Adaptive MLSDE Receivers for Wireless Communications

by

Hossein Zamiri-Jafarian

A thesis submitted in conformity with the requirements
for the Degree of Doctor of Philosophy,
Department of Electrical and Computer Engineering,
at the University of Toronto

Adaptive MLSDE Receivers for Wireless Communications

by

Hossein Zamiri-Jafarian

Ph.D. Thesis, 1998

Department of Electrical and Computer Engineering

University of Toronto

ABSTRACT

This thesis presents the structural design, performance evaluation and complexity reduction of adaptive maximum likelihood sequence detection and estimation (MLSDE) receivers for wireless communications.

The receiver structure is developed based on the maximum likelihood (ML) criterion for joint channel estimation and data detection and using the expectation and maximization (EM) algorithm. Generalized MLSDE (GMLSDE) algorithm, which alternates between estimation and detection and at the same time increases the likelihood iteratively, is developed using the on-line EM algorithm. The GMLSDE provides the theoretical ML base from which per-survivor processing (using different estimators for each survivor path) and conventional channel estimation (using only one estimator for all survivor paths) can be deduced for different channel models.

Numerous adaptive MLSDE receiver structures are developed for different channel models from the GMLSDE algorithm using causal estimation and detection methods which only guarantee increasing the likelihood. Although some structures are new, the method of deriving all the structures shows the power of the GMLSDE algorithm in unifying different ML-based joint channel estimation and data detection receivers.

Data detection part in the adaptive MLSDE receiver is implemented by the Viterbi algorithm while its estimation part is based on a modified Titterington's stochastic approximation method. The estimation part leads to a novel RLS/Kalman-type algorithm when the channel impulse response (CIR) is a deterministic process embedded in an additive colored Gaussian noise.

The bit error rate performances of the adaptive MLSDE receiver for a differentially-encoded quadrature phase shift keying signal transmitted over the frequency flat and selective Rayleigh fading channels are evaluated by simulations for different fading rates when the CIR is modeled at the receiver by deterministic and Gaussian random vectors/processes. Comparison between the performances shows that in many situations using the deterministic CIR model is sufficient.

The thesis proposes the adaptive state allocation (ASA) algorithm which uses adaptive threshold and adaptive state partitioning methods to reduce the receiver complexity at high and low signal-to-noise ratios respectively. In the adaptive threshold method, a threshold value is formulated to choose the more likely correct states at each time and in the adaptive state partitioning method, branch metrics of the trellis diagram are fused/diffused adaptively. Results for multipath Rayleigh fading channels show that the ASA algorithm greatly reduces the receiver complexity with negligible performance degradation.

Acknowledgments

I wish to express my sincere gratitude and appreciation to my supervisor, Professor S. Pasupathy, for his insightful guidance, invaluable advice and great encouragement throughout my research. I thank him deeply for his patience and for providing support during the exploration of my ideas.

I would like to thank Professor D. Falconer, external Ph.D. examiner, Professor D. Hatzinakos, internal thesis appraiser, and the other members of my Ph.D. committee, Professors D. Johns, E. S. Sousa, A. N. Venetsanopoulos, M. Eizenman and A. Leon-Garcia, for their invaluable comments.

I am also grateful to all the other professors from whom I have learned a lot during my studies at the University of Toronto and the Isfahan University of Technology.

Thanks should also go to the Communication Group students, especially Iranian students, who have created a friendly atmosphere which made my graduate work at the University of Toronto most enjoyable and beneficial.

I gratefully acknowledge the financial support of the Iranian Ministry of Culture and Higher Education and the Ferdowsi University through a graduate scholarship. As well, the financial support by CITR, ITRC and NSERC of Canada is highly appreciated.

I offer my heartfelt thanks to my wife, Parvin, for her intense support, endless patience and great devotion. Indeed, this journey would not have been possible without her encouragement and understanding and I am eternally indebted to her for what she has sacrificed during this work. The influence of my daughter, Yeganeh, and my son, Mohammadkia, is present in this thesis; they have filled my moments at home with so much joy.

I cannot complete these acknowledgements without expressing how appreciative I am of all the love and affection that my parents, especially my mother, have provided throughout my life. I could have never made it this far without them.

To the People of Iran.

Contents

List of Figures

List of Tables

List of Symbols

a_i	transmitted symbol	14
$\mathbf{a}$	vector of transmitted symbols	17
$\boldsymbol{a}_k$	vector of transmitted symbols which affect $y(k)$	18
$\mathcal{A}$	set of all possible transmitted symbol sequences	17
$\hat{\mathbf{a}}$	estimation of $\mathbf{a}$	17
b_i	upsampled transmitted symbol	16
B	bandwidth of transmitted signal, $s(t)$	14
B'	bandwidth of received signal	15
$c(\tau, t)$	equivalent low-pass impulse response of channel (continuous)	13
$\mathcal{C}$	complete data	33
$\mathcal{C}_k$	complete data up to time k	38
d_{min}	minimum distance error event	82
$d(\beta^u, \beta^c)$	Kullback-Leibler distance between incorrect and correct branch metrics	90
$d^l(\beta^u, \beta^c)$	Kullback- Leibler distance for lth error in minimum distance event	90
$\bar{d}(\beta^u, \beta^c)$	average Kullback-Leibler distance between incorrect and correct branch metrics	90
D_r	number indicating diversity	22
$\mathcal{D}$	desired data	33
$E[x]$	expectation of x	18
$f_c(\beta)$	probability density function of correct branch metric	90
$f_u(\beta)$	probability density function of incorrect branch metric	90
f_d	maximum Doppler frequency	14

f_dT	normalized fading rate	23
$\mathbf{f}^l$	lth row of $\boldsymbol{F}$ matrix	49
$\tilde{\mathbf{f}}^l_{\mid k}$	estimation of $\mathbf{f}^l$ based on received signal up to time k	50
$\boldsymbol{F}$	$M(L+1) \times M(L+1)$ constant transition matrix	49
$\boldsymbol{F}(k)$	$M(L+1) \times M(L+1)$ transition matrix at time k	38
$\mathbf{F}_k$	$MN(L+1) \times MN(L+1)$ transition matrix at time k	40
$g(t)$	impulse response of transmitter filter (continuous)	14
$g(k)$	impulse response of transmitter filter (discrete)	16
$\mathbf{g}(k)$	$(L+1) \times (L+1)$ diagonal matrix at time k	44
$\mathbf{g}$	$(L+1) \times (L+1)$ diagonal matrix	49
$\mathbf{g}^l$	lth diagonal element of $\mathbf{g}$	49
$G(k)$	$M(L+1) \times (L+1)$ matrix	44
$h_l(t)$	equivalent of channel impulse response at the output of receiver filter (continuous)	14
$h(l,k)$	impulse response of channel (discrete)	16
$h_l(k)$	impulse response of channel (discrete)	16
$\boldsymbol{h}$	$L+1$ vector of constant channel coefficients	41
$\boldsymbol{h}(k)$	vector of channel coefficients at time k	21
$\bar{\mathbf{h}}$	vector of channel coefficients	21
$\mathbf{h}^{\diamond}_k$	$MN(L+1)$ vector of channel coefficients	38
$\mathbf{h}(k)$	$M(L+1)$ vector of channel coefficients	38
$\mathbf{h}_k$	$M(L+1)$ vector of channel coefficients	38
$\check{h}_l(t)$	channel impulse response at the output of receiver filter (continuous)	14
$\tilde{\boldsymbol{h}}_{\mid k}$	estimation of $\boldsymbol{h}$ based on received signal up to time k	41
$\tilde{\mathbf{h}}^{\diamond}_{k\mid k}$	estimation of $\mathbf{h}^{\diamond}_k$ based on received signal up to time k	40
$\tilde{\mathbf{h}}_{k\mid k}$	estimation of $\mathbf{h}_k$ based on received signal up to time k	41
I	number of transmitted symbols	14
$Im[X]$	imaginary part of X	84
$\mathcal{I}$	incomplete data	33
$\mathcal{I}_k$	incomplete data up to time k	38

J	upsampling factor (ceiling function of T/T_s)	15
J_0	zero-order Bessel function	23
L	length of channel memory	15
L_f	frame length	24
L_R	length of memory of the fading process	18
L_s	duration of impulse response of transmitter filter	17
$L(\boldsymbol{a}_k)$	unit upper triangular matrix of Cholesky decomposition of $R_{\boldsymbol{y}}$	19
$L(\boldsymbol{\theta})$	log-likelihood function	34
M	$M-1$ is the order of fading process	38
M_b	number of multiplications per branch metric	80
N	$N-1$ is the order of colored Gaussian noise	80
N_0	variance of additive Gaussian noise (discrete)	15
N_{b_k}	number of selected states at time k	81
N_m	number of multiplications per symbol	80
N_s	number of states	24
N_{s_k}	number of states at time k	81
$\hat{N}_0$	variance of additive Gaussian noise (continuous)	15
$p(x)$	probability density function of x	17
p_c^r	probability of removing the correct state	83
$\mathbf{P}_{k\|k}$	inverse of second derivative of $Q_k(.\|.)$	40
q	number of points in constellation	21
$Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(l)})$	conditional expectation of complete data	34
$r(t)$	received signal (continuous)	14
$R_x(\tau)$	autocorrelation function of $x(t)$	14
R_X	autocorrelation matrix of X	19
$Re[X]$	real part of X	84
$s(t)$	equivalent low-pass transmitted signal (continuous)	14
$s(k)$	transmitted signal (discrete)	16
$\boldsymbol{s}(k)$	$L+1$ row vector of transmitted signal at time k	22
$\boldsymbol{s}_i^c(k)$	output of transmitter filter corresponding to the correct branch	84

$\boldsymbol{s}_i^u(k)$	output of transmitter filter corresponding to the incorrect branch	84
$\mathbf{s}(k)$	$M(L+1)$ row vector of transmitted signal at time k with $(M-1)(L+1)$ zero elements	38
$\bar{\mathbf{s}}(k)$	$(k+1)M(L+1)$ row vector of transmitted signal at time k with $kM(L+1)$ zero elements	44
$S(k)$	$N \times MN(L+1)$ matrix of received signal at time k	38
T	symbol period	14
T_f	period of changing fading rate	71
$T_{h_i}(k)$	threshold for removing states at time k	83
$T_{h_{\max}}$	maximum threshold for removing states	87
T_m	multipath spread of channel	15
T_s	sample period	14
$T_B(s)$	transfer function of Bessel fading filter	24
$\mathcal{U}$	combination of discrete and continuous unknown parameters	58
$\mathcal{U}_k$	combination of discrete and continuous unknown parameters up to time k	62
$\tilde{\mathcal{U}}_{k\|K}$	estimation of $\mathcal{U}_k$ based on received signal up to time k	62
$v(t)$	output signal of channel (continuous)	14
$V(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(l)})$	conditional expectation of desired data	34
$w(k)$	zero mean white Gaussian random process with unit variance	44
$\boldsymbol{w}(k)$	$L+1$ zero mean white Gaussian random vector with identity autocorrelation matrix	44
$y(t)$	received signal at the output of receiver filter (continuous)	15
$y(k)$	received signal (discrete)	15
$\boldsymbol{y}_k^{\diamond}$	N vector of received signal at time k	38
$\mathbf{y}$	vector of received signal	17
$\mathbf{y}_k$	vector of received signal up to time k	38
$z(t)$	noise signal at the output of receiver filter (continuous)	15
$z(k)$	noise signal (discrete)	15
$\boldsymbol{z}_k$	N vector of additive Gaussian noise	38

$\acute{\boldsymbol{z}}_k$	$N-1$ vector of additive Gaussian noise	39
$\mathcal{Z}(t)$	additive Gaussian noise (continuous)	14
$(\Delta f)_c$	coherence bandwidth	15
$(\Delta t)_c$	coherence time	110
$\Delta \boldsymbol{s}_i(k)$	difference between $\boldsymbol{s}_i^c(k)$ and $\boldsymbol{s}_i^u(k)$	84
$\phi_c(\tau, \hat{\tau}; \Delta t)$	autocorrelation function of $c(\tau, t)$	16
$\boldsymbol{\gamma}$	first column of lower triangular matrix of Cholesky decomposition of Σ_z^{-1}	39
$\Gamma(\boldsymbol{a}_k)$	unit upper triangular matrix of Cholesky factorization of $R_{\boldsymbol{y}}^{-1}$	19
$\boldsymbol{\varphi}$	vector of time-invariant unknown parameters	47
$\boldsymbol{\varphi}_k$	vector of unknown parameters at time k	35
λ	forgetting factor	41
$\boldsymbol{\mu}_{\mid k}$	conditional mean of $\boldsymbol{h}$	42
$\boldsymbol{\mu}_z$	mean of $\boldsymbol{z}_k$	38
$\bar{\boldsymbol{\mu}}_{k\mid k}$	conditional mean of $\bar{\mathbf{h}}$	44
ν	rolloff factor	24
$\boldsymbol{\theta}$	vector of unknown parameters	33
$\boldsymbol{\theta}_k$	vector of unknown parameters up to time k	35
$\hat{\boldsymbol{\theta}}^{(l)}$	estimation of $\boldsymbol{\theta}$ in lth iteration	34
$\tilde{\boldsymbol{\theta}}_{k\mid l}$	estimation of $\boldsymbol{\theta}_k$ based on received signal up to time l	36
$\beta(.)$	branch metric	19
$\beta_{ij}(k)$	branch metric between ith and jth states at time k	83
$\beta_i^c(k)$	correct branch metric emanates from ith state	83
$\beta_i^u(k)$	minimum cost incorrect branch metric emanates from ith state	83
$\rho_{i,j}(\boldsymbol{a}_k)$	ijth element of the $\Gamma(\boldsymbol{a}_k)$ matrix	20
σ_c^2	variance of correct branch metric	90
σ_u^2	variance of incorrect branch metric	90
$\Sigma_{\mid l}$	conditional covariance matrix of $\boldsymbol{h}$	43
Σ_z	covariance matrix of $\boldsymbol{z}_k$	39
$\Sigma_{\acute{z}}$	covariance matrix of $\acute{\boldsymbol{z}}_k$	39

$\mathbf{\Sigma}_{k\|k}$	conditional covariance matrix of $\bar{\mathbf{h}}$	44
$\tilde{\Sigma}_{\|l}$	estimation of conditional covariance matrix of $\boldsymbol{h}$	43
τ_m	mean of multipath delay spread	109
τ_{rms}	standard deviation of multipath delay spread	109
ϑ	memory length of transmitter and channel	21

List of Acronyms

AR(N-1)	$N-1$th order autoregressive process	38
ARMA	autoregressive moving average	5
ASA	adaptive state allocation	7
ASA-AP	adaptive state partitioning method in ASA	95
ASA-AT	adaptive threshold method in ASA	95
BMC	branch metric calculator	20
CIR	channel impulse response	4
CPM	continuous phase modulation	4
CSP	channel statistical parameters	23
DDFSE	delayed decision-feedback sequence estimation	7
DFSE	decision-feedback sequence estimation	79
DFE	decision feedback equalizer	3
DPA	dynamic programming algorithm	20
DQPSK	differentially-encoded quadrature phase shift keying	8
EM	expectation and maximization	2
FIR	finite impulse response	13
FNF	frequency nonselective (flat) fading	15
FSF	frequency selective fading	15
GMLSDE	generalized maximum likelihood sequence detection and estimation	6
ISI	intersymbol interference	1
LMS	least mean square	5
MAP	maximum *a posteriori*	42

ML	maximum likelihood	2
MLSD	maximum likelihood sequence detection	1
MLSDE	maximum likelihood sequence detection and estimation	6
MMSE	minimum-mean square error	20
PBP	per-branch processing	61
PSP	per-survivor processing	5
RLS	recursive least squares	5
RSSE	reduced-state sequence estimation	7
SNR	signal-to-noise ratio	11
VA	Viterbi algorithm	12

Chapter 1

Introduction

A swift evolution in communication technology is under development to shift the information network from wire to wireless and from voice oriented systems to multi-media systems (voice, video and data). Mobility and portability provide more freedom for users and expand the communication network in the business, industrial and personal sectors. The wireless evolution, which will be preceded by a rapidly growing demand for a wide variety of new services in speech, video and especially data transfer, requires high bit rate data transmission with a reliable communications. These requirements along with the dynamic environment in mobile communications which disperses the transmitted signal in the time and frequency domains show the need of sophisticated signal processing algorithms.

Adaptive equalization is a popular approach to achieve good performance for narrowband modulated signal transmitted over a channel with intersymbol interference (ISI). However, the conventional equalization techniques such as linear and decision feedback equalizations sometimes do not achieve satisfactory performance in multipath fading channels [1]. Hence maximum likelihood sequence detection (MLSD) which minimizes sequence error rate has been considered for IS-54 [2, 3], IS-136 [4], GSM [5] and Japanese personal digital cellular [6] in order to achieve better performance in mobile communication systems.

MLSD based on the Viterbi algorithm needs channel parameters which are unknown in real situations and need to be estimated. Hence joint channel estimation and data detection based on sequence detection schemes were proposed in the literature [7–13]. In most of these studies the channel coefficients have been modeled as constant values [7–9]. The performance of the MLSD receiver is affected by the quality of the channel estimation especially in fast fading; thus it needs to model the channel as a time-variant system. However studies of the MLSD receiver using time-variant channel models assume knowledge about the channel which is not available [10–13].

Also, the MLSD receivers proposed in the literature and based on joint channel estimation and data detection use practical ways to estimate unknown parameters which are needed in the detection; but these methods are not based on maximum likelihood (ML) criterion. Sometimes in the multipath fading channels, the received signal does not provide complete data in order to achieve the ML criterion in channel estimation and data detection simultaneously. The expectation and maximization (EM) algorithm [14] which achieves ML criterion iteratively is a good candidate method to deal with this obstacle.

Computational complexity of the MLSD receiver increases exponentially when channel memory is increased. Although reduction of computational complexity of the Viterbi-based receivers and decoders have been considered for ISI channels in the literature [15–18], this challenging issue has rarely been considered for the multipath fading channels [19].

In this thesis we concentrate on receiver design based on ML criterion for joint channel estimation and sequence data detection using a bandlimited signaling scheme when the channel parameters are unknown. Statistical detection and estimation theory is used to develop optimal (or near optimal) and adaptive receiver structures based on the EM algorithm for Rayleigh multipath fading channels. Also, we focus on reduction of computational complexity of the MLSD receiver in which signal-to-noise ratio is time-variant in fading channels. The use of the EM algorithm to unify the ML estimation and detection problem and the study of complexity reduction schemes in fading channels are the main novel goals of this thesis.

We summarize major detection methods for a signal transmitted over multipath fading channels in the following Section. A brief survey of the maximum likelihood sequence detection receiver along with thesis motivations will be presented in Section 1.2. After a thesis overview, the chapter concludes with a summary of the major contributions of the thesis.

1.1 Detection Methods to Mitigate Channel Distortion

Detection methods, which combat the distortion introduced by multipath fading channels in frequency and time domains, can be categorized into two basic classes corresponding to wideband and narrowband modulation schemes.

Wideband modulation schemes such as commutation signaling [20] and spread spectrum techniques [21] employ a transmission bandwidth that is several orders of magnitude greater than the baseband signal bandwidth. Although these modulation schemes are bandwidth inefficient for a single user, many users can simultaneously share the same bandwidth without significantly interfering with each other. Spread spectrum modulation has an inherent capability to mitigate interference. Also this method resists distortions introduced by multipath fading due to the uniformly-spread energy of the signal over a very large bandwidth. Wideband modulation schemes often use a RAKE receiver to detect the transmitted signal [22].

Narrowband or bandlimited modulation schemes employ a transmission bandwidth of the same order as the bandwidth of the baseband signal. In these modulation schemes, equalization methods are used to mitigate the distortion introduced by the channel. Generally there are two basic equalization classes; linear and nonlinear. In linear equalization using structures such as transversal and lattice equalizers the received signal is delayed and weighted by equalizer coefficients and summed to produce the output. Nonlinear equalizers are capable of combating severe channel distortion which cannot be handled by linear equalizers. Decision feedback equalizer (DFE) and MLSD are examples of nonlinear equalizers. The DFE consists of a feedforward and a feedback filter and has the capability to remove the ISI that results from previously detected symbols through the feedback section [23, 24].

Determination of an optimum receiver to detect a bandlimited signal transmitted through a channel (deterministic or random) and corrupted by an additive Gaussian noise was studied by Kailath [25]. The structure of the optimal receiver can be interpreted as a bank of estimator-correlator combinations. However the structure of the receiver, which was derived based on manipulating the set of *a posteriori* probabilities, is very complex and impractical. MLSD as an optimal receiver which was first proposed by Forney [26] for ISI channels can be practically implemented by a Viterbi algorithm based on a known channel impulse response (CIR) [27]. The Viterbi algorithm originally proposed by Viterbi [28] for maximum likelihood decoding of convolutional codes is a special case of forward dynamic programming [29, 30]. The Viterbi algorithm which finds optimal trajectories (surviver paths) at each stage for all states in the trellis diagram was used by Forney in the MLSD receiver to detect a digital signal transmitted through the ISI channel corrupted by an additive Gaussian noise [26].The MLSD algorithm computes the cost (probability of error) through all trajectories from each stage to the next stage for all possible states (computing the branch metrics) and then finds the minimum cost trajectory (survivor path) for each state. Therefore at the final stage the best survivor path corresponds to the data sequence with minimum probability of sequence error.

1.2 A Brief Survey and Motivations

After Forney's work [27], Ungerboeck [31] modified the MLSD receiver such that the computation of the metrics in the Viterbi algorithm is straightforward. Ungerboeck's structure does not need a whitening filter and can be implemented in an adaptive manner. The MLSD receiver for fading channels whose statistical parameters are known was considered in [10–12, 32]. The MLSD receiver proposed in [32] contains a bank of time-variant linear filters, which is difficult to realize, to estimate subsequence waveforms along with the Viterbi algorithm. Lodge and Moher [10] proposed a MLSD receiver for continuous phase modulation (CPM) signal transmitted over Rayleigh flat fading channels. The receiver structure is a combination of a bank of linear prediction filters and the Viterbi algorithm. Modeling the fading process by an autoregressive moving average (ARMA) system , Dai and Shwedyk [11] have derived a MLSD receiver for frequency selective Rayleigh fading channels. Assuming known statistical parameters of the channel, the receiver used Kalman filtering in order to estimate the CIR. Yu and Pasupathy [12] have developed a MLSD receiver using an innovation-based approach. Under assumption of the finite channel memory with known statistical parameters, the receiver can be implemented by a bank of prediction filters and the Viterbi algorithm.

Implementation of MLSD using the Viterbi algorithm requires knowledge of the channel (deterministic or stochastic parameters) which is generally unknown at the receiver and should be estimated. To solve this problem, joint channel estimation and data detection methods were proposed in [7, 11, 33–35] where some estimation algorithms such as least mean square (LMS), recursive least squares (RLS) and Kalman filtering were used in channel estimation. However, the inherent decision delay in such procedures, when the Viterbi algorithm is used causes poor channel tracking in a time-variant environment. The idea of per-survivor processing (PSP) was proposed to combat the decision delay problem [8], where each survivor path in the trellis diagram has its own estimator. The PSP approach has shown good tracking performance in comparison with using a single channel estimator [9, 36].

In the MLSD derivation for a fading channel modeled by a random process, it is assumed that perfect knowledge of channel statistical parameters (e.g. mean and auto-covariance for Gaussian random process) is available at the receiver (e.g. [10–12]); however, these must be estimated from the received signal in practice. Recently Hart [37] proposed a minimization algorithm to search for the mean and autocovariance of the channel. The scheme, however, is very complicated and nonrecursive.

The first major motivation for this thesis comes from *joint channel estimation and data detection*. The joint estimation and detection methods proposed in the literature [7, 8, 11, 33–35] are practical ways to estimate the unknown parameters needed in the detection; however, the nature and degree of optimality of the estimation procedure, the influence of such an estimate on the optimality of the MLSD receiver and couplings among estimation, detection and channel model are not clear. Also, issues such as a unified view of many MLSD receivers developed in the literature, and the relation between the channel model and structure of the receiver need more investigation.

Sometimes, due to unknown channel parameters, maximizing the likelihood function directly to achieve the ML criterion is infeasible. In this situation, which is common in multipath fading channels, the received signal does not provide the complete data necessary to estimate the unknown channel parameters and direct access to the necessary data is impossible or missing. *The EM algorithm which achieves the ML criterion in an iterative manner is ideally suited to this problem.* The EM algorithm has two steps, expectation and maximization [38]. The first step takes the expectation of the log-likelihood function of the complete data given the current estimated parameters and the incomplete data (e.g. received signal). The second step provides a new estimate of the unknown parameters by maximizing the expectation of the log-likelihood function over the unknown parameters. These steps are repeated iteratively to increase the likelihood till the new estimated parameters become (or get arbitrarily close to) the same parameter values estimated at the previous iteration [14].

The EM algorithm, which satisfies the ML criterion when complete data is not available, uses decision feedback inherently and also couples estimation and detection; this provides another strong motivation to use EM in the joint channel estimation and data detection problem. Throughout the thesis, the generalized maximum likelihood detection and estimation (GMLSDE) receiver structure will be considered based on maximum likelihood detection/estimation theory and using the EM algorithm. Some adaptive MLSDE are derived based on the GMLSDE for different channel models.

The other major motivation for the thesis arises from the complexity of the MLSD receivers, which increases exponentially with increasing channel memory length such that, sometimes, the receiver implementation seems impractical. Some efforts have been done to reduce the complexity of MLSD receivers. One group reduces the time-duration of the channel impulse response by using an equalizer as a pre-filter [15, 34, 39–41]. Another group decreases the number of states by using Ungerboeck set partitioning to fuse some states permanently. Important methods from this latter group are reduced-state sequence estimation (RSSE) [16] and delayed decision-feedback sequence estimation (DDFSE) [17]. The other group selects some more likely states such as M algorithm [42] which chooses M states of the trellis diagram at each time and T algorithm [18] which chooses the states whose costs are less than a threshold. Fundamentally, the above efforts to reduce the complexity have been done for MLSD receivers designed for ISI channels and not for fading channels and also most of them have not been completely formulated. Using the nature of the time-variant signal-to-noise ratio in fading channels, an adaptive state allocation (ASA) algorithm will be developed in this thesis so as to drastically reduce the complexity of the MLSD/MLSDE receivers with negligible loss in the bit error rate performance.

1.3 Thesis Overview

The MLSD receiver for multipath fading channels will be studied in chapter 2 through a review of related literature. Based on the the sampling theorem for linear time-variant systems, a discrete model for Rayleigh multipath fading channel will be developed for the bandlimited signaling scheme. It also describes the relation between the statistical parameters of discrete and analog channel models. The chapter will derive the structure of the MLSD receiver based on known channel autocorrelation function and known CIR along with diversity reception. For known channel autocorrelation function, the MLSD structure is implemented by a bank of predictor filters with constant coefficients where the biased power of prediction errors are used as branch metrics in the trellis diagram. Due to the potentially better performance we get when CIR is estimated directly from the received signal in comparison with estimating the channel autocorrelation function from the autocorrelation function of the received signal and also considering the complexity issue, the chapter recommends the CIR estimation method which is the motivation for Chapter 3.

Estimation of channel parameters embedded in white and colored Gaussian noise will be investigated in Chapter 3. The estimation procedure is developed based on ML criterion using the EM algorithm. In general, the EM algorithm which is an iterative procedure provides ML estimation of channel parameters when maximization of the likelihood function may not be feasible directly. In this Chapter, we will develop a novel RLS/Kalman-type algorithm which estimates the channel parameters, which are corrupted with additive colored noise modeled by an autoregressive process, sequentially in-time. The estimating algorithm is derived by following Titterington's approach of stochastic approximation [43]. We will show how the RLS/Kalman-type algorithm becomes the regular RLS algorithm by relaxing the colored noise assumption. For a CIR modeled by a Gaussian vector/process, combined adaptive RLS and Kalman algorithms are derived for estimating the CIR along with the deterministic parameters of the dynamic evolution of CIR through a unified recipe based on ML criterion using the on-line EM algorithm. The results of this Chapter provides a foundation for joint estimation and detection algorithm which will be considered in Chapter 4.

The basic effort in Chapter 4 is to develop a solid theoretical base for adaptive joint channel estimation and data detection. The maximum likelihood detection and estimation (MLSDE) will be considered in a general framework based on ML detection/estimation theory using the EM algorithm. Generalized MLSDE (GMLSDE) is presented based on the on-line EM algorithm which alternates between detection and estimation and still satisfies the ML criterion. It will be shown that in the GMLSDE receiver the concept of per-survivor processing (PSP), which was proposed to combat the decision delay in the joint detection/estimation problem and is also a practical way to achieve better performance, appears naturally as an integral part of a likelihood-increasing procedure.

In Chapter 4, based on different levels of channel knowledge available at the receiver, some adaptive receivers will be derived by using the GMLSDE in a unified way. Although all derived adaptive MLSDE schemes are not new , they show, as examples, the power of the GMLSDE algorithm to develop new MLSDE receivers based on channel models and available channel knowledge in a unified framework. Chapter 4 will contain computer simulations evaluating and comparing the performance of different derived adaptive MLSDE when differentially-encoded quadrature phase shift keying (DQPSK) modulation scheme is used and the transmitted signal pass through frequency flat and selective channels with $f_dT = 0.1, 0.01$ and 0.001.

The computational complexity of the MLSD/MLSDE receiver grows exponentially with the channel memory length. Thus, though MLSD/MLSDE receiver is an optimal scheme, unfortunately its computational complexity limits many applications and also the implementation of MLSD/MLSDE receiver for large channel memory seems impractical. In Chapter 5, we will address this point of weakness of the MLSD/MLSDE receiver and will propose an algorithm called adaptive state allocation (ASA) which greatly reduces the computational complexity of the MLSD/MLSDE receiver with negligible degradation in the error performance. The ASA algorithm selects a few states of the trellis diagram with an adaptive threshold value computed on the basis of the short-term power of CIR. Also, an adaptive partitioning method which fuses the branch metrics is employed in the ASA. The adaptive threshold method selects the states whose costs are less than the threshold plus the minimum cost. The adaptive partitioning method fuses/diffuses branch metrics based on the Kullback-Leibler distance between the probability density functions of the correct and incorrect branch metrics in the trellis diagram. In this Chapter, some issues about the implementation of the ASA will be considered and also a comparison between the performance and computational complexity of the ASA and the regular MLSDE receiver (without the computational complexity reduction) for frequency flat and selective fading channels will be presented through computer simulations.

Main results and conclusions of this thesis are summarized in Chapter 6 along with some suggestions for future research.

1.4 Summary of Contributions

We now present a summary of the major contributions achieved in this thesis.

Estimation:

- A novel RLS/Kalman-type algorithm, which estimates channel parameters whose output is corrupted by an additive colored Gaussian noise modeled by an autoregressive process, is developed [44, 45].

- An iterative algorithm which estimates unknown parameter sets separately and still increases the likelihood function is developed [45].

- A recursive algorithm is derived for estimating the statistical parameters of a Gaussian random process CIR based on the principle of increasing the likelihood [44, 46].

Detection:

- Generalized MLSDE is proposed for real time joint channel estimation and data detection, which alternates between estimation and detection procedures and still satisfies the ML criterion [47].

- Different adaptive MLSDE algorithms are formulated in a unified manner for different channel models and different amounts of channel knowledge available at the receiver [47].

- A solid theoretical foundation is presented for per-survivor processing (PSP) as an integral part of an EM-based ML detection and estimation procedure [47].

- The structure of the MLSD receiver with known channel statistical parameters is proposed for multipath Rayleigh fading channels. The receiver structure is implemented by a bank of fixed coefficients prediction filters processing the received signal samples which are statistically sufficient [48, 49].

Complexity Reduction:

- An adaptive threshold method is formulated such that the correct state is chosen with high probability among a few selected states of the trellis diagram in order to reduce the computational complexity of MLSD or MLSDE receivers at high SNR without a significant loss in performance [50].

- An adaptive partitioning scheme which fuses/diffuses the branch metrics is proposed based on Kullback-Leibler distance to reduce the complexity of the MLSD/

MLSDE receiver at low SNR for a time-variant CIR without sacrificing the performance significantly [50].

- Adaptive state allocation (ASA) algorithm combining adaptive threshold and partitioning methods is developed. This algorithm greatly reduces the computational complexity of regular MLSD/MLSDE with negligible degradation in performance in a multipath fading environment [50].

Chapter 2

MLSD Receiver for Multipath Fading Channels

Communication over multipath fading channels such as land-mobile radio and indoor wireless communication suffers from intersymbol interference (ISI) and fading phenomena. Both ISI and fading imply correlation among a sequence of symbols which are received by the receiver. It is well known that maximum likelihood sequence detection (MLSD) exploits these correlations and is the optimal receiver for these channels in the sense of minimizing the sequence error rate for equiprobable sequences [10–12, 48].

Forney [26] proposed the MLSD structure based on the Viterbi algorithm (VA) as the optimal receiver for channels with ISI and additive white Gaussian noise. The MLSD receiver for Rayleigh flat fading channels and continuous phase modulation was presented by Lodge and Moher [10]. Dai and Shwedyk [11] developed sequential sequence estimation for selective Rayleigh fading channels. Using the innovations approach, a MLSD algorithm which demodulates the received signal recursively was proposed by Yu and Pasupathy [12].

In this chapter, we derive the MLSD structure based on dynamic programing algorithm (Viterbi algorithm) for Rayleigh fading channels, both frequency nonselective and selective, without restriction on the fading rate and signaling scheme. The MLSD is developed by applying the maximum likelihood (ML) criterion to samples of the received signal which provide sufficient statistics. For known autocorrelation function of the channel impulse response (CIR), the structure of the MLSD is realized using a bank of predictor filters with constant coefficients for computing the branch metrics which are used in the dynamic programing. The coefficients of the predictor filters are obtained based on the transmitted signal shape and the autocorrelation function of the CIR. This structure is similar to the structure proposed in [12] where an innovation approach was used to derive the receiver structure. For known CIR, the structure of the MLSD is implemented by a time-variant FIR filter.

The motivations behind this chapter are to develop the structures of the MLSD receivers based on known channel parameters (stochastic or deterministic) which are really unknown; then by evaluating the performance of the receivers we will establish a strategy for estimating the channel parameters.

The models of Rayleigh fading channels are described in Section 2.1 and the MLSD algorithm is derived in Section 2.2 for known statistical parameters and known impulse response of the channel separately. Diversity reception is applied to the MLSD in Section 2.3. Knowing the channel parameters plays an important role in the MLSD algorithm derived in this chapter. We discuss the estimation of channel parameters and the complexity of the estimation procedure in Section 2.5 based on the performance of the receiver evaluated in Section 2.4.

2.1 Channel Models for Bandlimited Signals

It is well-known that fading multipath channels can be modeled as time-variant systems. The equivalent low-pass impulse response $c(\tau, t)$ represents the output of the fading multipath channel at time t due to an input impulse applied at time $t - \tau$. Based on the central limit theorem $c(\tau, t)$ is a complex-valued Gaussian random process. When there is no fixed path between the transmitter and the receiver the mean value of $c(\tau, t)$ is zero and $|c(\tau, t)|$ at any instant t is Rayleigh distributed and the channel is referred to as a Rayleigh fading multipath channel [22] [51] (see Appendix A for a brief description of the multipath fading channels).

A communication system over the multipath fading channel is shown in Fig. 2.1.

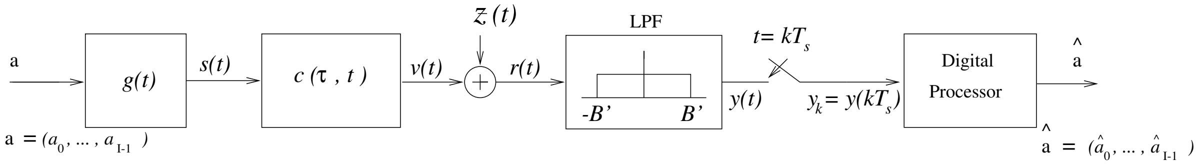

Figure 2.1: The communication system over fading multipath channels.

The equivalent low-pass transmitted signal is

$$s(t) = \sum_{i=0}^{\lfloor t/T \rfloor} a_i g(t - iT) \tag{2.1}$$

where $\{a_i\}_{i=0}^{I}$ are independent, identically distributed equiprobable symbols taken from a q-ary constellation, T is the symbol period, $g(t)$ is the impulse response of the transmitter filter, and $\lfloor . \rfloor$ is the floor function (the largest integer not exceeding the argument). The received waveform is

$$r(t) = v(t) + \mathcal{Z}(t) = \int_{-\infty}^{+\infty} c(\tau, t) s(t - \tau) d\tau + \mathcal{Z}(t) \tag{2.2}$$

where $c(\tau, t)$ is the equivalent low-pass impulse response of the fading multipath channel and $\mathcal{Z}(t)$ is a circularly symmetric zero-mean white complex Gaussian random process with an autocorrelation function $R_{\mathcal{Z}}(\tau) = \hat{N}_0 \delta(\tau)$ [52]. By using the sampling theorem when the bandwidth of $s(t)$ is B, $v(t)$, the signal component in $r(t)$, becomes

$$v(t) = \int_{-\infty}^{+\infty} c(\tau, t) \sum_l s(t - lT_s) \mathrm{sinc}(\frac{\tau - lT_s}{T_s}) d\tau = \sum_l s(t - lT_s) \check{h}_l(t) \tag{2.3}$$

where $\mathrm{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$, $T_s \leq \frac{1}{2B}$ is the sampling period and

$$\check{h}_l(t) \doteq \int_{-\infty}^{+\infty} c(\tau, t) \mathrm{sinc}(\frac{\tau - lT_s}{T_s}) d\tau \tag{2.4}$$

If f_d is defined as the maximum Doppler frequency of the multipath fading channel, the power spectral bandwidth of $\check{h}_l(t)$ is limited to f_d. Thus, from the statistical viewpoint, $h_l(t)$, the output of $\check{h}_l(t)$ filtered by an ideal low-pass filter of bandwidth f_d, is equal to $\check{h}_l(t)$.

$$h_l(t) \doteq \int_{-\infty}^{+\infty} 2f_d \check{h}_l(\lambda) \mathrm{sinc}(2f_d(t-\lambda))d\lambda = \check{h}_l(t) \quad \text{for all } l \tag{2.5}$$

When the coherence bandwidth, $(\Delta f)_c$ (see Appendix A for definition of the coherence bandwidth), of the fading multipath channel is $(\Delta f)_c > 2B$,the channel is called frequency nonselective (or flat) fading (FNF) and when $(\Delta f)_c < 2B$ the channel is called frequency selective fading (FSF). Since the essentially nonzero duration of the autocorrelation function of $c(\tau, t)$ with respect to τ is T_m, the multipath spread of the channel, a practical maximum of l is $L = \lceil \frac{T_m}{T_s} \rceil$, where $\lceil . \rceil$ is the ceiling function (the smallest integer exceeding the argument). For a frequency nonselective fading channel $L = 0$ and $h_0(t) = \int c(\tau, t)d\tau$. Therefore from (2.2)-(2.5) the received signal is

$$r(t) = \sum_{l=0}^{L} h_l(t)s(t - lT_s) + \mathcal{Z}(t) \quad \begin{array}{ll} L = 0 & \text{FNF} \\ L = \lceil \frac{T_m}{T_s} \rceil & \text{FSF} \end{array} \tag{2.6}$$

The power spectral bandwidth of $v(t)$ (2.3) is $B' = B + f_d$. Thus, $y(t)$, the output of an ideal low-pass filter with bandwidth B', is statistically sufficient to detect the data sequence (Fig. 2.1). Due to the flatness of the ideal low-pass filter and its bandwidth, the samples of $y(t)$, with sampling period $T_s = \frac{1}{2B'} \leq \frac{1}{2B}$ such that $J = T/T_s$ becomes the minimum integer number,are given by

$$\begin{aligned} y(k) = y(t)|_{t=kT_s} &= \sum_{l=0}^{L} h_l(t)s(t - lT_s) + z(t)|_{t=kT_s} \\ &= \sum_{l=0}^{L} h_l(kT_s)s(kT_s - lT_s) + z(kT_s) \\ &= \sum_{l=0}^{L} h(l,k)s(k-l) + z(k) \end{aligned} \tag{2.7}$$

where $z(t)$, the noise component at the output of the low-pass filter, is a bandlimited circularly symmetric complex Gaussian noise and $z(k) = z(t)|_{t=kT_s}$ is a circularly symmetric discrete complex Gaussian random process with variance $N_0 = E[|z(k)|^2] = 2B'\hat{N}_0$.

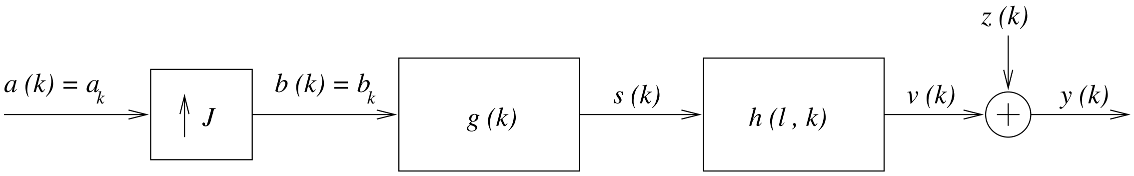

Figure 2.2: The discrete model of the transmitter and the fading channel, where $\uparrow J$ represents up-sampling by a factor J.

Also, $h(l,k) = h_l(t)|_{t=kT_s}$, and $s(k) = s(t)|_{t=kT_s}$ from (2.1) will be

$$\begin{aligned} s(k) &= \sum_{i=0}^{\lfloor \frac{k}{J} \rfloor} a_i g(kT_s - iJT_s) \\ &= \sum_{i=0}^{k} b_i g((k-i)T_s) = b(k) \star g(k) \end{aligned} \tag{2.8}$$

where $g(k) = g(t)|_{t=kT_s}$ and

$$b_k = b(k) = \begin{cases} a_{\frac{k}{J}} = a(\frac{k}{J}) & \text{when } \frac{k}{J} \text{ is an integer} \\ 0 & \text{otherwise} \end{cases} \tag{2.9}$$

Therefore, based on (2.7) the channel can be modeled as an L-order time-variant finite impulse response (FIR) filter. Fig. 2.2 shows the discrete model of the transmitter and the multipath fading channel. By uncorrelated scattering assumption for the channel with the impulse response $c(\tau, t)$, the autocorrelation function of $h(l,k)$ becomes

$$\begin{aligned} R_h(l_1, l_2; j) &= E[h(l_1, k+j)h^*(l_2, k)] \\ &= \int_{-\infty}^{+\infty} sinc(\frac{\tau - l_1 T_s}{T_s}) sinc(\frac{\tau - l_2 T_s}{T_s}) R_c(\tau, jT_s) d\tau \\ &\qquad -\infty \le j \le \infty \quad 0 \le l_1, l_2 \le L \end{aligned} \tag{2.10}$$

where the autocorrelation function of $c(\tau, t)$ is

$$\phi_c(\tau, \hat{\tau}; \Delta t) \;=\; E[c(\tau, t+\Delta t)c^*(\hat{\tau}, t)]$$

$$= \quad R_c(\tau, \Delta t)\delta(\tau - \hat{\tau}) \qquad -\infty \leq \Delta t \leq \infty \quad 0 \leq \tau, \hat{\tau} \leq \infty \quad (2.11)$$

As seen from (2.10), $h(l, k)$ for all l are wide-sense stationary random processes.

2.2 Maximum Likelihood Sequence Detection

The received signal $r(t)$ is filtered by an ideal low-pass filter with bandwidth $B' = B + f_d$ and then sampled at the Nyquist sampling rate $1/T_s$ (Fig. 2.1). The sampler outputs $y(kT_s)$ are sufficient statistics for the detection of $\mathbf{a} = (a_0, a_1, .., a_i, .., a_{I-1})$. By the assumption of equal probability for all possible input $\mathbf{a}$, the MLSD is the optimal decision rule for minimizing the sequence error rate [26] and the MLSD criterion is

$$C_{MLSD} \doteq p(\mathbf{y}|\hat{\mathbf{a}}) = \max_{\mathbf{a}\in\mathcal{A}} p(\mathbf{y}|\mathbf{a}) \tag{2.12}$$

where $\hat{\mathbf{a}} = (\hat{a}_0, \hat{a}_1, .., \hat{a}_i, .., \hat{a}_{I-1})$ is the detected symbol sequence, $\mathcal{A}$ is the set of all possible transmitted symbol sequences and $\mathbf{y} = (y_{K-1}, y_{K-2}, .., y_k, .., y_0)$ is the sequence of the output samples, where $K = JI$[1]. If the transmitter filter, whose impulse response duration is limited to L_sT_s ($L_s = \alpha J$, α is an integer), is a causal system, the output samples based on (2.7) and (2.8) are

$$y_k = y(t)|_{t=kT_s} = \sum_{l=0}^{L} h(l, k) \sum_{n=k-l-L_s+1}^{k-l} b_n g(k - l - n) + z(k) \tag{2.13}$$

From (2.12) the MLSD criterion becomes

$$\begin{aligned} C_{MLSD} &= \max_{\mathbf{a}\in\mathcal{A}} p(y_{K-1}, y_{K-2}, \ldots, y_k, \ldots, y_0|\mathbf{a}\} \\ &= \max_{\mathbf{a}\in\mathcal{A}} \{p(y_{K-1}, y_{K-2}, \ldots, y_{K-J}|y_{K-J-1}, \ldots, y_0, \mathbf{a}) \\ &\quad p(y_{K-J-1}, y_{K-J-2}, \ldots, y_{K-2J-1}|y_{K-2J-2}, \ldots, y_0, \mathbf{a}) \ldots\} \end{aligned}$$

[1]Due to the memory of the channel and the transmitter filter, the number of output samples is more than JI. However in practice, the transmitter sends some symbols, whose length is related to the whole system memory, known by the receiver. Hence the receiver processes only JI samples of the received signal.

$$= \max_{\mathbf{a}\in\mathcal{A}}\{\prod_{i=0}^{I-1}\prod_{j=0}^{J-1} p(y_{iJ+j}|y_{iJ+j-1}, y_{iJ+j-2}, ..., \mathbf{a})\} \tag{2.14}$$

Let us assume that $h(l,k)$ has a finite memory such that [12]

$$E[y(k)|y(k-1),...,y(0),\mathbf{a}] = E[y(k)|y(k-1),...,y(k-M+1),\mathbf{a}] \tag{2.15}$$

Since $y(k)$ is a Gaussian random process it can be shown that (2.15) leads to the relation given below[2].

$$p(y_k|y_{k-1},...,y_0,\mathbf{a}) = p(y_k|y_{k-1},...,y_{k-M+1},\boldsymbol{a}_k) \tag{2.16}$$

where $\boldsymbol{a}_k = [a_{\frac{k-L-L_s-L_R}{J}},...,a_{\frac{k}{J}}]$, $L_R = M-1$ and based on (2.9), only those elements of $\boldsymbol{a}_k$ whose subscripts are integers are included. Thus, the MLSD criterion (2.14) becomes

$$\begin{aligned} C_{MLSD} &= \max_{\mathbf{a}\in\mathcal{A}}\{\prod_{i=0}^{I-1}\prod_{j=0}^{J-1} p(y_{iJ+j}|y_{iJ+j-1},...,y_{iJ+j-M+1},\boldsymbol{a}_{iJ+j})\} \\ &= \max_{\mathbf{a}\in\mathcal{A}}\{\prod_{i=0}^{I-1}\prod_{j=0}^{J-1} \frac{p(y_{iJ+j},...,y_{iJ+j-M+1}|\boldsymbol{a}_{iJ+j})}{p(y_{iJ+j-1},...,y_{iJ+j-M+1}|\boldsymbol{a}_{iJ+j})}\} \\ &= \max_{\mathbf{a}\in\mathcal{A}}\{\prod_{i=0}^{I-1}\prod_{j=0}^{J-1} \frac{(det(R_{\boldsymbol{y}^{M-1}}))^{-1}}{\pi^{-1}(\det(R_{\boldsymbol{y}}))^{-1}} \frac{exp(-\boldsymbol{y}_{iJ+j-1}^{M-1^{\mathcal{H}}} R_{\boldsymbol{y}^{M-1}}^{-1} \boldsymbol{y}_{iJ+j-1}^{M-1})}{exp(-\boldsymbol{y}_{iJ+j}^{\mathcal{H}} R_{\boldsymbol{y}}^{-1} \boldsymbol{y}_{iJ+j})}\} \end{aligned} \tag{2.17}$$

After using logarithms in (2.17), we get[3]

$$\begin{aligned} C_{MLSD} &\equiv \min_{\mathbf{a}\in\mathcal{A}}\{\sum_{i=0}^{I-1}\sum_{j=0}^{J-1}[\log(\det(R_{\boldsymbol{y}})) - \log(\det(R_{\boldsymbol{y}^{M-1}})) + \boldsymbol{y}_{iJ+j}^{\mathcal{H}}\Omega(\boldsymbol{a}_{iJ+j})\boldsymbol{y}_{iJ+j}]\} \\ &\equiv \min_{\mathbf{a}\in\mathcal{A}}\{\sum_{i=0}^{I-1}\sum_{j=0}^{J-1}\beta(\boldsymbol{a}_{iJ+j})\} \end{aligned} \tag{2.18}$$

[2]Conditions in (2.15) and (2.16) can be interpreted as wide-sense and strict-sense Markov properties. It can be shown that if a Gaussian sequence is wide-sense Markov (2.15), it is also strict-sense Markov (2.16) [53].

[3]All the "log" in this thesis are to the base "e", unless we specify otherwise.

where $\beta(.)$ is the branch metric, $X^{\mathcal{H}}$ denotes the conjugate transpose of X and by defining $k = iJ + j$

$$
\begin{aligned}
\boldsymbol{y}_k &= [y_k, y_{k-1}, \ldots, y_{k-M+1}]^{\mathcal{T}} \\
\boldsymbol{y}_{k-1}^{M-1} &= [y_{k-1}, y_{k-2}, \ldots, y_{k-M+1}]^{\mathcal{T}} \\
R_{\boldsymbol{y}} &= E[\boldsymbol{y}_k \boldsymbol{y}_k^{\mathcal{H}} | \boldsymbol{a}_k] \\
R_{\boldsymbol{y}^{M-1}} &= E[\boldsymbol{y}_{k-1}^{M-1} {\boldsymbol{y}_{k-1}^{M-1}}^{\mathcal{H}} | \boldsymbol{a}_k] \\
\Omega(\boldsymbol{a}_k) &= [R_{\boldsymbol{y}}^{-1}] - \begin{bmatrix} 0 & \vdots 0_{1\times(M-1)} \\ \ldots\ldots & \ldots\ldots \\ 0_{(M-1)\times 1} \vdots & R_{\boldsymbol{y}^{M-1}}^{-1} \end{bmatrix}
\end{aligned}
$$

where $X^{\mathcal{T}}$ denotes the transpose of X and $0_{i\times j}$ is a $i \times j$ zero matrix. Using (2.13) the ijth element of $R_{\boldsymbol{y}}$ is

$$
\begin{aligned}
R_{\boldsymbol{y}}(i,j) &= E[y_{k-i} y_{k-j}^* | \boldsymbol{a}_k] = E[(\sum_{l_1=0}^{L} h(l_1, k-i) s(k-i-l_1) + z(k-i)) \\
&\quad (\sum_{l_2=0}^{L} h^*(l_2, k-j) s^*(k-j-l_2) + z^*(k-j))] \\
&= \sum_{l_1=0}^{L} \sum_{l_2=0}^{L} R_h(l_1, l_2; j-i) s(k-i-l_1) s^*(k-j-l_2) + N_0 \delta(j-i) \quad (2.19)
\end{aligned}
$$

From (2.10), it follows that $R_h(l_1, l_2; j) = R_h(l_2, l_1; j)$, $R_h(l_1, l_2; j) = R_h^*(l_1, l_2; -j)$ and from (2.19) $R_{\boldsymbol{y}}(i,j) = R_{\boldsymbol{y}}^*(j,i)$. Therefore $R_{\boldsymbol{y}}$ is a symmetric positive definite matrix and it can be factorized as $R_{\boldsymbol{y}} = L(\boldsymbol{a}_k) D(\boldsymbol{a}_k) L^{\mathcal{H}}(\boldsymbol{a}_k)$ [54], where $L(\boldsymbol{a}_k)$ is a unit upper triangular matrix and $D(\boldsymbol{a}_k)$ is a diagonal matrix, $D(\boldsymbol{a}_k) = \mathrm{diag}(d_0(\boldsymbol{a}_k), d_1(\boldsymbol{a}_k), \ldots, d_{M-1}(\boldsymbol{a}_k))$. Then the inverse of $R_{\boldsymbol{y}}$ is given as

$$
R_{\boldsymbol{y}}^{-1} = \Gamma^{\mathcal{H}}(\boldsymbol{a}_k) D^{-1}(\boldsymbol{a}_k) \Gamma(\boldsymbol{a}_k) \quad (2.20)
$$

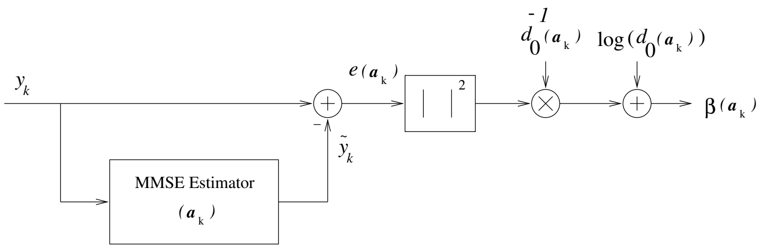

Figure 2.3: The block diagram of branch metric evaluation.

where $\Gamma(\boldsymbol{a}_k) = L^{-1}(\boldsymbol{a}_k)$. Similarly the inverse of $R_{\boldsymbol{y}^{M-1}}$ can be calculated. After some manipulations (see Appendix B), the branch metric $\beta(\boldsymbol{a}_k)$ in (2.18) becomes

$$\beta(\boldsymbol{a}_k) = \log(d_0(\boldsymbol{a}_k)) + \frac{|\sum_{m=0}^{M-1} y_{k-m}\rho_{m,0}(\boldsymbol{a}_k)|^2}{d_0(\boldsymbol{a}_k)} \tag{2.21}$$

where $\rho_{i,j}(\boldsymbol{a}_k)$ is the ijth element of the $\Gamma(\boldsymbol{a}_k)$ matrix. It can be shown that the factorization of $R_{\boldsymbol{y}}^{-1}$ based on (2.20) is equivalent to a Gram-Schmidt orthogonalization of $\boldsymbol{y}_k$ [55]. $L(\boldsymbol{a}_k)$ and $\Gamma(\boldsymbol{a}_k)$ are called innovations and whitening filters respectively [54, p. 207] such that by feeding the colored random process y_{k-m} to the FIR filter with coefficients $\rho_{m,0}(\boldsymbol{a}_k)$ for $m = 0, ..., M-1$, the output becomes a whitened random process. The branch metrics in (2.21) can also be written as

$$\beta(\boldsymbol{a}_k) = \log(d_0(\boldsymbol{a}_k)) + \frac{|e(\boldsymbol{a}_k)|^2}{d_0(\boldsymbol{a}_k)} \tag{2.22}$$

where, as shown in Fig.2.3, $e(\boldsymbol{a}_k) = y_k - \tilde{y}_k$ is the prediction error and $\tilde{y}_k = -\sum_{m=1}^{M-1} y_{k-j}\rho_{m,0}(\boldsymbol{a}_k)$ is the causal minimum-mean-square error (MMSE) estimate of y_k based on the assumption of sending $\boldsymbol{a}_k$ and given the previous samples $y_{k-1}, ..., y_{k-M+1}$ [54]. $d_0(\boldsymbol{a}_k)$ is the variance of the error $e(\boldsymbol{a}_k)$ and thus the branch metric can be interpreted as the normalized prediction error with a bias equal to $\log(d_0(\boldsymbol{a}_k))$.

Using the dynamic programming algorithm (DPA) based on (2.18), the receiver detects the information sequence $\mathbf{a}$. The MLSD structure receiver for known statistical channel parameters is shown in Fig.2.4 where each branch metric calculator (BMC) contains the predictor filter whose block diagram has been shown in Fig.2.3. When using

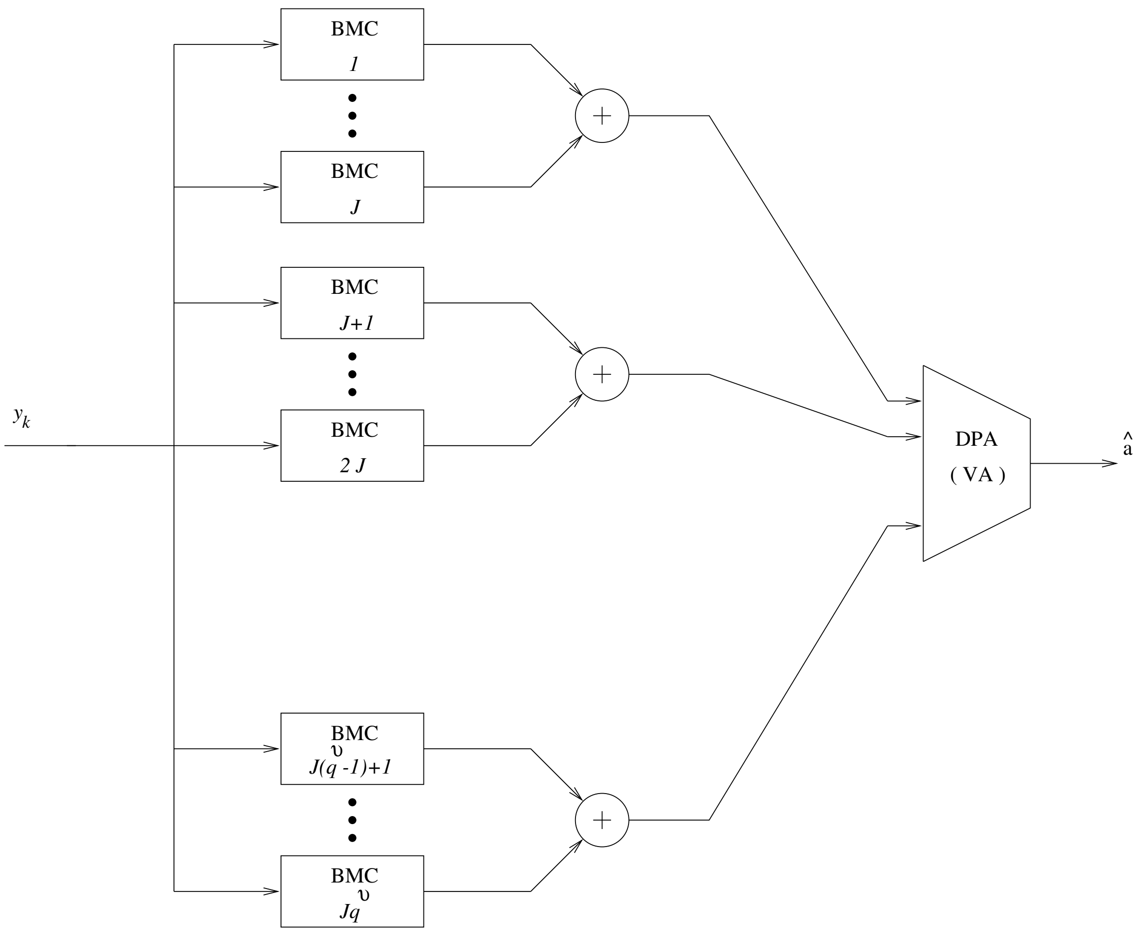

Figure 2.4: The Structure of the MLSD receiver

q-ary symbols, the number of states in the trellis diagram is $q^{\vartheta-1}$, where $\vartheta = \lceil \frac{L+L_s+L_R}{J} \rceil$ and the MLSD receiver for the detection of each symbol between i and $i+1$ stages in the trellis diagram calculates Jq^{ϑ} branch metrics from (2.18) and finds the survivor paths or minimum cost paths which terminate at each state in the trellis diagram.

The MLSD receiver whose branch metrics are computed from (2.21) is derived based on knowing the channel statistical parameters. However, the more informative situation is when the CIR is known by the receiver. By defining $\bar{\mathbf{h}} = [\boldsymbol{h}(K-1)^{\mathcal{T}}, ..., \boldsymbol{h}(0)^{\mathcal{T}}]^{\mathcal{T}}$ where $\boldsymbol{h}(k) = [h_0(k), ..., h_L(k)]^{\mathcal{T}}$, the MLSD criterion with known CIR is

$$C_{MLSD} \doteq p(\mathbf{y}|\bar{\mathbf{h}}, \hat{\mathbf{a}}) \quad = \quad \max_{\mathbf{a}\in\mathcal{A}} p(\mathbf{y}|\bar{\mathbf{h}}, \mathbf{a})$$

$$= \max_{\mathbf{a}\in\mathcal{A}}\{\prod_{i=0}^{I-1}\prod_{j=0}^{J-1} p(y_{iJ+j}|\boldsymbol{h}(iJ+j),\boldsymbol{a})\} \quad (2.23)$$

Defining $\boldsymbol{s}(k) = [s(k), ..., s(k-L)]$, from (2.7) and after taking logarithm from (2.23) it can be shown

$$C_{MLSD} \doteq \min_{\mathbf{a}\in\mathcal{A}}\{\sum_{i=0}^{I-1}\sum_{j=0}^{J-1} \beta(\boldsymbol{a}(iJ+j))\} \quad (2.24)$$

where the branch metric $\beta(\boldsymbol{a}(k))$ is given by[4]

$$\beta(\boldsymbol{a}(k)) = \frac{|y_k - \boldsymbol{s}(k)\boldsymbol{h}(k)|^2}{N_0} \quad (2.25)$$

$\boldsymbol{s}(k)$ is the output of the transmitter filter whose input is $\boldsymbol{a}(k)$. The structure of the MLSD receiver with known CIR is the same as the structure shown in Fig.2.4 when the branch metric calculator (BMC) blocks use (2.25) to compute the branch metrics. Meanwhile in this MLSD receiver $\vartheta = \lceil \frac{L_s+L}{J} \rceil$; therefore, the number of states in the trellis diagram is less than that in the trellis diagram of the MLSD receiver with known channel statistical parameters.

2.3 Diversity Reception

Errors in fading channels occur in a burst mode and the use of diversity is a well known method for improving the performance. The MLSD criterion for the reception of the same signal from D_r independently fading channels is

$$C_{MLSD} \doteq \max_{\mathbf{a}\in\mathcal{A}} p(\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, ..., \mathbf{y}^{(D_r)}|\mathbf{a}) \quad (2.26)$$

where $\mathbf{y}^{(d)} = \{y_{K-1}^{(d)}, y_{K-2}^{(d)}, ..., y_0^{(d)}\}$ is the received signal vector from channel d and $y_k^{(d)}$ is

$$y_k^{(d)} = \sum_{l=0}^{L} h^{(d)}(l,k) \sum_{n=k-l-L_s+1}^{k-l} b_n g((k-l-n)) + z^{(d)}(k) \quad d = 1, 2, ..., D_r \quad (2.27)$$

[4]Generally, knowing N_0 is not necessary for reception without diversity.

where $h^{(d)}(l,k)$ and $z^{(d)}(k)$ are the impulse response and additive Gaussian noise for the dth channel respectively. Based on the independence of D_r channels and the results of Section 2.2 we get

$$C_{MLSD} \equiv \min_{\mathbf{a}\in\mathcal{A}}\{\sum_{i=0}^{I-1}\sum_{j=0}^{J-1}\sum_{d=1}^{D_r}\beta^{(d)}(\boldsymbol{a}_{iJ+j})\} = \min_{\mathbf{a}\in\mathcal{A}}\{\sum_{i=0}^{I-1}\sum_{j=0}^{J-1}\boldsymbol{\beta}(\boldsymbol{a}_{iJ+j})\} \tag{2.28}$$

where the branch metric $\boldsymbol{\beta}(\boldsymbol{a}_k)$ is given by

$$\begin{aligned}\boldsymbol{\beta}(\boldsymbol{a}_k) &= \sum_{d=1}^{D_r} B^{(d)}(\boldsymbol{a}_k) = \sum_{d=1}^{D_r}\log(d_0^{(d)}(\boldsymbol{a}_k)) + \frac{\left|\sum_{m=0}^{M-1} y_{k-m}^{(d)}\rho_{m,0}^{(d)}(\boldsymbol{a}_k)\right|^2}{d_0^{(d)}(\boldsymbol{a}_k)} \\ &= \sum_{d=1}^{D_r}\log(d_0^{(d)}(\boldsymbol{a}_k)) + \frac{\left|e^{(d)}(\boldsymbol{a}_k)\right|^2}{d_0^{(d)}(\boldsymbol{a}_k)}\end{aligned} \tag{2.29}$$

Some performance evaluations, [49] [12], show that even using only one extra diversity decreases effectively the bit error rate of the MLSD receiver for fading channels. By doing similar manipulations, $\boldsymbol{\beta}(\boldsymbol{a}_k)$ for known CIR is given by

$$\boldsymbol{\beta}(\boldsymbol{a}_k) = \sum_{d=1}^{D_r}\frac{|y^{(d)}(k) - \boldsymbol{s}(k)\boldsymbol{h}^{(d)}(k)|^2}{N_0^{(d)}} \tag{2.30}$$

where $\boldsymbol{h}^{(d)}(k) = [h_0^{(d)}(k), ..., h_L^{(d)}(k)]^{\mathcal{T}}$ and $N_0^{(d)}$ denotes the variance of noise for the dth channel.

2.4 Computer Simulation

The performance of the MLSD receiver with known channel statistical parameters (known CSP) and known CIR have been evaluated by computer simulations for frequency flat (non-selective) and selective fading channels with $f_dT = 0.1$ and 0.01. The autocorrelation function of the CIR is modeled as

$$R_h(l_1,l_2;j) = \sum_{l_1=0}^{L} exp(-bl_1T_s)J_0(2\pi f_d jT_s)\delta(l_1-l_2) \quad 0\le l_1,l_2\le L \quad -\infty\le j\le\infty \tag{2.31}$$

where J_0 is the zero-order Bessel function, f_d is the maximum Doppler frequency. The delay rate in (2.31) was chosen as $b^{-1} = 2T_s$. $L = 0$ and $L = 2$ in (2.31) for flat fading and selective fading channels respectively (i.e. the channel was simulated with three paths in the selective channel). The impulse response of the transmitter filter is a raised-cosine pulse

$$g(t) = \mathrm{sinc}(\frac{t}{T})(\frac{cos(\nu\pi t/T)}{1-(2\nu t/T)^2}) \tag{2.32}$$

where the symbol duration $T = 1$ and $\nu = 0.35$. $g(k) = g(t)|_{t=kT_s}$ where $J = \frac{T}{T_s} = 2$. Meanwhile the differentially-encoded quadrature phase shift keying (DQPSK) modulation scheme was chosen. The Bessel fading filter for omnidirectional antenna is approximated by an all pole third order filter [56], $T_B(s)$, whose amplitude transfer function was shown in Fig.2.5. Therefore, the number of states is $N_s = 64$ and $N_s = 256$ for frequency flat and selective fading channels respectively in the MLSD receiver with known CSP and it is $N_s = 4$ and $N_s = 16$ respectively in the MLSD receiver with known CIR. The data sequence is divided into a sequence of frames with length L_f, where the overhead of each frame known by the receiver is three (one) and four (two) for flat and selective fading respectively in the MLSD receiver with known CSP (known CIR). The length of each frame was chosen as $L_f = 160$ data.

The bit error rate performance of the MLSD receivers with known CSP and known CIR in flat fading channel are shown in Fig.2.6 and Fig.2.7 for $f_dT = 0.1$ and $f_dT = 0.01$ respectively. As expected for both fading rates, the bit error rate for the known CIR is less than that for the known CSP. However by decreasing the fading rate the difference between the performances is decreased. Also, the performance of the MLSD receivers for selective fading channel are shown in Fig.2.8 and Fig.2.9 for $f_dT = 0.1$ and 0.01 respectively. As shown in the figures, in both cases the performance of known CIR is significantly better than that of known CSP. It is clear that knowing the CIR, which helps the receiver to achieve better performance with less number of states, is more informative than knowing CSP.

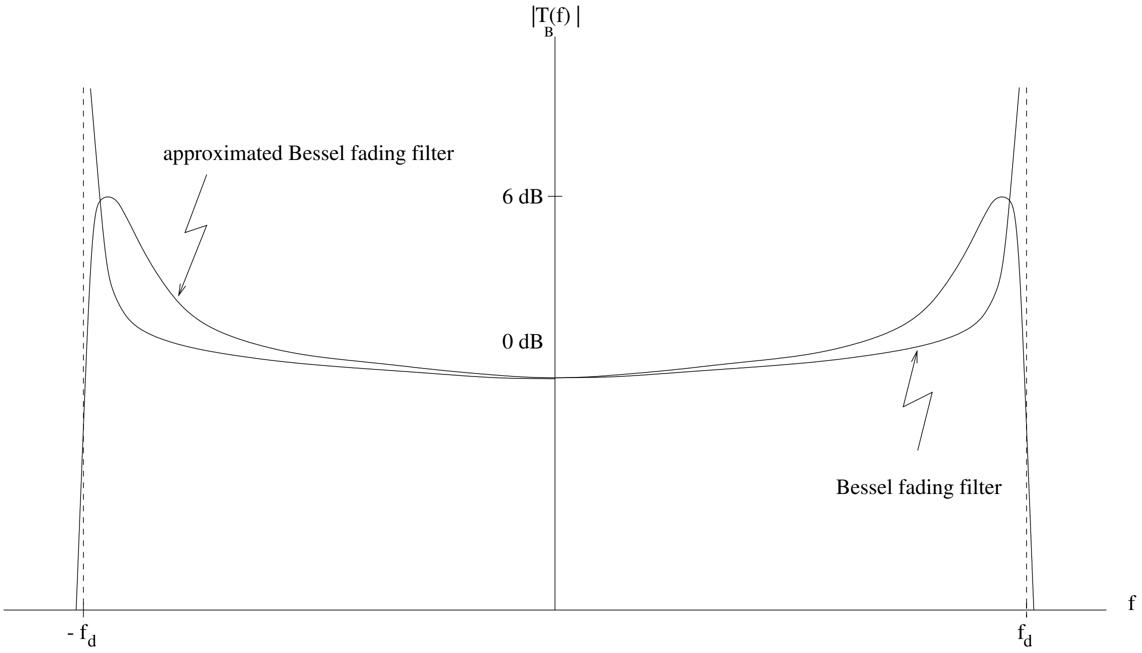

Figure 2.5: The frequency spectrum of the ideal and approximated Bessel fading filter.

2.5 Estimation of Channel Parameters

The branch metrics of the MLSD receiver derived in Sections 2.2 and 2.3 are computed based on known channel statistical parameters, $R_h(l_1, l_2, j)$, and the transmitted signal shape $s(k)$. The transmitted signal for different hypotheses of $\boldsymbol{a}_k$ is easy to obtain by knowing the impulse response of transmitter filter $g(k)$. However, the statistical parameters of the channel should be estimated based on the received signal. (2.19) shows the relation between the autocorrelation of the received signal and channel impulse response. The straightforward method to estimate $R_{\boldsymbol{y}}(i, j)$ is to use time-averaging method with a forgetting factor to accommodate the slow non-stationarity of the channel. However, there is no "best" method to estimate $R_h(l_1, l_2, j)$ from the estimation of $R_{\boldsymbol{y}}(i, j)$. By using iterative methods, such as Newton's algorithm for solving non-linear equations, $R_h(l_1, l_2, j)$ can be estimated based on the least squares criterion [57]. Since $R_{\boldsymbol{y}}(i, j)$ is computed for given $\boldsymbol{a}_k$, $R_h(l_1, l_2, j)$ should be estimated for all survivor paths at each time and then the coefficients of the predictor filters, $\rho_{m,0}(\boldsymbol{a})$, should be found for all the branch metrics emanating from each state.

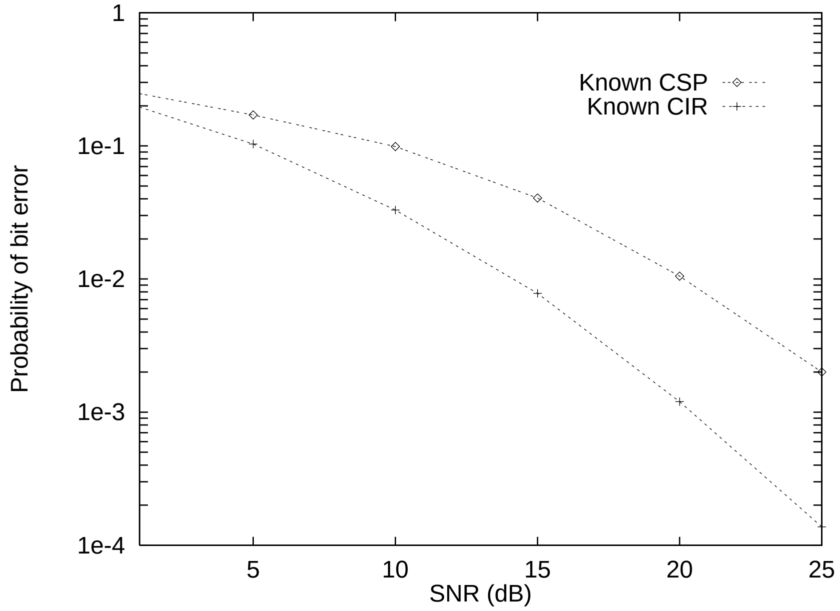

Figure 2.6: Bit error rate performance of MLSD receiver with known channel statistical parameters (CSP) and channel impulse response (CIR) for flat fading with $f_d T = 0.1$ and DQPSK signaling.

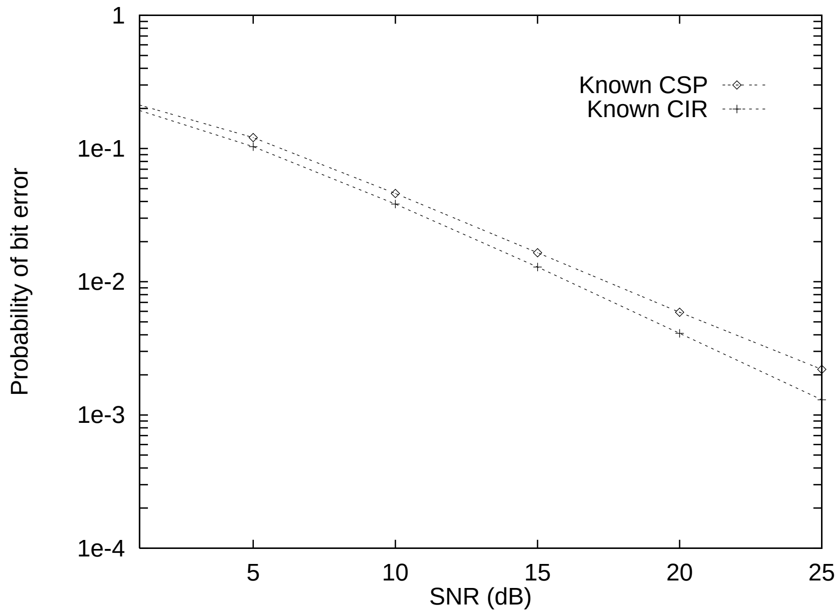

Figure 2.7: Bit error rate performance of MLSD receiver with known channel statistical parameters (CSP) and channel impulse response (CIR) for flat fading with $f_d T = 0.01$ and DQPSK signaling.

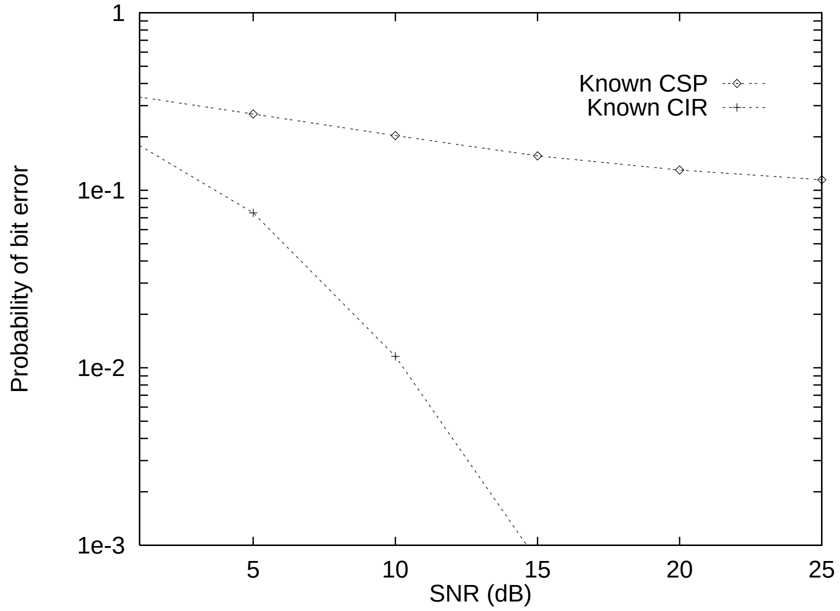

Figure 2.8: Bit error rate performance of MLSD receiver with known channel statistical parameters (CSP) and channel impulse response (CIR) for frequency selective fading with $f_d T = 0.1$ and DQPSK signaling.

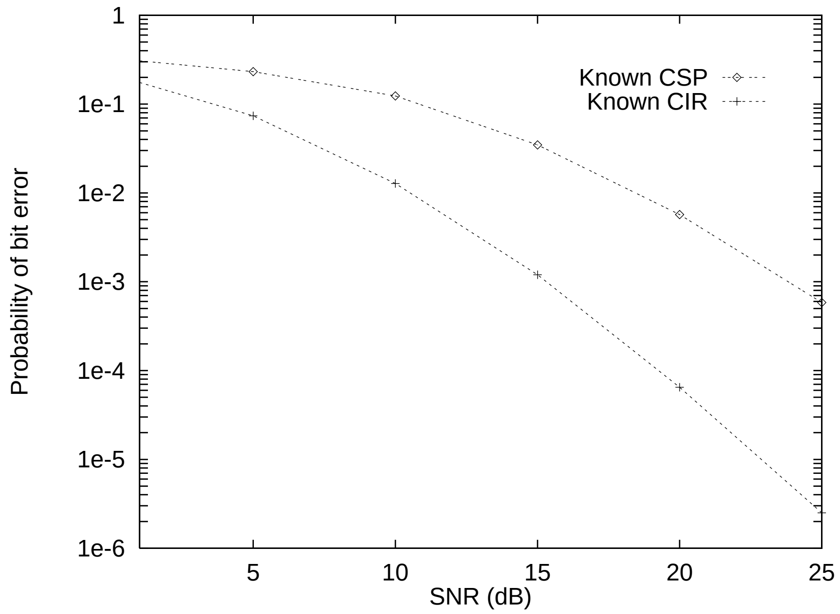

Figure 2.9: Bit error rate performance of MLSD receiver with known channel statistical parameters (CSP) and channel impulse response (CIR) for frequency selective fading with $f_d T = 0.01$ and DQPSK signaling.

It is clear that such an estimation scheme is very complicated and close to becoming impractical especially for channels with long memory. Moreover, for estimating $R_h(l_1, l_2, j)$ only the autocorrelation function of the received signal is necessary; however, the received signal $y(k)$, which is more informative, is available and it seems that by estimating $R_h(l_1, l_2, j)$ from the autocorrelation function of $y(k)$ some parts of the available information about the channel has not been used. Therefore the proposed MLSD structure in this chapter is suitable when the statistical parameters of channel are available and their estimation is not necessary.

The other method to derive a MLSD receiver is based on known CIR. Since the CIR is unknown, it should also be estimated by the receiver and instead of its real value, its estimate should be used in the detection. Since the MLSD receiver with known CIR has a potentially better performance and less computational complexity, we consider more details of such a receiver in the two subsequent chapters. We focus on the estimation part of the receiver based on maximum likelihood (ML) criterion and using the expectation and maximization (EM) algorithm in chapter 3 and then on the detection part of the receiver in chapter 4.

2.6 Summary

In this chapter, a discrete model for a Rayleigh multipath fading channel has been developed for a bandlimited signal based on the sampling theorem for linear time-variant systems. The relation between the statistical parameters of discrete and analog channel models has also been shown.

The structure of the MLSD receivers, based on known statistical parameters (autocorrelation function) and known impulse response of the channel, has been derived for frequency flat (non-selective) and selective Rayleigh fading channels. The structure of the MLSD receiver with known statistical parameters contains a bank of predictor filters with constant coefficients whose outputs provide the branch metrics for the dynamic programming algorithm to detect the transmitted signal. This detection method is blind and does not need extra information to be transmitted periodically to the receiver.

In real situations the channel parameters (autocorrelation function or impulse response) should be estimated. However, the procedure of estimating the autocorrelation function of the CIR directly from autocorrelation function of the received signal is very complicated and also its ideal performance is less than the potential performance of the MLSD receiver with CIR estimation. Therefore, we consider the MLSD scheme based on estimating the CIR and then finding its statistical parameters as a suitable candidate for a receiver whose details will be investigated in the following chapters. Moreover, estimating the CIR also provides the opportunity to model it as a deterministic vector/process as well.

Chapter 3

EM-Based Recursive Estimation of Channel Parameters

Estimation of the transmission channel plays a crucial role in many detection algorithms used in communication systems. Channel parameters which are embedded in white or colored Gaussian noise are observed indirectly through the received data. Some criteria such as minimum mean squared error (MMSE), maximum likelihood (ML) and min-max are generally used in estimating the unknown parameters. ML criterion is considered as an optimal detection method when the transmitted symbols are equiprobable and as a benchmark in estimation when unknown parameters are deterministic. However in many cases the received data does not provide complete information necessary to maximize the likelihood function. The expectation and maximization (EM) algorithm [38] [14] provides ML estimation of parameters when maximization of the likelihood function may not be feasible directly.

The EM algorithm is an iterative procedure with expectation and maximization steps. In the first step the conditional expectation of unobserved sufficient information (complete data) is taken under given observed insufficient information (incomplete data) and the current estimates of parameters. The second step provides new estimates of the parameters by maximizing the conditional expectation over the unknown parameters.

Applications of the EM algorithm in parameter estimation have been considered in [58–62]. Feder and Weinstein [58] developed a forward and backward algorithm for parameter estimation of superimposed signals. Iterative and sequential parameter estimation using the EM algorithm were proposed in [59] and [60] based on employing the RLS and Kalman algorithm in estimating Gaussian random vector/process. The on-line estimation of parameters based on the Kullback-Leibler information measure using stochastic approximations method was considered in [61] and [62]. Also several other applications of the EM algorithm in receiver design have been considered in [63–67].

EM algorithm is a batch-oriented approach which processes the entire received data. In order to eliminate the delay in decision-making, reduce storage and increase the computational efficiency in real-time applications, it is desirable and often necessary to process the received data in an on-line manner.

By modifying Titterington's stochastic approximation approach [43] and using the on-line EM algorithm, numerous recursive estimating procedures are developed in this chapter based on the ML criterion. Although achieving ML estimation is not always guaranteed, the recursive estimator increases the likelihood monotonically. Also due to on-line processing, the estimator is adaptive and can track the time-varying parameters of deterministic or locally stationary processes. In association with different channel impulse response (CIR) models such as unknown deterministic and Gaussian random vector/process with unknown stochastic characteristics embedded in Gaussian noise, the developed recursive estimation algorithms lead to RLS/Kalman-type, modified RLS, stochastic RLS, Kalman filtering or smoothing and combined RLS and Kalman-type algorithms. The RLS/Kalman-type algorithm, which estimates the unknown deterministic vector/process embedded in an additive colored Gaussian noise modeled by an autoregressive process, is novel. Also, the combined RLS and Kalman-type is a new algorithm which recursively estimates CIR and its ARMA model parameters.

In a recent comprehensive survey paper, Sayed and Kailath [68] showed how the entire family of RLS-type algorithms can be formulated in a unified manner within a state space model. Here we take such an approach one step further and show how we can derive many RLS/Kalman-type adaptive algorithms, where some of them are new, through a unified recipe based on ML criterion using the on-line EM algorithm. Meanwhile, although we concentrate on communication systems and channel estimation in this chapter, the results can also be applied to other estimation, detection and identification problems which can be modeled with state space equations.

Following the description of the system model in Section 3.1, ML estimation based on the EM algorithm is introduced in Section 3.2. In Section 3.3, a recursive estimation method for time-invariant/variant unknown parameters is developed based on the on-line EM algorithm and stochastic approximations. Estimation algorithms for different impulse response models in linear channels are derived in Section 3.4.

3.1 System Model

Data transmission through a linear noisy channel can be described as follows.

$$y(k) = \sum_{l} s(k-l)h(l,k) + z(k) \qquad k = 0, ..., K-1 \tag{3.1}$$

where $s(k)$ is the transmitted signal, $h(l,k)$ is the impulse response of the linear channel[1], $z(k)$ is additive noise generally modeled as a complex, circularly symmetric, white/colored Gaussian random process and $y(k)$ is the received signal. To detect the transmitted data, the receiver needs to know channel parameters such as the impulse response for deterministic channels or stochastic characteristics of the impulse response for stochastic channels. The channel parameters cannot be observed directly by the receiver and are considered as hidden parameters or hidden states.

The communication channel, usually suffering from impairment such as intersymbol interference (ISI) and multipath fading , is generally modeled as a discrete finite memory system whose impulse response length is limited to $L+1$, $h(l,k) = 0$ for $0 > l > L$. When $h(l,k)$ is considered as a random process it can be modeled by a hidden Markov chain and the receiver detects the transmitted data based on estimating the channel parameters.

In this chapter we focus on estimating the channel parameters ($h(l,k)$ or its statistical parameters at each time) and assume that the receiver has knowledge about $s(k)$.

[1] $h(l,k)$ is the time-variant channel response at time k due to an impulse applied at time $k-l$.

The knowledge about $s(k)$ can be achieved in the detection algorithm based on the estimation of $s(k)$ in a decision feedback equalization method or in association with different hypotheses of $s(k)$ in a maximum likelihood sequence detection (MLSD) method or when the communication system is in the training mode [7, 8, 11, 12, 49].

The deterministic parameters can be estimated based on different criteria. We choose the ML criterion which can be implemented by the EM algorithm.

3.2 ML Estimation

Let us consider $\boldsymbol{\theta}$ as a column vector of deterministic channel parameters to be estimated from the observed data vector $\mathbf{y} = [y(K-1), ..., y(0)]^{\mathcal{T}}$. ML estimation of $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}}\{p(\mathbf{y}|\boldsymbol{\theta})\} \equiv \arg\max_{\boldsymbol{\theta}}\{\log p(\mathbf{y}|\boldsymbol{\theta})\} \tag{3.2}$$

When $\mathbf{y}$ is incomplete information, the maximization of $p(\mathbf{y}|\boldsymbol{\theta})$, the conditional probability density function (pdf) of $\mathbf{y}$ given $\boldsymbol{\theta}$, is not tractable and does not lead to an explicit solution. Denoting $\mathbf{y}$ as the incomplete data $\mathcal{I} = \mathbf{y}$ and $\mathcal{D}$ as the desired additional information needed to complete $\mathcal{I}$, the ML estimation of $\boldsymbol{\theta}$ becomes

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}}\{\log(\int p(\mathcal{C}|\boldsymbol{\theta})d\mathcal{D})\} \tag{3.3}$$

where $\mathcal{C} = \{\mathcal{I}, \mathcal{D}\}$ is defined as the complete data. In many cases, it is hard to obtain a closed-form solution for (3.3). Under some regularity conditions an alternative solution is the EM algorithm [38] [14]. The EM algorithm is an iterative procedure to achieve the ML estimation of unknown parameters $\boldsymbol{\theta}$. From the definition of $\mathcal{C}$ and Bayes' rule one can write

$$\log p(\mathcal{I}|\boldsymbol{\theta}) = \log p(\mathcal{C}|\boldsymbol{\theta}) - \log p(\mathcal{D}|\mathcal{I}, \boldsymbol{\theta}) \tag{3.4}$$

Taking the conditional expectation of both side of (3.4) over $\mathcal{D}$ given $\mathcal{I}$ at parameter values $\hat{\boldsymbol{\theta}}^{(l)}$ (the estimation of $\boldsymbol{\theta}$ in lth iteration or initial value of $\boldsymbol{\theta}$), we get [38]

$$\log p(\mathcal{I}|\boldsymbol{\theta}) = E[\log p(\mathcal{C}|\boldsymbol{\theta})|\mathcal{I}, \hat{\boldsymbol{\theta}}^{(l)}] - E[\log p(\mathcal{D}|\mathcal{I}, \boldsymbol{\theta})|\mathcal{I}, \hat{\boldsymbol{\theta}}^{(l)}] \tag{3.5}$$

For convenience we define

$$\begin{aligned} L(\boldsymbol{\theta}) &= \log p(\mathcal{I}|\boldsymbol{\theta}) \\ Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(l)}) &= E[\log p(\mathcal{C}|\boldsymbol{\theta})|\mathcal{I}, \hat{\boldsymbol{\theta}}^{(l)}] \\ V(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(l)}) &= E[\log p(\mathcal{D}|\mathcal{I}, \boldsymbol{\theta})|\mathcal{I}, \hat{\boldsymbol{\theta}}^{(l)}] \end{aligned}$$

Jensen's inequality for a concave function implies

$$V(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(l)}) \leq V(\hat{\boldsymbol{\theta}}^{(l)}, \hat{\boldsymbol{\theta}}^{(l)}) \tag{3.6}$$

If $Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(l)}) > Q(\hat{\boldsymbol{\theta}}^{(l)}, \hat{\boldsymbol{\theta}}^{(l)})$, it can be guaranteed $L(\boldsymbol{\theta}) > L(\hat{\boldsymbol{\theta}}^{(l)})$. Therefore by maximizing $Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(l)})$ with respect to $\boldsymbol{\theta}$ causes the log-likelihood function $L(\boldsymbol{\theta})$ to be monotonically increasing .The idea of increasing the likelihood function iteratively is the main core of the EM algorithm which has two steps; expectation and maximization. In the first step the conditional expectation of the complete data given the most recent estimates of the parameters ,$Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(l)})$, is computed and in the second step the new estimation of the parameters, $\hat{\boldsymbol{\theta}}^{(l+1)}$, is chosen by maximizing $Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(l)})$ over $\boldsymbol{\theta}$. The steps of the EM algorithm at $(l+1)$th iteration are

E-step:

$$Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(l)}) = E[\log p(\mathcal{C}|\boldsymbol{\theta})|\mathcal{I}, \hat{\boldsymbol{\theta}}^{(l)}] \tag{3.7}$$

M-step:

$$\hat{\boldsymbol{\theta}}^{(l+1)} = \arg\max_{\boldsymbol{\theta}}\{Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(l)})\} \tag{3.8}$$

The steps of the EM algorithm are repeated iteratively until at $\acute{l}$th iteration $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}^{(\acute{l})} = \hat{\boldsymbol{\theta}}^{(\acute{l}-1)}$ or $\|\hat{\boldsymbol{\theta}}^{(\acute{l})} - \hat{\boldsymbol{\theta}}^{(\acute{l}-1)}\| < \epsilon_T$ where ϵ_T is a small threshold. If $L(\boldsymbol{\theta})$ has only one maximum, $\hat{\boldsymbol{\theta}}$ is the ML estimate of $\boldsymbol{\theta}$; however, when $L(\boldsymbol{\theta})$ has many local maxima, $\hat{\boldsymbol{\theta}}$ is a local ML estimate of $\boldsymbol{\theta}$ which may be different for different initial values ,$\hat{\boldsymbol{\theta}}^{(0)}$.

3.3 Recursive Estimation Using the On-Line EM Algorithm

The steps of the EM algorithm (3.7) and (3.8) have to be done iteratively based on the entire set of data and parameters and hence the algorithm is off-line. In communication (or real time processing) systems, the complete data, incomplete data and even sometimes the unknown parameters are sequentially evolving processes in time and estimation of the channel parameters needs real-time processing. Therefore, we are interested more in an on-line version of the EM algorithm. Meanwhile, due to the high dimensionality of $\mathcal{C}$ and $\boldsymbol{\theta}$ in the off-line EM algorithm, the required computational complexity and memory are also very high. The convergence of the EM algorithm is inversely related to the dimensionality of its complete data [69]; thus a less-informative complete data space improves the asymptotic convergence rate [70]. Furthermore, when the dimension of the $\boldsymbol{\theta}$-space is increasing with an increase in the dimension of $\mathcal{C}$-space, using the on-line algorithm leads to a decrease in $\boldsymbol{\theta}$-space as well and thus it causes faster convergence. This is similar to the space alternating generalized EM (SAGE) method in which the space of the unknown parameters is partitioned to achieve faster convergence [70].

Let us define $\mathcal{C}_k$, $\mathcal{I}_k$ and $\boldsymbol{\theta}_k = [\boldsymbol{\varphi}_k^{\mathcal{T}}, \boldsymbol{\varphi}_{k-1}^{\mathcal{T}}, ..., \boldsymbol{\varphi}_0^{\mathcal{T}}]^{\mathcal{T}}$ as the complete data, incomplete data and unknown parameters available up to time k respectively, where $\boldsymbol{\varphi}_l$ is a column vector of unknown parameters just at time l and $\boldsymbol{\varphi}^{\mathcal{T}}$ denotes $\boldsymbol{\varphi}$ transpose. Two steps of the on-line EM algorithm at time k are

1-E step

$$Q_k(\boldsymbol{\theta}_k|\tilde{\boldsymbol{\theta}}_{k|k-1}) = E[\log p(\mathcal{C}_k|\boldsymbol{\theta}_k)|\mathcal{I}_k, \tilde{\boldsymbol{\theta}}_{k|k-1}] \tag{3.9}$$

2-M step

$$\tilde{\boldsymbol{\theta}}_{k|k} = \arg\max_{\boldsymbol{\theta}_k}\{Q_k(\boldsymbol{\theta}_k|\tilde{\boldsymbol{\theta}}_{k|k-1})\} \tag{3.10}$$

where $\tilde{\boldsymbol{\theta}}_{k|l}$ is the estimate of $\boldsymbol{\theta}_k$ based on the signal received up to time l. Time-up-date vector $\tilde{\boldsymbol{\theta}}_{k+1|k}$ is given by

$$\tilde{\boldsymbol{\theta}}_{k+1|k} = \begin{bmatrix} \tilde{\boldsymbol{\varphi}}_{k+1|k} \\ \tilde{\boldsymbol{\theta}}_{k|k} \end{bmatrix} \tag{3.11}$$

where $\tilde{\boldsymbol{\varphi}}_{k+1|k}$ is the estimate of unknown parameters at time $k+1$ based on the entire signal received up to time k. In general $\tilde{\boldsymbol{\varphi}}_{k+1|k}$ can be a function of $\tilde{\boldsymbol{\theta}}_{k|k}$ and is obtained by using the dynamic evolution of the $\boldsymbol{\varphi}_{k+1}$ process.

The estimation procedure based on the on-line EM algorithm is more attractive when its maximization step can be done analytically in a recursive manner. Recursive estimation methods are computationally efficient and provide *up-to-date* results based on the new received information. Moreover, the recursive algorithm can track time-variant parameters or locally stationary processes in an adaptive manner.

Titterington proposed a recursive estimation method for constant unknown parameters based on the EM algorithm by observing independent sequential data [43]. We follow Titterington's approach of stochastic approximation, to develop a recursive algorithm for the on-line EM algorithm by relaxing the assumptions of constant parameters and independence of received data. The Taylor's expansion of $Q_k(\boldsymbol{\theta}_k|\tilde{\boldsymbol{\theta}}_{k|k-1})$ at point $\tilde{\boldsymbol{\theta}}_{k|k-1}$ is

$$\begin{aligned} Q_k(\boldsymbol{\theta}_k|\tilde{\boldsymbol{\theta}}_{k|k-1}) &= Q_k(\boldsymbol{\theta}_k|\tilde{\boldsymbol{\theta}}_{k|k-1})|_{\boldsymbol{\theta}_k=\tilde{\boldsymbol{\theta}}_{k|k-1}} + (\boldsymbol{\theta}_k - \tilde{\boldsymbol{\theta}}_{k|k-1})^{\mathcal{H}} \frac{\partial Q_k(\boldsymbol{\theta}_k|\tilde{\boldsymbol{\theta}}_{k|k-1})}{\partial \boldsymbol{\theta}_k^*}|_{\boldsymbol{\theta}_k=\tilde{\boldsymbol{\theta}}_{k|k-1}} \\ &+ (\boldsymbol{\theta}_k - \tilde{\boldsymbol{\theta}}_{k|k-1})^{\mathcal{H}} \frac{\partial^2 Q_k(\boldsymbol{\theta}_k|\tilde{\boldsymbol{\theta}}_{k|k-1})}{\partial^2 \boldsymbol{\theta}_k}|_{\boldsymbol{\theta}_k=\tilde{\boldsymbol{\theta}}_{k|k-1}} (\boldsymbol{\theta}_k - \tilde{\boldsymbol{\theta}}_{k|k-1}) + \ldots \end{aligned} \tag{3.12}$$

Where $\frac{\partial^2 Q_k(\boldsymbol{\theta}_k|\tilde{\boldsymbol{\theta}}_{k|k-1})}{\partial^2 \boldsymbol{\theta}_k} \doteq \frac{\partial^2 Q_k(\boldsymbol{\theta}_k|\tilde{\boldsymbol{\theta}}_{k|k-1})}{\partial \boldsymbol{\theta}_k^* \partial \boldsymbol{\theta}_k^T}$, $\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}^{\mathcal{H}}$ denote conjugate and conjugated transposed of $\boldsymbol{\theta}$ respectively. Approximation of $Q_k(\boldsymbol{\theta}_k|\tilde{\boldsymbol{\theta}}_{k|k-1})$ with three elements of its Taylor's series and maximizing it with respect to $\boldsymbol{\theta}_k^*$ at point $\tilde{\boldsymbol{\theta}}_{k|k-1}$ produces a recursive formula given by [2]

$$\tilde{\boldsymbol{\theta}}_{k|k} \simeq \tilde{\boldsymbol{\theta}}_{k|k-1} - \left(\frac{\partial^2 Q_k(\boldsymbol{\theta}_k|\tilde{\boldsymbol{\theta}}_{k|k-1})}{\partial^2 \boldsymbol{\theta}_k}\Big|_{\boldsymbol{\theta}_k=\tilde{\boldsymbol{\theta}}_{k|k-1}}\right)^{-1}\left(\frac{\partial Q_k(\boldsymbol{\theta}_k|\tilde{\boldsymbol{\theta}}_{k|k-1})}{\partial \boldsymbol{\theta}_k^*}\Big|_{\boldsymbol{\theta}_k=\tilde{\boldsymbol{\theta}}_{k|k-1}}\right) \tag{3.13}$$

where if $\boldsymbol{\varphi}_k$ is an $L+1$ vector, $\boldsymbol{\theta}_k$ and $\frac{\partial Q_k(\boldsymbol{\theta}_k|\tilde{\boldsymbol{\theta}}_{k|k-1})}{\partial \boldsymbol{\theta}_k^*}$ are $(k+1)(L+1)$ vectors and also $\frac{\partial^2 Q_k(\boldsymbol{\theta}_k|\tilde{\boldsymbol{\theta}}_{k|k-1})}{\partial^2 \boldsymbol{\theta}_k}$ is $(k+1)(L+1) \times (k+1)(L+1)$ matrix. Meanwhile the time-up-date vector $\tilde{\boldsymbol{\theta}}_{k+1|k}$ is obtained from(3.11). As seen from (3.13), entire unknown processes *up to* time k are estimated in the kth recursion based on their estimation at time $k-1$ and this is known as smoothing in the literature; estimating parameters only *for* time k is called filtering. Since only three elements of Taylor's expansion are used in developing the recursive formula, (3.13) is an approximation of ML estimation. However, when the third and higher derivatives of $Q_k(\boldsymbol{\theta}_k|\tilde{\boldsymbol{\theta}}_{k|k-1})$ are zero, as is usually true for the Gaussian case, the recursive formula (3.13) is exact [44].

3.4 Channel Estimation

In this section we consider channel estimation (CIR or its statistical parameters at each time) for different channel models using the recursive on-line EM algorithm developed in the previous section. It should be mentioned that the algorithms developed in this section can also be applied to other areas in estimation and detection.

a) Unknown Deterministic CIR

In this model CIR is considered as an unknown vector of deterministic parameters which may be time-invariant or time-variant. The received signal is expressed as

$$y(k) = \sum_{l=0}^{L} s(k-l)h(l,k) + z(k) = \boldsymbol{s}(k)\boldsymbol{h}(k) + z(k) \tag{3.14}$$

where $\boldsymbol{s}(k) = [s(k), ..., s(k-L)]$ and $\boldsymbol{h}(k) = [h(0,k), ..., h(L,k)]^{\mathcal{T}}$. Assume the dynamic

---

[2]When $\frac{\partial Q_k(.|.)}{\partial \boldsymbol{\theta}_k^*}|_{\boldsymbol{\theta}_k=\tilde{\boldsymbol{\theta}}_{k|k-1}} = 0$, one can conclude $\frac{\partial Q_k(.|.)}{\partial \boldsymbol{\theta}_k}|_{\boldsymbol{\theta}_k=\tilde{\boldsymbol{\theta}}_{k|k-1}} = 0$

change in linear time-variant CIR to be

$$\mathbf{h}(k) = \boldsymbol{F}(k)\mathbf{h}(k-1) \tag{3.15}$$

where $\mathbf{h}_k = \mathbf{h}(k) = [\boldsymbol{h}(k)^{\mathcal{T}}, ..., \boldsymbol{h}(k-M+1)^{\mathcal{T}}]^{\mathcal{T}}$ and $\boldsymbol{F}(k)$ is an $M(L+1) \times M(L+1)$ matrix. $z(k)$ is a stationary colored Gaussian noise modeled by an autoregressive process with order $N-1$ (AR(N-1)) and covariance matrix $\Sigma_z = \text{cov}(\boldsymbol{z}_k) = E[(\boldsymbol{z}_k - \boldsymbol{\mu}_z)(\boldsymbol{z}_k - \boldsymbol{\mu}_z)^{\mathcal{H}}]$, where $\boldsymbol{z}_k = [z(k), ..., z(k-N+1)]^{\mathcal{T}}$, $\boldsymbol{\mu}_z = E[\boldsymbol{z}_k]$. From (3.14) and (3.15) we have

$$\boldsymbol{y}_k^{\diamond} = \begin{bmatrix} y(k) \\ y(k-1) \\ \vdots \\ y(k-N+1) \end{bmatrix} = \begin{bmatrix} \mathbf{s}(k) & & & \\ & \mathbf{s}(k-1) & & 0 \\ & 0 & \ddots & \\ & & & \mathbf{s}(k-N+1) \end{bmatrix} \begin{bmatrix} \mathbf{h}_k \\ \mathbf{h}_{k-1} \\ \vdots \\ \mathbf{h}_{k-N+1} \end{bmatrix} + \boldsymbol{z}_k \tag{3.16}$$

$$= S(k)\mathbf{h}_k^{\diamond} + \boldsymbol{z}_k \tag{3.17}$$

where $\mathbf{s}(k) = [\boldsymbol{s}(k), \mathbf{0}]$ and $\mathbf{0}$ is an $(M-1)(L+1)$ zero row vector. The unknown parameter vector up to time k is $\bar{\mathbf{h}}_k = [\mathbf{h}_k^{\mathcal{T}}, ..., \mathbf{h}_0^{\mathcal{T}}]^{\mathcal{T}}$ in this model. Also, the complete and incomplete data are defined as $\mathcal{C}_k = \mathcal{I}_k = \mathbf{y}_k = [y(k), ..., y(0)]^{\mathcal{T}}$. Thus, $Q_k(\bar{\mathbf{h}}_k|\tilde{\bar{\mathbf{h}}}_{k|k-1})$ is given as

$$\begin{aligned} Q_k(\bar{\mathbf{h}}_k|\tilde{\bar{\mathbf{h}}}_{k|k-1}) &= E[\log p(\mathbf{y}_k|\bar{\mathbf{h}}_k)|\mathbf{y}_k, \tilde{\bar{\mathbf{h}}}_{k|k-1}] \\ &= \log p(y(k)|\bar{\mathbf{h}}_k, \mathbf{y}_{k-1}) + \log p(\mathbf{y}_{k-1}|\bar{\mathbf{h}}_{k-1}) \end{aligned} \tag{3.18}$$

Since $z(k)$ is an AR(N-1) model, we have

$$E[y(k)|\bar{\mathbf{h}}_k, \mathbf{y}_{k-1}] = E[y(k)|\mathbf{h}_k^{\diamond}, y(k-1), ..., y(k-N+1)] \tag{3.19}$$

$y(k)$ is a Gaussian process thus (3.19) leads to [53]

$$p(y(k)|\bar{\mathbf{h}}_k, \mathbf{y}_{k-1}) = p(y(k)|\mathbf{h}_k^{\diamond}, y(k-1), ..., y(k-N+1)) \tag{3.20}$$

Therefore (3.18) becomes

$$\begin{aligned} Q_k(\bar{\mathbf{h}}_k|\tilde{\bar{\mathbf{h}}}_{k|k-1}) &= \log\{\frac{p(\boldsymbol{y}_k^{\diamond}|\mathbf{h}_k^{\diamond})}{p(y(k-1), ..., y(k-N+1)|\mathbf{h}_k^{\diamond})}\} + Q_{k-1}(\bar{\mathbf{h}}_{k-1}|\tilde{\bar{\mathbf{h}}}_{k-1|k-2}) \\ &= -\{\log(\pi) + \log(\det(\Sigma_z)) - \log(\det(\Sigma_{\acute{z}})) + (\boldsymbol{y}_k^{\diamond} - S(k)\mathbf{h}_k^{\diamond} - \boldsymbol{\mu}_z)^{\mathcal{H}} \\ &\quad \times(\Sigma_z^{-1} - \begin{bmatrix} 0 & \vdots & \mathbf{0} \\ \dots & \dots & \dots \\ \mathbf{0} & \vdots & \Sigma_{\acute{z}}^{-1} \end{bmatrix})(\boldsymbol{y}_k^{\diamond} - S(k)\mathbf{h}_k^{\diamond} - \boldsymbol{\mu}_z)\} \\ &\quad +Q_{k-1}(\bar{\mathbf{h}}_{k-1}|\tilde{\bar{\mathbf{h}}}_{k-1|k-2}) \end{aligned} \tag{3.21}$$

where $\acute{\boldsymbol{z}}_k = [z(k-1), ..., z(k-N+1)]^{\mathcal{T}}$ and $\Sigma_{\acute{z}} = \text{cov}(\acute{\boldsymbol{z}}_k)$. Σ_z^{-1} and $\Sigma_{\acute{z}}^{-1}$ are symmetric positive definite matrices and by using Cholesky decomposition they can be factorized into a product of two triangular matrices which are complex conjugated transposes of each other [54]. After doing some manipulations we can show [49]

$$\begin{aligned} Q_k(\bar{\mathbf{h}}_k|\tilde{\bar{\mathbf{h}}}_{k|k-1}) &= -\{\log(\pi) - \log(\gamma_0\gamma_0^*) + (\boldsymbol{y}_k^{\diamond} - S(k)\mathbf{h}_k^{\diamond} - \boldsymbol{\mu}_z)^{\mathcal{H}}\boldsymbol{\gamma}\boldsymbol{\gamma}^{\mathcal{H}} \\ &\quad \times(\boldsymbol{y}_k^{\diamond} - S(k)\mathbf{h}_k^{\diamond} - \boldsymbol{\mu}_z)\} + Q_{k-1}(\bar{\mathbf{h}}_{k-1}|\tilde{\bar{\mathbf{h}}}_{k-1|k-2}) \end{aligned} \tag{3.22}$$

where $\boldsymbol{\gamma} = [\gamma_0, \gamma_1, ..., \gamma_{N-1}]^{\mathcal{T}}$ is the first column of the lower triangular matrix of the Cholesky decomposition of Σ_z^{-1}. Due to maximization step at time $k-1$, the first derivative of $Q_k(\bar{\mathbf{h}}_k|\tilde{\bar{\mathbf{h}}}_{k|k-1})$ with respect to $\mathbf{h}_l^{\diamond *}$ at point $\bar{\mathbf{h}}_k = \tilde{\bar{\mathbf{h}}}_{k|k-1}$ is

$$\frac{\partial Q_k(\bar{\mathbf{h}}_k|\tilde{\bar{\mathbf{h}}}_{k|k-1})}{\partial \mathbf{h}_l^{\diamond *}}|_{\bar{\mathbf{h}}_k=\tilde{\bar{\mathbf{h}}}_{k|k-1}} = \begin{cases} \frac{\partial Q_{k-1}(\bar{\mathbf{h}}_{k-1}|\tilde{\bar{\mathbf{h}}}_{k-1|k-2})}{\partial \mathbf{h}_l^{\diamond *}}|_{\bar{\mathbf{h}}_{k-1}=\tilde{\bar{\mathbf{h}}}_{k-1|k-1}} = 0 & 0 \leq l \leq k-1 \\ S(k)^{\mathcal{H}}\boldsymbol{\gamma}\boldsymbol{\gamma}^{\mathcal{H}}(\boldsymbol{y}_k^{\diamond} - S(k)\tilde{\mathbf{h}}_{k|k-1}^{\diamond} - \boldsymbol{\mu}_z) & l = k \end{cases} \tag{3.23}$$

Since the first derivative of $Q_k(.|.)$ with respect to $\mathbf{h}_l^{\diamond *}$ is zero for $l \neq k$, only the estimate of $\mathbf{h}_k^\diamond$ is needed at time k. The second derivative of $Q_k(.|.)$ at point $\bar{\mathbf{h}}_k = \tilde{\bar{\mathbf{h}}}_{k|k-1}$ is

$$\begin{aligned}
\frac{\partial^2 Q_k(\bar{\mathbf{h}}_k|\tilde{\bar{\mathbf{h}}}_{k|k-1})}{\partial^2 \mathbf{h}_k^\diamond}\Big|_{\bar{\mathbf{h}}_k=\tilde{\bar{\mathbf{h}}}_{k|k-1}} &= \left(\frac{\partial \mathbf{h}_{k-1}^\diamond}{\partial \mathbf{h}_k^\diamond}\right)^* \left(\frac{\partial^2 Q_{k-1}(\bar{\mathbf{h}}_{k-1}|\tilde{\bar{\mathbf{h}}}_{k-1|k-2})}{\partial^2 \mathbf{h}_{k-1}^\diamond}\Big|_{\bar{\mathbf{h}}_{k-1}=\tilde{\bar{\mathbf{h}}}_{k-1|k-1}}\right)\left(\frac{\partial \mathbf{h}_{k-1}^\diamond}{\partial \mathbf{h}_k^\diamond}\right)^{\mathcal{T}} \\
&\quad -S(k)^{\mathcal{H}}\boldsymbol{\gamma}\boldsymbol{\gamma}^{\mathcal{H}}S(k) \\
&= \mathbf{F}_k^{-\mathcal{H}}\left(\frac{\partial^2 Q_{k-1}(\bar{\mathbf{h}}_{k-1}|\tilde{\bar{\mathbf{h}}}_{k-1|k-2})}{\partial^2 \mathbf{h}_{k-1}^\diamond}\Big|_{\bar{\mathbf{h}}_{k-1}=\tilde{\bar{\mathbf{h}}}_{k-1|k-1}}\right)\mathbf{F}_k^{-1} \\
&\quad -S(k)^{\mathcal{H}}\boldsymbol{\gamma}\boldsymbol{\gamma}^{\mathcal{H}}S(k) \qquad (3.24)
\end{aligned}$$

where $\mathbf{F}_k$ is

$$\mathbf{F}_k = \begin{bmatrix} \boldsymbol{F}(k) & & & \\ & \boldsymbol{F}(k-1) & & 0 \\ & 0 & \ddots & \\ & & & \boldsymbol{F}(k-N+1) \end{bmatrix} \qquad (3.25)$$

By defining $\mathbf{P}_{k|k} = \left(-\frac{\partial^2 Q_k(\bar{\mathbf{h}}_k|\tilde{\bar{\mathbf{h}}}_{k|k-1})}{\partial^2 \mathbf{h}_k^\diamond}\Big|_{\bar{\mathbf{h}}_k=\tilde{\bar{\mathbf{h}}}_{k|k-1}}\right)^{-1}$ and using $(A+BC)^{-1} = A^{-1} - A^{-1}B(I+CA^{-1}B)^{-1}CA^{-1}$, we have

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}S(k)^{\mathcal{H}}\boldsymbol{\gamma}(1+\boldsymbol{\gamma}^{\mathcal{H}}S(k)\mathbf{P}_{k|k-1}S(k)^{\mathcal{H}}\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}^{\mathcal{H}}S(k)\mathbf{P}_{k|k-1} \qquad (3.26)$$

where from (3.24) $\mathbf{P}_{k|k-1} = \mathbf{F}_k\mathbf{P}_{k-1|k-1}\mathbf{F}_k^{\mathcal{H}}$. Therefore from (3.13) the recursive estimation of $\mathbf{h}_k^\diamond$ at time k becomes

$$\tilde{\mathbf{h}}_{k|k}^\diamond = \tilde{\mathbf{h}}_{k|k-1}^\diamond + \mathbf{P}_{k|k-1}S(k)^{\mathcal{H}}\boldsymbol{\gamma}(1+\boldsymbol{\gamma}^{\mathcal{H}}S(k)\mathbf{P}_{k|k-1}S(k)^{\mathcal{H}}\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}^{\mathcal{H}}(\boldsymbol{y}_k - S(k)\tilde{\mathbf{h}}_{k|k-1}^\diamond - \boldsymbol{\mu}_z) \qquad (3.27)$$

where $\tilde{\mathbf{h}}_{k|k-1}^\diamond = \mathbf{F}_k\tilde{\mathbf{h}}_{k-1|k-1}^\diamond$. It can be seen that the recursive formula (3.27) is similar to RLS/Kalman-type algorithm. According to [71, p. 248], when $z(k)$ is colored (correlated) noise, the recursive formula in a RLS algorithm needs to invert a non-diagonal matrix and the unknown parameters may not be estimated sequentially in time. In developing the recursive formula (3.27), however, we have shown that the inversion of a matrix is not necessary, and the unknown parameters can still be estimated sequentially in time. We believe this algorithm to be novel. This result can be interpreted as using the whitening filter along with a RLS/Kalman algorithm where the elements of $\boldsymbol{\gamma}^{\mathcal{H}}$ are the coefficients of the whitening filter. When $z(k)$ is a zero mean white Gaussian noise with variance N_0, we have $\boldsymbol{\gamma} = [(N_0)^{-\frac{1}{2}}, \overbrace{0, ..., 0}^{N-1}]^{\mathcal{T}}$ and it is easy to show that

$$\begin{aligned} \tilde{\mathbf{h}}_{k|k} &= \tilde{\mathbf{h}}_{k|k-1} + P_{k|k-1}\mathbf{s}(k)^{\mathcal{H}}(1 + \mathbf{s}(k)(k)P_{k|k-1}\mathbf{s}(k)^{\mathcal{H}})^{-1}(y(k) - \mathbf{s}(k)\tilde{\mathbf{h}}_{k|k-1}) \\ P_{k|k} &= P_{k|k-1} - P_{k|k-1}\mathbf{s}(k)^{\mathcal{H}}(1 + \mathbf{s}(k)P_{k|k-1}\mathbf{s}(k)^{\mathcal{H}})^{-1}\mathbf{s}(k)P_{k|k-1} \qquad (3.28) \end{aligned}$$

where $P_{k|k} = (-N_0 \frac{\partial^2 Q_k(\bar{\mathbf{h}}_k|\tilde{\bar{\mathbf{h}}}_{k|k-1})}{\partial^2 \mathbf{h}_k}|_{\bar{\mathbf{h}}_k = \tilde{\bar{\mathbf{h}}}_{k|k-1}})^{-1}$ and $P_{k|k-1} = \boldsymbol{F}(k)P_{k-1|k-1}\boldsymbol{F}(k)^{\mathcal{H}}$. Meanwhile selecting $\boldsymbol{F}(k) = \lambda^{-\frac{1}{2}}I$, where $0 < \lambda \leq 1$ and I is an $M(L+1) \times M(L+1)$ identity matrix, and defining unknown parameters $\boldsymbol{\varphi} = \boldsymbol{h} = [h(0), ..., h(L)]^{\mathcal{T}}$, the time-variant model leads to a modified RLS algorithm with a forgetting factor λ.

$$\tilde{\boldsymbol{h}}_{|k} = \tilde{\boldsymbol{h}}_{|k-1} + P_{|k-1}\boldsymbol{s}(k)^{\mathcal{H}}(\lambda + \boldsymbol{s}(k)P_{|k-1}\boldsymbol{s}(k)^{\mathcal{H}})^{-1}(y(k) - \boldsymbol{s}(k)\tilde{\boldsymbol{h}}_{|k-1}) \qquad (3.29)$$

where $P_{|k-1} = (-N_0 \frac{\partial^2 Q_{k-1}(\boldsymbol{h}|\tilde{\boldsymbol{h}}_{|k-2})}{\partial^2 \boldsymbol{h}}|_{\boldsymbol{h} = \tilde{\boldsymbol{h}}_{|k-1}})^{-1}$ and $P_{|k}$ is obtained from

$$P_{|k} = \lambda^{-1}(P_{|k-1} - P_{|k-1}\boldsymbol{s}(k)^{\mathcal{H}}(1 + \boldsymbol{s}(k)P_{|k-1}\boldsymbol{s}(k)^{\mathcal{H}})^{-1}\boldsymbol{s}(k)P_{|k-1}) \qquad (3.30)$$

When $\boldsymbol{F}(k) = I$, the time-variant impulse response becomes time-invariant and the estimation of $\boldsymbol{h}$ leads to the well known RLS algorithm. Although the criterion of the Kalman and RLS algorithms is minimum mean square error and it is different from the ML criterion satisfied by the EM algorithm , the two different criteria lead to the same results due to the linearity of the channel and the Gaussian model for the additive noise [46]. Since $z(k)$ is Gaussian the recursive estimation formula is also exact.

b) Gaussian Random CIR

The channel in a mobile communication system is generally modeled as a linear system whose impulse response is a random vector or random process. One of the common channel models in mobile communication is the Rayleigh multipath fading channel. In this model, the CIR is considered as a complex Gaussian random vector/process whose amplitude is Rayleigh distributed. Also, without loss of generality we assume that $z(k)$

is a zero mean white Gaussian noise with autocorrelation $R_z(k) = N_0\delta(k)$; for colored noise one can follow the same procedure developed in "a". In the following we consider the estimation of the Gaussian CIR parameters using the on-line EM algorithm.

b1- Gaussian Random Vector: The received signal is obtained from

$$y(k) = \boldsymbol{s}(k)\boldsymbol{h} + z(k) \tag{3.31}$$

where $\boldsymbol{h}$ is the Gaussian random vector. The maximum *a posteriori* (MAP) estimation of the CIR at time k is

$$\hat{\boldsymbol{h}} = \arg\max_{\boldsymbol{h}}\{\log p(\boldsymbol{h}|\mathbf{y}_k)\} = E[\boldsymbol{h}|\mathbf{y}_k] \tag{3.32}$$

Therefore the unknown deterministic parameter at time k is the conditional mean of $\boldsymbol{h}$, $\boldsymbol{\mu}_{|k} = E[\boldsymbol{h}|\mathbf{y}_k]$. Hence we consider $\boldsymbol{\varphi}_k$, the unknown deterministic parameters at time k, as $\boldsymbol{\varphi}_k = \boldsymbol{\mu}_{|k}$. The complete and incomplete data are defined as $\mathcal{C}_k = \{\mathbf{y}_k, \boldsymbol{h}\}$ and $\mathcal{I}_k = \mathbf{y}_k$ respectively at time k. From (3.9) the E-step of the on-line EM algorithm becomes

$$\begin{aligned} Q_k(\boldsymbol{\mu}_{|k}|\tilde{\boldsymbol{\mu}}_{|k-1}) &= E[\log p(\mathbf{y}_k, \boldsymbol{h}|\boldsymbol{\mu}_{|k})|\mathbf{y}_k, \tilde{\boldsymbol{\mu}}_{|k-1}] \\ &= E[\log p(\mathbf{y}_{k-1}, \boldsymbol{h}|\boldsymbol{\mu}_{|k}) + \log p(y(k)|\boldsymbol{\mu}_{|k}, \mathbf{y}_{k-1}, \boldsymbol{h})|\mathbf{y}_k, \tilde{\boldsymbol{\mu}}_{|k-1}] \\ &= Q_{k-1}(\boldsymbol{\mu}_{|k}|\tilde{\boldsymbol{\mu}}_{|k-1}) - \{\log(\pi N_0) \\ &\quad + E[(y(k) - \boldsymbol{s}(k)\boldsymbol{h})^{\mathcal{H}} N_0^{-1}(y(k) - \boldsymbol{s}(k)\boldsymbol{h})|\mathbf{y}_k, \tilde{\boldsymbol{\mu}}_{|k-1}]\} \end{aligned} \tag{3.33}$$

From the definition of $Q_{k-1}(\boldsymbol{\mu}_{|k}|\tilde{\boldsymbol{\mu}}_{|k-1})$, we have

$$\frac{\partial Q_{k-1}(\boldsymbol{\mu}_{|k}|\tilde{\boldsymbol{\mu}}_{|k-1})}{\partial \boldsymbol{\mu}_{|k}^*}\Big|_{\boldsymbol{\mu}_{|k}=\tilde{\boldsymbol{\mu}}_{|k-1}} = \frac{\partial Q_{k-1}(\boldsymbol{\mu}_{|k-1}|\tilde{\boldsymbol{\mu}}_{|k-2})}{\partial \boldsymbol{\mu}_{|k-1}^*}\Big|_{\boldsymbol{\mu}_{|k-1}=\tilde{\boldsymbol{\mu}}_{|k-1}} = 0 \tag{3.34}$$

Therefore by replacing $E[\boldsymbol{h}^{\mathcal{H}} A\boldsymbol{h}|\mathbf{y}_k] = E[(\boldsymbol{h} - \boldsymbol{\mu}_{|k})^{\mathcal{H}} A(\boldsymbol{h} - \boldsymbol{\mu}_{|k})|\mathbf{y}_k] + {\boldsymbol{\mu}_{|k}}^{\mathcal{H}} A\boldsymbol{\mu}_{|k}$, the first derivative of $Q_k(\boldsymbol{\mu}_{|k}|\tilde{\boldsymbol{\mu}}_{|k-1})$ at point $\boldsymbol{\mu}_{|k} = \tilde{\boldsymbol{\mu}}_{|k-1}$ becomes

$$\frac{\partial Q_k(\boldsymbol{\mu}_{|k}|\tilde{\boldsymbol{\mu}}_{|k-1})}{\partial \boldsymbol{\mu}_{|k}^*}|_{\boldsymbol{\mu}_{|k}=\tilde{\boldsymbol{\mu}}_{|k-1}} = \boldsymbol{s}(k)^{\mathcal{H}} N_0^{-1}(y(k) - \boldsymbol{s}(k)\tilde{\boldsymbol{\mu}}_{|k-1}) \tag{3.35}$$

Meanwhile $Q_k(\boldsymbol{\mu}_{|k}|\tilde{\boldsymbol{\mu}}_{|k-1})$ and $Q_{k-1}(\boldsymbol{\mu}_{|k}|\tilde{\boldsymbol{\mu}}_{|k-1})$ can be expanded as

$$\begin{aligned} Q_k(\boldsymbol{\mu}_{|k}|\tilde{\boldsymbol{\mu}}_{|k-1}) &= -\{\log(\pi)^{L+1} + \log(\det(\Sigma_{|k})) + \\ &\quad E[(\boldsymbol{h} - \boldsymbol{\mu}_{|k})^{\mathcal{H}}\Sigma_{|k}^{-1}(\boldsymbol{h} - \boldsymbol{\mu}_{|k})|\mathbf{y}_k, \tilde{\boldsymbol{\mu}}_{|k-1}]\} + \log p(\mathbf{y}_k|\boldsymbol{\mu}_{|k}) \quad (3.36) \\ Q_{k-1}(\boldsymbol{\mu}_{|k}|\tilde{\boldsymbol{\mu}}_{|k-1}) &= -\{\log(\pi)^{L+1} + \log(\det(\Sigma_{|k-1})) + \\ &\quad E[(\boldsymbol{h} - \boldsymbol{\mu}_{|k-1})^{\mathcal{H}}\Sigma_{|k-1}^{-1}(\boldsymbol{h} - \boldsymbol{\mu}_{|k-1})|\mathbf{y}_k, \tilde{\boldsymbol{\mu}}_{|k-1}]\} \\ &\quad + \log p(\mathbf{y}_{k-1}|\boldsymbol{\mu}_{|k}) \quad (3.37) \end{aligned}$$

where $\Sigma_{|l} = \mathrm{cov}(\boldsymbol{h}|\mathbf{y}_l)$. From (3.33) and (3.37), the second derivative of $Q_k(\boldsymbol{\mu}_{|k}|\tilde{\boldsymbol{\mu}}_{|k-1})$ at point $\boldsymbol{\mu}_{|k} = \tilde{\boldsymbol{\mu}}_{|k-1}$ becomes

$$\frac{\partial^2 Q_k(\boldsymbol{\mu}_{|k}|\tilde{\boldsymbol{\mu}}_{|k-1})}{\partial^2 \boldsymbol{\mu}_{|k}}|_{\boldsymbol{\mu}_{|k}=\tilde{\boldsymbol{\mu}}_{|k-1}} = -\Sigma_{|k-1}^{-1} - \boldsymbol{s}(k)^{\mathcal{H}} N_0^{-1}\boldsymbol{s}(k) \tag{3.38}$$

As shown in (3.38), the estimation of $\Sigma_{|k-1}$ is necessary to estimate $\boldsymbol{\mu}_{|k}$. By taking the second derivative of $Q_k(\boldsymbol{\mu}_{|k}|\tilde{\boldsymbol{\mu}}_{|k-1})$ at point $\boldsymbol{\mu}_{|k} = \tilde{\boldsymbol{\mu}}_{|k-1}$ from (3.36) and comparing it with (3.38) we have

$$\Sigma_{|k} = (\Sigma_{|k-1}^{-1} + \boldsymbol{s}(k)^{\mathcal{H}} N_0^{-1}\boldsymbol{s}(k))^{-1} \tag{3.39}$$

By choosing the initial value of the covariance matrix to be an estimate, (3.39) becomes

$$\begin{aligned} \tilde{\Sigma}_{|k} &= (\tilde{\Sigma}_{|k-1}^{-1} + \boldsymbol{s}(k)^{\mathcal{H}} N_0^{-1}\boldsymbol{s}(k))^{-1} \\ &= \tilde{\Sigma}_{|k-1} - \tilde{\Sigma}_{|k-1}\boldsymbol{s}(k)^{\mathcal{H}}(N_0 + \boldsymbol{s}(k)\tilde{\Sigma}_{|k-1}\boldsymbol{s}(k)^{\mathcal{H}})^{-1}\boldsymbol{s}(k)\tilde{\Sigma}_{|k-1} \end{aligned} \tag{3.40}$$

Hence, the recursive formula for estimating $\boldsymbol{\mu}_{|k}$ from (3.13) becomes

$$\begin{aligned} \tilde{\boldsymbol{\mu}}_{|k} &= \tilde{\boldsymbol{\mu}}_{|k-1} + (\tilde{\Sigma}_{|k-1}^{-1} + \boldsymbol{s}(k)^{\mathcal{H}} N_0^{-1}\boldsymbol{s}(k))^{-1}(\boldsymbol{s}(k)^{\mathcal{H}} N_0^{-1}(y(k) - \boldsymbol{s}(k)\tilde{\boldsymbol{\mu}}_{|k-1})) \\ &= \tilde{\boldsymbol{\mu}}_{|k-1} + \tilde{\Sigma}_{|k-1}\boldsymbol{s}(k)^{\mathcal{H}}(N_0 + \boldsymbol{s}(k)\tilde{\Sigma}_{|k-1}\boldsymbol{s}(k)^{\mathcal{H}})^{-1}(y(k) - \boldsymbol{s}(k)\tilde{\boldsymbol{\mu}}_{|k-1}) \end{aligned} \tag{3.41}$$

As seen from (3.40) and (3.41), the recursive relation is the same as the stochastic RLS algorithm [68]. The results show again how estimating a Gaussian random variable based on minimizing the mean square error and MAP criterion leads to the same result.

b2-Gaussian Random Process: It is very common to model the CIR as a Gaussian random process in a mobile communication system with a relatively fast fading rate. The dynamic changing of the CIR can be represented by

$$\mathbf{h}(k) = \boldsymbol{F}(k)\mathbf{h}(k-1) + G(k)\boldsymbol{w}(k) \tag{3.42}$$

where $\boldsymbol{w}(k) = [w_0(k), ..., w_L(k)]^{\mathcal{T}}$. Also, $F(k)$ and $G(k)$ are $M(L+1) \times M(L+1)$ and $M(L+1) \times (L+1)$ matrices defined by

$$\boldsymbol{F}(k) = \begin{bmatrix} F_1(k) & F_2(k) & \dots & F_M(k) \\ I & 0 & \dots & 0 \\ & \ddots & 0 & \\ & 0 & I & 0 \end{bmatrix} \qquad G(k) = \begin{bmatrix} \mathbf{g}(k) \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{3.43}$$

where I is $(L+1)\times(L+1)$ identity matrix and 0 is $(L+1)\times(L+1)$ zero matrix in (3.43). $\boldsymbol{w}(k)$ is a zero mean complex white Gaussian random vector with a $(L+1) \times (L+1)$ autocorrelation matrix $R_{\boldsymbol{w}}(k) = I\delta(k)$ and is independent of $z(k)$. Similar to the Gaussian vector case, it can be shown that in MAP estimation of the time-variant CIR, the conditional mean of $\bar{\mathbf{h}}_k = [\mathbf{h}_k^{\mathcal{T}}, \mathbf{h}_{k-1}^{\mathcal{T}}, ..., \mathbf{h}_0^{\mathcal{T}}]^{\mathcal{T}}$ at time k is necessary and should be considered as the unknown parameters, $\boldsymbol{\theta}_k = \bar{\boldsymbol{\mu}}_{k|k}$ where $\bar{\boldsymbol{\mu}}_{k|k} = E[\bar{\mathbf{h}}_k|\mathbf{y}_k]$. Let us define $\boldsymbol{\mu}_{k|k}$ and $\boldsymbol{\Sigma}_{k|k} = \mathrm{cov}(\bar{\mathbf{h}}_k|\mathbf{y}_k) = E[(\bar{\mathbf{h}}_k - \bar{\boldsymbol{\mu}}_{k|k})(\bar{\mathbf{h}}_k - \bar{\boldsymbol{\mu}}_{k|k})^{\mathcal{H}}|\mathbf{y}_k]$ based on their elements

$$\bar{\boldsymbol{\mu}}_{k|k} = \begin{bmatrix} \boldsymbol{\mu}_{k|k} \\ \boldsymbol{\mu}_{k-1|k} \\ \vdots \\ \boldsymbol{\mu}_{0|k} \end{bmatrix} \qquad \boldsymbol{\Sigma}_{k|k} = \begin{bmatrix} \Sigma_{k,k|k} & \Sigma_{k,k-1|k} & \dots & \Sigma_{k,0|k} \\ \Sigma_{k-1,k|k} & \Sigma_{k-1,k-1|k} & \dots & \Sigma_{k-1,0|k} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{0,k|k} & \Sigma_{0,k-1|k} & \dots & \Sigma_{0,0|k} \end{bmatrix} \tag{3.44}$$

where $\boldsymbol{\mu}_{j|k} = E[\mathbf{h}_j|\mathbf{y}_k]$ and $\Sigma_{i,j|k} = E[(\mathbf{h}_i - \boldsymbol{\mu}_{i|k})(\mathbf{h}_j - \mu_{j|k})^{\mathcal{H}}|\mathbf{y}_k]$. The complete and incomplete data at time k are defined as $\mathcal{C}_k = \{\mathbf{y}_k, \bar{\mathbf{h}}_k\}$ and $\mathcal{I}_k = \mathbf{y}_k$ respectively. From (3.9) the E-step of on-line EM algorithm becomes

$$\begin{aligned} Q_k(\bar{\boldsymbol{\mu}}_k|\tilde{\bar{\boldsymbol{\mu}}}_{k|k-1}) &= E[\log p(\mathbf{y}_k, \bar{\mathbf{h}}_k|\bar{\boldsymbol{\mu}}_k)|\mathbf{y}_k, \tilde{\bar{\boldsymbol{\mu}}}_{k|k-1}] \\ &= E[\log p(y(k)|\bar{\mathbf{h}}_k, \mathbf{y}_{k-1}, \bar{\boldsymbol{\mu}}_k) + \log p(\bar{\mathbf{h}}_k|\mathbf{y}_{k-1}, \bar{\boldsymbol{\mu}}_k) \\ &\quad + \log p(\mathbf{y}_{k-1}|\bar{\boldsymbol{\mu}}_k)|\mathbf{y}_k, \tilde{\bar{\boldsymbol{\mu}}}_{k|k-1}] \\ &= -\{\log(\pi N_0) + E[(y(k) - \mathbf{s}(k)\mathbf{h}_k)^{\mathcal{H}} N_0^{-1}(y(k) - \mathbf{s}(k)\mathbf{h}_k)|\mathbf{y}_k, \tilde{\bar{\boldsymbol{\mu}}}_{k|k-1}] \\ &\quad + \log((\pi)^{M(k+1)(L+1)}\det(\boldsymbol{\Sigma}_{k|k-1})) + E[(\bar{\mathbf{h}}_k - \bar{\boldsymbol{\mu}}_{k|k-1})^{\mathcal{H}} \boldsymbol{\Sigma}_{k|k-1}^{-1} \\ &\quad \times (\bar{\mathbf{h}}_k - \bar{\boldsymbol{\mu}}_{k|k-1})|\mathbf{y}_k, \tilde{\bar{\boldsymbol{\mu}}}_{k|k-1}]\} + E[\log p(\mathbf{y}_{k-1}|\bar{\boldsymbol{\mu}}_k)|\mathbf{y}_k, \tilde{\bar{\boldsymbol{\mu}}}_{k|k-1}] \end{aligned} \tag{3.45}$$

Similar to "b1", by replacing $E[\bar{\mathbf{h}}_k^{\mathcal{H}} A \bar{\mathbf{h}}_k|\mathbf{y}_k] = E[(\bar{\mathbf{h}}_k - \bar{\boldsymbol{\mu}}_{k|k})^{\mathcal{H}} A(\bar{\mathbf{h}}_k - \bar{\boldsymbol{\mu}}_{k|k})|\mathbf{y}_k] + \bar{\boldsymbol{\mu}}_{k|k}^{\mathcal{H}} A \bar{\boldsymbol{\mu}}_{k|k}$, the first and the second derivatives of $Q(\bar{\boldsymbol{\mu}}_k|\tilde{\bar{\boldsymbol{\mu}}}_{k|k-1})$ with respect to $\bar{\boldsymbol{\mu}}_{k|k}$ at point $\bar{\boldsymbol{\mu}}_k = \tilde{\bar{\boldsymbol{\mu}}}_{k|k-1}$ are given by

$$\frac{\partial Q_k(\bar{\boldsymbol{\mu}}_k|\tilde{\bar{\boldsymbol{\mu}}}_{k|k-1})}{\partial \bar{\boldsymbol{\mu}}_{k|k}^*}\Big|_{\bar{\boldsymbol{\mu}}_k = \tilde{\bar{\boldsymbol{\mu}}}_{k|k-1}} = \bar{\mathbf{s}}(k)^{\mathcal{H}} N_0^{-1}(y(k) - \mathbf{s}(k)\tilde{\boldsymbol{\mu}}_{k|k-1}) \tag{3.46}$$

$$\frac{\partial^2 Q_k(\bar{\boldsymbol{\mu}}_k|\tilde{\bar{\boldsymbol{\mu}}}_{k|k-1})}{\partial^2 \bar{\boldsymbol{\mu}}_{k|k}}\Big|_{\bar{\boldsymbol{\mu}}_k = \tilde{\bar{\boldsymbol{\mu}}}_{k|k-1}} = -\boldsymbol{\Sigma}_{k|k-1}^{-1} - \bar{\mathbf{s}}(k)^{\mathcal{H}} N_0^{-1}\bar{\mathbf{s}}(k) \tag{3.47}$$

where $\bar{\mathbf{s}}(k) = [\mathbf{s}(k), \mathbf{0}]$ and $\mathbf{0}$ is the $kM(L+1)$ zero row vector. Also similar to the Gaussian vector case, it can be shown from (3.45) that the estimation of $\boldsymbol{\Sigma}_{k|k}$ is

$$\begin{aligned} \tilde{\boldsymbol{\Sigma}}_{k|k} &= (\tilde{\boldsymbol{\Sigma}}_{k|k-1}^{-1} + \bar{\mathbf{s}}(k)^{\mathcal{H}} N_0^{-1}\bar{\mathbf{s}}(k))^{-1} \\ &= \tilde{\boldsymbol{\Sigma}}_{k|k-1} - \tilde{\boldsymbol{\Sigma}}_{k|k-1}\bar{\mathbf{s}}(k)^{\mathcal{H}}(N_0 + \bar{\mathbf{s}}(k)\tilde{\boldsymbol{\Sigma}}_{k|k-1}\bar{\mathbf{s}}(k)^{\mathcal{H}})^{-1}\bar{\mathbf{s}}(k)\tilde{\boldsymbol{\Sigma}}_{k|k-1} \end{aligned} \tag{3.48}$$

Thus the recursive formula for estimating $\bar{\boldsymbol{\mu}}_{k|k}$ becomes

$$\begin{aligned} \tilde{\bar{\boldsymbol{\mu}}}_{k|k} &= \tilde{\bar{\boldsymbol{\mu}}}_{k|k-1} + (\tilde{\boldsymbol{\Sigma}}_{k|k-1}^{-1} + \bar{\mathbf{s}}(k)^{\mathcal{H}} N_0^{-1}\bar{\mathbf{s}}(k))^{-1}(\bar{\mathbf{s}}(k)^{\mathcal{H}} N_0^{-1}(y(k) - \mathbf{s}(k)\tilde{\boldsymbol{\mu}}_{k|k-1})) \\ &= \tilde{\bar{\boldsymbol{\mu}}}_{k|k-1} + \tilde{\boldsymbol{\Sigma}}_{k|k-1}\bar{\mathbf{s}}(k)^{\mathcal{H}}(N_0 + \bar{\mathbf{s}}(k)\tilde{\boldsymbol{\Sigma}}_{k|k-1}\bar{\mathbf{s}}(k)^{\mathcal{H}})^{-1}(y(k) - \mathbf{s}(k)\tilde{\boldsymbol{\mu}}_{k|k-1}) \end{aligned} \tag{3.49}$$

From (3.42), the relation between $\bar{\mathbf{h}}_k$ and $\bar{\mathbf{h}}_{k-1}$ is given by

$$\bar{\mathbf{h}}_k = \begin{bmatrix} \boldsymbol{F}(k)\mathbf{h}_{k-1} \\ \cdots \\ \bar{\mathbf{h}}_{k-1} \end{bmatrix} + \begin{bmatrix} G(k)\boldsymbol{w}(k) \\ \cdots \\ 0 \end{bmatrix} \tag{3.50}$$

Obtaining $\tilde{\bar{\boldsymbol{\mu}}}_{k|k-1}$ and $\tilde{\boldsymbol{\Sigma}}_{k|k-1}$ from $\tilde{\bar{\boldsymbol{\mu}}}_{k-1|k-1}$ and $\tilde{\boldsymbol{\Sigma}}_{k-1|k-1}$ are straightforward by using (3.50).

$$\tilde{\bar{\boldsymbol{\mu}}}_{k|k-1} = \begin{bmatrix} \boldsymbol{F}(k)\tilde{\boldsymbol{\mu}}_{k-1|k-1} \\ \tilde{\bar{\boldsymbol{\mu}}}_{k-1|k-1} \end{bmatrix} \tag{3.51}$$

$$\tilde{\boldsymbol{\Sigma}}_{k|k-1} = \left[\begin{array}{c|cccc} \tilde{\Sigma}_{k,k|k-1} & \vdots & \boldsymbol{F}(k)\tilde{\Sigma}_{k-1,k-1|k-1} & \cdots & \boldsymbol{F}(k)\tilde{\Sigma}_{k-1,0|k-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \tilde{\Sigma}_{k-1,k-1|k-1}\boldsymbol{F}(k)^{\mathcal{H}} & \vdots & & & \\ \vdots & \vdots & & \tilde{\boldsymbol{\Sigma}}_{k-1|k-1} & \\ \tilde{\Sigma}_{0,k-1|k-1}\boldsymbol{F}(k)^{\mathcal{H}} & \vdots & & & \end{array}\right] \tag{3.52}$$

where $\tilde{\Sigma}_{k,k|k-1} = \boldsymbol{F}(k)\tilde{\Sigma}_{k-1,k-1|k-1}\boldsymbol{F}(k)^{\mathcal{H}} + G(k)G(k)^{\mathcal{H}}$. As can be seen the covariance matrix of $\bar{\mathbf{h}}_k$ at time k can be calculated directly based on selecting $\tilde{\Sigma}_{0,0|-1}$, the initial value of the covariance matrix of $\mathbf{h}_0$ at time zero without knowing the received signal $\mathbf{y}_k$. Meanwhile, since the first $L+1$ elements of the vector $\bar{\mathbf{s}}(k)$ are nonzero, only the first $L+1$ columns of $\tilde{\boldsymbol{\Sigma}}_{k|k-1}$ need to be calculated for obtaining $\tilde{\bar{\boldsymbol{\mu}}}_{k|k}$.

The recursive relation (3.49) estimates the entire unknown sequence of parameters in each recursion by processing over a new sample of the received signal. If only $\tilde{\boldsymbol{\mu}}_{k|k}$, part of $\tilde{\bar{\boldsymbol{\mu}}}_{k|k}$, is selected from (3.49), the recursive formula becomes Kalman filtering. In general, by defining the unknown process $\boldsymbol{\varphi}_k = \boldsymbol{\mu}_{j|k}$ and the complete data $\mathcal{C}_k = \{\mathbf{y}_k, \mathbf{h}_j\}$ at time k and following the same procedure done in this subsection, it can be shown that the on-line EM algorithm leads to predicting, filtering and smoothing algorithms for $j > k$, $j = k$ and $j < k$ respectively. The recursive formula (3.49) is the estimation procedure of smoothing for the entire unknown parameters up to time k based on the available information up to this time. By selecting $\mathcal{C}_k = \{\mathbf{y}_k, \bar{\mathbf{h}}_{k+j}\}$ and $\boldsymbol{\theta}_k = \bar{\boldsymbol{\mu}}_{k+j|k}$ where $j > 0$, it is straightforward to modify the estimation procedure to include the prediction problem as well.

In a smoothing algorithm it is common to assume that the estimator knows the entire received sequence and two Kalman algorithms, forward and backward, are applied in the estimating procedure . However, the recursive estimating method developed here based on the on-line EM algorithm is more general. Not only are the entire unknown processes re-estimated at each time, but also the boundary of the available data is changing with time. Once again estimation based on MMSE and MAP criteria for Gaussian process merged to the same procedure for the linear system case. Meanwhile, for colored Gaussian noise, one can be follow the same procedure done in subsection "a" and show that the relations (3.48) and (3.49) become similar to (3.26) and (3.27).

C) Hybrid of Unknown Parameters and Gaussian Process

Sometimes the unknown set of parameters is a combination of constant and sequential parameters. In this situation the estimation procedure needs both the methods developed in the subsections "a" and "b". In general, the complete data may also be different for the two types of unknown parameters and the estimating algorithm can contain two combined on-line EM algorithms. Although there is interaction between the unknown parameters in different times, estimating parameters at time k depends on the estimation of the other parameters at time $k-1$. Hence, not only the unknown parameters, but also the complete data may be partitioned to run the on-line EM algorithms. Let us focus on more details of this situation in channel estimation.

The estimating procedure in the Gaussian random process CIR (3.42) is based on knowing $\boldsymbol{F}(k)$ and $G(k)G(k)^{\mathcal{H}}$ matrices. These parameters are generally unknown and should also be estimated in the receiver. Assuming time-invariant $\boldsymbol{F}$ and G matrices, there are two unknown parameter sets at time k, $\boldsymbol{\theta}_{1,k} = \bar{\boldsymbol{\mu}}_{k|k}$ and $\boldsymbol{\varphi}_2$, the elements of $\boldsymbol{F}$ and $GG^{\mathcal{H}}$. The complete data for estimating $\boldsymbol{\theta}_{1,k}$ and $\boldsymbol{\varphi}_2$ is not the same. While the complete data for estimating $\boldsymbol{\theta}_{1,k}$ is $\mathcal{C}_{1,k} = \{\mathbf{y}_k, \bar{\mathbf{h}}_k\}$, the complete data for estimating $\boldsymbol{\varphi}_2$ at time k is $\mathcal{C}_{2,k} = \bar{\mathbf{h}}_k$; however the incomplete data for both is $\mathcal{I}_{1,k} = \mathcal{I}_{2,k} = \mathbf{y}_k$. Since less-informative complete data improves asymptotic convergence rate [70], all the unknown parameters $\{\boldsymbol{\theta}_{1,k}, \boldsymbol{\varphi}_2\}$ are estimated with two separate on-line EM algorithms based on the following theorem.

Theorem 1: Let the unknown parameter set $\mathcal{U}$ be divided into $\mathbf{i}$ separate sets, $\mathcal{U} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_\mathbf{i}\}$. $L(\hat{\mathcal{U}}^{(l+1)}) \geq L(\hat{\mathcal{U}}^{(l)})$ where $L(\mathcal{U}) = \log p(\mathcal{I}|\mathcal{U})$ (3.5) and the estimation of $\mathcal{U}$ at $(l+1)$th iteration, $\hat{\mathcal{U}}^{(l+1)} = \{\hat{\boldsymbol{\theta}}_1^{(l+1)}, \hat{\boldsymbol{\theta}}_2^{(l+1)}, ..., \hat{\boldsymbol{\theta}}_\mathbf{i}^{(l+1)}\}$, is estimated from the following steps using the EM algorithm,

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_1^{(l+1)} &= \max_{\boldsymbol{\theta}_1}\{Q(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2^{(l)}, ..., \hat{\boldsymbol{\theta}}_\mathbf{i}^{(l)}|\hat{\mathcal{U}}^{(l)})\} \\
\hat{\boldsymbol{\theta}}_2^{(l+1)} &= \max_{\boldsymbol{\theta}_2}\{Q(\hat{\boldsymbol{\theta}}_1^{(l+1)}, \boldsymbol{\theta}_2, \hat{\boldsymbol{\theta}}_3^{(l)}, ..., \hat{\boldsymbol{\theta}}_\mathbf{i}^{(l)}|\hat{\mathcal{U}}^{(l)})\} \\
&\vdots \\
\hat{\boldsymbol{\theta}}_\mathbf{i}^{(l+1)} &= \max_{\boldsymbol{\theta}_\mathbf{i}}\{Q(\hat{\boldsymbol{\theta}}_1^{(l+1)}, \hat{\boldsymbol{\theta}}_2^{(l+1)}, ..., \hat{\boldsymbol{\theta}}_{\mathbf{i}-1}^{(l+1)}, \boldsymbol{\theta}_\mathbf{i}|\hat{\mathcal{U}}^{(l)})\}
\end{aligned} \tag{3.53}$$

Proof: See Appendix C.[3] By defining $\tilde{\boldsymbol{\theta}}_{1,k|k-1} = \hat{\boldsymbol{\theta}}_1^{(l)}$, $\tilde{\boldsymbol{\theta}}_{1,k|k} = \hat{\boldsymbol{\theta}}^{(l+1)}$, $\tilde{\boldsymbol{\varphi}}_{2|k-1} = \hat{\boldsymbol{\theta}}_2^{(l)}$ and $\tilde{\boldsymbol{\varphi}}_{2|k} = \hat{\boldsymbol{\theta}}_2^{(l+1)}$, Theorem 1 can be applied to the unknown parameter set $\{\boldsymbol{\theta}_{1,k}, \boldsymbol{\varphi}_2\}$ in order to increase the likelihood by estimating $\boldsymbol{\theta}_{1,k}$ and $\boldsymbol{\varphi}_2$ with two separate on-line EM algorithms.

The first recursive formula for estimating $\boldsymbol{\theta}_{1,k}$ is the same as the Gaussian random process case where $Q_{1,k}(\boldsymbol{\theta}_{1,k}|\tilde{\boldsymbol{\theta}}_{1,k|k-1}, \tilde{\boldsymbol{\varphi}}_{2|k-1})$ is defined as

$$Q_{1,k}(\boldsymbol{\theta}_{1,k}|\tilde{\boldsymbol{\theta}}_{1,k|k-1}, \tilde{\boldsymbol{\varphi}}_{2|k-1}) = E[\log p(\mathbf{y}_k, \bar{\mathbf{h}}_k|\boldsymbol{\theta}_{1,k}, \tilde{\boldsymbol{\varphi}}_{2|k-1})|\mathbf{y}_k, \tilde{\boldsymbol{\theta}}_{1,k|k-1}, \tilde{\boldsymbol{\varphi}}_{2|k-1}] \tag{3.54}$$

and its derivatives are taken with respect to $\boldsymbol{\theta}_{1,k}^*$ at point $\boldsymbol{\theta}_{1,k} = \tilde{\boldsymbol{\theta}}_{1,k|k-1}$. The estimation procedure of $\boldsymbol{\theta}_{1,k}$ is the same as the procedure described in Gaussian random process "b2" using the estimates of $\boldsymbol{F}$ and $GG^{\mathcal{H}}$at time $k-1$ instead of their real values.

It is more convenient to estimate $\boldsymbol{\varphi}_2$ based on an ARMA model of $h_l(k) = h(l, k)$ instead of the state space model. It is clear from (3.43) that only F_i for $i = 1, ..., M$ and

[3]Iteration between some maximization steps is similar to Gauss-Seidel method [72] and is also called the cyclic coordinate ascent method [73]. Meng and Rubin [74] proposed a similar reduction model to achieve simple conditional maximization and called it expectation-conditional maximization.

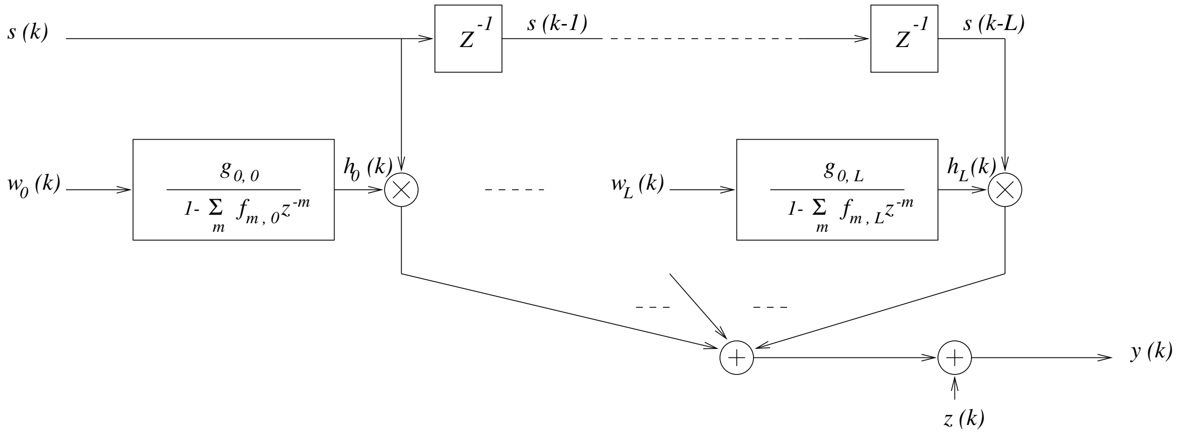

Figure 3.1: A discrete channel model for multipath fading.

the diagonal elements of $\mathbf{g}\mathbf{g}^{\mathcal{H}}$ need be estimated ($R_{\boldsymbol{w}}(k) = I\delta(k)$). When diag($\mathbf{g}$)= $\{\mathbf{g}^0, \mathbf{g}^1, ..., \mathbf{g}^L\}$, the ARMA model of $h_l(k)$ is

$$h_l(k) = \mathbf{f}^l \mathbf{h}_{k-1} + \mathbf{g}^l w_l(k) \qquad\qquad 0 \le l \le L \tag{3.55}$$

where $\mathbf{f}^l$ is the lth row of $\boldsymbol{F}$ matrix (Fig.3.1 shows the channel model when there is no correlation between the coefficients of the channel. For this figure the non zero elements of $\mathbf{f}^l$ are $\{f_{1,l}, ..., f_{M,l}\}$ and $\mathbf{g}^l = g_{0,l}$). By defining $\boldsymbol{\varphi}_2 = [\mathbf{f}^L, ..., \mathbf{f}^0, R_{g^L}, ..., R_{g^0}]$ where $R_{g^l} = \mathbf{g}^l \mathbf{g}^{l^{\mathcal{H}}}$ for $l = 0, ..., L$, the E-step of estimating $\boldsymbol{\varphi}_2$ at time k is given by

$$\begin{aligned}
Q_{2,k}(\boldsymbol{\varphi}_2|\tilde{\boldsymbol{\theta}}_{1,k|k-1}, \tilde{\boldsymbol{\varphi}}_{2|k-1}) &= E[\log p(\mathbf{y}_k, \bar{\mathbf{h}}_k|\tilde{\boldsymbol{\theta}}_{1,k|k}, \boldsymbol{\varphi}_2)|\mathbf{y}_k, \tilde{\boldsymbol{\theta}}_{1,k|k-1}, \tilde{\boldsymbol{\varphi}}_{2|k-1}] \\
&= E[\log p(\bar{\mathbf{h}}_k|\tilde{\boldsymbol{\theta}}_{1,k|k}, \boldsymbol{\varphi}_2)|\mathbf{y}_k, \tilde{\boldsymbol{\theta}}_{1,k|k-1}, \tilde{\boldsymbol{\varphi}}_{2|k-1}] + \log p(\mathbf{y}_k|\bar{\mathbf{h}}_k) \\
&= \sum_{l=0}^{L}\{\sum_{j=1}^{k} -\{\log(\pi R_{g^l}) \\
&+ E[(h_l(j) - \mathbf{f}^l \mathbf{h}_{j-1})^{\mathcal{H}} R_{g^l}^{-1}(h_l(j) - \mathbf{f}^l \mathbf{h}_{j-1})|\mathbf{y}_k, \tilde{\boldsymbol{\theta}}_{1,k|k-1}, \tilde{\boldsymbol{\varphi}}_{2|k-1}]\} \\
&+ E[\log p(h_l(0)|\tilde{\boldsymbol{\theta}}_{1,k|k}, \boldsymbol{\varphi}_2)|\mathbf{y}_k, \tilde{\boldsymbol{\theta}}_{1,k|k-1}, \tilde{\boldsymbol{\varphi}}_{2|k-1}]\} \\
&+ \log p(\mathbf{y}_k|\bar{\mathbf{h}}_k)
\end{aligned} \tag{3.56}$$

Since $\log p(\mathbf{y}_k|\bar{\mathbf{h}}_k)$ is not a function of $\boldsymbol{\varphi}_2$, thus as was mentioned earlier and also shown by (3.56), the complete data for estimating $\boldsymbol{\varphi}_2$ is $\mathcal{C}_{2,k} = \bar{\mathbf{h}}_k$. The first derivative of $Q_{2,k}(.|.)$ with respect to $\mathbf{f}^{l^*}$ at point $\boldsymbol{\varphi}_2 = \tilde{\boldsymbol{\varphi}}_{2|k-1}$ is

$$\left.\frac{\partial Q_{2,k}(\boldsymbol{\varphi}_2|\tilde{\boldsymbol{\theta}}_{1,k|k-1},\tilde{\boldsymbol{\varphi}}_{2|k-1})}{\partial \mathbf{f}^{l^*}}\right|_{\boldsymbol{\varphi}_2 = \tilde{\boldsymbol{\varphi}}_{2|k-1}} = \frac{1}{\tilde{R}_{g^l_{|k-1}}}\sum_{j=1}^{k}\{\tilde{\boldsymbol{\sigma}}^l_{j-1,j|k} - \tilde{\mathbf{f}}^l_{|k-1}\tilde{\Sigma}_{j-1,j-1|k} + \tilde{\boldsymbol{\mu}}^{\mathcal{H}}_{j-1|k}(\tilde{\mu}^l_{j|k} - \tilde{\mathbf{f}}^l_{|k-1}\tilde{\boldsymbol{\mu}}_{j-1|k})\} \qquad 0 \le l \le L \tag{3.57}$$

where $\boldsymbol{\sigma}^l_{j-1,j|k} = E[(\mathbf{h}_{j-1} - \boldsymbol{\mu}_{j-1|k})^{\mathcal{H}}(h_l(j) - \mu^l_{j-1|k})|\mathbf{y}_k]$. The second derivative of $Q_{2,k}(.|.)$ with respect to $\mathbf{f}^l$ becomes

$$\left.\frac{\partial^2 Q_{2,k}(\boldsymbol{\varphi}_2|\tilde{\boldsymbol{\theta}}_{1,k|k-1},\tilde{\boldsymbol{\varphi}}_{2|k-1})}{\partial^2 \mathbf{f}^{l}}\right|_{\boldsymbol{\varphi}_2 = \tilde{\boldsymbol{\varphi}}_{2|k-1}} = -\frac{1}{\tilde{R}_{g^l_{|k-1}}}\sum_{j=1}^{k}\tilde{\Sigma}_{j-1,j-1|k} + \tilde{\boldsymbol{\mu}}_{j-1|k}\tilde{\boldsymbol{\mu}}^{\mathcal{H}}_{j-1|k} \qquad 0 \le l \le L \tag{3.58}$$

From (3.57) and (3.58) and using (3.13), the recursion formula for estimating the row vector $\mathbf{f}^l$ becomes

$$\begin{aligned}\tilde{\mathbf{f}}^l_{|k} &= \tilde{\mathbf{f}}^l_{|k-1} + (\sum_{j=1}^{k}\tilde{\boldsymbol{\sigma}}^l_{j-1,j|k} - \tilde{\mathbf{f}}^l_{|k-1}\tilde{\Sigma}_{j-1,j-1|k-1} + \tilde{\boldsymbol{\mu}}^{\mathcal{H}}_{j-1|k}(\tilde{\mu}^l_{j|k} - \tilde{\mathbf{f}}^l_{|k-1}\tilde{\boldsymbol{\mu}}_{j-1|k}))\\ &\times \ (\sum_{j=1}^{k}\tilde{\Sigma}_{j-1,j-1|k} + \boldsymbol{\mu}_{j-1|k}\boldsymbol{\mu}^{\mathcal{H}}_{j-1|k})^{-1} \qquad 0 \le l \le L\end{aligned} \tag{3.59}$$

As (3.59) shows, the recursive estimation of $\mathbf{f}^l$ at each recursion needs to calculate the inverse of a matrix. Using a filtering approach in (3.59) instead of smoothing and assuming $\tilde{\Sigma}_{j-1,j-1|j} \ll \tilde{\boldsymbol{\mu}}_{j-1|j}\tilde{\boldsymbol{\mu}}^{\mathcal{H}}_{j-1|j}$ and after doing some manipulations we have

$$\tilde{\mathbf{f}}^l_{|k} \simeq \tilde{\mathbf{f}}^l_{|k-1} + \tilde{\boldsymbol{\mu}}^{\mathcal{H}}_{k-1|k}P^{\diamond}_{k-1|k}(1 + \tilde{\boldsymbol{\mu}}^{\mathcal{H}}_{k-1|k}P^{\diamond}_{k-1|k}\tilde{\boldsymbol{\mu}}_{k-1|k})^{-1}(\tilde{\mu}^l_{k|k} - \tilde{\mathbf{f}}^l_{|k-1}\tilde{\boldsymbol{\mu}}_{k-1|k}) \quad 0 \le l \le L \tag{3.60}$$

where $P^{\diamond}_{k-1|k} = (\sum_{j=1}^{k-1}\tilde{\boldsymbol{\mu}}_{j-1|j}\tilde{\boldsymbol{\mu}}^{\mathcal{H}}_{j-1|j})^{-1}$ and $P^{\diamond}_{k|k+1}$ is given by

$$P^{\diamond}_{k|k+1} = P^{\diamond}_{k-1|k} - P^{\diamond}_{k-1|k}\tilde{\boldsymbol{\mu}}_{k-1|k}(1 + \tilde{\boldsymbol{\mu}}^{\mathcal{H}}_{k-1|k}P^{\diamond}_{k-1|k}\tilde{\boldsymbol{\mu}}_{k-1|k})^{-1}\tilde{\boldsymbol{\mu}}^{\mathcal{H}}_{k-1|k}P^{\diamond}_{k-1|k} \tag{3.61}$$

Meanwhile, it can be seen from (3.59) that the estimation of $\mathbf{f}^l$ is independent of the estimation of R_{g^l}. The first and the second derivatives of $Q_{2,k}(.|.)$ with respect to $R_{g^l}^{-1}$ become

$$\frac{\partial Q_{2,k}(\boldsymbol{\varphi}_2|\tilde{\boldsymbol{\theta}}_{1,k|k-1}, \tilde{\boldsymbol{\varphi}}_{2|k-1})}{\partial R_{g^l|k}^{-1*}}\Bigg|_{\boldsymbol{\varphi}_2 = \tilde{\boldsymbol{\varphi}}_{2|k-1}} = [kR_{g^l|k-1} - \sum_{j=1}^{k} E(h_l(j) - \mathbf{f}^l\mathbf{h}_{j-1})(h_l(j) - \mathbf{f}^l\mathbf{h}_{j-1})^{\mathcal{H}}|\mathbf{y}_k]\Bigg|_{\boldsymbol{\varphi}_2 = \tilde{\boldsymbol{\varphi}}_{2|k-1}} \tag{3.62}$$

$$\frac{\partial^2 Q_{2,k}(\boldsymbol{\varphi}_2|\tilde{\boldsymbol{\theta}}_{1,k|k-1}, \tilde{\boldsymbol{\varphi}}_{2|k-1})}{\partial^2 R_{g^l|k}^{-1}}\Bigg|_{\boldsymbol{\varphi}_2 = \tilde{\boldsymbol{\varphi}}_{2|k-1}} = -k\tilde{R}_{g^l|k-1}^2 \tag{3.63}$$

By using a filtering approach, assuming $\tilde{\Sigma}_{j-1|j} \ll \tilde{\boldsymbol{\mu}}_{j-1|j}\tilde{\boldsymbol{\mu}}_{j-1|j}^{\mathcal{H}}$ and $\tilde{\Sigma}_{j|j} \ll \tilde{\boldsymbol{\mu}}_{j|j}\tilde{\boldsymbol{\mu}}_{j|j}^{\mathcal{H}}$ for $0 \leq j \leq k$ and doing some manipulations, it can be shown that

$$\tilde{R}_{g^l|k} \simeq \tilde{R}_{g^l|k-1} + \frac{1}{k}(\tilde{R}_{g^l|k-1} - |\tilde{\mu}_{k|k}^l - \tilde{\mathbf{f}}_{|k-1}^l\tilde{\boldsymbol{\mu}}_{k-1|k}|^2) \tag{3.64}$$

All conditional mean values in (3.60) and (3.64) can be obtained from (3.49). Combining Kalman (3.48 and 3.49) and RLS (3.60 and 3.61) algorithms, which recursively estimates CIR and its ARMA model parameters, $\boldsymbol{F}$ and $GG^{\mathcal{H}}$, is new. It can be shown that $\boldsymbol{F} = R_{\mathbf{h}}(1)R_{\mathbf{h}}^{-1}(0)$ and $GG^{\mathcal{H}} = R_{\mathbf{h}}(0) - \boldsymbol{F}R_{\mathbf{h}}^{\mathcal{H}}(1)$, where the autocorrelation function of $\mathbf{h}(k)$ is $R_{\mathbf{h}}(l) = E[\mathbf{h}(k)\mathbf{h}^{\mathcal{H}}(k-l)]$ [11]. Although the estimated value of $\mathbf{h}(k)$, $\tilde{\boldsymbol{\mu}}_{k|k-1}$, is related to a specific realization of $y(k)$, the estimation of $\boldsymbol{F}$ and $GG^{\mathcal{H}}$ should be the same for all realizations of $y(k)$. Because of this reason, the coefficients of the predictor filters in the MLSD receiver developed in Chapter 2 based on the autocorrelation function of $\mathbf{h}(k)$, $R_{\mathbf{h}}(l)$, are time-invariant . However, the coefficients of the FIR filter in the MLSD receiver based on known CIR are time-variant.

3.5 Summary

Recursive estimation of deterministic parameters embedded in white and colored Gaussian noise modeled by an autoregressive process was investigated in this chapter based on the ML criterion for different models of linear transmission channel using the on-line EM algorithm. By using Titterington's approach, which was modified for dynamic parameters, to estimate time-invariant/variant parameters in the on-line EM algorithm, we get different types of RLS, Kalman and combined RLS/Kalman-type algorithms. Some of these algorithms such as the RLS/Kalman-type algorithm , which estimates deterministic unknown CIR embedded in a colored Gaussian noise, and the combined RLS/Kalman algorithm, which estimates Gaussian CIR with its ARMA model parameters, are new. We will show later in Chapter 4 how these algorithms can be used in joint channel estimation and data detection. These algorithms were derived directly based on the EM approach which emerged as a powerful tool for unification of different types of adaptive algorithms. Applications of such estimation algorithms are not limited to channel estimation. These algorithms can play a natural and important role in adaptive estimation and detection schemes in other areas as well.

Chapter 4

Adaptive MLSDE Using the EM Algorithm

The detection of a signal transmitted through a communication channel having memory and additive Gaussian noise has been widely studied for different channel models. Equalization techniques have been used in communication systems to combat the intersymbol interference (ISI) induced by dispersive channels. When the transmitted data sequences are equiprobable, the maximum likelihood sequence detection (MLSD) minimizes sequence error probability and can, hence, be considered as an optimal equalization method. MLSD, implemented using the Viterbi algorithm for known finite channel impulse response (CIR), is well known [26]. The MLSD algorithm has also been studied for a mobile communication channel which disperses the transmitted signal in both time and frequency domains and whose impulse response is considered as a Gaussian random process [10–12, 37, 49].

Due to unknown CIR or unknown statistical parameters of the CIR, joint data detection and channel estimation methods were proposed by combining Viterbi algorithm for data detection with adaptive methods, such as least mean square (LMS), recursive least squares (RLS) and Kalman filtering, for estimating the CIR [7, 11, 34, 40]. However, the inherent decision delay in such procedures causes poor channel tracking in a time-variant environment. The idea of per-survivor processing (PSP) was proposed to combat the decision delay problem, where each survivor path of the trellis diagram in the MLSD structure has its own CIR estimation [8, 9, 36]. Although PSP is a practical way to achieve better performance in a time-variant channel, the nature and degree of optimality of such PSP-based channel estimation procedures, the influence of such estimates on the optimality of the MLSD criterion and the coupling among estimation, detection and channel models are not clear.

In this Chapter, the maximum likelihood detection and estimation (MLSDE) algorithm is considered in a general framework based on ML detection/estimation theory. To detect the transmitted signal, the receiver usually needs to know some other parameters. We show that the MLSDE criterion for detecting the data sequence and estimating the unknown parameters can be achieved by the EM algorithm which is an iterative method. The EM algorithm increases the likelihood of the detected/estimated parameters in each iteration with expectation and maximization steps till it achieves the global or a local maximum [38] [14].

Generalized MLSDE (GMLSDE) is presented as an EM-based algorithm, which alternates between detection and estimation and still satisfies the MLSDE criterion. GMLSDE is implemented based on the on-line EM algorithm for real time detection/estimation, where in each recursion the algorithm increases the likelihood function. It is shown that the concept of PSP emerges naturally from the EM aspect of the GMLSDE algorithm as an integral part of a likelihood-increasing procedure when the previously detected/estimated parameters are used as given conditions for the next expectation step. Some adaptive MLSDE receivers are derived based on the GMLSDE framework in a unified way for different levels of channel knowledge which are available at the receiver. The recursive estimation proposed by Titterington [43] is employed for the estimation of time-variant/invariant unknown deterministic parameters. Each adaptive receiver uses those steps of detection and estimation of GMLSDE generated by the selected channel model. Although two new adaptive MLSDE receivers along with some previously known ones are derived as examples, the power of GMLSDE is not limited to these particular algorithms and one can use the unified framework of GMLSDE to develop new algorithms based on different channel models and levels of knowledge available at the receiver.

Even though the concept behind the EM algorithm was known in early statistical literature [75, 76], it was the later, seminal paper by Dempster *et al* [38] that spurred much research on many applications of the EM algorithm including communication-related ones [14]. Kaleh *et al* [65] derived the iterative method for joint channel parameter estimation and symbol detection using an EM-based forward-backward algorithm. Georghiades and Han [67] used the EM algorithm to study sequence estimation in random phase and fading channels. We may point to [63,77,78] for some other applications of the EM algorithm in communication systems.

This Chapter is organized as follows. GMLSDE is developed via the EM algorithm and the relation to PSP is explained in Section 4.1. Different adaptive MLSDE algorithms are derived in Section 4.2 associated with different levels of channel knowledge available at the receiver. Section 4.3 contains computer simulations, results and comparisons for a DQPSK modulation scheme in flat and selective fading channels with different fading rates. The results of simulations with time-variant fading rate for a flat fading channel with different levels of channel knowledge are also presented in Section 4.3. Section 4.4 presents a summary of the Chapter.

4.1 Generalized MLSDE Algorithm

The main goal in digital communication systems is to detect the sequence of the transmitted symbols $\mathbf{a} = \{a_0, ..., a_{I-1}\}$ by observing the received signal $\mathbf{y} = [y(K-1), ..., y(0)]^{\mathcal{T}}$. The optimal receiver maximizes the joint probability density function (pdf) of $\mathbf{a}$ and $\mathbf{y}$ for detecting $\mathbf{a}$. When all the sequences $\mathbf{a}$ are equiprobable, the optimal sequence receiver accomplishes MLSD with the criterion given by

$$C_{MLSD} \doteq \max_{\mathbf{a}}\{p(\mathbf{y}|\mathbf{a})\} \equiv \max_{\mathbf{a}}\{\log p(\mathbf{y}|\mathbf{a})\} \tag{4.1}$$

where $\mathbf{a}$ is selected from $\mathcal{A}$, a set of all its possibilities, $\mathbf{a} \in \mathcal{A}$. The detected sequence $\hat{\mathbf{a}}$ is

$$\hat{\mathbf{a}} = \arg\max_{\mathbf{a}}\{\log p(\mathbf{y}|\mathbf{a})\} \tag{4.2}$$

The received signal is a function of the transmitted symbols $\mathbf{a}$, media parameters such as channel parameters $\hbar_c$ and additive noise $\mathbf{z} = [z(K-1), ..., z(0)]^{\mathcal{T}}$.

$$\mathbf{y} = f_{ch}(\mathbf{a}, \hbar_c) + \mathbf{z} \tag{4.3}$$

When the channel is modeled as a linear system, $f_{ch}(.)$ is a linear function and its arguments are $\mathbf{a}$ and CIR. To detect $\mathbf{a}$ by observing $\mathbf{y}$, the receiver should know the function $f_{ch}(.)$, the pdf of $\mathbf{z}$ and the CIR (if it is deterministic) or the pdf of CIR (if it is a random variable/process).

The structure of the MLSD receiver will be different for different channel models. The linear function $f_{ch}(.)$ is usually modeled as having finite memory (e.g. FIR system for ISI or multipath fading channels). Also it is very common to model additive noise $\mathbf{z}$ as a stationary, white, complex, zero mean, circularly symmetric, Gaussian random process with autocorrelation function $R_z(k) = N_0\delta(k)$. The CIR or its parameters (usually unknown to the receiver) should be estimated during the detection procedure.

The MLSD criterion (4.1) is suitable when the symbols, $\mathbf{a}$, are the only parameters which are unknown and the received signal, $\mathbf{y}$, provides complete information necessary for such a detection procedure. However, in general, other parameters which can be modeled as a set of unknown deterministic parameters, random variables/processes or both are also needed to complete the detection procedure. In detection theory, such problems are termed as composite hypothesis testing or detection with unwanted parameters [79].

If we define $\boldsymbol{\theta}$ as the needed unknown deterministic parameters which should be estimated, the criterion of joint maximum likelihood sequence detection and estimation (MLSDE) becomes

$$C_{MLSDE} \doteq \max_{\mathbf{a},\boldsymbol{\theta}}\{\log p(\mathbf{y}|\mathbf{a},\boldsymbol{\theta})\} \tag{4.4}$$

Since sometimes $\mathbf{y}$ does not provide the complete information necessary to obtain the ML estimates of the parameters $\mathcal{U} = \{\mathbf{a}, \boldsymbol{\theta}\}$ from (4.4) directly, we present the solution to (4.4) using the EM algorithm [38]. By considering $\mathcal{I} = \mathbf{y}$ as incomplete data and $\mathcal{D}$ as the random variables or processes needed to complete $\mathcal{I}$ for detecting/estimating $\mathcal{U}$, the log-likelihood function is given by

$$L(\mathcal{U}) = \log p(\mathcal{I}|\mathcal{U}) = \log p(\mathcal{C}|\mathcal{U}) - \log p(\mathcal{D}|\mathcal{U}, \mathcal{I}) \tag{4.5}$$

where $\mathcal{C} = \{\mathcal{I}, \mathcal{D}\}$ is the complete data. On taking the conditional expectation of both sides of (4.5) with respect to $\mathcal{D}$ given $\mathcal{I}$ and a parameter set $\hat{\mathcal{U}}^{(l)}$ (the estimation of $\mathcal{U}$ at lth iteration or initial estimate of $\mathcal{U}$) we have

$$L(\mathcal{U}) = E[\log p(\mathcal{C}|\mathcal{U})|\hat{\mathcal{U}}^{(l)}, \mathcal{I}] - E[\log p(\mathcal{D}|\mathcal{U}, \mathcal{I})|\hat{\mathcal{U}}^{(l)}, \mathcal{I}] \tag{4.6}$$

Following [38], one can show that $L(\hat{\mathcal{U}}^{(l+1)}) \geq L(\hat{\mathcal{U}}^{(l)})$ where

$$Q(\mathcal{U}|\hat{\mathcal{U}}^{(l)}) = E[\log p(\mathcal{C}|\mathcal{U})|\hat{\mathcal{U}}^{(l)}, \mathcal{I}] \tag{4.7}$$

and

$$\hat{\mathcal{U}}^{(l+1)} = \arg\max_{\mathcal{U}}\{Q(\mathcal{U}|\hat{\mathcal{U}}^{(l)})\} \tag{4.8}$$

The above iteration,(4.7) and (4.8), is repeated such that the estimation of $\mathcal{U}$ at each iteration is used as the given condition for the next iteration until at $\acute{l}$th iteration $\hat{\mathcal{U}} = \hat{\mathcal{U}}^{(\acute{l})} = \hat{\mathcal{U}}^{(\acute{l}-1)}$ or the estimates get arbitrarily close to each other. Hence, if $L(\mathcal{U})$ has only one maximum point, we have

$$\hat{\mathcal{U}} = \arg\max_{\mathcal{U}}\{L(\mathcal{U})\} \tag{4.9}$$

Otherwise, $\hat{\mathcal{U}}$ is a local maximum point. Thus, the MLSDE criterion can be achieved iteratively by following the EM algorithm.

The EM algorithm has two steps; expectation (4.7) and maximization (4.8). The first step, (4.7), is to take the expectation of the log-likelihood function of the complete data given the current detected/estimated parameters, $\hat{\mathcal{U}}^{(l)}$ and the incomplete (observed) data, $\mathcal{I}$. The second step, (4.8), provides a new estimate of the parameters, $\hat{\mathcal{U}}^{(l+1)}$, by maximizing the expectation of the log-likelihood function (computed in the first step) over the unknown parameters $\mathcal{U}$. The EM algorithm repeats the expectation and maximization steps iteratively in order to increase the likelihood of the parameters [38] [14]. Therefore instead of calculating the ML detection/estimation of $\mathcal{U} = \{\mathbf{a}, \boldsymbol{\theta}\}$ directly from (4.4) in a closed form or in one iteration which is very complex and impractical especially for a high dimensional problem, the EM algorithm uses the iterative method which increases the likelihood of detected/estimated parameters in each iteration until it achieves the MLSDE criterion (4.4).

When the unknown parameter set $\mathcal{U}$ contains different types of unknown variables/vectors (i.e. in our problem $\mathbf{a}$ and $\boldsymbol{\theta}$ are discrete and continuous respectively), the maximization step of the EM algorithm is a challenging job and often intractable. In order that the maximization step becomes tractable we partition the set of unknown parameters into disjoint sets and find the maximum of each partitioned set separately. Therefore in this method, each iteration contains more than one expectation and maximization step. By following Theorem 1 presented in Chapter 3 (3.53), generalized MLSDE (GMLSDE) procedure increases the likelihood at each iteration to satisfy the ML criterion[1] by dividing $\mathcal{U}$ into $\mathbf{a}$ and $\boldsymbol{\theta}$ and estimation of $\boldsymbol{\theta}$ (continuous unknown parameter set) and detection of $\mathbf{a}$ (discrete unknown parameter set) are alternated. The estimation and detection steps at the $(l+1)$th iteration for the GMLSDE are[2]:

Estimation part:

1-E step:

$$\begin{aligned} Q_1(\boldsymbol{\theta}|\hat{\mathbf{a}}^{(l)}, \hat{\boldsymbol{\theta}}^{(l)}) &= Q(\hat{\mathbf{a}}^{(l)}, \boldsymbol{\theta}|\hat{\mathbf{a}}^{(l)}, \hat{\boldsymbol{\theta}}^{(l)}) \\ &= E[\log p(\mathcal{C}|\hat{\mathbf{a}}^{(l)}, \boldsymbol{\theta})|\hat{\mathbf{a}}^{(l)}, \hat{\boldsymbol{\theta}}^{(l)}, \mathcal{I}] \end{aligned} \tag{4.10}$$

2-M step:

$$\hat{\boldsymbol{\theta}}^{(l+1)} = \arg\max_{\boldsymbol{\theta}}\{Q_1(\boldsymbol{\theta}|\hat{\mathbf{a}}^{(l)}, \hat{\boldsymbol{\theta}}^{(l)})\} \tag{4.11}$$

Detection part:

[1] Similar to the EM algorithm , GMLSDE may only achieve a local maximum point if $L(\mathcal{U})$ has many maxima.

[2] In general dividing the parameter set into two separate sets (Theorem 1) does not guarantee achieving ML criterion; however, since $\mathbf{a}$ is discrete and the detection part considers all the possibilities of $\mathbf{a}$, the algorithm is guaranteed to achieve ML criterion when there are no local maxima.

3-E step[3]:

$$\begin{aligned} Q_2(\mathbf{a}|\hat{\mathbf{a}}^{(l)}, \hat{\boldsymbol{\theta}}^{(l)}) &= Q(\mathbf{a}, \hat{\boldsymbol{\theta}}^{(l+1)}|\hat{\mathbf{a}}^{(l)}, \hat{\boldsymbol{\theta}}^{(l)}) \\ &= E[\log p(\mathcal{C}|\mathbf{a}, \hat{\boldsymbol{\theta}}^{(l+1)})|\hat{\mathbf{a}}^{(l)}, \hat{\boldsymbol{\theta}}^{(l)}, \mathcal{I}] \end{aligned} \tag{4.12}$$

4-M step:

$$\hat{\mathbf{a}}^{(l+1)} = \arg\max_{\mathbf{a}}\{Q_2(\mathbf{a}|\hat{\mathbf{a}}^{(l)}, \hat{\boldsymbol{\theta}}^{(l)})\} \tag{4.13}$$

Dividing the parameters as discrete and continuous decreases the dimension of the parameter space and usually increases the convergence of the algorithm [69]. Also it allows us to use different methods in the maximizing steps for discrete and continuous parameters. *GMLSDE shows a natural coupling between estimation and detection, which should allow us to get some insight about the influence of either one on the other.* This has not been possible with earlier approaches to the combined detection/estimation problem since channel estimation has been traditionally uncoupled from the detection problem when developing the detection algorithm.

The proposed GMLSDE algorithm,(4.10)-(4.13), has been developed using the entire received signal $\mathbf{y}$ or, in other words, the EM algorithm is off-line. Since the main purpose of an adaptive algorithm is to detect the sequentially emitted $\mathbf{a}$ in real time, we are interested in an on-line version of the EM algorithm. By defining $\mathcal{C}_k$ and $\mathcal{I}_k$ as the complete data and the incomplete data available up to time k respectively, the steps of GMLSDE algorithm at time k using the on-line scheme for estimating and detecting are given as

Estimation part:

1-E step:

$$Q_{1,k}(\boldsymbol{\theta}_k|\tilde{\mathbf{a}}_{k|k-1}, \tilde{\boldsymbol{\theta}}_{k|k-1}) = E[\log p(\mathcal{C}_k|\tilde{\mathbf{a}}_{k|k-1}, \boldsymbol{\theta}_k)|\tilde{\mathbf{a}}_{k|k-1}, \tilde{\boldsymbol{\theta}}_{k|k-1}, \mathcal{I}_k] \tag{4.14}$$

[3]Expectation step of the detection part is the same as the expectation of the estimation part, except that $\hat{\boldsymbol{\theta}}^{(l)}$ is used instead of $\boldsymbol{\theta}$ and $\mathbf{a}$ is considered as a parameter.

2-M step:

$$\tilde{\boldsymbol{\theta}}_{k|k} = \arg\max_{\boldsymbol{\theta}_k}\{Q_{1,k}(\boldsymbol{\theta}_k|\tilde{\mathbf{a}}_{k|k-1}, \tilde{\boldsymbol{\theta}}_{k|k-1})\} \tag{4.15}$$

Detection part:

3-E step:

$$Q_{2,k}(\mathbf{a}_k|\tilde{\mathbf{a}}_{k|k-1}, \tilde{\boldsymbol{\theta}}_{k|k-1}) = E[\log p(\mathcal{C}_k|\mathbf{a}_k, \tilde{\boldsymbol{\theta}}_{k|k})|\tilde{\mathbf{a}}_{k|k-1}, \tilde{\boldsymbol{\theta}}_{k|k-1}, \mathcal{I}_k] \tag{4.16}$$

4-M step:

$$\tilde{\mathbf{a}}_{k|k} = \arg\max_{\mathbf{a}_k}\{Q_{2,k}(\mathbf{a}_k|\tilde{\mathbf{a}}_{k|k-1}, \tilde{\boldsymbol{\theta}}_{k|k-1})\} \tag{4.17}$$

where $\{\mathbf{a}_k, \boldsymbol{\theta}_k\}$ is the parameter set up to time k. Also, $\tilde{\mathbf{a}}_{k|l}$ and $\tilde{\boldsymbol{\theta}}_{k|l}$ are the detected value of $\mathbf{a}_k$ and the estimated of $\boldsymbol{\theta}_k$ based on the data received up to time l respectively. Similar to Theorem 1 in Chapter 3, it can also be shown that the likelihood function is increased at each recursion[4] (increasing time). Meanwhile, the time-up-date vector $\tilde{\boldsymbol{\theta}}_{k+1|k}$ is given by

$$\tilde{\boldsymbol{\theta}}_{k+1|k} = \begin{bmatrix} \tilde{\boldsymbol{\varphi}}_{k+1|k} \\ \tilde{\boldsymbol{\theta}}_{k|k} \end{bmatrix}$$

where $\tilde{\boldsymbol{\theta}}_{k|k} = [\tilde{\boldsymbol{\varphi}}^{\mathcal{T}}_{k|k}, \tilde{\boldsymbol{\varphi}}^{\mathcal{T}}_{k-1|k}, ..., \tilde{\boldsymbol{\varphi}}^{\mathcal{T}}_{0|k}]^{\mathcal{T}}$ and $\tilde{\boldsymbol{\varphi}}_{k+1|k}$ is obtained generally from the dynamic evolution of the $\boldsymbol{\varphi}_{k+1}$ process,

$$\boldsymbol{\varphi}_{k+1} = f_\theta(\boldsymbol{\theta}_k, \mathbf{a}_k) \tag{4.18}$$

The elements of the transmitted sequential symbols, $\mathbf{a} = \{a_k\}_{k=0}^{I-1}$, are selected from a finite set (i.e. $a_k \in \{\pm 1 \pm j\}$ in QPSK) and generally (assuming no coding) the sequential symbols are independent of each other. In other words (unlike (4.18))there is no dynamic relation between the present symbol and the previous symbols. Therefore

[4] In general, each recursion (4.14-4.17) can be implemented with some iterations (4.10-4.13). In this case, $\{\tilde{\mathbf{a}}_{k|k-1}, \tilde{\boldsymbol{\theta}}_{k|k-1}\}$ is the set of initial values for the first iteration at time k and $\{\tilde{\mathbf{a}}_{k|k}, \tilde{\boldsymbol{\theta}}_{k|k}\}$ which is the detection/estimation of the unknown parameter set at recursion k (or the result of the the first iteration at time k) and is the initial value for the second iteration at time k. This procedure will always increase the likelihood.

the time-update $\tilde{\mathbf{a}}_{k+1|k}$ is defined by

$$\tilde{\mathbf{a}}_{k+1|k} = \begin{bmatrix} a_{k+1} \\ \tilde{\mathbf{a}}_{k|k} \end{bmatrix}$$

where a_{k+1} is selected from all the possibilities of the alphabet set (i.e. $\{\pm 1 \pm j\}$ in QPSK). Meanwhile, in general the estimation part of GMLSDE may contain more than one step of expectation and maximization. Based on Theorem 1 in Chapter 3, the unknown parameter set $\boldsymbol{\theta}$ may be divided into different sets and each set will have its own expectation and maximization steps. This idea is very useful in increasing the convergence of the algorithm especially when the complete data is different for each separate set [70]. Meanwhile from the view-point of increasing the likelihood, estimation and detection can be done in a different order in GMLSDE. However in order to decrease the complexity or due some other practical issues, it may be that one order (i.e. estimation after detection) is preferable to the other order. Doing the estimation part before the detection part leads to the concept of per-branch processing (PBP) which estimates different channel parameters for each branch metric. However, when the detection part is done before the estimation part, channel parameters are estimated for each survivor path and this method leads to the idea of PSP (i.e. see Section 4.3).

As the estimation and detection parts of GMLSDE (4.10)-(4.13) show, the idea of using the previous decisions and estimates as a tool for detecting and estimating the future arises naturally in implementing MLSDE based on the EM algorithm. The decision-based receiver was used in [17] for an ISI channel with infinite impulse response. Later it was proposed in many such joint detection and estimation methods and is generally known as per-survivor processing (PSP) [8]. The PSP was originally proposed as a practical way to implement joint detection and estimation; however in GMLSDE, due to the inherent coupling between the estimation and detection parts, the temporary decision of the data, $\tilde{\mathbf{a}}_{k|k-1}$, at time k is a given condition in the estimation part. Therefore the idea of PSP (estimating different CIR or other channel parameters for different survivor paths) comes up naturally as an integral part of the EM-based ML detection/estimation procedure. Thus one can say that the EM algorithm provides a solid theoretical foundation for using the PSP in ML-based receivers due to the inherent embedding of decision feedback in the EM approach. Meanwhile, if all survivor paths at time k have the same root in the trellis diagram at time $k - L_d$, and due to only one detected sequence $\tilde{\mathbf{a}}_{k-L_d}$, there will be only one estimation for $\boldsymbol{\theta}_{k-L_d}$ at time k, $\tilde{\boldsymbol{\theta}}_{k-L_d|k}$. If we assume that $\tilde{\boldsymbol{\theta}}_{k|k} \simeq \tilde{\boldsymbol{\theta}}_{k-L_d|k}$, the detection/estimation procedure leads to the MLSDE receiver which was proposed by Qureshi [34] before the idea of PSP emerged in the research literature.

The convergence of the EM algorithm is inversely related to the dimension of its complete data space. Less necessary and less informative complete data improves the asymptotic convergence rate. The on-line EM algorithm deals with unknown parameters and complete data only up to the process time and their dimensions increase linearly with time, thus causing faster convergence. This idea is similar to the space-alternating generalized EM (SAGE) method which achieves faster convergence by partitioning the parameters and the complete data [70]. In the on-line EM algorithm, the parameters and complete data are naturally partitioned through time. Meanwhile, due to the parameter space-time coupling, the on-line EM algorithm is a recursive algorithm (each recursion can generally be done by more than one iteration) whereas the off-line EM algorithm is only an iterative algorithm in general.

As we considered in Chapter 3, the estimation part in the on-line EM algorithm can be approximately implemented by a recursive formula based on the modified Titterington's approach [43]. Using a stochastic approximation in order to consider only three elements of a Taylor's series expansion of $Q_{1,k}(\boldsymbol{\theta}|\tilde{\mathcal{U}}_{k|k-1})$,(4.14), at point $\boldsymbol{\theta}_k = \tilde{\boldsymbol{\theta}}_{k|k-1}$, one can show (4.15) becomes

$$\tilde{\boldsymbol{\theta}}_{k|k} \simeq \tilde{\boldsymbol{\theta}}_{k|k-1} - \left(\frac{\partial^2 Q_{1,k}(\boldsymbol{\theta}_k|\tilde{\mathcal{U}}_{k|k-1})}{\partial^2 \boldsymbol{\theta}_k}\Big|_{\boldsymbol{\theta}_k=\tilde{\boldsymbol{\theta}}_{k|k-1}}\right)^{-1}\left(\frac{\partial Q_{1,k}(\boldsymbol{\theta}_k|\tilde{\mathcal{U}}_{k|k-1})}{\partial \boldsymbol{\theta}_k^*}\Big|_{\boldsymbol{\theta}_k=\tilde{\boldsymbol{\theta}}_{k|k-1}}\right)$$
$$\text{for all } \tilde{\mathcal{U}}_{k|k-1} \quad (4.19)$$

where $\tilde{\mathcal{U}}_{k|k-1} = \{\tilde{\mathbf{a}}_{k|k-1}, \tilde{\boldsymbol{\theta}}_{k|k-1}\}$ and $\boldsymbol{\theta}^*$ denotes conjugate of $\boldsymbol{\theta}$. If $\boldsymbol{\varphi}_k$ is an $L+1$ vector, $\boldsymbol{\theta}_k$ and $\frac{\partial Q_{1,k}(\boldsymbol{\theta}_k|\tilde{\mathcal{U}}_{k|k-1})}{\partial \boldsymbol{\theta}_k^*}$ are $k(L+1)$ vectors and also $\frac{\partial^2 Q_{1,k}(\boldsymbol{\theta}_k|\tilde{\mathcal{U}}_{k|k-1})}{\partial^2 \boldsymbol{\theta}_k}$ is an $k(L+1)\times k(L+1)$

matrix. It should be noted that when the third and higher derivatives of $Q_{1,k}(.|.)$ are zero, as is usually true for the Gaussian case, the recusive formula (4.19) is exact.

The detection part, due to the finite alphabet of the transmitted symbols, can be implemented with a dynamic programming (Viterbi) algorithm. The expectation step in the detection part (4.16) is obtained for all possibilities of $\{\tilde{\mathbf{a}}_{k|k-1}, \tilde{\boldsymbol{\theta}}_{k|k}\}$ at time k. If only one $\boldsymbol{\varphi}_{k|k-1}$ is estimated for each $\tilde{\mathcal{U}}_{k-1|k-1} = \{\tilde{\mathbf{a}}_{k-1|k-1}, \tilde{\boldsymbol{\theta}}_{k-1|k-1}\}$ and a_k is selected from a q-ary signaling scheme, there are q possible estimates for $\tilde{\mathcal{U}}_{k|k-1}$ and $\tilde{\boldsymbol{\theta}}_{k|k}$ in each estimated $\tilde{\mathcal{U}}_{k-1|k-1}$[5]. When the duration of the impulse response of the transmitter filter is one and the length of the channel memory is L, based on the trellis structure, the number of detection/estimation of $\tilde{\mathcal{U}}_{k|k}$ is q^L and the number of the possibilities of $\tilde{\mathcal{U}}_{k|k-1}$ is q^{L+1} (e.g. $\tilde{\mathcal{U}}^{(\vartheta)}_{k|k-1}$ for $\vartheta = 1, ..., q^{L+1}$). Therefore each $\tilde{\mathcal{U}}_{k|k}$ is maximized over q values of $\tilde{\mathcal{U}}_{k|k-1}$. To be more precise, each $\tilde{\mathcal{U}}_{k|k-1}$ should correspond with each hypothesis of $\boldsymbol{a}_k$, the sequence of transmitted symbols affecting $y(k)$. For easy presentation, however, we avoid using the $\tilde{\mathcal{U}}^{(\vartheta)}_{k|k-1}$ notation.

Meanwhile, although the EM algorithm is a method to achieve ML criterion with expectation and maximization steps, the algorithm does not tell us how to do these steps. We only considered some very general implementation aspects of the detection and estimation steps in this section. More details of the GMLSDE implementation by adaptive algorithms will be explored in the next section based on different channel models.

4.2 Adaptive MLSDE Based on Channel Models

The GMLSDE algorithm was derived in Section 4.1 in a general framework without specifying any channel model. In this section we show how some adaptive MLSD/MLSDE algorithms developed previously in the literature along with some new adaptive MLSDE algorithms can be derived from the generalized MLSDE. The main goal of this section

[5]In PBP method, there are q possible estimates for channel parameters, $\tilde{\boldsymbol{\theta}}_{k|k}$, for each state. However, in PSP method there is one estimate for each state and $\tilde{\boldsymbol{\theta}}_{k|k}$ is computed only for survivor paths.

is to show the power of the proposed generalized MLSDE algorithm in deriving joint detection and estimation algorithms in a unified way based on the different channel models and available channel knowledge.

We consider three different model categories for a linear channel:

1. Known CIR.

2. Unknown deterministic CIR (time-invariant/variant).

3. Stochastic CIR (random vector/process) $\begin{cases} \text{known statistical parameters.} \\ \text{unknown statistical parameters.} \end{cases}$

In all above models the additive noise, $z(k)$ in (4.3), is considered as a circularly, symmetric, zero mean, white, complex, Gaussian random process whose variance is N_0. The main step for developing an adaptive MLSDE algorithm is to define $\boldsymbol{\theta}_k$, $\mathcal{C}_k$ and $\mathcal{I}_k$ according to the channel model. Based on the definitions of these parameters, the procedure to implement adaptive MLSDE may contain only a detection part, an estimation part or both. Also, it may need to do only the maximization step in the detection/estimation part. We focus on the statistical CIR model which is suitable for mobile communications and briefly mention known CIR and unknown deterministic CIR models whose MLSD/MLSDE algorithms are well known in the literature. Meanwhile, in order to reduce the complexity and implement the algorithm in a causal manner (as we explain later), adaptive MLSDE may not achieve the maximum likelihood, but increases the likelihood function in each recursion.

A discrete model of the communication system for a linear channel is shown in Fig.2.2. $\mathbf{a} = \{a_k\}_{k=0}^{I}$ is the set of the transmitted symbols and b_k is upsampling of a_k by a factor $J = \lceil \frac{T}{T_s} \rceil$, where T is the symbol period and T_s is the sample period[6]. The received signal is

$$y(k) = \sum_{l=0}^{L} h(l,k)s(k-l) + z(k) = \boldsymbol{s}(k)\boldsymbol{h}(k) + z(k) \tag{4.20}$$

[6]When $J > 1$, each recursion in the detection part needs J recursions in the estimation part.

where the duration of $h(l,k)$ with respect to l is $L+1$, $s(k)$ is the output of the transmitter filter with impulse response $g(k)$, $z(k)$ is additive noise, $\boldsymbol{s}(k) = [s(k), ..., s(k-L)]$ and $\boldsymbol{h}(k) = [h(0,k), ..., h(L,k)]^{\mathcal{T}}$.

Model a: Known CIR

The MLSD receiver for known CIR was derived by Forney [26]. From the generalized MLSDE point of view the complete data is available in the receiver or $\mathcal{C}_k = \mathcal{I}_k = \mathbf{y}_k = [y(k), ..., y(0)]^{\mathcal{T}}$ and there is no unknown parameter set $\boldsymbol{\theta}$. Therefore only the maximization step of the detection part (4.17) is needed and it can be done recursively by the Viterbi algorithm [26].

Model b: Unknown Deterministic CIR It is common to consider the CIR as a vector of time-variant/invariant deterministic parameters in channels with ISI. In this model for a time-invariant CIR the vector of unknown continuous parameters is defined as $\boldsymbol{\varphi} = \boldsymbol{h} = [h(0), ..., h(L)]^{\mathcal{T}}$ and at time k the complete and incomplete data are equal $\mathcal{C}_k = \mathcal{I}_k = \mathbf{y}_k$. Thus only the maximization steps of estimation and detection parts, (4.15) and (4.17), are needed in this model. It can be shown that (4.15) leads to the RLS algorithm by using (4.19) [44]. If $\tilde{\boldsymbol{h}}_{|k}$ is the estimation of CIR at time k, (4.17) becomes

$$\begin{aligned}
\tilde{\mathbf{a}}_{k|k} &= \arg\max_{\mathbf{a}_k}\{Q_{2,k}(\mathbf{a}_k|\tilde{\mathbf{a}}_{k|k-1}, \tilde{\boldsymbol{h}}_{|k-1})\} = \arg\max_{\mathbf{a}_k}\{\log p(\mathbf{y}_k|\mathbf{a}_k, \tilde{\boldsymbol{h}}_{|k})\} \\
&= \arg\max_{\mathbf{a}_k}\{\sum_{l=0}^{k} -|y(l) - \boldsymbol{s}(l)\tilde{\boldsymbol{h}}_{|k}|^2\} \qquad \text{for all } \{\tilde{\mathbf{a}}_{k|k-1}, \tilde{\boldsymbol{h}}_{|k}\} \qquad (4.21)
\end{aligned}$$

Using the new estimation of $\boldsymbol{h}$ at time k, $\tilde{\boldsymbol{h}}_{|k}$ in order to detect $\mathbf{a}_k = \{a_k, ..., a_0\}$ from (4.21) is a very complex and non-causal process. In order to avoid complexity, a causal detection procedure[7] is considered. Therefore instead of $\tilde{\boldsymbol{h}}_{|k}$, $\tilde{\boldsymbol{h}}_{|l}$ is used in calculating $|y(l) - \boldsymbol{s}(l)\tilde{\boldsymbol{h}}_{|k}|^2$ for $0 \le l \le k$ in (4.21). Thus (4.21) becomes

$$\tilde{\mathbf{a}}_{k|k} = \arg\max_{\mathbf{a}_k}\{\sum_{l=0}^{k} -|y(l) - \boldsymbol{s}(l)\tilde{\boldsymbol{h}}_{|l}|^2\} \qquad (4.22)$$

[7]Detecting a_k from $y(l)$ for $l \le k$ is defined as a causal detecting procedure.

Since $\tilde{\mathbf{a}}_{k-1|k-1}$ maximizes $\sum_{l=0}^{k-1} -|y(l) - \boldsymbol{s}(l)\tilde{\boldsymbol{h}}_{|l}|^2$, (4.22) can be implemented in a recursive manner using the Viterbi algorithm. It should be noted that the causal detection procedure does not guarantee achieving the maximum of the likelihood function (global or local); however it guarantees to increase the likelihood function, $L(\mathcal{U})$, in each recursion. Meanwhile, the estimation and detection procedures shows how the idea of PSP [8] emerges from the GMLSDE algorithm as a recursive approach to increase the likelihood function.

For a time-variant deterministic CIR, the estimation part leads to Kalman type algorithm [44] where the causal detection procedure needs only a causal estimation procedure[8].

Model c: Stochastic CIR with known parameters

The CIR is often modeled as a random vector or random process in a mobile environment. For example, Rayleigh multipath fading whose impulse response is considered as a complex Gaussian random vector (for very slow fading) or process (for fast fading) is a very common model in mobile communication systems. In the Model "c" we focus on the random process CIR. The received signal is obtained from (4.20) with $h(l,k) = h_l(k)$ being considered a Gaussian random process for $0 \leq l \leq L$. By defining $\mathbf{h}_k = \mathbf{h}(k) = [\boldsymbol{h}(k)^{\mathcal{T}}, \boldsymbol{h}(k-1)^{\mathcal{T}}, ..., \boldsymbol{h}(k-M+1)^{\mathcal{T}}]^{\mathcal{T}}$, the received signal $y(k)$ and the dynamic changing of the CIR represented by the state-space formulation is given by

$$y(k) = \mathbf{s}(k)\mathbf{h}(k) + z(k) \tag{4.23}$$

$$\mathbf{h}(k) = \boldsymbol{F}\mathbf{h}(k-1) + G\boldsymbol{w}(k) \tag{4.24}$$

where $\mathbf{s}(k) = [\boldsymbol{s}(k), \mathbf{0}]$, $\mathbf{0}$ is an $(M-1)(L+1)$ zero row vector, $\boldsymbol{w}(k) = [w_0(k), ..., w_L(k)]^{\mathcal{T}}$ is a zero mean white complex stationary Gaussian vector process whose autocorrelation matrix is $R_{\boldsymbol{w}}(k) = I_{(L+1)}\delta(k)$, $I_{(L+1)}$ is an $(L+1) \times (L+1)$ identity matrix, and it is

[8]Estimating $\mathbf{h}(k) = [\boldsymbol{h}(k)^{\mathcal{T}}, ..., \boldsymbol{h}(k-M+1)^{\mathcal{T}}]^{\mathcal{T}}$ from $y(l)$ for $l \leq k$ is defined as a causal estimation procedure.

independent of $z(k)$. $\boldsymbol{F}$ and G matrices are defined by

$$\boldsymbol{F} = \begin{bmatrix} F_1 & F_2 & \dots & F_M \\ I_{(L+1)} & 0_{(L+1)} & \dots & 0_{(L+1)} \\ & \ddots & 0_{(L+1)} & \\ & 0_{(L+1)} & I_{(L+1)} & 0_{(L+1)} \end{bmatrix} \qquad G = \begin{bmatrix} \mathbf{g} \\ 0_{(L+1)} \\ \vdots \\ 0_{(L+1)} \end{bmatrix} \tag{4.25}$$

where $0_{(L+1)}$ is an $(L+1) \times (L+1)$ zero matrix.

In this channel model, the complete data and the incomplete data at time k are $\mathcal{C}_k = \{\mathbf{y}_k, \bar{\mathbf{h}}_k\}$ and $\mathcal{I}_k = \mathbf{y}_k$ respectively, where $\bar{\mathbf{h}}_k = [\mathbf{h}_k^{\mathcal{T}}, ..., \mathbf{h}_0^{\mathcal{T}}]^{\mathcal{T}}$. It is easy to show that the unknown continuous parameters at time k is $\Theta_{k|k} = \{\theta_{k|k}, \theta_{k-1|k}, ..., \theta_{0|k}\}$ where $\theta_{l|k} = \{\bar{\boldsymbol{\mu}}_{l|k}, \boldsymbol{\Sigma}_{l|k}\}$, $\bar{\boldsymbol{\mu}}_{l|k} = E[\bar{\mathbf{h}}_l|\mathbf{y}_k]$ and $\boldsymbol{\Sigma}_{l|k} = \text{cov}(\bar{\mathbf{h}}_l|\mathbf{y}_k)$ for $0 \leq l \leq k$. In this model all the steps in detecting and estimating parts are necessary. The expectation step of detection part (4.16) becomes

$$\begin{aligned} Q_{2,k}(\mathbf{a}_k|\tilde{\mathbf{a}}_{k|k-1}, \tilde{\Theta}_{k|k-1}) &= E[\log p(\mathbf{y}_k, \bar{\mathbf{h}}_k|\mathbf{a}_k, \tilde{\Theta}_{k|k})|\tilde{\mathbf{a}}_{k|k-1}, \tilde{\Theta}_{k|k-1}, \mathbf{y}_k] \\ &= \sum_{l=0}^{k}\{\log p(y(l)|\mathbf{a}_k, \tilde{\theta}_{l|k-1}, \mathbf{y}_{k-1}) \\ &+ E[\log p(\boldsymbol{h}(l)|\mathbf{a}_k, \tilde{\theta}_{l|k}, \mathbf{h}(l-1), \mathbf{y}_k|\tilde{\mathbf{a}}_{k|k-1}, \tilde{\Theta}_{l|k-1}, \mathbf{y}_k]\} \end{aligned} \tag{4.26}$$

Due to the non-causal estimation of $\theta_{l|k-1}$ for $0 \leq l \leq k$, detecting $\mathbf{a}_k$ from (4.26) is very complex. Considering causal estimation of $\boldsymbol{\mu}_{l|k-1} = E[\mathbf{h}_l|\mathbf{y}_{k-1}]$ and $\Sigma_{l|k-1} = \text{cov}(\mathbf{h}_l|\mathbf{y}_{k-1})$ the detecting procedure at time k becomes

$$\tilde{\mathbf{a}}_{k|k} = \arg\max_{\mathbf{a}_k}\{\sum_{l=0}^{k} -\beta(l)\} \tag{4.27}$$

such that the branch metric $\beta(l)$ is given by

$$\begin{aligned} \beta(l) &= \frac{|y(l) - \tilde{\mathbf{s}}(l)\tilde{\boldsymbol{\mu}}_{l|l-1}|^2}{\tilde{\mathbf{s}}(l)\tilde{\Sigma}_{l|l-1}\tilde{\mathbf{s}}(l)^{\mathcal{H}} + N_0} + \log(\tilde{\mathbf{s}}(l)\tilde{\Sigma}_{l|l-1}\tilde{\mathbf{s}}(l)^{\mathcal{H}} + N_0) \\ &+ \log(\det(\bar{\boldsymbol{F}}\tilde{\Sigma}_{\boldsymbol{x}_{l|l}}\bar{\boldsymbol{F}}^{\mathcal{H}}) \end{aligned} \tag{4.28}$$

where $\boldsymbol{x} = [\boldsymbol{h}(k)^{\mathcal{T}}, ..., \boldsymbol{h}(k-M)^{\mathcal{T}}]^{\mathcal{T}}$, $\bar{\boldsymbol{F}} = [I_{(L+1)}, -F_1, ..., -F_M]$ and $\Sigma_{\boldsymbol{x}_{l|l}} = \text{cov}(\boldsymbol{x}_l|\mathbf{y}_l)$. By using the Viterbi algorithm (4.27) can be implemented in a recursive manner where $\beta(l)$ is the branch metric of the trellis diagram at time l. It can be shown that causal estimation leads the estimating part of Model "c" to the Kalman algorithm by using (4.19) where $\tilde{\boldsymbol{\mu}}_{l+1|l}$ and $\tilde{\Sigma}_{l+1|l}$ are updated for all possibilities of $\{\tilde{\mathbf{a}}_{l|l-1}, \tilde{\boldsymbol{\theta}}_{l|l-1}\}$ using the recursions given by [44, 80]

$$\tilde{\boldsymbol{\mu}}_{l+1|l} = \boldsymbol{F}\tilde{\boldsymbol{\mu}}_{l|l-1} + \boldsymbol{F}\tilde{\Sigma}_{l|l-1}\tilde{\mathbf{s}}(l)^{\mathcal{H}}(N_0 + \tilde{\mathbf{s}}(l)\tilde{\Sigma}_{l|l-1}\tilde{\mathbf{s}}(l)^{\mathcal{H}})^{-1}(y(l) - \tilde{\mathbf{s}}(l)\tilde{\boldsymbol{\mu}}_{l|l-1}) \tag{4.29}$$

$$\tilde{\Sigma}_{l+1|l} = \boldsymbol{F}\tilde{\Sigma}_{l|l}\boldsymbol{F}^{\mathcal{H}} + GG^{\mathcal{H}} \tag{4.30}$$

where

$$\tilde{\Sigma}_{l|l} = \tilde{\Sigma}_{l|l-1} + \tilde{\Sigma}_{l|l-1}\tilde{\mathbf{s}}(l)^{\mathcal{H}}(N_0 + \tilde{\mathbf{s}}(l)\tilde{\Sigma}_{l|l-1}\tilde{\mathbf{s}}(l)^{\mathcal{H}})^{-1}\tilde{\mathbf{s}}(l)\tilde{\Sigma}_{l|l-1} \tag{4.31}$$

Meanwhile, it is straightforward to compute $\tilde{\Sigma}_{\boldsymbol{x}_{l|l}}$ from $\tilde{\Sigma}_{l|l}$ using (4.24).

The branch metric measure derived in (4.28) is different from the branch metric measure proposed in [10, 11, 13, 37]. Although the same state-space model was chosen in the above references, the last term in (4.28) is extra in comparison with the branch metric proposed in [10, 11, 13, 37] which maximizes the logarithm of the pdf of $\mathbf{y}_k$ at time k to compute the branch metrics. In GMLSDE using Model "c", however, the expectation of the logarithm of the joint pdf of $\mathbf{y}_k$ and $\bar{\mathbf{h}}_k$ over $\bar{\mathbf{h}}_k$ is maximized. Thus the last term in (4.28) can be interpreted as the contribution of the error in estimating $\boldsymbol{h}(k)$ given $y(k)$ to the branch metric at time k. Meanwhile since $\bar{\boldsymbol{F}}\tilde{\Sigma}_{\boldsymbol{x}_{l|l-1}} = 0$, if $\tilde{\Sigma}_{\boldsymbol{x}_{l|l}}$ is replaced with $\tilde{\Sigma}_{\boldsymbol{x}_{l|l-1}}$ (using PSP method instead of PBP method), the last term of (4.28) vanishes in computing the branch metric.

Model d: Stochastic CIR with unknown parameters

In the estimation part of the Model "c", it was assumed that $\boldsymbol{F}$ and $GG^{\mathcal{H}}$ matrices were known in the updating equations (4.29) and (4.30). Generally, these matrices are also unknown and should be estimated from the received signal. The elements of $\boldsymbol{F}$ and $GG^{\mathcal{H}}$ can be estimated using (3.60) and (3.64) in Chapter 3. It should be mentioned that the detection part of model "d" is similar to the detection part of model "c" [47]. The steps of the adaptive MLSDE receivers needed for different channel models are

Channel Model	Estimation Part	Detection Part
Known CIR	-	"M" step (VA)
Unknown Deterministic CIR	"M" step (RLS or RLS-type)	"M" step (VA)
Stochastic CIR Known $\boldsymbol{F}$, G	"E" and "M" steps (Kalman)	"E" and "M" steps (VA)
Stochastic CIR Unknown $\boldsymbol{F}$, G	Two "E" and "M" steps (RLS-type and Kalman)	"E" and "M" steps (VA)

Table 4.1: The needed steps of the adaptive MLSDE receivers for different channel models.

summarized in Table 4.1.

Hart has recently proposed a method to estimate the unknown statistical parameters of Ricean fading channels [37]. The method in [37] estimates the mean vector and autocovariance matrix of CIR based on computing the mean and autocovariance of the received signal; however, it is complex and non-recursive [57]. The method proposed in Model "d" estimates the unknown parameters based on increasing the likelihood function in a recursive manner.

4.3 Computer Simulations and Comparisons

Computer simulations have been done for different channel models to evaluate the performance of the generalized MLSDE algorithm. The channel model is selected as flat fading and frequency-selective fading with three fading rates, $f_d T = 0.1, 0.01$ and 0.001. The channel is simulated with three paths in frequency selective channel, Fig. 4.1. The autocorrelation function, impulse response of the transmitter filter, modulation scheme and the number of samples per symbol are selected to be the same as the ones selected in chapter 2. Also, similar to chapter 2, the Bessel fading filter for omnidirectional antenna is approximated by an all pole third order filter, therefore the $\boldsymbol{F}$ in (4.25) becomes 3×3 and 9×9 matrix in the flat fading and the selective fading channels respectively. The number of states in the trellis diagram therefore is four and sixteen for flat fading and selective fading channels respectively. The data sequence is divided

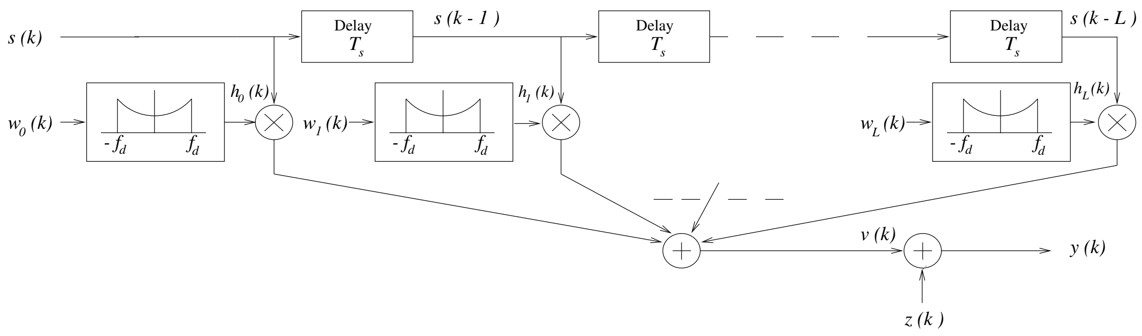

Figure 4.1: The baseband tapped delay line model of the multipath fading channel.

into a sequence of frames with length L_f, where the overhead of each frame known by the receiver is one and two symbols for flat and selective fading respectively and the length of each frame was chosen as $L_f = 160$ data.

The bit error rate performance of the adaptive MLSDE algorithm based on the on-line EM algorithm is considered in association with different levels of channel knowledge which is available at the receiver. Although the impulse response of the fading channel was simulated as a stochastic random process, we consider four levels of available channel knowledge at the receiver : a) known CIR, b) unknown deterministic CIR, c) stochastic CIR with known channel parameters, F and $GG^{\mathcal{H}}$ matrices and d) stochastic CIR with unknown F and $GG^{\mathcal{H}}$ matrices.

Figures 4.2, 4.3 and 4.4 show the error performances of the adaptive MLSDE in the flat fading channel for $f_dT = 0.1, 0.01$ and 0.001 respectively. Also, the performances of the adaptive MLSDE for the selective fading channel are shown in figures 4.5, 4.6 and 4.7 for $f_dT = 0.1, 0.01$ and 0.001 respectively.

In these figures, the simulation results in curves (a), (b), (c) and (d) correspond to the Models "a", "b", "c" and "d" in Section 4.2 respectively. Curves (a) show the performance of the receiver when the CIR is known; curves (b) show how the receiver performs when it assumes a deterministic CIR and estimates it by the modified RLS algorithm with a proper forgetting factor, λ; for different fading rates (in order to optimize the RLS performance, see Chapter 3); the performance of the receiver in curves (c) are achieved based on the assumption of a stochastic CIR, whose stochastic parameters are estimated by Kalman filtering, with known channel parameters $\boldsymbol{F}$ and

$GG^{\mathcal{H}}$ and finally curves (d) indicate the receiver performance with a stochastic model for CIR and unknown channel parameters $\boldsymbol{F}$ and $GG^{\mathcal{H}}$, where both Kalman and RLS algorithms are used in the estimation part.

As can be seen, the performances of the receiver in (b) is close to the performance of (c) only for slow flat fading and very slow selective channels. By increasing the fading rate the difference in performance between (b) and (c) increases due to an insufficient degree of freedom in the RLS up dating equation ($\boldsymbol{F} = \lambda^{-\frac{1}{2}}I$, see Chapter 3) to track the dynamic channel changes.

The performances of (c) and (d) are close for $f_dT = 0.01$ and $f_dT = 0.001$. It should be mentioned that in order to limit the complexity of the algorithm and time of simulations in (d), the $\boldsymbol{F}$ and $GG^{\mathcal{H}}$ matrices were updated in each frame of data instead of each recursion. However in (b), (c) and (d) cases, the CIR is estimated in each recursion accompanied by PSP which estimates different CIR for each survivor path in the trellis diagram. Fig. 4.8 shows the bit error rate of a flat fading channel for (a), (b), (c) and (d) cases when the fading rate is changing periodically between $f_dT = 0.1$ and $f_dT = 0.01$. One period of $f_dT(t)$ is

$$f_dT(t) = \begin{cases} 0.1 & 0 \leq t \leq T_f/4 \\ -\frac{1}{800000T}(t - \frac{T_f}{4}) + 0.1 & T_f/4 \leq t \leq T_f/2 \\ 0.01 & T_f/2 \leq t \leq 3T_f/4 \\ \frac{1}{800000T}(t - \frac{3T_f}{4}) + 0.01 & 3T_f/4 \leq t \leq T_f \end{cases} \tag{4.32}$$

where the period time $T_f = 28800T$. The linear change in fading rate corresponds to a linear change in vehicle speed. In this situation the modified RLS algorithm with forgetting factor $\hat{\lambda} = 0.999$ is used in estimating the $\boldsymbol{F}$ and $GG^{\mathcal{H}}$ matrices. As Fig. 4.8 shows, the performance of (c), with known $\boldsymbol{F}$ and $GG^{\mathcal{H}}$ and (d), with unknown $\boldsymbol{F}$ and $GG^{\mathcal{H}}$, are close. However the performance of (b), with the assumption of unknown deterministic CIR, is far from (c). Therefore the maximum fading rate dominates the performance when the fading rate is time-variant.

When the CIR is modeled as a random process (Model "c") or time-variant deterministic process (Model "b"), the estimation part can be done through a Kalman type algorithm. However, the branch metrics are computed in a different manner. For a CIR modeled as a Gaussian random process, the branch metric is computed from (4.28) as shown in Model "c" or from (4.33) which was derived in [10, 11, 13, 37] by maximizing $\log p(\mathbf{y}_k|\mathbf{a}_k)$.

$$\beta(l) = \frac{|y(l) - \mathbf{s}(l)\tilde{\boldsymbol{\mu}}_{l|l-1}|^2}{\mathbf{s}(l)\tilde{\Sigma}_{l|l-1}\mathbf{s}(l)^{\mathcal{H}} + N_0} + \log(\mathbf{s}(l)\tilde{\Sigma}_{l|l-1}\mathbf{s}(l)^{\mathcal{H}} + N_0) \tag{4.33}$$

For deterministic time-variant CIR, similar to Model "b" the branch metric becomes

$$\beta(k) = |y(l) - \mathbf{s}(l)\tilde{\mathbf{h}}_{l|l}|^2 \tag{4.34}$$

Simulation results (which have not been shown) for flat and selective fading do not show a significant difference in performance between computing branch metrics from (4.28) and (4.33). Also for flat and selective fading with $f_dT = 0.01$ and 0.001, difference in performance between computing branch metrics from (4.34) and (4.28) or (4.33) is negligible; however, for flat fading with $f_dT = 0.1$ this difference is around 5% and the branch metric computed from (4.33) achieves better performance than the branch metric computed from (4.34) (not shown in Fig.4.2). For selective fast fading ($f_dT = 0.1$) the performance of receiver which computes the branch metrics from (4.34) is shown as curve (e) in Fig.4.5.

These results show that when the variance of estimation error is small, computing the branch metrics from (4.34) is sufficient. Moreover in this situation, by doing detection before estimation in the generalized MLSDE, where $\tilde{\mathbf{h}}_{l|l-1}$ is substituted with $\tilde{\mathbf{h}}_{l|l}$ in (4.34), the complexity of estimation is decreased since the estimation procedure needs to compute only q^L branches instead of q^{L+1} or in other words using PSP method instead of PBP method.

Although we have not considered a specific mobile communication system in this

		IS-136		GSM	
Speed		$120km/h$	$200km/h$	$120km/h$	$200km/h$
f_d	$f_c = 900MHz$	$100Hz$	$167Hz$	$100Hz$	$167Hz$
	$f_c = 1.8GHz$	$200Hz$	$334Hz$	$200Hz$	$334Hz$
f_dT	$f_c = 900MHz$	4.1×10^{-3}	6.8×10^{-3}	3.7×10^{-4}	6.2×10^{-4}
	$f_c = 1.8GHz$	8.2×10^{-3}	1.37×10^{-2}	7.4×10^{-4}	1.23×10^{-3}
Number of paths	$T_m = 1\mu\sec$	1	1	2	2
	$T_m = 10\mu\sec$	2	2	6	6

Table 4.2: Fading rate and number of paths in IS-136 and GSM mobile systems for $J = 2$.

Section, the simulation results show directions for receiver design based on current mobile systems such as IS-136 and GSM, whose fading rate and number of paths have been shown in Table 4.2, and also for future systems.

4.4 Summary

In this Chapter we derived GMLSDE which generated coupled estimation and detection procedures based on the EM algorithm. In each recursion, estimation and detection were done alternately in order to increase the likelihood function. Also PSP appears naturally in GMLSDE. Adaptive MLSDE algorithms were derived in the framework of GMLSDE in a unified way for some important channel models. In association with the channel model and the level of knowledge available at the receiver, the adaptive MLSDE contains all or some of the steps of the GMLSDE in the estimation and detection parts. Detection part of adaptive MLSDE algorithms is implemented by the Viterbi algorithm through the trellis structure. The modified Titterington's approach, stochastic approximation, was used to implement the estimation part of the adaptive MLSDE. Although Titterington's approach is generally an approximate method based on a Taylor's expansion, when the third and higher order elements of Taylor's expansion are zero (as was true for the models considered), it is exact.

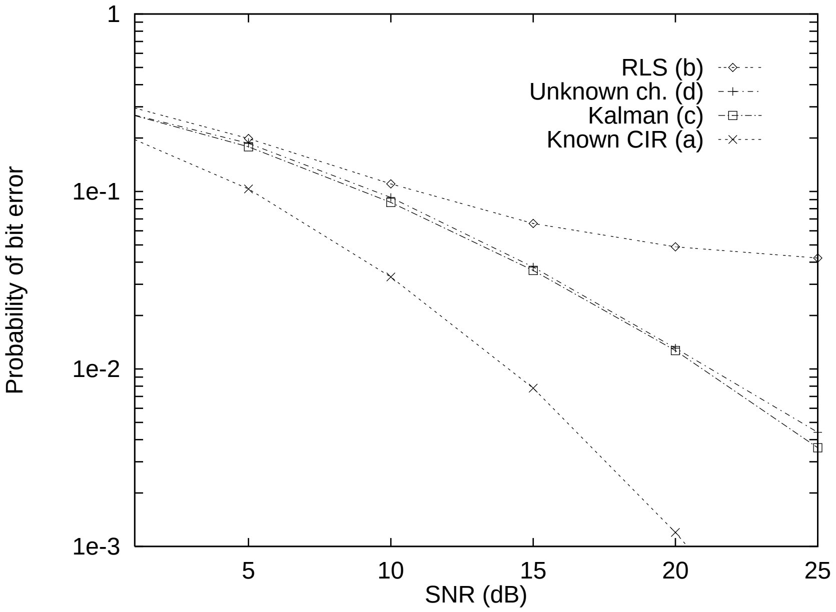

Figure 4.2: Bit error rate performance for flat fading with $f_d T = 0.1$ and DQPSK signaling. (a) CIR is known. (b) Estimation of CIR using RLS algorithm. (c) Estimation of stochastic CIR with known parameters using Kalman filtering. (d) Stochastic CIR with unknown parameters.

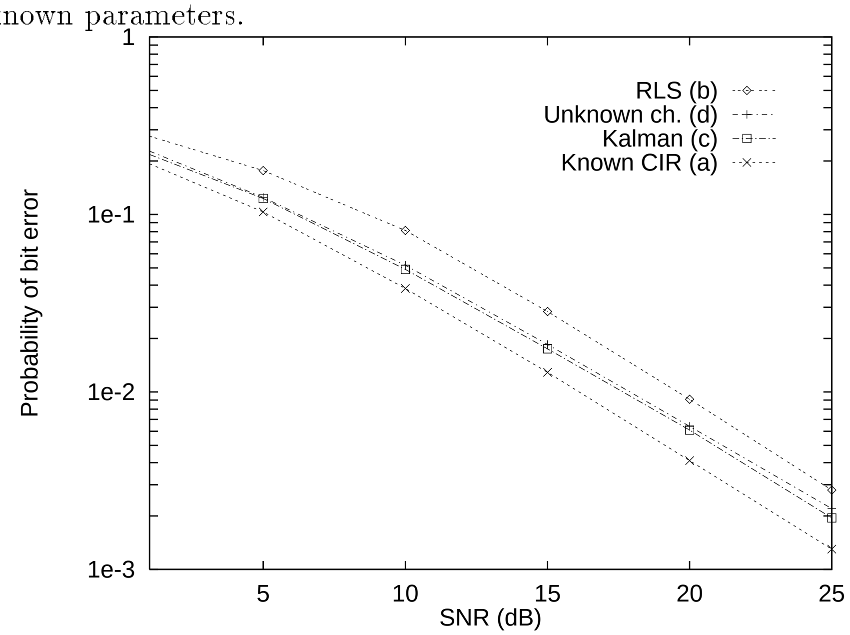

Figure 4.3: Bit error rate performance for flat fading with $f_d T = 0.01$ and DQPSK signaling. (a) CIR is known. (b) Estimation of CIR using RLS algorithm. (c) Estimation of stochastic CIR with known parameters using Kalman filtering. (d) Stochastic CIR with unknown parameters.

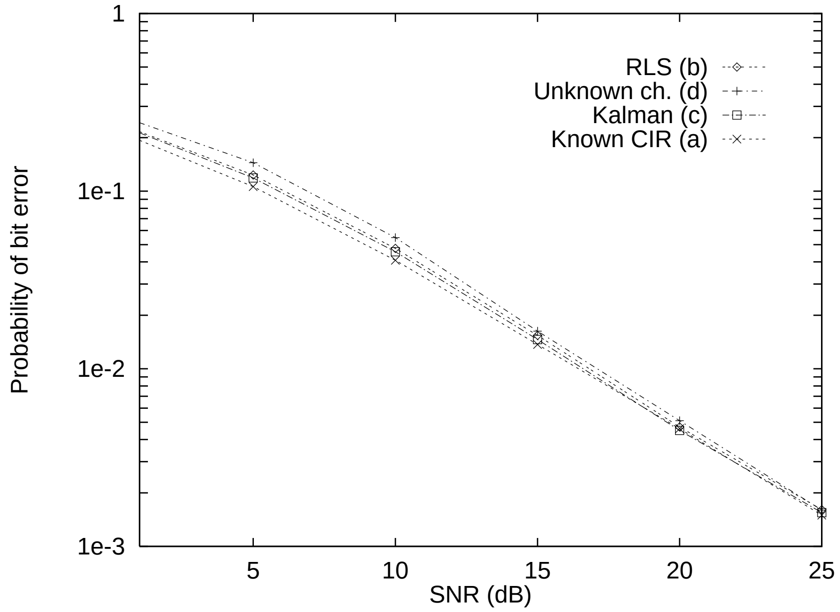

Figure 4.4: Bit error rate performance for flat fading with $f_d T = 0.001$ and DQPSK signaling. (a) CIR is known. (b) Estimation of CIR using RLS algorithm. (c) Estimation of stochastic CIR with known parameters using Kalman filtering. (d) Stochastic CIR with unknown parameters.

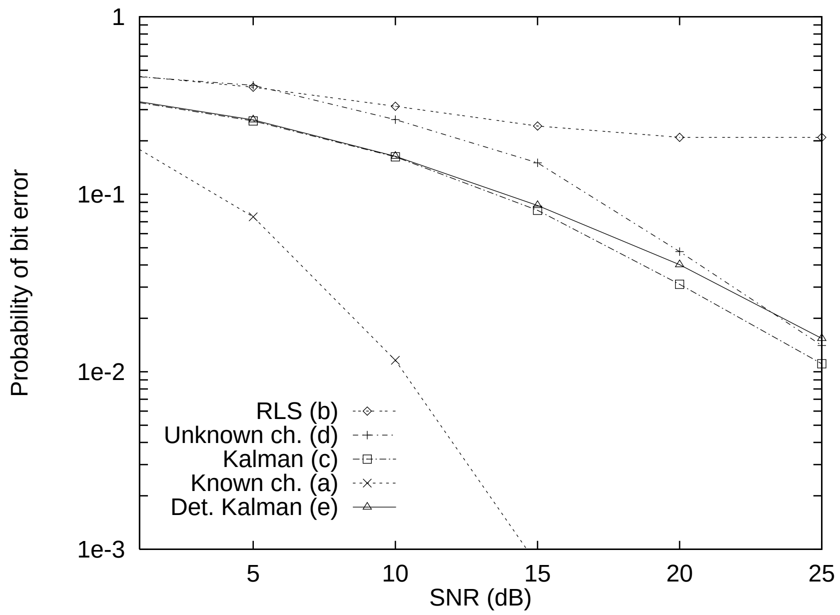

Figure 4.5: Bit error rate performance for selective fading with $f_d T = 0.1$ and DQPSK signaling. (a) CIR is known. (b) Estimation of CIR using RLS algorithm. (c) Estimation of stochastic CIR with known parameters using Kalman filtering. (d) Stochastic CIR with unknown parameters. (e) Assuming time-variant deterministic CIR and estimating CIR with Kalman type algorithm.

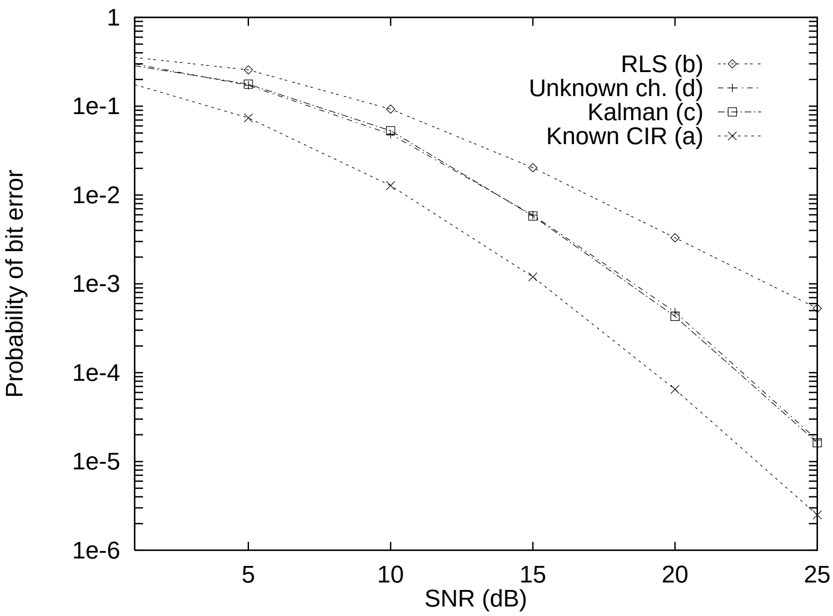

Figure 4.6: Bit error rate performance for selective fading with $f_d T = 0.01$ and DQPSK signaling. (a) CIR is known. (b) Estimation of CIR using RLS algorithm. (c) Estimation of stochastic CIR with known parameters using Kalman filtering. (d) Stochastic CIR with unknown parameters.

Adaptive MLSDE algorithm, though not achieving the maximum likelihood , but increasing the likelihood function in each recursion, was simulated for frequency flat and selective fading channels with three different fading rates based on four channel model assumptions at the receiver; known CIR, unknown deterministic CIR and Gaussian CIR with known and unknown statistical parameters. The comparison between the simulation results showed that the deterministic unknown time-invariant CIR, whose estimation leads to the RLS algorithm, achieves a performance close to known CIR in slow fading channels. However in fast or relatively fast fading channels, the deterministic unknown time-variant CIR, whose estimation can be done with Kalman filtering and RLS algorithms for obtaining impulse response and the unknown constant parameters of channels respectively, shows good performance. Only for fast selective fading channels, the use of a Gaussian process model for CIR achieves better performance. Meanwhile, by replacing smoothing instead of filtering in many of above procedures, one can derive forward-backward versions of the corresponding estimation and detection algorithms [75] [65].

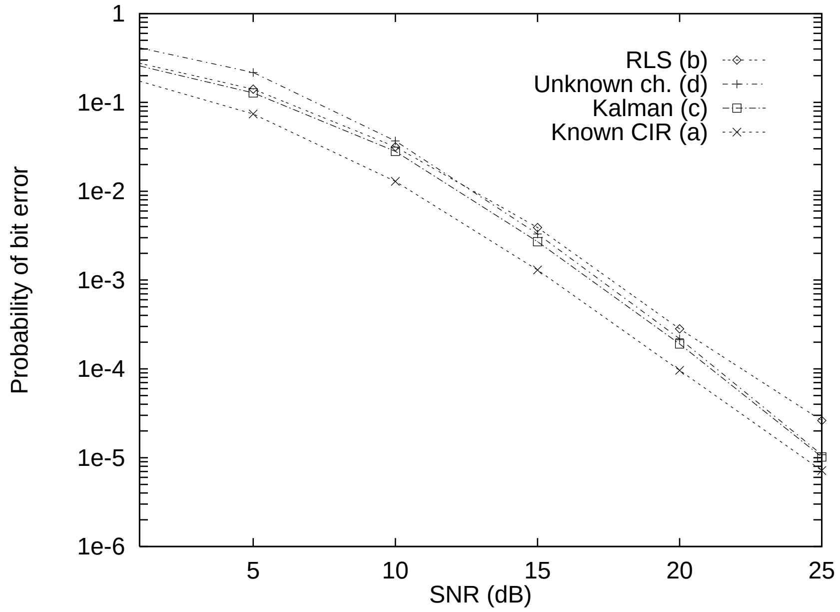

Figure 4.7: Bit error rate performance for selective fading with $f_d T = 0.001$ and DQPSK signaling. (a) CIR is known. (b) Estimation of CIR using RLS algorithm. (c) Estimation of stochastic CIR with known parameters using Kalman filtering. (d) Stochastic CIR with unknown parameters.

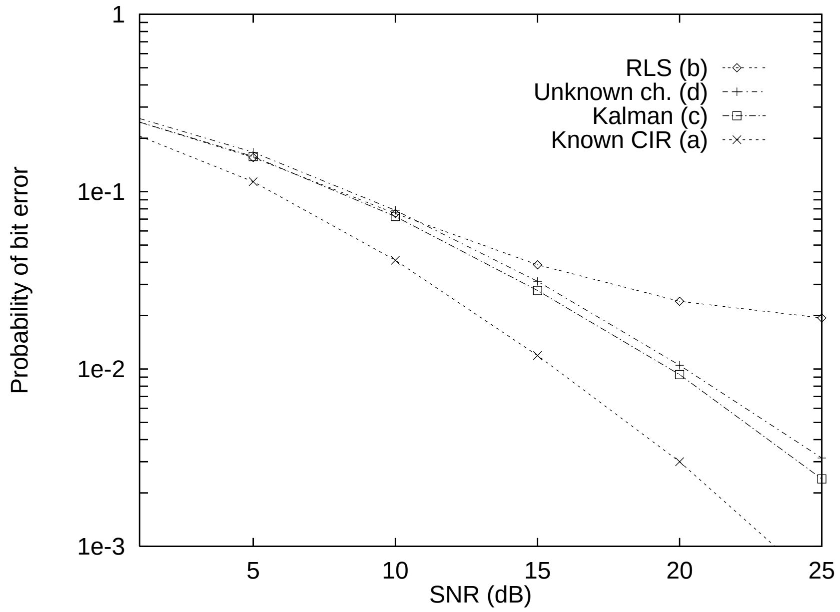

Figure 4.8: Bit error rate performance of flat fading for linear changing in the normalized fading rate between $f_d T = 0.1$ and 0.01 for DQPSK signaling. (a) CIR is known. (b) Estimation of CIR using RLS algorithm. (c) Estimation of stochastic CIR with known parameters using Kalman filtering. (d) Stochastic CIR with unknown parameters.

Chapter 5

Adaptive State Allocation Algorithm

The structure and performance of the MLSD/MLSDE receiver, as an optimal equalizer, for the frequency flat (nonselective) and selective fading channels were considered in the previous three chapters. The MLSD/MLSDE receiver minimizes sequence error rate with the assumption of equiprobable transmitted sequences. However, the major challenge with such a nonlinear receiver is its computational complexity which has motivated a considerable amount of research to find less complex algorithms or structures with near-optimal performance.

One group of efforts concentrated on using a conventional equalizer before the Viterbi algorithm in order to reduce the duration of the overall impulse response of the channel/equalizer such that the number of states in the trellis diagram decreases [34]-[41]. Qureshi and Newhall [34] used the linear equalizer to achieve a desired overall impulse response and in [15,39] the optimized desired impulse response was investigated. Lee and Hill [40] and Wesolowski [41] proposed to use decision feedback equalization (DFE) to shorten the overall impulse response.

Another group of efforts focused on reducing the computational complexity by selecting only a subset of the total number of states in the trellis diagram. There are two main strategies for selecting the subset. The first strategy selects some states which are in the trajectories of the more likely correct paths [81]-[84]. The M algorithm [42] is perhaps the most famous among these algorithms and serves as the main core for other algorithms using this strategy. At every stage in the trellis diagram, the M algorithm allows only M paths with lower costs to extend into the next stage. Trade-off between computational complexity and performance of the M algorithm is controlled by the value of M. Anderson [83] evaluated the least M in the M algorithm for decoding of convolutional codes. Simmons [18] [84] proposed to select the states whose costs were below a fixed value in the decoding of convolutional codes.

The second strategy proposed in [16]-[86] is known as reduced-state sequence estimation (RSSE). In this method, by using the Ungerboek set partition, only a fixed subset of states is selected and, in contrast to the M algorithm, the subset in the RSSE is always fixed. The delayed decision-feedback sequence estimator in [17] and the decision-feedback sequence estimator (DFSE) in [16] are special cases of RSSE. In contrast to conventional DFE, where only one detected sequence of symbols is considered as a decision sequence for detecting the new received symbol, more than one detected sequence are considered in DFSE as decision sequences. The idea of optimum partitioning, based on keeping the minimum distance error rate of the partitioned set close to that of the unpartitioned set, is considered in [87] and [88].

In this chapter we consider the computational complexity issue of the MLSD/MLSDE receiver and propose an algorithm called adaptive state allocation (ASA) algorithm which greatly reduces the computational complexity of the MLSDE receiver with a negligible degradation in the error performance. In the ASA algorithm, based on the short-term power of the CIR , an adaptive threshold for selecting only a few states of the trellis diagram is employed along with an adaptive partitioning method which fuses branch metrics to reduce the computational complexity of the MLSDE receiver. The adaptive threshold method is derived in Subsection 5.2.1 such that the states whose costs are higher than the threshold plus the minimum cost ,which is the cost of the best survivor path at each time, are removed. The threshold value is formulated based on the probability of removing the correct state or suspending the correct path in the trellis diagram. The adaptive partitioning method developed in Subsection 5.2.2 is based on the Kullback-Leibler distance between the probability density functions of the correct and the incorrect branch metrics in the trellis diagram. In the adaptive partitioning method, due to time-variant changes in the channel, it is assumed that a CIR sample whose short-term power is less than a threshold is zero for calculating the branch metrics. Although the ASA algorithm intends to reduce the computational complexity of the detection part of the MLSDE receiver, due to reduction in the number of branches for which estimation of CIR is not necessary, the computational complexity of the estimation part is also decreased. Some issues about the implementation of the ASA algorithm are considered in Section 5.3. At the end of this chapter a comparison between the performance and computational complexity of the ASA and the regular MLSDE algorithms for frequency flat and selective Rayleigh fading channels is presented through simulations. We use the term "regular" to denote the MLSDE algorithm proposed in chapter 4 with full complexity. Also when we use the term "using the ASA" or "applying the ASA" it means "using ASA along with MLSDE to reduce the number of states".

5.1 Computational Cost in the MLSD/MLSDE

In general the number of multiplications per symbol, N_m, in the MLSD/MLSDE receiver is

$$N_m = qM_bN_s \tag{5.1}$$

where q is the size of the signal set (q branches emanate from each state), M_b is the number of multiplications used in the calculation of each branch metric (for the MLSDE receiver M_b contains the number of multiplications needed to estimate the CIR as well) and N_s is the number of states in the trellis. $N_s = q^{\vartheta-1}$, where ϑ denotes the system memory which depends on the parameters of the channel and the shaping filter in the transmitter . The computational complexity can be reduced by decreasing any one of the parameters in (5.1) or a combination of them.

The trellis states at time k are denoted by $\boldsymbol{a}_k = (a_{k-(\vartheta-2)}, ..., a_{k-j}, ..., a_k)$, where the number of states depends on the size of the signal alphabet that every symbol a_{k-j} can select at time k. The maximum size is q and the maximum number of states is $q^{\vartheta-1}$. If q_{kj} is the size of the signal set that symbol a_{k-j} can select at time k, the total number of states in the trellis will be

$$N_{s_k} = \prod_{j=0}^{\vartheta-2} q_{kj} \tag{5.2}$$

Three different ways exist for reducing the number of states at time k.

i) The first way is to select a subset N_{b_k} states where $N_{b_k} \leq N_{s_k}$. This method is similar to the M algorithm [42]. However the number of selected states is not necessarily the same at each stage.

ii) The second way is to fuse or combine some branch metrics. For example, in a binary alphabet two branch metrics which correspond to $(a_{k-2}, a_{k-1}, a_k) \equiv 110$ and $(a_{k-2}, a_{k-1}, a_k) \equiv 100$ can be fused to $(a_{k-2}, a_k) \equiv 10$ when the contribution of the channel coefficient which multiplies the transmitter filter output corresponding to a_{k-1} in computing the branch metrics is very small. This is similar to the partitioning idea proposed in [16], where only a fixed or static partitioning is used for all the stages in the trellis diagram. However, in contrast, our proposed method uses adaptive partitioning and is the generalization of the method proposed in [49] where only the branch metrics with common latest symbol set are fused. We will explore the details of this method in Subsection 5.2.2.

iii) The third way is the combination of the above two methods "i" and "ii" at the same time. Not only is a subset of states chosen, but also the number of branches is reduced by fusion. This method called the ASA algorithm is suitable for time-variant channels like fading channels, where the signal-to-noise ratio (SNR) is time-variant. When the channel goes to fade the SNR is low and it is very likely that an error will occur. When the channel is out of fade the SNR is high and the error rate is very much less than the average error rate. Therefore using the same computational complexity for all SNR situations is not efficient [19, 48, 49].

The other parameter, which is important in computational complexity, is M_b in (5.1). Conventionally, the number of multiplications for calculating branch metrics is related to ϑ or more precisely to the number of symbols contributing to the computation of the branch metrics. By fusion, the size of the symbol set contributing to the branch metrics is reduced; thus the number of multiplications is reduced as well.

5.2 The ASA Algorithm

The main strategy of the ASA algorithm is to keep the performance of the ASA close to that of the regular MLSD/MLSDE with a reduction in computational complexity. It is well known that the error probability in the MLSD for ISI channels depends on d_{min} the minimum distance over all error events in the trellis structure [26]. In contrast to ISI channels, d_{min} is not constant in fading channels. When the channel goes to fade d_{min} is smaller than when the channel is out of fade. Therefore the strategy of the ASA algorithm should be designed to keep the minimum distance error event of the ASA algorithm close to that of the regular MLSDE. Based on this strategy the adaptive threshold and adaptive partitioning methods are developed in the following two subsections.

5.2.1 Adaptive Threshold Method

One method to reduce the computational complexity of the MLSD/MLSDE is to remove some states whose costs are larger than a threshold value. This idea called T-algorithm was proposed in [18] for reducing the decoding complexity of convolutional codes; however [18] did not suggest a formula for selecting the threshold value. In [18] the relation between the performance and decoding complexity was studied based on computer simulations by selecting different threshold values. We use the concept proposed in [18] to reduce the computational complexity of the MLSD/MLSDE receiver for detecting the transmitted signal over the multipath fading environment such that the performance of the reduced complexity receiver remains close to the performance of the regular one. At each time the states whose costs are larger than a threshold plus a minimum cost (the cost of the best survivor path) are removed from the trellis diagram. The strategy of choosing the threshold value is considered in the remainder of this Subsection.

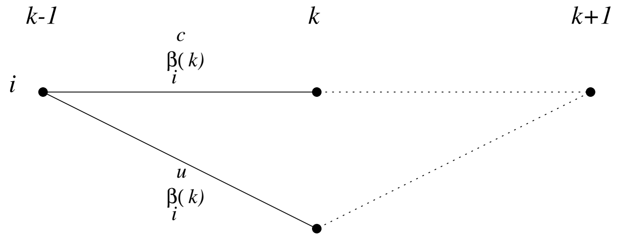

Figure 5.1: The minimum distance error event with two branches originating from ith state

Assuming that the CIR is known to the receiver, the branch metric between ith state and jth state at time k is

$$\beta_{ij}(k) = |y(k) - \boldsymbol{s}_{ij}(k)\boldsymbol{h}(k)|^2 \tag{5.3}$$

where $\boldsymbol{s}_{ij}(k)$ is a row vector of the transmitter filter output samples corresponding to transmission symbols relating the ith state to jth state. Let us assume that the ith state is the correct state at which the correct path and incorrect paths begin to diverge. We define $\beta_i^c(k)$ as the correct branch metric and $\beta_i^u(k)$ as the minimum incorrect branch metric which diverge from ith state at time k (Fig. 5.1). Since both correct and incorrect branches are diverging from the same state, the probability of removing the correct state (or suspending the correct path) is[1]

$$p_c^r = p(\beta_i^c(k) > \beta_i^u(k) + T_{h_i}(k)) \tag{5.4}$$

where $T_{h_i}(k)$, the threshold, is a non-negative value. The more likely correct states are selected based on $T_{h_i}(k)$ when the minimum distance error event has only two branches (Fig. 5.1). From (5.3) the $\beta_i^c(k)$ and $\beta_i^u(k)$ are

$$\begin{aligned} \beta_i^c(k) &= \beta_{ij}(k) = |z(k)|^2 \\ \beta_i^u(k) &= \beta_{il}(k) = |z(k)|^2 + |\Delta \boldsymbol{s}_i(k)\boldsymbol{h}(k)|^2 \end{aligned} \tag{5.5}$$

[1]We assume that the correct state is not the minimum cost state in (5.4). Therefore, computing $T_{h_i}(k)$ from (5.4) leads to less probability of removing the correct state.

$$+2(z_R(k)Re[\Delta\boldsymbol{s}_i(k)\boldsymbol{h}(k)] + z_I(k)Im[\Delta\boldsymbol{s}_i(k)\boldsymbol{h}(k)]) \tag{5.6}$$

where $z_R(k)$ and $z_I(k)$ are the real and imaginary parts of $z(k)$ respectively, $\Delta\boldsymbol{s}_i(k) = \boldsymbol{s}_i^c(k) - \boldsymbol{s}_i^u(k)$ such that $\boldsymbol{s}_i^c(k) = \boldsymbol{s}_{ij}(k)$ and $\boldsymbol{s}_i^u(k) = \boldsymbol{s}_{il}(k)$ are the transmitter filter outputs corresponding to the correct and the incorrect branch metrics and $Re[X]$ and $Im[X]$ denote real and imaginary parts of X respectively. If we assume $z(k)$ is a circularly symmetric zero-mean white complex Gaussian random process with autocorrelation $R_z(\tau) = N_0\delta(\tau)$, $z_R(k)$ and $z_I(k)$ are zero-mean Gaussian random processes which are independent of each other with $N_0/2$ variance. From (5.5) and (5.6) p_c^r becomes

$$\begin{aligned} p_c^r &= p(2(z_R(k)Re[\Delta\boldsymbol{s}_i(k)\boldsymbol{h}(k)] \\ &\quad + z_I(k)Im[\Delta\boldsymbol{s}_i(k)\boldsymbol{h}(k)]) \leq -|\Delta\boldsymbol{s}_i(k)\boldsymbol{h}(k)|^2 + T_{h_i}(k)) \\ &= \frac{1}{2}\mathrm{erfc}(\sqrt{\frac{(|\Delta\boldsymbol{s}_i(k)\boldsymbol{h}(k)|^2 + T_{h_i}(k))^2}{4N_0|\Delta\boldsymbol{s}_i(k)\boldsymbol{h}(k)|^2}}) = \frac{1}{2}\mathrm{erfc}(C_1) \end{aligned} \tag{5.7}$$

where erfc(.) denotes the complementary error function. By selecting a value for p_c^r, which should be very much smaller than one, it easy to find C_1, the argument of the complementary error function. Thus the non-negative threshold value at time k ,$T_{h_i}(k)$, becomes

$$T_{h_i}(k) = \max(2C_1\sqrt{N_0|\Delta\boldsymbol{s}_i(k)\boldsymbol{h}(k)|^2} - |\Delta\boldsymbol{s}_i(k)\boldsymbol{h}(k)|^2, 0) \tag{5.8}$$

In developing (5.8), it was assumed that the CIR, $\boldsymbol{h}(k)$, was known and $\Delta\boldsymbol{s}_i(k)$ is a deterministic vector. However only the estimate of CIR is available in adaptive MLSDE and the value of $\Delta\boldsymbol{s}_i(k)$ depends on the state and branch metric which are assumed to be the correct ones. When the estimation is reliable, one can substitute $\hat{h}(k)$, the estimate of $h(k)$, for $h(k)$ in (5.8); then the expectation of $|\Delta\boldsymbol{s}_i(k)\boldsymbol{h}(k)|^2$ can be approximated by

$$E[|\Delta\boldsymbol{s}_i(k)\boldsymbol{h}(k)|^2] \simeq E[|\Delta\boldsymbol{s}_i(k)\hat{\boldsymbol{h}}(k)|^2] = \hat{\boldsymbol{h}}^{\mathcal{H}}(k)E[\Delta\boldsymbol{s}_i^{\mathcal{H}}(k)\Delta\boldsymbol{s}_i(k)]\hat{\boldsymbol{h}}(k) \tag{5.9}$$

where $E[\Delta\boldsymbol{s}_i^{\mathcal{H}}(k)\Delta\boldsymbol{s}_i(k)]$ can easily be calculated. Therefore the average threshold for selecting the more likely correct states is approximated by

$$\bar{T}_h(k) \simeq \max(2C_1\sqrt{N_0 E[|\Delta \boldsymbol{s}_i(k)\hat{\boldsymbol{h}}(k)|^2]} - E[|\Delta \boldsymbol{s}_i(k)\hat{\boldsymbol{h}}(k)|^2], 0) \tag{5.10}$$

As mentioned before, the error probability of the MLSDE for channels with memory is dominated by the minimum distance error event in the trellis diagram. The $\bar{T}_h(k)$ value in (5.10) is obtained based on the distance of one branch between the correct path and the incorrect path. In other words it is assumed that the minimum distance error event in the trellis diagram contains only two branches. When the length of the minimum distance error event contains $\bar{M}+1$ branches , the probability of removing the correct state at mth branch of the minimum distance error event is

$$p(\bar{\varepsilon}_1, ..., \bar{\varepsilon}_{m-1}, \varepsilon_m) = p(\varepsilon_m|\bar{\varepsilon}_1, ..., \bar{\varepsilon}_{m-1})p(\bar{\varepsilon}_1, ..., \bar{\varepsilon}_{m-1}) \qquad m \leq \bar{M} \tag{5.11}$$

where

$$\begin{aligned} \varepsilon_m &= \text{removing the correct path at } m\text{th branch of the minimum distance error event} \\ \bar{\varepsilon}_m &= \text{retaining the correct path at } m\text{th branch of the minimum distance error event} \end{aligned}$$

It can be seen from Fig 5.2 that $p(\varepsilon_{\bar{m}}|\bar{\varepsilon}_1, ..., \bar{\varepsilon}_{\bar{m}-1})$, the probability of removing the correct path at $\bar{m}$th branch of the minimum distance error event given that it is retained till $\bar{m}-1$th branch, becomes

$$p(\varepsilon_{\bar{m}}|\bar{\varepsilon}_1, ..., \bar{\varepsilon}_{\bar{m}-1}) = p(\sum_{m=0}^{\bar{m}-1}\beta_i^c(k-m) > \sum_{m=0}^{\bar{m}-1}\beta_i^u(k-m) + T_{h_i}(k)) \quad \bar{m} \leq \bar{M} \tag{5.12}$$

Based on the definition of the correct and the incorrect branch metrics it is easy to show that

$$\begin{aligned} \sum_{m=0}^{\bar{m}-1}\beta_i^c(k-m) &= \sum_{m=0}^{\bar{m}-1}|z(k-m)|^2 \\ \sum_{m=0}^{\bar{m}-1}\beta_i^u(k-m) &= \sum_{m=0}^{\bar{m}-1}|z(k-m)|^2 + |\Delta \boldsymbol{s}_i(k-m)\boldsymbol{h}(k-m)|^2 \end{aligned} \tag{5.13}$$

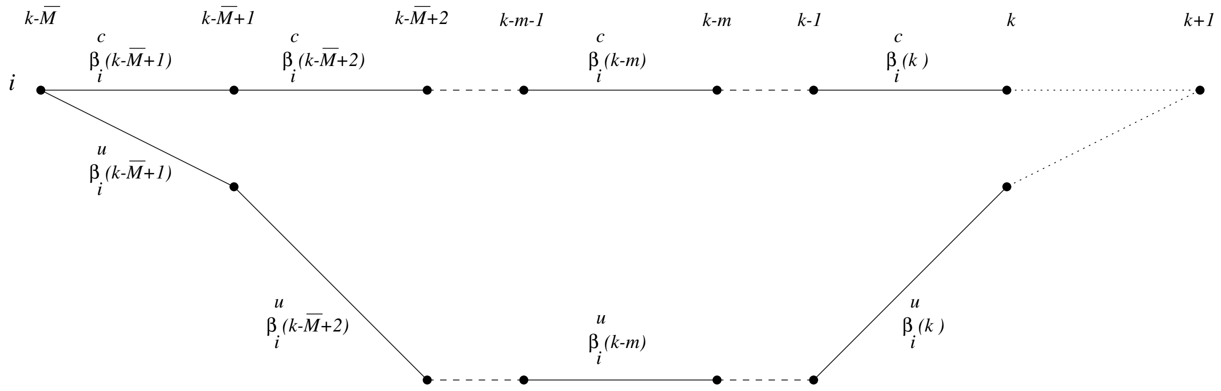

Figure 5.2: The minimum distance error event originating from ith state in the trellis diagram

$$
\begin{aligned}
&+2(z_R(k-m)Re[\Delta \boldsymbol{s}_i(k-m)\boldsymbol{h}(k-m)] \\
&+z_I(k-m)Im[\Delta \boldsymbol{s}_i(k-m)\boldsymbol{h}(k-m)]) \qquad (5.14)
\end{aligned}
$$

At each time the threshold value should be computed for all $1 \leq \bar{m} \leq \bar{M}$ and then the maximum value is used as the threshold. The average threshold value can be approximated at time k by following the same procedure which has been done for $\bar{M} = 1$ case and it becomes

$$
\begin{aligned}
\bar{T}_h(k) \;\simeq\; & \max\Big(\max_{1\leq\bar{m}\leq\bar{M}}\Big\{2C_1\sqrt{N_0\sum_{m=0}^{\bar{m}-1}E[|\Delta \boldsymbol{s}_i(k-m)\hat{\boldsymbol{h}}(k-m)|^2]} \\
& -\sum_{m=0}^{\bar{m}-1}E[|\Delta \boldsymbol{s}_i(k-m)\hat{\boldsymbol{h}}(k-m)|^2]\Big\}, 0\Big) \qquad (5.15)
\end{aligned}
$$

where similar to $\bar{M} = 1$ case, the constant C_1 is selected based on the probability of removing the correct states in the trellis diagram. Meanwhile, $p(\varepsilon_{\bar{m}}|\bar{\varepsilon}_1, ..., \bar{\varepsilon}_{\bar{m}-1}) \geq p(\bar{\varepsilon}_1, ..., \bar{\varepsilon}_{\bar{m}-1}, \varepsilon_{\bar{m}})$; therefore, since we obtain a larger value by using $p(\varepsilon_{\bar{m}}|\bar{\varepsilon}_1, ..., \bar{\varepsilon}_{\bar{m}-1})$, the value of $T_h(k)$ based on $p(\varepsilon_{\bar{m}}|\bar{\varepsilon}_1, ..., \bar{\varepsilon}_{\bar{m}-1})$ is an upper-bound. The dynamic range of $\bar{T}_h(k)$ changing according to the condition of the channel is shown in Fig 5.3 where $x = (\sum_{m=0}^{\bar{m}-1}E[|\Delta \boldsymbol{s}_i(k-m)\hat{\boldsymbol{h}}(k-m)|^2])^{1/2}$. As Fig 5.3 shows, the threshold value becomes very small in two different situations in the channel; when the channel is very good or the power of CIR is very high and when the channel is very bad or the power of CIR

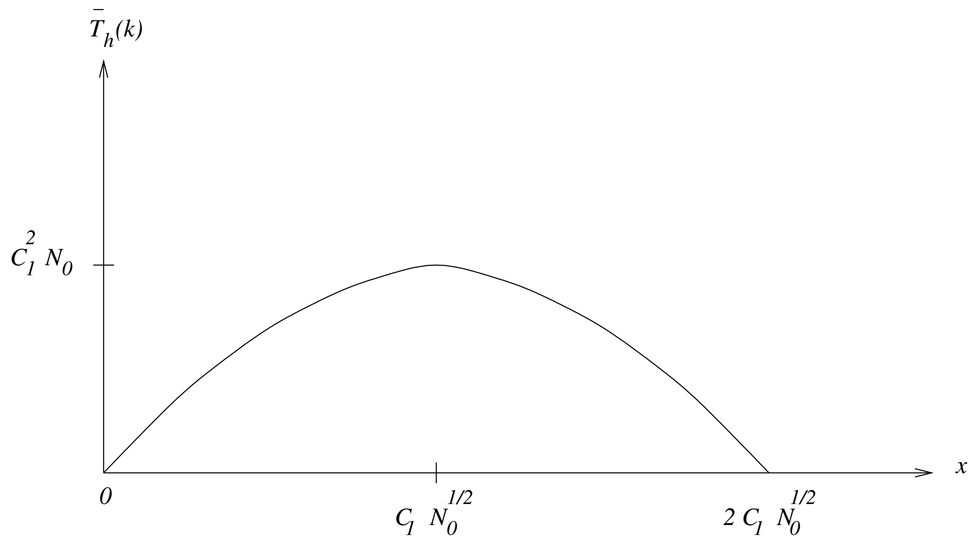

Figure 5.3: $\bar{T}_h(k)$ as a function of channel condition.

is very low in the time duration of the minimum distance error event. In other words when the channel stays in a very good or very bad condition for a while greater than the time in which a minimum distance error event can happen, it is enough to keep only one state (minimum cost) in the trellis diagram and remove the other states.

The straightforward way to avoid the calculation of $T_h(k)$ at each time is to select the maximum possible value of $T_h(k)$. From (5.15), it is easy to show that by considering the threshold as a function of x, its maximum value, $T_{h_{max}}$, becomes[2]

$$T_{h_{max}} = C_1^2 N_0 \tag{5.16}$$

The value of $T_{h_{max}}$ is not related to the CIR and the transmitted signal. It is only a function of C_1, which is found from the probability of removing the correct state, and N_0, the variance of the additive Gaussian noise. Therefore, $T_{h_{max}}$ can be used to reduce the computational complexity of the MLSDE receiver not only for multipath fading channels but also for ISI channels.

[2] $T_{h_{max}}$ is given by (5.16) even if $|\Delta \boldsymbol{s}_i(k-m)\boldsymbol{h}(k-m)|^2$ is used in (5.15) instead of the expectation of its estimate.

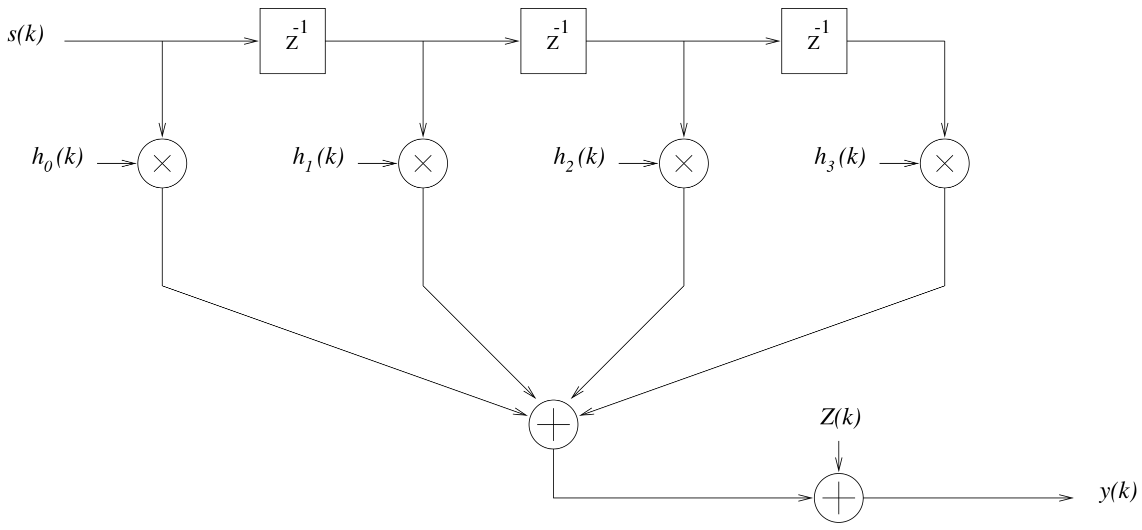

Figure 5.4: The digital channel model with $L = 3$.

5.2.2 Adaptive State Partitioning

The CIR at each sample, $h_l(k)$, is modelled as a Gaussian random process in the multipath fading channel. The number of states increases exponentially with "l" or the memory of the channel. One way to reduce the computational complexity of convolutional decoding and MLSD receiver is to use the idea of state partitioning which is to consider a subset of all the possibilities for all the symbols [16]. Based on this idea we propose fusing and diffusing of branch metrics in the trellis diagram in association with the power of $h_l(k)$ for $l = 0, ..., L$. When the power of $h_l(k)$ for a specific "l" is less than a threshold , it is assumed to be zero. Therefore some branch metrics , which are different due to non-zero value of the $h_l(k)$, become the same and they are fused. As an example Fig. 5.4 shows a discrete multipath fading channel whose impulse response duration is $L = 3$. The corresponding trellis diagram of this channel for $q = 2$ is shown in Fig. 5.5. In transition from time $k-1$ to time k, the sequence $(a_{k-3}, a_{k-2}, a_{k-1}, a_k)$ corresponding to $\boldsymbol{s}(k) = [s(k), s(k-1), s(k-2), s(k-3)]$, and $h_l(k)$ for $l = 0, ..., 3$ contribute in the calculation of each branch metric[3].

[3]For the sake of simplicity,we have not considered the effects of the transmitter filter memory and have considered one sample per symbol situation.

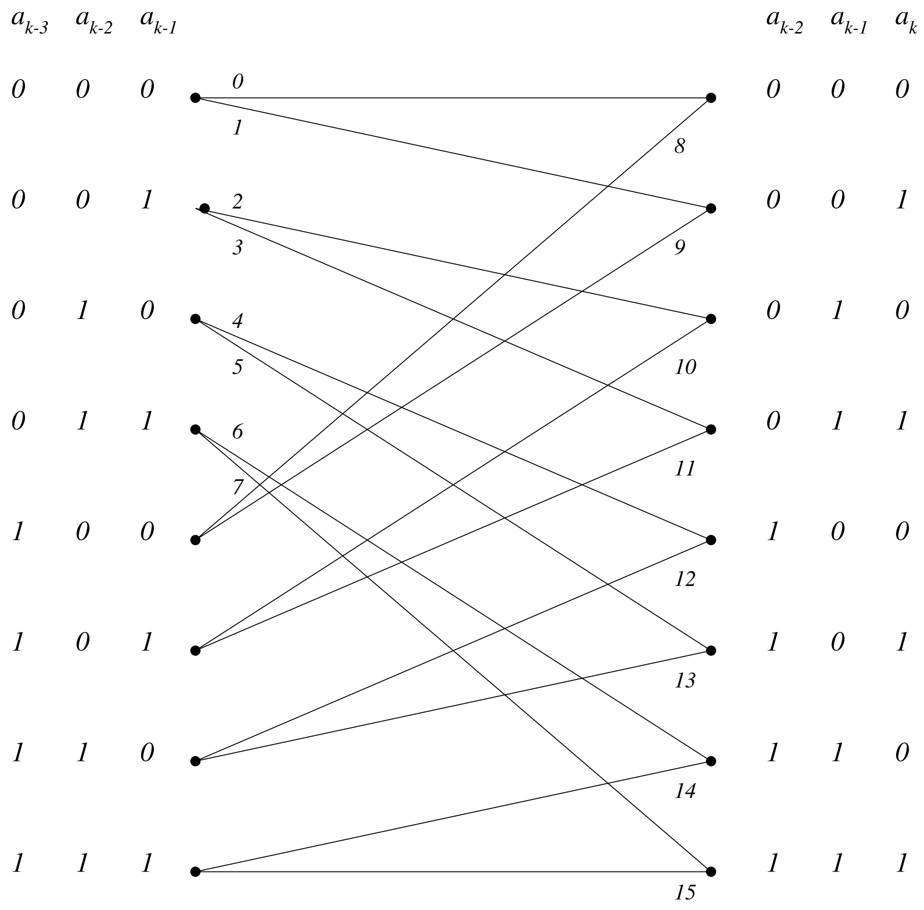

Figure 5.5: The trellis diagram for channel model shown in Fig 5.4 With $q = 2$. The branches are indicated by the numbers 0,1,...,16.

Assuming $h_l(k) = 0$ for each $0 \leq l \leq L$ reduces the number of branch metrics by a factor q and also decreases the number of multiplications in calculating each branch metric. For example, in Fig. 5.5, if it is assumed that $h_1(k) = 0$ the branch metrics (0,2), (1,3), (4,6), (5,7), (8,10), (9,11), (12,14) and (13,15) are fused and become equal[4] or if $h_3(k) = 0$, the branch metrics (0,8), (1,9), (2,10), (3,11), (4,12), (5,12), (6,14) and (7,15) become equal. Therefore all the possibilities of the symbol (or symbols), which correspond to $h_l(k) = 0$, are ignored in computing the branch metrics. Since the power of $h_l(k)$ is time-variant, the partitioning of states is also time-variant based on the situation of the channel. Hence we call the method"adaptive state partitioning" or briefly "adaptive partitioning".

The strategy of the adaptive partitioning method is to keep the error probabilities of the partitioned and the unpartitioned MLSDE receiver close to each other. The branch metric in the MLSDE receiver is considered as a random process which combines two stationary random processes $z(k)$ and $\boldsymbol{h}(k)$. Since $z(k)$ and $\boldsymbol{h}(k)$ are circularly symmetric zero-mean complex Gaussian random processes, which are independent of each other, the distribution of branch metric is a two degree chi-square random process whose density function for the correct and the incorrect branch metrics at each time k are given by

$$f_c(\beta) = \frac{1}{\sigma_c^2} e^{-\frac{\beta(k)}{\sigma_c^2}} \tag{5.17}$$

$$f_u(\beta) = \frac{1}{\sigma_u^2} e^{-\frac{\beta(k)}{\sigma_u^2}} \tag{5.18}$$

where σ_c^2 and σ_u^2 are

$$\sigma_c^2 = E[\beta^c(k)] = N_0 \tag{5.19}$$

$$\sigma_u^2 = E[\beta^u(k)] = N_0 + \Delta\boldsymbol{s}(k) E[\boldsymbol{h}(k)\boldsymbol{h}^{\mathcal{H}}(k)] \Delta\boldsymbol{s}^{\mathcal{H}}(k) \tag{5.20}$$

where $\Delta\boldsymbol{s}(k) = [\Delta s_0(k), ..., \Delta s_L(k)]$ is the transmitted signal-difference between the

[4]Branch metrics (i, j) are said to be fused when ith branch metric = jth branch metric; see Fig. 5.5.

correct and the incorrect branch metrics. From the view-point of error event probability the worst case can happen when the incorrect branch metric becomes the correct branch metric by fusion. In other words the difference symbol (symbols) between the correct and the incorrect branch metrics corresponding to the channel coefficient (coefficients), $h_l(k)$, becomes zero. We would like to find the relation between the distance of the correct and incorrect branch metrics and the power of CIR samples. Due to the randomness of the branch metrics, we consider the distance between the probability density functions of the incorrect and the correct branch metrics. There are different definitions for the distance between two probability density functions [89]. Here we use the Kullback-Leibler distance[5] between $f_u(\beta)$ and $f_c(\beta)$.

$$d(\beta^u, \beta^c) = \int \log(\frac{f_u(\beta)}{f_c(\beta)}) f_u(\beta) d\beta = \frac{\Delta\sigma^2}{\sigma_c^2} - \log(1 + \frac{\Delta\sigma^2}{\sigma_c^2}) \tag{5.21}$$

where based on the mutual independence of elements $\boldsymbol{h}(k)$ and from (5.19) and (5.20) one can show

$$\Delta\sigma^2 = \sigma_u^2 - \sigma_c^2 = \sum_{l=0}^{L} |\Delta s_l(k)|^2 E[|h_l(k)|^2] \tag{5.22}$$

In the minimum distance error event, only one symbol is different between the correct and the incorrect branch metrics or in other words only one element of $\Delta\boldsymbol{s}(k)$ is not zero in each branch. If we assume that the lth element of $\Delta\boldsymbol{s}(k)$, $\Delta s_l(k)$, is not zero and use the average value of $|\Delta s_l(k)|^2$ and since $\Delta s_l(k)$ will be different for different symbols selected from q possibilities, we have

$$\Delta\sigma_l^2 = \overline{\Delta\sigma^2} = \overline{|\Delta s_l(k)|^2} E[|h_l(k)|^2] \tag{5.23}$$

Therefore the Kullback-Leibler distance for lth error in the minimum distance error event becomes

$$d^l(\beta^u, \beta^c) = \frac{\Delta\sigma_l^2}{\sigma_c^2} - \log(1 + \frac{\Delta\sigma_l^2}{\sigma_c^2}) \tag{5.24}$$

$\bar{d}(\beta^u, \beta^c)$, the average Kullback-Leibler distance between the correct and the incorrect

[5]Even though the Kullback-Leibler measure does not satisfy all the properties of a proper distance, it has nevertheless been used widely in many similar problems (see, e.g., [61, 62]).

branch metrics in the minimum distance error event, is given by

$$\bar{d}(\beta^u, \beta^c) = \frac{1}{L+1} \sum_{l=0}^{L} d^l(\beta^u, \beta^c) \tag{5.25}$$

$\bar{d}(\beta^u, \beta^c)$ is calculated based on $E[|h_l(k)|^2]$ which is the long-term power of $h_l(k)$. In the MLSDE receiver where the CIR is estimated as $\hat{h}_l(k)$, the short-term power of $h_l(k)$ can be approximated with $|\hat{h}_l(k)|^2$ at time k. Based on the short-term power of $\hat{h}_l(k)$, the Kullback-Leibler distance for lth branch between the correct and the incorrect branch metric in the minimum distance error event is defined by

$$d_k^l(\beta^u, \beta^c) = \frac{\overline{|\Delta s_l(k)|^2}|\hat{h}_l(k)|^2}{N_0} - \log(1 + \frac{\overline{|\Delta s_l(k)|^2}|\hat{h}_l(k)|^2}{N_0}) \tag{5.26}$$

In the adaptive partitioning method, it is assumed that $h_l(k)$ is zero for calculating the branch metrics or the branch metrics are fused when $d_k^l(\beta^u, \beta^c)$ is very much smaller than $\bar{d}(\beta^u, \beta^c)$.

$$d_k^l(\beta^u, \beta^c) \leq C_2 \bar{d}(\beta^u, \beta^c) \tag{5.27}$$

The coefficient C_2 is a constant whose value indicates a trade-off between the performance and the computational complexity of the receiver. In order to keep the performances of the partitioned and unpartitioned receiver close, C_2 should be very much smaller than one. By calculating $\bar{d}(\beta^u, \beta^c)$ and doing some manipulations, one can find from (5.27) a threshold for the short-term power of $\hat{h}_l(k)$ to fuse the branch metrics (see Section 5.4).

5.3 Implementation

The number of states and branch metrics are time-variant in the ASA algorithm. Thus the computational complexity of ASA algorithm is also time-variant and is much lower than that of the regular MLSDE only in the mean sense. From the view-point of real time implementation of the ASA algorithm, the software or hardware should be able to handle the ASA algorithm with the highest possible complexity which may reach the complexity of the regular MLSDE receiver in a special situation. It means that the full complexity of software or hardware is not used in the ASA algorithm and this points to a weakness in this algorithm.

For making the computational complexity of software or hardware constant, we propose to use a buffering shift register in the receiver after the sampler (Fig. 2.2), where the samples of the received signal are held in the shift register and then are used by the digital processor. Although the input rate to the shift register is $\frac{1}{T_s}$ and is constant, the output rate from the shift register is time-variant and depends on the situation of the channel. Therefore, from the view-point of software or hardware, the computational complexity is constant and only the rate of the signal received by the digital processor is time-variant. In this situation, the software or hardware can be designed to handle the mean value of the time-variant computational complexity of the ASA algorithm.

It is clear that the length of the queue in the shift-register is time-variant. The shift-register length should be selected based on the statistical characteristics of the queue length in the shift-register which depends on the channel coherence time, channel multipath spread and sampling rate. The design of the shift-register is beyond the scope of our goal in this thesis.

The other point in the ASA algorithm is the extra computational complexity for calculating the threshold value and the short-term power of the estimated CIR samples in adaptive threshold and adaptive partitioning methods respectively. Since the maximum Doppler shift is usually very much smaller than the transmitted signal bandwidth, the CIR variation is very much slower than the sampling rate of the received signal. Therefore, it is not necessary to up-date the threshold value and short-term power of the CIR at each time. Based on the fading rate of channel, one can select a frame of the received signal in whose duration the channel variation is very small and select a constant threshold value for fusing/diffusing the branch metrics at each frame. In this way the extra computational complexity in the implementation of the ASA algorithm is negligible especially when the PSP method [8] is used in the MLSDE receiver. Meanwhile for slow fading, the short-term power of the CIR can be estimated in association with a window of the estimated CIR to reduce the effects of estimation errors.

5.4 Computer Simulations

To evaluate the bit error rate performance and the computational complexity of the ASA algorithm, computer simulations have been done based on the selected channel model in Chapter 4 for flat and frequency selective fading channel with fading rate $f_d T = 0.01$. Also, the modulation scheme, the characteristic of the transmitter filter and sampling rate are similar to the ones in Chapter 4. We apply the ASA algorithm in the regular MLSDE receiver when Kalman estimation with known channel parameters is used for CIR estimation and compare the bit error rate performance and the computational complexity of the ASA algorithm with this MLSDE type receiver.

Fig. 5.6 shows the bit error rate performance of the regular MLSDE and the ASA algorithm when only the adaptive threshold method is used for the flat fading channel. The adaptive threshold value is calculated from (5.15) and C_1 is chosen such that the probability of removing the correct state becomes 0.001. Meanwhile, due to fading rate, the threshold value is renewed only after every ten symbols and it remains constant for the duration of ten symbols. As seen in Fig. 5.6, the difference in performance between the regular MLSDE and the ASA algorithms is very small and negligible. However,as seen in Fig. 5.7, the computational complexity (number of multiplications) of the ASA algorithm is about 75% less than that of the regular MLSDE algorithm at SNR=25dB. In the adaptive threshold method the minimum computational complexity reduction is achieved when only one state is selected at each time in the trellis diagram. Since in the flat fading channel model the number of total states is $N_s = 4$, the maximum computational complexity reduction is 75%. As Fig. 5.7 shows the proposed adaptive threshold method is very close to this level of reduction at SNR=25dB.

The error rate performances of the regular MLSDE and the ASA algorithms for either adaptive threshold (ASA-AT) or adaptive partitioning (ASA-AP) are shown in Fig. 5.8 for frequency selective fading channel with three paths, $L = 2$. In the adaptive threshold method, due to less dynamic changes in CIR power in the frequency selective channel and ease in obtaining the threshold value, the maximum threshold value which is constant is calculated from (5.16) and C_1 is chosen such that the probability of removing the correct state becomes 0.001. In the adaptive partitioning method, since the sensitivity of receiver performance to the difference between the correct and incorrect branch metrics is increased by increasing SNR, we experimentally chose $C_2 = 0.1 \times 10^{-\frac{SNR}{10}}$ (C_2 is approximately proportional to the error probability, $C_2 \approx 0.1p(\varepsilon)$) in (5.27), where $C_2 = 0.1$ and 0.001 for SNR=$0dB$ and SNR=$20dB$ respectively. As seen in Fig. 5.8, the error rate performances of ASA-AT and ASA-AP are close to that of the regular MLSDE receiver.

The computational complexities of the regular MLSDE, ASA-AT and ASA-AP algorithms are shown in Fig. 5.9. The complexity of the ASA-AT is about 93.5% less than that of the regular MLSDE at SNR=25 dB and it is increased by decreasing SNR. ASA-AT achieves approximately the maximum complexity reduction which is 93.75% or selecting only one state from among sixteen states of the trellis diagram, $N_s = 16$. In contrast to ASA-AT, the complexity of ASA-AP is decreased at low SNR and for high SNR it is close to the complexity of the regular MLSDE. This phenomenon can be explained in the extreme case when the noise power is zero or SNR=∞. In this situation the correct branch metric is always zero and we should consider all the CIR samples for calculating the branch metrics in order to select the correct path. Therefore the computational complexity of the ASA-AP becomes the same as the computational complexity of the regular MLSDE at high SNR.

Fig. 5.10 shows the error rate performance of the ASA algorithm when both adaptive threshold and adaptive partitioning methods are used. The degradation in the performance of the ASA algorithm is seen to be negligible in comparison with that of the regular MLSDE. Due to the application of both adaptive threshold and adaptive partitioning methods the computational complexity of the ASA algorithm shown in Fig. 5.11 is very much lower than that of the regular MLSDE at high and low SNR. From Fig. 5.9, one can conclude that for having low computational complexity only in high or low SNR using the adaptive threshold method or adaptive partitioning method is enough.

5.5 Summary

We have described a new adaptive algorithm called the ASA algorithm which greatly reduces the computational complexity of the regular MLSDE receiver based on measuring the short-term power of the estimated CIR. In this algorithm, only a few states of the trellis diagram are chosen as the more likely correct states with adaptive threshold and adaptive state partitioning methods. Although the ASA algorithm is suboptimal, simulation results show that its performance is very close to that of the regular MLSDE.

In the adaptive threshold method which decreases the computational complexity of the regular MLSDE at high SNR the threshold value was formulated based on the probability of removing the correct state. The threshold value changes according to the short-term power of the estimated CIR. The maximum value of the threshold which depends on CIR and transmitted signal was proposed and can be used for ISI channels as well. The adaptive state partitioning method which decreases the computational complexity of the regular MLSDE at low SNR was developed based on the Kullback-Leibler distance between the pdf's of the correct and incorrect branch metrics. Therefore using one of theses methods is enough when the receiver is designed only for high or low SNR. However if the probability of removing the correct state is selected to be larger for low SNR the complexity reduction of the adaptive threshold method will be more at low SNR at the expense of a decrease in performance.

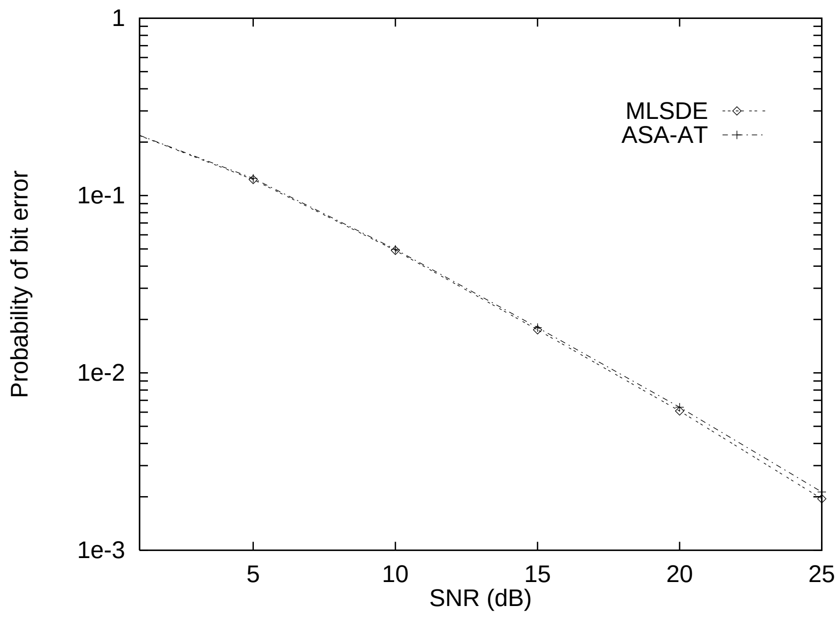

Figure 5.6: Bit error rate performance of the regular MLSDE and the ASA algorithms for flat fading channel with $f_d T = 0.01$ and DQPSK signaling. The adaptive threshold method is used in the ASA algorithm.

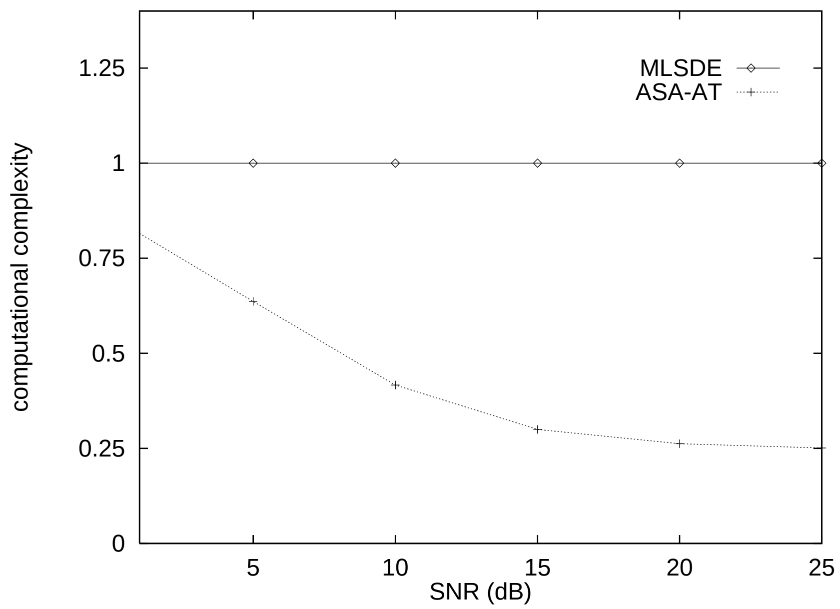

Figure 5.7: Computational complexity of the regular MLSDE and the ASA algorithm for flat fading channel with $f_d T = 0.01$ and DQPSK signaling. The computational complexity of algorithms is normalized to that of the regular MLSDE. The adaptive threshold method is used in the ASA algorithm.

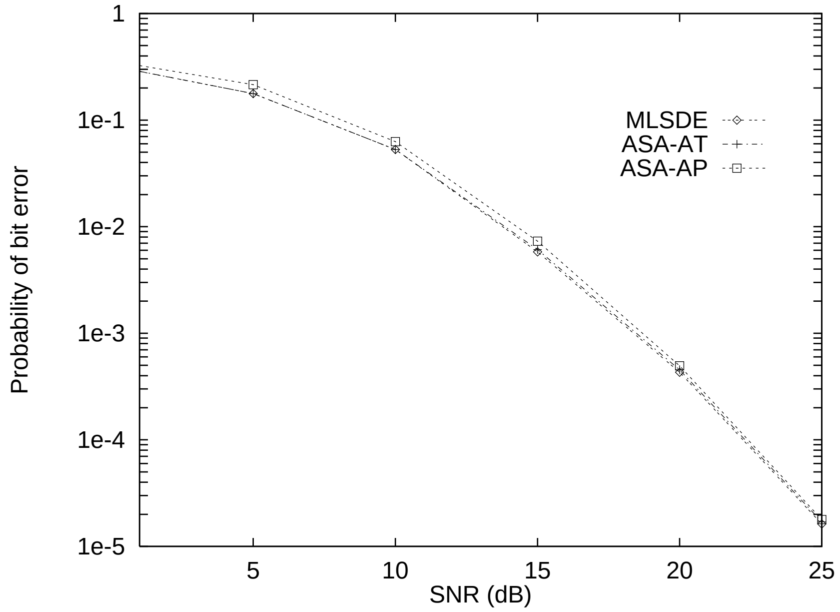

Figure 5.8: Bit error rate performance of the regular MLSDE and the ASA algorithms for frequency selective fading channel with $f_dT = 0.01$ and DQPSK signaling. The adaptive threshold and adaptive state partitioning methods are used in ASA-AT and ASA-AP respectively.

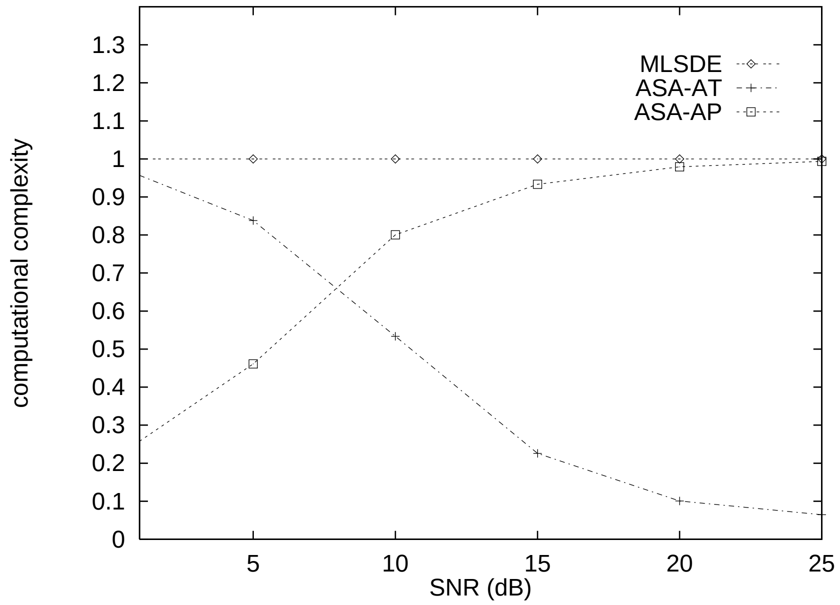

Figure 5.9: Computational complexity of the regular MLSDE and the ASA algorithms for frequency selective fading channel with $f_dT = 0.01$ and DQPSK signaling. The computational complexity of algorithms is normalized to that of the regular MLSDE. The adaptive threshold and adaptive state partitioning methods are used in ASA-AT and ASA-AP respectively.

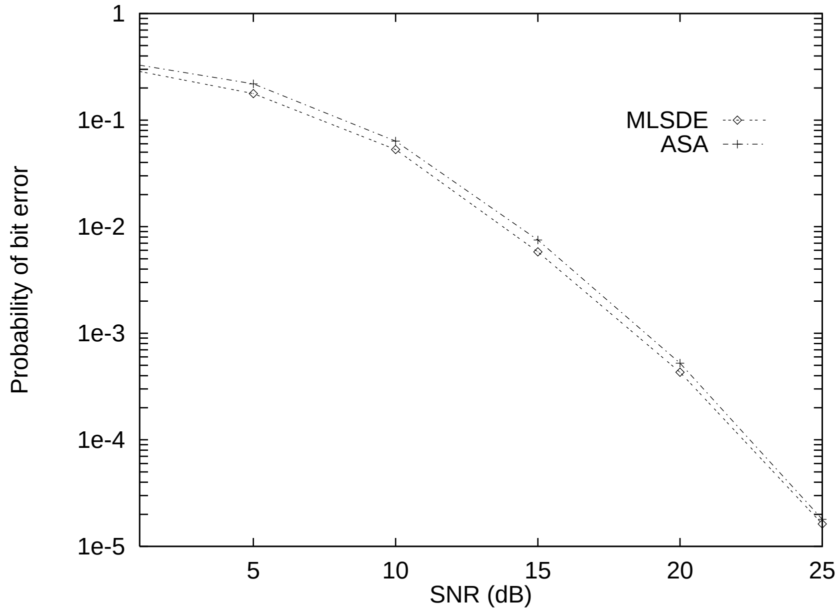

Figure 5.10: Bit error rate performance of the regular MLSDE and the ASA algorithms for frequency selective fading channel with $f_d T = 0.01$ and DQPSK signaling. Both adaptive threshold and adaptive state partitioning methods are used in the ASA algorithm.

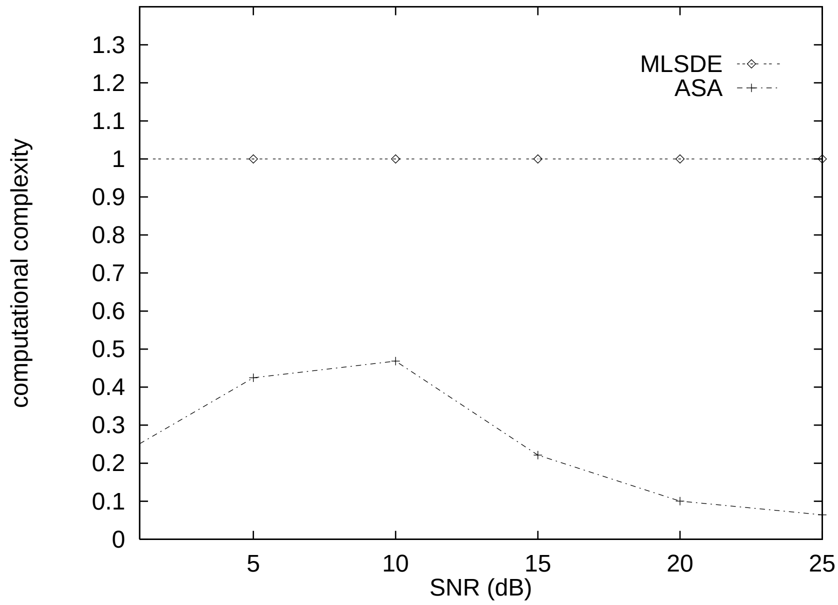

Figure 5.11: Computational complexity of the regular MLSDE and the ASA algorithms for frequency selective fading channel with $f_d T = 0.01$ and DQPSK signaling. The computational complexity of algorithms is normalized to that of the regular MLSDE. Both adaptive threshold and adaptive state partitioning methods are used in the ASA algorithm.

Chapter 6

Summary, Conclusions and Future Research

This thesis was devoted to the design of optimal or near-optimal receiver for bandlimited signaling transmitted over multipath fading channels. Structure, performance and complexity of the adaptive maximum likelihood sequence detection and estimation (MLSD/MLSDE) were investigated for different frequency flat and selective fading channels under different levels of channel knowledge which are available at the receiver. A detailed summary and conclusions of this thesis are given at the following section. This thesis will conclude with some suggestions for future research directions in Section 6.2.

6.1 Summary and Conclusions

In Chapter 1, major challenging issues in the design and development of mobile communication systems such as spectral resources, power consumption, reliable communications and complexity were pointed out. A brief survey of the MLSD receiver as an optimal receiver and a good candidate for detection of a bandlimited signal transmitted over multipath fading channels was presented.

The fist major motivation for this thesis, which was expressed in Chapter 1, arose from the need to develop a solid theory for joint channel estimation and data detection based on ML criterion . Also another motivation was to generalize the MLSDE algorithm so that different MLSDE receivers developed in literature can be derived in a unified way and this may open new windows in the area of receiver design. Computational complexity of the MLSD/MLSDE receivers is a challenging issue in implementing these receivers for detecting a signal transmitted over multipath fading channel. Reducing the complexity of the MLSDE receiver was another major motivation for this thesis.

The EM algorithm is an iterative method which can satisfy the ML criterion when the complete data is not available. This algorithm uses decision feedback inherently and is able to couple the estimation and detection parts adaptively in a ML-based receiver. The abilities of the EM algorithm motivated us to use it in joint channel estimation and data detection. The rest of Chapter 1 gave an overview and summarized the major contributions of the thesis.

A discrete finite impulse response (FIR) linear system model was developed for Rayleigh multipath fading channels in Chapter 2. The structures of the MLSD receivers were derived using ML criterion based on known channel statistical parameters (CSP) and channel impulse response (CIR). The assumption of known CSP, which is the autocorrelation function of CIR for a Rayleigh fading channel, leads to the MLSD receiver structure implemented by a bank of FIR perdition filters with constant coefficients and the Viterbi algorithm. The order of prediction filters was related to the channel memory length and the coefficients of each prediction filter were computed based on the statistical parameters of the channel and hypothesis sequence assumption for each branch in the trellis diagram. The MLSD structure for known CIR was implemented by a time-variant FIR filter similar to the one proposed by Forney [27].

The simulation results in Chapter 2 show that the performance of the MLSD receiver for known CIR is superior to that of the MLSD with known channel autocorrelation function. In practice, neither channel autocorrelation function nor CIR is known and should be estimated. However due to the complexity of estimating the autocorrelation function of channel directly from the autocorrelation function of received signal without estimating CIR [37] and also due to the superior performance of the MLSD receiver for known CIR, the chapter concluded with the recommendation to use CIR estimation method in joint channel estimation and data detection receiver. Also, the Chapter has highlighted the view that the CIR estimation method provides the opportunity to consider fading process either as a random process or as an unknown deterministic process.

Chapter 3 focused on the issue of channel estimation within the framework of the EM algorithm. Estimating channel parameters using the EM algorithm, which increases the likelihood function iteratively, is a suitable method of presenting the problem in a general framework which can consider different channel models and levels of available knowledge at the receiver. In this Chapter, the EM algorithm was introduced and it was shown how instead of maximizing the likelihood function directly, which sometimes is not feasible due to the complete data being not available, the expectation of another likelihood function related to the complete data, which is tractable, is maximized iteratively to reach the maximum of the original likelihood function.

The recursive channel estimation was crystalized based on the on-line EM algorithm which is suitable for real-time processing in communication systems and also Titterington's approach of stochastic approximation was modified in order to estimate time-variant complex variables adaptively. Chapter 3 continued by modeling the CIR with deterministic time-invariant/variant parameters and a Gaussian random vector/process. A novel RLS/Kalman-type algorithm was developed to estimate a deterministic time-invariant/variant CIR corrupted by an additive colored Gaussian noise which was modeled by an autoregressive process. The stochastic RLS and Kalman algorithms emerged from the maximization part of the EM algorithm using modified Titterington's approach when CIR was considered as a Gaussian vector and process respectively. Also a combination of RLS and Kalman algorithms was derived to estimate two separate sets of unknown channel parameters adaptively based on the principle of increasing the likelihood. The principle of increasing the likelihood of divided unknown parameter set which was used in the EM framework, provided the foundation for joint channel estimation and data detection methods in chapter 4.

Developing adaptive likelihood sequence detection and estimation (MLSDE) was the main intention of Chapter 4. The idea of using the EM algorithm to estimate separate parameter sets was extended in this Chapter such that the generalized MLSDE (GMLSDE) was developed which alternated between estimation and detection along with a facility to increase the likelihood function. The GMLSDE guarantees inherent coupling between channel estimation and data detection procedures. It has been shown in Chapter 4 how the per-survivor processing (PSP) which was considered more as a practical way of channel estimation in the MLSDE [8], emerges naturally from the ML-based GMLSDE and how the traditional channel estimation method, which was employed in sequence detection using the Viterbi algorithm [34], arises when CIR is a very slow process. Numerous adaptive MLSDE receivers were explored from the GMLSDE algorithm; these receivers may not guarantee reaching the maximum of likelihood due to using casual estimation and detection procedures, but they satisfy the principle of increasing likelihood. According to the channel model and the available knowledge at the receiver, the adaptive MLSDE accommodated all or some steps of the GMLSDE in estimation and detection parts. Although only some structures of adaptive receivers derived in this Chapter are new and others have been derived in the literature previously, deriving all receiver structures from GMLSDE in a unified way shows the powerful ability of the GMLSDE in unifying different ML-based joint channel estimation and data detection receivers. Also deriving adaptive MLSDE from GMLSDE clarifies the relation between the receiver structure and the channel model.

The performance of the adaptive MLSDE receiver for DQPSK modulated signal transmitted over frequency flat and selective Rayleigh fading channels corrupted with additive white Gaussian noise were evaluated by computer simulations in Chapter 4. The performances of different adaptive MLSDE receivers, which were developed based on different channel model assumptions and level of available knowledge, were compared with each other such that the simulated channel and the received signal for all receivers were the same.

The simulation results show that for very slowly varying environment, $f_dT \leq 0.001$, the performance of the adaptive MLSDE receiver in which CIR was modeled by an unknown deterministic vector is close enough to the performance of the MLSDE receiver with known CIR (complete knowledge about the channel). When the channel is not very fast ($f_dT \leq 0.01$) and its impulse response is simulated by a Gaussian process to generate the received signal, the adaptive MLSDE receiver with a time-variant unknown deterministic CIR achieves the same performance as that of the adaptive MLSDE receiver with the CIR modeled by a Gaussian random process which is the real model. For very fast fading rate ($f_dT > 0.1$) and frequency selective channels, the adaptive MLSDE receiver in which CIR is modeled by a Gaussian random process shows superior performance compared with the adaptive MLSDE receiver with unknown deterministic CIR model. Generally, the more the received signal is dispersed in time and frequency domains by the channel, the more the gap that exists between the performances of the receivers using the real and approximated channel models.

The methods to reduce the computational complexity of the adaptive MLSDE receiver which is the major limitation in implementing this receiver was investigated in Chapter 5. Generally, in fading channels due to time-varying SNR nature, considering a detection algorithm with the same computational complexity for different values of SNR to achieve a fixed performance is wasteful. When the power of the received signal decreases due to fading, the receiver should use a more complicated detection algorithm in order to achieve a performance near to that of the receiver when the power of the received signal is high or the channel is out of fade. In Chapter 5, the adaptive state allocation (ASA) algorithm was developed such that the number of states allocated to the trellis diagram and the memory length of the channel in computing the branch metrics are time-variant. Due to variations of the received signal power, the computational complexity of ASA algorithm is time-variant.

Chapter 5 derived two computational complexity reduction methods, namely adaptive threshold and adaptive state partitioning, embedded in the ASA algorithm. A threshold value in the adaptive threshold method was formulated based on the short-term power of the CIR such that a few states, whose costs are less than minimum cost plus the threshold value, are selected as the more likely correct states at each time. The threshold value was chosen in association with some approximations such that the probability of loss of the correct path in the trellis diagram becomes less than a selected value. Although the threshold value is time-variant for multipath fading channels, a maximum value of the threshold was found in order to reduce the computational complexity of the MLSD and MLSDE receivers for both multipath fading and ISI channels.

Computer simulations show that the adaptive threshold greatly reduces the computational complexity of the adaptive MLSDE for frequency flat and selective fading channels at high SNR with a negligible loss in the performance. For example, the minimum cost state plus a fraction of one state were chosen on the average at each time as more likely states in this method at SNR=25 dB such that it caused about 75% and 93.5% reduction of computational complexity for frequency flat fading channel realized by four states and frequency selective fading channel implemented with sixteen states at SNR=25 dB respectively. Moreover, the performances of reduced complexity receives are very close to the performances of the regular ones. However, this method does not work well at low SNR such that, as has been shown by simulation, the computational complexity of the adaptive MLSDE using adaptive threshold becomes close to that of the regular adaptive MLSDE at low SNR.

Adaptive state partitioning method was developed in Chapter 5 to decrease the computational complexity of the adaptive MLSDE by fusing the branch metrics and also by decreasing the computation of the branch metrics. For a time-variant CIR, the contributions of some estimated CIR samples in computing the branch metrics are negligible in comparison with the role of the other samples and additive Gaussian noise. In this situation the CIR samples whose contributions in computing branch metrics are insignificant can be assumed to be zero so that it decreases the channel memory length leading to lower complexity in computing the branch metrics and to fusing some branch metrics which have different values when these samples are not zero. When the Kullback-Leibler distance between the incorrect and correct branch metrics in the minimum error event becomes very much smaller than average value, the CIR samples which may make a difference between the correct and incorrect branch metrics become zero. As simulation results for selective fading show in Chapter 5, this method reduces the computational complexity at low SNR.

The ASA algorithm, which contains both reduction complexity methods, drastically reduces the computational complexity of the selective frequency fading channel at high and low SNR while the degradation in performance is insignificant, according to the simulation results in Chapter 5. Although the ASA algorithm was mainly developed to reduce the complexity of the detection part of the adaptive MLSDE, it reduces the complexity of estimation part as well since it removes some paths for which channel estimations are not needed.

6.2 Future Research Directions

The following items are among a number of directions in which the outcomes of this thesis can be extended.

1. The GMLSDE receiver was developed in Chapter 4 without specifying the channel model. However, we mainly focused on deterministic and Rayleigh fading channels in this thesis. Considering other channel models such as Ricean and non-Gaussian channels is a direction in which the results of the GMLSDE can be extended.

2. Using Theorem 1 along with RLS/Kalman-type algorithm developed in Chapter 3, one can extend the results of Chapter 3 for estimating unknown parameters embedded in colored Gaussian noise with unknown statistics. Some efforts has been done in [90]; however, more work is needed especially in the processing of up-sampled signals with more bandwidth efficient signaling schemes such as 16-QAM and 64-QAM.

3. Throughout the thesis we always assumed that perfect symbol synchronization was available. However symbol or frame synchronization is an important part of the receiver and impairment of that has a great impact on the performance of the detection and estimation algorithm. Development of the synchronization scheme using the EM algorithm and considering its effects on the performance of the adaptive MLSDE is another possible direction.

4. The results of simulations in Chapter 4 show that the performance of the adaptive MLSDE receiver for data detection is very poor in fast multipath fading channels.

Using coding is an effective way to improve the receiver performance. Joint estimating, detecting and decoding along with using the ASA algorithm introduced in Chapter 5 can be another way to enhance the achievements of this thesis.

5. We limited ourself to causal estimating and detecting procedures in developing adaptive MLSDE receivers from the GMLSDE in Chapter 4. However, by dividing the sequence data into sequence frames with known headers and using the procedure of the GMLSDE, one can develop non-casual estimation and detection in order to achieve or get closer to the ML goal. Meanwhile this method also provides a proper foundation for using turbo coding schemes along with the ASA algorithm to reduce the computational complexity.

Appendix A: Multipath Fading Channels

In wireless mobile communication systems, the transmitted signal reaches the receiver over multiple propagation paths which causes intersymbol interference (ISI). The signal is reflected and scattered by cluttered environment with buildings, vehicles and trees. The effect can cause fluctuations in the amplitude, phase and arrival angle of the received signal. Thus due to random fluctuations, the signal arrives at the receiver with time-variant attenuation, delay and frequency shift in each path and it may be destructive or constructive for the received signal. Time-variant phenomenon in multipath channels causes signal fading at the receiver. Wireless mobile communication channel which has both phenomena, ISI and a time-variant nature, is categorized as a multipath fading channel.

Two type of fading affects communication over mobile environment; large-scale and small-scale. Large scale fading refers to random fluctuations over large area and it is modeled by a log-normal distribution [91]. The mean path loss in large-scale fading is a function of distance between receiver and transmitter and variation is about 6-10 dB. However small-scale fading represents the time-variant changes in the received signal due to small changes (as small as a half-wave length) in the transmission environment. The signal envelope for small-scale fading in multipath channels is usually modeled by a Rayleigh distribution when there is no line-of-sight between the transmitter and the receiver. When there is a line-of-sight propagation path the small-scale fading of signal is described by a Rice distribution [91]. Wireless communication channels are usually named after their amplitude distributions. The Rayleigh fading channel represents the worst case of fading due to the absence of a specular component of the signal between the transmitter and receiver, and its power variation is about 20-30 dB [91].

The received signal, $r(t)$, is generally described by convolving a transmitted signal, $s(t)$, and the impulse response of channel, $c(t, \tau)$.

$$r(t) = \int_{-\infty}^{+\infty} s(t-\tau)c(t,\tau)d\tau \tag{A-1}$$

$c(t,\tau)$ is the equivalent low-pass response of the channel at time t due to a impulse applied at time $t-\tau$ and it is usually complex. For a Rayleigh fading channels, $c(t,\tau)$ is a zero mean complex valued Gaussian process. The phase and amplitude distributions of $c(t,\tau)$ is uniform over $(0, 2\pi)$ and Rayleigh respectively.

$$p(c) = \begin{cases} \frac{c}{\sigma_c^2} e^{(-\frac{c^2}{2\sigma_c^2})} & \text{for } c > 0 \\ 0 & \text{otherwise} \end{cases} \tag{A-2}$$

where $c = |c(t,\tau)|$ and $2\sigma_c^2$ is the mean power of the $c(t,\tau)$. Bello [92] specified the Rayleigh fading channels by its second order statistic under wide-sense stationary uncorrelated scattering (WSSUS) assumption. The model assumes that the attenuation and phase shift of the channel impulse response associated with different delays (e.g. τ and $\hat{\tau}$) is uncorrelated. The autocorrelation function, under WSSUS assumption, is defined by

$$\begin{aligned} \phi_c(\tau, \hat{\tau}; \Delta t) &= E[c(\tau, t+\Delta t) c^*(\hat{\tau}, t)] \\ &= R_c(\tau, \Delta t)\delta(\tau - \hat{\tau}) \quad -\infty \le \Delta t \le \infty \quad 0 \le \tau, \hat{\tau} \le \infty \end{aligned} \tag{A-3}$$

where $R_c(\tau, \Delta t) = E[c(\tau, t+\Delta t)c^*(\tau, t)]$, "$*$" denotes conjugation operator and E is the expectation operator. $R_c(\tau) = R_c(\tau, 0)$ called the multipath intensity profile shows how the average power of the channel impulse response varies as a function of time delay τ. The duration of τ over which $R_c(\tau)$ is essentially nonzero is denoted by T_m and is called the multipath spread of the channel or maximum excess delay. For defining T_m , the threshold level below the strongest component might be chosen around 10-20 dB. The other parameters which may be more useful in showing the behaviour of the multipath intensity profile of channel are mean, τ_m, and standard deviation, τ_{rms} (rms delay spread) which are defined as

$$\tau_m = \frac{\int_0^\infty \tau R_c(\tau) d\tau}{\int_0^\infty R_c(\tau) d\tau} \tag{A-4}$$

$$\tau_{rms} = \sqrt{\frac{\int_0^\infty (\tau - \tau_m)^2 R_c(\tau) d\tau}{\int_0^\infty R_c(\tau) d\tau}} \tag{A-5}$$

T_m or τ_{rms} can be viewed as a measurement which categorizes the channel as frequency selective or nonselective fading. When T, the symbol duration, is very much larger than T_m ($T \gg T_m$ or $T \gg \tau_{rms}$), the ISI is negligible and channel is called frequency-nonselective or flat fading. If $T < T_m$ (or $T < \tau_{rms}$), the channel disperses the transmitted signal in frequency domain and is called frequency selective fading. Let $R_c(\Delta f)$ define the Fourier transform of $R_c(\tau)$. The bandwidth over which the channel passes all spectral components with approximately equal gain and linear phase is called coherence bandwidth, $(\Delta f)_c$. T_m and $(\Delta f)_c$ are reciprocally related. For frequency nonselective and selective fading channels the bandwidth of the transmitted signal is less and larger than $(\Delta f)_c$ respectively. The two dimensional Fourier transform of $R_c(\tau, \Delta t)$ is defined $S_c(\Delta f, \zeta)$ and by selecting Δf to zero, $S_c(\zeta) = S_c(0, \zeta)$ is called the Doppler power spectrum of channel. $S_c(\zeta)$ shows the dispersion of a transmitted signal tone at the receiver. The inverse Fourier transform of $S_c(\zeta)$ shows how the received signal disperses in time-domain. The time duration over which the behaviour of the channel is constant is called coherence time $(\Delta t)_c$. When there is no time variations in the channel, $S_c(\zeta)$ becomes a delta function, $S_c(\zeta) = \delta(\zeta)$. The duration over which $S_c(\zeta)$ is essentially nonzero is called Doppler spread which is the reciprocal of coherence time. For mobile communication channel, $S_c(\zeta)$ is given by [93, 94]

$$S_c(\zeta) = \begin{cases} \frac{K_c}{\sqrt{1-(\frac{\zeta}{f_d})^2}} & |\zeta| < f_d \\ 0 & \text{otherwise} \end{cases} \tag{A-6}$$

where K_c is a constant, f_d is the maximum (one-sided) Doppler frequency shift. $f_d T$ is an important value which shows the time variations of fading channel. The channel is referred to as slow fading if the channel coherence time is larger than the symbol duration or $f_d T < 0.01$. The fading channel is called fast when the symbol duration is larger than the coherence time; $f_d T \geq 0.1$ is generally considered fast fading.

Appendix B

The $R_{\boldsymbol{y}^{M-1}}$ is a symmetric positive definite matrix and similar to $R_{\boldsymbol{y}}$, it can be factorized as $R_{\boldsymbol{y}^{M-1}} = L_{M-1}(\boldsymbol{a}_k)D_{M-1}(\boldsymbol{a}_k)L^{\mathcal{H}}_{M-1}(\boldsymbol{a}_k)$, where $L_{M-1}(\boldsymbol{a}_k)$ is a unit upper triangular matrix and $D_{M-1}(\boldsymbol{a}_k)$ is a diagonal matrix, $D_{M-1}(\boldsymbol{a}_k) = \mathrm{diag}(d_1(\boldsymbol{a}_k), d_2(\boldsymbol{a}_k), \ldots, d_{M-1}(\boldsymbol{a}_k))$. The inverse of $R_{\boldsymbol{y}^{M-1}}$ is $R^{-1}_{\boldsymbol{y}^{M-1}} = \Gamma^{\mathcal{H}}_{M-1}(\boldsymbol{a}_k)D^{-1}_{M-1}(\boldsymbol{a}_k)\Gamma_{M-1}(\boldsymbol{a}_k)$, where $\Gamma_{M-1}(\boldsymbol{a}_k) = L^{-1}_{M-1}(\boldsymbol{a}_k)$. From (2.19) one can easily show that all ijth elements of $\Gamma_{M-1}(\boldsymbol{a}_k)$ and $D^{-1}_{M-1}(\boldsymbol{a}_k)$ are equal to the all $(i-1)(j-1)$th elements of $\Gamma(\boldsymbol{a}_k)$ and $D^{-1}(\boldsymbol{a}_k)$ for $1 \le i, j \le M-1$ respectively. Then

$$\begin{bmatrix} 0 & \vdots\ 0_{1\times(M-1)} \\ \cdots\cdots & \cdots\cdots \\ 0_{(M-1)\times 1}\vdots & R^{-1}_{\boldsymbol{y}^{M-1}} \end{bmatrix} = \Gamma^{\mathcal{H}}(\boldsymbol{a}_k) \begin{bmatrix} 0 & \vdots\ 0_{1\times(M-1)} \\ \cdots\cdots & \cdots\cdots \\ 0_{(M-1)\times 1}\vdots & D^{-1}_{\boldsymbol{y}^{M-1}}(\boldsymbol{a}_k) \end{bmatrix} \Gamma(\boldsymbol{a}_k) \tag{B-1}$$

and $\Omega(\boldsymbol{a}_k)$ becomes

$$\begin{aligned} \Omega(\boldsymbol{a}_k) &= \Gamma^{\mathcal{H}}(\boldsymbol{a}_k) \begin{bmatrix} d_0^{-1}(\boldsymbol{a}_k) & & \\ & 0 & 0 \\ 0 & & \ddots \\ & & 0 \end{bmatrix} \Gamma(\boldsymbol{a}_k) \\ &= \begin{bmatrix} 1 \\ \rho^*_{1,0}(\boldsymbol{a}_k) \\ \vdots \\ \rho^*_{M-2,0}(\boldsymbol{a}_k) \\ \rho^*_{M-1,0}(\boldsymbol{a}_k) \end{bmatrix} d_0^{-1}(\boldsymbol{a}_k)\,[1, \rho_{1,0}(\boldsymbol{a}_k), \ldots, \rho_{M-2,0}(\boldsymbol{a}_k), \rho_{M-1,0}(\boldsymbol{a}_k)] \end{aligned} \tag{B-2}$$

where $\rho_{i,j}(\boldsymbol{a}_k)$ is the ijth element of $\Gamma(\boldsymbol{a}_k)$. Meanwhile the determinants $R_{\boldsymbol{y}}$ and $R_{\boldsymbol{y}^{M-1}}$ are

$$\begin{aligned}\det(R_{\boldsymbol{y}}) &= \prod_{m=0}^{M-1} d_m(\boldsymbol{a}_k) \\ \det(R_{\boldsymbol{y}^{M-1}}) &= \prod_{m=1}^{M-1} d_m(\boldsymbol{a}_k)\end{aligned}$$

Therefore by substituting $\Omega(\boldsymbol{a}_k)$, $\det(R_{\boldsymbol{y}^M})$ and $\det(R_{\boldsymbol{y}^{M-1}})$ in (2.18) the branch metrics $\beta(\boldsymbol{a}_k)$ become (2.21)

$$\beta(\boldsymbol{a}_k) = \log(d_0(\boldsymbol{a}_k)) + \frac{\left|\sum_{m=0}^{M-1} y_{k-m}\rho_{m,0}(\boldsymbol{a}_k)\right|^2}{d_0(\boldsymbol{a}_k)} \tag{B-3}$$

Appendix C

Proof of Theorem 1: Without loss of generality, first we assume that $\mathbf{i} = 2$. From the maximization steps of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$,(3.53), we conclude

$$Q(\hat{\boldsymbol{\theta}}_1^{(l+1)}, \hat{\boldsymbol{\theta}}_2^{(l)} | \hat{\mathcal{U}}^{(l)}) \geq Q(\hat{\boldsymbol{\theta}}_1^{(l)}, \hat{\boldsymbol{\theta}}_2^{(l)} | \hat{\mathcal{U}}^{(l)}) \tag{C-1}$$

$$Q(\hat{\boldsymbol{\theta}}_1^{(l+1)}, \hat{\boldsymbol{\theta}}_2^{(l+1)} | \hat{\mathcal{U}}^{(l)}) \geq Q(\hat{\boldsymbol{\theta}}_1^{(l+1)}, \hat{\boldsymbol{\theta}}_2^{(l)} | \hat{\mathcal{U}}^{(l)}) \tag{C-2}$$

The left side of (C-1) is equal to the right side of (C-2), therefore From (C-1) and (C-2) we have

$$Q(\hat{\mathcal{U}}^{(l+1)} | \hat{\mathcal{U}}^{(l)}) \geq Q(\hat{\mathcal{U}}^{(l)} | \hat{\mathcal{U}}^{(l)}) \tag{C-3}$$

Defining $V(\mathcal{U} | \hat{\mathcal{U}}^{(l)}) = E[\log p(\mathcal{D} | \mathcal{U}, \mathcal{I}) | \hat{\mathcal{U}}^{(l)}, \mathcal{I}]$ and from Jensen's inequality (see Lemma 1 in [38]) we can show

$$V(\hat{\mathcal{U}}^{(l+1)} | \hat{\mathcal{U}}^{(l)}) \leq V(\hat{\mathcal{U}}^{(l)} | \hat{\mathcal{U}}^{(l)}) \tag{C-4}$$

Thus from (3.5) and considering inequalities (C-3) and (C-4) it is easy to show

$$L(\hat{\mathcal{U}}^{(l+1)}) \geq L(\hat{\mathcal{U}}^{(l)}) \tag{C-5}$$

By following the same procedure, (C-1-C-4), one can achieve (C-5) for $\mathbf{i} > 2$.

References

[1] G. L. Stuber, *Principles of Mobile Communication.* Boston: Kluwer Academic Publishers, 1996.

[2] W. P. Chou and P. J. McLane, "16-state nonlinear equalizer for IS-54 digital cellular channels," *IEEE Trans. Veh. Technol.*, vol. 45, pp. 12–20, Feb. 1996.

[3] J. Lin, F. Ling and J. Proakis, "Joint data and channel estimation for TDMA mobile," *International Journal of Wireless Information Networks*, vol. 1, pp. 229–237, 1994.

[4] G. E. Bottomley and S. Chennakeshu, "Unification of MLSE receivers and estension to time-varying channels," *IEEE Trans. Commun.*, vol. 46, pp. 464–472, April 1998.

[5] B. Sklar, "Rayleigh fading channels in mobile digital communication systems, Part II: Mitigation," *IEEE Communications Magazine*, pp. 102–109, July 1997.

[6] K. Hamied and G. Stuber, "An adaptive truncated MLSE receiver for Japanese personal digital cellular," *IEEE Trans. Veh. Technol.*, pp. 41–50, Feb. 1996.

[7] N. Seshadri, "Joint data and channel estimation using blind trellis search techniques," *IEEE Trans. Commun.*, pp. 1000–1011, Feb. 1994.

[8] R. Raheli and A. Polydoros, "Per-survivor processing: A general approach to MLSE in uncertain environments," *IEEE Trans. Commun.*, vol. COM-43, pp. 354–364, Feb./March/April 1995.

[9] K. M. Chugg and A. Polydoros, "MLSE for an unknown channel- Part II: Tracking performance," *IEEE Trans. Commun.*, vol. 44, pp. 949–958, August 1996.

[10] J. H. Lodge and M. L. Moher, "Maximum likelihood sequence estimation of CPM signals transmitted over Rayleigh flat-fading channels," *IEEE Trans. Commun.*, vol. 38, pp. 787–794, June 1990.

[11] Q. Dai and E. Shwedyk, "Detection of bandlimited signals over frequency selective Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 42, pp. 941–950, Feb. 1994.

[12] X. Yu and S. Pasupathy, "Innovations-based MLSE for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. COM-43, pp. 1534–1544, Feb./Mar./Apr. 1995.

[13] M. E. Rollins and S. J. Simmons, "Simplified per-survivor Kalman processing in fast frequency-selective fading channels," *IEEE Trans. Commun.*, vol. 45, pp. 544–553, May 1997.

[14] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Magazine*, pp. 47–60, Nov. 1996.

[15] D. D. Falconer and F. R. Magee, "Adaptive channel memory truncation for maximum likelihood sequence estimation," *Bell Syst. Tech. J.*, vol. 9, pp. 1541–1562, Nov. 1973.

[16] M. V. Eyuboglu and S. U. H. Qureshi, "Reduced-state sequence estimation with set partitioning and decision feedback," *IEEE Trans. Commun.*, vol. COM-36, pp. 13–20, Jan. 1988.

[17] A. Duel-Hallen and C. Heegard, "Delayed decision-feedback sequence estimation," *IEEE Trans. Commun.*, vol. COM-37, pp. 428–436, May 1989.

[18] S. J. Simmons, "Breadth-first trellis decoding with adaptive effort," *IEEE Trans. Commun.*, vol. COM-38, pp. 3–12, Jan. 1990.

[19] J. P. Seymour and M. P. Fitz, "Near-optimal symbol-by-symbol detection schemes for flat Rayleigh fading," *IEEE Trans. Commun.*, vol. 43, pp. 1525–1533, February/March/April 1995.

[20] G. L. Turin, "Commutation signaling - an antimultipath technique," *IEEE Journal on Selected Areas in Communications*, vol. 4, pp. 548–562, July 1984.

[21] A. J. Viterbi, *CDMA Principles of Spread Spectrum Communication.* New York: Addison Wesley, 1995.

[22] J. G. Proakis, *Digital Communications.* New York: McGraw-Hill, 1989. 2nd edition.

[23] S. U. H. Qureshi, "Adaptive equalization," *IEEE Proceedings*, vol. 73, pp. 1340–1387, Sept. 1985.

[24] D. D. Falconer, "Signal processing," *Proceedings of Vehicular Technology Conference*, pp. 11–15, Atlanta, April 1996.

[25] T. Kailath, "Correlation detection of signals perturbed by a random channel," *IRE Transactions on Information Theory*, pp. 361–366, June 1960.

[26] G. D. Forney, "Maximum-likelihood sequence estimation of digital sequences in presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 363–378, May 1972.

[27] G. D. Forney, "The Viterbi algorithm," *IEEE Proceedings*, vol. 61, pp. 268–278, March 1973.

[28] A. J. Viterbi, "Error bound for convolutional codes an asymptotically optimum decoding algorithm," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 260–269, April 1967.

[29] J. K. Omura, "On the Viterbi decoding algorithm," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 177–179, January 1969.

[30] R. E. Bellman, *Dynamic Programing.* Princeton, NJ.: Princeton University Press, 1957.

[31] G. Ungerboeck, “Adaptive maximum-likelihood receiver for carrier-modulated data-transmission systems,” *IEEE Trans. Commun.*, vol. 22, pp. 624–636, May 1974.

[32] R. E. Morley and D. L. Snyder, “Maximum likelihood sequence estimation for randomly dispersive channels,” *IEEE Trans. Commun.*, vol. 27, pp. 833–839, June 1979.

[33] F. R. Magee and J. G. Proakis, “Adaptive maximum-likelihood sequence estimation for digital signaling in presence of intersymbol interference,” *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 120–124, Jan. 1973.

[34] S. U. H. Qureshi and E. E. Newhall, “An adaptive receiver for data transmission over time-dispersive channels,” *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 448–457, July 1973.

[35] M. Ghosh and C. L. Weber, “Maximum-likelihood blind equalization,” *Optical Engineering*, vol. 31, pp. 1224–1228, July 1992.

[36] K. M. Chugg and A. Polydoros, “MLSE for an unknown channel- Part I: optimality considerations,” *IEEE Trans. Commun.*, vol. 44, pp. 836–846, July 1996.

[37] B. D. Hart and D. P. Taylor, “Maximum-likelihood synchronization, equalization, and sequence estimation for unknown time-varying frequency-selective Rician channels,” *IEEE Trans. Commun.*, vol. 46, pp. 211–221, Feb. 1998.

[38] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum-likelihood from incomplete data via the EM algorithm,” *J. Roy. Statist. Soc.*, vol. 39, pp. 1–38, 1977.

[39] C. T. Beare, “The choice of the desired impulse response in combined linear-Viterbi algorithm equalizers,” *IEEE Trans. Commun.*, vol. COM-26, pp. 1301–1307, Aug. 1978.

[40] W. U. Lee and F. S. Hill, Jr., "A maximum-likelihood sequence estimator with decision-feedback equalization," *IEEE Trans. Commun.*, vol. COM-25, pp. 971–979, Sep. 1977.

[41] K. Wesolowski, "An efficient DFE and ML suboptimum receiver for data transmission over dispersive channels using two-dimensional signal constellations," *IEEE Trans. Commun.*, vol. COM-35, pp. 336–339, March 1987.

[42] J. B. Anderson and S. Mohan, "Sequential coding algorithm: a survey and cost analysis," *IEEE Trans. Commun.*, vol. COM-32, pp. 169–176, Feb. 1984.

[43] D. M. Titterington, "Recursive parameter estimation using incomplete data," *J. Roy. Stat. Soc. B*, vol. 46, no. 2, pp. 257–267, 1984.

[44] H. Zamiri-Jafarian and S. Pasupathy, "Recursive channel estimation for wireless communication via the EM algorithm," *Conference Proceedings of ICPWC'97*, pp. 33–37, Bombay, India, Dec. 1997.

[45] H. Zamiri-Jafarian and S. Pasupathy, "EM-based recursive estimation of channel parameters," *IEEE Trans. Commun.* ,(submitted).

[46] H. Zamiri-Jafarian and S. Pasupathy, "Adaptive MLSD receiver with identification of flat fading channels," *Proceedings of Vehicular Technology Conference*, pp. 695–699, Phoenix, May 1997.

[47] H. Zamiri-Jafarian and S. Pasupathy, "Adaptive MLSDE using the EM algorithm," *IEEE Trans. Commun.* ,(submitted).

[48] H. Zamiri-Jafarian and S. Pasupathy, "Adaptive state allocation in MLSE for Rayleigh fading channels," *Proceedings of Vehicular Technology Conference*, pp. 691–695, Atlanta, April 1996.

[49] H. Zamiri-Jafarian and S. Pasupathy, "Adaptive state allocation algorithm in MLSD receiver for multipath fading channels: Structure and strategy," *IEEE Trans. Veh. Technol.* ,(to be published).

[50] H. Zamiri-Jafarian and S. Pasupathy, "Complexity reduction of adaptive MLSDE receiver," ,(under preparation).

[51] S. Stein, "Fading channel issues in system engineering," *IEEE Journal on Selected Areas in Communications*, vol. 40, pp. 68–89, Feb. 1987.

[52] E. A. Lee and D. G. Messerschmitt, *Digital Communications.* Boston: Kluwer Academic Publishers, 1994, 2nd edition.

[53] A. Papoulis, "Markoff and wide-sense Markoff sequences," *Proc.IEEE*, p. 1661, Oct. 1965.

[54] A. Papoulis, *Probability, random variables, and stochastic processes.* New York: McGraw-Hill, 1991. 2nd edition.

[55] J. Makhoul, "A class of all-zero lattice digital filters: properties and applications," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-26, pp. 304–314, August 1978.

[56] W. H. A. Guston and E. H. Walker, "A multipath fading simulator for radio," *IEEE Trans. Veh. Technol.*, pp. 241–244, Nov. 1973.

[57] B. Hart, "MLSE diversity receiver structures," Ph.D. thesis, University of Canterbury, Christchurch, New Zealand, 1996.

[58] M. Feder and E. Weinstein, "Parameter estimation of superimposed signal using the EM algorithm," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, pp. 477–489, April 1988.

[59] E. Weinstein, M. Feder and A. Oppenheim, "Sequential algorithm for parameter estimation based on the Kullback-Leibler information measure," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 38, pp. 1652–1654, Sep. 1990.

[60] E. Weinstein, A. Oppenheim, M. Feder and J. R. Buck, "Iterative and sequential algorithms for multisensor signal enhancement," *IEEE Trans. on Signal Processing*, vol. 42, pp. 846–859, August 1994.

[61] V. Krishnamurthy and J. B. Moore, "On-line estimation of hidden Markov model parameters based on the Kullback-Leibler information measure," *IEEE Trans. on Signal Processing*, pp. 2557–2573, Aug. 1993.

[62] V. Krishnamurthy, "On-line estimation of dynamic shock-error models based on the Kullback-Leibler information measure," *IEEE Trans. Automatic Control*, pp. 1129–1135, May 1994.

[63] H. V. Poor, "On parameter estimation in DS/SSMA formats," *in Advances in Communication and Signal Processing*, pp. 59–70, Springer Verlag, 1988.

[64] C. N. Georghiades and D. L. Snyder, "The expectation-maximization algorithm for symbol unsynchronized sequence detection," *IEEE Trans. Commun.*, vol. 39, pp. 54–61, Jan. 1991.

[65] G. K. Kaleh and R. Vallet, "Joint parameter estimation and symbol detection for linear or nonlinear unknown channels," *IEEE Trans. Commun.*, vol. COM-42, pp. 2406–2413, July 1994.

[66] L. B. Nelson and H. V. Poor, "Iterative multiuser receivers for CDMA channels: an EM-based approach," *IEEE Trans. Commun.*, vol. 44, pp. 1700–1710, Dec. 1996.

[67] C. N. Georghiades and J. C. Han, "Sequence estimation in presence of random parameters via the EM algorithm," *IEEE Trans. Commun.*, vol. 45, pp. 300–308, March 1997.

[68] A. H. Sayed and T. Kailath, "A state-space approach to adaptive RLS filtering," *IEEE Signal Processing Magazine*, pp. 18–60, July 1994.

[69] J. A. Fessler, N. H. Clinthorne and W. L. Rogers, "On complete-data space for PET reconstruction algorithm," *IEEE Trans. on Nuclear Science*, vol. 40, pp. 1055–1061, August 1993.

[70] J. A. Fessler and A. O. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Trans. on Signal Processing*, vol. 42, pp. 2664–2677, Oct. 1994.

[71] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Prentice Hall, 1993.

[72] K. Sauer and C. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. on Signal Processing*, vol. 41, pp. 534–548, Feb. 1993.

[73] W. Zangwill, *Nonlinear Programming- a Unified Approach.* Englewood Cliffs: Prentice Hall, 1969.

[74] X. L. Meng and D. B. Rubin, "Maximum likelihood estimation via the ECM algorithm: a general framework," *Biometrika*, vol. 80, pp. 267–278, 1993.

[75] L. E. Baum, T. Petrie, G. Soules and N. Weiss, "A maximization technique occurring in statistical analysis of probabilistic functions of Markov chains," *The Ann. Math. Stat.*, vol. 41, pp. 164–171, 1970.

[76] L. E. Baum, "An inequality and associated maximization technique in statistical estimation for probabilistic function of Markov processes," *Inequalities*, vol. 3, pp. 1–8, 1972.

[77] S. M. Zabin and H. V. Poor, "Efficient estimation of class A noise parameters via the EM algorithm," *IEEE Trans. Inform. Theory*, vol. 37, pp. 60–72, Jan. 1991.

[78] A. Ansari and R. Viswanathan, "Application of expectation-maximization algorithm to the detection of direct-sequence signal in pulsed noise jamming," *IEEE Trans. Commun.*, vol. 41, pp. 1151–1154, August 1993.

[79] H. L. Van Trees, *Detection, Estimation and Modulation Theory; Part I.* John Wiley, 1968.

[80] S. Haykin, *Adaptive Filter Theory.* New York: Prentice Hall, 1991, 2nd edition.

[81] F. L. Vermeulen and M. E. Hellman, "Reduced state Viterbi decoders for channels with intersymbol interference," *Proceedings of Intl. Conference on Communications*, pp. 37B–1 to 37B–4, June 1974, Minneapolis, MN.

[82] G. J. Foschini, "A reduced state variant of maximum likelihood sequence detection attaining optimum performance for high signal-to-noise ratios," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 605–609, Sep. 1977.

[83] J. B. Anderson, "Limited search trellis decoding of convolutional codes," *IEEE Trans. Inform. Theory*, vol. IT-35, pp. 944–955, Sep. 1989.

[84] S. J. Simmons and P. Senyshyn, "Reduced-search trellis decoding of coded modulation over ISI channels," *Proceedings of IEEE Global Telecommunications Conference*, pp. 401.7.1–401.7.4, Nov. 1990, San Diego.

[85] M. V. Eyuboglu and S. U. H. Qureshi, "Reduced-state sequence estimation for coded modulation on intersymbol interference channels," *IEEE Journal on Selected Areas in Communications*, vol. 7, pp. 989–995, Aug. 1989.

[86] P. R. Chevillat and E. Eleftheriou, "Decoding of trellis-encoded signals in the presence of intersymbol interference and noise," *IEEE Trans. Commun.*, vol. COM-37, pp. 669–676, July 1989.

[87] S. Olcer, "Reduced-state sequence detection of multilevel partial-response signals," *IEEE Trans. Commun.*, vol. COM-40, pp. 3–6, Jan. 1992.

[88] B. E. Spinnler and J. B. Huber, "Design of hyper states for reduced-state sequence estimation," *Proceedings of IEEE Global Telecommunications Conference*, pp. 1–6, Nov. 1995, San Diego.

[89] T. L. M. Longo and R. M. Gray, "Quantization for decentralized hypothesis testing under communication constraints," *IEEE Trans. Inform. Theory*, vol. IT-36, pp. 241–255, March 1990.

[90] H. Zamiri-Jafarian and S. Pasupathy, "EM-approach to estimation of wireless channels with known/unknown noise statistics," *ICT'98*, pp. 94–98, Vol. II, Porto Carras, Greece, June 1998.

[91] B. Sklar, "Rayleigh fading channels in mobile digital communication systems, Part I: Characterization," *IEEE Communications Magazine*, pp. 90–100, July 1997.

[92] P. A. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun.*, vol. 11, pp. 360–393, Dec. 1963.

[93] W. C. Jakes, *Microwave Mobile Communications.* New York: Wiley, 1974.

[94] R. H. Clark, "A statistical theory of mobile-radio reception," *Bell Syst. Tech. J.*, pp. 957–999, July-Aug. 1968.