable ER estimate. As a direct consequence, a significant portion of the bandwidth is wasted for transmitting the training sequences or the feedback sequences that do not carry any useful information for the users. Moreover, since the ER estimate is the result of averaging over the whole observation period, when a long observation is involved, the error monitor is not able to detect abrupt changes of the ER and thus fails to provide a prompt alert of the variation of the system performance.

As the development of technology has made telecommunications systems more robust against the transmission errors than ever before, this class of error monitoring methods is subsiding gradually. However, due to its simple implementation, it is still appealing for applications with no strict demand on real-time estimation, such as offline experiments and out-of-service system maintenance.

3.5 Kernel Real-Error Monitor

The RECM is extremely inefficient in the sense that only the observations that have been identified as transmission errors could participate in the calculation of the ER. To improve the data efficiency and thereby reduce the observation time, various strategies have been explored; among them, the most significant improvement is achieved by incorporating a nonparametric density estimator (Chapter 2). When a kernel density estimator is

CHAPTER 3. KERNEL REAL-ERROR MONITOR

---

employed, the resultant error monitor is known as a kernel real-error monitor (KREM). The effectiveness and efficiency of kernel real-error monitoring has already been confirmed by computer simulations in [1–4]. However, it remains unclear how the kernel real-error monitor can be tuned to achieve optimum performance. The objective of this chapter is to develop effective optimization approaches for kernel real-error monitoring.

3.5.1 Principle of Kernel Real-Error Monitoring

In general, a kernel real-error monitor for a M -ary communications system can be constructed as follows

$$\hat{P}_0 = \sum_{m=0}^{M-1} \left[\dot{P}_m \int_{\epsilon_m} \hat{f}_m(x_m) dx_m \right] \quad (3.7)$$

where $\dot{P}_m$ is the probability of occurrence of the m^{th} symbol, x_m denotes the corresponding decision statistic, ϵ_m represents the error region of x_m , and $\hat{f}_m$ denotes the kernel estimate of the PDF of x_m . It follows from Eq. (2.4) that

$$\hat{f}_m(x_m) = \frac{1}{n_m h} \sum_{i=1}^{n_m} K\left(\frac{x_m - X_{mi}}{h}\right) \quad (3.8)$$

where X_{mi} , $i = 1, \dots, n_m$, denotes the i^{th} observation of x_m , and $\sum_{m=0}^{M-1} n_m = n$.

Rather than relying on some special events as the real-error counting monitor does, the kernel real-error monitor, as shown in Eq. (3.7), uses all the observations to construct an ER estimate. The direct benefit of this high efficient data exploitation strategy is that the kernel real-error monitor can provide more ER estimate within a given observation period, or alternatively, it can cut down the observation cost greatly without compromising the monitoring capability.

The error region ϵ_m can be expressed in a general form as $(-\infty, r_{m1}) \cup (r_{m2}, \infty)$, where r_{m1} and r_{m2} , $-\infty \leq r_{m1} < r_{m2} \leq \infty$, are respectively the lower and upper thresholds of x_m . Using Eq. (3.8), we can rewrite Eq. (3.7) as follows

$$\hat{P}_0 = \sum_{m=0}^{M-1} n_m^{-1} \dot{P}_m \sum_{i=1}^{n_m} T(r_{m1}, r_{m2}, X_{mi}) \quad (3.9)$$

CHAPTER 3. KERNEL REAL-ERROR MONITOR

---

where $T(\cdot)$ denotes the area under a two-sided tail of the kernel, i.e.,

$$T(r_{m1}, r_{m2}, X_{mi}) = \int_{-\infty}^{\frac{1}{h}(r_{m1}-X_{mi})} K(x) dx + \int_{\frac{1}{h}(r_{m2}-X_{mi})}^{+\infty} K(x) dx_m. \quad (3.10)$$

For the choice of standard Gaussian kernel,

$$T(r_{m1}, r_{m2}, X_{mi}) = Q\left(\frac{X_{mi} - r_{m1}}{h}\right) + Q\left(\frac{r_{m2} - X_{mi}}{h}\right) \quad (3.11)$$

where Q denotes the Gaussian error function, i.e.,

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} e^{-t^2/2} dt. \quad (3.12)$$

Assume that the source symbols have equal probability of occurrence and n is sufficiently large such that $\dot{P}_m$ approaches n_m/n . Accordingly, Eq. (3.9) becomes

$$\hat{P}_0 = \frac{1}{n} \sum_{m=0}^{M-1} \sum_{i=1}^{n_m} T(r_{m1}, r_{m2}, X_{mi}) \quad (3.13)$$

Therefore, the kernel ER estimate can be interpreted as the mean coverage of n two-sided tails. Each tail is associated with a specific observation, and its coverage of indicates the probability that the observation could be an error.

As can be seen from Eq. (3.10), for a fixed choice of kernel and a given observation sequence, the coverage of a tail is solely determined by smoothing parameter h . Because the choice of the smoothing parameter is much more important than the choice of the kernel [5], and because the standard Gaussian kernel is so prevalent that it has almost become a real standard, the optimization of kernel ER estimation is essentially the problem of optimizing the smoothing effect. The smoothing parameter given in Eq. (2.8), which aims to minimize the MISE of the density estimate, although widely adopted in data analysis, does not fit well into the context of ER estimation. This is due to the fact that in error monitoring we are more concerned with the tail properties of the underlying

CHAPTER 3. KERNEL REAL-ERROR MONITOR

density but not the density itself. Accordingly, the MISE criterion defined by Eq. (2.6), which highlights the quality of the density estimate, loses its justification. This problem has been noted in [1] and [2], where a relatively subjective choice of $n^{-1/2}$ is recommended. Unfortunately, this parameter is often insufficient in smoothing the observations, leading to greater estimation variance. Therefore, there is a need to develop new smoothing optimization approaches for kernel ER estimation.

Note that, as h approaches zero, the tail area function Eq. (3.10) tends to a rectangular function, i.e.,

$$\lim_{h \rightarrow 0} T(r_1, r_2, X_i) = \begin{cases} 0 & \text{if } r_1 < X_i < r_2, \\ 0.5 & \text{if } X_i = r_1 \text{ or } r_2, \\ 1 & \text{if } X_i < r_1 \text{ or } X_i > r_2, \end{cases} \quad (3.14)$$

and a real-error monitor is obtained. Therefore, the RECM can be regarded as a special KREM with zero smoothing effect.

It should also be noted that the kernel ER estimation in Eq. (3.9) and (3.13) involves M separate ER estimations, for each symbol in the alphabet. This solution is inconvenient in design and analysis, considering that the M estimators need to be configured separately. In the following discussion, we will demonstrate how to simplify the ER estimation by constructing a new random variable, referred to as pseudo decision statistic. The pseudo decision statistic is sufficient for kernel ER estimation, unless we are concerned with some specific types of transmission errors.

3.5.2 Optimal Smoothing Effect

As a natural and reasonable measure of the quality of parameter estimation, the mean square error (MSE) of ER estimate (see Eq. (3.1c)) has been selected to guide the optimization of the smoothing effect. In this section, our discussion will be limited to BPSK

CHAPTER 3. KERNEL REAL-ERROR MONITOR

---

modulation. However, as will be illustrated in Section 3.5.3, the results obtained can be easily extended to other modulation schemes.

3.5.2.1 Statistical Properties

In BPSK, the error regions are defined by a unique threshold r ($r = 0$ usually). Accordingly, the kernel real-error monitor Eq. (3.13) with the choice of the standard Gaussian kernel becomes

$$\hat{P}_0 = \frac{1}{n} \left[\sum_{i=1}^{n_0} Q\left(\frac{X_{0i} - r}{h}\right) + \sum_{i=1}^{n_1} Q\left(\frac{r - X_{1i}}{h}\right) \right] \quad (3.15)$$

where $\{X_{0i}\}$ and $\{X_{1i}\}$ are respectively the observation sequences of symbols ‘0’ and ‘1’, and n_0 and n_1 are their sizes. We can construct a new sequence, called pseudo observation sequence, as follows

$$\{\tilde{X}_i\} = \{X_{1i}\} \cup \{2r - X_{0i}\}. \quad (3.16)$$

The pseudo decision statistic, herein, refers to the random variable associated with observations $\{\tilde{X}_i\}$, denoted as $\tilde{x}$. The density of the pseudo decision statistic, $f(\tilde{x})$, is equal to $f(x)$, provided that the observations are properly labeled and the noise is symmetrically distributed. The latter condition can be dropped, since the difference between $f(\tilde{x})$ and $f(x)$ caused by the asymmetry of the noise distribution has little effect on the ER estimation (see Eq. (3.13)).

Accordingly, Eq. (3.39) can be simplified to

$$\hat{P}_0 = \frac{1}{n} \sum_{i=1}^n Q\left(\frac{r - \tilde{X}_i}{h}\right). \quad (3.17)$$

Assume that $\{\tilde{X}_i\}$ are independent identically distributed and the number of observations, n , is sufficiently large. It follows that

$$\mathbb{E} \hat{P}_0 = \mathbb{E} Q(\cdot) = \int_{-\infty}^{+\infty} Q(\cdot) f(\tilde{x}) d\tilde{x} \quad (3.18a)$$

CHAPTER 3. KERNEL REAL-ERROR MONITOR

$$\text{var } \hat{P}_0 = \frac{1}{n} \text{var } Q(\cdot) = \frac{1}{n} [\text{E } Q^2(\cdot) - \text{E}^2 Q(\cdot)], \quad (3.18b)$$

where, for clear presentation, we have discarded the argument of $Q\left(\frac{r-\tilde{X}_i}{h}\right)$. Substituting $\tilde{x}$ with $r - hy$ in Eq. (3.18a) produces

$$\begin{aligned} \text{E } Q(\cdot) &= h \int_{-\infty}^{\infty} Q(y) f(r - hy) dy \\ &= h \int_0^{\infty} \{[1 - Q(y)] f(r + hy) + Q(y) f(r - hy)\} dy. \end{aligned}$$

Note that

$$P_0 = \int_r^{\infty} f(\tilde{x}) d\tilde{x} = h \int_0^{\infty} f(r + hy) dy.$$

The previous expression can thus be simplified to

$$\text{E } Q(\cdot) = P_0 - h \int_0^{\infty} Q(y) [f(r + hy) - f(r - hy)] dy.$$

Expanding $f(r + hy)$ and $f(r - hy)$ with the Taylor series, i.e.,

$$f(r \pm hy) = \sum_{k=0}^{\infty} \frac{1}{k!} (\pm hy)^k f^{(k)}(r)$$

and assuming that all terms of order higher than four are negligible, we obtain

$$\text{E } Q(\cdot) \approx P_0 - 2h^2 f'(r) \int_0^{\infty} y Q(y) dy - \frac{1}{3} h^4 f^{(3)}(r) \int_0^{\infty} y^3 Q(y) dy. \quad (3.19)$$

The variance of the kernel ER estimate can be derived similarly. Substituting $\tilde{x}$ with $r - hy$ in the first term of the right side of Eq. (3.18b) gives

$$\begin{aligned} \text{E } Q^2(\cdot) &= h \int_{-\infty}^{\infty} Q^2(y) f(r - hy) dy \\ &= P_0 + h \int_0^{\infty} Q^2(y) [f(r + hy) + f(r - hy)] dy \\ &\quad - 2h \int_0^{\infty} Q(y) f(r + hy) dy. \end{aligned}$$

CHAPTER 3. KERNEL REAL-ERROR MONITOR

Again, using the Taylor series expansions of $f(r + hy)$ and $f(r - hy)$, we get

$$\begin{aligned} \mathbb{E} Q^2(\cdot) \approx & P_0 - 2hf(r) \int_0^\infty Q(y)Q(-y) dy - 2h^2 f'(r) \int_0^\infty yQ(y) dy \\ & - h^3 f''(r) \int_0^\infty y^2 Q(y)Q(-y) dy - \frac{1}{3}h^4 f^{(3)}(r) \int_0^\infty y^3 Q(y) dy. \end{aligned} \quad (3.20)$$

Let $\Omega_i(k) \equiv \int_0^\infty y^k Q^i(y) dy$, $i = 1, 2$. It can be verified that for arbitrary nonnegative integer k

$$\Omega_1(k) = \frac{2^{k/2}}{(1+k)\sqrt{2\pi}} \Gamma(1 + \frac{k}{2}), \quad (3.21)$$

where Γ is the gamma function, i.e., $\Gamma(z) \equiv \int_0^\infty t^{z-1} e^{-t} dt$. For ease of use, rewrite Eq. (3.21) as follows

$$\Omega_1(k) = \begin{cases} \frac{2^{k/2}(k/2)!}{\sqrt{2\pi}(k+1)} & \text{if } k \text{ is even,} \\ \frac{\prod_{i=0}^{(k-1)/2} (2i+1)}{2(k+1)} & \text{if } k \text{ is odd.} \end{cases} \quad (3.22)$$

It can also be verified that

$$\Omega_2(k) = \begin{cases} \frac{\Omega_1(k)}{2} - \frac{\sum_{i=0}^{k/2} \left[2^{-i} \prod_{j=i+1}^{k/2} (2j) \prod_{l=1}^i (2l-1) \right]}{2(k+1)\sqrt{\pi}} & \text{if } k \text{ is even,} \\ \frac{\Omega_1(k)}{2} - \frac{\sum_{i=0}^{(k-1)/2} \left[i! \prod_{j=i+1}^{(k-1)/2} (2j+1) \right]}{2(k+1)\pi} & \text{if } k \text{ is odd,} \end{cases} \quad (3.23)$$

where, for ease of presentation, $\prod_{j=k/2+1}^{k/2} (2j)$ and $\prod_{l=1}^0 (2l-1)$ have been defined to be one.

Applying Eq. (3.22) and (3.23) to Eq. (3.19) and (3.20) produces

$$\mathbb{E} Q(\cdot) \approx P_0 - \frac{h^2}{2} f'(r) - \frac{h^4}{8} f^{(3)}(r), \quad (3.24a)$$

$$\begin{aligned} \mathbb{E} Q^2(\cdot) \approx & P_0 - \frac{h}{\sqrt{\pi n}} f(r) - \frac{h^2}{2n} f'(r) \\ & - \frac{5h^3}{12\sqrt{\pi n}} f''(r) - \frac{h^4}{8n} f^{(3)}(r), \end{aligned} \quad (3.24b)$$

CHAPTER 3. KERNEL REAL-ERROR MONITOR

---

from which it readily follows that the estimation bias and variance of the KREM are respectively

$$\text{bias } \hat{P}_0 \approx -\frac{h^2}{2}f'(r) - \frac{h^4}{8}f^{(3)}(r), \quad (3.25a)$$

$$\begin{aligned} \text{var } \hat{P}_0 \approx & \frac{1}{n}(P_0 - P_0^2) - \frac{h}{\sqrt{\pi n}}f(r) - \frac{h^2}{2n}f'(r) \\ & - \frac{5h^3}{12\sqrt{\pi n}}f''(r) - \frac{h^4}{8n}f^{(3)}(r). \end{aligned} \quad (3.25b)$$

Eq. (3.25a) suggests that the estimation bias of the KREM does not rely on the sample size; therefore, taking longer observation solely cannot reduce the bias. Note that, as h approaches zero, the bias and variance of kernel ER estimate tend to zero and $n^{-1}(P_0 - P_0^2)$ respectively, identical with those of the real-error counting monitor (Eq. (3.6)), verifying again that the latter can be categorized into kernel real-error monitor.

By forcing the most significant neglected terms in the right sides of Eq. (3.25) (not given in the equations) to be far smaller than the least significant survivals, it can be found that Eq. (3.25a) is a valid approximation of the bias of kernel ER estimate only if

$$h \ll \sqrt{6 \left| \frac{f^{(3)}(r)}{f^{(5)}(r)} \right|} \approx 2.45 \sqrt{\left| \frac{f^{(3)}(r)}{f^{(5)}(r)} \right|}, \quad (3.26)$$

and Eq. (3.25b) is a valid approximation of the estimation variance only if

$$h \ll \frac{60\sqrt{\pi}}{43} \left| \frac{f^{(3)}(r)}{f^{(4)}(r)} \right| \approx 2.47 \left| \frac{f^{(3)}(r)}{f^{(4)}(r)} \right|. \quad (3.27)$$

3.5.2.2 Optimum Smoothing Effect

Substituting the bias and variance of the kernel ER estimate, Eq. (3.25), into Eq. (3.1c) and forcing the derivative of $\text{MSE}(\hat{P}_0)$ with respect to h to be zero, we get a cubic function

$$h^3 + ah^2 + bh + c = 0 \quad (3.28)$$

CHAPTER 3. KERNEL REAL-ERROR MONITOR

---

where

$$\begin{aligned} a &= \frac{5\pi^{-0.5}f''(r)}{2f^{(3)}(r) - 4nf'(r)^2} \\ b &= \frac{2f'(r)}{f^{(3)}(r) - 2nf'(r)^2} \\ c &= \frac{2\pi^{-0.5}f(r)}{f^{(3)}(r) - 2nf'(r)^2}. \end{aligned}$$

The optimum smoothing parameter in the sense of minimizing the MSE can then be obtained by solving Eq. (3.28). Let

$$p = b - \frac{1}{3}a^2 \quad (3.29)$$

$$q = \frac{2}{27}a^3 - \frac{1}{3}ab + c \quad (3.30)$$

It is well known that [42], under the condition

$$D = (q/2)^2 + (p/3)^3 > 0, \quad (3.31)$$

Eq. (3.28) has a unique real solution

$$h_1 = -\frac{a}{3} + \sqrt[3]{D^{1/2} - \frac{q}{2}} + \sqrt[3]{-D^{1/2} - \frac{q}{2}}. \quad (3.32)$$

So far, we have derived an optimum smoothing parameter for minimizing the MSE of the ER estimate resulted from kernel real-error monitoring. It is valid whenever the standard Gaussian kernel applies. However, this smoothing parameter cannot be used directly since it depends on the density being estimated. A conventional solution to this problem is to assume a Gaussian statistical model as the reference. This solution seems to be more reasonable in the context of error monitoring, since the typical distributions of the decision statistics can be expected to exhibit more similarity to the Gaussian distribution

---

CHAPTER 3. KERNEL REAL-ERROR MONITOR

---

in their tails. It follows from Eq. (3.26) and (3.27) that, for a Gaussian distribution with mean μ and variance σ^2 , h should satisfy

$$h < \alpha \left| \frac{\sigma^2}{r - \mu} \right| \quad (3.33)$$

to guarantee the validity of the approximations in Eq. (3.25). Parameter α should be selected to be much smaller than 2.45, $\alpha = 0.8$ for instance. Usually, it is unnecessary to verify the smoothing parameter produced by Eq. (3.32). Eq. (3.33) nearly always holds when α is set to 0.8; even if it does fail, the resultant smoothing parameter can still produce an acceptable smoothing effect.

If h is sufficiently small such that all terms on the right side of Eq. (3.25b) of order higher than one can be neglected, Eq. (3.32) can be simplified to

$$h_1 \approx 2^{1/6} e^{(r-\mu)^2/6\sigma^2} |r - \mu|^{-2/3} \sigma^{5/3} n^{-1/3}. \quad (3.34)$$

This approximation tends to oversmooth the observations at a low ER or a small sample size and thus should be used cautiously.

3.5.2.3 σ^2 Estimator

As can be seen from Eq. (3.31) and (3.32), the resultant optimum smoothing parameter is a function of the sample size n and the variance of the Gaussian model σ^2 . Since σ^2 is generally unknown, it is necessary to derive it from the observed statistics. The simplest σ^2 estimator can be formed as the sample variance of $\tilde{x}$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (\tilde{X}_i - \hat{\mu})^2 \quad (3.35)$$

where $\hat{\mu}$ denotes the sample mean, i.e. $\hat{\mu} = n^{-1} \sum_{i=1}^n \tilde{X}_i$. In [5], it is suggested to use the interquartile range of the samples, i.e., the distance between the first and third quartiles,

CHAPTER 3. KERNEL REAL-ERROR MONITOR

as $\hat{\sigma}$. A detailed discussion on robust variance estimation in the context of nonparametric density approximation can be found in [43]. All these variance estimators attempt to precisely represent the underlying distribution. In another words, the resultant Gaussian model optimizes a certain measure of the overall quality of the density estimate (Gaussian model), e.g., the MISE, the likelihood, and etc. Since our concern in this study is the accuracy of the tail(s) that the model reproduces, it is more appropriate to use some local measure to optimize the variance estimate, e.g. the likelihood of the samples falling into the tail(s) of the distribution. Let us consider the Gaussian model of the pseudo decision statistic $\tilde{x}$. Denote the boundary of the left tail of $f(\tilde{x})$ as r_a , $r_a \geq r$, and without loss of generality, assume $\tilde{X}_i \in (-\infty, r_a)$, $i = 1, \dots, n_a$, where n_a is the number of observations within $(-\infty, r_a)$. The log-likelihood function of $\{\tilde{X}_i \mid \tilde{X}_i \in (-\infty, r_a)\}$ is given by

$$\sum_{i=1}^{n_a} \log \left\{ \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(\tilde{X}_i - \hat{\mu})^2}{2\sigma^2} \right] \right\},$$

where $\hat{\mu}$ still represents the sample mean. By forcing the derivative of the log-likelihood with respect to σ to be zero, it can be obtained that

$$\hat{\sigma}^2 = \frac{1}{n_a} \sum_{i=1}^{n_a} (\tilde{X}_i - \hat{\mu})^2. \quad (3.36)$$

In general, smaller r_a suggests more accurate tail representation but increases the risk of getting erratic result (i.e. more dynamic $\hat{\sigma}^2$). As indicated in [44], more than 20 erroneous events are needed to derive reliable ER estimate (tail area). Thus, we can select r_a to accommodate 20 samples in the range of $(-\infty, r_a)$.

Although it is feasible to adapt the smoothing effect to the time variation of $\hat{\sigma}^2$, we recommend to fix the smoothing parameter for the sake of interpretation and comparison of the resultant ER estimates. The smoothing parameter can be selected to optimize the estimate of the critical ER, for example, the ER that delimits acceptable and unacceptable system performances.

3.5.3 Kernel Real-Error Monitoring for M -ary Modulation

By selecting appropriate pseudo decision statistic, KREM can be readily applied to other modulation schemes (see also Section 3.5.1).

In general, a symbol-by-symbol detector attempts to minimize a Euclidean distance metric [29]

$$D(X_i) = \|X_i - \bar{X}_j\| \quad (3.37)$$

where $\bar{X}_j$, $j = 1, \dots, M$, denotes the expected value of the decision statistic x when the j^{th} symbol is transmitted. Accordingly, the pseudo decision statistic can be selected to be the deviation of an observation from its expected value. The pseudo observation sequence can then be constructed as follows

$$\{\tilde{X}_i\} = \{X_i - \bar{X}(X_i)\} \quad (3.38)$$

where $\bar{X}(X_i)$ represents the expected value of X_i .

Let r_1 and r_2 be the boundaries of the error region of $\tilde{x}$, $\epsilon_{\tilde{x}}$, that is, $\epsilon_{\tilde{x}} = (-\infty, r_1) \cup (r_2, \infty)$. Let $P_0(\tilde{x})$ be the ER of $\tilde{x}$, i.e., the probability of an observation of $\tilde{x}$ falling in $\epsilon_{\tilde{x}}$. It follows from Eq. (3.11) that

$$\hat{P}_0(\tilde{x}) = \frac{1}{n} \sum_{i=1}^n \left[Q\left(\frac{\tilde{X}_i - r_1}{h}\right) + Q\left(\frac{r_2 - \tilde{X}_i}{h}\right) \right] \quad (3.39)$$

M -ary PAM

In M -ary PAM, $\{X_i\}$ are the received symbols (the output of the demodulator), and the desired ER estimate is given by

$$\hat{P}_0 = \frac{M-1}{M} \hat{P}_0(\tilde{x}). \quad (3.40)$$

CHAPTER 3. KERNEL REAL-ERROR MONITOR

The scale factor $(M - 1)/M$ is necessary since the error regions of the leftmost and rightmost decision statistics, x_0 and x_{M-1} , are one-sided. The boundaries of $\epsilon_{\tilde{x}}$ are respectively

$$r_1 = -d \quad (3.41a)$$

$$r_2 = d \quad (3.41b)$$

where $2d$ is the minimum distance of the constellation (i.e., the distance between adjacent signal amplitudes).

M -ary PSK

For M -ary PSK ($M > 4$), it is more convenient to consider the phase detection, which is equivalent to the minimum distance detection in error rate performance [29]. In that case, the observation X_i is the phase of the complex received symbol $y_i = y_{Ri} + jy_{Ii}$, i.e.,

$$X_i = \tan^{-1} \frac{y_{Ri}}{y_{Ii}}. \quad (3.42)$$

The error region $\epsilon_{\tilde{x}}$ is now defined by

$$\epsilon_{\tilde{x}} = (-\pi, r_1) \cup (r_2, \pi) \quad (3.43)$$

where

$$r_1 = -\frac{\pi}{M} \quad (3.44a)$$

$$r_2 = \frac{\pi}{M}. \quad (3.44b)$$

Noting that the integral of the kernel over $(-\infty, \pi) \cup (\pi, \infty)$ is negligible, we can obtain

$$\hat{P}_0 \approx \hat{P}_0(\tilde{x}). \quad (3.45)$$

The ER estimation for BPSK is identical to that for BPAM. Based on the fact a QPSK signal is simply two BPSK signals in phase quadrature, the ER estimation for QPSK can be easily obtained.

CHAPTER 3. KERNEL REAL-ERROR MONITOR

 M -ary QAM

In M -ary QAM, the observations $\{X_i\}$ are the complex received symbols. For ease of presentation, we can construct the pseudo observation sequence as follows

$$\{\tilde{X}_i\} = \{\Re[X_i - \bar{X}(X_i)]\} \cup \{\Im[X_i - \bar{X}(X_i)]\}, \quad (3.46)$$

where $\Re[\cdot]$ and $\Im[\cdot]$ are respectively the real and imaginary parts of the argument. It is reminded that $\{\tilde{X}_i\}$ constructed in this way consists of $2n$ elements.

In the case of square signal constellation, the error region $\epsilon_{\tilde{x}}$ is the same as in M -ary PAM, and the ER estimate can be derived from $\hat{P}_0(\tilde{x})$ as follows (see [29])

$$\hat{P}_0 = \frac{\sqrt{M} - 1}{\sqrt{M}} \left\{ 1 - [1 - \hat{P}_0(\tilde{x})]^2 \right\} \approx \frac{2(\sqrt{M} - 1)}{\sqrt{M}} \hat{P}_0(\tilde{x}). \quad (3.47)$$

For non-square QAM, we can obtain the ER estimate similarly. However, the scale factor has to be determined specifically for each constellation. For non-rectangular constellations, the pseudo decision statistic, the pseudo observation sequence, and the error regions may also need to be designed specifically.

Now, we have arrived at the conclusion that the the ER estimate for an M -ary modulation can be expressed as the product of a constant (the scale factor) and the ER estimate of a single random variable (the pseudo decision statistic), except for the non-rectangular QAM. It follows that the results obtained in the previous section for BPSK modulation are also valid for M -ary modulation schemes.

It should be noted that, in the previous section, the pseudo observation sequence has been constructed in a different way, where the observations associated with different sides of the error region $\epsilon_{\tilde{x}}$, i.e., $(-\infty, r_1)$ and (r_2, ∞) , have been processed separately, so as to obtain a continuous error region. The two methods of constructing the pseudo observation sequence are essentially equivalent for ER estimation.

3.5.4 Blind Estimation Inability

Kernel real-error monitoring is very efficient in exploiting the received signal in the sense that all the observations are involved in the construction of the ER estimate. However, its performance in blind ER estimation is still unsatisfactory. Without the knowledge of the source signal, the kernel real-error monitor cannot properly distinguish the observations of different source symbols. Due to the classification (detection) error, the target of PDF approximation, i.e. the density of the pseudo decision statistic $\tilde{x}$, is not the true density $f(x)$ any more, but a distorted version with tail(s) folded, partially or fully², onto the main body (see Eq. (3.16), and note that $\{X_{0i}\}$ and $\{X_{1i}\}$ are now the observations that are *classified* as ‘0’ and ‘1’ respectively). The distortion is not a serious problem for PDF approximation because the tail area is negligible compared to the main body, but for ER estimation it can be a disaster.

Again, consider BPSK modulation. For clear presentation, let us denote the distorted PDF and the corresponding ER estimate as f_B and $\hat{P}_B$ respectively. The expected value of $\hat{P}_B$ is given by

$$\mathbb{E} \hat{P}_B = \int_{-\infty}^{\infty} Q\left(\frac{r - \tilde{x}}{h}\right) f_B(\tilde{x}) d\tilde{x}.$$

Similarly as in Section 3.5.2, it can be deduced that

$$\text{bias } \hat{P}_B \approx \frac{h f_B(r)}{\sqrt{2\pi}} - P_0 \quad (3.48a)$$

$$\text{var } \hat{P}_B \approx \frac{h f_B(r)}{\sqrt{2\pi n}}, \quad (3.48b)$$

where $f_B(r)$ is equal to $2f(r)$ under the assumption that symbols ‘0’ and ‘1’ are equiprobable and the tail is completely folded. By comparing the bias and variance of $\hat{P}_B$ to those

---

²“Partial folding” occurs when some error correction mechanism, e.g. FEC (forward error correction), is employed. In that case, certain information of the source is revealed to the receiver, and thus, the ER estimation should be regarded as non-blind estimation (more precisely, semi-blind).

CHAPTER 3. KERNEL REAL-ERROR MONITOR

Fig. 3.2. Inability of kernel real-error monitor in blind ER estimation. The shadowed areas represent the respective error probabilities.

of the nonblind estimate $\hat{P}_0$ (Eq. (3.25)), we can find that the blind ER estimate is distinctly inferior. As such, it is more reasonable to group kernel real-error monitoring into nonblind ER estimation techniques.

Fig. 3.2 illustrates the inability of kernel real-error monitor in blind estimation in a more straightforward manner, by comparing $\hat{f}_B(\tilde{x})$ and $\hat{P}_B$ to $f(x)$ and P_0 directly.

3.5.5 Computer Simulations

To demonstrate the influence of appropriate smoothing on the performance of kernel real-error monitoring, computer simulations have been conducted. Let us first consider BPSK signals transmitted over AWGN (additive white Gaussian noise) channel.

Fig. 3.3 compares the optimal KREM with the RECM by their normalized rms (root

CHAPTER 3. KERNEL REAL-ERROR MONITOR

Fig. 3.3. Normalized rms errors of the optimum kernel real-error monitor ($h = h_1$) and the real-error counting monitor ($h = 0$); $n = 5000$, and the results are derived from 1000 repetitions.

CHAPTER 3. KERNEL REAL-ERROR MONITOR

---

mean square) errors, i.e., $\sqrt{\text{MSE } \hat{P}_0}/P_0$. One may be disappointed to find that the use of optimal smoothing does not decrease the estimation error significantly. However, this is unnecessary since the rms error just depicts one aspect of error monitoring performance. The most attractive feature of optimal smoothing is the reduction of the minimum observation time to satisfy the consistency³ requirement of error monitoring. In Fig. 3.3, 5000 samples (observations) have been used to calculate the ER. Actually, such a sample size is far less than adequate for the real-error counting monitor to give useful ER estimate at low ER, as can be seen from Fig. 3.4, where the ER estimates produced by the two error monitors are directly compared. The large number of vacancies appeared in Fig. 3.4(b) signifies the frequent occurrence of trivial estimate ($\hat{P}_0 = 0$). Clearly, when the ER is below 10^{-3} , the method of counting the transmission errors within the observation time (5000 sampling intervals) cannot generate useful ER estimate for system monitoring. Fig. 3.5 compares the the percent coverage of the two estimators for a given tolerance interval $[0.2P_0, 5P_0]$. It follows the conclusion that undersmoothing the samples has a disastrous effect on the reliability of KREM in the sense of tolerance.

In Fig. 3.6, the smoothing parameter is purposely set to triple of the optimal value. As can be seen, oversmoothing the observations increases the estimation error significantly. Fig. 3.7 displays the ER estimates obtained with optimal and over-smoothing effect, from which one can easily observe that as a result of oversmoothing a positive bias is introduced into the ER estimate.

Fig. 3.8 illustrates the performance of the optimum KREM for 16-QAM. As expected, the KREM works effectively for M -ary modulation as well. It is reminded that, although the number of samples of the received signal is still $n = 5000$, we now have $2n$ observations

---

³Herein, “consistency” refers to the ability of the estimator to generate consistently usable estimate, i.e. the reproducibility mentioned in pp. 21.

CHAPTER 3. KERNEL REAL-ERROR MONITOR

(a)

(b)

Fig. 3.4. ER estimates produced by (a) the optimum kernel real-error monitor ($h = h_1$), and (b) the real-error counting monitor ($h = 0$). The dotted lines show the theoretical ERs. $n = 5000$.

CHAPTER 3. KERNEL REAL-ERROR MONITOR

Fig. 3.5. Percent coverage of the optimum kernel real-error monitor ($h = h_1$) and the real-error counting monitor ($h = 0$) for tolerance interval $[0.2P_0, 5P_0]$. $n = 5000$.

Fig. 3.6. Normalized rms errors of the optimum kernel real-error monitor ($h = h_1$) and the monitor with triple smoothing effect ($h = 3h_1$). $n = 5000$.

CHAPTER 3. KERNEL REAL-ERROR MONITOR

(a)

(b)

Fig. 3.7. ER estimates produced by (a) the optimum kernel real-error monitor ($h = h_1$), and (b) the monitor with triple smoothing effect ($h = 3h_1$). The dotted lines show the theoretical ERs. $n=5000$.

CHAPTER 3. KERNEL REAL-ERROR MONITOR

Fig. 3.8. ER estimates produced by the optimum kernel real-error monitor for 16-QAM. $n = 5000$.

of the pseudo decision statistic $\tilde{x}$ to estimate the ER (see Eq. (3.46)). Therefore, the ER estimates are more reliable than those for BPSK (compared with Fig. 3.7(a)).

Conclusively, appropriate smoothing is essential for the performance of KREM. Undersmoothing the observations leads to more accurate ER estimation, but degrades the consistency seriously; oversmoothing improves the consistency, but causes significant bias; the proposed optimum smoothing, by minimizing the overall estimation error, can achieve a balance between the accuracy and consistency. It should be mentioned that, in the cases where the observation time is extremely concerned and the absolute value of the ER estimate is relatively less important, it may be desirable to sacrifice a certain degree of accuracy for lower observation cost, and oversmoothing can be considered. Nevertheless, there is a better approach to further reduce the observation time, which will be introduced in the subsequent chapter.

3.6 Summary

The use of nonparametric density estimate to improve the efficiency and consistency of *a posteriori* ER estimation is not a new technique. There have been a number of simulation studies dedicated to this topic. However, the optimization of the density estimator for ER estimation has never been investigated. By minimizing the MSE of the ER estimate, we have developed a simple optimization approach for kernel real-error monitoring in this chapter. The resultant optimum kernel error monitor, as verified by computer simulations, can provide good performance when the transmitted signal is known *a priori*. However, its blind estimation capability is unsatisfactory. In addition, its data efficiency is inadequate for the estimation of a small ER. In the next chapter, we will see how to achieve efficient blind ER estimation by combining density estimation and ER extrapolation techniques.

Chapter 4

Kernel Pseudo-Error Monitor

4.1 Introduction

As mentioned in Chapter 3, the real-error counting approach discards all the information buried in the sample sequence other than the erroneous events in the calculation of the ER. Consequently, the observation time required to fulfill the consistency expectation is unbearable. The pseudo-error monitoring approach, instead, attempts to derive the ER from a set of pseudo-error rate (PER) estimates. Usually, a PER estimate is obtained by enumerating the occurrences of the pseudo errors, i.e., the events with relatively higher risk to be erroneous. Since the pseudo errors occur more frequently than the real errors, the pseudo-error monitoring approach is comparatively more efficient. In Chapter 3 we have already illustrated how the nonparametric density estimation technique can be exploited to enhance the efficiency of a real-error counting monitor. It's a reasonable conjecture that the nonparametric density estimation technique can be incorporated into a pseudo-error monitor to achieve higher data efficiency. This chapter introduces our effort in this direction.

4.2 Traditional Pseudo-Error Monitor

The technique of pseudo-error monitoring was developed in the 1960s to address the need for blind and efficient ER estimators to replace the awkward real-error counting monitor. Nowadays, it has become the most popular *a posteriori* ER estimation algorithm.

4.2.1 Principle of Pseudo-Error Monitoring

The essence of pseudo-error monitoring is to amplify the ER by relaxing the error criterion or degrading the quality of the received signal. In the former case, a pseudo-error region is defined around the threshold for each type of pseudo error. Once an event is observed to fall into this region, the corresponding pseudo-error counter is triggered. In the latter case, the received signal is purposely degraded by reducing the signal power, raising the noise level, introducing new source of interference, and etc. The resultant signal is subsequently fed into a decoder, whose output is compared to the original decoded sequence to identify the pseudo errors. No matter how the pseudo errors are generated, by counting the number of the pseudo errors, a PER estimate can be obtained, and from the extrapolation of a set of PER estimates, the desired ER estimate can be derived.

A common practice is to treat the logarithmic ER as a linear function of the quality of the received signal. The validity of this approximation has been discussed in, e.g., [45–48], and it has been pointed out that, in most cases, the linear approximation is sufficiently accurate for error monitoring [38]. As an example, in Fig. 4.1, we illustrate the relationship between the logarithmic ER and the SNR for BPSK transmission over an AWGN channel. As can be seen, the error of linear approximation is insignificant over a wide range of SNR.

CHAPTER 4. KERNEL PSEUDO-ERROR MONITOR

Fig. 4.1. Relationship between the logarithmic ER and the SNR for BPSK transmission over an AWGN channel

Accordingly, the ER estimate can be extrapolated from two PER estimates as follows

$$\log \hat{P}_0 = \frac{d_1 \log \hat{P}_{d2} - d_2 \log \hat{P}_{d1}}{d_1 - d_2} \quad (4.1)$$

where d_i , $i = 1, 2$, is the signal degradation factor, $\hat{P}_{di}$ denotes the corresponding PER estimate, and $\hat{P}_0$ denotes the desired ER estimate. The signal degradation factor is defined to be

$$d_i = 1 - \frac{\gamma_i}{\gamma_0} \quad (4.2)$$

where γ_0 and γ_i are respectively the SNRs per bit (E_b/N_0) before and after the received signal is degraded. Because relaxing the error criterion is in effect equivalent to introducing a certain amount of degradation, extrapolation formula Eq. (4.1) is generally valid. For a BPSK system that suffers from the AWGN, the equivalent degradation factor d_{rk}

CHAPTER 4. KERNEL PSEUDO-ERROR MONITOR

---

corresponding to the modified threshold r_k , $r_k > 0$, can be calculated as follows

$$d_{rk} = 1 - \left(\frac{\mu - r_k}{\mu - r_0} \right)^2, \quad (4.3)$$

where μ is the mean value of a decision statistic, r_0 denotes the original threshold and is usually set to zero. Higher order curve fitting can be considered if the linear extrapolation cannot fulfill the accuracy requirement [45]. This solution, however, is not recommended, since the improvement on estimation accuracy is achieved at the cost of degradation of the estimation consistency, while the latter factor is usually considered to be more important in error monitoring.

If the ratio of the PER to the ER is sufficiently large, the real transmission errors can be neglected in the calculation of the PER. As such, the pseudo-error monitor successfully alleviates the dependence on the *a priori* knowledge of the source signal.

4.2.2 Statistical Properties

The bias and variance of the logarithmic ER estimate given in Eq. (4.1) can be expressed as follows

$$\text{bias}(\log \hat{P}_0) = \text{E}(\log \hat{P}_0) - \log P_0 = \frac{d_1 \text{E}(\log \hat{P}_{d2}) - d_2 \text{E}(\log \hat{P}_{d1})}{d_1 - d_2} - \log P_0 \quad (4.4a)$$

$$\text{var}(\log \hat{P}_0) = \text{E}(\log^2 \hat{P}_0) - \text{E}^2(\log \hat{P}_0) = \frac{d_1^2 \text{var}(\log \hat{P}_{d2}) + d_2^2 \text{var}(\log \hat{P}_{d1})}{(d_1 - d_2)^2} \quad (4.4b)$$

Let X and Y be two random variables that satisfies $Y = \log X$. Assume that the PDF, mean value and variance of X are respectively f_x , m_x and σ_x^2 . Using the Taylor series of $\log x$ at m_x , i.e., $\log x = \log m_x - \log e \cdot \sum_{k=1}^{\infty} k^{-1} (1 - x/m_x)^k$, we obtain

$$\text{E} Y = \int_0^{\infty} \log x f_x(x) dx = \log m_x - \log e \sum_{k=1}^{\infty} k^{-1} \mu_{xk} (-m_x)^{-k} \quad (4.5)$$

CHAPTER 4. KERNEL PSEUDO-ERROR MONITOR

---

where μ_{xk} is the k^{th} central moment of X . For a Gaussian distribution,

$$\mu_{xk} = \begin{cases} \prod_{i=1}^{k/2} (2i-1) \cdot \sigma_x^k & \text{if } k \text{ is even,} \\ 0 & \text{if } k \text{ is odd.} \end{cases}$$

Accordingly, Eq. (4.5) becomes

$$\mathbb{E} Y = \log m_x - \frac{1}{2} \log e \sum_{k=1}^{\infty} k^{-1} \prod_{i=1}^k (2i-1) \cdot \sigma_x^{2k} m_x^{-2k}.$$

If $\sigma_x^2 \ll m_x^2$,¹ the higher order terms of σ_x^2 can be neglected, leading to

$$\mathbb{E} Y \approx \log m_x - \frac{\sigma_x^2}{2m_x^2} \log e. \quad (4.6)$$

Similarly, it can be deduced that when $\sigma_x^2 \ll m_x^2$

$$\mathbb{E} Y^2 \approx \log^2 m_x + \frac{\sigma_x^2}{m_x^2} \log e (\log e - \log m_x) \quad (4.7)$$

In traditional pseudo-error monitoring, the PER estimate $\hat{P}_d$ corresponding to signal degradation factor d can be interpreted as the mean of a set of independent and identically distributed random variables, i.e.,

$$\hat{P}_d = \frac{1}{n} \sum_{i=1}^n E_i$$

where n is the sample size and E_i , $i = 1, 2, \dots, n$, is equal to 1 or 0, with probabilities P_d and $1 - P_d$ respectively. The mean and variance of $\hat{P}_d$ are approximately

$$m_d \approx P_d \quad (4.8)$$

$$\sigma_d^2 \approx n^{-1} P_d (1 - P_d) \quad (4.9)$$

---

¹which suggests that the sample size n should be much greater than the reciprocal of the PER (see Eq. (4.9)), a condition that can be easily satisfied.

CHAPTER 4. KERNEL PSEUDO-ERROR MONITOR

Fig. 4.2. Normalized rms error of the logarithmic ER estimate in traditional pseudo-error monitoring ($d_1 = 0.4$, $\gamma_0 = 6$ dB, $n = 5000$).

Substituting Eq. (4.6), (4.7), (4.8) and (4.9) into Eq. (4.4) and assuming that $n \gg P_d^{-1}$, we arrive at

$$\text{bias}(\log \hat{P}_0) \approx \frac{d_1 \log P_{d2} - d_2 \log P_{d1}}{d_1 - d_2} - \log P_0 \quad (4.10a)$$

$$\text{var}(\log \hat{P}_0) \approx \frac{d_1^2 (P_{d2}^{-1} - 1) + d_2^2 (P_{d1}^{-1} - 1)}{n(d_1 - d_2)^2} \log^2 e. \quad (4.10b)$$

In Fig. 4.2, the normalized rms error of $\log \hat{P}_0$, i.e., $\sqrt{\text{MSE}(\log \hat{P}_0)} / \log \hat{P}_0$, is shown versus d_2 , where an AWGN channel is investigated and d_1 is fixed to be 0.4. It can be observed that certain degree of deviation from the optimal setting of the pseudo-error monitor does not severely degrade the quality of the ER estimate.

4.2.3 Traditional Pseudo-Error Monitors

A number of pseudo-error monitors have been proposed in literature, mainly identified by the methods of generating and detecting the pseudo errors. As mentioned previously, one feasible method is to simulate the transmissions occurred in worse-conditioned channels by either reducing the signal power or increasing the noise or interference level. Particular realizations of this method include the noise addition monitor [49], the filter imperfection monitor [50] and etc. Another important method is to relax the error criteria, and the resultant pseudo-error monitor is referred to as threshold modification monitor (TMM) [45]. Some ER estimators exploit the technique of automatic eye-pattern analysis [51, 52]. In a wide sense, they can be interpreted as special threshold modification monitors that adopt more complicated error classification systems.

In spite of the success in relaxing the dependence on the *a priori* knowledge of the transmitted signal, traditional pseudo-error monitoring does not show sufficient efficiency in exploiting the received signal. The PER estimates are obtained by enumerating each type of pseudo error occurred within the observation period. Clearly, to avoid the trivial result of zero, the sample size cannot be less than the reciprocal of the PER. In fact, it has been pointed out in [38, 44, 49] that at least ten pseudo errors are required to guarantee the consistency of the PER estimate. Such a sample size is prohibitive in practice, considering that modern communication systems normally work at high fidelity. This problem can be partially solved by increasing the level of degradation or further relaxing the error criteria to have a greater amplification of the ER. Unfortunately, since the discrepancy between the extrapolation formula and the actual relationship that underlies the extrapolation formula is unavoidably aggrandized, this solution may suffer from serious loss of estimation accuracy.

4.3 Kernel Pseudo-Error Monitor

Traditional pseudo-error monitoring approach exhibits excellent blind estimation capability, but it is inadequate in reducing the observation time; the kernel real-error monitoring approach exploits the received signal in a more efficient manner, but its performance in blind state is disappointing. The objective of kernel pseudo-error monitoring is to combine the strengths of the two approaches to generate fast and reliable blind ER estimate.

4.3.1 Principle of Kernel Pseudo-Error Monitoring

In kernel real-error monitoring, the ER estimate is derived from direct integration of the density approximate over the error region. Consequently, its performance is highly relied on the *a priori* knowledge of the transmitted signal (see Section 3.5.4). In kernel pseudo-error monitoring (KPEM), this problem is overcome by using the density approximate to calculate the PERs, which are then extrapolated as in traditional pseudo-error monitoring to form the desired ER estimate. Without loss of generality (see Section 3.5.3), we consider BPSK modulation.

KPEM Based on Threshold Modification

If the pseudo errors are generated by modifying the thresholds, the kernel pseudo-error monitor using linear extrapolation Eq. (4.1) can be expressed as follows

$$\log \hat{P}_0 = \frac{d_{r1} \log \hat{P}_{r2} - d_{r2} \log \hat{P}_{r1}}{d_{r1} - d_{r2}} \quad (4.11)$$

where d_{rk} , $k = 1, 2$, is the equivalent signal degradation factor corresponding to modified threshold r_k (Eq. (4.3)), and $\hat{P}_{rk}$ is the PER estimate obtained by integrating the kernel PDF approximate over the pseudo-error region.

CHAPTER 4. KERNEL PSEUDO-ERROR MONITOR

For BPSK modulation, the pseudo-error region can be written as $(-r_k, r_k)$. It follows that

$$\hat{P}_{rk} = \int_{-r_k}^{r_k} \hat{f}(x) dx \quad (4.12)$$

where

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right). \quad (4.13)$$

Note that in the above equation $\{X_i\}$ have been treated as the samples of a unique random variable, while in KREM, it is necessary to discriminate the observations of different source symbols. Substituting Eq. (4.13) into Eq. (4.12) yields

$$\hat{P}_{rk} = \frac{1}{n} \sum_{i=1}^n Q\left(\frac{X_i - r_k}{h}\right) - \frac{1}{n} \sum_{i=1}^n Q\left(\frac{X_i + r_k}{h}\right) \quad (4.14)$$

As can be seen, the PER estimates can be derived directly from the samples. Therefore, the explicit expression of the kernel PDF approximate is unnecessary. Fig. 4.3 depicts the structure of this kind of KPEM, where the kernel PDF estimators are shown for illustration only.

KPEM Based on Signal Degradation

If the pseudo errors are generated by degrading the received signal, the kernel pseudo-error monitor based on the linear extrapolation can be expressed in the form of Eq. (4.1). The PER estimates are now the integrals of a series of PDF approximates, corresponding to different signal degradation factors, over an identical error region. Once again, we have to discriminate the observations of the source symbols.

Consider the BPSK modulation. Let x_{k0} and x_{k1} be the decision statistics corresponding to the transmission of symbols '0' and '1' at degradation factor d_k . Clearly, the error regions of x_{k0} and x_{k1} remain to be $(-\infty, r_0)$ and (r_0, ∞) respectively. It follows that

$$\hat{P}_{dk} = \dot{P}_0 \int_{-\infty}^{r_0} \hat{f}_{k0}(x) dx + (1 - \dot{P}_0) \int_{r_0}^{\infty} \hat{f}_{k1}(x) dx \quad (4.15)$$

CHAPTER 4. KERNEL PSEUDO-ERROR MONITOR

Fig. 4.3. Block diagram of the kernel pseudo-error monitor based on threshold modification and linear extrapolation

CHAPTER 4. KERNEL PSEUDO-ERROR MONITOR

---

where $\dot{P}_0$ is the probability of occurrence of symbol ‘0’ and $\hat{f}_{kj}(x)$, $j = 0, 1$, is the PDF approximate of x_{kj} , i.e.,

$$\hat{f}_{k0}(x) = \frac{1}{n_0 h} \sum_{i=1}^{n_0} K\left(\frac{x - X_{0i}^k}{h}\right) \quad (4.16a)$$

$$\hat{f}_{k1}(x) = \frac{1}{n_1 h} \sum_{i=1}^{n_1} K\left(\frac{x - X_{1i}^k}{h}\right) \quad (4.16b)$$

where X_{ji}^k , $i = 1, \dots, n_j$, denotes an observation of x_{kj} . Under the assumption that n is sufficiently large such that n_0/n accurately approximates $\dot{P}_0$, by substituting Eq. (4.16) into Eq. (4.15), we obtain

$$\hat{P}_{dk} = \frac{1}{n} \left[\sum_{i=1}^{n_0} Q\left(\frac{X_{0i}^k - r_0}{h}\right) + \sum_{i=1}^{n_1} Q\left(\frac{r_0 - X_{1i}^k}{h}\right) \right]. \quad (4.17)$$

Alternatively, the same as in Section 3.5.2, we can construct a pseudo observation sequence as follows

$$\{\tilde{X}_i^k\} = \{X_{1i}^k\} \cup \{2r_0 - X_{0i}^k\}. \quad (4.18)$$

The PER estimate $\hat{P}_{dk}$ can then be calculated as the integral of the kernel PDF approximate of $\tilde{x}$ over $\{r_0, \infty\}$,

$$\hat{P}_{dk} = \frac{1}{n} \sum_{i=1}^n Q\left(\frac{r_0 - \tilde{X}_i^k}{h}\right). \quad (4.19)$$

Fig. 4.4 illustrates the structure of the KP EM based on signal degradation and linear extrapolation, where the kernel density estimators are used to approximate the PDFs of the pseudo decision statistics and, same as previous, they are presented for illustration only. As can be seen, compared to that shown in Fig. 4.3, this error monitor comes with an increase in implementation complexity. The rest of the chapter will focus on the scheme based on threshold modification.

CHAPTER 4. KERNEL PSEUDO-ERROR MONITOR

Fig. 4.4. Block diagram of the kernel pseudo-error monitor based on signal degradation and linear extrapolation

4.3.2 Properties and Optimization

Combining the results presented in Section 3.5.2 and 4.2.2, we can obtain the statistical properties of the kernel pseudo-error monitor Eq. (4.11)

$$\text{bias}(\log \hat{P}_0) \approx \text{bias}_0(\log \hat{P}_0) - \frac{d_{r1}P_{r2}^{-1}f'(r_2) - d_{r2}P_{r1}^{-1}f'(r_1)}{2(d_{r1} - d_{r2})} \log e \cdot h^2 \quad (4.20a)$$

$$\text{var}(\log \hat{P}_0) \approx \text{var}_0(\log \hat{P}_0) - \frac{d_{r1}^2 f(r_2) P_{r2}^{-2} + d_{r2}^2 f(r_1) P_{r1}^{-2}}{\sqrt{\pi} n (d_{r1} - d_{r2})^2} \log^2 e \cdot h \quad (4.20b)$$

where bias_0 and var_0 denote respectively the estimation bias and variance of the traditional threshold modification monitor using the same pseudo error classification system, i.e., the results of replacing d_i with d_{ri} , $i = 1, 2$, in Eq. (4.4). Usually, it is unnecessary to design separate smoothing processes for each PER estimator, since, as shown by Eq. (4.20), the quality of the ER estimate is ultimately determined by the quality of the estimate of the smaller PER. Therefore, we can implement a shared smoothing process and optimize it to favor the estimate of the smaller PER.

4.3.3 Performance Advantages

The error counting estimation method maps the ER domain, $[0, 1]$, to a set of discrete values, $\{k/n, k = 0, 1, \dots, n\}$, where k is the number of the errors. Apparently, in this solution, the sample size n must be far greater than the reciprocal of the ER to avoid trivial results of zero. As mentioned before, more than ten errors should be recorded to enable a reliable ER estimation, which places a very high demand on the observation time at a low ER. The traditional pseudo-error monitors rely on the error counting method to estimate the PERs (by counting the pseudo errors), and accordingly inherit its disadvantages as well. Although the introduction of ER extrapolation helps to relax the requirement for long observation time, it is still inadequate at extremely low ERs. Consider a BPSK

CHAPTER 4. KERNEL PSEUDO-ERROR MONITOR

---

system solely corrupted by the AWGN and assume that the SNR per bit is 12 dB, corresponding to $P_0 = 9.0 \times 10^{-9}$. Let 0.1 be the modified threshold, corresponding to an ER amplification factor of 22.4. It can be easily verified that at least 5.0×10^7 samples are needed. Even if a wider pseudo-error region is defined to have an ER amplification factor as large as 1000, the pseudo-error counting approach still needs about 1.1×10^6 samples to produce acceptable PER estimate. The KREM, on the contrary, maps the ERs to a continuous range of values, $[P_c, 1 - P_c]$, where $P_c \equiv Q(|r - \mu|h^{-1})$ is the ER of clean signal (followed from Eq. (3.17)). Theoretically, a kernel error monitor always produces nontrivial estimate, regardless the observation time and the ER. This attractive feature is inherited by the KPEM and makes it distinctively more competitive than traditional designs in fast-varying channels.

In addition, by mapping the infinite ER domain to a finite number of values, the error counting approach and hence the traditional pseudo-error monitoring unavoidably results in ER ambiguity, i.e., the inability to discriminate closely-spaced ERs. The minimum ER distance that can be resolved is n^{-1} . This problem is, at least theoretically, obviated from the methods based PDF approximation, in which, one-to-one mappings are built up between the ERs and their estimates.

The superiority of the KPEM is also evident by its flexibility in adjusting the behavior of the monitor. Since the objective of error monitoring is to provide a reliable indicator of the system performance, the consistency of the ER estimate is usually considered to be more important than the absolute value of the ER itself. In traditional pseudo-error monitoring, other than increasing the observation time, the only method of improving the estimation consistency is to define wider pseudo-error regions, or equivalently, to add in larger amounts of signal degradation. As has been mentioned earlier, this solution may introduce unbearable bias, and in some cases, may even lead to misjudgement of the

CHAPTER 4. KERNEL PSEUDO-ERROR MONITOR

---

system performance. In kernel pseudo-error monitoring, however, a better consistency is the immediate outcome of choosing a larger smoothing parameter. Although also suffers certain loss of accuracy, this solution is advantageous in that the orders of the ER estimates will not be altered, i.e., lower ERs are still mapped to smaller values and vice versa. Consequently, the increment of the estimation bias will not show a significant effect on the final evaluation of system performance. Moreover, the definition of a narrower PER can help to reduce the error introduced by ER extrapolation.

4.3.4 Computer Simulations

In Fig. 4.5 and 4.6, the KPEM is compared to the KREM and the TMM by, respectively, their ER estimation errors and the percent coverage for a given tolerance interval, where the modified thresholds are set to 0.2 and 0.4, the KREM is optimized to minimize the MSE of $\hat{P}_0$, the KPEM is optimized to minimize the MSE of $\hat{P}_{r_1}$ (letting $r_1 < r_2$), the TMM and the KPEM work in blind estimation mode while the KREM works in nonblind mode. As can be observed, the proposed KPEM approach successfully combines the advantages of the KREM and the TMM designs to achieve better blind estimation performance. The superiority of the proposed KPEM over the TMM is more clearly illustrated in Fig. 4.7, where the settings of the two monitors remain unchanged and the ER estimates produced by them are displayed directly. It can be seen that the KPEM is more efficient than the TMM in using the information carried by the received signal.

The proposed ER estimation approach can also be applied to non-Gaussian noise. Fig. 4.8 gives an example, where the noise distribution is assumed to be an equal mixture of two Gaussian components with the same variance σ^2 and different means $\pm d$, corresponding to the case where the transmitted signal is corrupted by the AWGN and an uncorrelated

CHAPTER 4. KERNEL PSEUDO-ERROR MONITOR

Fig. 4.5. Normalized rms errors of the threshold modification monitor and the kernel real- and pseudo-error monitors. The results are derived from 1000 repetitions. $n = 5000$.

CHAPTER 4. KERNEL PSEUDO-ERROR MONITOR

Fig. 4.6. Percent coverage of the threshold modification monitor and the kernel real- and pseudo-error monitors for tolerance interval $[0.2P_0, 5P_0]$. The results are derived from 1000 repetitions. $n = 2000$.

CHAPTER 4. KERNEL PSEUDO-ERROR MONITOR

(a)

(b)

Fig. 4.7. ER estimates produced by (a) the optimum kernel pseudo-error monitor, and (b) the threshold modification monitor. The dotted lines show the theoretical ERs. $n = 5000$.

CHAPTER 4. KERNEL PSEUDO-ERROR MONITOR

Fig. 4.8. ER estimates produced by the optimum kernel pseudo-error monitor in non-Gaussian environment. The dotted lines show the theoretical ERs. $n = 2000$.

Fig. 4.9. ER estimates produced by the optimum kernel pseudo-error monitor in the presence of ISI. The dotted lines show the theoretical ERs. $n = 5000$.

CHAPTER 4. KERNEL PSEUDO-ERROR MONITOR

---

interference (i.e., co-channel interference). The variance σ^2 is set to 0.0315 and the distance between the two components is changed so as to illustrate the performance of the KPEM at various ER values.

Fig. 4.9 shows the performance of the KPEM in the presence of ISI, where the channel impulse response is assumed to be $[f_1, f_2]^T$, f_1 is fixed at 1, and f_2 varies from 0 to 1.2 with a step of 0.1. The noise variance is set to 0.0443, corresponding to $P_0 = 10^{-6}$ when $f_2 = 0$. The modified thresholds are still 0.2 and 0.4. As illustrated, the KPEM works satisfactorily as long as the eye diagram of the received signal is open, i.e., $f_1 > f_2$ in this case. However, when the eye diagram is totally closed, the KPEM cannot provide reliable ER estimate any more. This is due to the fact that the KPEM exploits only the observed statistics, from which it is unable to determine whether or not the eye diagram is open. For example, two channel impulse responses, $[1, 1.2]^T$ and $[1.2, 1]^T$, make no difference for the KPEM (and any blind ER estimator based on the observed statistics). Fortunately, in that case (when the ER is very large), accurate ER estimation is of little interest.

4.4 Summary

This chapter presents a novel solution to *a posteriori* ER estimation, the kernel pseudo-error monitoring method. The proposed approach is a hybrid of the conventional pseudo-error monitoring and the nonparametric density estimation techniques. By extrapolating the ER from a set of PER estimates, the proposed KPEM can avoid the dependence on the *a priori* knowledge of the transmitted signal experienced by the KREM. By using the nonparametric density approximate of the decision statistic to calculate the PER estimates, the KPEM can achieve higher data efficiency and thus better estimation reliability than conventional pseudo-error monitors. Computer simulations confirm that the performance of the proposed KPEM is superior over the KREM and the conventional pseudo-error monitor.

Chapter 5

Origin Analysis Equalizer

5.1 Introduction

As an effective method of estimating an unknown statistical distribution based on a set of observations, the density estimation technique has the potential to be applied in applications where the knowledge of the distribution is of interest. In the previous two chapters, we have shown the benefit of using this technique in error monitoring. In this chapter, we will see its application in *channel equalization*.

In general, communications signals suffer some degree of distortion during transmission. In high speed data communications, a significant portion of the signal distortion is ascribed to the presence of the ISI (Inter-Symbol Interference). This chapter presents a novel statistical approach to ISI mitigation. This approach attempts to adapt the decision rule in accordance with the time-varying channel characteristics. The channel characteristics are learnt through a process termed “*origin analysis*”, in which the density estimation technique plays a critical role.

5.2 Overview of Channel Equalization

The past decade has witnessed a fast growing interest in deploying high speed data communications systems for broadband multimedia services. One major concern in the design and implementation of such systems is the presence of ISI, which can severely distort the transmitted signal, leading to significant degradation in BER performance. The ISI effect can be mitigated through a process termed *channel equalization*.

The Maximum Likelihood Sequence Estimation (MLSE) approach, implemented commonly with the Viterbi algorithm [53, 54], is known as the optimum solution (in BER performance) to ISI mitigation [29, 55, 56]. However, this approach has a computational complexity that grows exponentially with the span of the ISI and the size of the signal constellation, and is therefore unattractive for practical applications.

In order to reduce the computational complexity, various suboptimum approaches have been developed in the literature. A common suboptimum equalization strategy is to reverse the process through which the signal is contaminated and thus to recover the original signal waveform. Particular realizations of this strategy include Linear Equalization (LE), Decision-Feedback Equalization (DFE), etc. An LE attempts to use a linear filter (usually a FIR) to eliminate the ISI components [57–59], while a DFE uses a nonlinear process to mitigate the ISI effect [60–67]. Both approaches have a computational complexity linearly proportional to the span of the ISI. However, in some cases, the loss of ISI mitigation ability is so severe that a satisfactory BER performance cannot be maintained [29]. Other suboptimum equalizers based on reduced-complexity sequence estimation have also been studied in the literature [68–71].

Equalization algorithms operate by processing a set of parameters, normally the Channel Impulse Response (CIR), to mitigate the ISI. These parameters are generally unknown

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

---

and often time-varying in nature (e.g. in wireless mobile systems). Channel estimation offers a solution for estimating the channel parameters based on the channel output. Depending on whether or not a training sequence is used, channel estimation techniques can be broadly classified into two categories: non-blind and blind. In non-blind estimation, the problem of CIR estimation is equivalent to the design of a filter that best matches the given input (training sequence or pilot symbols) and the corresponding output (received signal) [72–75]. Non-blind estimation is effective when the channel characteristics remain relatively unchanged over time. However, for fast varying channels, the effective signal bandwidth can be greatly reduced because of the need to transmit training sequences frequently. Blind solutions, on the other hand, try to use the received signal alone to estimate the channel [72, 76–82]. Blind schemes are bandwidth efficient. However, their convergence is in general not guaranteed especially when they are applied to single-input single-output (SISO) systems [72]. Furthermore, for ill-conditioned channels such as those with zeros on the unit circle, most blind methods have limited performances [29]. The *expectation-maximization* (EM) algorithm has been used in the development of a stochastic optimization process for blind channel estimation. Using a reasonable initialization guess, the EM algorithm is almost always convergent [79]. Unfortunately, when the ISI span is large, the high computational cost can make this approach impractical.

A channel equalizer is said to be adaptive if it has the ability to automatically adjust its design parameters according to the time-varying the channel characteristics. A wide variety of adaptive equalization techniques have been proposed in the literature [29, 30, 83–93]. Instead of using a separate structure, the channel estimation function is often incorporated into the design of adaptive equalizers, where a joint estimation of the CIR and the source symbol sequence is performed. For a SISO system, this kind of technique is

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

---

generally more effective and efficient (than separating channel estimation from equalization). Nevertheless, an adaptive equalizer suffers from not only the equalization errors but also channel estimation errors. There is always a dilemma between having a good BER performance and a low computational complexity.

A novel approach to ISI mitigation has been developed in this study. Instead of adopting the conventional strategy of recovering the transmitted signal directly, the proposed approach, termed *origin analysis equalization*, attempts to mitigate the ISI by adapting the decision rule in accordance with the channel characteristics. The remaining parts of this chapter will demonstrate the feasibility of this new equalization approach.

5.3 Principle of Origin Analysis Equalization

A frequency-selective fading channel can be modeled as a time-varying filter with impulse response $\mathbf{f} \equiv [f_i]^T$, $i = 0, 1, \dots, L-1$, where the superscript ‘ T ’ denotes transpose and L is the number of resolvable multipath components. For simplicity, we only consider BPSK modulation. The k^{th} sample at the output of the coherent demodulator, $r(k)$, can now be expressed as follows

$$r(k) = \sum_{i=0}^{L-1} f_i a(k-i) + n(k) \quad (5.1)$$

where $a(\cdot)$, the transmitted symbol (the signal constellation point for source symbol $s(\cdot)$), is either 1 or -1, and $n(\cdot)$, the noise component, is assumed to have a Gaussian statistics with zero mean and variance σ_n^2 . It is clear that the signal component

$$x(k) \equiv \sum_{i=0}^{L-1} f_i a(k-i) \quad (5.2)$$

has 2^L possible values and each corresponds to a particular *state* of L -bit sequence $\mathbf{s}(k) \equiv [s(k), s(k-1), \dots, s(k-L+1)]$. Denote these values in descending order as $[x_i]^T$, i.e.,

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

$x_0 \geq x_1 \geq \dots \geq x_{2^L-1}$, and label the *states* of the L -bit sequence with $[\mathbf{s}_i]^T$ according to their distances to the all-zero *state* $\mathbf{s}_0 \equiv [0, \dots, 0, 0]$, i.e., $\mathbf{s}_1 \equiv [0, \dots, 0, 1]$, $\mathbf{s}_2 \equiv [0, \dots, 1, 0]$ and so forth. The density of $r(k)$ can thus be written as

$$p\{r(k)\} = \sum_{i=0}^{2^L-1} \frac{P_i(k)}{\sqrt{2\pi}\sigma_n} \exp\left\{-\frac{[r(k) - x_i(k)]^2}{2\sigma_n^2}\right\} \quad (5.3)$$

where $P_i(k)$ is the probability of $x(k)$ taking the i^{th} possible value.

Assume that the source symbol $s(k)$ is equally probable to be 0 or 1 and the CIR changes at a rate much slower than that of the baseband transmitted signal (i.e., slow fading). It follows that $P_i(k)$ approaches 2^{-L} as the sample size increases. $p\{r(k)\}$ can thus be interpreted as the mean density of 2^L Gaussian distributions with identical spreads and different centers $[x_i]^T$. These 2^L distributions are referred to as *sample origins* and are denoted as $[\mathbf{o}_i]^T$. It should be noted that, if the previous *state*, $\mathbf{s}(k-1)$, is already known, only two *states* can occur at the k^{th} sampling instance, termed Concurrent State Pair (CSP). Correspondingly, given the previous *state*, there are only two possible *origins* for the current sample. These two *origins* are separated by $|2f_0|$ and are referred to as a Concurrent Origin Pair (COP). Fig. 5.1 illustrates these notions with a dual-path transmission model.

For a given $\mathbf{s}(k-1)$, we can express the conditional probability density of $r(k)$ as follows

$$p\{r(k) \mid \mathbf{s}(k-1)\} = \frac{1}{2\sqrt{2\pi}\sigma_n} \sum_{a(k)=\pm 1} \exp\left\{-\frac{[r(k) - T(k) - f_0 a(k)]^2}{2\sigma_n^2}\right\} \quad (5.4)$$

where $T(k) \equiv \sum_{i=1}^{L-1} f_i a(k-i)$ is the midpoint of the COP.

As suggested by Eq. (5.1), the ISI can be mitigated by applying a fixed decision rule after the multipath components have been removed from the received signal. This is the underlying principle of the family of equalization approaches that use some forms of

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

Fig. 5.1. Notions of Concurrent Origin Pair (COP) and Concurrent State Pair (CSP).

(a) The PDF of $r(k)$ in a channel with impulse response $[1, 0.4]^T$. (b) The corresponding state transition diagram.

Fig. 5.2. Block diagram of *origin analysis equalizer*

inverse filtering, e.g., the LE, the DFE, and their derivatives. Eq. (5.4), however, gives clues for a different approach. It can be easily understood that given the previous *state* $\mathbf{s}(k-1)$ the maximum likelihood (ML) decision of $s(k)$ can be made as follows: if $f_0 > 0$

$$\hat{s}(k) = \begin{cases} 0 & \text{if } r(k) > T(k), \\ 1 & \text{if } r(k) < T(k); \end{cases} \quad (5.5)$$

and if $f_0 < 0$

$$\hat{s}(k) = \begin{cases} 0 & \text{if } r(k) < T(k), \\ 1 & \text{if } r(k) > T(k). \end{cases} \quad (5.6)$$

Hence, a ML solution can be obtained by adapting the decision rule in accordance with the time-varying parameter $T(k)$. This idea has led to the birth of the Origin Analysis Equalization (OAE) approach to be introduced in this chapter.

Fig. 5.2 shows the main structure of an OAE. As can be seen, the *origin analysis* block consists of two processes: an *origin localization* process for generating real-time

estimate for $[x_i]^T$, and an *origin identification* process for building up the correspondence between $[x_i]^T$ and $[\mathbf{s}_i]^T$. The decision logic evaluates $T(k)$ based on the results of the *origin analysis* and then implements Eq. (5.5) or (5.6) to decode the sample sequence.

5.4 Origin Localization

Since multipath effect is generally unknown and time-varying, direct determination of $[x_i]^T$ is often infeasible. The purpose of *origin localization* is to derive $[x_i]$ from the observed channel statistics. This section presents a nonparametric localization approach, which is based on nonparametric density estimation. The approach consists of three consecutive steps: nonparametric PDF estimation, peak extraction, and peak verification.

5.4.1 Nonparametric PDF Estimation

The first step of nonparametric *origin localization* is to estimate the PDF of the sample distribution. As has been introduced in Chapter 2, a general nonparametric PDF estimator can be expressed as follows

$$\hat{p}(r) = \frac{1}{nh} \sum_{k=1}^n W\left[\frac{r - r(k)}{h}\right] \quad (5.7)$$

where n represents the sample size, $W(\cdot)$ is the smoothing function, and h is the smoothing parameter that is used to fine-tune the smoothing effect. A variety of nonparametric PDF estimators have been proposed in the literature, and most of them are based on the idea of sample smoothing (see Chapter 2). The naive PDF estimator that implements the rectangular smoothing function Eq. (2.3) has been selected for this study, for its low computational complexity and relatively simple implementation. It should be noted, however, that the naive method is inefficient as it requires a relatively long sample sequence

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

---

in order to generate a good PDF estimate. In the following discussions, the sample size is assumed to be sufficiently large for naive density estimation.

For a given smoothing function, the value of h in Eq. (5.7) can be suitably selected to have a good balance between the resolution of the PDF estimator and its robustness against spurious. It is well known that a large h is more effective in suppressing spurious. However, an inappropriately large h can obscure essential features and result in insufficient compensation to ISI. In this study, h is specified by the expected number of samples falling within the coverage of a peak bin (i.e., the bin centered at a theoretical density peak), denoted as n_p , i.e.,

$$h = \frac{n_p}{np_p} \quad (5.8)$$

where p_p is the theoretical peak density of an *origin*, i.e., $p_p = 1/(2^L \sqrt{2\pi}\sigma_n)$.

5.4.2 Peak Extraction

The next step is to extract the peaks of the PDF approximate $\hat{p}(r)$. A simple scheme is used in our method, where an observation window is slid along the r axis and the location of the local density maximum is determined continuously. If the location occurs at the center of the window, then it is taken to be the location of the peak. The width of the observation window, w , should be selected carefully so that the real peaks can be extracted. One way of selecting w value is

$$w = \kappa d_{\min} \quad (5.9)$$

where $d_{\min} \equiv 2\sigma_n$ is the minimum distance between distinguishable *origins* and $\kappa \in (0, 1)$ is introduced to reduce the chance of true peaks being merged. Unless the sample size is too small, a value of 0.6 for κ works fine. Alternatively, the density peaks can be identified by searching along the derivative of the smoothed density estimate for sign changes. However, this solution can let in too many false peaks, and consequently, make peak verification a harder task.

5.4.3 Peak Verification

The peaks extracted may contain false peaks due to inappropriate setting of h and w . Peak verification deals with this problem. Complete removal of peak redundancy can sometimes be achieved by merging adjacent peaks of separation less than $d_{\min}$ apart. In some cases, however, false peaks occur in the tails of the density. In such cases, it is necessary to verify the peak values and remove those below a certain level.

On the other hand, peak overlap can occur if the minimum requirement for distinguishable *origins* is not satisfied, i.e.,

$$d_i > d_{\min}, \quad i = 0, 1, \dots, 2^L - 2 \quad (5.10)$$

fails at some d_i , where $d_i \equiv x_i - x_{i+1}$ is the distance between adjacent peaks. In that case, an additional process, peak decomposition, has to be incorporated. The peak overlaps can be identified by observing the peak densities. It can be deduced that a peak with density above $2e^{-0.5}p_p$ is highly likely to be an overlap. However, this method does not reveal the number of overlapped peaks. An indirect but more reliable method is to identify *origin* overlaps. The density valleys can be localized in a similar way as that for the density peaks. Let $M (\leq 2^L)$ be the number of distinguishable *origins*, $\{\hat{v}_i \mid 0 \leq i \leq M, \hat{v}_0 = \infty, \hat{v}_M = -\infty\}$ be the estimated valley locations (see Fig. 5.1), and n_i be the number of samples falling within $(\hat{v}_{i+1}, \hat{v}_i)$. The number of *origins* making up the i^{th} distinguishable *origin*, n_{ci} , can then be determined as

$$n_{ci} = \mathcal{I}\left(\frac{2^L n_i}{n}\right) \quad (5.11)$$

where $\mathcal{I}(\cdot)$ denotes the operation of rounding to the nearest integer. The peak decomposition process is applied whenever $n_{ci} > 1$. This process involves replacing an overlapped peak with n_{ci} component peaks where the later have equal density p_p and different offsets

---

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

$\Delta x_{ij}, j = 0, 1, \dots, n_{ci} - 1$. By assuming that the n_{ci} component peaks are equally spaced, we can approximate Δx_{ij} as

$$\Delta x_{ij} \approx \frac{\hat{d}_{vvi} - 2\hat{d}_{pv0}}{2} - j \frac{\hat{d}_{vvi} - 2\hat{d}_{pv0}}{n_{ci} - 1} \quad (5.12)$$

where $\hat{d}_{vvi} = \hat{v}_i - \hat{v}_{i+1}$ and $\hat{d}_{pv0} = \hat{x}_0 - \hat{v}_1$.¹ Accurate evaluation of Δx_{ij} is difficult, and meaningless as the estimated peak locations are themselves inaccurate. In fact, we can equate it to an arbitrary small value, e.g. zero, without needing to worry too much about the negative effect on the system performance.

5.4.4 L and σ_n^2 Estimation

In previous discussions, L , the number of resolvable multipath components, has been assumed to be known *a priori*. In practice, however, this parameter is often unknown. Consequently, we need a method to estimate L based on the observations made at the receiver.

As can be seen from Fig. 5.1, the left portion of the leftmost *origin* ($\mathbf{o}_{2L-1}$) and the right portion of the rightmost *origin* ($\mathbf{o}_0$) suffer the least from *origin* superposition. They can thus be approximately regarded as being shaped by the noise component alone. By processing the sample sequence in the way described in Section 5.4.1 and 5.4.2, and with large predefined h and w , we can obtain the rough locations of $\mathbf{o}_0$ and $\mathbf{o}_{2L-1}$, denoted as $\tilde{x}_0$ and $\tilde{x}_{2L-1}$ respectively. Let n_l and n_r be the numbers of observations falling within $(-\infty, \tilde{x}_{2L-1})$ and $(\tilde{x}_0, \infty)$ respectively. An L estimator can then be formed as follows

$$\hat{L} = \mathcal{I} \left(\log_2 \frac{n}{n_l + n_r} \right). \quad (5.13)$$

---

¹It is worthy to note that $\hat{d}_{pv0}$ is an estimate of f_{L-1} .

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

As mentioned earlier, the noise variance, once known, can help suppress the peak redundancy caused by inappropriate selection of h and w . Let $\{r_{li} \mid i = 1, \dots, n_l\}$ and $\{r_{ri} \mid i = 1, \dots, n_r\}$ denote the observations falling within $(-\infty, \tilde{x}_{2L-1})$ and $(\tilde{x}_0, \infty)$ respectively, and let $\{\tilde{r}_{oi}\}$ be the union of $\{r_{li} - \tilde{x}_{2L-1}\}$ and $\{r_{ri} - \tilde{x}_0\}$. A rough σ_n^2 estimate can be derived from $\{\tilde{r}_{oi}\}$ as follows

$$\hat{\sigma}_n^2 = \sum_{i=1}^{n_l+n_r} \tilde{r}_{oi}^2. \quad (5.14)$$

Both the L and the σ_n^2 estimators described above use only a small fraction of the observations $((n_l + n_r)/n)$. Consequently, a relatively long observation time is required for the estimators to produce accurate estimates. Fortunately, high accuracy is not needed here— L is a discrete quantity and σ_n^2 is used to select appropriate values for h and w only. Note that we can still localize the *origins* even without the knowledge of the noise variance, although with more difficulty.

5.4.5 Localization Examples

Fig. 5.3 illustrates the *origin localization* process. Under the condition that the channel vector $\mathbf{f}$ is $[1, 0.5, 0.25]^T$ and the SNR (i.e., $\sum_{i=0}^{L-1} |f_i|^2/N_0$) is 13 dB, the ideal PDF of the channel statistics is as depicted in Fig. 5.3(a), where the dotted lines are the *origins*. Using the method described in Section 5.4.4 gives $\hat{\sigma}_n^2 = 0.029$. Setting h to 8.5×10^{-3} (forcing n_p to be 5 in Eq. (5.8)), we get the PDF estimate in Fig. 5.3(b). Fig. 5.3(c) shows the extracted density peaks, where w has been set to 0.20 (Eq. (5.9)). As a result of properly selecting the parameters, the identified peak sequence contains no redundant members (Eq. (5.10) is always true) and thus peak verification can be skipped.

Fig. 5.4 illustrates the localization process when peak overlaps occur. The channel vector $\mathbf{f}$ is now assumed to be $[0.4, 1, 0.5]^T$ and the SNR remains at 13 dB. The subfigures,

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

Fig. 5.3. *Origin localization* for a channel with impulse response $[1, 0.5, 0.25]^T$, based on naive PDF estimation. (a) Ideal PDF. (b) Naive PDF estimate. (c) Extracted density peaks.

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

Fig. 5.4. *Origin localization* for a channel with impulse response $[0.4, 1, 0.5]^T$, based on naive PDF estimation. (a) Ideal PDF. (b) Naive PDF estimate. (c) Extracted density peaks. (d) Verified density peaks.

from the top down, are respectively, the ideal density function, the naive estimate derived from 2000 samples, the extracted peaks, and the outcome of peak verification. It can be observed that the three highest peaks shown in Fig. 5.4(c) are overlaps of two components (derived from Eq. (5.11)), and they have been decomposed into pairs of smaller peaks through the peak verification process, as shown in Fig. 5.4(d).

5.4.6 Kernel Localization

The implementation of advanced nonparametric PDF estimators can help to significantly increase the accuracy of *origin localization*. For example, peak extraction (Section 5.4.2) can be performed using the kernel PDF approximate (see Chapter 2) rather than the naive estimator. Since the channel has been assumed to be Gaussian, the smoothing parameter given in Eq. (2.8) can be applied to give

$$h_0 \approx 1.06\sigma_n (2^L/n)^{0.2}. \quad (5.15)$$

In order to reduce the computational load, we can decimate the sample sequence before calculating the kernel PDF approximate. Fig. 5.5 compares the localization accuracy of the nonparametric approaches based on kernel and naive PDF approximations. The localization accuracy is calculated as the normalized rms error of the *origin* location estimates, i.e.

$$\sqrt{\frac{\sum_{i=0}^{2^L-1} (\hat{x}_i - x_i)^2}{\sum_{i=0}^{2^L-1} x_i^2}}. \quad (5.16)$$

The channel is assumed to be the same as that considered in Fig. 5.3. Fig. 5.5 shows that the kernel-based approach can significantly reduce the estimation error.

Fig. 5.5. Accuracy of nonparametric *origin* localizers

5.4.7 Limitations

In this section, a nonparametric *origin localization* approach has been presented. This approach is computationally efficient as it involves only logical and basic arithmetic operations, such as addition and subtraction. However, as the approach relies on the shape of the observed distribution for *origin localization*, it has several inherent limitations. Firstly, as the noise level increases, accurate localization becomes more difficult. When the noise level is so high that the *origins* are no longer “visible”, the underlying distribution tends to become unimodal and the nonparametric approach will fail. Secondly, the noise distribution has been assumed to be Gaussian, and the presence of other interferences in practice, e.g. co-channel interference and adjacent channel interference, can easily render the assumed Gaussian model invalid. Consequently, the performance of the nonparametric localizer will degrade too. Thirdly, for systems employing multi-level

modulation schemes, the number of *origins* will grow exponentially, and a much longer sample sequence will be needed. Although advanced PDF estimator such as the kernel estimator can be used to relax the requirement for large sample size, the dependence on the shape of the distribution remains. To overcome these problems, more robust and efficient localization algorithms are needed. This is the research focus of the next chapter.

5.5 Origin Identification

In general, the correspondences between the *origins* and the *sequence states* are not known *a priori*. This is due to the fact that a particular channel PDF such as that shown in Fig. 5.3(a) can represent up to $2^L \times L!$ ² different channel impulse responses. It is therefore necessary to perform *origin identification*.

5.5.1 Non-Blind Identification

A straightforward method of identifying the *origins* is to use training sequences. Clearly, the *state* with the maximum likelihood to be the counterpart of *origin* $\mathbf{o}_i$ can be found as follows

$$\mathbf{s}(\mathbf{o}_i) = \arg \min_{\mathbf{s}_j} |\tilde{x}(\mathbf{s}_j) - x_i|, \quad j = 0, \dots, 2^L - 1 \quad (5.17)$$

where $\tilde{x}(\mathbf{s}_i)$ is the mean of the samples with *state* $\mathbf{s}_i$. By repeatedly sending each possible L -bit sequence and evaluating the 2^L sample means, the whole mapping system can be constructed. The main problem of this solution lies in its inability to track time variation of the channel characteristics. Even a subtle change of the transmission conditions, e.g.,

---

²which is the total number of permutations of the sequences $\{\tilde{f}_i \mid i = 0, \dots, L-1\}$, where $\tilde{f}_i$ is equal to f_i or $-f_i$.

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

---

fade-in or fade-out, can completely invalidate the previous mapping system constructed. As a result, the detector has to be retrained frequently, especially in relatively fast-varying channels, resulting in a serious waste of bandwidth. Nevertheless, the training-based solution serves as a good means to initiate the detector because of its simplicity and reliability.

5.5.2 Blind Identification

To avoid the wastage of bandwidth, a blind identification method is proposed in this section. Since the *state-origin* correspondence system is essentially the reflection of the correlation of continuous observations made at the receiver, it is reasonable to conjecture that the *origins* can be identified by exploiting the temporal information carried by the samples (order of occurrence).

5.5.2.1 Identifiability

Let us begin with an investigation of *rational state sequences*. Each *state* sequence can be treated as a scheme of *origin* arrangement. It should be noted that not all schemes are feasible. For instance, it is impossible to design a communication system to realize $[\mathbf{s}_1, \mathbf{s}_0, \mathbf{s}_2, \mathbf{s}_3]$. A *state* sequence is said to be rational if it corresponds to a feasible way of *origin* arrangement. The following theorem gives necessary conditions for a *state* sequence to be rational.

Theorem 5.1. *An ordered state sequence of dimension 2^L is rational only if $\forall i \in \{0, \dots, 2^L - 1\}$, $l \in \{0, \dots, L - 1\}$, and $j, k \in \Omega_l$*

$$\Omega_l = \{0, \dots, 2^l - 1\} + 2^{l+1}\{0, \dots, 2^{L-l-1} - 1\},$$

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

---

both of the following equations

$$\phi(i) + \phi(2^L - 1 - i) = 2^L - 1 \quad (5.18)$$

$$\text{sign}[\phi(j) - \phi(2^l + j)] = \text{sign}[\phi(k) - \phi(2^l + k)] \quad (5.19)$$

can be satisfied, where $\phi(i) \in \{0, \dots, 2^L - 1\}$ represents the order of occurrence of state $\mathbf{s}_i$ in this sequence, $\text{sign}(\cdot)$ denotes the sign of the argument.

Proof: Determining the rationality of a *state* sequence is equivalent to evaluating the solvability of the following equation system

$$\mathbf{A}\mathbf{f} = \tilde{\mathbf{x}} \quad (5.20)$$

where $\mathbf{f}$ denotes the channel coefficient vector, $\tilde{\mathbf{x}} = [x_{\phi(0)}, x_{\phi(1)}, \dots, x_{\phi(2^L-1)}]^T$, and $\mathbf{A} = [\mathbf{a}_0^T, \mathbf{a}_1^T, \dots, \mathbf{a}_{2^L-1}^T]^T$, $\mathbf{a}_i$ is the modulated signal vector corresponding to $\mathbf{s}_i$. It is well known that this equation system is solvable if and only if

$$\text{rank}(\mathbf{A}) = \text{rank}([\mathbf{A}, \tilde{\mathbf{x}}]). \quad (5.21)$$

For clear presentation, let $\bar{i}$ denote $2^L - 1 - i$. Note that

$$\mathbf{a}_i + \mathbf{a}_{\bar{i}} = \mathbf{0}_{1 \times L} \quad (5.22)$$

holds $\forall i \in \{0, 1, \dots, 2^L - 1\}$, where $\oplus$ denotes bitwise exclusive-or, and $\mathbf{0}_{1 \times L}$ represents the $1 \times L$ zero vector. Thus, when Eq. (5.21) is satisfied, we have

$$x_{\phi(i)} + x_{\phi(\bar{i})} = 0. \quad (5.23)$$

Making the substitution of $x_{\phi(i)} = -x_{\phi(\bar{i})}$ in the above equation, we can easily arrive at Eq. (5.18). It should also be noted that $\forall l \in \{0, 1, \dots, L - 1\}$ and $j, k \in \Omega_l$,

$$\mathbf{a}_j - \mathbf{a}_{2^l+j} = \mathbf{a}_k - \mathbf{a}_{2^l+k}. \quad (5.24)$$

Accordingly, we have

$$x_{\phi(j)} - x_{\phi(2^l+j)} = x_{\phi(k)} - x_{\phi(2^l+k)} \quad (5.25)$$

Eq. (5.19) then follows. $\square$

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

Let $\mathbf{P} \equiv [p_{ij}]$, $i, j = 0, 1, \dots, 2^L - 1$, be the transition matrix of ordered *state* sequence $[\mathbf{s}_k]$, $k = 0, 1, \dots, 2^L - 1$, i.e., p_{ij} is equal to 1 if the transition from $\mathbf{s}_i$ to $\mathbf{s}_j$, written as $\mathbf{s}_i \rightarrow \mathbf{s}_j$, is possible and 0 otherwise. Note that $\mathbf{s}_i \rightarrow \mathbf{s}_j$ can only occur if the oldest $L - 1$ bits of $\mathbf{s}_j$ exactly reproduces the latest $L - 1$ bits of $\mathbf{s}_i$, i.e.,

$$\mathbf{s}_i \rightarrow \mathbf{s}_j \quad \text{if } j = (i \gg 1) \text{ or } 2^{L-1} + (i \gg 1) \quad (5.26)$$

where $i \gg k$ denotes the right shift of i by k bits. Therefore, $\mathbf{P}$ is solely determined by L , the number of resolvable paths. Hereinafter, $\mathbf{P}$ will be referred to as the elementary transition matrix. Let $\mathbf{Q} \equiv [q_{ij}]$, be the *origin* transition matrix, i.e., q_{ij} is equal to 1 if the transition from $\mathbf{o}_i$ to $\mathbf{o}_j$ is possible and 0 otherwise. Since the *origin* transition matrix can be interpreted as the transition matrix of permuted *state* sequence $[\mathbf{s}(\mathbf{o}_k)]$, where $\mathbf{s}(\mathbf{o}_k)$ denotes the *state* corresponding to the k^{th} *origin* $\mathbf{o}_k$, we have the following theorem:

Theorem 5.2. *Given the elementary state and the origin transition matrices of an arbitrary communication system, $\mathbf{P}$ and $\mathbf{Q}$ respectively, there exists at least one orthogonal, symmetric and centrosymmetric transform matrix $\mathbf{T}$ that satisfies*

$$\mathbf{Q} = \mathbf{TPT}. \quad (5.27)$$

Proof: It is well known that a permutation like $[\mathbf{s}(\mathbf{o}_k)]$ can be derived from the ordered sequence $[\mathbf{s}_k]$ through a series of element exchanges. In addition, the exchange of the i^{th} and j^{th} elements of a *state* sequence is equivalent to the exchanges of the i^{th} and j^{th} rows and columns on the transition matrix of that sequence. Thus, $\mathbf{Q}$ can be obtained from $\mathbf{P}$ as follows

$$\mathbf{Q} = \mathbf{E}_l \mathbf{E}_{l-1} \cdots \mathbf{E}_1 \mathbf{P} \mathbf{E}_1 \cdots \mathbf{E}_{l-1} \mathbf{E}_l = \mathbf{TPT} \quad (5.28)$$

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

---

wherein $\mathbf{E}_k$, $k \in \{1, 2, \dots, l\}$, is the matrix representation of a row or column exchange. Clearly, transform matrix $\mathbf{T}$ is orthogonal and symmetric. Note that the *state* sequence $[\mathbf{s}(\mathbf{o}_k)]$ has a practical embodiment only if $\mathbf{s}(\mathbf{o}_i)$ and $\mathbf{s}(\mathbf{o}_{\bar{i}})$ are bitwise distinct for arbitrary $i \in \{0, 1, \dots, 2^L - 1\}$, i.e.,

$$\mathbf{s}(\mathbf{o}_i) \oplus \mathbf{s}(\mathbf{o}_{\bar{i}}) = \mathbf{s}_{2^L-1}$$

where $\bar{i} \equiv |2^L - 1 - i|$. Thus, the (i, j) row or column exchange operation, if involved in Eq. (5.28), must be succeeded by an $(\bar{i}, \bar{j})$ row or column exchange. The centrosymmetry of $\mathbf{T}$ follows. $\square$

From the definition of the *origin* transition matrix and Theorem 5.2, it can be easily derived that the *origin* transition matrix of an arbitrary communication system with L transmission paths has the following properties.

Property 5.1. *The rank of the origin transition matrix is 2^{L-1} , and each row and column has a duplicate in the matrix.*

Property 5.2. *The origin transition matrix is centrosymmetric, i.e., $q_{ij} = q_{\bar{i}\bar{j}}$ is true for arbitrary $i, j \in \{0, 1, \dots, 2^L - 1\}$.*

Property 5.3. *There are two nonzero elements in each row and column of the origin transition matrix. The L -bit sequences associated with them are differentiated by the first bit in the row direction and the last in the column direction, that is, if $q_{ij} = q_{ik} = 1$ and $\mathbf{s}(\mathbf{o}_j) = \mathbf{s}_u$, $\mathbf{s}(\mathbf{o}_k)$ is either $\mathbf{s}_{u+2^L-1}$ or $\mathbf{s}_{u-2^L-1}$; if $q_{ij} = q_{kj} = 1$ and $\mathbf{s}(\mathbf{o}_i) = \mathbf{s}_u$, $\mathbf{s}(\mathbf{o}_k)$ is either $\mathbf{s}_{u+1}$ or $\mathbf{s}_{u-1}$.*

Property 5.4. *There are two nonzero elements on the main diagonal of the origin transition matrix, and they correspond to $\mathbf{s}_0$ and $\mathbf{s}_{2^L-1}$ respectively.*

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

Since each *rational state sequence* corresponds to a practical scheme of *origin* arrangement, Theorem 5.2 and the above properties apply to the transition matrix of an arbitrary rational sequence. It should also be noted that the mapping from the *state* sequences to the *state* transition matrices is non-injective. For example, $[\mathbf{s}_0, \mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3]$ and $[\mathbf{s}_3, \mathbf{s}_2, \mathbf{s}_1, \mathbf{s}_0]$ have the same transition matrix

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}.$$

Theorem 5.3. *Two state sequences of equal length 2^L have the same state transition matrix if and only if for arbitrary $i \in \{0, 1, \dots, 2^L - 1\}$*

$$\mathbf{s}_i^{(1)} \oplus \mathbf{s}_i^{(2)} = \mathbf{s}_{2^L-1}, \quad (5.29)$$

where $\mathbf{s}_i^{(1)}$ and $\mathbf{s}_i^{(2)}$ represent the i^{th} elements of the first and second sequences respectively.

Proof: The sufficiency of Eq. (5.29) follows immediately from the fact that if transition $\mathbf{s}_i \rightarrow \mathbf{s}_j$ is possible, transition $\mathbf{s}_{\bar{i}} \rightarrow \mathbf{s}_{\bar{j}}$ is also possible.

To demonstrate the necessity of Eq. (5.29), let us assume $\mathbf{s}_{\phi_1(k)}^{(1)} = \mathbf{s}_k$, $k = 0, \dots, 2^L - 1$, where $\phi_1(k)$ is the order of occurrence of *state* $\mathbf{s}_k$ in sequence $[\mathbf{s}_k^{(1)}]$. From Property 5.4 we know that $\mathbf{s}_{\phi_1(0)}^{(2)}$ is either $\mathbf{s}_0$ or $\mathbf{s}_{2^L-1}$. Let us consider the latter case first, i.e., force $\mathbf{s}_{\phi_1(0)}^{(2)}$ to be $\mathbf{s}_{2^L-1}$. Since both $\mathbf{s}_0$ and $\mathbf{s}_{2^L-1}$ can only transit to, other than themselves, $\mathbf{s}_{2^L-1}$ and $\mathbf{s}_{2^L-1-1}$ respectively, we have $\mathbf{s}_{\phi_1(2^L-1)}^{(2)} = \mathbf{s}_{2^L-1-1}$. As shown in Eq. (5.26), the pairs of *states* associated with the transitions out of *states* $\mathbf{s}_{2^L-1}$ and $\mathbf{s}_{2^L-1-1}$ are $\{\mathbf{s}_{2^L-2}, \mathbf{s}_{2^L-1+2^L-2}\}$ and $\{\mathbf{s}_{2^L-2-1}, \mathbf{s}_{2^L-1+2^L-2-1}\}$ respectively. From Theorem 5.1 we know $\mathbf{s}_{\phi_1(2^L-2)}^{(2)} = \mathbf{s}_{2^L-2-1}$ and $\mathbf{s}_{\phi_1(2^L-1+2^L-2)}^{(2)} = \mathbf{s}_{2^L-1+2^L-2-1}$. Repeating this procedure sequentially leads to the conclusion that Eq. (5.29) is true for arbitrary $i \in \{0, 1, \dots, 2^L - 1\}$. In the case of $\mathbf{s}_{\phi_1(0)}^{(2)} = \mathbf{s}_{2^L-1}$, it can be easily verified that a duplicate of $[\mathbf{s}_i^{(1)}]$ is obtained at the end. $\square$

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

This study is concerned only with *rational state sequences*. Combining Theorem 5.1 and 5.3, we arrive at the following corollary.

Corollary 5.1. *Two rational state sequences of equal length have the same state transition matrix if and only if they are inverse to each other.*

In accordance with Corollary 5.1, the transform matrix $\mathbf{T}$ that satisfies Eq. (5.27) is not unique. This problem is considered in the following theorem.

Theorem 5.4. *Given the elementary state and the origin transition matrices of an arbitrary communication system, $\mathbf{P}$ and $\mathbf{Q}$ respectively, two qualified transform matrices can be found and they are related as follows*

$$\mathbf{T}_1 \mathbf{T}_2 = \tilde{\mathbf{I}} \quad (5.30)$$

where $\mathbf{T}_1$ and $\mathbf{T}_2$ denote the transform matrices and $\tilde{\mathbf{I}}$ is a matrix with ‘1’ on the secondary diagonal and ‘0’ elsewhere.

Proof: It follows from Theorem 5.2 that

$$\mathbf{T}_1 \mathbf{T}_2 \mathbf{P} \mathbf{T}_2 \mathbf{T}_1 = \mathbf{P}.$$

This suggests that, through the series of element interchanges that are associated with transform matrix $\mathbf{T}_1 \mathbf{T}_2$, a new *state* sequence is obtained and it has the same *state* transition matrix with $[\mathbf{s}_i]$. From Corollary 5.1 we know that the new sequence is the inverse of $[\mathbf{s}_i]$. The assertion follows. $\square$

The temporal information contained in the *origin* transition matrix is insufficient to guarantee the uniqueness of the configuration of the *state-origin* mapping system. Fortunately, this ambiguity can be overcome with knowledge of an arbitrary *state-origin* correspondence.

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

Theorem 5.5. *Given an arbitrary state-origin correspondence, the state-origin mapping system is uniquely determined by the origin transition matrix.*

Proof: We know from Corollary 5.1 that only two *rational state sequences* can be found to match a given *origin* transition matrix and they are the inverse of each other. It follows that the two sequences are different at each position. Therefore, once an arbitrary *origin* has been identified, the two sequences can be distinguished and the *state-origin* mapping system is uniquely determined. $\square$

Clearly, given an arbitrary *state-origin* correspondence, the qualified transform matrix $\mathbf{T}$ is also uniquely determined.

In general, the training-based method introduced in Section 5.5.1 can be used to construct a *state-origin* correspondence. However, in many cases, this is unnecessary.

Corollary 5.2. *The state-origin mapping system is uniquely determined by the origin transition matrix under a loose assumption that*

$$\sum_{i=0}^{L-1} f_i > 0. \quad (5.31)$$

Proof: Using $\mathbf{a}_0 \mathbf{f} = -\mathbf{a}_{2^L-1} \mathbf{f} = \sum_{i=0}^{L-1} f_i$ in Eq. (5.31) yields $x(\mathbf{a}_0) = \mathbf{a}_0 \mathbf{f} > \mathbf{a}_{2^L-1} \mathbf{f} = x(\mathbf{a}_{2^L-1})$. Recalling Property 5.4, we know that the nonzero element on the main diagonal of the *origin* transition matrix with the smaller index corresponds to *state* $\mathbf{s}_0$ and the other one corresponds to $\mathbf{s}_{2^L-1}$. The assertion then follows from Theorem 5.5. $\square$

With the theorems, properties and corollaries discussed above, we have illustrated the feasibility of identifying the *origins* based on the temporal correlation of continuous samples, represented by the *origin* transition matrix. We also see that *origin identification* can be conducted in three steps: identification of an arbitrary *origin*, estimation of the *origin* transition matrix, and determination of the configuration of the *state-origin* mapping system. In the subsequent discussions, we assume that Eq. (5.31) holds. Accordingly, the first step can be skipped.

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

5.5.2.2 Estimation of the Origin Transition Matrix

To obtain an estimate of the *origin* transition matrix, we need to set up an *origin* transition record first. A rough decision of the *origin* of $r(k)$ can be made as follows

$$\hat{\mathbf{o}}[r(k)] = \mathbf{o}_i \quad \text{if } r(k) \in (z_{i+1}, z_i) \quad (5.32)$$

where

$$z_i = \begin{cases} \infty & \text{if } i = 0, \\ \frac{1}{2}(\hat{x}_{i-1} + \hat{x}_i) & \text{if } i = 1, \dots, 2^L - 1, \\ -\infty & \text{if } i = 2^L. \end{cases} \quad (5.33)$$

Enumerating the occurrences of each type of one-step *origin* transition within the observation period produces the *origin* transition record. Expressing it as a matrix and formalizing the outcome with respect to Property 5.3, i.e., setting the two most significant entries of each row and column to one and the rest to zero, yields the desired estimate of the *origin* transition matrix $\mathbf{Q}$.

5.5.2.3 Determination of the State-Origin Correspondences

Once the *origin* transition matrix $\mathbf{Q}$ is known, the *state-origin* mapping system can be easily derived from $\mathbf{P}$ and $\mathbf{Q}$.

It is difficult to obtain the mapping system by solving Eq. (5.27) directly. This is because the solution of this equation is not unique (although only one is qualified). Let $\mathbf{A}$ be an arbitrary nonsingular matrix that commutes with $\mathbf{P}$. Clearly, if $\mathbf{T}$ satisfies Eq. (5.27), so does $\mathbf{TA}$. Fortunately, starting from the available *state-origin* correspondence and applying Property 5.3 Eq. (5.26) and Theorem 5.1 repeatedly, we can reveal all the *state-origin* correspondences successively.

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

Step 1. Without loss of generality, let $\mathbf{o}_i$ be the *origin* associated with $\mathbf{s}_0$, i.e., $\mathbf{s}(\mathbf{o}_i) = \mathbf{s}_0$, and let q_{ij} be the other nonzero element of the i^{th} row of the *origin* transition matrix $\mathbf{Q}$. From Property 5.3 we know that $\mathbf{s}(\mathbf{o}_j) = \mathbf{s}_{2^{L-1}}$.

Step 2. Let q_{jm} and q_{jn} , $m < n$, be the nonzero elements of the j^{th} row of matrix $\mathbf{Q}$, i.e., $\mathbf{o}_j \rightarrow \{\mathbf{o}_m, \mathbf{o}_n\}$. From Eq. (5.26) we know that $\mathbf{s}_{2^{L-1}} \rightarrow \{\mathbf{s}_{2^{L-2}}, \mathbf{s}_{2^{L-1}+2^{L-2}}\}$. In addition, it follows from Theorem 5.1 that $\mathbf{s}_{2^{L-2}}$ and $\mathbf{s}_{2^{L-1}+2^{L-2}}$ occur in the same relative order as $\mathbf{s}_0$ and $\mathbf{s}_{2^{L-1}}$ in permuted *state* sequence $[\mathbf{s}(\mathbf{o}_k)]$. *Origins* $\mathbf{o}_m$ and $\mathbf{o}_n$ can then be identified as follows

$$\begin{cases} \mathbf{s}(\mathbf{o}_n) = \mathbf{s}_{2^{L-1}+2^{L-2}}, \mathbf{s}(\mathbf{o}_m) = \mathbf{s}_{2^{L-2}}, & \text{if } i < j; \\ \mathbf{s}(\mathbf{o}_m) = \mathbf{s}_{2^{L-1}+2^{L-2}}, \mathbf{s}(\mathbf{o}_n) = \mathbf{s}_{2^{L-2}}, & \text{otherwise.} \end{cases}$$

Step 3. Each new correspondence can be exploited in the same way as that in the previous step to give another pair of *state-origin* correspondences. In this way, the whole *state-origin* mapping system can be constructed sequentially.

It should be noted that, in the above procedures, only valid transitions out of a *state* are exploited, i.e., the nonzero elements occurring in pairs in the row direction (see the following example). Alternatively, we can derive the *state-origin* correspondences from the transitions into a *state*, i.e., the nonzero elements occurring in pairs in the column direction, or both.

Let us illustrate the identification process with an example. Table 5.1 shows an *origin* transition record obtained from a channel with coefficient vector $[1, 0.25, 0.5]^T$, at SNR=13 dB. The *origins* are localized as described in Section 5.4, and the samples are classified using Eq. (5.32). The (i, j) entry of the table is the number of transitions from *origin* $\mathbf{o}_i$ to $\mathbf{o}_j$ detected within the observation period (2000 sampling intervals). For

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

Table 5.1. *Origin* transition record obtained for a communication channel with impulse response $[1, 0.25, 0.5]^T$ and at SNR 13 dB (2000 symbols)

	$j=0$	1	2	3	4	5	6	7
$i=0$	90	7	12	20	101	5	14	0
1	17	12	85	10	15	7	71	6
2	95	4	11	8	80	13	16	1
3	18	20	82	19	15	21	101	12
4	10	82	17	14	12	93	12	11
5	2	14	6	89	8	19	8	115
6	16	77	8	17	6	94	7	12
7	1	7	6	111	14	9	8	106

Table 5.2. *Origin* transition record obtained for a communication channel with impulse response $[1, 0.5, 0.25]^T$ and at SNR 13 dB (2000 symbols)

	$j=0$	1	2	3	4	5	6	7
$i=0$	131	2	0	23	118	4	0	0
1	91	14	1	14	86	10	2	0
2	28	83	4	2	27	91	5	0
3	26	89	29	0	21	95	21	1
4	2	19	97	13	0	19	91	13
5	0	11	86	19	0	15	80	25
6	0	0	12	99	0	1	22	98
7	0	0	11	112	2	0	11	123

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

Table 5.3. *Origin* transition record obtained for a communication channel with impulse response $[-0.5, 0.25, 1]^T$ and at SNR 13 dB (2000 symbols)

	$j=0$	1	2	3	4	5	6	7
$i=0$	19	95	11	94	16	15	2	19
1	18	10	28	6	9	85	15	98
2	77	18	72	7	8	4	8	6
3	10	11	7	12	88	23	69	17
4	16	85	13	94	28	7	11	10
5	15	17	6	9	7	93	11	104
6	103	28	47	10	4	12	7	7
7	13	5	15	5	104	23	96	17

Table 5.4. *Origin* transition record obtained for a communication channel with impulse response $[-1, -0.25, -0.5]^T$ and at SNR 13 dB (2000 symbols)

	$j=0$	1	2	3	4	5	6	7
$i=0$	130	6	13	22	100	4	17	2
1	25	15	79	6	16	12	86	5
2	97	9	19	20	82	2	20	1
3	10	29	90	6	10	29	81	13
4	15	79	16	18	23	75	9	25
5	3	11	13	83	4	12	8	83
6	14	86	7	18	20	74	4	15
7	1	8	13	95	5	9	13	84

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

---

ease of comparison, the *origin* transition records obtained with channel coefficient vectors $[1, 0.5, 0.25]^T$, $[-0.5, 0.25, 1]^T$ and $[-1, -0.25, -0.5]^T$ are given in Table 5.2, 5.3 and 5.4 respectively. Note that the arrangements of the $2L$ most significant entries associated with the legitimate *origin* transitions, are different in the three tables. If we display and compare the transition records of all the 48 ($2^3 \times 3! = 48$. See footnote 2 on page 83.) channel impulse responses with the same statistical representation (channel PDF) as that shown in Fig. 5.3(a), it can be found that two channel impulse responses, $\mathbf{f}_1$ and $\mathbf{f}_2$, have the same *origin* transition matrix if and only if they are the inverse of each other, i.e. $\mathbf{f}_1 = -\mathbf{f}_2$. This confirms that the *state-origin* correspondences can be uniquely determined when Eq. (5.31) holds.

Formalizing Table 5.1 as described in Section 5.5.2.2 leads to

$$\hat{\mathbf{Q}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

As can be seen, the diagonal entry of $\hat{\mathbf{Q}}$, q_{ii} , is nonzero only at $i = 0$ and 7. It follows from Property 5.4 and Eq. (5.31) that $\mathbf{s}(\mathbf{o}_0) = \mathbf{s}_0$ and $\mathbf{s}(\mathbf{o}_7) = \mathbf{s}_7$. By starting from the correspondence of $\mathbf{s}(\mathbf{o}_0) = \mathbf{s}_0$ and noting that the elementary *state* transition matrix for

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

Fig. 5.6. Identification of the *state-origin* correspondences

this channel is given by

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

the whole *state-origin* mapping system can be constructed, as illustrated in Fig. 5.6.

The *origin identification* process is based on the results of *origin localization*. If the

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

---

origin location estimates contain large error, the reliability of the *origin* transition matrix obtained will become questionable. To prevent false identification, we can use a ‘soft’ method to formalize the *origin* transition record. Instead of giving a hard decision for each entry of the transition matrix (transition path), we can temporarily retain the entries above a certain level and then use Property 5.1 and 5.3 to derive the transition matrix. For example, let us assume that the $(0, 1)$ entry of Table 5.1 is 100 and the others remain unchanged. By comparing the sums of the $\{(0, 0), (0, 4), (2, 0), (2, 4)\}$ entries and the $\{(0, 1), (0, 4), (2, 1), (2, 4)\}$ entries, the permitted transitions out of *origin* $\mathbf{o}_0$ can be properly determined. Similarly, a ML estimation of the *origin* transition matrix can be obtained if necessary.

It should also be noted that when some *origins* are spaced very closely, such as that shown in Fig. 5.4, even the ML estimator cannot guarantee the correctness of the identification results. Fortunately, in that case, the identification error can be totally neglected as it has little effect on the calculation of the decision threshold $T(k)$.

5.5.3 Blind Identification for M -ary Modulation

Note that the derivations in Section 5.5.2.1 are essentially based on Eq. (5.22) and (5.24). It can be easily verified that both equations hold for M -ary modulation too (after replacing 2^L with M^L), if we label the channel states so that (see Eq. (5.18))

$$\phi(i) + \phi(M^L - 1 - i) = M^L - 1. \quad (5.34)$$

Therefore, by replacing 2^L with M^L , the conclusions (theorems and corollaries) drawn for BPSK, except Eq. (5.31) (Corollary 5.2), can be applied to other modulation schemes. To include complex modulation, Eq. (5.31) should be generalized as follows

$$\sum_{i=0}^{L-1} \Re(f_i) > 0 \quad \text{and} \quad \sum_{i=0}^{L-1} \Im(f_i) > 0. \quad (5.35)$$

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

That is, the uniqueness of the *state-origin* mapping system is guaranteed if the sums of the real and the imaginary parts of the CIR are positive. The *state* transition matrix $\mathbf{Q}$ has different properties: $\mathbf{Q}$ has M nonzero elements in each row and column and the main diagonal. However, as to be shown in the following example, this does not affect the *origin* identification.

To illustrate the origin identification for M -ary modulation, we consider 4-QAM signal transmitted over a two-tap channel. The constellation of the received signal is as shown in Fig. 5.7. Let us assume that $\hat{\mathbf{Q}}$ and $\mathbf{o}(\mathbf{s}_0)$ have been obtained ($\mathbf{o}(\mathbf{s}_0)$ can be derived from $\hat{\mathbf{Q}}$ if Eq. (5.35) holds). It should be noted that the presence of ISI does not change the relative order of $\mathbf{s}_0$, $\mathbf{s}_1$, $\mathbf{s}_2$ and $\mathbf{s}_3$ (corresponding to symbol sequences “00”, “01”, “02” and “03” respectively) with respect to their centroid, that is, the order of occurrence along the circle centered at their centroid.³ Therefore, once $\mathbf{o}(\mathbf{s}_0)$ is known, $\mathbf{o}(\mathbf{s}_1)$, $\mathbf{o}(\mathbf{s}_2)$ and $\mathbf{o}(\mathbf{s}_3)$ (corresponding to the *origin* transitions into $\mathbf{o}(\mathbf{s}_0)$) are also determined. Similarly, the origins associated with “10”, “20” and “30” can also be identified from $\hat{\mathbf{Q}}$ and $\mathbf{o}(\mathbf{s}_0)$. The whole *state-origin* mapping system can be successively revealed in this way. The above procedure can be directly applied to other modulation schemes.

5.6 Decision Logic

As is shown in Fig. 5.2, the outcomes of the *origin analysis*, including the estimated locations of the *origins* obtained from *origin localization* and the *state-origin* correspondences obtained from *origin identification*, are fed into a decision logic, to calculate a set of decision thresholds:

$$T_i = \frac{1}{2} [\hat{x}(\mathbf{s}_i) + \hat{x}(\mathbf{s}_{2^{L-1}+i})], \quad i = 0, 1, \dots, 2^{L-1} - 1 \quad (5.36)$$

---

³Note that this property holds for arbitrary modulation scheme. It can be regarded as the generalization of Property 5.3 and 5.4.

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

Fig. 5.7. Origin identification for 4-QAM. Assume that the data symbols, “0” to “3”, are mapped to $[1, i]$, $[-1, i]$, $[-1, -i]$ and $[1, -i]$ respectively.

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

---

where $\hat{x}(\mathbf{s}_i)$ denotes the estimated location of the *origin* corresponding to *state* $\mathbf{s}_i$. Whenever a new sample $r(k)$ is received, the decision logic selects the appropriate threshold based on the previous *state* $\mathbf{s}(k-1)$, and implements Eq. (5.5) or (5.6) as a binary quantizer to detect the transmitted symbol. Without loss of generality, let us assume that $\mathbf{s}(k-1)$ is $\mathbf{s}_j$. It follows that at the k^{th} sampling interval the decision threshold should be switched to

$$T(k) = T_{j \gg 1}, \quad (5.37)$$

i.e., midpoint of the current COP.

The OAE is in effect equivalent to a DFE with one feedforward tap [64]. Accordingly, the OAE suffers from error propagation. The effect of error propagation is more severe in cases where the energy of the first multipath component is relatively weak. This usually occurs when the light-of-sight (LOS) between the transmitter and the receiver is blocked. To solve this problem, we can implement a sequence detector instead of the symbol detector in Eq. (5.5) and (5.6). A detailed discussion of error propagation in OAE will be given in the next chapter.

It should also be noted that, the feedback filtering of the decoded signal, required by the DFE for evaluating the postcursor distortion, has been implicitly incorporated into the *origin analysis*. Therefore, the OAE is free from the trouble of implementing the feedback filter.

5.7 Computer Simulations

Monte Carlo simulations have been conducted to compare the proposed OAE algorithm with the classical LE, DFE and MLSE algorithms, on their BER performances. Each simulation consists of 50 repetitions of the same process at given SNR. Within each

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

---

process, a random sequence of 1×10^4 symbols is transmitted repeatedly until at least 50 transmission errors have been detected. The OAE calculates the decision thresholds $\{T_i\}$ based on the first 5000 samples, and maintains their values throughout the remaining process. The LE, with 41 taps, and the DFE, with feedforward and feedback orders of 21, use the first 500 symbols as the training sequence for channel estimation. The MLSE, with a trellis length of 30, is assumed to have perfect knowledge of the CIR.

Fig. 5.8 and 5.9 illustrates the BER performance of the OAE algorithm in the presence of mild and relatively severe ISI respectively. The CIR is assumed to be $[1, 0.5]^T$ in Fig. 5.8 and $[1, -1]^T$ in Fig. 5.9⁴. As can be seen, the OAE provides comparable performance to the DFE in the former scenario. However, as a consequence of increased error propagation, it is inferior to the DFE in the presence of relatively severe ISI.

As has been discussed earlier, the nonparametric *origin localization* process is highly sensitive to the noise level and statistics. The presence of strong noise can easily obscure important features of the underlying distribution, and result in false discrimination of the *origins*. An example is given in Fig. 5.10, where the CIR is assumed to be $[1, -0.5, 0.7]^T$. Since the nonparametric *origin localizer* cannot completely separate the *origins* in the SNR range of interest (from 7 to 11 dB), the BER performance of the OAE is inferior to that of the DFE in this scenario. It should be mentioned that, when the SNR is below 5.4 dB, the underlying distribution becomes unimodal and the *origin* location estimates produced by the nonparametric localizer tend to be inaccurate. As a result, the nonparametric OAE fails.

Fig. 5.11 investigates the performance of the proposed OAE approach for a non-Gaussian noise environment, where the CIR is assumed to be $[1, 0.5]^T$ and the noise

---

⁴The performance of the proposed OAE algorithm in relatively more realistic channels will be investigated in the next chapter.

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

Fig. 5.8. BER performance of nonparametric *origin analysis equalizer* for a channel with impulse response $[1, 0.5]^T$.

Fig. 5.9. BER performance of nonparametric *origin analysis equalizer* for a channel with impulse response $[1, -1]^T$.

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

Fig. 5.10. BER performance of nonparametric *origin analysis equalizer* for a channel with impulse response $[1, -0.5, 0.7]^T$.

Fig. 5.11. BER performance of nonparametric *origin analysis equalizer* for non-Gaussian noise.

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

---

distribution (*origin*) is modeled as an equal mixture of two Gaussian components with different means $\pm d$ and same variance σ_n^2 . The variance σ_n^2 is set to 0.0496 throughout the simulation, and the SNR, defined as

$$\frac{\sum_{i=0}^{L-1} |f_i|^2}{N_0 + 2d^2},$$

is varied by changing the the distance between the two components $2d$. When the SNR is greater than 7 dB, the noise distribution remains unimodal, and the proposed OAE works satisfactorily without the need of special adjustment. However, if the two components are separated clearly such that the *origins* become multimodal, the process will need to know the noise distribution for grouping the density peaks that belong to the same *origins*.

5.8 Summary

This chapter introduces a new approach to channel equalization, where the decision rule is adapted according to time variation of the ISI. The adaptive decision is achieved via a process termed *origin analysis*. The *origin analysis* consists of two processes: *origin localization* and *identification*. The *origin localization* process aims to build up a channel statistical model, and the *origin identification* process reinforces the model to fulfil the requirement of adaptation of the decision rule. Specifically, the *origin localization* process estimates the amplitude of the channel impulse response, and the *origin identification* process provides the phase information, i.e., the order and sign of each channel coefficient. One advantage of this new channel characterization approach is that the wastage of bandwidth needed for conventional channel estimation can be reduced greatly. In general, a very short training sequence is sufficient; and under a loose condition that the sum of the channel coefficients is positive, a blind *origin analysis* can work effectively too.

CHAPTER 5. ORIGIN ANALYSIS EQUALIZER

A particular realization of the proposed approach has been described and investigated, where the *origin* locations are derived from the nonparametric PDF estimate of the channel statistics. Simulations show that this OAE can provide comparable BER performance to the DFE in some special cases. However, as the nonparametric *origin localizer* relies on the shape of the observed distribution to separate the *origins*, the OAE cannot work well in the presence of strong noise. Furthermore, due to error propagation, the decision-directed symbol-by-symbol detection cannot effectively mitigate the ISI, especially when the eye diagram of the signal is closed. Robust *origin analysis* will be addressed in the next chapter, together with a study on sequence detection for OAE, as a solution to the error propagation problem.

Chapter 6

Semi-Parametric Origin Analysis Equalizer

6.1 Introduction

The nonparametric *origin localization* approach developed in the previous chapter has a number of inherent limitations. This is because the nonparametric PDF estimate contains much irrelevant detail, which in some cases can overwhelm the desired information. As the objective of *origin localization* is to provide a set of parameters, i.e. the *origin* locations, to represent the channel statistics, it seems more appropriate to use semi-parametric PDF estimate. This chapter will focus on *origin localization* based on semi-parametric density estimation. In addition, we will investigate sequence detection, as a way to overcome the error propagation problem encountered in the proposed OAE (Origin Analysis Equalization).

6.2 Finite Gaussian Mixture Model

As in the previous chapter, we consider the transmission of BPSK signal over a frequency-selective slowly-fading channel, and use the tapped delay line channel model, i.e.,

$$r(k) = \sum_{i=0}^{L-1} f_i a(k-i) + n(k) \quad (6.1)$$

where L is the number of resolvable multipaths, $\mathbf{f} \equiv [f_i]^T$ is the channel coefficient vector, and $\{a(k)\}$, $\{n(k)\}$ and $\{r(k)\}$ denote the transmitted symbols, noise components and samples of the received signal respectively. It follows that the channel statistics can be expressed as

$$p\{r(k)\} = \sum_{i=0}^{2^L-1} \frac{P_i(k)}{\sqrt{2\pi}\sigma_n} \exp \left\{ -\frac{[r(k) - x_i(k)]^2}{2\sigma_n^2} \right\} \quad (6.2)$$

where $x_i(k)$ is the location of the i^{th} *origin* at the k^{th} sampling interval, σ_n is the standard deviation of the Gaussian noise component, and $P_i(k)$ is the probability that the sample $r(k)$ comes from the i^{th} *origin*. The objective of *origin localization* is to estimate $\mathbf{x}(k) \equiv [x_0(k), x_1(k), \dots, x_{2^L-1}(k)]^T$ based on the available sample sequence $\mathbf{r} \equiv [r_1, r_2, \dots, r_n]^T$, where n represents the sample size and $r_j = r(j+k-n) \forall j \in \{1, 2, \dots, n\}$. Clearly, this is a finite Gaussian mixture problem.

6.3 Maximum Likelihood Localization

Assuming that $\mathbf{x}(k)$ and $\mathbf{P}(k) \equiv [P_0(k), P_1(k), \dots, P_{2^L-1}(k)]^T$ are stationary over the observation period, we can then leave out the time labels and rewrite Eq. (6.2) as follows

$$p(r) = \sum_{i=0}^{2^L-1} P_i g_i(r; x_i, \sigma_n^2) \quad (6.3)$$

where $g_i(r; x_i, \sigma_n^2)$ denotes the i^{th} *origin*, i.e.,

$$g_i(r; x_i, \sigma_n^2) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp \left\{ -\frac{(r - x_i)^2}{2\sigma_n^2} \right\}. \quad (6.4)$$

CHAPTER 6. SEMI-PARAMETRIC ORIGIN ANALYSIS EQUALIZER

The joint probability density of $\mathbf{r}$ (likelihood function) is

$$\mathcal{P}(\mathbf{r}; \mathbf{P}, \mathbf{x}, \sigma_n^2) = \prod_{j=1}^n p\{r_j\} = \prod_{j=1}^n \sum_{i=0}^{2^L-1} P_i g_i(r_j; x_i, \sigma_n^2). \quad (6.5)$$

The ML estimates of σ_n , $\mathbf{x}$ and $\mathbf{P}$ follow from the maximization of the likelihood under the condition that $\sum_{i=0}^{2^L-1} P_i = 1$. In many cases, it is more convenient to use the log-likelihood function that is given by

$$\log \mathcal{P}(\mathbf{r}; \mathbf{P}, \mathbf{x}, \sigma_n^2) = \sum_{j=1}^n \log \left\{ \sum_{i=0}^{2^L-1} P_i g_i(r_j; x_i, \sigma_n^2) \right\}. \quad (6.6)$$

Let us define

$$\mathcal{L}(\mathbf{r}; \mathbf{P}, \mathbf{x}, \sigma_n^2) = \log \mathcal{P}(\mathbf{r}; \mathbf{P}, \mathbf{x}, \sigma_n^2) - \lambda \left(\sum_{i=0}^{2^L-1} P_i - 1 \right) \quad (6.7)$$

where λ is a Lagrange multiplier used to introduce the constraint $\sum_{i=0}^{2^L-1} P_i = 1$. The ML estimation problem can be equivalently solved by maximizing this metric function.

6.3.1 Unconstrained EM Localization

Let us first consider the *origin localization* based on unconstrained EM. By equating the derivatives of $\mathcal{L}$ with respect to each independent parameter to zero, we can obtain

$$\frac{\partial \mathcal{L}}{\partial P_i} = \sum_{j=1}^n \frac{g_i(r_j; x_i, \sigma_n^2)}{p(r_j)} - \lambda = 0 \quad (6.8)$$

$$\frac{\partial \mathcal{L}}{\partial \sigma_n^2} = \sum_{j=1}^n \sum_{i=0}^{2^L-1} \frac{P_i}{p(r_j)} \frac{\partial g_i(r_j; x_i, \sigma_n^2)}{\partial \sigma_n^2} = 0 \quad (6.9)$$

and

$$\frac{\partial \mathcal{L}}{\partial x_i} = \sum_{j=1}^n \frac{P_i}{p(r_j)} \frac{\partial g_i(r_j; x_i, \sigma_n^2)}{\partial x_i} = 0. \quad (6.10)$$

Multiplying Eq. (6.8) by P_i and summing over i shows

$$\lambda = n.$$

CHAPTER 6. SEMI-PARAMETRIC ORIGIN ANALYSIS EQUALIZER

Moreover, from the Bayes' theorem, we know that the posterior probability that a given sample r_j belongs to the i^{th} origin is

$$P(i | r_j) = \frac{P_i g_i(r_j; x_i, \sigma_n^2)}{p(r_j)}. \quad (6.11)$$

Multiplying Eq. (6.8) by P_i and using the above equation, we can express the ML estimate of P_i as follows

$$\hat{P}_i = \frac{1}{n} \sum_{j=1}^n P(i | r_j), \quad i = 0, 1, \dots, 2^L - 1. \quad (6.12)$$

Applying Eq. (6.11) to Eq. (6.10) and noting that

$$\frac{\partial g_i(r_j; x_i, \sigma_n^2)}{\partial x_i} = g_i(r_j; x_i, \sigma_n^2)(r_j - x_i)\sigma_n^{-2},$$

we obtain the ML estimate of x_i

$$\hat{x}_i = \frac{1}{nP_i} \sum_{j=1}^n P(i | r_j) r_j, \quad i = 0, 1, \dots, 2^L - 1. \quad (6.13)$$

Similarly, applying Eq. (6.11) to Eq. (6.9) and noting that

$$\frac{\partial g_i(r_j; x_i, \sigma_n^2)}{\partial \sigma_n^2} = \frac{g_i(r_j; x_i, \sigma_n^2)}{2\sigma_n^2} \left[\frac{(r_j - x_i)^2}{\sigma_n^2} - 1 \right],$$

we obtain the ML estimate of σ_n^2

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{j=1}^n \sum_{i=0}^{2^L-1} P(i | r_j) (r_j - x_i)^2. \quad (6.14)$$

It can be seen that, the ML estimators for the parameters of the finite Gaussian mixture model are closely analogous to those for the parameters of a Gaussian distribution, except that each data is weighted by the posterior probability $P(i | r_j)$. In the ideal case, $P(i | r_j)$ is equal to one if r_j comes from the i^{th} origin and zero otherwise, and the finite mixture modelling becomes 2^L separate Gaussian estimation.

CHAPTER 6. SEMI-PARAMETRIC ORIGIN ANALYSIS EQUALIZER

Since the posterior probabilities involve the parameters being estimated, it is impossible to calculate the ML estimates of the parameters directly by solving Eq. (6.13) and (6.14). A classical solution to this problem is the EM (Expectation-Maximization) algorithm that has been introduced in Chapter 2. The EM algorithm iteratively alternates between an E-step and an M-step until some predefined convergence criterion is met. For this particular mixture model, the E-step on the $(k+1)^{\text{th}}$ iteration calculates the Bayesian estimates of the posterior probabilities based on the parameter estimates obtained at the k^{th} iteration; and the subsequent M-step updates the parameter estimates by applying the results of the E-step to Eq. (6.12), (6.13) and (6.14) successively. If the convergence requirement is not satisfied, the new parameter estimates will be used to initiate another round of iteration.

The EM algorithm needs to be provided with some initial values, either the parameters being estimated or the posteriori probabilities. For multivariate data analysis, the selection of initial values is almost essential if one is to avoid misleading solutions. For univariate data analysis, it is not critical but still important [20, 21]. In this study, the EM algorithm is initialized by assuming that the mixing proportions $\{P_i\}$ are equal and the *origins* are clearly separated

$$\begin{aligned}\hat{P}_i^{[0]} &= \frac{1}{2^L} \\ \hat{x}_i^{[0]} &= \tilde{r}_{(2i+1)n_c}, \quad i = 0, 1, \dots, 2^L - 1 \\ \hat{\sigma}_n^{2[0]} &= \frac{1}{n} \sum_{i=0}^{2^L-1} \sum_{j=1}^{2n_c} (\tilde{r}_{2in_c+j} - \hat{x}_i^{[0]})^2,\end{aligned}\tag{6.15}$$

where $n_c \equiv n/2^{L+1}$ and $\{\tilde{r}_j\}$ is a permutation of $\{r_j\}$, obtained by rearranging the samples in descending order. To be specific, the ordered sequence $\{\tilde{r}_j\}$ is divided equally into 2^L groups. The medians of each group are then used as the initial *origin* locations and the mean sample variance is used as the initial variance.

CHAPTER 6. SEMI-PARAMETRIC ORIGIN ANALYSIS EQUALIZER

The EM algorithm is considered to have converged if the relative change in the log-likelihood or the parameter estimates in consecutive iterations is smaller than a given tolerance. This study adopts the latter criterion, i.e., stop the iterations if

$$\frac{\sum_{i=0}^{2^L-1} |\hat{x}_i^{[k+1]} - \hat{x}_i^{[k]}|}{\sum_{i=0}^{2^L-1} |\hat{x}_i^{[k]}|} < \delta_x, \quad (6.16)$$

where δ_x is the desired tolerance, 0.3% for instance. Note that this criterion is a measure of the progress of convergence. Thus, the convergence to a maximum cannot be guaranteed.

Fig. 6.1 and 6.2 investigate the performance of constrained EM *origin localization* for the channel statistics shown in Fig. 5.3(a) (CIR = $[1, 0.5, 0.25]^T$, SNR = 13 dB). Fig. 6.1 gives the histogram of the normalized rms error of the location estimate (see Eq. (5.16)) derived from 100 samples, and Fig. 6.2 plots the normalized rms errors versus the sample size. By comparing Fig. 6.2 to Fig. 5.5, one can conclude that the unconstrained EM localization approach can provide more accurate location estimates than the nonparametric localization approaches. However, its performance is still unsatisfactory, especially when the sample size is relatively small, as can be seen from Fig. 6.2.

6.3.2 Constrained EM Localization

The relatively poor performance of unconstrained localization can be easily explained. Without any guidance, the EM localizer will attempt to reproduce the observed statistics accurately. The observed statistics is, however, highly dynamic when the sample size is very small. By introducing some constraints, the accuracy of the EM localizer can be improved greatly.

Recall that the *origins* are located symmetrically, i.e., $\forall i \in [0, 2^L - 1]$, $x_i = -x_{\bar{i}}$, where $\bar{i} = 2^L - 1 - i$. Thus, the *origin* location vector $\mathbf{x}$ has only 2^{L-1} independent elements. As a result, the likelihood function Eq. (6.5) becomes

$$\mathcal{P}(\mathbf{r}; \mathbf{P}, \mathbf{x}, \sigma_n^2) = \prod_{j=1}^n \sum_{i=0}^{2^{L-1}-1} [P_i g_i(r_j; x_i, \sigma_n^2) + P_{\bar{i}} g_{\bar{i}}(r_j; -x_i, \sigma_n^2)], \quad (6.17)$$

CHAPTER 6. SEMI-PARAMETRIC ORIGIN ANALYSIS EQUALIZER

Fig. 6.1. Histogram of the *origin localization* error of unconstrained EM localizer for a channel with impulse response $[1, 0.5, 0.25]^T$

Fig. 6.2. Error performance of unconstrained EM *origin localization*

CHAPTER 6. SEMI-PARAMETRIC ORIGIN ANALYSIS EQUALIZER

and accordingly, the location estimator Eq. (6.13) can be rewritten as

$$\hat{x}_i = -\hat{x}_{\bar{i}} = \frac{1}{n(P_i + P_{\bar{i}})} \sum_{j=1}^n [P(i | r_j) - P(\bar{i} | r_j)] r_j, \quad i = 0, \dots, 2^{L-1} - 1. \quad (6.18)$$

Compared with Eq. (6.13), this location estimator is advantageous in that the samples belonging to the i^{th} and $\bar{i}^{\text{th}}$ *origins* can be used together to construct the estimate of x_i . Consequently, the number of samples required for the localizer to produce accurate *origin* location estimates can be reduced greatly. Fig. 6.3 shows the histogram of the normalized rms error of this *origin localizer* (denoted as “Ia”). By comparing it to Fig. 6.1, one can see the advantage of constrained EM localization easily.

As has been mentioned in Section 5.5.2, the *origins* can be divided into 2^{L-1} pairs with equal distance. This property can also be integrated into EM localization, resulting in the following *origin* location estimator

$$\begin{aligned} \hat{x}_{2i} = -\hat{x}_{\bar{2i}} &= \frac{\sum_{j=1}^n [P(2i | r_j) - P(\bar{2i} | r_j)] r_j}{n(P_{2i} + P_{2i+1} + P_{\bar{2i}} + P_{\bar{2i+1}})} \\ &+ \frac{\sum_{j=1}^n [P(2i+1 | r_j) - P(\bar{2i+1} | r_j)] (r_j - d)}{n(P_{2i} + P_{2i+1} + P_{\bar{2i}} + P_{\bar{2i+1}})}, \quad (6.19) \\ \hat{x}_{2i+1} = -\hat{x}_{\bar{2i+1}} &= \hat{x}_{2i} - d, \quad i = 0, 1, \dots, 2^{L-2} - 1, \end{aligned}$$

where d denotes the distance between the paired *origins*. The ML estimate of d can be obtained by forcing the derivative of likelihood function with respect to d to be zero, and the result is

$$\hat{d} = \frac{\sum_{j=1}^n \sum_{i=0}^{2^{L-2}-1} [P(\bar{2i+1} | r_j)(r_j + x_{2i}) - P(2i+1 | r_j)(r_j - x_{2i})]}{n \sum_{i=0}^{2^{L-2}-1} (P_{2i+1} + P_{\bar{2i+1}})}. \quad (6.20)$$

Fig. 6.4 displays the histogram of the normalized rms error of this constrained EM localizer (IIa). As can be seen, the incorporation of an additional constraint does not bring a noticeable performance advantage. If the strongest multipath component is weaker than the sum of the others (i.e., the eye diagram is totally closed), the performance of the

CHAPTER 6. SEMI-PARAMETRIC ORIGIN ANALYSIS EQUALIZER

Fig. 6.3. Histogram of the *origin localization* error of constrained EM localizer Ia for a channel with impulse response $[1, 0.5, 0.25]^T$

Fig. 6.4. Histogram of the *origin localization* error of constrained EM localizer IIa for a channel with impulse response $[1, 0.5, 0.25]^T$

CHAPTER 6. SEMI-PARAMETRIC ORIGIN ANALYSIS EQUALIZER

---

location estimator can even be degraded, due to the increased chance of local convergence. An example is given in Fig. 6.6, where the CIR has been set to $[1, 0.5, 0.7]^T$. Compared with the results of EM localization Ia shown in Fig. 6.5, the location estimates produced by this localizer are inferior in accuracy. The underlying reason is that Eq. (6.19) places restrictions on the order of $\{\hat{x}_i\}$. In order to comply with these restrictions, the localization accuracy is often sacrificed when the eye diagram is closed. In the previous example, the localization Ia tends to converge to $\hat{\mathbf{x}} = [2.2, 1.2, 0.8, 0.2, -0.2, -0.8, -1.2, -2.2]^T$, which is the best EM localization result that can be obtained with the start value of Eq. (6.15). However, this location estimate will not be accepted by the localizer IIa, since the equal distance condition Eq. (6.19) is not satisfied.

If all symbols are transmitted with equal probability, it is not necessary to estimate the mixing proportions $\{P_i\}$, and accordingly, the EM *origin* localizers can be simplified. However, since the EM localization will have a higher probability to be trapped in a local maximum, this simplification is not recommended, unless the sample size n is sufficiently large such that the equal mixing model matches the observed statistics very well.

The performance of these EM localization approaches is compared in Table 6.1 and 6.2¹, where “constrained Ib” and “I Ib” refer to the simplifications of the localization approaches Ia and IIa by assuming equal mixing proportions, “ErrX” represents the normalized localization error, “ErrVar” represents the normalized variance estimation error, and “RepNum” and “Time” represent respectively the average number of iterations and the average computation time for these approaches to converge. The tolerance level δ_x (Eq. (6.16)) has been set to 0.3%. As can be observed from Table 6.2, the introduction of the equal mixing property degrades the localization performance in the presence of relatively severe ISI. In the following discussions, we will focus on localization approach Ia, which has shown the best overall performance in the two scenarios.

---

¹The results shown in the tables are obtained on a personal computer with an Intel Pentium IV 3.06 GHz processor and 1 GB of RAM.

CHAPTER 6. SEMI-PARAMETRIC ORIGIN ANALYSIS EQUALIZER

Fig. 6.5. Histogram of the *origin localization* error of constrained EM localizer Ia for a channel with impulse response $[1, 0.5, 0.7]^T$

Fig. 6.6. Histogram of the *origin localization* error for constrained EM localizer IIa for a channel with impulse response $[1, 0.5, 0.7]^T$

CHAPTER 6. SEMI-PARAMETRIC ORIGIN ANALYSIS EQUALIZER

Table 6.1. Performance indices of various EM *origin localizers* for a channel with impulse response $[1, 0.5, 0.25]^T$ and at SNR 13 dB (100 samples)

	ErrX	ErrVar	RepNum	Time (ms)
unconstrained	0.260	0.352	6.4	2.41
constrained Ia	0.059	0.254	5.4	1.70
constrained Ib	0.114	1.271	14.9	4.48
constrained IIa	0.055	0.275	4.6	1.60
constrained IIb	0.088	0.969	8.4	2.22

Table 6.2. Performance indices of various EM *origin localizers* for a channel with impulse response $[1, 0.5, 0.7]^T$ and at SNR 13 dB (100 samples)

	ErrX	ErrVar	RepNum	Time (ms)
unconstrained	0.263	0.254	7.4	2.23
constrained Ia	0.062	0.280	8.3	2.34
constrained Ib	0.124	1.497	12.3	3.41
constrained IIa	0.144	0.761	13.2	3.84
constrained IIb	0.133	1.656	14.2	3.69

6.3.3 Rate of Convergence

One critical concern in the development of an adaptive equalization algorithm is its adaptability to the time-varying channel characteristics. This, on the one hand, highlights the importance of data efficiency, i.e., the observation time required for the equalizer to maintain a satisfactory performance. On the other hand, it is concerned with the rate of convergence, i.e., the computation time taken by the equalizer to reach a relatively stable state. This section investigates the second problem in the context of OAE.

For the cases considered in Table 6.1 and 6.2, the rate of convergence is not a serious problem. As can be seen from Table 6.2, localizer Ia needs about 2.34 ms to converge. It follows that, this localizer can support a transmission rate of up to 42.7 Kbps.² However, in more complicated cases, e.g., multidimensional modulation or large multipath spread, more samples are required and speed-up of convergence becomes a necessity.

A natural thinking is to reduce the number of parameters to be estimated. In the previous section, we have attempted to use the equal distance property to remove the redundant parameters. Unfortunately, this effort has been proven to be a failure (compare constrained Ia and IIa in Table 6.2). An alternative method to achieve faster convergence is to separate the variance estimation from the EM algorithm.

In order to obtain a reliable variance estimate, we can construct a feedback mechanism to remove the signal component from the received signal and then calculate the variance of the remainder (ideally, the pure noise component). The noise component comprising sample $r(k)$ can be estimated as

$$\hat{n}(k) = r(k) - \hat{x}(k), \quad (6.21)$$

where $\hat{x}(k)$ denotes the location estimate of the most probable *origin* of $r(k)$ (see Section 5.6). The variance of $\{\hat{n}(k)\}$ can then be used to simplify the EM algorithm.

CHAPTER 6. SEMI-PARAMETRIC ORIGIN ANALYSIS EQUALIZER

Fig. 6.7. Convergence of constrained EM *origin localizers* for a channel with impulse response $[1, 0.5, 0.7]^T$

CHAPTER 6. SEMI-PARAMETRIC ORIGIN ANALYSIS EQUALIZER

Table 6.3. Performance indices of constrained EM *origin localizer* Ia^+ for a channel with impulse response $[1, 0.5, 0.7]^T$ and at SNR 13 dB (100 samples)

δ_x	ErrX	RepNum	Time (ms)
0.3%	0.050	5.6	1.52
0.5%	0.052	4.4	1.27
1.0%	0.053	2.9	0.85
1.5%	0.054	2.2	0.64
2.0%	0.056	1.9	0.53
5.0%	0.062	1.0	0.41

Fig. 6.7 illustrates the improvement on convergence achieved by implementing a separate variance estimator, wherein the resultant new *origin localizer* is denoted as ‘ Ia^+ ’. Clearly, this method can be used to reduce the number of iterations without degrading the quality of the location estimate. It should be noted that, the method does not accelerate the convergence to the potential (maximum accuracy) of the EM localizer—it improves the potential itself. Therefore, it is more appropriate to treat this method as a means to enhance the localization accuracy. Table 6.3 lists the performance indices of localizer Ia^+ with different choices of tolerance δ_x for the channel model considered in Table 6.2. It can be seen that, with as less as one EM iteration, the accelerated *origin localizer* Ia^+ can achieve a comparable performance to Ia . Correspondingly, it can support a higher transmission rate (243.9 Kbps).

General acceleration approaches have been studied in, e.g., [20]. They are not considered in this study, because most of them compromise performance for faster convergence.

---

²Assume that each run of EM *origin localization* is performed independently.

However, in the cases where the rate of convergence is of critical concern and the localization accuracy is comparatively less important, these approaches can apply.

6.3.4 Spurious Convergence Control

The likelihood function is multimodal, and there is no guarantee that the EM algorithm will always converge to the global maximum likelihood. In the cases that the EM algorithm is trapped in a spurious maximum, the obtained *origin* location estimates may contain large error, and consequently, the performance of the OAE is degraded significantly.

Fig. 6.8 plots the normalized localization error versus the normalized variance estimation error for constrained EM localizer Ia. It can be observed that the localization error is in principle proportional to the variance estimation error. Therefore, the method of separating variance estimation from the EM algorithm described in the previous section is also helpful for overcoming spurious convergences caused by poor variance estimation. However, Fig. 6.8 also clearly illustrates that most spurious convergences have no direct relation with the accuracy of the variance estimator.

In general, inappropriate choice of initial values is the main cause of spurious convergence [20]. Accordingly, a common strategy to deal with spurious convergence is to start the EM algorithm with a variety of initial values and locate the highest likelihood maximum. The initial location values can be generated by treating them as independent realizations of a random variable $X \sim N(\mu_r, \sigma_r^2)$, where μ_r and σ_r^2 denote the sample mean and variance respectively. It should be noted that a higher likelihood does not always suggest a better parameter estimate, since the likelihood measures the degree that a parameter estimate is consistent with the observed statistics but not their true values. Therefore, in the context of *origin localization*, it is more reasonable to grade the parameter estimates according to their risks of being spurious rather than the likelihoods. Let

Fig. 6.8. Correspondence between the *origin localization* error and the variance estimation error in constrained EM localization Ia for a channel with impulse response $[1, 0.5, 0.7]^T$ and at SNR 13 dB

CHAPTER 6. SEMI-PARAMETRIC ORIGIN ANALYSIS EQUALIZER

$\{\hat{\mathbf{x}}_i \mid i = 1, \dots, M_I\}$ be the location estimates resulted from M_I different initializations. Usually, a majority of them are concentrated in a small area. We can remove the highly suspicious members and calculate the centroid of the remainders to form a reliable location estimate. A suspicious member, known as an outlier in statistics, can be identified by verifying if its distance from the other members is above a certain value, e.g. $\lambda \bar{d}$, where $\bar{d} \equiv M_I^{-1} \sum_{i=1}^{M_I} d_i$ is the mean of the distance $d_i \equiv \sum_{j=1}^{M_I} \|\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j\|$, and λ is a threshold. In favor of Spurious Convergence Control (SCC), a relatively small threshold is recommended, e.g. $\lambda = 1.5$.

Table 6.4 compares the rates of occurrence of spurious convergence with and without SCC for the channel statistics investigated in Fig. 6.8, where spurious convergence has been defined to be that ending with a normalized localization error greater than 0.18, and ‘SCC I’ and ‘SCC II’ refer respectively to the parameter selections based on the maximum likelihood and minimum risk criteria. It should be mentioned that, without any constraint, the SCC will promote but not suppress the occurrence of spurious convergence for localization Ia. To solve this problem, in this simulation we have limited the range of variance estimate as follows

$$\hat{\sigma}_n^2 < \frac{\sum_{i=0}^{L-1} f_i^2}{2\gamma_{\max}}, \quad (6.22)$$

where $\gamma_{\max}$ denotes the maximum expectation for the SNR, 3.2 (5 dB) in this instance. As is illustrated in the table, the minimum distance criterion is more reliable in preventing spurious convergence. It should also be noted that increasing the number of random initializations in SCC I does not always reduce the occurrence of spurious convergence, which sufficiently demonstrates the difference between the most likely and most desirable parameter estimates. For ease of comparison, the histograms of localization errors of Ia⁺ without SCC and with SCC II are given in Fig. 6.9 and 6.10.

CHAPTER 6. SEMI-PARAMETRIC ORIGIN ANALYSIS EQUALIZER

Fig. 6.9. Histogram of the *origin localization* error of constrained EM localizer Ia^+ with SCC I for a channel with impulse response $[1, 0.5, 0.7]^T$

Fig. 6.10. Histogram of the *origin localization* error of constrained EM localizer Ia^+ with SCC II for a channel with impulse response $[1, 0.5, 0.7]^T$

Table 6.4. Rates of occurrence of spurious convergence with and without spurious convergence control (100 samples)

M_I	1	6	10
Ia	3.54×10^{-2}	—	—
Ia, SCC I	—	1.08×10^{-2}	1.22×10^{-2}
Ia, SCC II	—	2.38×10^{-3}	1.51×10^{-3}
Ia ⁺	2.88×10^{-2}	—	—
Ia ⁺ , SCC I	—	7.31×10^{-3}	8.05×10^{-3}
Ia ⁺ , SCC II	—	5.65×10^{-4}	4.56×10^{-4}

Instead of using the maximum likelihood or minimum distance criterion, one can develop his own measure of the quality of *origin localization*. For example, we can implement a nonparametric estimator for the underlying density, e.g. the kernel density estimator introduced in Section 2.3.2, and select the localization result that best matches the nonparametric density estimate. In this way, spurious convergence can be almost totally avoided (at least, those containing large localization error are unlikely to occur). This approach is not investigated because of the computational complexity.

6.3.5 EM Localization for M -ary Modulation

The EM localization can be easily extended to M -ary modulation. However, it should be noted that the channel statistics is now a mixture of M^L Gaussian components (*origins*) and, for complex modulation, the i^{th} *origin* is a complex Gaussian distribution, i.e.,

$$g_i(r; x_i, \sigma_n^2) = \frac{1}{2\pi\sigma_n^2} \exp \left\{ -\frac{\|r - x_i\|^2}{2\sigma_n^2} \right\}, \quad (6.23)$$

CHAPTER 6. SEMI-PARAMETRIC ORIGIN ANALYSIS EQUALIZER

where r and x_i are both complex variables. Therefore, the number of samples required for EM localization Ia is linearly proportional to M^L . For large M and L , the observation cost and computational complexity associated with this localization scheme can be prohibitive. This problem can be overcome by exploiting the correlations of $\{x_i\}$, such as in localization IIa. However, as illustrated in Section 6.3.2, introducing more constraints increases the risk of spurious convergence and, in some cases, can lead to significant degradation of the localization accuracy. The exploration for more robust and efficient localization approach should be the focus of the future research on *origin analysis equalization*.

It should be mentioned that, for convenience, we can label the origins of a complex signal according to the phases of their means, in ascending order from $-\pi$ to π . Consequently, the symmetry of the origins can be clearly expressed as follows

$$x_i = x_{\bar{i}}, \quad (6.24)$$

where $\bar{i} = M^L - 1 - i$. This equation can then be used to simplify the localization process, the same as for BPSK modulation.

6.4 EM-Based Origin Identification

As has been mentioned in the previous chapter, the *origin identifier* relies on the estimated sample *origins* to derive the *origin* transition matrix. The *origin* estimates (sample labels) can be obtained easily as a byproduct of EM *origin localization* as follows

$$\hat{\mathbf{o}}(r_j) = \mathbf{o}_i, \quad i = \arg \max_i \hat{P}(i | r_j) \quad (6.25)$$

where $\hat{P}(i | r_j)$ is the estimate of the probability that r_j belongs to the i^{th} *origin*. This labeling approach attends to the possible deviation from the equal mixing assumption. Therefore, it is more reliable than that given in Eq. (5.32). By using this approach, we can avoid the trouble of designing a separate sample classifier. Moreover, the reliability of *origin identification* is improved (the difference, however, is subtle).

Fig. 6.11. Illustrations of error propagation in OAE

6.5 Sequence Detection

As indicated in Section 5.6, due to the use of a feedback mechanism to decide the current COP, a decision error is highly likely to be followed by a series of decision errors in OAE. This phenomenon is referred to as error propagation in the literature. Two examples are given in Fig. 6.11 to illustrate the effect of error propagation in OAE. Fig. 6.11(a) shows a scenario in which error propagation can be neglected and Fig. 6.11(b) shows a scenario in which error propagation causes serious performance degradation. Various approaches can be developed to solve this problem. Leaving out the consideration of implementation complexity, the best approach is to use a sequence detector to replace the hard-decision detector given by Eq. (5.5).

CHAPTER 6. SEMI-PARAMETRIC ORIGIN ANALYSIS EQUALIZER

Let $x^{(0)}(k)$ and $x^{(1)}(k)$ be the locations of the *origins* of $r(k)$, corresponding to $s(k) = 0$ and 1 respectively. Let $\{x^{(i)}(k+1) \mid i = 0, \dots, 3\}$ be the locations of the possible *origins* of $r(k+1)$, corresponding to $\mathbf{s} \equiv [s(k+1), s(k)]^T = [0, 0]^T, [0, 1]^T, [1, 0]^T$, and $[1, 1]^T$ respectively. For convenience, label the *states*³ of $\mathbf{s}$ with respect to their distances to all-zero vector $[0, 0]^T$, i.e., $\mathbf{s}_0 = [0, 0]^T$, $\mathbf{s}_1 = [0, 1]^T$, $\mathbf{s}_2 = [1, 0]^T$, and $\mathbf{s}_3 = [1, 1]^T$. Under the assumption that $f_0 > 0$, the locations of the *origins* shown in Fig. 6.11(a) and (b), from the right to the left, are $[x^{(i)}(k+1) \mid i = 0, 2, 1, 3]$ and $[x^{(i)}(k+1) \mid i = 0, 1, 2, 3]$ respectively. A two-symbol sequence detector can then be constructed as follows

$$\hat{s}(k) \in \mathbf{s}_i, \quad \mathbf{s}_i = \arg \min_{\mathbf{s}_i} D_2(k, \mathbf{r}; \mathbf{s}_i) \quad (6.26)$$

where $\mathbf{r} = [r(k+1), r(k)]^T$ and $D_2(k, \mathbf{r}; \mathbf{s})$ is a distance metric given by

$$D_2(k, \mathbf{r}; \mathbf{s}) = [\mathbf{r} - \mathbf{x}(\mathbf{s})]^T [\mathbf{r} - \mathbf{x}(\mathbf{s})], \quad (6.27)$$

where $\mathbf{x}(\mathbf{s})$ is a vector consisting of the *origin* locations associated with $\mathbf{s}$. For example

$$D_2(k, \mathbf{r}; \mathbf{s}_0) = \sum_{i=0}^1 [r(k+i) - x^{(0)}(k+i)]^2.$$

A K -symbol sequence detection can be obtained by minimizing the distance metric $D_K(k, \mathbf{r}; \mathbf{s})$, where $\mathbf{r}$ and $\mathbf{s}$ are $K \times 1$ vectors. When $K = 1$, the symbol-by-symbol detector in Eq. (5.5) is obtained. If $K > L$, the Viterbi algorithm [53, 54] can be used to reduce the computational complexity.

Since decision feedback is essentially a technique of truncating the CIR, the OAE based on sequence estimation (OASE) can also be interpreted as a reduced-complexity MLSE [68–71], where the channel is identified through *origin analysis*. In conventional MLSE, K cannot be smaller than L , while in the OASE, K can be selected freely. In many

---

³Note that the term “*state*” used here is different from that used in the previous chapter.

Fig. 6.12. Possible state transitions in a 3-tap channel; $M = 2$, $K = 3$, $\hat{s}(k - 1) = s_0$.

cases, the choice of $K = L - L_p$, where L_p is the number of postcursors, is sufficient for the OASE to maintain a satisfactory performance. Clearly, the advantage of the OASE (and other reduced-complexity estimation approaches) is more significant when the post-cursor response is relatively long. Fig. 6.12 exemplifies the possible state transitions in the OASE, from which, one can easily see how the complexity reduction is achieved.

Due to the residual error propagation, the OASE suffers performance degradation. However, as indicated in the previous studies [68–71], compared to the substantial reduction in complexity, the loss in performance is usually negligible. It is reminded that the error propagation can always be reduced by increasing the traceback length K . If needed, we can even discard the decision feedback, to obtain a conventional MLSE based on *origin analysis*.

6.6 Adaptive Origin Analysis Equalization

In most communication systems, the CIRs are time varying. Therefore, a well-designed channel equalizer should have the ability to adjust itself with respect to the time variation of the channel characteristics.

Adaptive OAE can be achieved readily by periodically updating the *origin localization* and *identification* results. For example, we can divide the data stream into frames of fixed size, and perform *origin analysis* once a new frame is received. This approach can still be applied even if the CIR changes significantly over successive frames. However, its computational complexity is often prohibitive, especially when SCC (Spurious Convergence Control) is needed.

An alternative approach is to further divide a data frame into a certain number of blocks, letting be n_B . When a new data block is received, the previous parameter estimates are used to evaluate the posterior probabilities $\{P(i | r_j) | i = 0, \dots, 2^L - 1\}$ for each sample of this block (see Eq. (6.11)). The results, together with the posterior probabilities of the latest $n_B - 1$ data blocks, are then used to update the parameter estimates. As a benefit of using the previous localization results to initiate the new localization, the problem of spurious convergence caused by using random start or unreliable guess can be generally avoided. Moreover, since a partial E-step is performed, the computational cost can be reduced greatly.

One may have noticed that the constrained EM localizer Ia^+ described in Section 6.3.3 requires the previously decoded signal to estimate the noise variance. In Section 6.3.2, we have already compared a variety of approaches that do not need the decoded signal and have shown that constrained localization Ia provides the best overall performance among these approaches. If there is no other means to learn the variance, we can implement

localization Ia and SCC I at the initial stage (for the first data frame), and switch to localization Ia⁺ once the decoded signal is sufficient for reliable variance estimation. It should be reminded again that the localizer Ia must be provided with some constraint, the range of variance for instance, when it works together with SCC

6.7 Error Performance Analysis

6.7.1 Error Performance of Symbol Detection

To simplify the analysis, let us assume that the *origins* are distributed sparsely such that the labels (*origins*) of the samples involved in EM localization can be estimated perfectly (it is equivalent to assume that the training signal is known to the receiver). Accordingly, the analysis of Gaussian mixture is translated into the problem of 2^L separate Gaussian estimations, each responsible for the localization of a particular *origin*. Without loss of generality, let us consider the estimation of x_i based on n_i independent realizations of $\tilde{r} \sim N(x_i, \sigma_n^2)$. It is known that under the previous assumption, the estimate $\hat{x}_i$ follows a standard Gaussian distribution $N(x_i, \sigma_x^2)$, where

$$\sigma_x^2 = \frac{\sigma_n^2}{n_i}. \quad (6.28)$$

Let $\mathbf{o}_j$ be the *origin* paired with $\mathbf{o}_i$ (i.e., $\mathbf{o}_i$ and $\mathbf{o}_j$ form a COP) and assume $x_j > x_i$. Thus, $x_j = x_i + 2|f_0|$. For the observations of $\tilde{r}$, $\{\tilde{r}_k \mid k = 1, \dots, n_i\}$, a decision error occurs only if $|\tilde{r}_k - \hat{x}_i| < |\tilde{r}_k - \hat{x}_j|$, or equivalently, $\tilde{r}_k > (\hat{x}_i + \hat{x}_j)/2$. It follows that the probability of erroneous decision for the samples coming from the i^{th} *origin*, $P_{\text{E}i}$, is given by

$$P_{\text{E}i} = P\{\tilde{r} - (\hat{x}_i + \hat{x}_j)/2 > 0\}. \quad (6.29)$$

CHAPTER 6. SEMI-PARAMETRIC ORIGIN ANALYSIS EQUALIZER

Since $\tilde{r}$, $\hat{x}_i$ and $\hat{x}_j$ are independent Gaussian random variables and

$$\tilde{r} \sim N(x_i, \sigma_n^2)$$

$$\hat{x}_i \sim N(x_i, \sigma_x^2)$$

$$\hat{x}_j \sim N(x_i + 2|f_0|, \sigma_x^2),$$

we know that the composite variable $\tilde{r} - (\hat{x}_i + \hat{x}_j)/2$ is also Gaussian

$$\tilde{r} - \frac{\hat{x}_i + \hat{x}_j}{2} \sim N\left(-|f_0|, \sigma_n^2 + \frac{\sigma_x^2}{2}\right).$$

Accordingly, Eq. (6.29) becomes

$$P_{\text{E}i} = Q\left(\frac{|f_0|}{\sqrt{\sigma_n^2 + \sigma_x^2/2}}\right). \quad (6.30)$$

Substituting Eq. (6.28) into the above equation, we obtain

$$P_{\text{E}i} = Q\left(\frac{|f_0|\sigma_n^{-1}}{\sqrt{1 + (2n_i)^{-1}}}\right). \quad (6.31)$$

Let us further assume that the mixing proportions $\{P_i\}$ are equal and the sample size n is sufficiently large so that $n_i = n/2^L$ for an arbitrary $i \in \{0, 1, \dots, 2^L - 1\}$. Accordingly, the probability of independent error for symbol-based detection Eq. (5.5), P_{IE} , can be expressed as follows

$$\begin{aligned} P_{\text{IE}} &= \sum_{i=0}^{2^L-1} P_i P_{\text{E}i} = P_{\text{E}i} = Q\left(\frac{|f_0|\sigma_n^{-1}}{\sqrt{1 + 2^{L-1}n^{-1}}}\right) \\ &= Q\left(\sqrt{\frac{2\gamma_0}{1 + 2^{L-1}n^{-1}}}\right), \end{aligned} \quad (6.32)$$

where $\gamma_0 \equiv f_0^2/(2\sigma_n^2)$ denotes the ratio of the energy of the first multipath component to that of the noise. Recall that, for a BPSK signal transmitted over an AWGN channel with SNR γ_0 , the error probability of matched-filter (correlation) detection is $Q(\sqrt{2\gamma_0})$ [29].

CHAPTER 6. SEMI-PARAMETRIC ORIGIN ANALYSIS EQUALIZER

Thus, the SNR loss in an ideal OAE, i.e., that with perfect sample labels and suffering no error propagation, is $10 \log_{10}(1 + 2^{L-1}n^{-1})$ dB. Clearly, if the training signal is known *a priori* or if the *origins* are separated sufficiently, the loss of SNR caused by localization error can be neglected, unless the sample size n is very small.

In the above discussions, it has been assumed that a decision error will not affect the successive decisions, i.e., the decision errors are independent of each other. However, as has been mentioned earlier, the OAE is subject to error propagation, which can deteriorate the performance of OAE significantly in some cases.

Use $\mathbf{e}(k)$ to denote the occurrence of a decision error at the k^{th} moment. The conditional probability of $\mathbf{e}(k+1)$, given $\mathbf{e}(k)$ and $\mathbf{s}_0$, $P\{\mathbf{e}(k+1) \mid \mathbf{e}(k), \mathbf{s}_0\}$, can be expressed as follows

$$\begin{aligned} P\{\mathbf{e}(k+1) \mid \mathbf{e}(k), \mathbf{s}_0\} &= P\{\hat{s}(k+1) = 1 \mid \hat{s}(k) = 1\} \\ &= P\{|r(k+1) - x^{(3)}(k+1)| < |r(k+1) - x^{(1)}(k+1)|\}. \end{aligned} \quad (6.33)$$

Neglecting the *origin localization* error, we have

$$r(k+1) \sim N(x^{(0)}(k+1), \sigma_n^2).$$

It is also known that

$$\begin{aligned} x^{(0)}(k+1) - x^{(1)}(k+1) &= 2f_1 \\ x^{(0)}(k+1) - x^{(3)}(k+1) &= 2f_0 + 2f_1. \end{aligned}$$

Using these expressions to simplify Eq. (6.33), we get

$$P\{\mathbf{e}(k+1) \mid \mathbf{e}(k), \mathbf{s}_0\} = Q\left(\frac{f_0 + 2f_1}{\text{sign}(f_0)\sigma_n}\right). \quad (6.34)$$

CHAPTER 6. SEMI-PARAMETRIC ORIGIN ANALYSIS EQUALIZER

Similarly, it can be deduced that

$$P\{\mathbf{e}(k+1) \mid \mathbf{e}(k), \mathbf{s}_1\} = Q\left(\frac{f_0 - 2f_1}{\text{sign}(f_0)\sigma_n}\right), \quad (6.35)$$

$$P\{\mathbf{e}(k+1) \mid \mathbf{e}(k), \mathbf{s}_2\} = Q\left(\frac{f_0 - 2f_1}{\text{sign}(f_0)\sigma_n}\right), \quad (6.36)$$

$$P\{\mathbf{e}(k+1) \mid \mathbf{e}(k), \mathbf{s}_3\} = Q\left(\frac{f_0 + 2f_1}{\text{sign}(f_0)\sigma_n}\right). \quad (6.37)$$

Assume that $\mathbf{s}_i$, $i = 0, \dots, 3$, occurs in the transmitted stream with the same probability $P(\mathbf{s}_i) = 1/4$. Thus, the conditional error probability of $\hat{s}(k+1)$ given erroneous $\hat{s}(k)$ has the form of

$$\begin{aligned} P\{\mathbf{e}(k+1) \mid \mathbf{e}(k)\} &= \sum_{i=0}^3 P\{\mathbf{e}(k+1) \mid \mathbf{e}(k), \mathbf{s}_i\} P(\mathbf{s}_i) \\ &= \frac{1}{2} \left[Q\left(\frac{|f_0| + 2f_1}{\sigma_n}\right) + Q\left(\frac{|f_0| - 2f_1}{\sigma_n}\right) \right]. \end{aligned} \quad (6.38)$$

For convenience, let us rewrite the above equation as follows

$$P\{\mathbf{e}(k+1) \mid \mathbf{e}(k)\} = \frac{1}{2} \left[Q\left(\sqrt{2\gamma_0}(1 + 2\beta)\right) + Q\left(\sqrt{2\gamma_0}(1 - 2\beta)\right) \right] \quad (6.39)$$

where we have used the following relation

$$|f_1| = \beta|f_0|. \quad (6.40)$$

Fig. 6.13 plots the error probability of $\hat{s}(k+1)$, conditioned on $\mathbf{e}(k)$, versus β . As can be seen, when β is very small, i.e., the energy of the first multipath component is much greater than the second, the effect of error propagation is negligible (this is the case illustrated in Fig. 6.11(b)). As β increases, the probability of consecutive errors will increase rapidly, and eventually, when the energy of the second multipath component goes above the first, it will approach 0.5 (this is the case illustrated in Fig. 6.11(a)). It should also be noted that the error propagation effect is relatively insensitive to the noise level (γ_0).

Fig. 6.13. Probability of consecutive errors

Thus, we can derive an upper bound for the error probability of OAE based on symbol detection, denoted as P_E ,

$$P_E < P_{IE} \sum_{l=1}^{+\infty} l \left(\frac{1}{2} \right)^l = 2P_{IE}. \quad (6.41)$$

It is well known that the Chernoff bound of P_{IE} is given by

$$P_{IE} \leq \exp \left\{ -\frac{\gamma_0}{1 + 2^{L-1}n^{-1}} \right\}. \quad (6.42)$$

By combining Eq. (6.41) and (6.42), we arrive at the conclusion that the error propagation introduces another SNR loss of up to 1.6 dB. Altogether, in symbol-detection OAE with perfect sample labeling but no error propagation control, the SNR degradation due to the presence of ISI is less than $1.6 + 10 \log_{10}(1 + 2^{L-1}n^{-1})$ dB.

6.7.2 Error Performance of Sequence Detection

Now, let us investigate the effect of using two-symbol sequence detection on the error performance of OAE. It is very difficult to obtain an explicit expression for the error probability. In this section, we will give an error bound instead, which is sufficient to show the advantages of sequence detection.

We know that, when $s(k) = 0$, the decision $\hat{s}(k)$ obtained with Eq. (6.26) is erroneous if

$$\min\{D_2(\mathfrak{s}_0), D_2(\mathfrak{s}_2)\} > \min\{D_2(\mathfrak{s}_1), D_2(\mathfrak{s}_3)\}, \quad (6.43)$$

and when $s(k) = 1$, $\hat{s}(k)$ is erroneous if

$$\min\{D_2(\mathfrak{s}_1), D_2(\mathfrak{s}_3)\} > \min\{D_2(\mathfrak{s}_0), D_2(\mathfrak{s}_2)\}, \quad (6.44)$$

where the parameters of the distance metrics, k and $\mathfrak{x}$, have been dropped for brevity.

Note that if the *origins* (component distributions) are centrosymmetric, we have

$$P\{D_2(\mathfrak{s}_0) \mid \mathfrak{s}_0\} = P\{D_2(\mathfrak{s}_3) \mid \mathfrak{s}_3\}$$

$$P\{D_2(\mathfrak{s}_2) \mid \mathfrak{s}_0\} = P\{D_2(\mathfrak{s}_1) \mid \mathfrak{s}_3\}$$

$$P\{D_2(\mathfrak{s}_1) \mid \mathfrak{s}_0\} = P\{D_2(\mathfrak{s}_2) \mid \mathfrak{s}_3\}$$

$$P\{D_2(\mathfrak{s}_3) \mid \mathfrak{s}_0\} = P\{D_2(\mathfrak{s}_0) \mid \mathfrak{s}_3\}$$

and

$$P\{D_2(\mathfrak{s}_0) \mid \mathfrak{s}_2\} = P\{D_2(\mathfrak{s}_3) \mid \mathfrak{s}_1\}$$

$$P\{D_2(\mathfrak{s}_2) \mid \mathfrak{s}_2\} = P\{D_2(\mathfrak{s}_1) \mid \mathfrak{s}_1\}$$

$$P\{D_2(\mathfrak{s}_1) \mid \mathfrak{s}_2\} = P\{D_2(\mathfrak{s}_2) \mid \mathfrak{s}_1\}$$

$$P\{D_2(\mathfrak{s}_3) \mid \mathfrak{s}_2\} = P\{D_2(\mathfrak{s}_0) \mid \mathfrak{s}_1\}.$$

CHAPTER 6. SEMI-PARAMETRIC ORIGIN ANALYSIS EQUALIZER

Thus, for transmission over a Gaussian mixture channel that can be expressed in the form of Eq. (6.2), symbols ‘0’ and ‘1’ have equal error probability. Without loss of generality, consider the detection of $s(k) = 0$. It follows from Eq. (6.43) that

$$\begin{aligned}
 P\{\mathbf{e}(k) \mid s(k) = 0\} = & \\
 & P\{D_2(\mathfrak{s}_0) > D_2(\mathfrak{s}_2); D_2(\mathfrak{s}_1) > D_2(\mathfrak{s}_3); D_2(\mathfrak{s}_2) > D_2(\mathfrak{s}_3)\} \quad [\mathbf{e}_1] \\
 & + P\{D_2(\mathfrak{s}_0) > D_2(\mathfrak{s}_2); D_2(\mathfrak{s}_1) < D_2(\mathfrak{s}_3); D_2(\mathfrak{s}_2) > D_2(\mathfrak{s}_1)\} \quad [\mathbf{e}_2] \\
 & + P\{D_2(\mathfrak{s}_0) < D_2(\mathfrak{s}_2); D_2(\mathfrak{s}_1) > D_2(\mathfrak{s}_3); D_2(\mathfrak{s}_0) > D_2(\mathfrak{s}_3)\} \quad [\mathbf{e}_3] \\
 & + P\{D_2(\mathfrak{s}_0) < D_2(\mathfrak{s}_2); D_2(\mathfrak{s}_1) < D_2(\mathfrak{s}_3); D_2(\mathfrak{s}_0) > D_2(\mathfrak{s}_1)\}, \quad [\mathbf{e}_4]
 \end{aligned}$$

where the decision errors have been classified according to their conditions of occurrence.

Let us first assume that $s(k+1) = 0$, i.e., $\mathfrak{s} = \mathfrak{s}_0$. By applying the definition of decision metric Eq. (6.27), the third condition for $\mathbf{e}_4$ to occur, $D_2(\mathfrak{s}_0) > D_2(\mathfrak{s}_1)$, can be expressed as

$$\sum_{i=0}^1 [r(k+i) - x^{(0)}(k+i)]^2 > \sum_{i=0}^1 [r(k+i) - x^{(1)}(k+i)]^2. \quad (6.45)$$

Note that, once again, the *origin localization* error has been neglected. Using the following relations to simplify Eq. (6.45)

$$\begin{aligned}
 x^{(0)}(k) - x^{(1)}(k) &= 2f_0 \\
 x^{(0)}(k+1) - x^{(1)}(k+1) &= 2f_1,
 \end{aligned}$$

we get

$$f_0 y_k + f_1 y_{k+1} < -(f_0^2 + f_1^2), \quad (6.46)$$

where $y_k \equiv r(k) - x^{(0)}(k)$ and $y_{k+1} \equiv r(k+1) - x^{(0)}(k+1)$ are two Gaussian random variables with mean zero and variance σ_n^2 . In a similar way, it can be deduced that $\mathbf{e}_1$

CHAPTER 6. SEMI-PARAMETRIC ORIGIN ANALYSIS EQUALIZER

---

occurs if

$$f_0 y_{k+1} < -f_0^2$$

$$f_0 y_{k+1} < -f_0^2 - 2f_0 f_1$$

$$f_0 y_k + f_1 y_{k+1} < -(f_0 + f_1)^2$$

hold simultaneously, $\mathbf{e}_2$ occurs if

$$f_0 y_{k+1} < -f_0^2$$

$$f_0 y_{k+1} > -f_0^2 - 2f_0 f_1$$

$$f_0 y_k + (f_1 - f_0) y_{k+1} < -f_1^2$$

hold simultaneously, $\mathbf{e}_3$ occurs if

$$f_0 y_{k+1} > -f_0^2$$

$$f_0 y_{k+1} < -f_0^2 - 2f_0 f_1$$

$$f_0 y_k + (f_0 + f_1) y_{k+1} < -f_0^2 - (f_0 + f_1)^2$$

hold simultaneously, and $\mathbf{e}_4$ occurs if

$$f_0 y_{k+1} > -f_0^2$$

$$f_0 y_{k+1} < -f_0^2 - 2f_0 f_1$$

$$f_0 y_k + f_1 y_{k+1} < -f_0^2 - f_1^2$$

hold simultaneously. In the case that both f_0 and f_1 are positive, event $\mathbf{e}_1$ occurs if

$$y_{k+1} < -f_0 - 2f_1$$

$$f_0 y_k + f_1 y_{k+1} < -(f_0 + f_1)^2$$

CHAPTER 6. SEMI-PARAMETRIC ORIGIN ANALYSIS EQUALIZER

---

hold simultaneously, $\mathbf{e}_2$ occurs if

$$\begin{aligned} -f_0 - 2f_1 &< y_{k+1} < -f_0 \\ f_0 y_k + (f_1 - f_0) y_{k+1} &< -f_1^2 \end{aligned}$$

hold simultaneously, $\mathbf{e}_4$ occurs if

$$\begin{aligned} y_{k+1} &> -f_0 \\ f_0 y_k + f_1 y_{k+1} &< -f_0^2 - f_1^2 \end{aligned}$$

hold simultaneously, and the conditions for event $\mathbf{e}_3$ to occur cannot be satisfied. Thus,

$$P\{\mathbf{e}(k) \mid \mathbf{s}_0\} = P\{\mathbf{e}_1 \mid \mathbf{s}_0\} + P\{\mathbf{e}_2 \mid \mathbf{s}_0\} + P\{\mathbf{e}_4 \mid \mathbf{s}_0\}. \quad (6.47)$$

To simplify the analysis, let us further assume $f_1 > f_0$, that is, consider the case when symbol-detection OAE suffers severe error propagation. It follows that

$$\begin{aligned} P\{\mathbf{e}_1 \mid \mathbf{s}_0\} &= P\{y(k+1) < -f_0 - 2f_1; f_0 y(k) + f_1 y(k+1) < -(f_0 + f_1)^2\} \\ &< P\{y(k+1) < -f_0 - 2f_1; f_0 y(k) + f_1 y(k+1) < -(f_0^2 + f_1^2)\} \\ P\{\mathbf{e}_2 \mid \mathbf{s}_0\} &= P\{-f_0 - 2f_1 < y(k+1) < -f_0; f_0 y(k) + (f_1 - f_0) y(k+1) < -f_1^2\} \\ &< P\{-f_0 - 2f_1 < y(k+1) < -f_0; f_0 y(k) + f_1 y(k+1) < -(f_0^2 + f_1^2)\} \\ P\{\mathbf{e}_3 \mid \mathbf{s}_0\} &= P\{y(k+1) > -f_0; f_0 y(k) + f_1 y(k+1) < -(f_0 + f_1)^2\}. \end{aligned}$$

Consequently,

$$\begin{aligned} P\{\mathbf{e}(k) \mid \mathbf{s}_0\} &< P\{f_0 y(k) + f_1 y(k+1) < -(f_0 + f_1)^2\} \\ &= Q\left(\sqrt{\frac{f_0^2 + f_1^2}{\sigma_n^2}}\right) = Q\left(\sqrt{2\gamma_0(1 + \beta^2)}\right) \end{aligned} \quad (6.48)$$

It can be verified that this error bound applies when $f_0 < f_1$ or $f_0 f_1 < 0$. Moreover, it is also valid for detection of s_1 , s_2 and s_3 . Therefore, the probability of independent decision error in two-symbol sequence detection is upper bounded by

$$P_{\text{IE}} < Q\left(\sqrt{2\gamma_0(1+\beta^2)}\right). \quad (6.49)$$

and it follows from Eq. (6.41) that the overall error probability satisfies

$$P_{\text{E}} < 2P_{\text{IE}} < 2Q\left(\sqrt{2\gamma_0(1+\beta^2)}\right). \quad (6.50)$$

Since $Q\left(\sqrt{2\gamma_0(1+\beta^2)}\right)$ is the error probability of matched-filter detection for an AWGN channel with SNR $\gamma_0(1+\beta^2)$, we can conclude that, compared to a detector that completely removes the ISI and fully exploits the energy of the first two multipath components, the SNR loss experienced by the OAE based on two-symbol detection is less than 1.6 dB.

It should be noted that, although sequence detection has been claimed to be an effective way to deal with error propagation, it does not alleviate the dependence of consecutive detection errors. What it really does is to reduce the probability of independent error in the cases where severe error propagation will be observed otherwise. Fig. 6.14 plots the probability of independent error in symbol-based and two-symbol sequence detection versus the parameter β , where γ_0 is set to 5 dB. As can be easily seen, the independent error probability of sequence detection decreases rapidly as β increases (or equivalently, as the error propagation effect becomes more significant). This is in contrast to the symbol detector, which maintains the same error probability at various β values.

6.7.3 Effect of Channel Length Misestimation

In the previous discussions, the channel length L has been assumed *a priori*. When L is unknown, we can select an empirical value, denoted as L_E , by considering some typical

Fig. 6.14. Probability of independent error in symbol and sequence detection

transmission environment, such as those described in [94]. However, it should be attended that an empirical determination of L may lead to significant performance degradation.

In the case of overestimation of L ($L_E > L$), the EM *localizer* attempts to decompose each *origin* into 2^{L_E-L} Gaussian components. As a result, spurious convergence is more likely to occur, and accordingly, the performance of the OAE will be degraded (recall that the start value given in Eq. (6.15) is based on the assumption that the *origins* are clearly separated). However, in general, the effect of overestimation can be overcome by using the SCC methods introduced in Section 6.3.4.

On the contrary, in the case of underestimation ($L_E < L$), the EM *localizer* attempts to use one Gaussian density to represent 2^{L-L_E} *origins*. The underestimation effect is determined by the energy of the multipath components that have been left out of consideration. If $\sum_{i=L_E+1}^{L-1} |f_i|$ is much smaller than $|f_0|$, the underestimation of L will have little

CHAPTER 6. SEMI-PARAMETRIC ORIGIN ANALYSIS EQUALIZER

Fig. 6.15. Effect of underestimation of L . $L_E = L - 1$; the solid lines show the actual origins, and the dotted lines show the localization results.

effect on the performance of the OAE. However, when $\sum_{i=L_E+1}^{L-1} |f_i| > f_0$, the consequence can be disastrous.

The effect of underestimation is illustrated in Fig. 6.15, where only one COP is shown and L_E is assumed to be $L - 1$. As can be observed, in this case, the underestimation of L introduces a constant localization error $|f_{L-1}|$. Accordingly, as $|f_{L-1}|$ increases, the performance degradation will become more serious. In the extreme case when $|f_{L-1}|$ is much greater than $|f_0|$, the symbol decisions will have an error rate of up to 0.5 if the previous decision is erroneous (see Fig. 6.11 for a similar scenario). Fig. 6.15 can also be regarded as an illustration of the effect of overestimation of L , in which case, however, a large localization error is highly unlikely to occur.

Comparatively, the underestimation of L is more disturbing and intractable. Therefore, when it is necessary to determine L based on empirical knowledge, relatively large values should be considered first (the complexity issue is not considered here).

6.8 Computer Simulations

In this section, we will investigate the performance of the semi-parametric OAE in some relatively realistic scenarios. As in the previous chapter, each simulation consists of 50 repetitions of the same process at given SNR. Within each process, a random sequence of 1×10^4 symbols is transmitted repeatedly until at least 50 detection errors have been recorded. The OAE performs *origin analysis* based on the first n samples, and maintains the results throughout the remaining process. The LE, with 41 taps, and the DFE, with feedforward and feedback orders of 21, use the first 500 symbols as the training sequence for channel estimation. The MLSE, with trellis length of 30, is assumed to have perfect knowledge of the CIR.

To illustrate the advantages of the semi-parametric *origin localization* approach, let us first consider the channel used in Fig. 5.10, i.e., the CIR is $[1, -0.5, 0.7]^T$. Fig. 6.16 shows the BER performance of the LE, DFE, MLSE and semi-parametric OAE approaches, where the sample size n is set to 100. Compared to the nonparametric OAE, whose BER performance is given in Fig. 5.10, the semi-parametric OAE can reduce the BER noticeably with much less observation time. Furthermore, it does not rely on the shape of the underlying distribution, and thus can work in the presence of strong noise.

The channel models used in the following simulations are consistent with the COST 207 standard [94]. Fig. 6.17 considers a typical urban (TU) propagation, where the CIR is assumed to be $[1, -0.66]^T$ and n is set to 100. Fig. 6.18 considers a bad urban (BU) environment, where the CIR is assumed to be $[1, 0.21, 0.78]^T$ and n is set to 200. Fig. 6.19 uses a hilly terrain (HT) channel, where the CIR is assumed to be $[1, 0, 0, 0, -0.52, 0.17]^T$ and n is set to 200. In Fig. 6.19 the number of the trivial CIR coefficients has been assumed *a priori*. Accordingly, the *origin localizer* needs to localize 2^3 *origins* only. If this

CHAPTER 6. SEMI-PARAMETRIC ORIGIN ANALYSIS EQUALIZER

---

parameter is unknown, the *localizer* has to estimate 2^6 origin locations. The dashed line in Fig. 6.19 shows the results of using 800 samples to perform full-size *origin localization*. As can be observed, although more samples have been used, the BER performance of the OAE is significantly degraded⁴.

The performance of the OASE is illustrated in Fig. 6.20 and 6.21. In Fig. 6.20, the CIR is assumed to be $[-0.73, 1]^T$, n is set to 100, and the solid lines, from the top down, are the performance of the OASE with the traceback length $K = 1, 3, 5$ and 10 respectively. In Fig. 6.21, a relatively long CIR, $[0.9978, 0.9185, 0.7304, 0.4881, 0.2674, 0.1112, 0.0310]^T$ [71], is considered, n is set to 1000, and the solid lines, from the top down, correspond to $K = 1, 3$ and 5 respectively. As can be observed, in Fig. 6.21 the OASE can dramatically reduce the computational complexity (with $K = 5$) at the cost of a slight loss of performance, while in Fig. 6.20 the complexity and performance tradeoff is less appealing. It follows the conclusion that the OASE can offer an excellent balance between complexity and performance when the postcursor response is relatively long.

Fig. 6.22 shows the effect of underestimation of L on the performance of the OAE. In Fig. 6.22(a) and (b), the CIRs are assumed to be $[0.9978, 0.9185, 0.7304, 0.4881, 0.2674, 0.1112, 0.0310]^T$ and $[0.9978, 0.9185, 0.7304, 0.4881, 0.1112, 0.0310, 0.2674]^T$ respectively, the channel length L is set to 5, n is set to 1000, the solid lines, from the top down, are respectively the performance of the OASE with $K = 1, 3$ and 5, and the dashed lines are the results obtained with $L = 7$ and $K = 5$. As can be seen from Fig. 6.22(a),

---

⁴Actually, the performance degradation is caused by inappropriate start value for the EM localization. Recall that Eq. (6.15) assumes that the *origins* are clearly separated. As such, when *origin* overlap occurs, the EM localization is highly likely to converge to spurious maximum. This problem can be solved by using a nonparametric density estimate as the reference in spurious convergence control (see Section 6.3.4).

if the energy of the multipaths that have been neglected is quite low, the performance degradation caused by the underestimation of L is not obvious. However, if the energy is relatively high, the degradation can be serious (Fig. 6.22(b)). The effect of overestimation of L has already been illustrated in Fig. 6.19, where the dashed line shows the performance of the OAE with $L = 6$ while the actual number of multipath components is three. It is reminded that the degradation of performance caused by the overestimation of L can be overcome by implementing SCC (see Footnote 4).

The effectiveness of the OAE for M -ary modulation is confirmed in Fig. 6.23, where 4-QAM is considered, the CIR is assumed to be $[-0.72 + 1.52j, -0.56 + 1.62j, 0.51 - 0.76j]^T$, n is set to 500, and the solid lines, from the top down, are respectively the performance of the OASE with $K = 1, 3$ and 5.

6.9 Summary

This chapter investigates the possibility of using semi-parametric density estimate, or to be more specific, finite mixture model, to improve the performance of the proposed OAE algorithm. Compared with the nonparametric *origin analysis* approach introduced in the previous chapter, the semi-parametric approach can greatly improve the *origin localization* accuracy while at the same time reducing the observation time considerably. Furthermore, not relying on the shape of the underlying distribution to perform *origin analysis*, the semi-parametric approach is less sensitive to noise energy.

As a way to overcome error propagation, sequence estimation has been studied in the framework of OAE. An approximate performance analysis has also been conducted, for OAE based on symbol and sequence detections respectively. Computer simulations show that the OASE can achieve an excellent balance between complexity and performance.

CHAPTER 6. SEMI-PARAMETRIC ORIGIN ANALYSIS EQUALIZER

Fig. 6.16. BER performance of semi-parametric *origin analysis equalizer* for a channel with impulse response $[1, -0.5, 0.7]^T$.

Fig. 6.17. BER performance of semi-parametric *origin analysis equalizer* for a COST 207 TU channel with impulse response $[1, -0.66]^T$.

CHAPTER 6. SEMI-PARAMETRIC ORIGIN ANALYSIS EQUALIZER

Fig. 6.18. BER performance of semi-parametric *origin analysis equalizer* for a COST 207 BU channel with impulse response $[1, 0.21, 0.78]^T$.

Fig. 6.19. BER performance of semi-parametric *origin analysis equalizer* for a COST 207 HT channel with impulse response $[1, 0, 0, 0, -0.52, 0.17]^T$.

CHAPTER 6. SEMI-PARAMETRIC ORIGIN ANALYSIS EQUALIZER

Fig. 6.20. BER performance of semi-parametric *origin analysis equalizer* for a COST 207 TU channel with impulse response $[-0.73, 1]^T$.

Fig. 6.21. BER performance of semi-parametric *origin analysis equalizer* for a channel with impulse response $[0.9978, 0.9185, 0.7304, 0.4881, 0.2674, 0.1112, 0.0310]^T$.

CHAPTER 6. SEMI-PARAMETRIC ORIGIN ANALYSIS EQUALIZER

(a)

(b)

Fig. 6.22. Effect of underestimation of channel length. (a) CIR= [0.9978, 0.9185, 0.7304, 0.4881, 0.2674, 0.1112, 0.0310]^T; (b) CIR= [0.9978, 0.9185, 0.7304, 0.4881, 0.1112, 0.0310, 0.2674]^T. In both cases, $L = 5$.

CHAPTER 6. SEMI-PARAMETRIC ORIGIN ANALYSIS EQUALIZER

Fig. 6.23. Performance of semi-parametric *origin analysis equalizer* for 4-QAM; CIR $= [-0.72 + 1.52j, -0.56 + 1.62j, 0.51 - 0.76j]^T$.

CHAPTER 6. SEMI-PARAMETRIC ORIGIN ANALYSIS EQUALIZER

Finally, it should be mentioned that, although the tapped delay line model has been assumed throughout the discussions, the proposed OAE approach can also be applied to sparse channels. In general, a sparse channel can be represented as a CIR with a number of zero coefficients. Therefore, the problem encountered in sparse channel equalization is similar to the problem of overestimating the channel length, which has been discussed in Section 6.7.3. It is reminded that in Fig. 6.19, we have already illustrated the performance of the OAE for a sparse channel.

Chapter 7

Conclusions and Recommendations

7.1 Conclusions

Parametric density estimates are widely used in telecommunications to represent the channel statistics. However, in many practical problems, a parametric model is unable to reveal all features of interest. The main objective of this thesis work is to investigate the properties and possible applications of the semi- and non-parametric density estimation techniques to communications for problems that cannot be handled easily by the parametric approaches. Two particular problems have been investigated, namely, the *a posteriori* ER estimation and the ISI mitigation.

The ultimate goal of *a posteriori* ER estimation to provide reliable and fast-responsive blind ER estimate for system monitoring purposes. The most straightforward method of counting the transmission errors is extremely data-inefficient in the sense that only the erroneous events are used to derive the ER estimate. As such, it requires a very long observation time. Moreover, this approach cannot work with blind estimation. Several authors have already demonstrated how the received signal can be more efficiently utilized through the use of nonparametric density approximation technique. However, the

CHAPTER 7. CONCLUSIONS AND RECOMMENDATIONS

---

optimization of the density estimator for ER estimation has never been addressed. In this study, we have developed an optimization algorithm that minimizes the mean square error of the ER estimate. Conventional error monitors based on density approximation also suffers from poor blind estimation capability. By extrapolating the ER from a set of pseudo-error rates, the pseudo-error monitoring approach can overcome the dependence on the *a priori* knowledge of the transmitted signal. However, for fast time-varying systems, its data efficiency is inadequate. By combining the strengths of both the density approximation and the ER extrapolation techniques, we have proposed a new error monitoring approach. Computer simulations have shown that the proposed approach is able to provide better performance than existing methods.

The second problem is concerned with the ISI. It is well known that the presence of ISI can seriously distort the transmitted signal, leading to great degradation in BER performance. Conventional approaches to ISI mitigation, e.g. the LE and the DFE algorithms, rely on accurate estimation of the CIR. Reliable CIR estimate is usually obtained through a training process, which can cause a significant reduction in bandwidth efficiency. To solve the difficulties associated with CIR estimation, we have proposed a blind channel estimation approach, the origin analysis approach, and based on that, developed a new channel equalization algorithm. The origin analysis consists of two processes. The first process, termed origin localization, attempts to provide a set of parameters that are directly related to the amplitude of the CIR, while the second process, termed origin identification, is responsible for the retrieval of the phase information. Origin localization can be achieved by non- or semi-parametric density estimation. Both solutions have been studied extensively, and the semi-parametric technique is found to be more suitable for the task of modeling an ISI channel for ISI mitigation. As demonstrated by Monte Carlo

CHAPTER 7. CONCLUSIONS AND RECOMMENDATIONS

---

simulations, the proposed equalizer based on semi-parametric channel modeling has the potential to provide better BER performance than the LE and the DFE.

Comparatively, the nonparametric density estimate is more reliable in representing the underlying distribution. Thus, such estimate is more appropriate in cases where high accuracy is required, e.g., in *a posteriori* ER estimation. On the other hand, since the nonparametric estimate does not have a closed-form expression, the semi-parametric estimate may be a better choice if the details of the shape of the density function is of little concern, e.g., in ISI mitigation.

7.2 Recommendations for Future Research

In this study, we have investigated two applications of the density estimation techniques in telecommunications. Some directions for future research are as follows.

- The optimization approach presented in Chapter 3 is designed to minimize the MSE of the ER estimation. However, since the performance of a communication system is generally a nonlinear function of the ER, the MSE criterion, which equally attends to the estimation bias and variance, may not fit well into the context of performance evaluation. As a result, the optimal ER estimate in the sense of MSE may not always be a good performance index. The system performance is usually treated as a linear function of the logarithm of the ER. Thus, we can design the kernel ER estimator to minimize the MSE of the estimation of the logarithmic ER. An alternative criterion is the probability of decision error of the error monitor, that is, the probability that an unacceptable performance is identified as being acceptable and an acceptable performance is identified as being unacceptable.

CHAPTER 7. CONCLUSIONS AND RECOMMENDATIONS

- The kernel pseudo-error monitor design introduced in Chapter 4 has been optimized to minimize the MSE of the estimate of the smaller PER. As mentioned above, by using other optimization criteria, better monitoring performance may be achieved. Moreover, the parameters that affect the accuracy of error extrapolation should also be taken into account in the development of the optimization algorithm.
- In Chapter 6, it has been shown that the exploitation of the equal distance property of origin distribution degrades the performance of the EM origin localizer. However, we can use this property to postprocess the origin location estimates so as to achieve better localization. Theoretically, in this way, the sample size required to satisfy a given localization accuracy can be reduced by a factor of M^{L-1}/L .
- In Chapter 5 and 6, we have assumed that the number of component distributions, or equivalently, the number of resolvable multipath components L , is known *a priori*. However, this parameter is seldom available in practice. Although it has been shown by simulations that an empirical choice of L is normally sufficient, a better solution to this problem is to incorporate an estimator of the number of mixing components into the EM localizer (see, e.g., [20] and the references therein).
- In Chapter 5 and 6, the component distribution has been assumed to be Gaussian. When dealing with a non-Gaussian noise, the proposed OAE algorithm unavoidably suffers from performance degradation. To obviate this problem, we can implement a nonparametric density estimator for the component distribution, and use the result to modify the EM localization algorithm accordingly.
- The proposed origin analysis process is essentially a blind channel estimation method. From the results of *origin analysis*, a CIR estimate can be easily derived.

CHAPTER 7. CONCLUSIONS AND RECOMMENDATIONS

Therefore, it can be suitably combined with conventional channel equalization approaches, such as MLSE and DFE, to form more effective ISI mitigation techniques. It would be interesting to investigate the performance of using the *origin analysis* approach as a CIR estimator.

- In this study, the density estimation techniques have been used to improve the data efficiency in *a posteriori* ER estimation and to provide channel information for ISI mitigation. It should be reminded that these techniques have great potential to be applied in other fields in communications. For example, it is possible to modify the OAE algorithm to deal with other sources of interferences, such as cochannel interference and adjacent channel interference. Development of new applications for the density estimation techniques should be pursued further.

Author's Publications

Journal Papers

- [1]. Peng Wang and Wee Ser, "Origin analysis—a novel channel equalization approach," to be submitted to *IEEE Transactions on Communications*.
- [2]. Peng Wang and Wee Ser, "Adaptive origin analysis equalization" submitted to *IEEE Transactions on Signal Processing*.
- [3]. Peng Wang and Wee Ser, "Channel estimation and equalization via origin analysis," *EURASIP Journal on Applied Signal Processing* (in revision).
- [4]. Peng Wang and Wee Ser, "Performance optimization for kernel error monitoring," *WSEAS Transactions on Communications*, vol. 4, no. 9, pp. 858-865, Sep. 2005.
- [5]. Peng Wang and Wee Ser, "A novel pseudoerror monitor," *EURASIP Journal on Applied Signal Processing*, vol. 2004, no. 3, pp. 412-417, Mar. 2004.

Conference Papers

- [1]. Peng Wang and Wee Ser, "A novel equalization algorithm for frequency-selective channels," to appear in *Proc. International Symposium on Information Theory and its Applications 2006*, Seoul, Korea, Oct. 2006.

CHAPTER 7. CONCLUSIONS AND RECOMMENDATIONS

- [2]. Peng Wang and Wee Ser, “Origin analysis for blind equalization,” to appear in *Proc. 9th Annual International Symposium on Wireless Personal Multimedia Communications*, San Diego, CA, Sep. 2006.
- [3]. Peng Wang and Wee Ser, “Optimum kernel error rate estimation,” in *Proc. 9th WSEAS International Conference on Communications*, Athens, Greece, July 2005.
- [4]. Peng Wang and Wee Ser, “A novel statistical approach to ISI mitigation,” in *Proc. 4th IEEE Wireless Telecommunications Symposium*, pp. 155-161, Pomona, CA, Apr. 2005.
- [5]. Peng Wang and Wee Ser, “Kernel pseudo-error monitor,” in *Proc. 7th WSEAS International Conference on Communications*, Corfu Island, Greece, July 2003. (also appears in *Recent Advances in Communications and Computer Science*, pp. 1-5, WSEAS Press, 2003.)

Bibliography

- [1] J. D. Laster, “Robust GMSK demodulation using demodulator diversity and BER estimation,” Ph.D. dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA, 1997.
- [2] J. D. Laster, J. H. Reed, and W. H. Tranter, “Bit error rate estimation using probability density function estimators,” *IEEE Trans. Veh. Technol.*, vol. 51, no. 1, pp. 260–267, Jan. 2003.
- [3] K. A. Phillips, “Probability density function estimation applied to minimum bit error rate adaptive filtering,” Master’s thesis, Virginia Polytechnic Institute and State University, Blacksburg, VA, 1999.
- [4] K. A. Phillips, J. H. Reed, and W. H. Tranter, “Minimum BER adaptive filtering,” in *Proc. IEEE Int. Conf. Commun.*, vol. 3, New Orleans, LA, June 2000, pp. 1675–1680.
- [5] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. New York, NY: Chapman and Hall, 1986.
- [6] E. Parzen, “On estimation of a probability density function and mode,” *Ann. Math. Statist.*, vol. 33, no. 3, pp. 1065–1076, Sept. 1962.

BIBLIOGRAPHY

- [7] M. E. Tarter and R. A. Kronmal, “An introduction to the implementation and theory of nonparametric density estimation,” *Amer. Statist.*, vol. 30, no. 3, pp. 105–112, Aug. 1976.
- [8] D. W. Scott, “On optimal and data-based histograms,” *Biometrika*, vol. 66, no. 3, pp. 605–610, Dec. 1976.
- [9] M. P. Wand, “Data-based choice of histogram bin width,” *Amer. Statist.*, vol. 51, no. 1, pp. 59–64, Feb. 1997.
- [10] G. R. Terrell and D. W. Scott, “Oversmoothed nonparametric density estimates,” *J. Amer. Statist. Assoc.*, vol. 80, no. 389, pp. 209–214, Mar. 1985.
- [11] M. Rosenblatt, “Remarks on some nonparametric estimates of a density function,” *Ann. Math. Statist.*, vol. 27, no. 3, pp. 832–837, Sept. 1956.
- [12] A. J. Izeman, “Recent development in nonparametric density estimation,” *J. Amer. Statist. Assoc.*, vol. 86, no. 413, pp. 205–224, Mar. 1991.
- [13] D. O. Loftsgaarden and C. P. Quesenberry, “A nonparametric estimate of a multivariate density function,” *Ann. Math. Statist.*, vol. 36, no. 3, pp. 1049–1051, June 1965.
- [14] D. S. Moore and J. W. Yackel, “Consistency properties of nearest neighbor density function estimators,” *Ann. Statist.*, vol. 5, no. 1, pp. 143–154, Jan. 1977.
- [15] L. Breiman, W. Meisel, and E. Purcell, “Variable kernel estimates of multivariate densities,” *Technometrics*, vol. 19, no. 2, pp. 135–144, May 1977.
- [16] G. R. Terrell and D. W. Scott, “Verkenel kernel density estimation,” *Ann. Statist.*, vol. 20, no. 3, pp. 1236–1265, Sept. 1992.

BIBLIOGRAPHY

- [17] R. P. Boas, “Representation of probability distributions by charlier series,” *Ann. Math. Statist.*, vol. 20, no. 3, pp. 376–392, Sept. 1949.
- [18] G. L. Anderson and R. J. P. de Figueiredo, “An adaptive orthogonal-series estimator for probability density function,” *Ann. Statist.*, vol. 8, no. 2, pp. 347–376, Mar. 1980.
- [19] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*. Hoboken, NJ: John Wiley & Sons, 2003.
- [20] G. J. McLachlan and D. Peel, *Finite Mixture Models*. New York: John Wiley & Sons, 2000.
- [21] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. New York: John Wiley & Sons, 1997.
- [22] M. Watanabe and K. Yamaguchi, Eds., *The EM Algorithm and Related Statistical Models*. New York: Marcel Dekker, 2004.
- [23] R. A. Redner and H. F. Walker, “Mixture densities, maximum likelihood and the em algorithm,” *SIAM Rev.*, vol. 26, no. 2, pp. 195–239, 1984.
- [24] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *J. R. Statist. Soc. B*, vol. 39, no. 1, pp. 1–38, 1977.
- [25] X. L. Meng and D. Dyk, “The EM algorithm—an old folk song sung to a fast new tune,” *J. R. Statist. Soc. B*, vol. 59, no. 3, pp. 511–567, 1997.
- [26] *Determination of the Performance of Data Communication Systems*, ANSI Std. X3.44, 1974.

BIBLIOGRAPHY

- [27] P. M. McManamon, R. K. Rosich, J. A. Payne, and M. J. Miles, "Performance criteria for digital networks," U.S. Dept. Of Commerce, Washington, DC, Office of Telecommun. Rep. 75-54, Jan. 1975.
- [28] N. B. Seitz and D. Bodson, "Data communication performance assessment," *Telecommun.*, vol. 14, no. 2, pp. 35–89, Feb. 1980.
- [29] J. G. Proakis, *Digital Communications*, 3rd ed. New York, NY: McGraw-Hill, 1995.
- [30] A. F. Molisch, Ed., *Wideband Wireless Digital Communications*. Upper Saddle River, NJ: Prentice-Hall, 2001.
- [31] J. G. Proakis, "Probabilities of error for adaptive reception of M -phase signals," *IEEE Trans. Commun. Technol.*, vol. 16, no. 1, pp. 71–81, Feb. 1968.
- [32] Y. Miyagaki, N. Morinaga, and T. Namekawa, "Error probability characteristics for CPSK signal through m -distributed fading channel," *IEEE Trans. Commun. Technol.*, vol. 26, no. 1, pp. 88–100, Jan. 1978.
- [33] S. Chennakeshu and J. B. Anderson, "Error rates for Rayleigh fading multichannel reception of MPSK signals," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 338–346, Feb./Mar./Apr. 1995.
- [34] M. G. Shayesteh and A. Aghamohammadi, "On the error probability of linearly modulated signals on frequency-flat Ricean, Rayleigh, and AWGN channels," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 1454–1466, Feb./Mar./Apr. 1995.
- [35] M. K. Simon and M. S. Alouini, "A unified approach to the probability of error for noncoherent and differentially coherent modulations over generalized fading channels," *IEEE Trans. Commun.*, vol. 46, no. 12, pp. 1625–1638, Dec. 1998.

BIBLIOGRAPHY

- [36] M. K. Simon and M. S. Alouini, Eds., *Digital Communications Over Generalized Fading Channels: A Unified Approach to Performance Analysis*. New York, NY: John Wiley & Sons, 2000.
- [37] J. Zhang, “Performance analysis for coded CDMA systems,” *IEEE Trans. Inform. Theory*, vol. 49, no. 12, pp. 3282–3288, Dec. 2003.
- [38] E. A. Newcombe and S. Pasupathy, “Error rate monitoring for digital communications,” *Proc. IEEE*, vol. 70, no. 8, pp. 805–822, Aug. 1982.
- [39] J. R. Raol, G. Girija, and J. Singh, *Modelling and Parameter Estimation of Dynamic Systems*. London, United Kingdom: Institution of Electrical Engineers, 2004.
- [40] D. C. Montgomery and G. C. Runger, *Applied Statistics and Probability for Engineers*, 3rd ed. New York, NY: John Wiley & Sons, 2003.
- [41] L. C. Scholley and J. W. McCurnin, “Measurement of bit errors in data transmission systems,” *IEEE Trans. Commun.*, vol. 24, no. 1, pp. 144–146, Jan. 1976.
- [42] D. Zwillinger, Ed., *CRC Mathematical Tables and Formulae*, 31st ed. Boca Raton, FL: CRC Press, 2003.
- [43] H. Dette, A. Munk, and T. Wagner, “Estimating the variance in nonparametric regression—what is a reasonable choice?” *J. R. Statist. Soc. B*, vol. 60, no. 4, pp. 751–764, 1998.
- [44] S. E. Portny, “Large sample confidence limits for binary error probabilities,” *Proc. IEEE*, vol. 54, no. 12, p. 1993, Dec. 1966.

BIBLIOGRAPHY

- [45] D. J. Gooding, "Performance monitor technique for digital receivers based on extrapolation of error rate," *IEEE Trans. Commun. Technol.*, vol. 16, no. 3, pp. 380–387, Mar. 1968.
- [46] S. B. Weinstein, "Estimation of small probabilities by linearization of the tail of a probability distribution function," *IEEE Trans. Commun. Technol.*, vol. 19, no. 6, pp. 1149–1155, June 1971.
- [47] B. J. L. *et al.*, "A bit error rate monitor for digital PSK links," *IEEE Trans. Commun.*, vol. 23, no. 5, pp. 518–525, May 1975.
- [48] I. M. Kostić, "Pseudo error rate of a PSK system with hardware imperfections, noise and cochannel interference," *IEE Proc. -I*, vol. 136, no. 5, pp. 333–338, Oct. 1989.
- [49] J. M. Keelty and J. Feher, "On line pseudo-error monitors for digital transmission systems," *IEEE Trans. Commun.*, vol. 26, no. 8, pp. 1275–1282, Aug. 1978.
- [50] E. A. Newcombe and S. Pasupathy, "Error rate monitoring in a partial response system," *IEEE Trans. Commun.*, vol. 28, no. 7, pp. 1052–1061, July 1980.
- [51] T. A. Fitch and M. L. Steinberger, "Eye violation detectors," U.S. Patent 4 633 465, 1986.
- [52] T. J. Nohara, A. Premji, and W. R. Seed, "A new signal quality degradation monitor for digital transmission channels," *IEEE Trans. Commun.*, vol. 43, no. 2, pp. 1333–1336, Feb. 1995.
- [53] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inform. Theory*, vol. 13, no. 2, pp. 260–269, Apr. 1967.

BIBLIOGRAPHY

- [54] G. D. Forney, "The Viterbi algorithm," *Proc. IEEE*, vol. 61, no. 3, pp. 268–278, Mar. 1973.
- [55] ———, "Maximu-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. 18, no. 3, pp. 363–378, May 1972.
- [56] B. D. Hart and D. P. Taylor, "Extended MLSE diversity receiver for the time- and frequency-selective channel," *IEEE Trans. Commun.*, vol. 45, no. 3, pp. 322–333, Mar. 1997.
- [57] R. W. Lucky, "Techniques for adaptive equalization of digital communication systems," *Bell Syst. Tech. J.*, vol. 45, no. 2, pp. 255–268, Feb. 1966.
- [58] B. Widrow, "Adaptive filters, I: Fundamentals," Stanford Electronics Laboratory, Stanford University, Standford, CA, Tech. Rep. 6764-6, Dec. 1966.
- [59] M. Barton and D. W. Tufts, "A suboptimum linear receiver based on a parametric channel model," *IEEE Trans. Commun.*, vol. 39, pp. 1328–1334, Sept. 1991.
- [60] M. E. Austin, "Decision feedback equalization for digital communication over dispersive channels," Lincoln Laboratory, MIT, Cambridge, MA, Tech. Rep. 461, Aug. 1967.
- [61] D. A. George, R. R. Bowen, and J. R. Storey, "An adapive decision feedback equalizer," *IEEE Trans. Commun. Technol.*, vol. 19, no. 3, pp. 281–293, June 1971.
- [62] C. A. Belfiore and J. H. Park, "Decision-feedback equalization," *Proc. IEEE*, vol. 67, no. 8, pp. 1143–1156, Aug. 1979.

BIBLIOGRAPHY

- [63] D. Williamson, R. A. Kennedy, and G. W. Pulford, "Block decision feedback equalization," *IEEE Trans. Commun.*, vol. 40, no. 2, pp. 255–264, Feb. 1992.
- [64] S. Ariyavisitakul, "A decision feedback equalizer with time-reversal structure," *IEEE J. Select. Areas Commun.*, vol. 10, no. 4, pp. 599–613, Apr. 1992.
- [65] G. K. Kaleh, "Channel equalization for block transmission systems," *IEEE J. Select. Areas Commun.*, vol. 13, no. 1, pp. 110–121, Jan. 1995.
- [66] N. Al-Dhahir and A. H. Sayed, "The finite-length multi-input multi-output MMSE-DFE," *IEEE Trans. Signal Processing*, vol. 48, no. 10, pp. 2921–2936, Oct. 2000.
- [67] D. Mattera, L. Paura, and F. Sterle, "Widely linear decision-feedback equalizer for time-dispersive linear MIMO channels," *IEEE Trans. Signal Processing*, vol. 53, no. 7, pp. 2525–2536, July 2005.
- [68] M. V. Eyuboglu and S. U. H. Qureshi, "Reduced-state sequence estimation with set partitioning and decision feedback," *IEEE Trans. Commun.*, vol. 36, no. 1, pp. 12–20, Jan. 1988.
- [69] A. Duel-Hallen and C. Heegard, "Delayed decision-feedback sequence estimation," *IEEE Trans. Commun.*, vol. 37, no. 5, pp. 428–436, May 1989.
- [70] S. Ariyavisitakul and L. J. Greenstein, "Reduced-complexity equalization techniques for broadband wireless channels," *IEEE J. Select. Areas Commun.*, vol. 15, no. 1, pp. 5–15, Jan. 1997.
- [71] M. Magarini, A. Spalvieri, and G. Tartara, "The mean-square delayed decision feedback sequence detector," *IEEE Trans. Commun.*, vol. 50, no. 9, pp. 1462–1470, Sept. 2002.

BIBLIOGRAPHY

- [72] J. K. Tugnait, L. Tong, and Z. Ding, "Single-user channel estimation and equalization," *IEEE Signal Processing Mag.*, vol. 17, no. 3, pp. 17–28, May 2000.
- [73] S. N. Crozier, D. D. Falconer, and S. A. Mahmoud, "Least sum of squared errors (LSSE) channel estimation," *IEE Proc. -F*, vol. 138, no. 4, pp. 371–378, Aug. 1991.
- [74] H. Arslan and G. E. Bottomley, "Channel estimation in narrowband wireless communication systems," *Wirel. Commun. Mob. Comput.*, vol. 1, no. 2, pp. 201–219, Apr./Jun. 2001.
- [75] D.-Y. Chen, Y. G. Hong, and K. S. Trivedi, "Optimal estimation of training interval for channel equalization," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1844–1853, Sept. 2004.
- [76] H. Liu, G. Xu, L. Tong, and T. Kailath, "Recent developments in blind channel equalization: From cyclostationarity to subspaces," *Signal Process.*, vol. 50, no. 1-2, pp. 83–99, Apr. 1996.
- [77] K. Abed-Meraim, W. Qiu, and Y. Hua, "Blind system identification," *Proc. IEEE*, vol. 85, no. 8, pp. 1310–1322, Aug. 1997.
- [78] D. Boss, K.-D. Kammeyer, and T. Petermann, "Is blind channel estimation feasible in mobile communication systems? A study based on GSM," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1479 – 1492, Oct. 1998.
- [79] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: A time domain approach," *IEEE Trans. Inform. Theory*, vol. 40, no. 2, pp. 340–349, Mar. 1994.

BIBLIOGRAPHY

- [80] G. K. Kaleh and R. Vallet, "Joint parameter estimation and symbol detection for linear or nonlinear unknown channels," *IEEE Trans. Commun.*, vol. 42, no. 7, pp. 1514–1525, July 1994.
- [81] M. K. Tsatsanis and G. B. Giannakis, "Equalization of rapidly fading channels: Self-recovering methods," *IEEE Trans. Commun.*, vol. 44, no. 5, pp. 619–630, May 1996.
- [82] M. C. Necker and G. L. Stüber, "Totally blind channel estimation for OFDM on fast varying mobile radio channels," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1514–1525, Sept. 2004.
- [83] D. A. Geoge, R. R. Bowen, and J. R. Storey, "An adaptive decision feedback equalizer," *IEEE Trans. Commun. Technol.*, vol. 19, no. 3, pp. 281–293, June 1971.
- [84] P. Monsen, "Adaptive equalization of the slow fading channel," *IEEE Trans. Commun.*, vol. 22, no. 8, pp. 1064–1075, Aug. 1974.
- [85] S. U. H. Qureshi, "Adaptive equalization," *Proc. IEEE*, vol. 73, no. 9, pp. 1349–1387, Sept. 1985.
- [86] J. G. Proakis, "Adaptive equalization for TDMA digital mobile radio," *IEEE Trans. Veh. Technol.*, vol. 40, no. 2, pp. 333–341, May 1991.
- [87] H. Kubo, K. Murakami, and T. Fujino, "An adaptive maximum-likelihood sequence estimator for fast time-varying intersymbol interference channels," *IEEE Trans. Commun.*, vol. 42, no. 2/3/4, pp. 1872–1880, Feb./Mar./Apr. 1994.
- [88] F. B. Ueng and Y. T. Su, "Adaptive blind equalization using second- and higher order statistics," *IEEE J. Select. Areas Commun.*, vol. 13, no. 1, pp. 132–140, Jan. 1995.

BIBLIOGRAPHY

- [89] R. Raheli, A. Polydoros, and C.-K. Tzou, "Per-survivor processing: A general approach to MLSE in uncertain environments," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 354–364, Feb./Mar./Apr. 1995.
- [90] G. Castellini, F. Conti, E. D. Re, and L. Pierucci, "A continuously adaptive MLSE receiver for mobile communications: Algorithm and performance," *IEEE Trans. Commun.*, vol. 45, no. 1, pp. 80–89, Jan. 1997.
- [91] E. Baccarelli, R. Cusani, and S. Galli, "A novel adaptive receiver with enhanced channel tracking capability for TDMA-based mobile radio communications," *IEEE J. Select. Areas Commun.*, vol. 16, no. 9, pp. 1630–1639, Dec. 1998.
- [92] H. Zamiri-Jafarian and S. Pasupathy, "Adaptive MLSDE using the EM algorithm," *IEEE Trans. Commun.*, vol. 47, no. 8, pp. 1181–1193, Aug. 1999.
- [93] C.-C. Yeh and J. R. Barry, "Adaptive minimum bit-error rate equalization for binary signaling," *IEEE Trans. Commun.*, vol. 48, no. 7, pp. 1226–1235, July 2000.
- [94] *COST 207: Digital land mobile radio communications: final report*, Office for Official Publications of the European Communities, 1989.