THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par *Institut National Polytechnique de Toulouse*
Discipline ou spécialité : *Signal, Image, Acoustique et Optimisation*

Présentée et soutenue par *Anchalee PUENGNIM*
Le *26 septembre 2008*

Titre : *Classification de modulations linéaires et non-linéaires à l'aide de méthodes Bayésiennes*

JURY

Mohamed NAJIM	Professeur à l'ENSEIRB, Bordeaux	Président
David BRIE	Professeur à l'Université Henri Poincaré, Nancy 1	Rapporteur
Guillaume GELLE	Professeur à l'Université de Reims	Rapporteur
Josep VIDAL	Profesor Titular à l'UPC, Espagne	Examinateur
Thierry ROBERT	Ingénieur de Recherche, CNES, Toulouse	Examinateur
Jean-Yves TOURNERET	Professeur à l'INPT, Toulouse	Directeur de thèse
Nathalie THOMAS	Maître de Conférence à l'INPT, Toulouse	Co-directrice de thèse

---

Ecole doctorale : *Mathématiques, Informatique, Télécommunications de Toulouse*
Unité de recherche : *Institut de Recherche en Informatique de Toulouse (IRIT)*
Directeur de Thèse : *Jean-Yves TOURNERET*

Acknowledgement

This PhD thesis has been prepared at the Télécommunications Spatiales et Aéronautiques (TéSA) in collaboration with the Centre National d'Etudes Spatiales (CNES).

First of all, I would like to thank the director of laboratory TéSA, Professor Francis Castanié for providing an excellent atmosphere during three years. I also would like to thank every member of the jury:

- Professor Mohamed Najim from the Université de Bordeaux-1 for being the president of the jury.
- Professor Guillaume Gelle from the Université de Reims Champagne-Ardenne and Professor David Brie from the Université Henri Poincaré for being my rapporteurs and for their informative comments to my thesis.
- Monsieur Thierry Robert from the CNES for his guidance and interest to the research project.
- Associate Professor Josep Vidal from the Technical University of Catalunya, Spain for his helpful discussion and valuable ideas on the application of the Baum-Welch algorithm.
- Professor Jean-Yves Tourneret, Head of the Signal and Communication team of the Institut de Recherche en Informatique de Toulouse, ENSEEIHT, Institut National Polytechnique de Toulouse, my advisor for his guidance, encouragement, and distinctive discussion. I am in debt of his careful proofreading and fruitful contribution to every publication. I am appreciated for his interest in my work and in exchanging resourceful ideas. Apart from his excellent competence in scientific work, I am very grateful for his kindness, thoughtfulness, and friendliness.

- Associate Professor Nathalie Thomas, Telecommunication and Network Department, ENSEEIHT, Institut National Polytechnique de Toulouse, my co-advisor for her generous and countless help. Her scientific supports and ideas have contributed enormously to this thesis. I appreciate her availability for work discussion and her amicability.

I also thank Monsieur Hervé Guillon from CNES. Within a short period of time that we worked together has produced an interesting project on GMSK.

I wish to express my gratitude to Corinne Mailhes and her family. It was an unforgettable and special way to learn French while being a baby-sitter during two years for such a wonderful family.

My special thanks go to my colleagues in the laboratory, who have made things interesting and fun, who offered helps when I needed, and who have provided interesting topics of conversation. All of these have given me the opportunities to practice French, to learn French and other cultures, and more importantly to feel integrated.

To my best friend, Kritsapon Leelavattananon, who helped me to execute lots of Matlab simulations using his powerful resources, I am very thankful. Without his help I would not be able to get simulation results in time.

I would like to specially thank Martin Blödt from my heart, who has made my life in foreign lands full of fantastic and interesting experiences, who has motivated me to work, and who has introduced me to his friends and they finally became my good friends.

Finally, I would like to thank my beloved parents for their love and support. I thank my lovely sister for her encouragement and understanding. I appreciate that she and her husband take such a good care of our parents during my long years of absence from home.

Résumé

La reconnaissance de modulations numériques consiste à identifier, au niveau du récepteur d'une chaîne de transmission, l'alphabet auquel appartiennent les symboles du message transmis. Cette reconnaissance est nécessaire dans de nombreux scénarios de communication, afin, par exemple, de sécuriser les transmissions pour détecter d'éventuels utilisateurs non autorisés ou bien encore de déterminer quel terminal brouille les autres.

Le signal observé en réception est généralement affecté d'un certain nombre d'imperfections, dues à une synchronisation imparfaite de l'émetteur et du récepteur, une démodulation imparfaite, une égalisation imparfaite du canal de transmission. Nous proposons plusieurs méthodes de classification qui permettent d'annuler les effets liés aux imperfections de la chaîne de transmission. Les symboles reçus sont alors corrigés puis comparés à ceux du dictionnaire des symboles transmis. Plus précisément, nous étudions trois techniques permettant d'estimer la loi a posteriori d'une modulation au niveau du récepteur. La première technique estime les paramètres inconnus associés aux diverses imperfections affectant le récepteur à l'aide d'une approche Bayésienne couplée avec une méthode de simulation MCMC (Markov Chain Monte Carlo). Une deuxième technique utilise l'algorithme de Baum Welch qui permet d'estimer de manière récursive la loi a posteriori du signal reçu et de déterminer la modulation la plus probable parmi un catalogue donné. La dernière méthode étudiée dans cette thèse consiste à corriger les erreurs de synchronisation de phase et de fréquence avec une boucle de phase.

Les algorithmes considérés dans cette thèse ont permis de reconnaître un certain nombre de modulations linéaires de types QAM (Quadrature Amplitude Modulation) et PSK (Phase Shift Keying) mais aussi des modulations non linéaires de type GMSK (Gaussian Minimum Shift Keying).

Abstract

This thesis studies classification of digital linear and nonlinear modulations using Bayesian methods. Modulation recognition consists of identifying, at the receiver, the type of modulation signals used by the transmitter. It is important in many communication scenarios, for example, to secure transmissions by detecting unauthorized users, or to determine which transmitter interferes the others.

The received signal is generally affected by a number of impairments. We propose several classification methods that can mitigate the effects related to imperfections in transmission channels. More specifically, we study three techniques to estimate the posterior probabilities of the received signals conditionally to each modulation. The first technique estimates the unknown parameters associated with various imperfections using a Bayesian approach coupled with Markov Chain Monte Carlo (MCMC) methods. A second technique uses the Baum Welch (BW) algorithm to estimate recursively the posterior probabilities and determine the most likely modulation type from a catalogue. The last method studied in this thesis corrects synchronization errors (phase and frequency offsets) with a phase-locked loop (PLL).

The classification algorithms considered in this thesis can recognize a number of linear modulations such as Quadrature Amplitude Modulation (QAM), Phase Shift Keying (PSK), and nonlinear modulations such as Gaussian Minimum Shift Keying (GMSK).

Contents

List of Figures

List of Tables

List of Abbreviations

ACM	adaptive coding and modulation
ALRT	average likelihood ratio test
APSK	amplitude phase shift keying
AWGN	additive white Gaussian noise
BPSK	binary phase shift keying
BW	Baum-Welch
CCSDS	consultative committee for space data system
CPM	continuous phase modulation
DT	decision-theoretic
DVB	digital video broadcasting
dB	decibel
FEC	forward-error-correction
FSK	frequency shift keying
GLRT	generalized likelihood ratio test
GMSK	Gaussian minimum shift keying
HLRT	hybrid likelihood ratio test
HMM	hidden Markov model
HOS	high-order statistic
ISI	intersymbol interference
ITU	International Telecommunications Union
LMS	least mean square
LP	lowpass
MAP	maximum a posteriori
MCMC	Markov Chain Monte Carlo

MFSK	M-ary frequency shift keying
MH	Metropolis-Hastings
MMSE	minimum mean square error
ML	maximum likelihood
MOM	method of moments
MPSK	M-ary phase shift keying
MQAM	M-ary quadrature amplitude modulation
MQAM	M-ary quadrature amplitude modulation
MSE	mean square error
NRZ	nonreturn-to-zero
OQPSK	offset quadrature phase shift keying
PAM	pulse amplitude modulation
PLL	phase-locked loop
PR	pattern recognition
PSK	phase shift keying
PSP	per-survivor processing
pdf	probability density function
QAM	quadrature amplitude modulation
QPSK	quadrature phase shift keying
qLLR	quasi log-likelihood ratio
SNR	signal-to-noise ratio
TDMA	time-division multiple access
UW	unmodulated waveform

List of Symbols

$\Phi(t, \mathbf{a})$	time-varying phase
Ω	Markov chain state space
α	unknown real amplitude factor
$\alpha_i(n)$	normalized forward variable
$\beta_i(n)$	normalized backward variable
θ	unknown parameter vector
λ_i	ith constellation or ith model
μ	mean
μ_m	LMS step-size parameter for *signal means*
μ_s	LMS step-size parameter for noise variance
ξ	damping factor
π_i	initial state distribution at state ith
σ^2	variance
ϕ	phase offset
ω_n	normalized natural frequency of the loop
B	3dB bandwidth of the lowpass Gaussian filter
$B_l T_{loop}$	normalize loop noise bandwidth
BT	bandwidth of GMSK signal
C_{4x}	fourth-order cumulant
$E[\,]$	expectation operator
$\mathrm{E_b}$	bit energy
$\mathrm{E_b/N_0}$	ratio of bit energy to noise power spectral density
$\mathrm{E_s}$	symbol energy

$\mathrm{E_s/N_0}$	ratio of signal energy to noise power spectral density
M_{km}	moments of order k
M	number of points in a constellation
M_j	number of points in the jth constellation
N_s	number of symbols in the observation interval
N	number of states of a HMM
N_i	number of iterations of MCMC methods
N_{bi}	number of burn-in iterations
N_0	noise power spectral density
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution
$P(\lambda_j \vert x)$	posteriori probability
$P(\lambda_j)$	prior probability
P_{cc}	average probability of correct classification
$P^{i\vert i}$	probability of correct classification
$R(\lambda_i \vert x)$	cost function
R	roll-off factor
$\mathbf{S}$	state matrix
$\boldsymbol{S}^\sharp$	pseudo-inverse of $\mathbf{S}$
T	symbol duration
$1/T_{loop}$	sampling rate of phase-locked loop
$\mathbf{a}$	transmitted data sequence
a_k	data symbol
a_{ij}	state transition probability distribution
$c_{i,j}$	cost
$\boldsymbol{m}$	vector containing all *signal means*
$d(n)$	i.i.d. symbol sequence
f_r	normalized residual carrier frequency, frequency offset
f	frequency
f_c	carrier frequency
f_d	Doppler shift
$f(t)$	impulse response of the lowpass filter

$g(t)$	frequency shape pulse
h	residual channel coefficient vector
$h(t)$	impulse response of the pulse-shaping filter
$l(x\|\lambda_j)$	logarithm of the likelihood function
m	modulation index
$p(\theta\|x)$	posteriori distribution
$p(x\|\lambda_i)$	likelihood function
$q(y\|\theta^n)$	proposal distribution
q	channel memory
$u(t)$	baseband complex envelope
$w(t)$	lowpass-equivalent complex Gaussian noise
$w_{BP}(t)$	additive white Gaussian noise with spectral density $N_0/2$ W/Hz
$x(n)$	baseband complex envelope of the received signal sampled at $t = nT$
$z(n)$	i.i.d. complex Gaussian noise sequence

Chapter 1

Introduction

Contents

Digital modulation classification consists of identifying the type of a modulated signal corrupted by noise and impairments. It is required in many communication applications such as interference identification, spectrum management, signal confirmation in non-cooperative scenarios and intelligent modems in cooperative scenarios. As the frequency spectrum becomes more and more packed, national and international regulators confront the increasingly complicated issue of managing and monitoring spectrum usage. For example, radio traffic is in general controlled by international regulations such as International Telecommunications Union (ITU). Spectrum monitoring and management systems provide the essential administrative and planning tools for regulatory entities to validate that licensees conform with the approved standards. These systems can automatically identify transmitters that are not included in the user's license database (illegal operators) as well as transmitters that are not working within their licensed parameters, e.g., channel bandwidth, roll-off factor, modulation type, bit rate, forward-error-correction (FEC) rate. Compliance with regulations is

also an advantage in ensuring mutual interference free and fighting against organized crime. Measuring the spectrum occupancy is crucial for planning future allocations of frequencies.

Another application of automatic modulation classification can be found in the adaptive coding and modulation (ACM) transmission available in digital video broadcasting standard (DVB-S2). For interactive point-to-point applications such as Internet navigation, ACM optimizes the transmission parameters for each user depending on path condition. In clear sky conditions, the transmitter may transmit data using a 16APSK modulation and switch to a 8PSK modulation under heavy fading. The scheme is achieved on frame to frame basic. The selected parameters should be known to the receiver so that data can be recovered correctly. The transmitter needs to send supplementary signaling data causing inefficiency in transmission. However, the loss can be alleviated using automatic modulation classification which the modulation type is identified at the receiver without redundant data from the transmitter.

In recent years, software defined radio and other reconfigurable communication systems, which the transceiver hardware can be reconfigured via software, have reinforced the important role of automatic modulation recognition. All reconfigurable communication systems must reconfigure every technical parameter particulary modulation format to be able to demodulate any type of input signal automatically.

A modulation classification algorithm decides what modulation is being transmitted based on information of the received signal such as amplitude, phase, frequency, and noise power. In general, the amount of information known to the receiver is very little. The nature of communication channels relating to applications also contributes to the uncertainty of the information. This makes the problem of modulation classification very difficult and challenging. There is always a trade-off amongst performance, speed, memory requirement, and complexity of classification algorithms depending on application purposes. For instance, spectrum monitoring and signal confirmation systems can afford high complexity and large memory requirement in exchange of very accurate results. In software defined radio, a fast algorithm may be preferred.

1.1 Problem formulation and signal model

Assume that we can operate in a coherent and synchronous environment and that the carrier, timing, and waveform recovery have been accomplished. All problems concerning signal bandwidth, baud rate, pulse-shaping filter and noise variance estimations are not addressed in this work.

1.1.1 Ideal case

In an ideal case, after preprocessing, the baseband complex envelope of the received modulated signal sampled at one sample per symbol at the output of the matched-filter corrupted by additive Gaussian noise can be written as:

$$x(n) = d(n) + z(n), \quad n = 1, 2, ..., N_s \tag{1.1}$$

where

- $x(n)$ is the baseband complex envelope of the received signal,
- $d(n)$ is an independent and identically distributed (i.i.d.) symbol sequence drawn from one of c constellations denoted $\{\lambda_1, \lambda_2, ..., \lambda_c\}$, where λ_j is a set of M_j complex numbers $\{S_1, S_2, ..., S_{M_j}\}$,
- M_j is the number of symbols of the j^{th} constellation,
- N_s is the number of symbols in the observation interval,
- $z(n)$ is an i.i.d. complex Gaussian noise sequence which has zero-mean and variance σ_z^2 (the real and imaginary components of $z(n)$ are independent and identically distributed).[1]

In this situation, the maximum likelihood (ML) classifier developed by Wei and Mendel [WM00], minimizes the probability of classification error, thus it achieves the optimum performance. Bayes theory provides a minimum error-rate classifier by finding the maximum a posteriori probabilities $P(\lambda_j|x), j = 1, 2, ..., c$. The Bayes classifier, which is a multi-hypothesis statistical testing, applies the Bayes rule:

$$\text{assign } x \text{ to } \lambda_i \text{ if } R(\lambda_i|x) \leq R(\lambda_j|x), \forall j = 1, ..., c, \tag{1.2}$$

[1] The parameter σ_z^2 is assumed to be known without loss of generality, as explained in [SS00].

with $R(\lambda_i|x) = \sum_{j=1}^{c} c_{i,j}\ P(\lambda_j|x)$ is the cost function and $c_{i,j}$ is the cost of deciding λ_i given that $x \in \lambda_j$. Furthermore, the maximum a posteriori (MAP) classifier can be obtained in the special case of $0-1$ loss functions:

$$c_{i,j} = \begin{cases} 0 & \text{if } i = j \\ 1 & \text{if } i \neq j \end{cases}$$

and expressed as

$$\text{assign } x \text{ to } \lambda_i \text{ if } P(\lambda_i|x) \geq P(\lambda_j|x), \forall j. \tag{1.3}$$

If all modulations are equally-likely

$$P(\lambda_j) = \frac{1}{c}\ ,\ \forall j,$$

the MAP classifier reduces to the ML classifier:

$$\text{assign } x \text{ to } \lambda_i \text{ if } p(x|\lambda_i) \geq p(x|\lambda_j), \forall j. \tag{1.4}$$

The ML classifier selects the modulation of the samples $x = [x(1), ..., x(N_s)]$ as the one that maximizes the probability density function (pdf) $p(x|\lambda_j)$ using the I and Q samples, where $x(n) = I(n) + jQ(n)$, as sufficient statistics. Such problem was studied in [WM00] in the ideal situation where all parameters are known a priori. The ML classifier can be rewritten as followed:

$$\text{assign } x \text{ to } \lambda_i \text{ if } l(x|\lambda_i) \geq l(x|\lambda_j), \quad \forall j, \tag{1.5}$$

where $l(x|\lambda_j)$ is the logarithm of the likelihood associated to class λ_j (whose constellation consists of M_j symbols $S_1, S_2, ..., S_{M_j}$) up to additive and multiplicative constants

$$l(x|\lambda_j) = \sum_{k=1}^{N_s} \ln \left\{ \frac{1}{M_j} \sum_{i=1}^{M_j} \exp\left(-\frac{1}{\sigma_z^2} \| \ x(k) - S_i \ \|^2 \right) \right\}. \tag{1.6}$$

Although this situation is unrealistic in noncooperative communications, it can be used as a reference to which suboptimal classifiers can be compared. This ideal classifier provides an upper bound of the expected performance for a digital modulation classifier.

1.1.2 More realistic case

In real scenarios, the received signal suffers from various impairments such as the imperfections of synchronization (the frequency, phase, and timing offsets), the intersymbol interference (ISI) from dispersive channel characteristic, the residual channel (from the imperfect equalization), and fading. All signal impairments lead to more complex problems in classification. As an example, the noisy received communication signal may be affected by the frequency and phase offsets and residual channel. In this case, the baseband complex envelope of the received signal sampled at one sample per symbol at the output of a matched filter can be written as in [SS00]:

$$x(n) = e^{j(\pi \frac{n}{N_s} f_r + \phi)} \sum_{l=0}^{q} h_l d(n-l) + z(n), \quad n = 1, 2, ..., N_s \tag{1.7}$$

where

- $f_r = 2N_s(f_c - \hat{f}_c) \in (-1/2, 1/2]$ is a normalized residual carrier frequency also called frequency offset (f_c is the carrier frequency and $\hat{f}_c$ is the frequency of the local oscillator, $\Delta f = f_c - \hat{f}_c$). Note that these notations imply that f_r is the constellation rotation whose maximum value is $\pi/2$ for $n = N_s$),
- $h = [1, h_1, ..., h_q]$ is the residual channel coefficient vector,
- ϕ is the phase offset.

The signal model in (1.7) takes into account the mismatch effect. The effects of the different parameters associated to model (1.7) are illustrated in Figure 1.1. 4QAM signals suffering from Gaussian noise only are shown in Figure 1.1 (a). The number of symbols N_s is 500. In the presence of the normalized frequency offset of 0.3 ($\Delta f = 3 \times 10^{-4}$) and the phase offset of $\pi/8$, the effects on the signals together with Gaussian noise are plotted in Figure 1.1 (b) and (c), respectively. Figure 1.1 (d) depicts the effect of a residual channel with $h = [1, 0.35, 0.25]$ in absence of Gaussian noise. The problem is more difficult when all imperfections (frequency offset, phase offset, and residual channel) are present as shown in Figure 1.2. The aim of this dissertation is to study classifiers that are robust to model mismatch, practical to implement, and achieve near optimum performance.

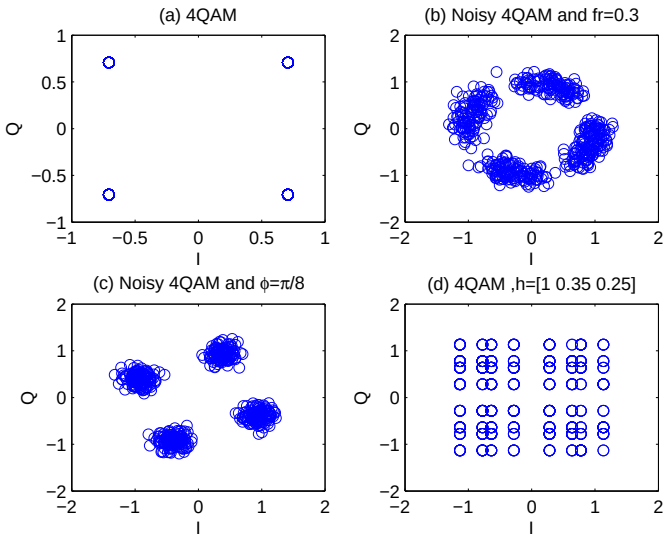


Figure 1.1: Model mismatch effect.

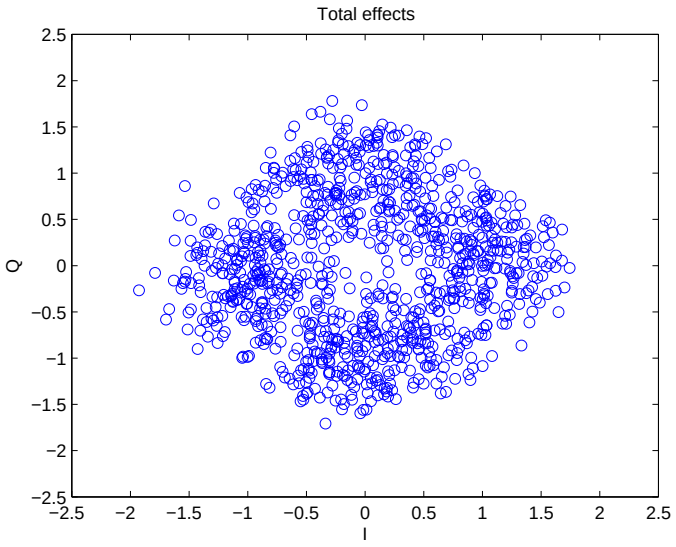


Figure 1.2: Total model mismatch effect.

1.2 Literature review

Two main classes of modulation classifications are decision-theoretic (DT) and statistical pattern recognition (PR). In the DT approach, modulation classification is a multiple hypothesis-testing problem and is solved mainly by the Bayes or ML principle. The core idea is that the posterior distributions of the possible constellations, conditioned on the modulated signal, provides all information for classification. A classifier based on Bayes rule achieves the optimal solution in the sense that it minimizes the probability of error (or an appropriate average cost function). However, the Bayes classifier may be difficult to implement due to its high computational complexity. This is particularly true for the classification of digital modulations, because averaging over the data symbols leads to an exponential computational complexity, when there are too many parameters unknown at the receiver. Also, the Bayes classifier is not robust to model mismatch. Most decision-theoretic approaches already used for modulation classification are summarized in Table 1.1. To overcome the difficulties inherent to the Bayesian strategy, several suboptimal likelihood based classifiers have been proposed in the signal processing and communication literature (see for instance [PK90; CLP95; ADC+04]). The main idea of these classifiers is to avoid the costly integration required to derive the posterior distribution of the unknown parameters. The integration can be avoided by approximating the average likelihood ratio test (ALRT) [PK90; CLP95], leading to the quasi log-likelihood ratio (qLLR), or by estimating every unknown parameter by the ML estimation and using the generalized likelihood ratio test (GLRT) [LP95]. The qLLR classifier does not approximate the likelihood ratio accurately enough for the optimum threshold (e.g., zero for two hypothesis tests) to be utilized. Replacing the unknown parameters with their ML estimates in GLRT no longer retains the optimum threshold. Therefore, these two techniques are sensitive to threshold setting which in turn depends on the number of observation symbols and the SNR. Hybrid likelihood ratio test (HLRT), which considers some unknown parameters as discrete random variables and the others as deterministic unknown variables, is another solution [HH02; PAP00]. The plug-in classifier that will be addressed in Section 1.3 is equivalent to HLRT classifier. The data symbols are treated as discrete random variables whereas the rest of the unknown parameters is treated as deterministic unknown variables. Thus estimating these deterministic unknown variables and replacing the estimates in the likelihood function are crucial steps for

successful classification. Pattern recognition approach is an alternative to likelihood based classifiers. The idea is to extract interesting features from the observations and use the features for classification. In this case, the key point is to find the "appropriate" set of features depending on the considered communication system. Classifiers based on the PR approach are generally much less complex than those of DT approach but suboptimal. Many features have been proposed in the literature including statistical moments [SH92] or higher-order statistics [SS00]. Table 1.1 resumes some DT based classifiers. Table 1.2 and 1.3 summarize some of existing pattern recognition classifiers. Details of classifiers in Table 1.1-1.3 can be found in Appendix A.

1.3 Plug-in classifier

A classifier based on Bayes or MAP rule achieves the optimal solution but may be impractical to implement due to high complexity and difficulty in computing the exact posterior probability $P(\lambda_i|x)$ in the presence of many unknown parameters. To obtain a near optimum performance and realizable classifier, the estimated posterior probability $\hat{P}(\lambda_i|x)$ can be an interesting solution. From the MAP classifier in (1.3) and introducing the unknown parameter θ, we can write

$$\text{assign } x \text{ to } \lambda_i \text{ if } p(x|\theta_i, \lambda_i)P(\lambda_i) \geq p(x|\theta_j, \lambda_j)P(\lambda_j), \forall j. \tag{1.8}$$

Let $\hat{p}(x|\theta_i, \lambda_i)$ be the estimated pdf of observation sequence conditioned on the class λ_i and the estimate $\hat{\theta}_i$. If all modulations have equal prior probabilities, (1.8) reduces to

$$\text{assign } x \text{ to } \lambda_i \text{ if } \hat{p}(x|\theta_i, \lambda_i) \geq \hat{p}(x|\theta_j, \lambda_j), \forall j. \tag{1.9}$$

The classifier that will be considered intensively in this study consists of replacing the unknown parameters (e.g., f_r, ϕ, h) in the pdf by their estimated values, resulting in the so-called plug-in classifier in (1.9). Proposed techniques to calculate the estimated posterior probability $\hat{P}(\lambda_i|x)$ will be presented in Chapter 2.

1.4 HOS classifier

The high-order statistic (HOS) classifier developed by Swami [SS00] will be considered with a particular interest in this study. The features used by the authors in [SS00] are the

normalized fourth-order cumulants. These statistics characterize the shape of the noisy signal constellations or noisy baseband samples. This method can easily be applied in a hierarchical manner to classify various digital signaling formats. It is particularly effective for discriminating format subclasses, such as phase shift keying (PSK) versus pulse amplitude modulation (PAM) and quadrature amplitude modulation (QAM). It may also be applied within each subclass to determine the exact modulation type if sufficient signal-to-noise ratio (SNR) and sample size are available.

For a complex-valued stationary random process x_k, the mixed moments of order k are defined as [DWW02], [LB97]:

$$M_{km} \triangleq E[x_k^{k-m}(x_k^*)^m]. \tag{1.10}$$

Thus the second-order moment can be written as

$$M_{20} = E[x_k^2] \quad \text{or} \quad M_{21} = E[|x_k|^2]. \tag{1.11}$$

Many different stationary fourth-order cumulants can be defined for complex signals. These cumulants include

$$\begin{aligned} C_{40} &= cum(x_k, x_k, x_k, x_k), \\ C_{41} &= cum(x_k, x_k, x_k, x_k^*), \\ C_{42} &= cum(x_k, x_k, x_k^*, x_k^*). \end{aligned} \tag{1.12}$$

For zero-mean random variables a, b, c, and, d, the operator $cum(\centerdot, \centerdot, \centerdot, \centerdot)$ is defined as

$$cum(a, b, c, d) = E[abcd] - E[ab]E[cd] - E[ac]E[bd] - E[ad]E[bc]. \tag{1.13}$$

The fourth-order cumulants can be estimated by standard estimators [SS00]:

$$\begin{aligned} \hat{C}_{40} &= \frac{1}{N_s}\sum_{k=1}^{N_s} x_k^4 - 3\left(\frac{1}{N_s}\sum_{k=1}^{N_s} x_k^2\right)^2, \\ \hat{C}_{41} &= \frac{1}{N_s}\sum_{k=1}^{N_s} x_k^3 x_k^* - 3\left(\frac{1}{N_s}\sum_{k=1}^{N_s} x_k^2\right)\left(\frac{1}{N_s}\sum_{k=1}^{N_s} |x_k|^2\right), \\ \hat{C}_{42} &= \frac{1}{N_s}\sum_{k=1}^{N_s} |x_k|^4 - \left|\frac{1}{N_s}\sum_{k=1}^{N_s} x_k^2\right|^2 - 2\left(\frac{1}{N_s}\sum_{k=1}^{N_s} |x_k|^2\right)^2. \end{aligned} \tag{1.14}$$

Theoretical cumulant statistics C_{40} and C_{42} have been computed for various constellation types (see Table I of [SS00]). C_{42} is used first to decide whether the constellation is real-valued (BPSK/PAM), circular (PSK), or rectangular (QAM). Then, if the unknown phase rotation can be assumed to be small, $|C_{40}|$ may be used for classification within each subclass. If the unknown phase rotation cannot be ignored, then $|C_{40}|$ must be employed rather than C_{40}. Table I of [SS00] also shows that it is advantageous to use $|C_{40}|$ for the test statistic (rather than C_{42}) because $|C_{40}| = 0$ for 8PSK. For a given SNR, one can compute the optimal threshold under the assumption that $\hat{C}_{40}$ is Gaussian. Let μ_k and σ_k^2 denote the mean and variance of the statistic, $S = \hat{C}_{40}$, under the k^{th} hypothesis; and assume that the M hypotheses are ordered so that $\mu_1 < \mu_2 < ... < \mu_M$. A simplifying approximation considers that the variances of S are all equal under the M hypothesis. In this case, the decision rule consists of choosing H_k if

$$(\mu_{k-1} + \mu_k)/2 < S < (\mu_k + \mu_{k+1})/2, \tag{1.15}$$

with $\mu_0 = -\infty$ and $\mu_{M+1} = \infty$.
Consider the four-class problem based on the following set of modulation types

$$\lambda = \{\text{BPSK}, \text{4PAM}, \text{8PSK}, \text{16QAM}\}. \tag{1.16}$$

The decision rule for the four-class problem Ω_4 can then be summarized as:

$$|\hat{C}_{40}| < 0.34 \Rightarrow \text{8PSK}$$

$$0.34 \leq |\hat{C}_{40}| < 1.02 \Rightarrow \text{16QAM}$$

$$1.02 \leq |\hat{C}_{40}| < 1.02 \Rightarrow \text{4PAM}$$

$$|\hat{C}_{40}| \geq 1.68 \Rightarrow \text{BPSK}$$

The cumulants $\hat{C}_{40}$ and $\hat{C}_{42}$ are moderately robust to model mismatch. Even if the thresholds have been determined by assuming no frequency offset, phase offset, and residual channel, the classifier works well in presence of imperfection.

1.5 Performance measure

A basic performance measure is the probability of correct classification. Denotes as $P^{(i|i)}$ the (classification) probability to declare the signal format i has been sent, where indeed the

format i is present. The average probability of correct classification is defined by

$$P_{cc} = \frac{1}{c}\sum_{i=1}^{c} P^{(i|i)}, \tag{1.17}$$

where c is the number of possible modulations (i.e., the number of classes). When the theoretical probability of error cannot be determined in closed-form, (1.17) provides a simple measure of classification performance and is used throughout this thesis.

1.6 Chapter organization and contribution

The organization of this thesis is described in the following. Chapter 2 presents a method to mitigate signal impairments such as synchronization errors due to the imperfection of local oscillator, residual channel from imperfect equalization, and intersymbol interference (ISI) from dispersive channel. The method estimates unknown parameters related to the impairments of the transmission channel in order to compensate their effects by de-rotating, de-spreading and de-convolving. We have studied three techniques to estimate the unknown parameters based on Markov Chain Monte Carlo (MCMC) methods, the Baum-Welch (BW) algorithm, and phase-locked loop (PLL). Our work in this chapter leads to:

- A modification of the MCMC parameter estimation to approximate complex residual channels as an extension of the work in [LTD01].

Chapter 3 studies a plug-in MAP classifier based on the MCMC parameter estimation. We study the classification performance of the proposed classifier in the presence of mismatch effects. The same methodology is extended to a slow Rayleigh fading environment. The key findings in this chapter are:

- The phase and frequency offsets, residual channels, as well as fading amplitude and phase can be estimated using MCMC methods.
- The proposed MCMC plug-in classifier can identify BPSK/4PAM/8PSK/16QAM modulations subjected to model mismatch effects and outperforms the ML and HOS classifiers studied by [WM00] and [SS00], respectively.
- For our study, the computation complexity involving in drawing samples of Metropolis-Hastings (MH) algorithm can be reduced using the relationship: $\ln \sum_j e^{a_j} \simeq \max_j a_j$.

- Our simulation results demonstrate that the proposed MCMC plug-in classifier can distinguish BPSK/QPSK/8PSK/16QAM modulations under a slow Rayleigh fading and outperforms the MOM classifier in [ADC+04].

Chapter 4 studies the utilization of the BW algorithm to a plug-in MAP classifier. The BW algorithm is used to compute the posterior probabilities which are then plugged into the optimal Bayes decision rule. The classification performance of the proposed MAP classifier based on the BW algorithm is presented for different scenarios. Note that Chapter 3 and Chapter 4 focus on linear modulations. The main contributions include:

- Our proposed classifier can identify OQPSK from QPSK modulations even they have the same constellations. It can be generalized to classify between $\pi/4$-QPSK and 8PSK modulations, which also share the constellation shapes.
- Compare to the qLLR classifier proposed by [CLP95] to recognize BPSK/QPSK/OQPSK modulations, our proposed classifier obtains better classification performance.
- Our proposed classifier can recognize 16PSK/16QAM modulations in unknown ISI channels and outperforms the PSP/GLRT classifier studied by [LP95].
- It can be used to discriminate BPSK/QPSK/8PSK/16QAM modulations in unknown ISI channels.
- From our simulations, it is found that the initialization of the BW algorithm by the method according to [Men91] improves classification performance compare with random initialization.
- From simulations, we find that the LMS-update type BW algorithm yields better classification results than that of the standard batch mode BW algorithm.
- We demonstrate the use of the BW algorithm in conjunction with phased-lock loops (PLL) to classify BPSK/QPSK/OQPSK/8PSK/16QAM modulations in the presence of large frequency offsets.

Chapter 5 deals with nonlinear modulations. It describes the importance of nonlinear modulations and motivation to recognize linear and nonlinear modulations. We explain the principle of nonlinear Gaussian minimum shift keying (GMSK) modulations and how the BW

algorithm can be applied to them. We use the proposed MAP classifier studied in Chapter 4 for unknown ISI channels to recognize two GMSK modulations with different bandwidths on AWGN channels. Then we apply the same strategy to identify linear modulations from nonlinear modulations. The important results are:

- The proposed MAP classifier based on the BW algorithm can be applied to recognize GMSK modulations with different bandwidth ($BT = 0.25$, $BT = 0.5$) and provides good classification performance. This is based on the assumption that other linear modulations have been identify by other existing methods.
- We can classify linear modulations (BPSK, QPSK, 8PSK) and nonlinear (GMSK $BT = 0.25$, GMSK $BT = 0.5$) modulations using the conventional receiver structure for linear modulations and the proposed MAP classifier based on the BW algorithm.
- The proposed classifier achieves good classification performance at small SNRs as required by satellite/space communication applications.

Chapter 6 concludes the importance of our work and proposes directions for future research.

Table 1.1: A summary of DT classifiers in AWGN.

Author(s)	*Modulations*	*Unknown parameters*
Wei et al. [WM00]	16QAM, V29	-
Sapiano et al. [SM96]	UW, BPSK, QPSK, 8PSK	-
Sills [Sil99]	BPSK, QPSK, 16QAM, V29, 32QAM, 64QAM	Phase offset
Kim et al. [KP88], [PK90]	BPSK, QPSK	Phase offset
Long et al. [LCP94]	16PSK, 16QAM, V29	Phase offset
Hong et al. [HH03]	BPSK, QPSK	Signal level
Beidas et al. [BW96], [BW98]	32FSK, 64FSK	Phase jitter, timing offset
Pannagiotu et al. [PAP00]	16PSK, 16QAM, V29	Phase offset
Hong et al. [HH00]	BPSK, QPSK	Signal level
Chugg et al. [CLP95]	BPSK, QPSK, OQPSK	Phase offset, signal power, noise power

Table 1.2: A summary of PR classifiers under the ideal situation in AWGN.

Author(s)	*Features*	*Modulations*
Azzouz et al. [AN96b], [AN96a]	Maximum power spectral density of normalized centered amplitude, standard deviation of normalized centered amplitude, phase and frequency	2ASK, 4ASK, BPSK, QPSK, 2FSK, 4FSK
Hsue et al. [HS89], [HS90]	variance of the zero-crossing interval sequence, phase difference and zero-crossing interval histrograms	UW, BPSK, QPSK, 8PSK, BFSK, 4FSK, 8FSK
Yang et al. [YS91a], [YS97], [YL98]	PDF of phase	UW, BPSK, QPSK, 8PSK
Soliman et al. [SH92], [YS95]	Statistical moments phase	UW, BPSK, QPSK, 8PSK
Sapiano et al. [SMH95]	DFT of phase PDF	UW, BPSK, QPSK, 8PSK
HO et al. [HPC95], [HPC00]	Variance of haar WT magnitude, haar WT magnitude and peak magnitude histograms	BPSK, QPSK, 8PSK, 2FSK, 4FSK, 8FSK, CP2
LeMartret et al. [**?**]	Forth- and second-order moments of the received signal	QPSK, 16QAM
LeMartret et al. [MLL98]	Forth- and second-order cyclic cumulants of the received signal	QPSK, 16QAM, 64QAM
Dobre et al. [DBNS03]	Eigth-order cyclic cumulants of the received signal	BPSK, QPSK, 8PSK, 4ASK, 8ASK, 16QAM, 64QAM, 256QAM
Yu et al. [YSS03]	DFT of the received signal	2FSK, 4FSK, 8FSK, 16FSK, 32FSK

Table 1.3: A summary of PR classifiers.

Author(s)	*Features*	*Modulations*	*Unknown parameters*	*Channel(s)*
Swami et al. [SS00]	Normalized fourth-order cumulants of the received signal	BPSK, 4ASK, 16QAM, 8PSK, V32, V29, V29c	Phase, frequency and timing offset	AWGN, impulsive noise, cochannel interference
Dobre et al. [DBNS04]	Eighth-, sixth, and fourth-order cyclic cumulants of the received signal	4QAM, 16QAM	Phase and frequency offset, phase jitter	AWGN, impulsive noise
Spooner et al. [Spo95]	Fourth- and second-order cyclic cumulants of the received signal	MSK, QPSK, BPSK, 8PSK, 8QAM	-	AWGN, cochannel interference
Spooner [Spo01]	Sixth-, fourth- and second-order cyclic cumulants of the received signal	QPSK, 16QAM, 64QAM, V29	-	AWGN, cochannel interference

Chapter 2

Impairment Mitigation

Contents

2.1 Introduction

In coherent system, synchronization errors such as frequency and phase offsets due to the imperfection of local oscillator lead to system impairments. The frequency offset spreads the received signal constellation points whereas the phase offset rotates them, see Figure 1.1

(b)-(c). The residual channel from imperfect equalization and ISI from dispersive channel are also sources of deterioration. To mitigate the effects, several unknown parameters can be estimated and then the estimated values are used to de-rotate and de-spread the received signals. This step is the key point of the plug-in classifiers studied intensively in this thesis. To implement the plug-in classifiers in (1.9), we have to estimate the posterior probability $\hat{P}(\lambda_i|x)$ or equivalently the pdf of observation sequence given each possible modulation format. We study three approaches to estimate the probability of observation sequence given the model as well as unknown parameters.

- The first approach consists of two steps. The parameter estimation step is carried out by a Bayesian method coupled with Markov Chain Monte Carlo (MCMC) methods. The second step replaces the unknown parameters by their estimates in the likelihood function and calculate the posterior probability of each possible modulation.

- The second approach relies on hidden Markov models (HMMs) whose model parameters and the probability of observation sequence given the model are approximated at one step using the forward/backward Baum-Welch (BW) algorithm. The estimation of the probability of observation sequence conditioned on each possible modulation type will be used in the rest of the work for classification.

MCMC methods and the BW algorithm are appropriate for packet transmission because they work on the principle of feed-forward estimation. The packet mode of operation is typical of time-division multiple access (TDMA), where several users share the capacity of the communication channel by transmitting bursts of data in non-overlapping time intervals. Further distinction will be made between short packet operation and long packet operation. In the case of short packet communication, the number of data symbols per packet is so small that the carrier phase and frequency offsets can be considered as constant over the entire burst. As a result, it is sufficient to acquire a single parameter estimation per packet and apply these parameters for detecting all data symbols within the packet. In the case of long packet communication, the variation of these parameters over the packet cannot be neglected. In this case, operation in long packet mode is similar to operation in continuous mode. Thus, it is necessary to make multiple carrier phase and frequency estimates per packet in order to track the fluctuations of these parameters. In the packet transmission, the synchronization parameters (particularly the carrier phase) can change significantly between

packets from the same user, so that these parameters must be acquired again for each packet. In the case of short packet communication, the information sequence per packet might be only about a hundred symbols long, so that for a high efficiency there should be only a very short preamble or preferably no preamble at all. Therefore, a feed-forward synchronizer or so called estimator is required for this mode of transmission due to a very short acquisition time requirement. In the continuous transmission, this requirement can be relaxed. Applications of the continuous transmission are found in Satellite downlink communications such as Digital Video Broadcasting (DVB) systems. Feedback synchronizers such as phase-locked loops (PLLs) are generally preferred because of their abilities to track any change in the carrier phase and frequency.

- The last approach also has two steps. A phase-locked loop (PLL) is applied to compensate the frequency and phase offsets first. Then the probability of observation sequence given the model is estimated by the BW algorithm.

2.2 MCMC Parameter estimation

The requirement of the integration of high-dimensional functions to obtain the posterior distribution is the main drawback of Bayesian approaches. An attempt to alleviate the disadvantage is to compute complex integrals by expressing them as expectation of some distribution and then estimate this expectation by drawing samples from that distribution. This is referred to as Monte Carlo integration. MCMC methods are techniques that simulate samples from some complex distribution of interest. MCMC methods are so-named because one uses the previous sample values to randomly generate the next sample value, generating a Markov chain (as the transition probabilities between sample values are only a function of the most recent sample value).

The signal model is expressed as in (1.7). The unknown parameter vector $\theta = (f_r, \phi, h)$ is estimated according to the minimum mean square error (MMSE) principle (which minimizes the standard quadratic cost function $E[(\hat{\theta} - \theta)^2]$)

$$\hat{\theta}_{\text{MMSE}} = E[\theta|x]. \tag{2.1}$$

Obviously, a closed-form expression for the MMSE estimator of θ cannot be obtained. However, the MMSE estimate (which is the mean of the *a posteriori* density) can be approximated as follows

$$\hat{\theta}_{\text{MMSE}} = \int \theta p(\theta|x) d\theta \simeq \frac{1}{N_i} \sum_{n=1}^{N_i} \theta^n, \tag{2.2}$$

where $\theta^n, n = 1, ..., N_i$ are samples drawn from $\theta^n \sim p(\theta|x)$ and N_i is the number of iterations. This result can be used to approximate the MMSE estimator $\hat{\theta}_{\text{MMSE}}$, as soon as it is possible to generate samples θ^n distributed according to $p(\theta|x)$. This work proposes to generate θ^n using the Metropolis-Hastings (MH) algorithm, which is one of the most popular MCMC methods. The MH algorithm consists of drawing samples distributed according to $p(\theta|x)$ by running an ergodic Markov chain whose stationary distribution is the posteriori distribution $p(\theta|x)$. The reader is invited to consult [GRS96] for more details. It is proved that the following ergodic theorem holds ([GRS96, p.47])

$$\frac{1}{N_i} \sum_{k=1}^{N_i} \eta(\theta^k) \underset{N_i \to \infty}{\overset{as}{\to}} E_{p(\theta|x)}[\eta(\theta)], \tag{2.3}$$

for every function η defined on Ω, where $E_{p(\theta|x)}[\eta(\theta)]$ exists. In (2.3), $\underset{N_i \to \infty}{\overset{as}{\to}}$ denotes almost-sure convergence. Consequently, after discarding the so-called "burn-in" samples, the mean of the *a posteriori* distribution (2.1) is estimated by the time average of the remaining Markov chain samples (which converges to the MMSE estimator according to (2.3)).

2.2.1 Posteriori distribution $p(\theta|x)$

By Bayes theorem, the posterior distribution can be expressed as

$$p(\theta|x) \propto p(x|\theta)p(\theta),$$

where $\propto$ means *proportional to*, $p(x|\theta)$ is the likelihood of the observed data conditioned on the unknown vector θ, and $p(\theta)$ is the prior knowledge about θ. The following priors for the MMSE estimation of θ are used:

- Uninformative independent uniform priors for the frequency and phase offsets: $p(f_r, \phi) = p(f_r)p(\phi)$ where $p(f_r) = I_{(-1/2,1/2]}(f_r)$, $p(\phi) = \frac{M}{2\pi} I_{[-\pi/M,\pi/M]}(\phi)$ for an M-PSK modulation, $p(\phi) = \frac{2}{\pi} I_{[-\pi/4,\pi/4]}(\phi)$ for other modulations, and I is the indicator function.

- Independent normal priors distribution $\mathcal{N}(0, \sigma_h^2)$ are selected for the residual channel FIR filter taps, $\sigma_h^2 = 0.01$ (see [SS00]). A suitable choice of parameter σ_h^2 allows to incorporate vague prior information about the parameters h_l.

From the signal model in (1.1), we know that the conditional pdf of the observation symbol given the transmitted symbol has the normal distribution and the logarithm of the likelihood associated to class λ_j has the form in (1.6). Thus reversing the effect of the unknown parameters θ results in an approximated likelihood function of (1.6). This can be achieved by de-spreading and de-rotating the observation symbol by the estimated frequency and phase offsets, respectively. Then the observation sequence is passed through the inverse filter constructed from the estimated coefficients. Given a modulation type λ_j with M_j constellation points, the approximated logarithm of the likelihood can be written as

$$\hat{l}(x|\theta) = \sum_{k=1}^{N_s} \ln \left[\frac{1}{M_j} \sum_{i=1}^{M_j} \exp \left(-\frac{1}{\sigma_z^2} \parallel x(k)^\theta - S_i \parallel^2 \right) \right], \tag{2.4}$$

where $x(k)^\theta$ is the output of the filter having the transfer function $H^{-1}(z) = 1/\left(\sum_{l=1}^{q} \hat{h}_l z^{-l}\right)$ driven by the input $x(k)e^{-j(\pi k \hat{f}_r/N_s + \hat{\phi})}$.

2.2.2 Metropolis-Hastings (MH) algorithm

The Markov chain state space and current state are denoted by Ω and $\theta^n = (f_r^n, \phi^n, h^n) \in \Omega$, respectively. At each iteration, a candidate y is drawn according to a proposal distribution $q(y|\theta^n)$. This candidate is accepted with the following probability:

$$\alpha(\theta^n, y) = \min \left\{ 1, \frac{p(y|x)q(\theta^n|y)}{p(\theta^n|x)q(y|\theta^n)} \right\}. \tag{2.5}$$

Equivalently, if *rand* is the outcome of a uniform drawing on [0,1] and θ^n is the value of θ at iteration n, the next value of θ is chosen as follows:

$$\begin{cases} \theta^{n+1} = y \text{ if } rand < \frac{p(y|x)q(\theta^n|y)}{p(\theta^n|x)q(y|\theta^n)}, \\ \theta^{n+1} = \theta^n \text{ otherwise.} \end{cases} \tag{2.6}$$

A fundamental property of the MH algorithm is that any proposal distribution $q(y|\theta^n)$ can be chosen, provided that the support of $p(\cdot|x)$ is contained in the support of $q(y|\theta^n)$ [GRS96]. In this work, it is appropriate to use the normal distribution $\mathcal{N}(\theta^n, \sigma^2)$ as a

proposal distribution, where σ^2 is optimized to provide a suitable acceptance rate (1/4 to 1/2, see [Rob98, p.8]. We choose to draw y from a local perturbation of the previous sample, i.e., $y = \theta^n + \epsilon$, leading to the well-known random-walk MH algorithm. In this case, the proposal distribution is of the form $q(y|\theta^n) = g(y - \theta^n)$. Interestingly, the choice of a symmetric distribution for g leads to an acceptance probability which is independent on q.

Instead of updating the whole of θ *en bloc*, it is often more convenient and computationally efficient to divide θ into k blocks and to update each block one-at-a-time. This procedure has been suggested by many authors (see [GRS96] for more details) and has been shown to improve the mixing property of the sampler. Here we propose to update θ one component at-a-time. Such strategy, indeed, exhibits good performance in classification of digital modulations as shown in [LTD01].

2.2.3 Simulation results

This section shows some results of MCMC parameter estimation. A 4QAM signal is transmitted through a linear FIR channel and corrupted by Gaussian noise. At the receiver, the signal is further deteriorated by synchronization errors. The signal-to-noise ratio (SNR) in decibels for the normalized (unit energy) constellation is defined as

$$\text{SNR} = 10 \log_{10} \left(\frac{1}{\sigma_z^2} \right). \tag{2.7}$$

Simulation parameters

- λ_j =4QAM, $M_j = 4$.
- $f_r = 0.1$, $\phi = 0.2$, otherwise stated in the figures.
- $h = [1,\ 0.35 + 0.25j]$.
- SNR = 10 dB
- Current state: $\theta^n = (f_r^n, \phi^n, h^n)$.
- Proposal distribution: random walk $q(y|\theta^n) \sim \mathcal{N}(\theta^n, \sigma^2)$ where $\sigma = 0.03$.

Figure 2.1 plots the histogram of the estimated posteriori distribution of the frequency offset f_r and phase offset ϕ. The number of samples N_s is 250 and the number of iterations

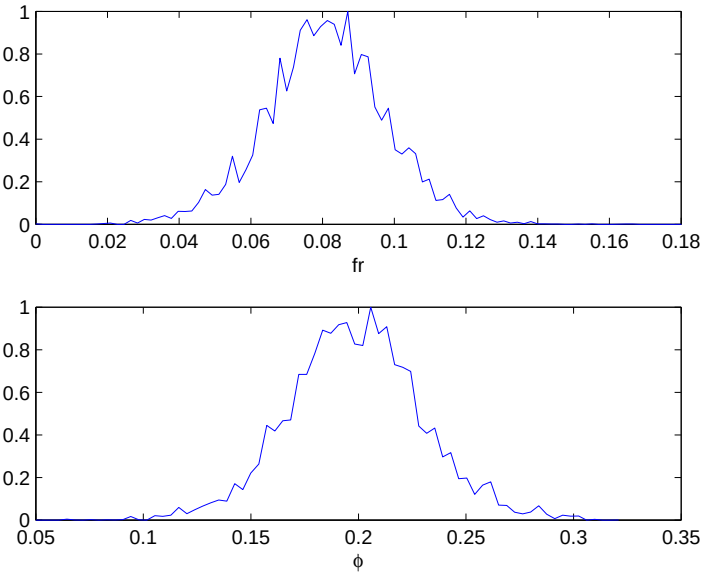


Figure 2.1: Normalized histograms of f_r and ϕ.

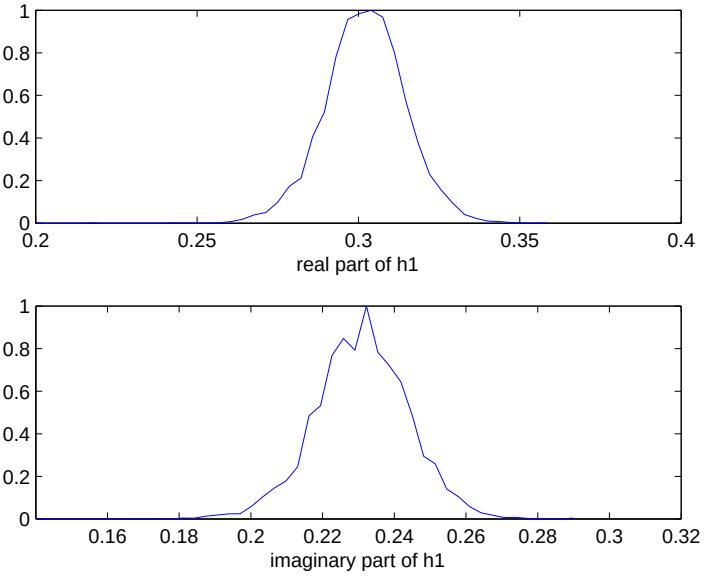
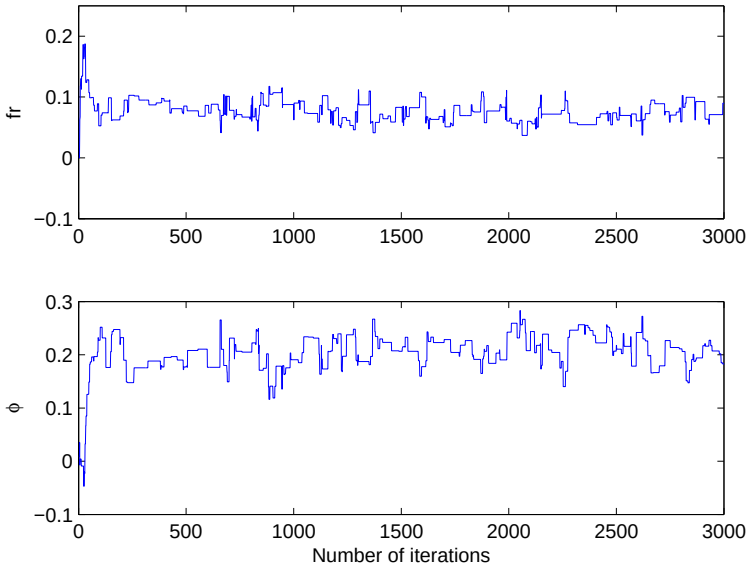
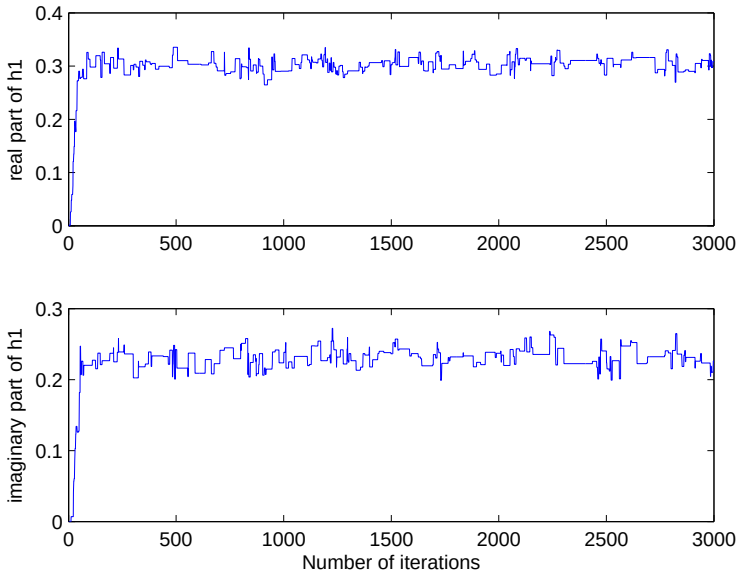


Figure 2.2: Normalized histograms of h_1.

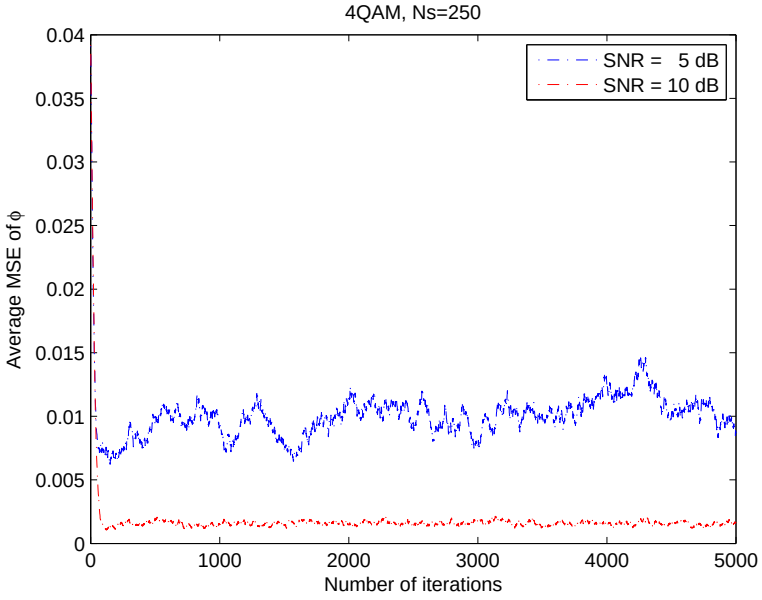


(a) Estimated fr and ϕ.

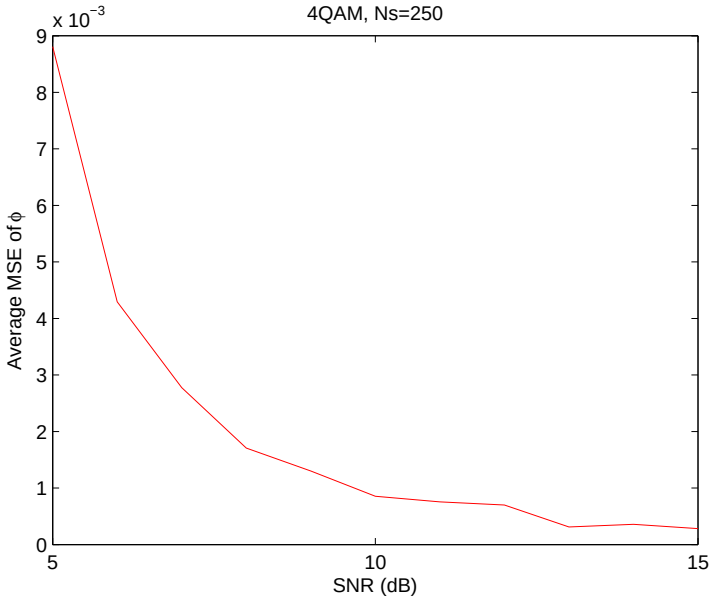


(b) Estimated real and imaginary parts of h_1.

Figure 2.3: MCMC samples

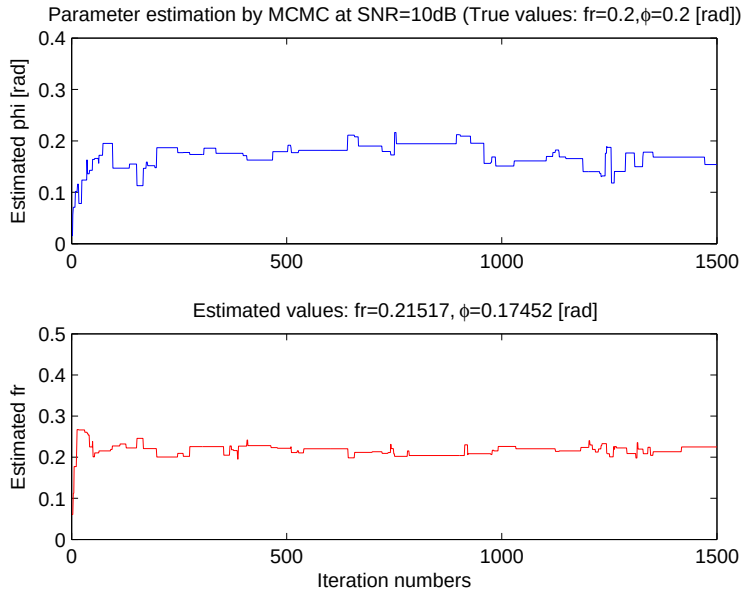


(a) Average MSE of ϕ versus the number of iterations.

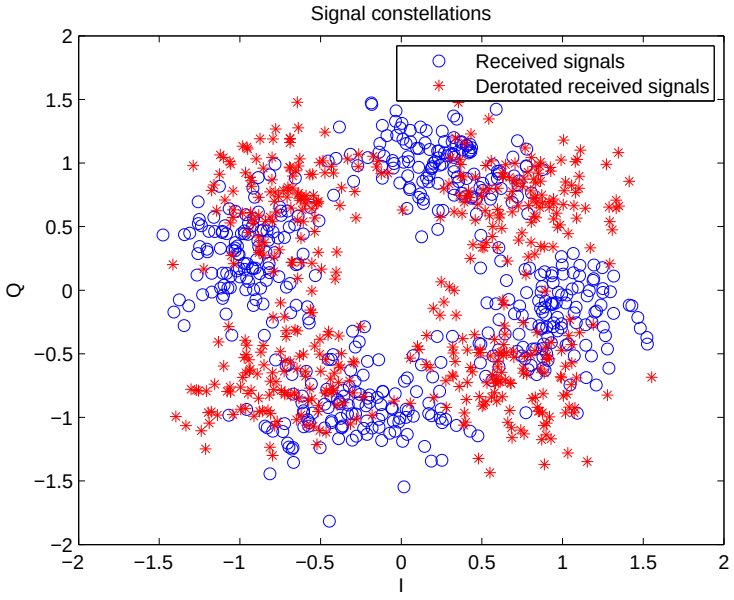


(b) Average MSE of ϕ versus SNR (dB).

Figure 2.4: Convergence of MCMC samples



(a) Estimated ϕ and fr.



(b) Signal constellation before and after impairment mitigation.

Figure 2.5: Impairment mitigation using MCMC methods

N_i is 50000 including burn-in samples. Similar histograms can be produced for h_1 and are plotted in Figure 2.2. Markov Chain convergence of each estimated parameter are shown in Figure 2.3. Here, the number of iteration N_i is 3000. It can be observed that samples reach the target values after hundreds of burn-in samples.

Figure 2.4 (a) shows the average mean square error (MSE) of ϕ versus the number of iterations after 100 Monte Carlo runs for 2 values of SNR. At SNR = 10 dB, it is clearly seen that $N_i = 1000$ samples are sufficient to approximate the MMSE estimator. At SNR = 5 dB, the value of average MSE oscillates around 0.01 till $N_i = 5000$. Figure 2.4 (b) displays the average MSE versus SNR given the iteration number of 1000 and burn-in samples of 500.

Figure 2.5 (a) plots the estimated frequency and phase offsets of 4QAM signal. The number of symbol is 500 and the number of iterations is 1500 including the burn-in of 500 whereas figure 2.5 (b) shows the signal constellation before and after frequency and phase compensation.

MCMC methods have the advantage that the accuracy of the estimation can be improved by increasing the number of samples and iterations. More importantly, choosing the proposal distribution corresponding exactly to the target posteriori distribution is a key success of the accuracy of the methods. Certainly in some circumstances the target posteriori distribution may not be known. MCMC methods can estimate many unknown parameters $(f_r, \phi, h, ...)$. However the complexity and speed of calculation also grow with the number of samples, iterations, and unknown parameters.

2.3 BW Parameter estimation

Assume that we have a perfect synchronization but the signal is transmitted through a dispersive channel. The signal model in (1.7) can be represented in another way as

$$\begin{aligned} x(n) &= \sum_{l=0}^{q} h_l d(n-l) + z(n), \quad n = 1, 2, ..., N_s, \\ &= h s^T(n) + z(n), \end{aligned} \tag{2.8}$$

where

- $d(n) \in \{d_1, d_2, ..., d_M\}$ is an i.i.d. symbol sequence of M-values drawn from one of c constellations denoted $\{\lambda_1, \lambda_2, ..., \lambda_c\}$,
- $s(n) = [d(n), d(n-1), ..., d(n-q)]$,
- $h = [h_0, ..., h_q]$ is a vector containing the $q+1$ taps of the linear finite impulse response (FIR) channel and $h_0, ..., h_q$ are channel coefficients,
- q is the channel memory.

The received signal $x(n)$ can be modeled as a probabilistic function of an hidden state at time n which is represented by a first order HMM with the following characteristics:

1. The state of the HMM at the nth time instant is $s(n)$. Thus, $s(n)$ takes its values in $\{s_1, s_2, ..., s_N\}$ of size $N = M^{q+1}$, where s_j is the jth possible value of $s(n)$.
2. The state transition probability distribution is
$$a_{ij} = P[s(n+1) = s_j | s(n) = s_i],$$
which equals $1/M$ when all symbols are equally likely.
3. The initial state distribution vector $\pi = (\pi_1, ..., \pi_N)^T$ is defined by $\pi_i = P[s(1) = s_i] = 1/N$.
4. Based on (2.8), the pdf of the observation $x(n)$ conditioned on state j, denoted as $p_j[x(n)] \triangleq p[x(n)|s_j]$ can be written
$$p_j[x(n)] = \frac{1}{\pi\sigma_z^2} \exp\left(-\frac{|x(n) - m_j|^2}{\sigma_z^2}\right), \tag{2.9}$$

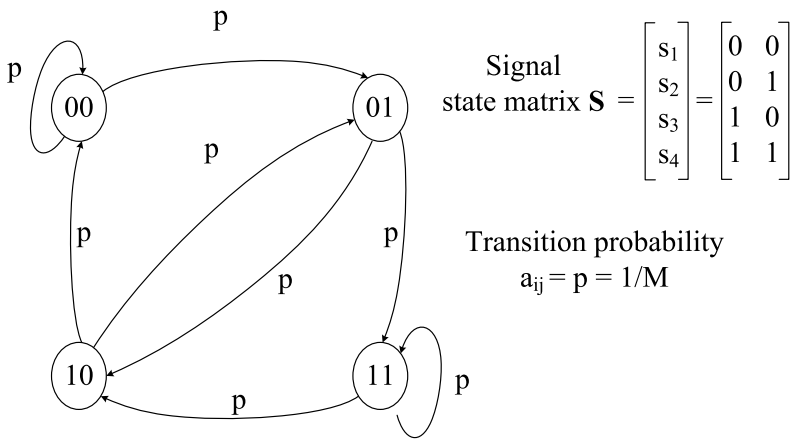


Figure 2.6: Received signal modeled by a HMM.

for $j = 1, ..., N$, where $m_j = \sum_{l=0}^{q} h_l d_j(n-l)$. We denote as $\boldsymbol{m} = [m_1, ..., m_N]^T$ the vector containing all *signal means*.

For example, if M=2, q=1 and all symbols are equally likely, we can draw the state diagram and write the signal state matrix as in Figure 2.6.

Given the above HMM, the BW algorithm can be used to determine the probability of the observation sequence given the model and estimate the unknown model parameters. It is based on a forward-backward procedure which estimates iteratively the unknown model parameters maximizing the posterior probability of the unknown parameters. After convergence, the BW algorithm provides MAP estimates of $\boldsymbol{m}$ and σ_z^2 such that:

$$(\hat{\boldsymbol{m}}, \hat{\sigma}_z^2) = \arg\max_{\boldsymbol{m}, \sigma_z^2} P(\boldsymbol{m}, \sigma_z^2 | \boldsymbol{x}, \lambda). \tag{2.10}$$

The algorithm needs a forward operation to compute $P(\boldsymbol{m}, \sigma_z^2|\boldsymbol{x}, \lambda)$ whereas a forward/backward algorithm is necessary to estimate the unknown parameters m_j and σ_z^2. This section describes the principles of the standard BW algorithm. A least mean square type (LMS-type) update BW algorithm is also discussed.

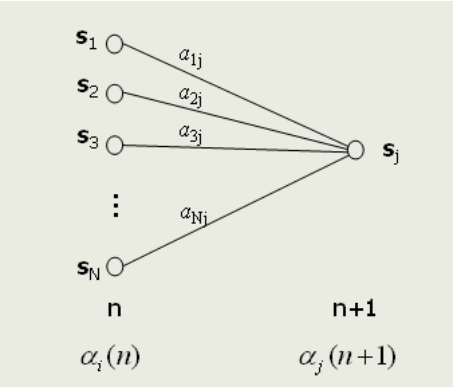


Figure 2.7: Computation of the forward variable.

2.3.1 The standard BW algorithm

The standard BW algorithm [Rab89] estimates $P(\boldsymbol{x}|\boldsymbol{m}, \sigma_z^2, \lambda)$ by using the following three-step procedure iteratively:

1) Compute the normalized forward variable $\alpha_i(n)$, see Figure 2.7

- Initialization:

$$\alpha_i(1) = \pi_i p_i(x(1)), \quad 1 \leq i \leq N \tag{2.11}$$

- Induction:

$$\alpha_j(n+1) = c(n) p_j(x(n+1)) \sum_{i=1}^{N} \alpha_i(n) a_{ij}, \tag{2.12}$$

for $n = 1, 2, ..., N_s - 1$, $j = 1, ..., N$, and where $c(n) = \left(\sum_{i=1}^{N} \alpha_i(n)\right)^{-1}$,

2) Compute the normalized backward variable $\beta_i(n)$, see Figure 2.8

- Initialization:

$$\beta_i(N_s) = c(N_s), \quad 1 \leq i \leq N \tag{2.13}$$

- Induction:

$$\beta_i(n) = c(n) \sum_{j=1}^{N} a_{ij} p_j(x(n+1)) \beta_j(n+1), \tag{2.14}$$

for $n = N_s - 1, ..., 1$ and $i = 1, ..., N$.

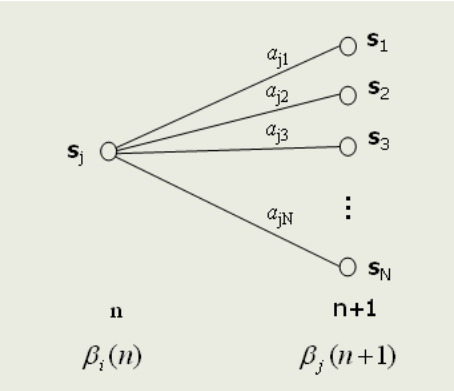


Figure 2.8: Computation of the backward variable.

3) Estimate the model parameters as follows

$$\hat{m}_i = \frac{\sum_{n=1}^{N_s} \gamma_i(n) x(n)}{\sum_{n=1}^{N_s} \gamma_i(n)}, \tag{2.15}$$

$$\hat{\sigma}_z^2 = \frac{1}{N_s} \sum_{n=1}^{N_s} \sum_{i=1}^{N} \gamma_i(n) |m_i - x(n)|^2, \tag{2.16}$$

where $\gamma_i(n) = \alpha_i(n)\beta_i(n)$.

In a batch mode implementation, steps 1 to 3 are carried out iteratively with updated values of $p_j[x(n)]$ until convergence. Thus, the estimated probability of the observation sequence given the model is computed as follows

$$\widehat{P}(\boldsymbol{x}|\boldsymbol{m}, \sigma_z^2, \lambda) = \frac{\sum_{i=1}^{N} \alpha_i(N_s)}{\sum_{i=1}^{N_s} c(i)}. \tag{2.17}$$

Different modifications have been applied to the standard BW algorithm to improve estimation/classification performance or reduce computation complexity. These modifications are presented in Section 2.3.2 and 2.3.3.

2.3.2 Regularization

For a linear channel, we have the relationship $\mathbf{m} = \mathbf{S}h^T$, where $\mathbf{S}$ is the state matrix defined as $\mathbf{S} = [s_1, s_2, ..., s_N]^T$. This information can be used to regularize the estimated

mean values. For this, at each iteration, the estimated means are projected into the space spanned by the columns of $\mathbf{S}$ [FV94]:

$$\mathbf{m} \leftarrow \mathbf{S}\,\mathbf{S}^{\sharp}\mathbf{m} \tag{2.18}$$

where $\boldsymbol{S}^{\sharp}$ is the pseudo-inverse of $\mathbf{S}$ and $\hat{h} = [\mathbf{S}^{\sharp}\mathbf{m}]^T$.

2.3.3 The LMS-type update algorithm

The standard BW algorithm suffers from the "curse of dimensionality" because the computation complexity and memory requirement are proportional to the square of the number of the states. Furthermore the convergence rate is rather slow. Thus, it is worth seeking improvements in terms of memory and computation speed. In this thesis, we have implemented the LMS-update type algorithm initially presented in [FV94]:

$$m_i(n) = m_i(n-1) + \mu_m \gamma_i(n) e_i(n), \tag{2.19}$$

$$\sigma_z^2(n) = (1-\mu_s)\sigma_z^2(n-1) + \mu_s \left(\sum_{i=1}^{N} \gamma_i(n) |e_i(n)|^2 \right), \tag{2.20}$$

where $e_i(n) = x(n) - m_i(n-1)$ for $i = 1, ..., N$.

The initialization and time-induction calculation for the forward variable can be computed as in the standard BW algorithm. The calculation of backward variable can be obtained by using the fixed-lag or sawtooth-lag schemes [KM93]. In this work, we have implemented the fixed-lag case where $\Delta > q+1$ and apply for each n the normalized backward recursion from $n+\Delta$ to n. However, for the normalized backward recursion from $n+\Delta$ to n, the calculation of the normalized forward variable from n to $n+\Delta$ is required for the fixed-lag Δ. This means that the calculation can be started as soon as the observation symbols are greater than $2+\Delta$.

2.3.4 Simulation results

This section studies the convergence and tracking characteristics of the LMS-type update algorithms. The signal-to-noise ratio (SNR) in decibels is defined as

$$\text{SNR} = 10 \log_{10} \left(\frac{|h|^2}{\sigma_z^2} \right).$$

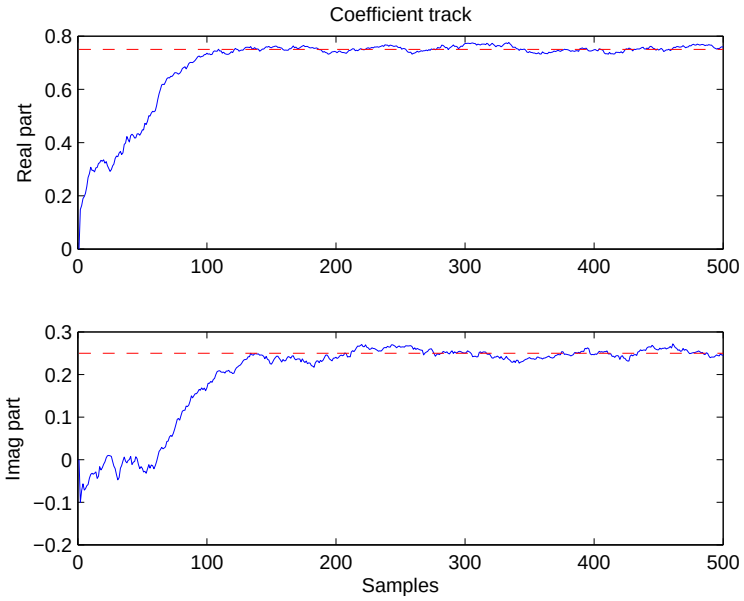


Figure 2.9: Estimated real and imaginary part of h_1.

A 4QAM signal is transmitted through a linear channel whose complex impulse response is $h = [1, 0.75 + 0.25j]$. The output of the filtered sequence is then contaminated by an additive complex white Gaussian noise with variance $\sigma_z^2 = 0.01$. The initial values and the step-sizes of the LMS-type update algorithm have been adjusted as follows:

$$\mu_m = 0.6, \mu_s = 0.1, \sigma_{\text{init}}^2 = 1, \Delta = 5.$$

Figures 2.9 and 2.10 display typical estimates for the real and imaginary parts of h_1 and the variance σ_z^2 for a single run. Figure 2.11 shows the average MSE versus SNR for the estimated real and imaginary parts of h_1. Of course, better performance can be achieved for high SNRs, as expected.

Figure 2.12 (a) shows the received 4QAM signals propagated through a dispersive channel and corrupted by Gaussian noise, *signal means*, and *estimated means* obtained from the modified BW algorithm. It can be seen that the received signal constellation bears no resemblance to the original 4QAM constellation. Figure 2.12 (b) plots the signals after deconvolution of the estimated channel coefficients (calculated from the estimated *means*) with the received signals.

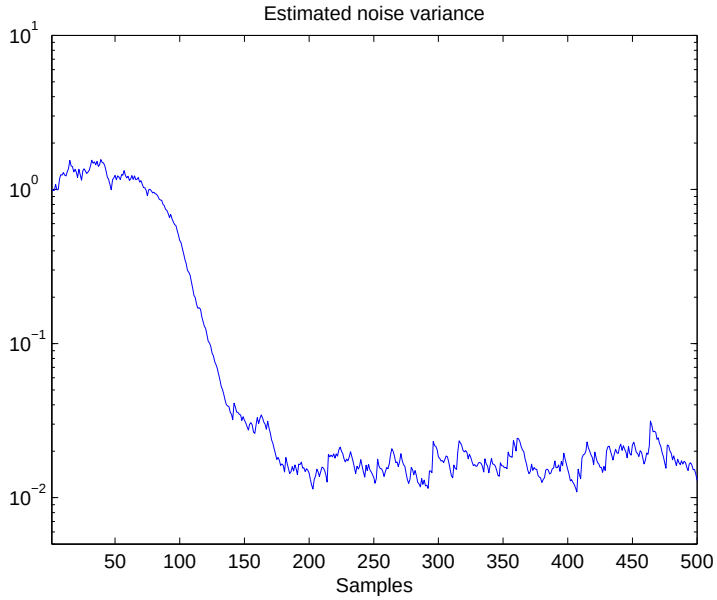


Figure 2.10: Estimated noise variance.

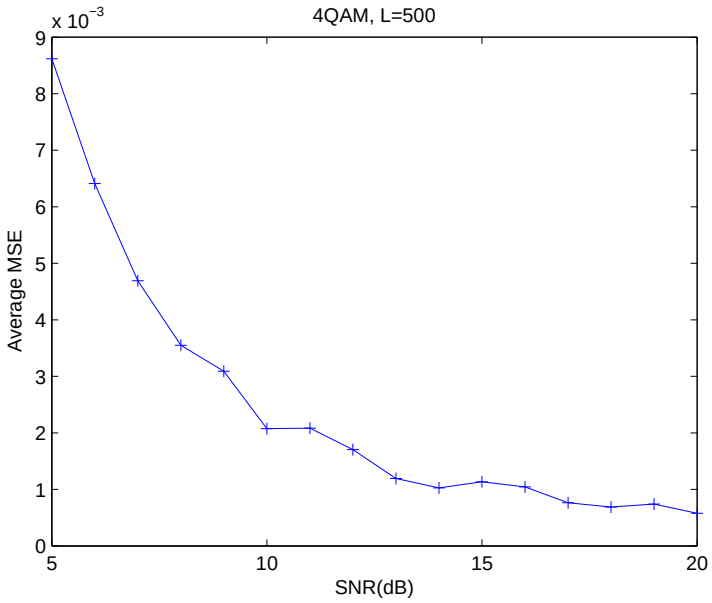
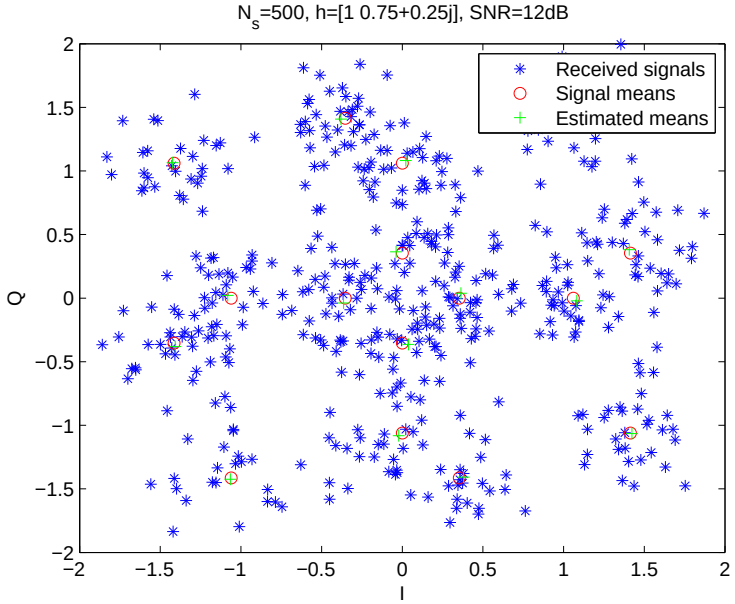
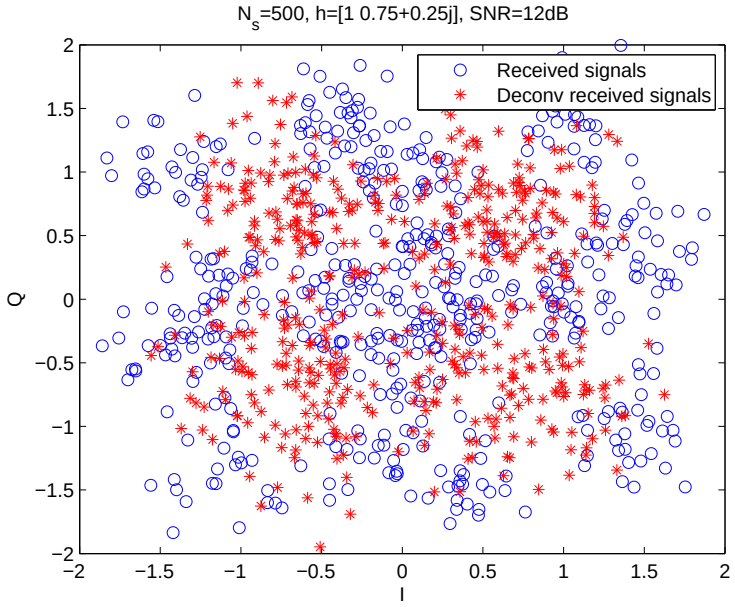


Figure 2.11: Average MSE versus SNR.



(a) Signal *means* and estimated *means*.



(b) Signal constellation before and after impairment mitigation.

Figure 2.12: Dispersive channel mitigation using the modified BW algorithm.

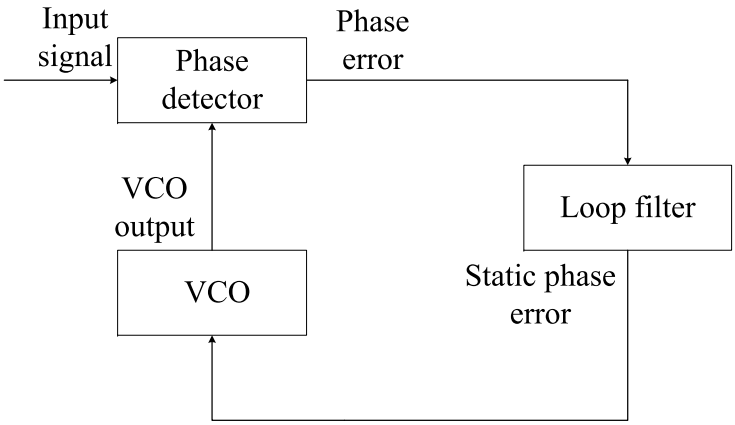


Figure 2.13: Analog phase locked loop.

2.4 Phase-locked loop

A PLL is generally used to synchronize the frequency and phase of a reference or input signal with an output signal which is usually generated by an oscillator. In the synchronized or so-called locked state, the difference (error) between the reference and the oscillator output is zero or very small. The overall circuit consists of three main parts. They are phase detector (PD), loop filter (LP) and voltage control oscillator (VCO), see Figure 2.13. Details on PLLs are rich and can be found in the literature. We are interested in implementing a digital phase-locked loop as in figure 2.14 in which the VCO is replaced by numerically voltage controlled oscillator (NCO).

2.4.1 Signal model

Consider the baseband received signal at the output of the matched filter affected by a phase error θ which is assumed to be constant over N_sT, where T is the symbol duration.

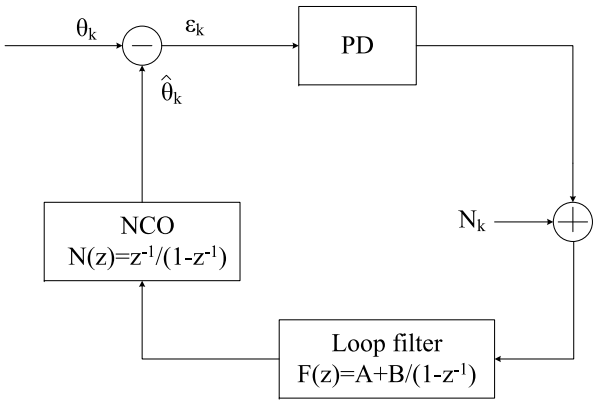


Figure 2.14: A digital phase-locked loop.

Assuming perfect timing recovery, the received samples can be expressed as:

$$x(k) = d(k)e^{j\theta(k)} + z(k), \quad n = 1, 2, ..., N_s \tag{2.21}$$

where $d(k) = d_I(k) + jd_Q(k)$ and $z(n)$ is an i.i.d. complex Gaussian noise sequence which has zero-mean and variance σ_z^2.

Let $\phi = \theta - \hat{\theta}$ be the phase difference between the phase error and the correction phase $\hat{\theta}$ from the VCO output signal. Applying the ML principle, we obtain the logarithm of the likelihood function [Bou97]

$$\Lambda(\phi) = \frac{1}{N_0}\left(\sum_{k=1}^{N_s} \mathrm{Re}[x(k)\hat{d}^*(k)e^{-j\hat{\phi}}]\right). \tag{2.22}$$

2.4.2 Phase detector : Decision-Directed (DD) algorithm

The role of the phase detector is to provide an information representing the phase error. This information will be filtered and integrated in order to generate $\hat{\theta}$ in attempt to make zero error. The DD algorithm uses the estimated data symbol to calculate (2.22). Let ϕ_{opt} be the optimal phase value that makes the derivative of (2.22) with respect to ϕ equal to

error signal	output of VCO
2	increase
0	not change
-2	decrease

Table 2.1: Polarity-type decision-feedback phase detector.

zero, we have

$$\frac{d}{d\phi}[\Lambda(\phi_{opt})] = 0 \quad \Leftrightarrow \quad \mathrm{Im}\left[\sum_{k=1}^{L} \hat{d}^*(k)x(k)e^{-j\phi_{opt}}\right] = 0, \tag{2.23}$$

where $\hat{d}^*(k)$ is the conjugate of the maximum likelihood data estimate. After the phase correction, we can write

$$w(k) = x(k)e^{-j\hat{\theta}}. \tag{2.24}$$

Thus (2.23) can be written as

$$\mathrm{Im}\left[\sum_{k=0}^{L} \hat{d}^*(k)x(k)e^{-j\phi_{opt}}\right] = \sum_{k=1}^{L} \mathrm{Im}[\hat{d}^*(k)w(k)] = 0. \tag{2.25}$$

Let ϵ_t be the error signal before loop filtering and $\epsilon(\phi)$ be the equivalent phase detector characteristic (also called loop S-curve),e.g., $\epsilon(\phi) = E[\epsilon_t(\phi)]$. Thus (2.25) can be expressed as

$$\epsilon_t(\phi(k)) = \mathrm{Im}[\hat{d}^*(k)w(k)]. \tag{2.26}$$

We can write various expressions for the phase detector according to [LV83]. However the author suggested the use of the error signal which is in the form

$$\epsilon_t(\phi(k)) = \mathrm{Im}[\mathrm{csgn}(w^*(k))\mathrm{csgn}(w(k) - \hat{d}(k))], \tag{2.27}$$

where $csgn(x) = sgn(\mathrm{Re}[x]) + jsgn(\mathrm{Im}[x])$.

The error signal in (2.27) provides a polarity-type decision-feedback phase detector. A phase detector applied this error signal outputs three values, i.e., $2, 0$, and -2. The positive signal voltage pulse will be sent to the LP to slow the VCO whereas the negative signal voltage pulse is induced to speed up the VCO. This can be summarize in Table 2.1.

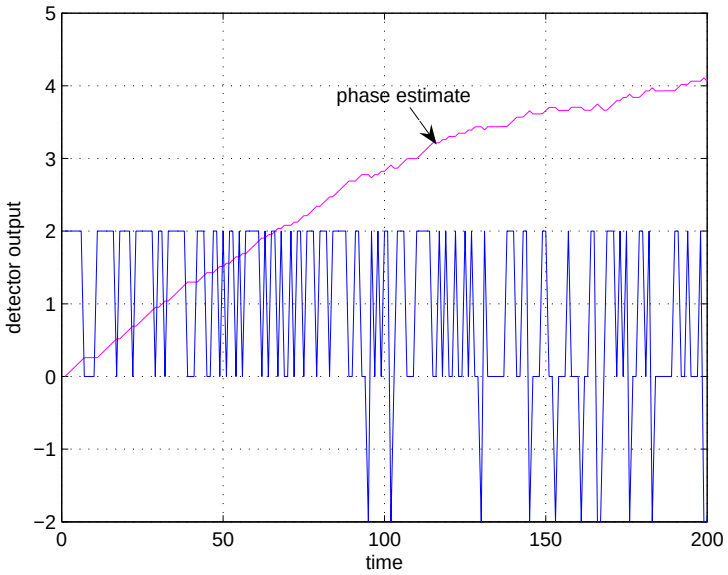


Figure 2.15: 4QAM at 20 dB.

Figure 2.15 plots the phase detector output and the phase estimate. The nature of this error signal has the following comments.

- There is no modulation noise at the equilibrium point $\phi = 0$.
- The implementation requires only EXOR gates and adders.
- This error signal applies to all modulation schemes as long as the quadrant symmetry is satisfied.

In figure 2.16, we compare the equivalent phase detector characteristic of the error signal in (2.26) represented by the solid line to (2.27) represented by the broken line. The signal constellation type is 16QAM operating at SNR = 22dB.

The gain of the detector G_d can be defined by the slope at the origin of its phase detector characteristic. The slope at the origin is an important element to determine the performance of a detector in term of the variance [Bou97]. It is necessary to know the slope in order to calculate the loop noise bandwidth, e.g., two loops with different detectors but the same low-pass filters will have different loop noise bandwidths. The slope at the origin of the

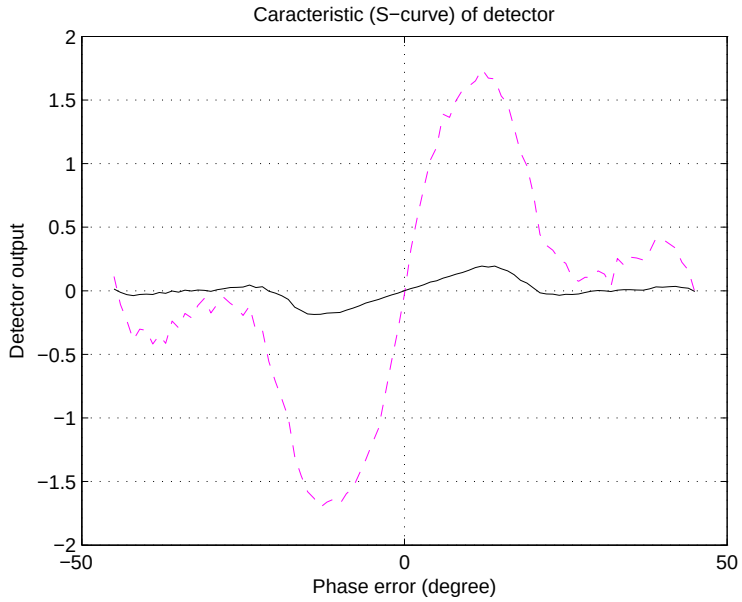


Figure 2.16: The phase detector characteristic of the error signals.

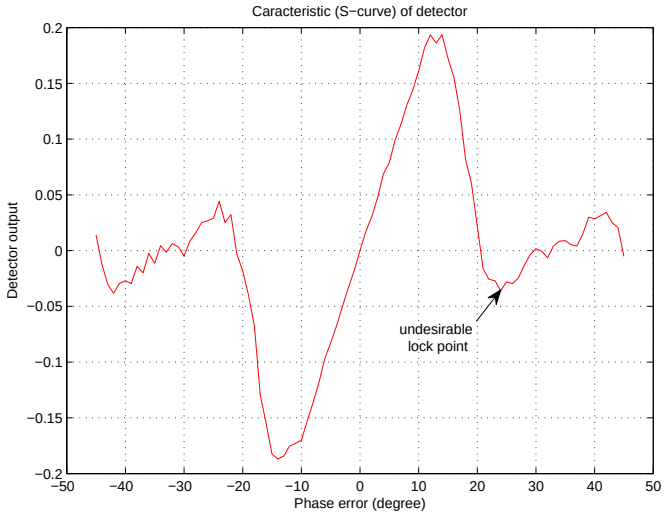


Figure 2.17: The phase detector characteristic of the error signal in (2.26).

detector using (2.27) is greater than that of (2.26). Thus the variance associated with this detector is smaller. Moreover, the area under curve of (2.26) is small causing the increase of the probability of cycle jump. The phase detector characteristic of (2.26) is enlarged in Fig 2.17 which exhibits undesirable lock points. Thus we choose to implement our PLL using the phase detector (2.27).

2.4.3 Loop filter

The choice of the loop filter affects the behavior of the PLL. Generally, the transfer function of the second order loop filter of the form $F(z) = A + \frac{B}{1-z^{-1}}$ is used where $B = 0$ yields the first order loop filter. The transfer function of the second order PLL is given by [LC81]

$$H(z) = \frac{GF(z)}{(z-1) + GF(z)}, \tag{2.28}$$

where G is the overall gain defined by $G = G_d.G_o$ ($G_o = 1$ is the NCO gain). Putting it another way, we obtain

$$H(z) = \frac{GA(z-1) + GBz}{(z-1)^2 + GA(z-1) + KBz}. \tag{2.29}$$

This is equivalent to its analog counterpart

$$H(s) = \frac{2\xi\omega_n s + \omega_n^2}{s^2 + 2\xi\omega_n s + \omega_n^2}, \tag{2.30}$$

where ξ is the damping factor ($\xi = 0.7$ is a common choice) and ω_n is the normalized natural frequency of the loop. The normalize loop noise bandwidth is given by

$$B_l T_{loop} = \frac{\omega_n T_{loop}}{2}\left(\xi + \frac{1}{4\xi}\right), \tag{2.31}$$

where $1/T_{loop}$ is the sampling rate of the PLL. The sampling rate $1/T_{loop}$ is typically chosen to minimize loop tracking error due to thermal noise and input phase dynamics. The phase error variance augments as the value of $1/T_{loop}$ decreases. To avoid threshold degradation, $B_l T_{loop} \lesssim 0.1$ is required [LC81]. The value of $B_l T_{loop}$ is also related to the acquisition time (the transition from a large initial uncertainty to a small steady-state estimation error variance). Widely speaking, a PLL with a small value of $B_l T_{loop}$ has large acquisition time.

By equating the denominator of (2.29) and (2.30), then letting $z = \exp(sT_{loop})$ and applying Taylor expansion of order 2, we obtain

$$\frac{GB}{1 + \frac{GA+GB}{2}} = (\omega_n T_{loop})^2, \tag{2.32}$$

and

$$\frac{GA+GB}{1+\frac{GA+GB}{2}} = 2\xi\omega_n T_{loop}. \tag{2.33}$$

Writing them another way,

$$GA = \frac{(2\xi - \omega_n T_{loop})\omega_n T_{loop}}{1-\xi\omega_n T_{loop}}, \tag{2.34}$$

and

$$GB = \frac{(\omega_n T_{loop})^2}{1-\xi\omega_n T_{loop}}. \tag{2.35}$$

Once $B_l T_{loop}$ is chosen, we can determine the parameter A and B of the loop filter.

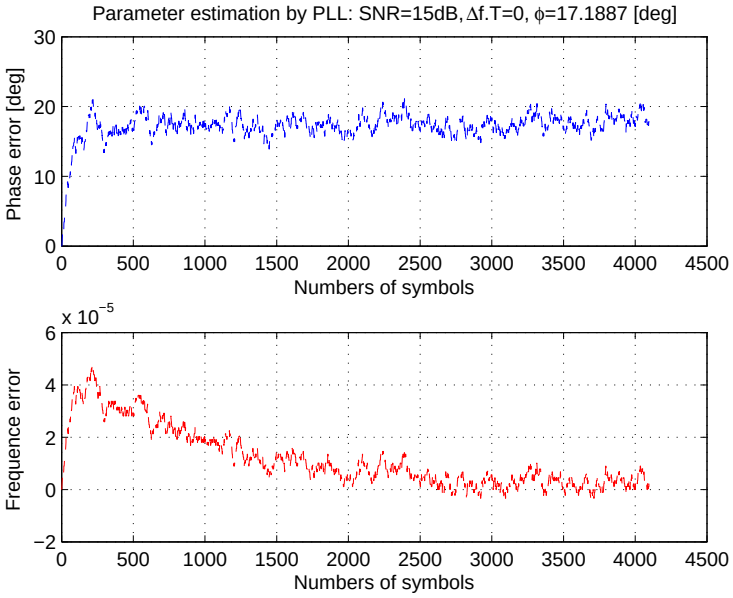
2.4.4 Simulation results

This section presents some results regarding the technique to combat the effects of the frequency and phase offsets using PLL. To simulate continuous transmission, the number of observation symbols of the transmitted 16QAM signals is 4096.

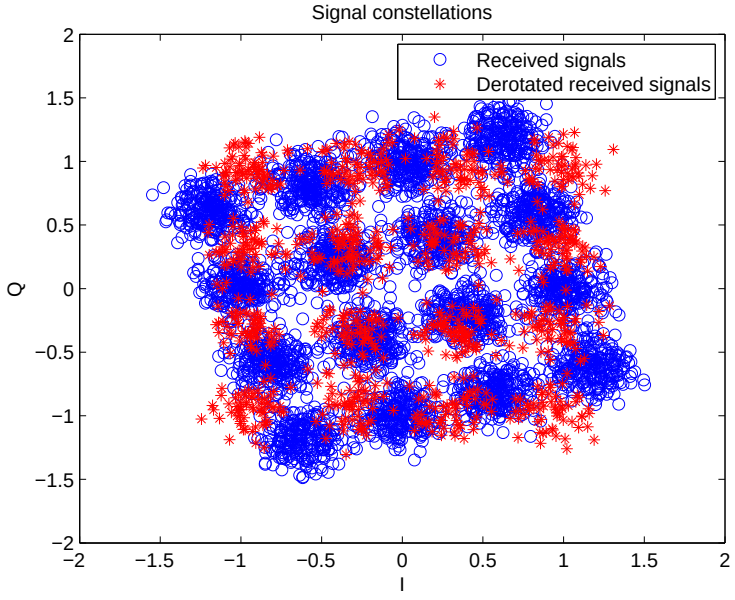
Figure 2.18 (a)-2.19 (a) show the influence of $B_l T_{loop}$ to the error variance of the estimated values in the presence of only the phase offset. In case of $B_l T_{loop} = 0.001$, the error variance is less than that of $B_l T_{loop} = 0.01$ but the acquisition time is longer. To compensate the phase offset, these two cases provide the correct signal constellations, see Figure 2.18 (b)-2.19 (b). When a large frequency offset of 0.02 is introduced, it is more prominent to see the advantage of $B_l T_{loop} = 0.01$. Figure 2.20 shows that the PLL fails to track the frequency offset and the correct signal constellation cannot be recovered. Better results are shown in Figure 2.21.

2.5 Conclusions

Three techniques to mitigate different impairments were addressed. The estimation of the unknown parameters (frequency and phase offset, residual channel) was conducted by simulating samples distributed according to the likelihood under all hypothesises. This operation was achieved by the Metropolis-Hastings algorithm which is one of the most popular MCMC methods. Exploiting the memory structure of ISI channels, which can be modeled by a first order HMM, is the core idea behind the parameter estimation using the modified BW algorithm. Since the computational complexity increases with the number of samples, these

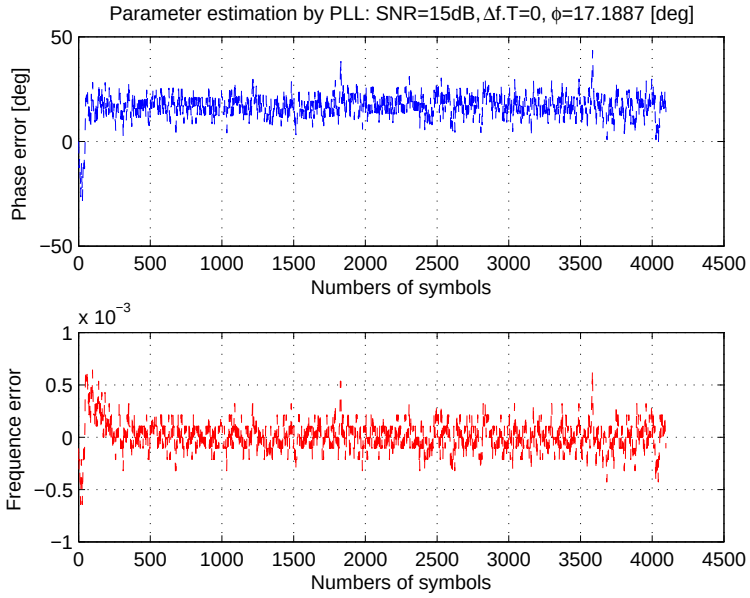


(a) Estimated ϕ and ΔfT.

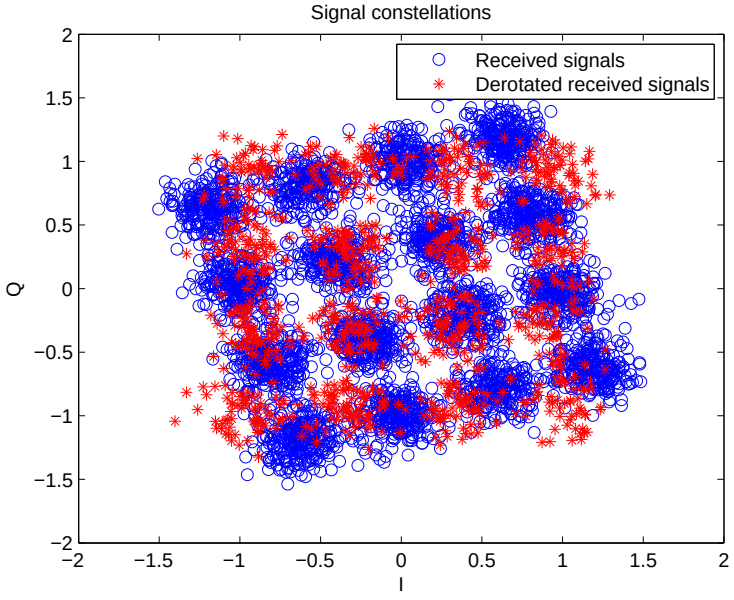


(b) Signal constellation before and after impairment mitigation.

Figure 2.18: Phase impairment mitigation using PLL, $B_l T_{loop} = 0.001$.

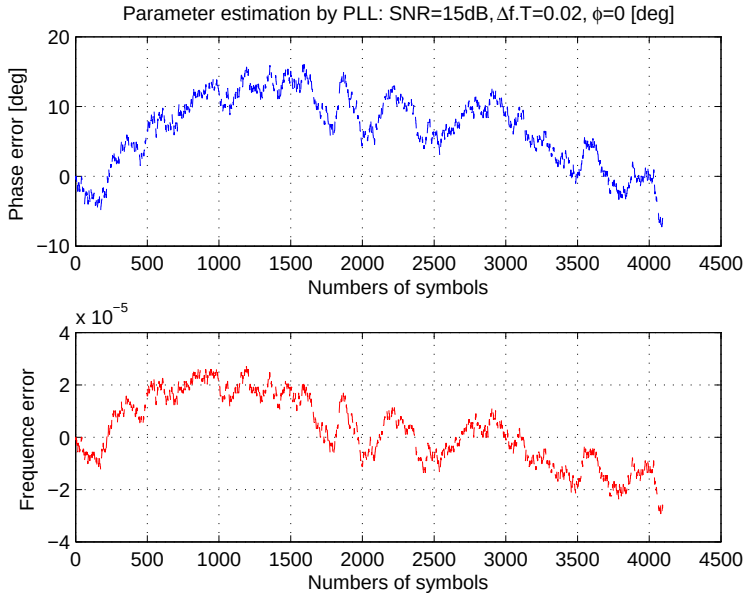


(a) Estimated ϕ and ΔfT.

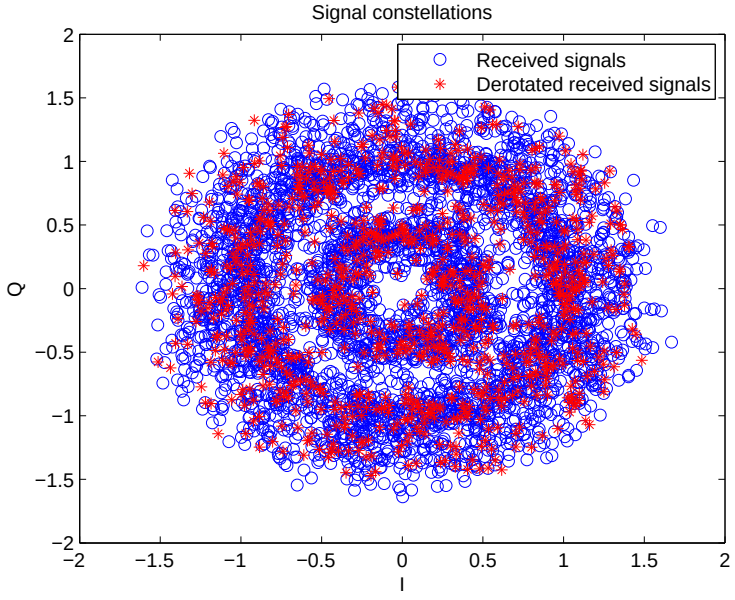


(b) Signal constellation before and after impairment mitigation.

Figure 2.19: Phase impairment mitigation using PLL, $B_l T_{loop} = 0.01$.

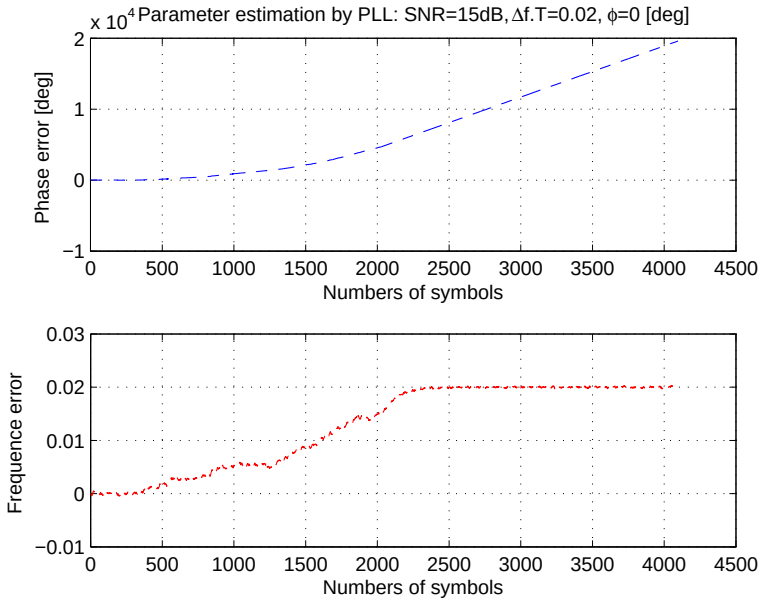


(a) Estimated ϕ and ΔfT.

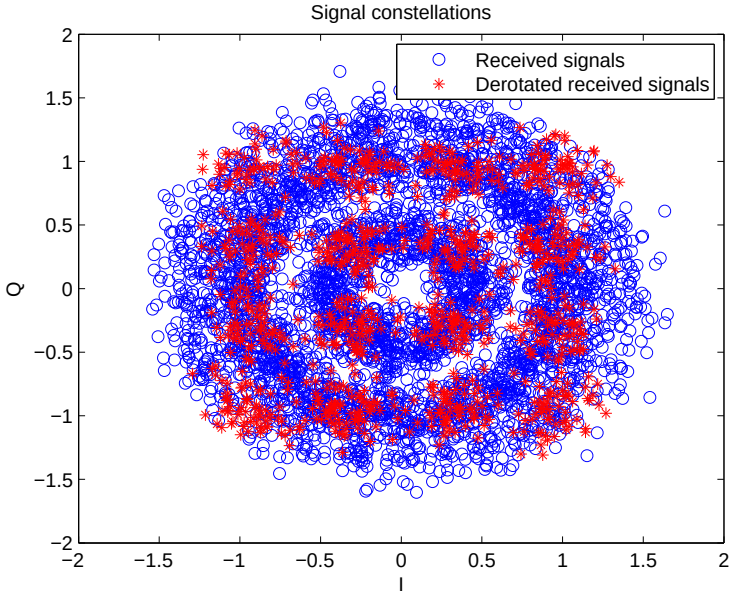


(b) Signal constellation before and after impairment mitigation.

Figure 2.20: Frequency impairment mitigation using PLL, $B_l T_{loop} = 0.001$.



(a) Estimated ϕ and ΔfT.



(b) Signal constellation before and after impairment mitigation.

Figure 2.21: Frequency impairment mitigation using PLL, $B_l T_{loop} = 0.01$.

two feed-forward approaches are suitable for a small packet transmission. The feedback synchronizer is more attractive in case that the transmitted signal of continuous transmission suffers from a large frequency offset.

Chapter 3

Classification of Linear Modulations Using MCMC Methods

Contents

3.1 Introduction

A considerable number of researches has been carried out on modulation classification mitigating only the effect of the additive white Gaussian noise (AWGN). The ML classifier studied by [WM00] reaches the optimal performance. However, it requires knowledge of all parameters such as the phase offset, the frequency offset, or the residual channel. The ML classifier provides an upper bound of performance and can be used as a reference for any classifier that performs under some non-ideal cases. In [PK90], the classifier based on

a quasi log-likelihood ratio (qLLR) was derived by approximating the average likelihood-ratio function of MPSK modulated signals in non-coherent carrier phase and asynchronous timing environments. The same principle can classify BPSK/QPSK/OQPSK signals in the presence of the phase offset [CLP95]. The author in [Sil99] also proposed a pseudo-ML modulation classifier to discriminate BPSK/QPSK/8PSK/16QAM/32QAM/64QAM signals in non-coherent carrier phase case. It was based on approximating the PDF of the phase difference at high SNR. However, none of these classifiers considered the situation where the modulated signal is deteriorated by the frequency offset and a residual channel. In this chapter we study an MCMC plug-in classifier which can recognize the modulated signal format taking into account all the effects mentioned above. Simulation experiments and comparison with the ML and HOS classifiers will be included.

We also extend the proposed MCMC plug-in classifier in a slow fading scenario. In some practical situations, the transmitted signal may propagate through various additional impairment environments including fading. The problem of classifying communication signals in presence of fading has received less attention in the literature. The Bayes classifier was studied in [HH03] to discriminate between BPSK and QPSK modulations. This method has high computation cost since it computes the expectation of the likelihood function over the fading amplitude. Hybrid likelihood-based solutions were proposed in [HH00; ADC+04]. The proposed methodologies consist of estimating the unknown parameters and putting these estimates in the likelihoods. The authors in [HH00] used numerical techniques to find the maximum likelihood estimate of the fading amplitude for BPSK and QPSK modulations. Approximating the fading amplitude and phase by the method of moments (MOM) was also studied in [ADC+04] for QAM modulations.

3.2 Modulation classification in presence of mismatch effects

This section studies the performance of an MCMC plug-in classifier in some realistic scenarios. The signal model is expressed as in (1.7) in which the synchronization errors (the frequency and phase offsets) and the residual channel (from the imperfect equalization) lead to more complex models for the received signals. Given N_s samples $x(k)$, $k = 1, 2, ..., N_s$ of a modulated signal, the classification problem is to identify the underlying modulation λ_i represented by the sample $x(k)$ where $\lambda_i \in \{\lambda_1, \lambda_2, ..., \lambda_c\}$. The plug-in classifier estimates

the unknown parameters $\theta = (f_r, \phi, h)$ by MCMC methods assuming the i^{th} hypothesis is true, then uses these estimates in a likelihood ratio test as if they were correct. The details of MCMC parameter estimation have been reported in Section 2.2. The MCMC plug-in rule is defined as

$$\text{assign } x \text{ to } \lambda_i \text{ if } \hat{l}(x|\theta_i, \lambda_i) \geq \hat{l}(x|\theta_j, \lambda_j), \forall j. \tag{3.1}$$

where

$$\hat{l}(x|\theta_j, \lambda_j) = \sum_{k=1}^{N_s} \ln \left[\frac{1}{M_j} \sum_{i=1}^{M_j} \exp\left(-\frac{1}{\sigma_z^2} \parallel x(k)^{\theta} - S_i \parallel^2 \right) \right],$$

and $x(k)^{\theta}$ is the output of the filter having the transfer function $H^{-1}(z) = 1/\left(\sum_{l=1}^{q} \hat{h}_l z^{-l}\right)$ driven by the input $x(k)e^{-j(\pi k \hat{f}_r/N_s + \hat{\phi})}$.

The four-class problem for linear modulations $\lambda = \{\text{BPSK}, \text{4PAM}, \text{8PSK}, \text{16QAM}\}$ studied in [SS00] is considered in this thesis for comparison. The classification performance is the average probability of correct classification P_{cc}. Various simulation experiments are presented for three classifiers: ML, HOS, and MCMC Plug-in classifiers. Note that the ML classifier in (1.5) and HOS classifier in Section 1.4 do not take into account the mismatch effects. To investigate the average probability of correct classification, simulation experiments were carried out using 1000 trials, i.e., 4000 trials for the four-class problem. The number of symbols in the observation period is $N_s = 250$. All constellation symbols have unit energy. The signal-to-noise ratio (SNR) in decibels is defined as in (2.7). Note that our simulation is based on one sample per symbol, thus SNR per symbol is equivalent to the energy per symbol to noise spectral density ($\text{E}_\text{s}/\text{N}_0$). The MCMC sampler has the following properties:

- Number of burn-in iterations: $N_{bi} = 500$,
- Number of iterations: $N_i = 3000$,
- Proposal distributions: $q(z|\theta_i^n) \sim \mathcal{N}(\theta_i^n, \sigma^2)$ where $\sigma = 0.03$.

3.2.1 Simulation results

Table 3.1 shows confusion matrices for the four-class problem Ω_4 obtained for SNR = 5 dB, $f_r = 0.4$, $\phi = 0$, and $h = [1, 0, 0]$. It is obviously seen that the MCMC plug-in classifier outperforms the ML and HOS classifiers.

Table 3.1: Confusion matrices for three classifiers of 100 trials at SNR = 5dB.

Classifier	Classifier Input	Classifier Output			
		BPSK	4PAM	8PSK	16QAM
ML	BPSK	0	0	99	1
	4PAM	0	0	0	100
	8PSK	0	0	99	1
	16QAM	0	0	18	82
HOS	BPSK	16	84	0	0
	4PAM	11	78	0	11
	8PSK	0	0	83	17
	16QAM	0	0	61	39
MCMC	BPSK	100	0	0	0
	4PAM	24	76	0	0
	8PSK	0	0	100	0
	16QAM	0	0	37	63

Influence of f_r

The robustness of the MCMC Plug-in classifier to the frequency offset is illustrated in Figure 3.1 for an ideal channel (i.e., no residual channel) $h = [1, 0, 0]$, $\phi = 0$, and SNR = 5 dB. Although we were interested in the effect of the frequency offset only for this experiment, we also estimated the values of h. The effect of the phase offset, keeping other parameters constant, is not presented because the HOS classifier is insensitive to the phase offset. Figure 3.2 and 3.3 display the performance of the three classifiers versus the frequency offset in the presence of a three-tap FIR channel $h = [1, 0.25, 0.15]$ for SNR = 0 dB and SNR = 5 dB, respectively. The ML classifier is not sensitive to a small range (i.e.,$f_r \leq 0.2$) of the frequency offset. After that its performance drops rapidly. The HOS classifier is also robust to a small range of the frequency offset but does not perform as good as the ML and MCMC Plug-in classifiers in this range. The performance of the HOS classifier decreases slowly compared to the ML classifier.

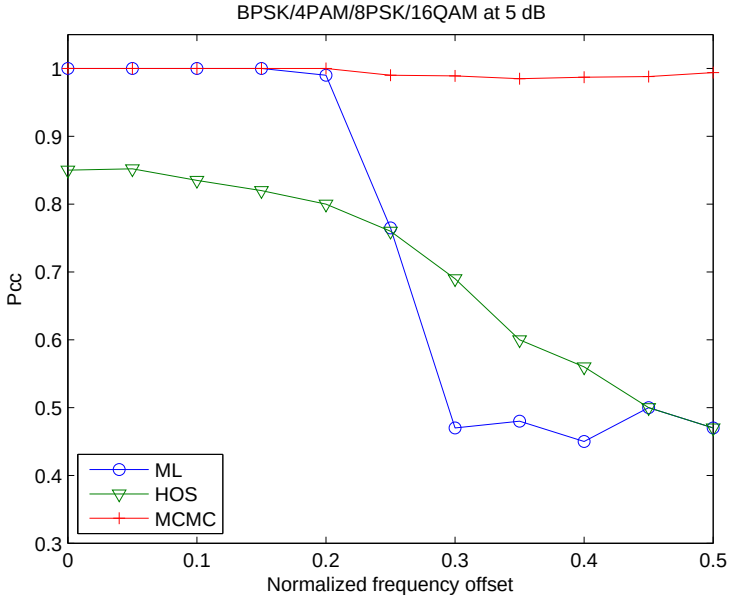


Figure 3.1: Performance versus f_r, $h = [1, 0, 0]$.

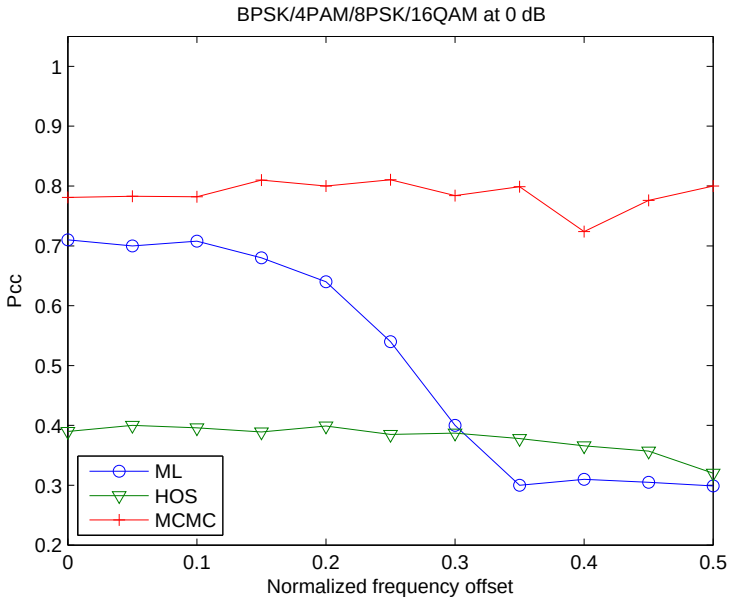


Figure 3.2: Performance versus f_r, $h = [1, 0.25, 0.15]$, SNR $= 0$ dB.

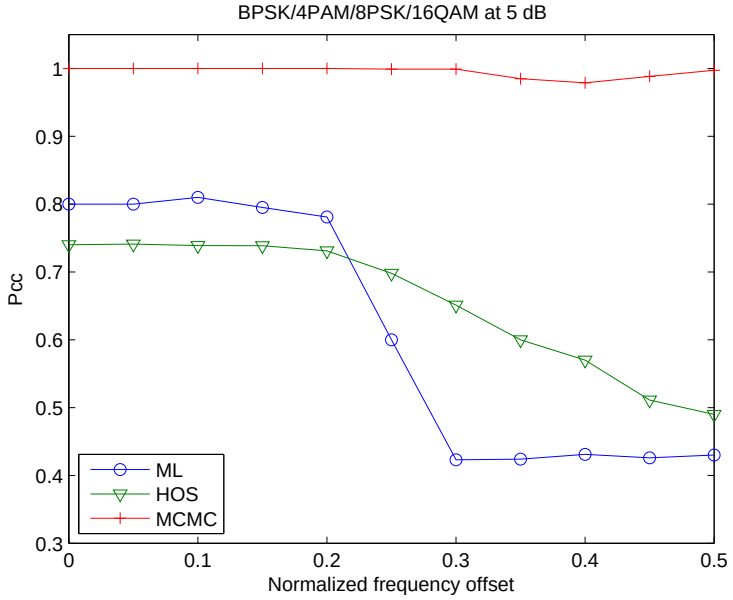


Figure 3.3: Performance versus f_r, $h = [1, 0.25, 0.15]$, SNR $= 5$ dB.

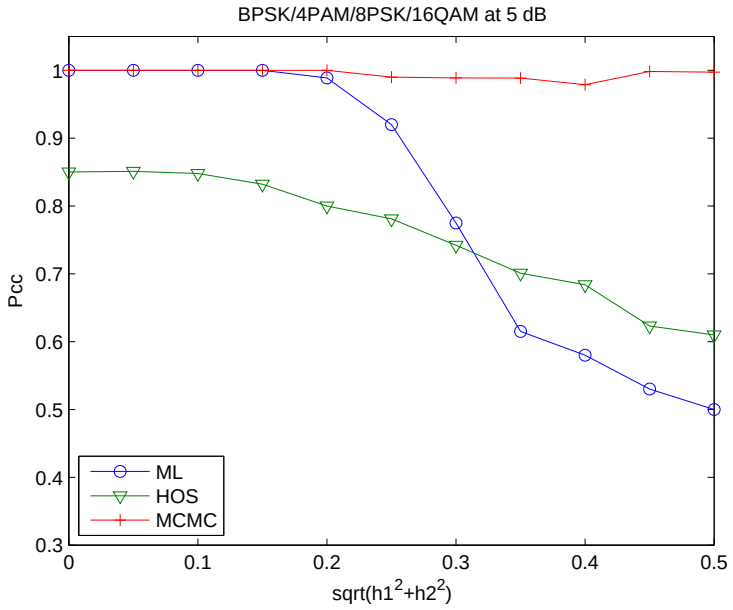


Figure 3.4: Performance versus residual channel modulus (only h is estimated).

Influence of residual channels

Figure 3.4 shows that the MCMC Plug-in classifier is not affected by residual channels, contrary to the ML and HOS classifiers. Recall here that the ML and HOS classifiers assume implicitly no residual channel. This explains their poor performance. The effect of a residual channel can be also seen when comparing Figure 3.1 with Figure 3.3. The ML and HOS classifiers have the same performance behavior trend as a function of f_r in the two figures but with smaller values of P_{cc} in the presence of a residual channel, $h = [1, 0.25, 0.15]$. However, the performance of the MCMC Plug-in classifier is unchanged. The price to pay with this MCMC Plug-in classifier is an increasing of computational cost compared to the ML and HOS classifiers.

3.3 Modulation classification in Rayleigh fading environment

This section extends the classification rule studied in the Section 3.2 for modulations subjected to Rayleigh fading. The proposed strategy is similar to the one developed in [ADC+04] except the unknown parameters (residual carrier frequency, phase offset, and fading amplitude) are estimated by using the MMSE estimator. The numerical problems related to this estimator are circumvented by using MCMC methods. Note that the main novelty of the proposed classification rule with respect to [LTD01] is that the fading amplitude and phase are estimated.

3.3.1 Signal model and assumptions

This work considers a synchronous transmission scheme over a Rayleigh fading channel. This kind of transmission yields residual carrier frequency and phase offsets due to imperfect coherent downconversion. We assume here that there is no residual channel effects and that the amplitude factor is random due to fading as in [HH02; HH03]. However, this study could be extended to more general models including a residual channel and timing errors (as in [VTK02]). After preprocessing, the baseband complex envelope of the received signal sampled at one sample per symbol at the output of a matched filter can be written as:

$$x(k) = \alpha(k) e^{j(\pi \frac{k}{N_s} f_r + \phi(k))} d(k) + z(k), \quad k = 1, 2, ..., N_s, \tag{3.2}$$

where

- N_s is the number of symbols in the observation interval,
- $d(k)$ is an i.i.d. symbol sequence drawn from one of c constellations denoted $\{\lambda_1, \lambda_2, ..., \lambda_c\}$, where λ_j is a set of M_j complex numbers $\{S_1, S_2, ..., S_{M_j}\}$,
- $\phi(k)$ is a phase offset (resulting from fading phase and synchronization errors) having a uniform distribution in the $[0, 2\pi]$ interval,
- $f_r = 2N_s(f_c - \hat{f}_c) \in (-1/2, 1/2]$ is a normalized residual carrier frequency also called frequency offset (f_c is the carrier frequency and $\hat{f}_c$ is the frequency of the local oscillator). Note that these notations imply that f_r is the constellation rotation whose maximum value is $\pi/2$ for $k = N_s$),

- $\alpha(k)$ is the unknown real amplitude factor,
- $z(k)$ is an i.i.d. complex Gaussian noise sequence which has zero-mean and variance σ_z^2 (the real and imaginary components of $z(k)$ are independent and identically distributed).

When the signal is transmitted through a slow flat fading channel, the attenuation factor α can be regarded as a random variable whose pdf is Rayleigh

$$p(\alpha) = \frac{\alpha}{\sigma_\alpha^2} \exp\left(-\frac{\alpha^2}{2\sigma_\alpha^2}\right) \mathbb{I}_{\mathbb{R}^+}(\alpha), \tag{3.3}$$

where $\mathbb{I}_{\mathbb{R}^+}(.)$ is the indicator function on $\mathbb{R}^+$ (i.e. $\mathbb{I}_{\mathbb{R}^+}(\alpha) = 1$ if $\alpha > 0$ and 0 else).

Fading channel

Flat fading mobile radio channels are usually characterized by the following frequency response:

$$S(f) = \frac{1}{2\pi f_d} \left[1 - \left(\frac{f}{f_d}\right)^2\right]^{-1/2} \mathbb{I}_{[-f_d, f_d]}(f). \tag{3.4}$$

The output of this channel can be generated by filtering a complex white Gaussian sequence with a low-pass Butterworth filter. The cutoff frequency of this filter is the product of the symbol duration T by the Doppler shift f_d due to vehicle motion. It is possible to generate slow or fast fading channels, depending on the value of f_dT. As an example, Fig. 3.5 shows the output of the Butterworth filter for $f_dT = 0.01$ (Top) and $f_dT = 0.001$ (Bottom), where $T = 1$. These figures clearly show that a large (resp. small) value of f_dT induces fast (low) fading amplitude variations.

The simulations performed in this work have been obtained for $f_dT = 0.001$. In this case, the fading amplitude α and phase ϕ are different in each symbol resulting from the variations as a function of time. However, for a slow fading, they can be assumed approximately constant for each group of $N_s = 100$ consecutive symbols $d(k)$. We can rewrite (3.2) as:

$$x(k) = \alpha e^{j(\pi \frac{k}{N_s} f_r + \phi)} d(k) + z(k), \quad k = 1, 2, ..., N_s. \tag{3.5}$$

At the receiver, the estimated values of parameters $\hat{\alpha}$, $\hat{\phi}$ obtained for MCMC methods represent the averaging of the fading amplitude and phase over 100 consecutive samples. This point is illustrated on the bottom of figure 3.5 which compares the real fading amplitude with its piecewise constant estimated value.

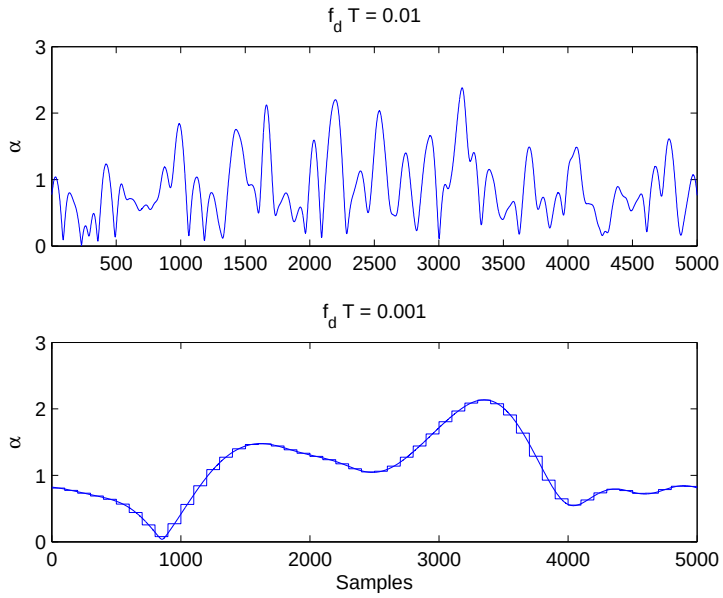


Figure 3.5: Fading amplitude versus time.

3.3.2 MCMC plug-in classifier

This section studies a plug-in classifier for classifying digital modulations subjected to Rayleigh fading. We assume that the received signal amplitude and phase vary from one observation interval to another and are unknown to the receiver. This assumption is realistic in a slow fading context and has been used in [HH03]. In this case, the classifier has to mitigate the amplitude and phase changes to yield good classification performance. One solution to this problem is to assign some prior distribution to fading amplitude and phase, then integrate out these parameters from the likelihood. However, this strategy yields classification rules with exponential implementation complexity. An alternative is to estimate the unknown parameters and then replace the unknown parameters by their estimates in the likelihood. This plug-in rule strategy has shown good classification properties in fading environment [ADC+04]. This section studies the MCMC plug-in classifier which estimates the phase offset, residual carrier frequency and fading parameters for modulation classification purposes. Denote as $\theta = (f_r, \phi, \alpha)$ the unknown parameter vector. The plug-in rule is

defined as follows:

$$\text{assign } x \text{ to } \lambda_i \text{ if } \hat{l}(x|\theta_i) \geq \hat{l}(x|\theta_j), \forall j = 1, ..., c, \tag{3.6}$$

where $\hat{l}(x|\theta_j)$ is the logarithm of the likelihood associated to class λ_j (whose constellation consists of M_j symbols $S_1, S_2, ..., S_{M_j}$)

$$\hat{l}(x|\theta_j) = \sum_{k=1}^{N_s} \ln \left[\frac{1}{M_j} \sum_{i=1}^{M_j} \exp \left(-\frac{1}{\sigma_z^2} \parallel x(k) - w_i \parallel^2 \right) \right], \tag{3.7}$$

and

$$w_i = \hat{\alpha} S_i e^{j(\pi \frac{k}{N_s} \hat{f}_r + \hat{\phi})}.$$

The plug-in rule can be used as soon as estimates of the unknown parameter vector θ can be obtained conditionally upon each class λ_j. A Bayesian estimation technique was studied in [LTD01] to estimate the unknown phase offset, carrier frequency and residual channel in absence of fading. The method can be extended to signals subjected to fading. More precisely, the unknown parameter vector $\theta = (f_r, \phi, \alpha)$ is estimated conditionally to each possible class λ_j using MCMC methods, see more details in Section 2.2. Note that the Markov chain state space and current state are denoted by Ω and $\theta^n = (f_r^n, \phi^n, \alpha^n) \in \Omega$, respectively.

Reducing the computational complexity

The acceptance probability (2.5) depends on the pdfs $p(y|x)$ and $p(\theta^n|x)$ whose computation requires to evaluate summations of logarithm functions. This operation can be easily and efficiently conducted on MATLAB. However, in practical applications where a Digital Signal Processor (DSP) has to be used, the evaluation of a logarithm function is too expensive. Therefore, an approximation version which reduces the calculation cost is preferable. The approximation is obtained from the following well-known relation:

$$\ln(e^{a_1} + e^{a_2}) = \max(a_1, a_2) + \ln(1 + e^{-|a_1 - a_2|}) \tag{3.8}$$

or it can be written as

$$\ln \sum_j e^{a_j} \simeq \max_j a_j \tag{3.9}$$

By applying this result to $a_i = -\frac{1}{\sigma_z^2} \parallel x(k) - w_i \parallel^2$, the following result can be obtained

$$\begin{aligned}\hat{l}(x|\lambda_j) &= \sum_{k=1}^{N_s}\Big[-\ln M_j + \ln\left(\sum_{i=1}^{M_j} e^{a_i}\right)\Big],\\ &\simeq -\ln M_j - \sum_{k=1}^{N_s}\max_i a_i,\\ &\simeq -\ln M_j - \frac{1}{\sigma_z^2}\sum_{k=1}^{N_s}\left(\max_{i=1}^{M_j} \parallel x(k) - w_i \parallel^2\right). \end{aligned} \tag{3.10}$$

This last expression reduces the computational cost required to evaluate the likelihood. The corresponding loss of performance is not critical in most simulations that have been conducted. This point will be illustrated in Section 3.3.4.

3.3.3 Method of moments

Estimating the parameter vector θ can be made by using the method of moments as in [ADC+04]. The phase offset is approximated by

$$\hat{\phi} = \frac{1}{M}\arg\left(\sum_{k=1}^{N_s} x(k)^M\right), \tag{3.11}$$

where M is the number of points in the constellations for the MPSK signal and $M = 4$ for the QAM signal. The amplitude can be found from

$$\hat{\alpha}^2 = \sqrt{\frac{M_{42} - 2M_{12}^2}{k_2 - 2}}, \tag{3.12}$$

where $M_{mn} \triangleq E\left[x^{m-n}(x^*)^n\right]$, $k_2 = 1$ for BPSK, 4QAM, 8PSK and $k_2 = 1.32$ for 16QAM.

3.3.4 Simulation results

Simulations have been carried out to evaluate the performance of the plug-in classifier. This section focuses on a four-class problem $\lambda = \{\text{BPSK}, \text{QPSK}, \text{8PSK}, \text{16QAM}\}$ which has already been considered in the literature [SS00]. All constellations have been normalized (unit energy) yielding the signal-to-noise ratio (SNR) defined as in (2.7). The MCMC sampler has the following properties:

- Current state: $\theta^n = (f_r^n, \phi^n, \alpha^n)$.
- Proposal distribution: random walk $q(y|\theta^n) \sim \mathcal{N}(\theta^n, \sigma^2)$ where $\sigma = 0.03$.
- Uninformative independent uniform priors for the frequency and phase offsets : $p(f_r, \phi) = p(f_r)p(\phi)$ where $p(f_r) = I_{(-1/2,1/2]}(f_r)$, $p(\phi) = I_{[0,2\pi]}(\phi)$, and I is the indicator function.
- Prior knowledge about α is defined in (3.3).
- Number of burn-in iterations: $N_{bi} = 500$,
- Number of iterations: $N_i = 1000$,

Parameter estimation

This section illustrates the performance of the MCMC-based MMSE estimator summarized in Section 3.3.2. The unknown parameter vector θ has been estimated on each burst of 100 symbols by running a Markov chain with 1000 samples including 500 burn-in samples (i.e., the first 500 samples generated by the MH algorithm have not been used for the estimation). The simulation has been conducted for a BPSK constellation with a signal to noise ratio SNR = 5 dB. Moreover, the residual carrier frequency is constant ($f_r = 0$ without loss of generality), the random phase ϕ is uniformly distributed on the interval $[-\pi/4, \pi/4]$ and the fading amplitude is distributed according to a Rayleigh distribution (see 3.3.1 for more details). The actual values of the unknown parameters (continuous lines) and the corresponding estimates (circles) for fading amplitude (top), phase offset (middle), and residual carrier frequency (bottom) are depicted on Fig. 3.6. Comparison to the MOM estimation at SNR $= -2$ dB is plotted in Figure 3.7. These results clearly show the accuracy of the proposed estimation methodology particularly at slow SNRs and in deep fade.

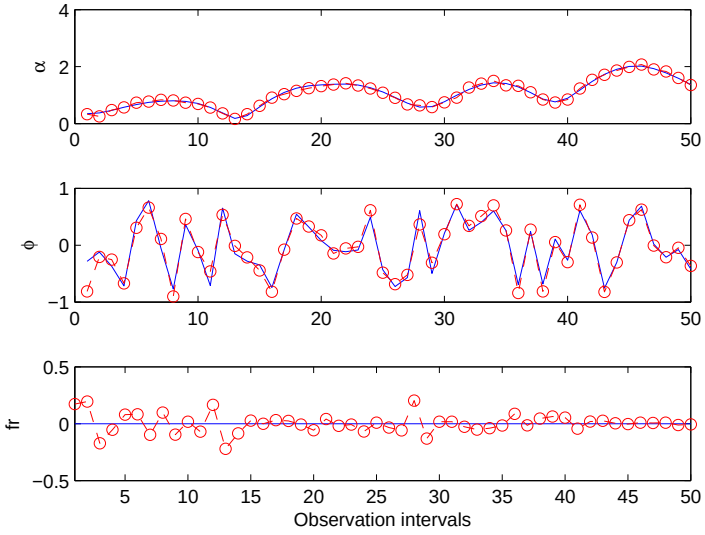


Figure 3.6: Unknown parameters and their estimates.

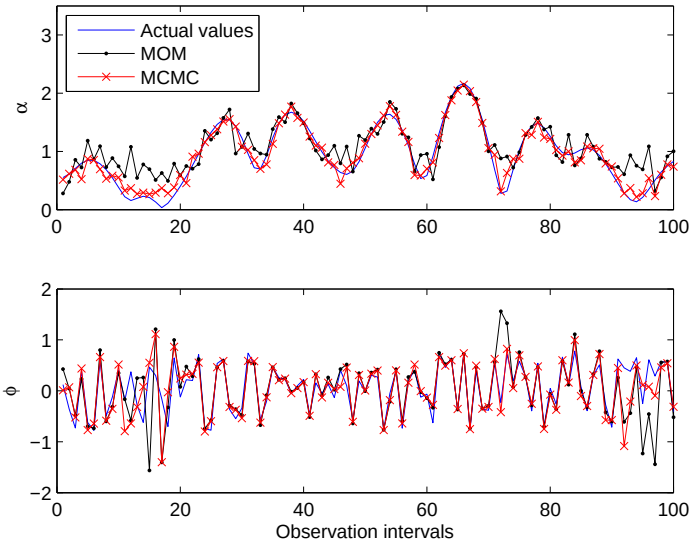


Figure 3.7: Comparison between the MOM and MCMC estimation methods.

Performance versus SNR

The first simulation results depicted on Figure 3.8 compare the average probability of correct classification for different classifiers as a function of SNR:

- the circle curve corresponds to the ML classifier (labeled Ref) which assumes the fading amplitude and phase of each symbol are known,
- the star curve is obtained for the MCMC plug-in classifier (labeled MCMC),
- the diamond curve stands for the classifier derived in [ADC+04] (labeled MOM).

The simulation scenario is similar to the examples of Figure 3.6-3.7 except that the uniform phase offset is in $[0, 2\pi]$ interval. Note again that the ML classifier cannot be implemented in practical applications since it assumes that the fading amplitude and phase are perfectly known. Thus, it provides an upper bound of classification performance. Figure 3.8 shows that the MCMC plug-in classifier outperforms the MOM classifier in a slow flat fading scenario. The figure also shows that the average probability of correct classification for the MCMC plug-in and MOM classifiers approaches the optimal one provided by the ML classifier for high SNRs.

Figure 3.9 shows the probability of correct classification of the MCMC plug-in classifier for each candidate modulation (BPSK, QPSK, 8PSK, and 16QAM). This figure indicates that modulations with large numbers of constellation points (8PSK and 16QAM) are more difficult to classify than modulations with small numbers of points (BPSK, QPSK) for the same SNR.

Performance versus f_r

Figure 3.10(b) and 3.10(a) show the effect of frequency offset (due to inaccuracies of the local oscillators) on classification performance for SNR = 10 and 15 dB. When the frequency offset is less than 0.2, the MCMC classifier still performs reasonably well. The classification performance drops very slightly at the frequency offset $f_r = 0.3$ and tends to degrade much further. However, it is important to note that the MCMC-based classifier is more robust to frequency offset than the MOM classifier particularly for $f_r > 0.2$.

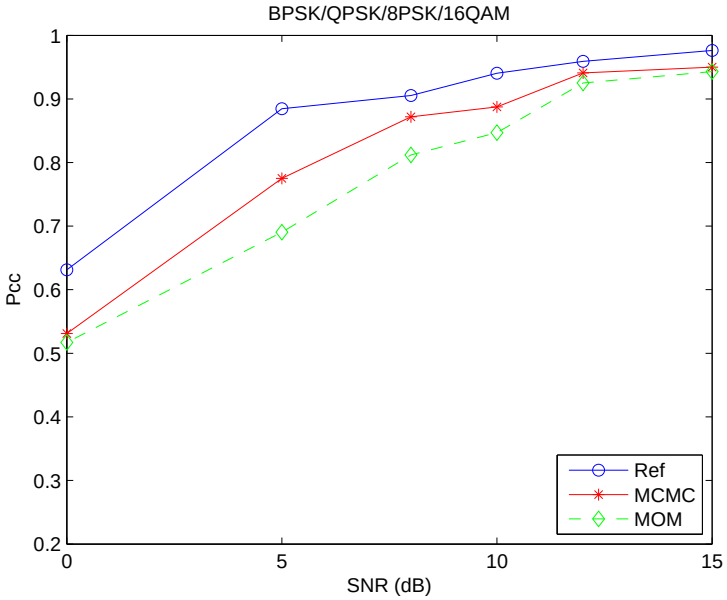


Figure 3.8: Performance versus SNR.

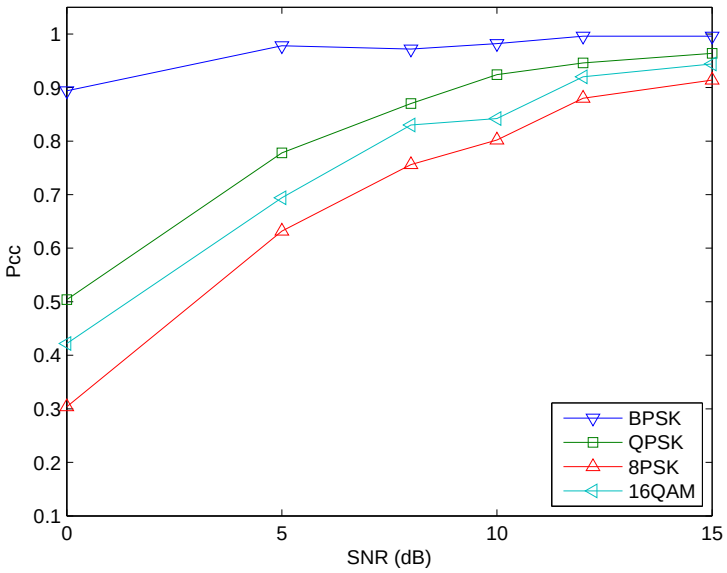
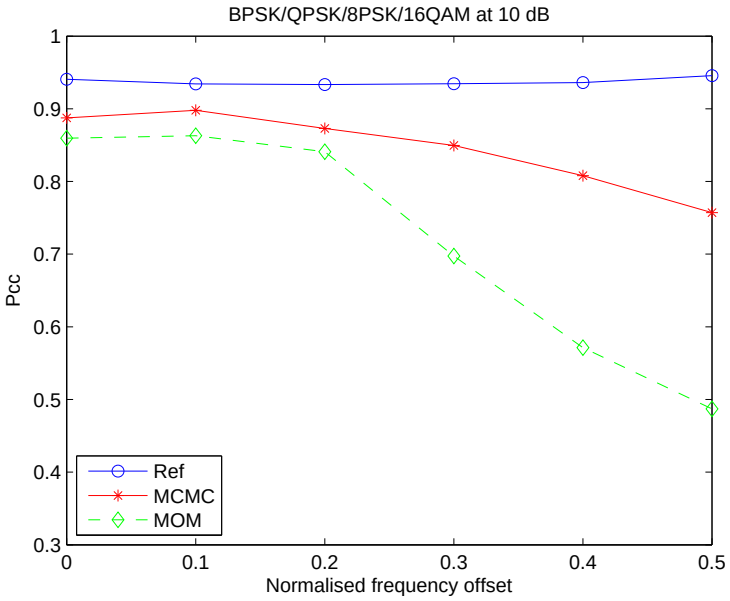
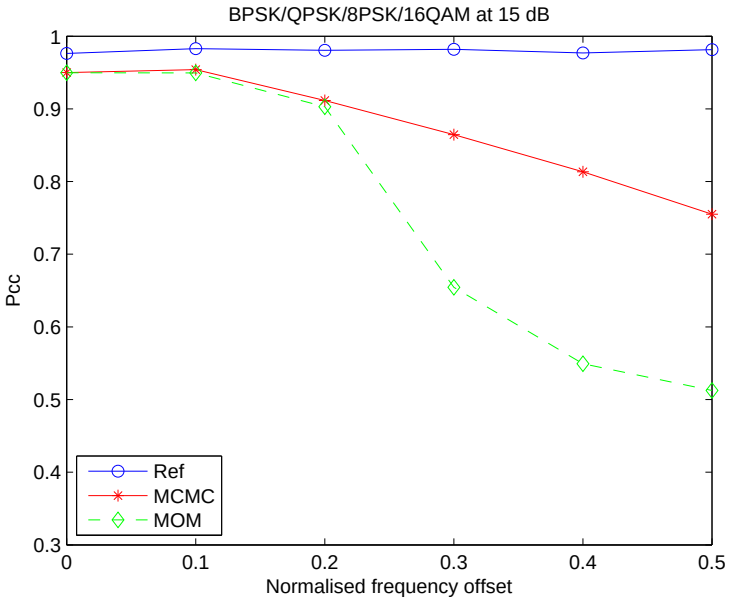


Figure 3.9: Probability of correct classification versus SNR.



(a) SNR = 10 dB



(b) SNR = 15 dB

Figure 3.10: Performance versus f_r in a slow flat fading scenario for different SNRs.

Approximated classification rule

The last simulation results illustrate the performance of the approximate MCMC classifier which uses (3.10) instead of (3.7). Figure 3.11 compares the estimated posterior distributions of the residual carrier frequency f_r obtained by using the exact (dotted line) and approximate (continuous line) MCMC samplers. The number of burn-in iterations N_{bi} for this example is 500 and the posteriors have been estimated by using the 2500 last Markov chain samples. The two distributions are clearly in good agreement, showing that the approximate MCMC sampler can be used if the computational cost of the algorithm is an important issue.

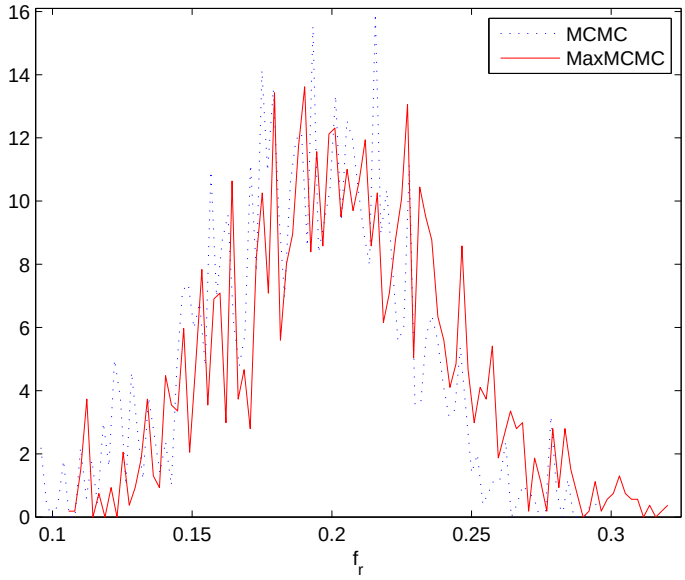


Figure 3.11: Estimated posteriors for $f_r = 0.2$.

3.4 Conclusions

The application of MCMC sampling to modulation classification in the presence of frequency offset, phase offset, and residual channel was presented. The proposed MCMC plug-in classifier estimated the unknown parameters characterizing the model mismatch. The estimates were then plugged into the class-conditional densities. The proposed classifier outperformed ML and HOS classifiers. Furthermore, the proposed strategy was extended to the problem

of digital modulation classification in a Rayleigh fading environment and showed good performance. Reducing the computational complexity of the proposed classifier is an important problem which will make the classifier more attractive and useful. It would be interesting to work on this problem in future studies.

Chapter 4

Classification of Linear Modulations using the BW Algorithm

Contents

4.1 Introduction

For classification purpose, our main interest is to determine the posterior probabilities of the received signals conditionally to each class. We propose to use the BW algorithm to compute these probabilities which are then plugged into the optimal Bayes decision rule. Applications of the BW algorithm to three different scenarios are studied in this chapter.

- Firstly (Section 4.2.2), we apply our proposed methodology to classify the offset quadrature phase shift keying (OQPSK) from QPSK signals in an AWGN channel.

We then compare our proposed classifier with the qLLR classifier studied by [CLP95] to discriminate BPSK/QPSK/OQPSK modulation formats.

- Secondly (Section 4.2.3), the BW algorithm in conjunction with PLLs are applied to the classification problem that includes the frequency offset. The role of PLL is to compensate the phase and frequency offsets. The BW algorithm computes the posterior probabilities of the signal sequence conditionally to each class after the compensation. We will study the advantage of using PLLs to mitigate large frequency offsets.

- Lastly (Section 4.3), modulation recognition in the presence of intersymbol interference (ISI) is investigated. In a non-cooperative scenario, the classification of digitally modulated signals propagating through an ISI environment has been studied by many researchers [LP95; BSSS00; LTD01]. However, it still presents a great deal of issues. Indeed, without some kind of ISI mitigation, the performance of current classification techniques designed for AWGN channels degrades significantly. All of the proposed classifiers addressed in this chapter use the classification rule defined in the following subsection.

4.1.1 Classification rule

The proposed classifiers apply the following classification rule

$$\text{Assign } \boldsymbol{x} \text{ to } \lambda_i \text{ if } \widehat{P}(\boldsymbol{x}|\lambda_i) \geq \widehat{P}(\boldsymbol{x}|\lambda_j), \forall j = 1, ..., C, \tag{4.1}$$

where $\widehat{P}(\boldsymbol{x}|\lambda_i) \triangleq \widehat{P}(\boldsymbol{x}|\boldsymbol{m}, \sigma_z^2, \lambda_i)$ is obtained from the BW algorithm (see (2.17)). Note that the whole sequence (of length N_s) is required to estimate $\widehat{P}(\boldsymbol{x}|\lambda_i)$ even if the online LMS-update type algorithm has been used for the the computation of $m_i(n)$ and $\sigma_z^2(n)$. Note also that the observation length N_s required to properly identify the modulation constellations should be greater than the maximum number of states $N_{\max} = M_{\lambda_c}^{q+1}$ (i.e. $N_s > N_{\max}$) so that every possible state can be reached by the algorithm.

4.2 Modulation classification in AWGN channels

An AWGN channel is a communication channel that can be modeled as the linear addition of white noise (with a constant power spectral density $\frac{N_0}{2}$ W/Hz) whose the amplitude has

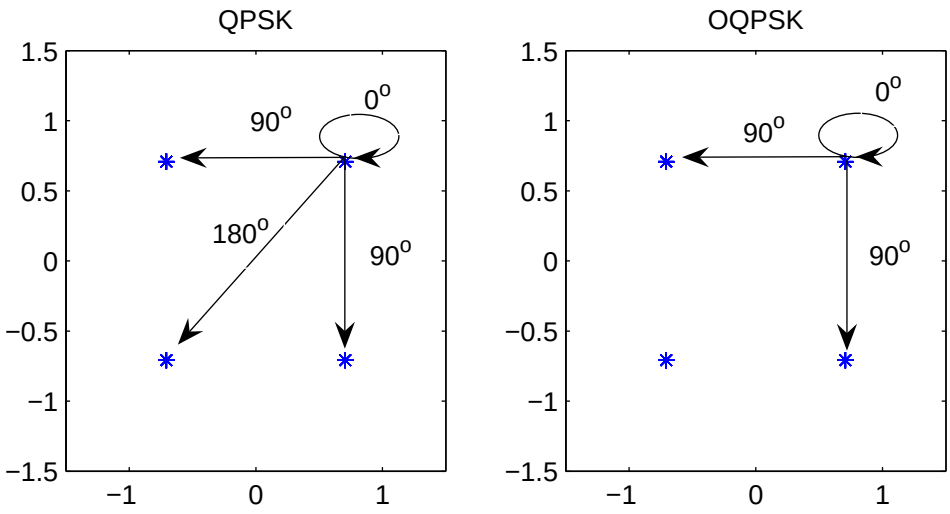


Figure 4.1: Constellations and phase changes of QPSK and OQPSK.

a Gaussian distribution. The model does not take into account of fading, nonlinearity or dispersion., etc.

The OQPSK modulation will be included in the dictionary. In satellite transmitters, OQPSK signals are less sensitive to spectral sidelobe spreading than QPSK signals, thus the out-of-band interfering due to band limiting and the nonlinearity of the amplifier is decreased. OQPSK is effectively the same as QPSK except that the I- and Q-channel pulses are offset in time by $T/2$ seconds. Unlike QPSK signal whose the phase changes at the symbol boundaries can be 0^o, $\pm 90^o$, and 180^o, the phase changes at the symbol boundaries of OQPSK signal can only be 0^o and $\pm 90^o$, see Figure 4.1. This property can be exploited via the BW algorithm to discriminate OQPSK signals from QPSK signals even though they have the same constellations. The fact that some transitions are not allowed is also useful to discriminate between $\frac{\pi}{4}$-QPSK and 8PSK signals using the same methodology. The constellations and phase changes of $\frac{\pi}{4}$-QPSK and 8PSK signals are illustrated in Figure 4.2. Note that in the figure we show all possible phase changes but 0^o for 8PSK.

4.2.1 Simulation results: ideal case

This section studies a three-class problem $\lambda = \{\text{BPSK}, \text{QPSK}, \text{OQPSK}\}$. This simulation considers that BPSK, QPSK, and OQPSK signals have a common baud-time (defined as the

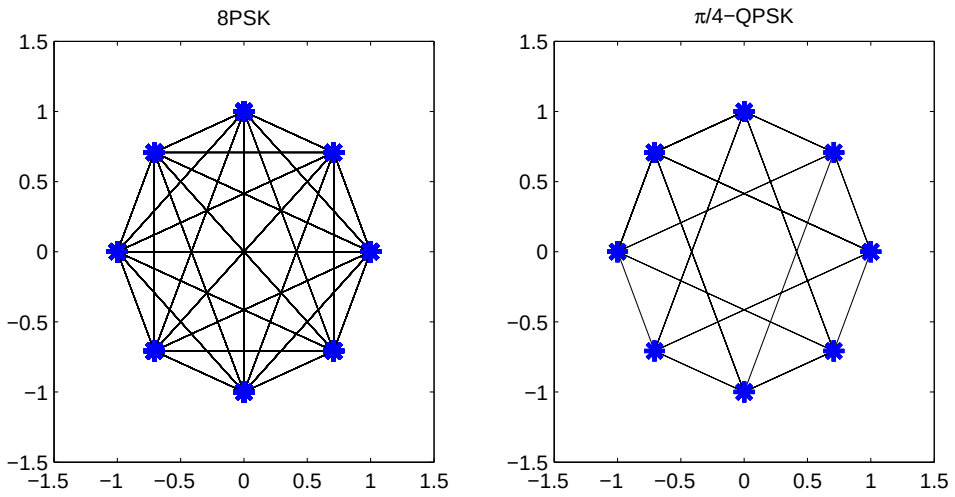


Figure 4.2: Constellations and phase changes of 8PSK and $\pi/4$-QPSK.

minimum time between data transition). Also the same observation interval is simulated as it is more realistic. Thus the number of symbols of BPSK and QPSK signals is twice the number of symbols of OQPSK signals, e.g., $N_{QPSK} = 2N_O$. The SNR in this experiment is defined as the SNR per bit. It is required to adjust the value of the LMS step-size parameter μ_m for each constellation. The values of μ_m used in this research have been obtained by minimizing the average MSE of the estimated parameters. The following results have been obtained: $\mu_m = 0.3$ for BPSK, $\mu_m = 0.6$ for QPSK and OQPSK. Figure 4.3 shows the performance comparison of our strategy with the method proposed by Chugg and Polydoros [CLP95]. The authors proposed a qLLR classifier to identify BPSK/QPSK/OQPSK modulation types. Due to difficulty in setting thresholds resulting from the approximation of ALRT, the classifier works in two stages. First, it distinguishes between {OQPSK} and {BPSK, QPSK}. If the received signal is not OQPSK type, then it classifies between BPSK and QPSK formats. Note that there is no phase offset in this simulation. The threshold of the qLLR classifier in Figure 4.3 is an *ideal threshold* obtained by maximizing P_{cc} over a large number of data and noise realizations. Although this threshold setting is not practical, it gives the best performance for the qLLR classifier. It is obviously seen that our classifier outperforms the qLLR classifier.

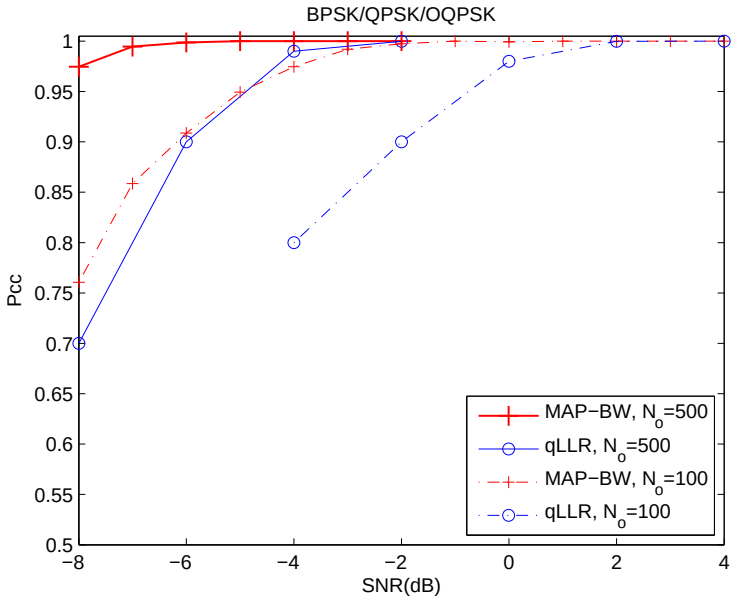


Figure 4.3: Average probability of correct classification versus SNR.

4.2.2 Simulation results: phase offset

The effect of a phase offset to our classifier is also investigated. The phase offset represents a synchronization error of the local oscillator at the receiver and is obtained by rotating the constellation with an angle ϕ. Figure 4.4 plots the performance of the classifier with (estimate σ_z^2 and $\mathbf{m}$) and without (estimate σ_z^2 only) the phase estimation at SNR $= 0$ dB and $N_O = 100$. From the figure, our proposed classifier is robust to the phase errors. The classification performance for different SNRs versus ϕ for $N_O = 100$ is illustrated in Figure 4.5.

4.2.3 Simulation results: large frequency offset

This section addresses an application of BW algorithm in conjunction with PLLs, as shown in Figure 4.6, to modulation classification when the received signal is affected by a large frequency offset. The role of PLL is to mitigate the phase and frequency impairments. After the phase and frequency correction, recognition is achieved by the same strategy as in Section 4.2. The baseband received signal in (2.21) can be rewritten as

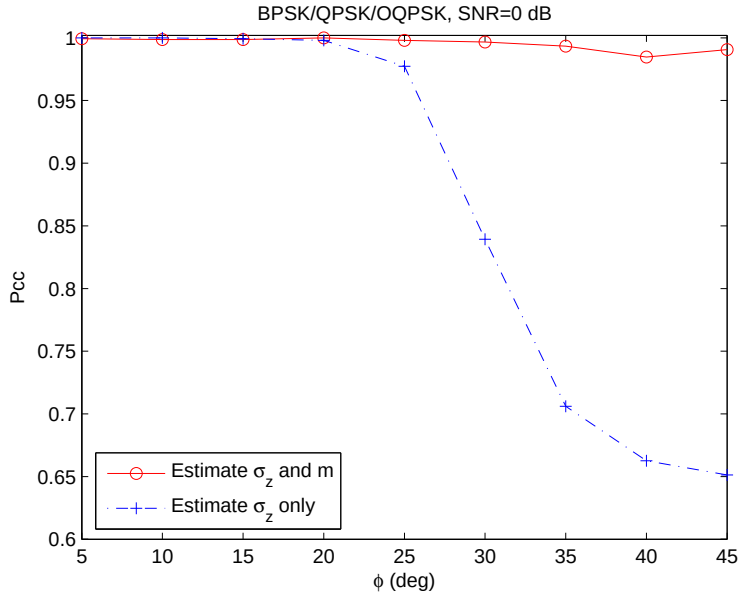


Figure 4.4: Average probability of correct classification versus ϕ.

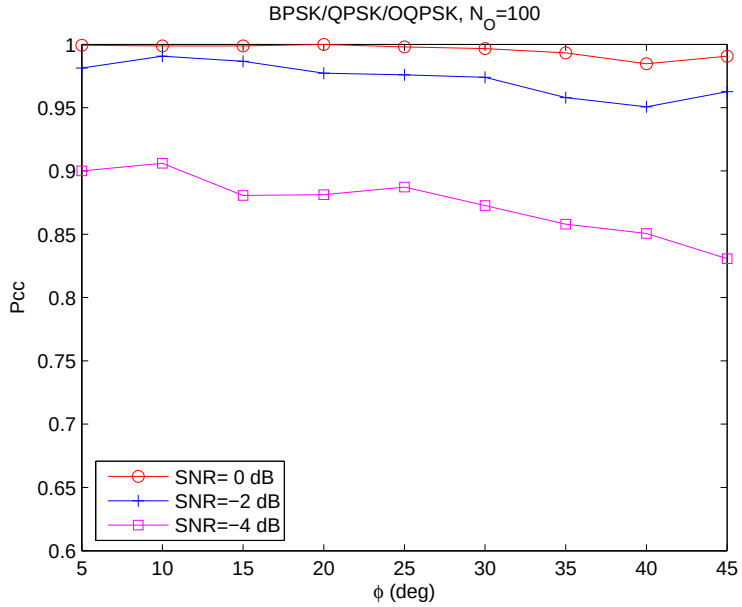


Figure 4.5: Average probability of correct classification versus ϕ for three values of SNR.

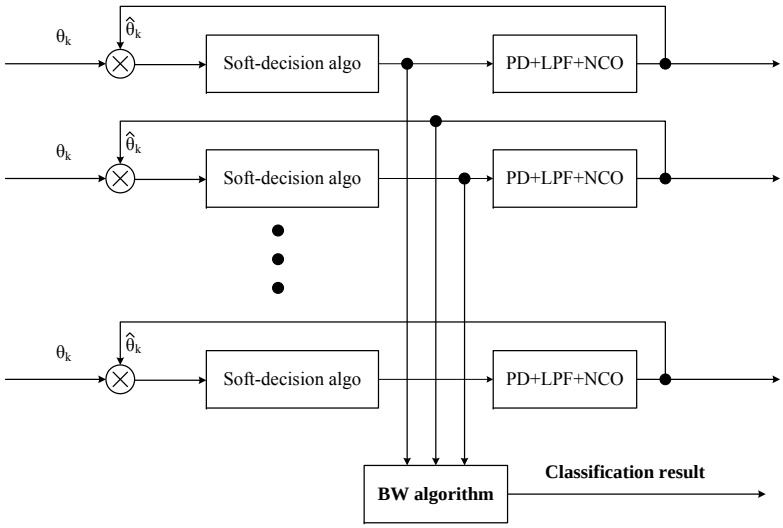


Figure 4.6: Proposed classifier for large frequency offset.

$$x(k) = e^{j(2\pi\Delta fT+\phi)}d(k) + z(k), \quad k = 1, 2, ..., N_s \tag{4.2}$$

where $\Delta fT = (f_c - \hat{f}_c)T$ is the frequency offset. In the situation where the frequency offset is very small $\Delta fT \ll 1$ or the maximum value of constellation rotation is $\pi/2$ for $k = N_s$, the feed-forward frequency estimation presented in Section 2.2 and 2.3 can be used. However, the two estimators cannot cope with large frequency offsets, $\Delta fT \geq 0.01$. This motivates us to use PLLs to counteract the synchronization errors.

We study the classification performance of the proposed classifier for a five-class problem $\lambda = \{\text{BPSK}, \text{QPSK}, \text{OQPSK}, \text{8PSK}, \text{16QAM}\}$. The number of samples is 4096 (for OQPSK $N_o = 2048$). The PLLs are implemented as explain in Section 2.4. The value of $B_l T_{loop}$ employed in this experiment is 0.01. The classification rule in (4.1.1) is applied for the last 200 samples to identify the transmitted signal format. This is based on the assumption that, after thousands of samples, one of the PLLs corresponding to the transmitted modulation format is locked with very high probability. Figure 4.7 plots the average probability of correct classification against SNR for three values of frequency offset. The performance is much deteriorated when $\Delta fT = 0.02$. In this case the average probability of correct classification

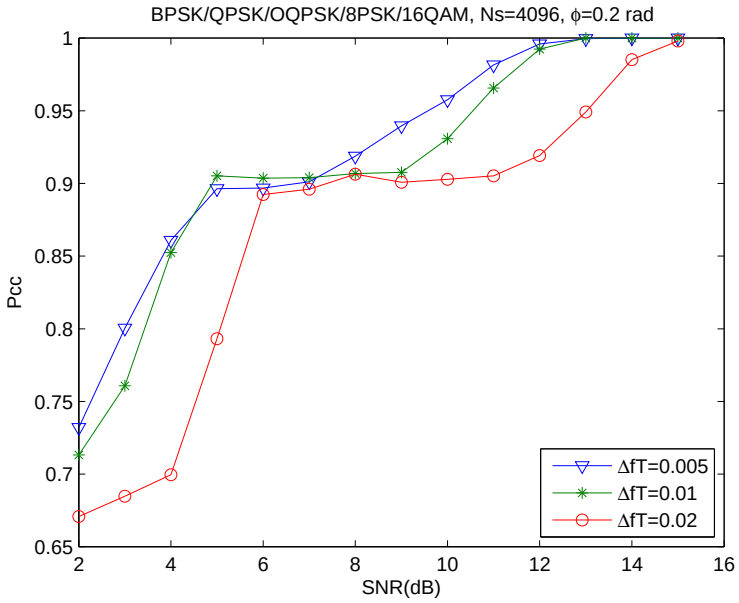


Figure 4.7: Average probability of correct classification versus SNR.

does not even increase when the SNR changes from SNR = 5 to 10 dB. This result is confirmed by the probability of correct classification (P_c) for each modulation format in Figure 4.8. It can be noticed that BPSK, QPSK, OQPSK and 8PSK signals can be identified with 100% of confidence starting from SNR = 6 dB. However, the proposed classifier requires high values of SNR (15 dB for this experiment) to obtain perfect classification when 16QAM signals are included.

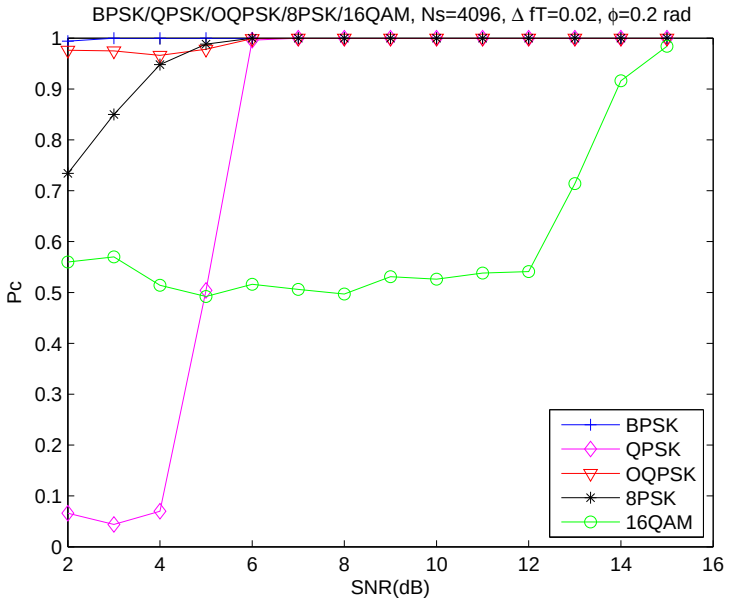


Figure 4.8: Probability of correct classification versus SNR.

4.3 Modulation classification in unknown ISI channels

A communication channel that has a nonideal frequency response characteristic causes the signal amplitude and delay distortion. As a result, the overlapping of successive pulses occurs and the peaks of the pulses are no longer distinguishable. This phenomenon is called intersymbol interference (ISI). Transmission of a signal through a band-limited channel also introduces ISI. In a band-limited channel case, ISI can be avoided using the transmitting and receiving filters that satisfy Nyquist criterion.

This section studies the performance of the digital modulation classifier based on HMMs in the presence of ISI channels. The classifier estimates HMM posterior probabilities as well as model parameters by using the forward/backward BW algorithm [Rab89]. The estimated posterior probabilities are then used for classification via the usual MAP rule. Here, we are interested in classifying linear modulation types transmitted through an unknown finite memory channel and corrupted by AWGN. As a result, our main goal is to determine the posterior probabilities that the received communication signal corresponds to modulation types belonging to a known dictionary. However, channel coefficient estimates can also be

obtained as side results.

The proposed classifier will be compared with the per-survivor processing (PSP) technique introduced in [LP95]. This technique estimates the data sequence and the unknown parameters of a communication signal, and classifies this communication signal by using the generalized likelihood ratio test (GLRT). It tackles the problem by using the PSP to estimate the channel coefficients and the data sequence in order to calculate the test statistic. Note that the classification thresholds of this method have to be determined empirically depending on the operating SNR. Also, PSP requires good initialization. Another practical approach which might be used for comparison is based on constant modulus and alphabet-matched algorithms followed by cumulant-based classifiers [BSSS00]. However, this technique relies on the performance of blind equalizers which usually operate at high SNRs. Furthermore, the decision after the cumulant-based classifier requires one to measure the erratic behavior of the cumulant estimates, which could be dubious and complicated.

Our assumptions regarding the operating system and the signal model are similar to [SS00]. After preprocessing, the baseband complex envelope of the received signal sampled at one sample per symbol at the output of a matched filter can be represented by (2.8).

4.3.1 Simulation results

Many simulations have been carried out to evaluate the performance of the proposed classifier. All constellations have been normalized (unit energy). The SNR in decibels is defined as

$$\text{SNR} = 10\log_{10}\left(\frac{|h|^2}{\sigma_z^2}\right).$$

Since the iterative BW algorithm may converge to a local maximum of the likelihood function, one important issue is parameter initialization.

BW algorithm initialization

The impulse response of the unknown channel can be estimated using HOS of the received signal. According to [Men91], the impulse response of a qth-order moving average (MA) system can be calculated from the estimated fourth-order cumulants of its output as

$$\widehat{h}_k = \frac{\widehat{c}_{4,x}(q,0,k)}{\widehat{c}_{4,x}(q,0,0)}, \quad k = 0,...,q, \tag{4.3}$$

where $\widehat{c}_{4,x}(t_1, t_2, t_3)$ is an estimate of

$$c_{4,x}(t_1, t_2, t_3) = \text{cum}(x^*(t), x(t+t_1), x(t+t_2), x^*(t+t_3))$$

with

$$\text{cum}(w, x, y, z) = E(wxyz) - E(wx)E(yz) - E(wy)E(xz) - E(wz)E(xy).$$

This procedure generally yields good estimations at reasonably high operating SNRs.

Classification performance

This section examines the performance of the plug-in MAP classifier defined in (3.6). All simulations have been obtained from 1000 trials belonging to each class λ_i (i.e. a total of 4000 signals for the four-class problem, and 2000 signals for the two-class problem). For our experiments, the mean vector $\boldsymbol{m}$ was initialized randomly or by (4.3) whereas the initial noise variance was set to $\sigma^2_{\text{init}} = 1$. The step-size for the LMS algorithm was set to $\mu_s = 0.1$ and $\Delta = 5$ for the fixed-lag scheme.

- **2-class problem**

  Consider a set of two modulation formats $\lambda = \{16\text{PSK}, 16\text{QAM}\}$. For the LMS-type algorithm, we apply $\mu_m = 20$ for both formats. This particular example is interesting because the two modulation formats 16PSK and 16QAM have the same number of states and are difficult to distinguish in the presence of ISI and noise. Figure 4.9 shows the average probability of correct classification versus SNR for this problem. Note that two different initializations of the channel coefficients have been considered, namely HOS initialization using (4.3) and random initialization. Of course, the performance improves when the HOS initialization is used. Figure 4.10 displays the average probability of correct classification versus the number of observations for different SNRs. This allows one to adjust the number of observations required to achieve a given classification performance. For instance, at SNR = 9dB, the observation length should satisfy $N_s \geq 500$ to ensure $P_{\text{cc}} \geq 0.9$. When operating at lower SNRs, larger values of N_s are necessary to ensure $P_{\text{cc}} \geq 0.9$. For comparison, we consider a two-tap FIR channel with impulse response $h = [0.707, 0.707]$ studied in [LP95]. The frequency response characteristics of this channel is compared to that of $h = [1, 0.75 + 0.25j]$

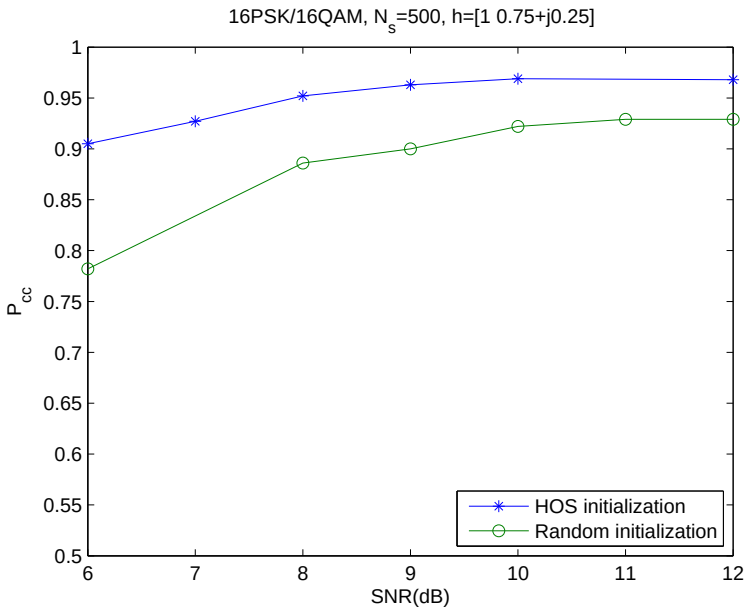


Figure 4.9: Average probability of correct classification versus SNR.

in Fig. 4.11. This figure shows that this new channel exhibits a severe ISI due to its strong attenuation in the signal bandwidth. Figure 4.12 compares the performance of the MAP classifier (4.1) and the PSP/GLRT classifier as a function of the observation length. The proposed classifier provides better performance for $300 < N_s < 900$. Note that the two classifiers achieve the same performance for SNR = 9dB and $N_s > 900$.

- **4-class problem**

  This section considers a set of four modulations which have been studied in [LTD01; BSSS00], i.e., $\lambda = \{\text{BPSK}, \text{4QAM}, \text{8PSK}, \text{16QAM}\}$. The values of μ_m used are: $\mu_m = 0.3$ for BPSK, $\mu_m = 0.6$ for 4QAM, $\mu_m = 10$ for 8PSK, and $\mu_m = 20$ for 16QAM. The average probabilities of correct classification obtained with the classifier (4.1) for random and HOS initializations are displayed in Fig. 4.13. Again, this result shows the necessity of having a good channel initialization. The probabilities of correct classification of each candidate modulation type are plotted in Fig. 4.14. This figure indicates that 4QAM and 16QAM are more difficult to classify than BPSK and 8PSK for the same SNR.

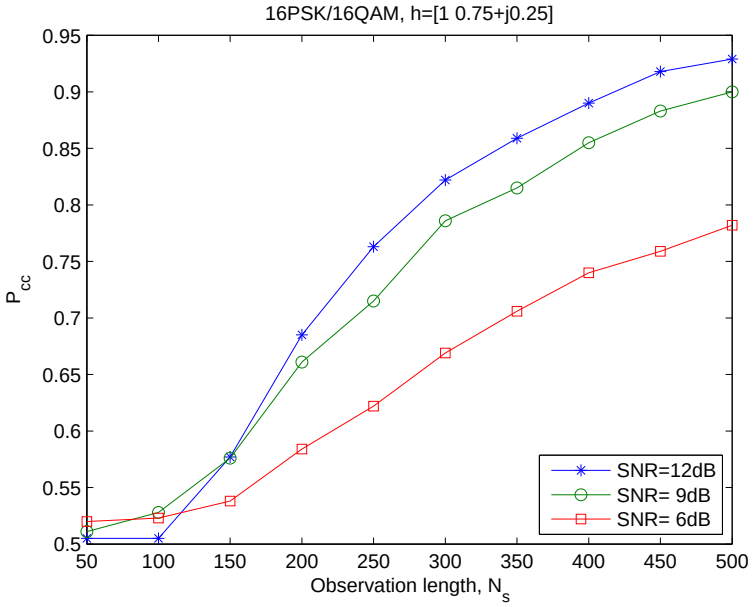


Figure 4.10: Average probability of correct classification versus observation length.

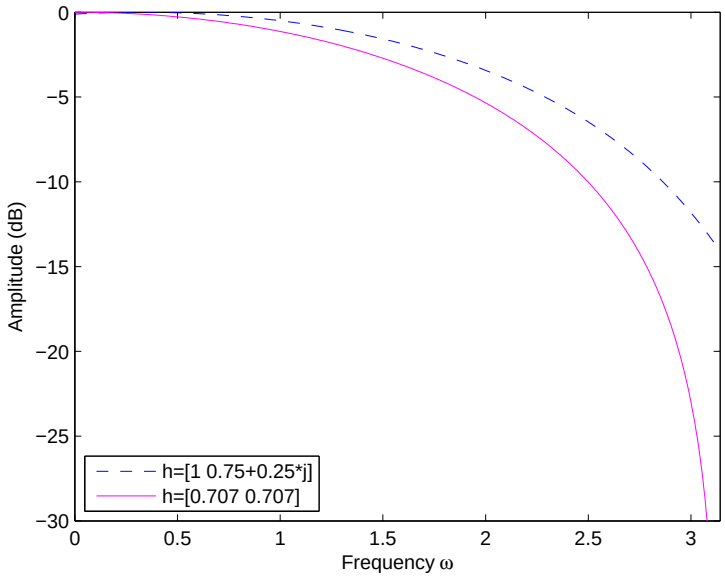


Figure 4.11: Amplitude spectra for two channels with ISI.

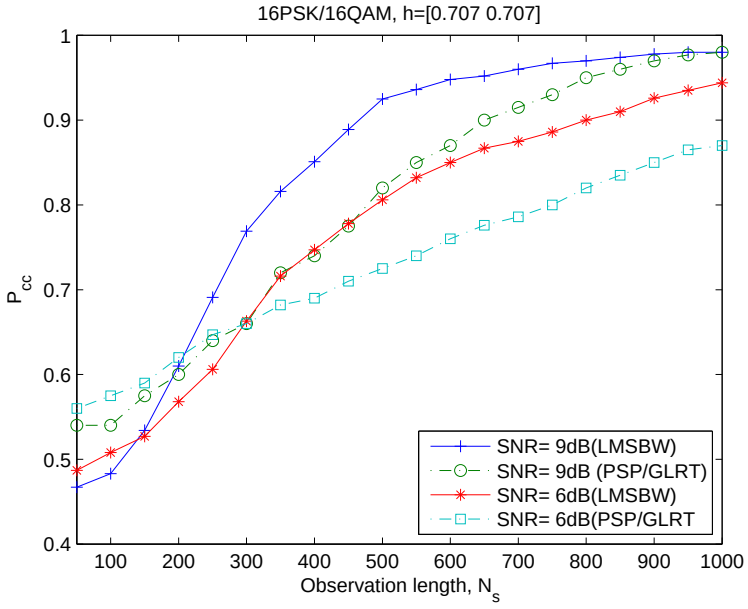


Figure 4.12: Average probability of correct classification versus observation length.

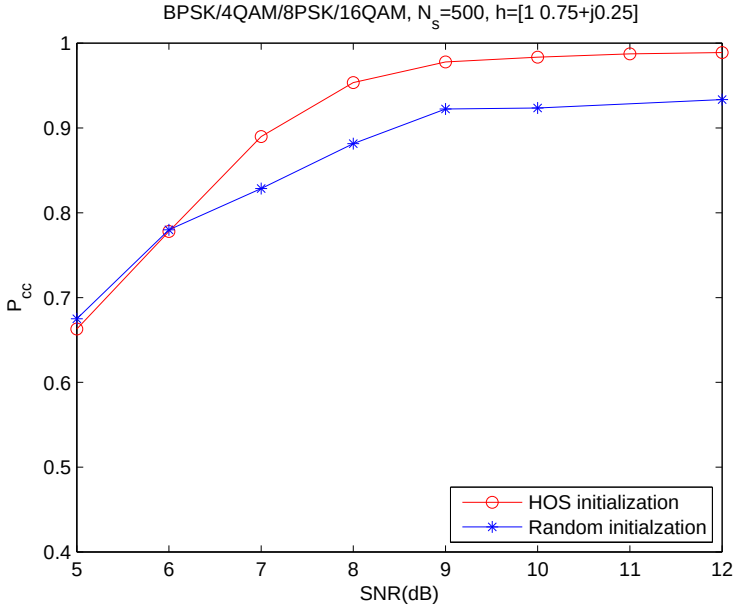


Figure 4.13: Average probability of correct classification versus SNR.

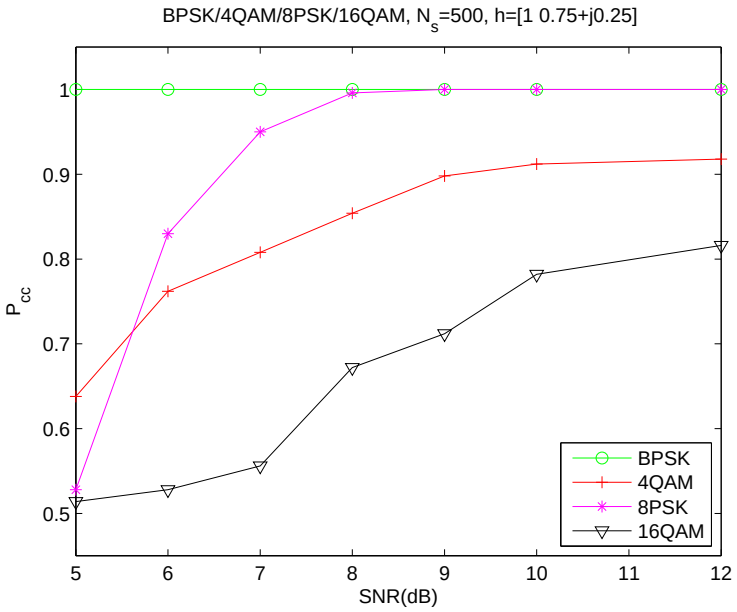


Figure 4.14: Probability of correct classification versus SNR.

4.4 Conclusions

The problem of digital modulation recognition in the presence of AWGN and finite memory unknown channels was addressed. The received communication signal was classified according to a plug-in MAP rule. This rule required to estimate the posterior distribution of the received communication signal conditionally to each modulation belonging to a known dictionary. This estimation was conducted by using the BW algorithm for hidden Markov models which has shown interesting properties for speech recognition. The performance of the proposed classifier was assessed by means of several simulation results. It is important to note that the proposed classifier is insensitive to phase offsets.

To alleviate the performance degradation from large frequency offsets, the classifier constructed from PLLs followed by the BW algorithm was proposed. Recognition was achieved via the plug-in MAP rule. Simulation results showed good classification performance.

Chapter 5

Classification of Nonlinear Modulations

Contents

5.1 Introduction

In this chapter, we consider nonlinear modulation methods in which the principle of superposition does not apply in the mapping of the digital sequence into successive waveforms. Nonlinear modulations are preferable in communication applications which use amplifier devices operating in a nonlinear mode or at near saturation such as in satellite communications. Indeed, the spectral spreading due to the nonlinearity of the amplifier is reduced

when using such modulations. The Gaussian minimum shift keying (GMSK) modulation is an important nonlinear modulation that has become a new modulation standard for telemetry/telecommand (TM/TC) satellite links. The consultative committee for space data system (CCSDS) for future space missions standardized two different GMSK signals [Con01]. More precisely, for the packet telemetry for space-to-earth links, the CCSDS recommends the GMSK ($BT = 0.25$) modulation for spacecrafts orbiting at the altitude below 2×10^6 km and the GMSK ($BT = 0.5$) modulation at the altitude above. The choice of GMSK modulation can be motivated by many interesting properties including spectrum efficiency, capacity of supporting several receivers, and high immunity against interference (see [VV02] and references therein). These new schemes will have to co-exist with other space systems using different linear modulation schemes (BPSK, QPSK, 8PSK). As a consequence, it is important to be able to identify the authorized and non authorized systems. Equivalently, the problem consists of recognizing the modulation associated to a received communication signal.

In this work, we study two classification problems:

- A Bayesian classifier which recognizes GMSK signals with different bandwidths $BT = 0.25$ and $BT = 0.5$ as recommended by the CCSDS. Figure 5.1 shows the power spectrum obtained with these two values of BT.
- A Bayesian classifier which recognizes linear modulations used in satellite systems (BPSK, QPSK, 8PSK) as well as the non-linear CCSDS standardized GMSK modulation schemes ($BT = 0.25$ and $BT = 0.5$).

5.2 GMSK signals

The GMSK modulation was originally proposed by [MH81] and can be realized by filtering a nonreturn-to-zero (NRZ) binary data stream with a Gaussian filter, then passed through an FM analog modulator as shown in Fig. 5.2. GMSK signals are partial continuous phase modulation (CPM) signals (with modulation index $m = 0.5$ and Gaussian frequency shaping) defined as [Pro01]:

$$y(t) = A\cos\left[2\pi f_c t + \Phi(t, \mathbf{a})\right], \qquad t \in \mathbb{R}, \tag{5.1}$$

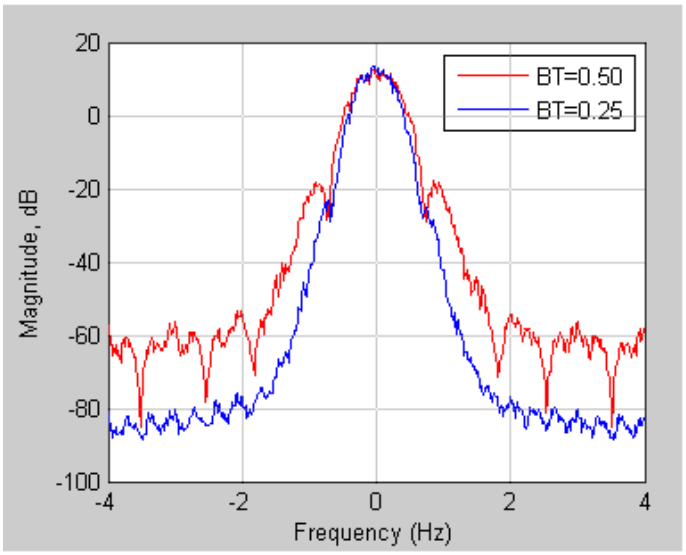


Figure 5.1: Normalized power spectrum at the output of the GMSK modulator.

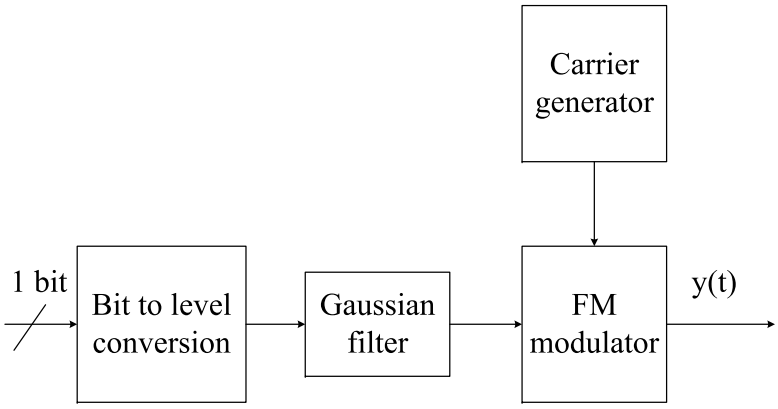


Figure 5.2: GMSK transmitter (FM implementation).

where f_c is the carrier frequency and $\Phi(t, \mathbf{a})$ is the time-varying phase. The transmitted data sequence of M-ary symbols selected from the alphabet $\pm 1, \pm 3, \ldots, \pm(M-1)$ denoted as $\mathbf{a} = \{a_k\}$ is embedded in the time-varying phase

$$\Phi(t, \mathbf{a}) = 2\pi m \sum_{k=-\infty}^{\infty} a_k q(t - kT), \tag{5.2}$$

where $q(t) = \int_{-\infty}^{t} g(\tau) d\tau$ and T is the symbol duration. The frequency shape pulse $g(t)$ has a smooth phase shape over a finite time interval $0 \leq t \leq LT$, where $L \in \mathbb{N}$, and is approximately zero outside this interval. For a GMSK signal, $g(t)$ is defined as

$$g(t) = \frac{1}{2T}\left[Q\left(2\pi B \frac{t - \frac{T}{2}}{\sqrt{\ln 2}}\right) - Q\left(2\pi B \frac{t + \frac{T}{2}}{\sqrt{\ln 2}}\right)\right], \tag{5.3}$$

where B is the 3dB bandwidth of the lowpass Gaussian filter (with $0 \leq BT \leq 1$) and $Q(t) = \int_t^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\tau^2}{2}\right) d\tau$. The time-varying phase during interval $[kT, (k+1)T]$ can be written as

$$\Phi(t, \mathbf{a}) = \theta_k(t, \mathbf{a}) + \phi_k, \tag{5.4}$$

where

$$\theta_k(t, \mathbf{a}) = 2\pi m \sum_{i=k-L+1}^{k} a_i q(t - iT), \tag{5.5}$$

and

$$\phi_k = m\pi \sum_{i=-\infty}^{k-L} a_i \pmod{2\pi}. \tag{5.6}$$

$\theta_k(t, \mathbf{a})$ is determined by the data symbol a_k and the previous $L-1$ symbols. ϕ_k represents the memory of all symbols up to time $k - L$. It also represents the constant part of the total time-varying phase in $[kT, (k+1)T]$, and is equal to the sum of the maximum phase changes contributed to each symbol, accumulated along the time axis up to the $(k-L)^{th}$ symbol interval. It can be recursively computed as

$$\phi_{k+1} = \phi_k + m\pi a_{k-L+1}. \tag{5.7}$$

If m is rational, i.e. $m = 2q/p$, the number of distinct values of ϕ_k is p. The state of a CPM signal at $t = kT$ is classically defined as the vector

$$s(k) = (\phi_k, a_{k-1}, a_{k-2}, \ldots, a_{k-L+1}). \tag{5.8}$$

Each state corresponds to a specific value of the excess phase. An example of the state trellis is shown in Figure 5.3. The signal constellations or scattering diagrams of two GMSK transmitted signals are shown in Figure 5.4. After constructing the state trellis associated to a GMSK signal, the BW algorithm can be applied to estimate the posterior probability of the received modulated signal as done in Section 4.3 for linear modulations in presence of residual channel interferences.

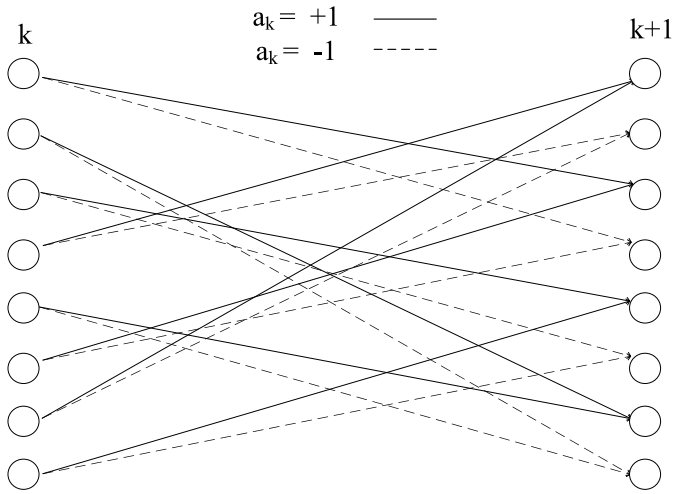


Figure 5.3: State trellis diagram of GMSK signal, $BT = 0.5$.

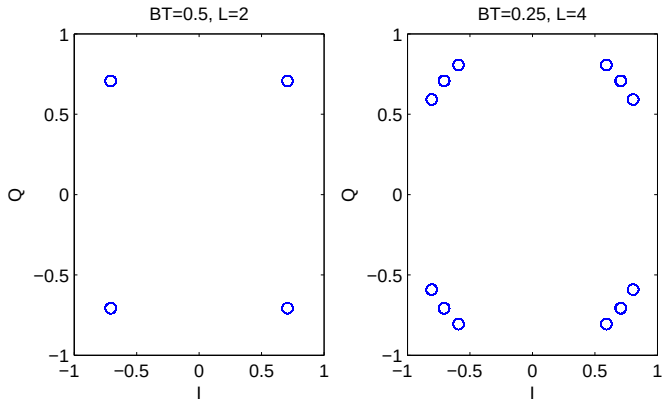


Figure 5.4: Constellations of GMSK transmitted signals.

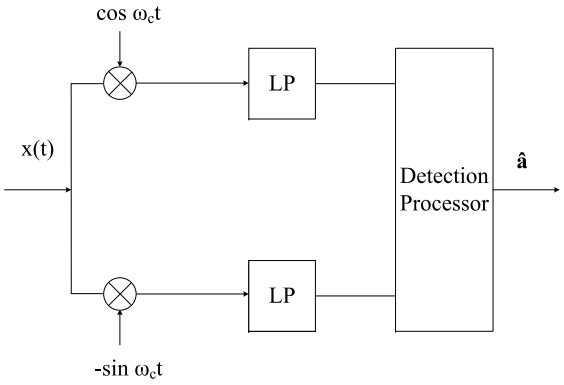


Figure 5.5: Basic quadrature receiver from [AAS86].

5.2.1 GMSK receivers

Satellite/space communication channel is modeled as an AWGN channel and coherent receivers are suitable to obtain the best performance. The coherent receivers can be implemented in two different ways.

- The received signal is multiplied with the synchronous carrier, followed by lowpass (LP) filters, see Figure 5.5. Then a phase generator constructs all the phase transitions depending on the states and compares the received signal with each of the possible transmitted signal, in order to compute a branch metric used by a subsequent Viterbi detector. This receiver has high complexity but is optimum.

- Based on the Laurent decomposition [Lau86], a GMSK signal can be approximated in a form of a superposition of two OQPSK signal with baseband pulses $C_0(t)$ and $C_1(t)$, respectively, and offset in time by T. Depending on the BT values ($BT = 0.25$ or $BT = 0.5$), the authors in [VMP+02] proposed two different receiver structures for each BT value. For instance, Figure 5.6 shows the suboptimum receiver for GMSK signals with $BT = 0.5$. For this value of BT, using only the pulse $C_0(t)$ is sufficient. The signal constellation (one sample per symbol) at the output of the suboptimum receiver displayed in 5.6 in the absence of noise is plotted in Figure 5.7. These suboptimum receivers will destroy the memory of GMSK modulation. Thus they are not suitable for our method to be applied.

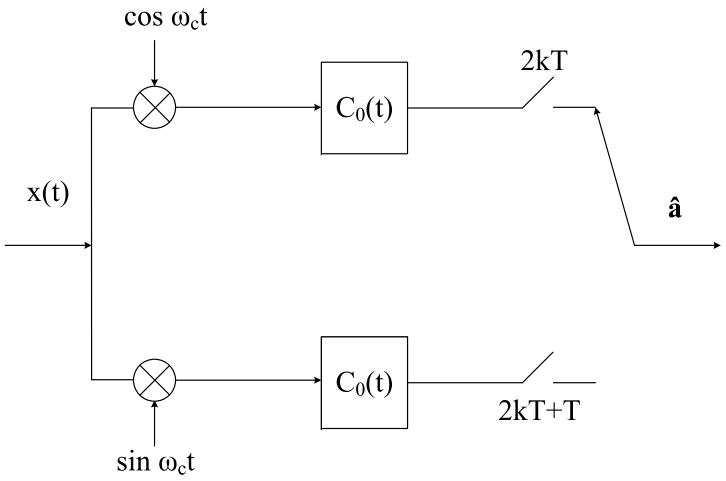


Figure 5.6: Simple GMSK receiver based on OQPSK receiver structure.

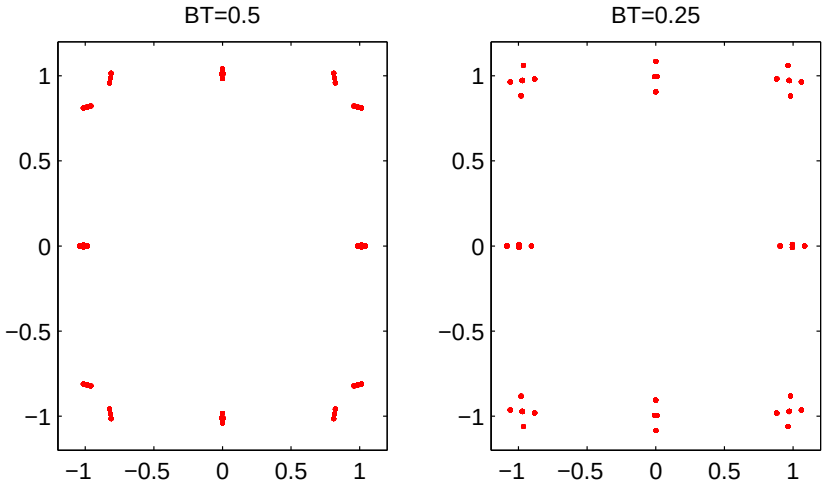


Figure 5.7: GMSK constellations at the output of the suboptimum receiver.

5.3 Classification of GMSK signals with different bandwidths

The classification of nonlinear modulations has received less attention in the literature though these modulations play a great deal in modern communications. Polydoros studied different methods for classifying nonlinear modulations with different modulation indexes [HP92; CP94]. A classifier based on an approximate likelihood function for a multiple M-ary frequency shift keying (MFSK) signal (transmitted through a Rayleigh fading channel) was also studied in [EMN02]. However, classification problems involving GMSK modulations have not been considered in the literature (to the best of our knowledge), despite the popularity of GMSK signals. The classification of linear modulation signals propagating via unknown ISI channels has been studied in Section 4.3. The first step of the proposed algorithm estimated the channel coefficients (which are related to the signal means) and noise variance using the BW algorithm. The received communication signal was then identified according to the MAP rule as in (4.1). In this chapter, we modify the algorithm proposed in Section 4.3 to handle non-linear modulations transmitted through an AWGN channel.

The proposed algorithm assumes that these two non-linear modulations have been pre-identified from other linear modulation candidates. This preprocessing step might be achieved by feature-based classifiers that discriminate constant and nonconstant envelope signals. For instance, the maximum of the squared Fourier transform of the normalized signal amplitudes has been used for this purpose in [AN96a]. Note that the classifier performance will be studied especially at small SNRs as required by GMSK modulation applications.

5.3.1 Signal and hidden Markov model

The baseband GMSK signal can be written as $u(t) = \exp[j\Phi(t, \mathbf{a})]$, where the phase $\Phi(t, \mathbf{a})$ has been defined in (5.2). The transmitted signal is modulated by a local oscillator $\exp(j\omega_c t)$ and is corrupted by additive white Gaussian noise $w_{BP}(t)$ with spectral density $\mathrm{N}_0/2$. At the receiver side, the received signal is multiplied by the synchronous carrier $\exp(-j\omega_c t)$, followed by low pass filters to generate the real and imaginary parts of the complex envelope of the received signal, as illustrated in Figure 5.5.

After downconversion, we obtain the received baseband signal

$$x(t) = u(t) \otimes f(t) + z(t), \quad t \in \mathbb{R}, \tag{5.9}$$

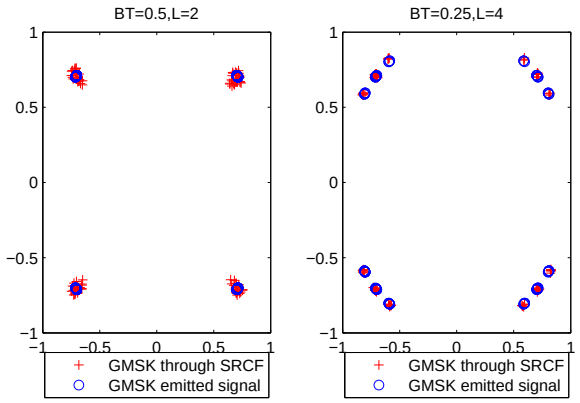


Figure 5.8: GMSK constellations (one sample per symbol).

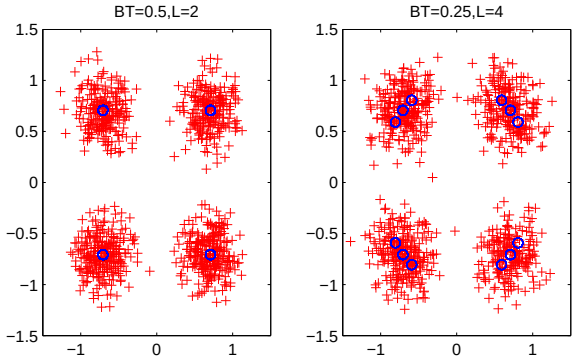


Figure 5.9: Noisy GMSK constellations at SNR = 2 dB.

where $f(t)$ is the impulse response of the LP filters, $z(t) = w(t) \otimes f(t)$ is a normalized complex-valued additive Gaussian noise process with variance σ_z^2 and " $\otimes$ " denotes convolution. Note that $w(t)$ represents the lowpass-equivalent complex Gaussian noise of $w_{BP}(t)$. The baseband complex envelope of the received modulated signal sampled at one sample per symbol ($t = kT$) at the output of the lowpass filters can be written as:

$$x(k) = u(k) \otimes f(k) + z(k), \quad k = 1, ..., N_s, \tag{5.10}$$

where N_s is the number of symbols in the observation interval. Two GMSK signal constellations obtained at the output of a square root raised cosine filter (roll-off factor $R = 0.35$ and cutoff frequency adapted to symbol duration) in the absence of noise marked by "+"

are shown in Figure 5.8. The two constellations are clearly similar even if they are obtained from two distinct GMSK modulations. In the presence of noise as shown in Figure 5.9, it is difficult to distinguish between the two modulations.

The received signal $x(k)$ can be modeled as a probabilistic function of an hidden state at time k which is represented by a first order HMM. This model will be used efficiently for classifying the two nonlinear GMSK modulations with different bandwidths (denoted as λ_1, λ_2). The main HMM characteristics are summarized below:

1. The state of the HMM at time instant k is $s(k)$ which belongs to an alphabet denoted as $\{s_1, s_2, ..., s_N\}$ of size $N = 4M^{L-1}$, where s_j is the jth possible value of $s(k)$. As an example, for binary symbols and GMSK modulation with $BT = 0.5, L = 2$, hence $N = 8$ different states. For binary symbols and GMSK modulation with $BT = 0.25, L = 4$, yielding $N = 32$ different states.

2. The state transition probability distribution is

$$a_{ij} = P[s(k+1) = s_j | s(k) = s_i],$$

which equals $1/M$ when all symbols are equally likely.

3. The initial state distribution vector $\pi = (\pi_1, ..., \pi_N)^T$ is defined by $\pi_i = P[s(1) = s_i] = 1/N,\ i = 1, ..., N$.

4. Based on (5.10), the pdf of the observation $x(k)$ conditioned on state i, denoted as $p_i[x(k)] \triangleq p[x(k)|s_i]$ can be written

$$p_i[x(k)] = \frac{1}{\pi\sigma_z^2} \exp\left(-\frac{|x(k) - m_i|^2}{\sigma_z^2}\right), \tag{5.11}$$

for $i = 1, ..., N$, where $m_i = u(k) \otimes f(k)$ and is approximated by the ith value of $\exp[j\Phi(kT, \mathbf{a})]$. We denote as $\boldsymbol{m} = [m_1, ..., m_N]^T$ the vector containing all possible constellations points.

5.3.2 Simulation results

Many simulations have been carried out to evaluate the performance of the proposed classifier. All constellations have been normalized to unit energy and generated with the bit

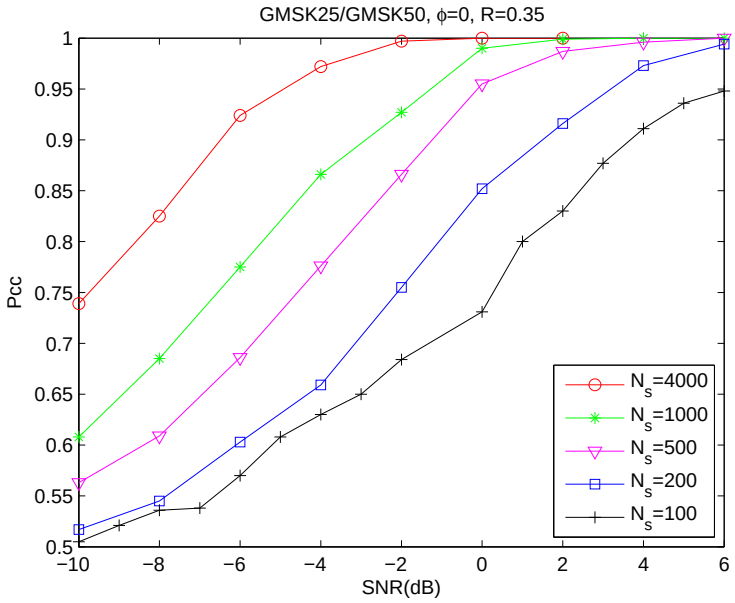


Figure 5.10: Classification performance versus SNR for different N_s.

duration $T = 1$ and the sampling rate $F_e = 10$. The SNR is defined as $\mathrm{E_b/N_0}$, where $\mathrm{E_b}$ is the energy per bit at the input of the receiver. Note that for our GMSK modulations, $\mathrm{E_s/N_0} = \mathrm{E_b/N_0}$, where $\mathrm{E_s}$ is the energy per symbol.

Figure 5.10 displays the classification performance as a function of SNR for the two GMSK modulations (five different values of the number of observations N_s are considered). This figure allows one to appreciate good classification performance even for small SNRs. Figure 5.11 shows the classification performance versus SNR for different values of roll-off factor R. Clearly, the roll-off factor has an impact on the performance and it should be adjusted as a function of signal bandwidth in practical scenarios.

The last simulations study the effect of a phase offset obtained by rotating the constellation with an angle ϕ (this phase offset is due to synchronization errors at the receiver). Figure 5.12 shows that the classification performance seems to be robust to moderate synchronization errors especially for SNR $\geq$ 0dB.

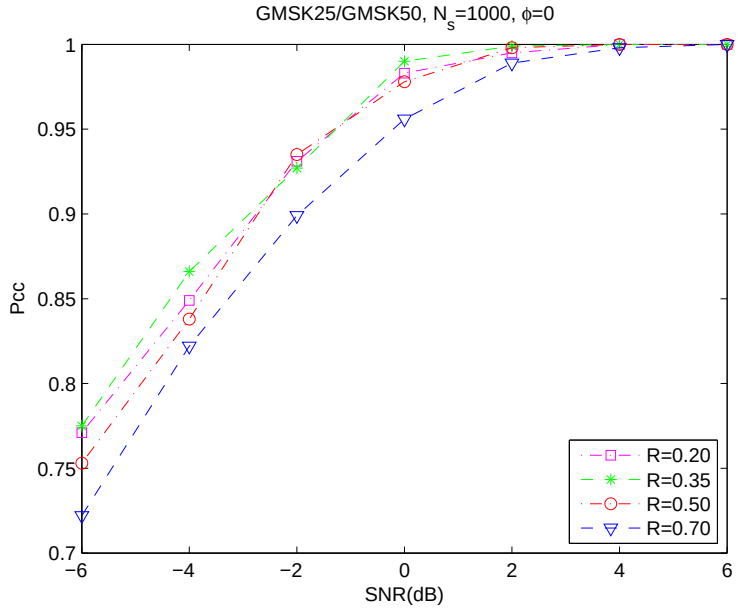


Figure 5.11: Classification performance versus SNR for different R.

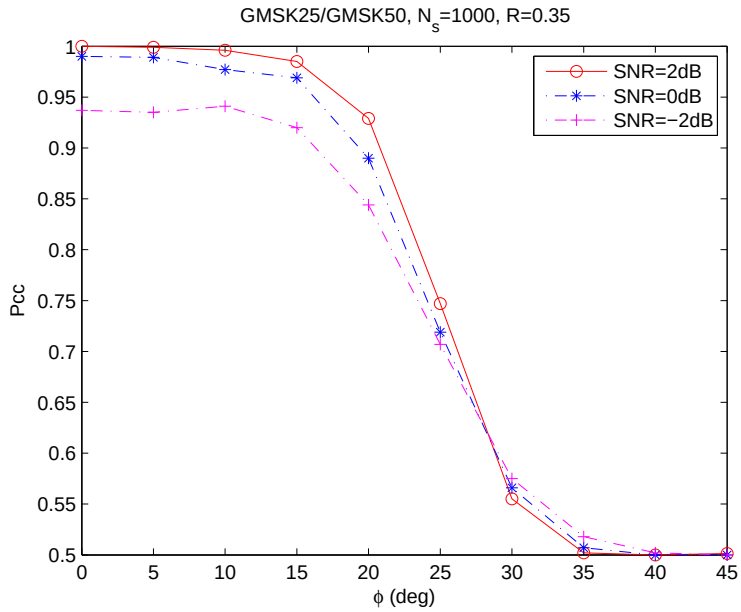


Figure 5.12: Classification performance versus phase offset.

5.4 Classification of linear and nonlinear modulations

A new methodology for classifying the two non-linear GMSK modulations recommended by CCSDS was proposed in Section 5.3. The classifier was based on a state trellis representation (exploiting the fact that the GMSK modulation is a modulation with memory) allowing the use of a modified version of the BW algorithm. The BW algorithm was used to estimate the posterior probabilities of the received modulated signal (conditionally to each class). These posterior probabilities were then plugged into the optimal Bayes decision rule.

However, the algorithm proposed in Section 5.3 assumed that non-linear modulations were pre-identified from other linear modulation candidates, which is not always a simple task. We take a step further and show that linear modulations used in satellite systems (BPSK, QPSK, 8PSK) as well as the non-linear standardized GMSK modulation schemes can be identified using the same recognition process. The emitted linearly or nonlinearly modulated signals are assumed to be corrupted by an additive Gaussian noise whose variance is estimated by the BW algorithm. The performance of the proposed classifier is assessed through several simulation results.

5.4.1 Linear M-PSK modulations

This section recalls some details of linear MPSK modulations to make this chapter self contained. This also allow us to introduce the notations used in Section 5.4.2.The baseband complex envelope of a linearly modulated signal can be written as

$$u(t) = \sum_k d_k h(t - kT), \tag{5.12}$$

where $h(t)$ is the impulse response of the pulse-shaping filter and T represents the symbol duration. The i.i.d. complex symbol sequence $\mathbf{d} = \{d_k\}$ to be transmitted takes its values from a set of M complex numbers $\{S_1, S_2, \ldots, S_M\}$ called constellation representing a particular modulation. MPSK modulations are defined by

$$S_m = \exp\left(j2\pi\frac{m-1}{M}\right), \quad m = 1, \ldots, M. \tag{5.13}$$

For instance, BPSK, QPSK and 8PSK constellations are displayed in Figure 5.13.

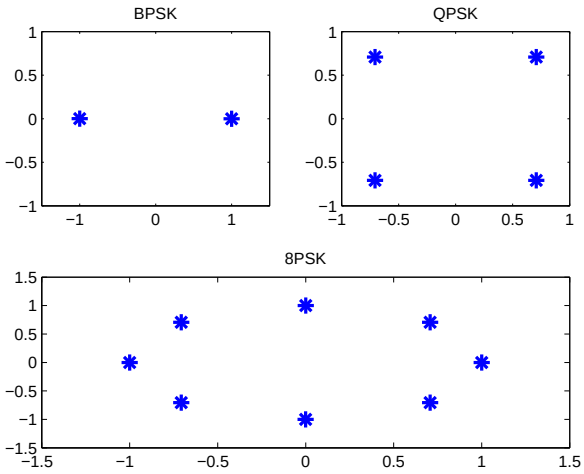


Figure 5.13: Classical linear modulation constellations.

5.4.2 Signal and Hidden Markov model

The signal model and receiver structure in Section 5.3.1 are applied here. The received baseband signal $x(k)$ can be modeled as a probabilistic function of an hidden state at time k which is represented by a first order HMM model whose characteristics are summarized below:

- The state of the HMM at time instant k is $s(k) = d_k$ for MPSK modulated signals (memoryless linear modulation) whereas $s(k) = (\phi_k, a_{k-1}, a_{k-2}, ..., a_{k-L+1})$ for GMSK modulated signals (non-linear modulation with memory). The state $s(k)$ takes its values in $\{s_1, s_2, ..., s_N\}$ where s_j is the jth possible value of $s(k)$. The size of this state matrix is $N = 4M^{L-1}$ for GMSK signals and $N = M$ for linear MPSK modulations.

- The state transition probability is defined by $a_{ij} = P[s(n+1) = s_j | s(n) = s_i]$, and equals $1/M$ when all symbols are equally likely.

- The initial state distribution vector $\pi = (\pi_1, ..., \pi_N)^T$ is defined by $\pi_i = P[s(1) = s_i] = 1/N$ for $i = 1, \ldots, N$.

- Based on (5.10), the pdf of the observation $x(k)$ conditioned on state i, denoted as

$p_i(x(k)) \triangleq p(x(k)|s(i))$ can be written

$$p_i[x(k)] = \frac{1}{\pi\sigma_z^2} \exp\left(-\frac{|x(k) - m_i|^2}{\sigma_z^2}\right),$$

where $i = 1, ..., N$ and m_i is the ith constellation point (ith possible value for $u(k) \otimes f(k)$). Note that $m_i = S_i$ for MPSKs when the transmitter and receiver filters are matched. For GMSK signals, m_i is approximated by the ith value of $e^{j\Phi(kT,\mathbf{a})}$. We denote as $\boldsymbol{m} = [m_1, ..., m_N]^T$ the vector containing all possible constellation points.

5.4.3 Simulation results

Many simulations have been carried out to evaluate the performance of the proposed plug-in MAP classifier. All constellations have been normalized to unit energy and generated with the bit duration $T = 1$ and the sampling rate $F_e = 10$. The SNR is defined as $\mathrm{E_b/N_0}$.

Tables 5.1-5.3 present the confusion matrices of the proposed classifier for different SNR values (the number of samples is $N_s = 500$ for these examples). It can be observed that the two GMSK signals as well as the MPSK signals can be distinguished even at very low values of SNR (even if the constellations of GMSK and QPSK signals are very similar). However, to distinguish among linear modulations, the required operating SNR is much higher especially when 8PSK modulations are present in the dictionary.

Figure 5.14 displays the classification performance as a function of SNR, for different values of the number of observations N_s. A good classification performance can be observed especially for small values of SNR which are typical for satellite space communications. The effect of roll-of mismatch on classification performance was also studied. Figure 5.15 displays the classification performance for several values of the roll-off factor R of the square root raised cosine filters used at the transmitter R_{Tx} and at the receiver R_{Rx}. The proposed classifier seems to be robust to roll-off mismatch. The last simulations study the effect of a phase offset obtained by rotating the constellation with an angle ϕ (this phase offset is due to synchronization errors at the receiver). Figure 5.16 shows that the classification performance is robust to moderate synchronization errors.

In/Out	GMSK25	GMSK50	BPSK	QPSK	8PSK
GMSK25	449	51	0	0	0
GMSK50	13	487	0	0	0
BPSK	0	0	500	0	0
QPSK	0	0	0	498	2
8PSK	0	0	0	0	500

Table 5.1: Confusion matrix for SNR=0dB.

In/Out	GMSK25	GMSK50	BPSK	QPSK	8PSK
GMSK25	406	94	0	0	0
GMSK50	46	454	0	0	0
QPSK	0	0	500	0	0
4QAM	0	0	0	457	43
8PSK	0	0	0	5	495

Table 5.2: Confusion matrix for SNR=−2dB.

In/Out	GMSK25	GMSK50	BPSK	QPSK	8PSK
GMSK25	334	164	1	0	1
GMSK50	123	375	0	1	1
BPSK	0	0	488	4	8
QPSK	0	0	0	313	187
8PSK	0	0	0	81	419

Table 5.3: Confusion matrix for SNR=−6dB.

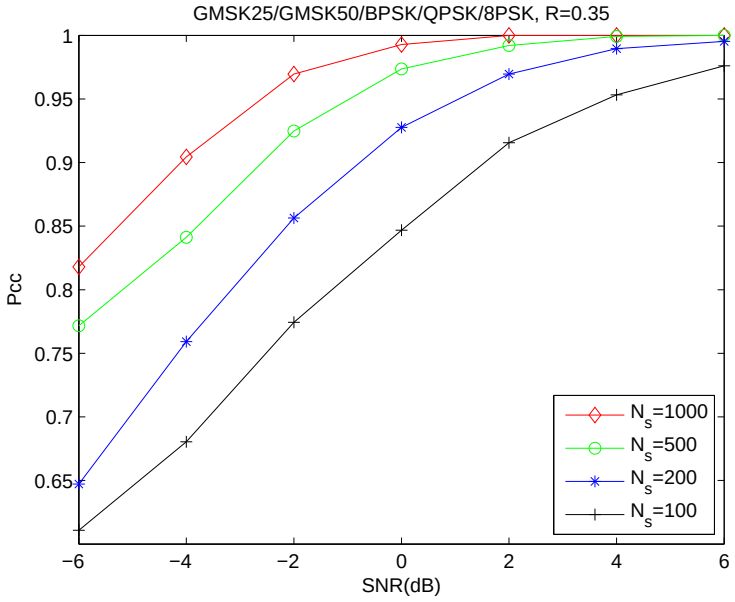


Figure 5.14: Classification performance versus SNR.

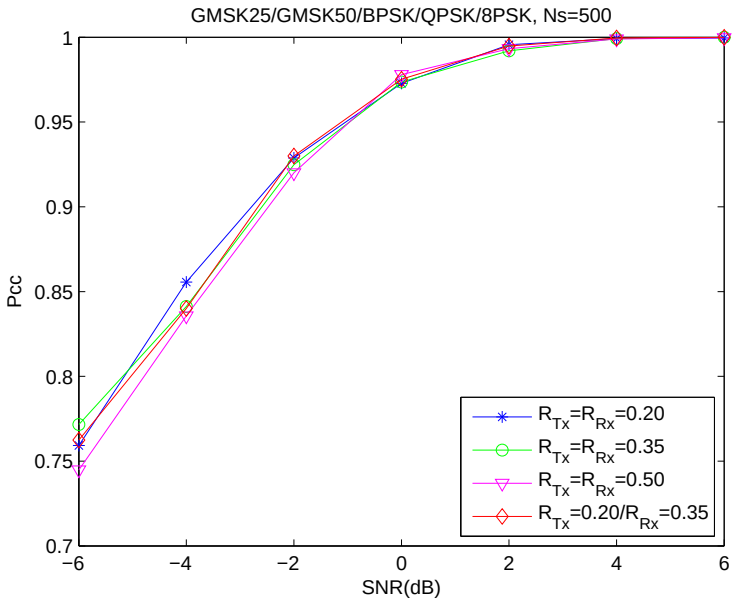


Figure 5.15: Classification performance versus SNR for different roll-off factor.

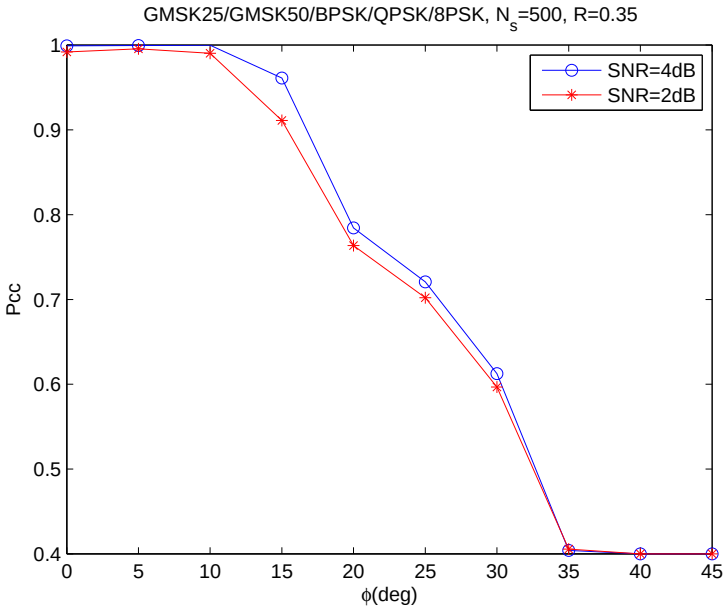


Figure 5.16: Classification performance versus phase offset.

5.5 Conclusions

The problem of identifying non-linear GMSK signals with different values of BT transmitted through AWGN channels has been considered first. The proposed algorithm assumed that these two non-linear modulations have been pre-identified from other linear modulation candidates. Then, the algorithm has been extended to the problem of classifying linear and nonlinear modulations transmitted through AWGN channels. Note that the classification was achieved using the same algorithm.

The received signal was classified according to a MAP rule. This rule required to estimate the posterior distribution of the received communication signal conditionally to each modulation belonging to a known dictionary. This estimation was conducted by using the BW algorithm for HMM. The performance of the proposed classifier was assessed by means of several simulation results and showed good classification performance.

Chapter 6

Conclusions and Perspectives

We have studied the plug-in MAP classifier for digital linear and nonlinear modulations. The key idea of the plug-in MAP classifier is to replace the unknown parameters in the likelihood of the observed data by their estimated value. We have studied three techniques to estimate the posterior probabilities of the received signals conditionally to each modulation, namely MCMC methods, the BW algorithm, and PLL.

The key idea of the MCMC plug-in classifier is to replace the unknown parameters in the likelihood of the observed data by their estimates. These values are obtained from averaging samples drawn by MH algorithm which is one of the most popular MCMC methods. An interesting advantage of MCMC methods is that the accuracy of the estimation can be improved by increasing the number of samples and iterations. More importantly, choosing the proposal distribution corresponding exactly to the target posteriori distribution increases the accuracy of the methods. However, in some circumstances the target posteriori distribution may not be known. In general, MCMC methods are capable of estimating many unknown parameters at the expense of the complexity and speed of calculation. The MCMC plug-in classifier was investigated for linear modulations. From simulation results, it is concluded that the MCMC plug-in classifier outperformed ML and HOS classifiers in the presence of mismatched effects where the received signal is subjected to synchronization errors and residual channels. In a slow Rayleigh fading scenario, we showed by simulation that the MCMC plug-in classifier can be employed and achieve better classification performance than the MOM plug-in classifier.

Due to the high computational cost of MCMC parameter estimation, the BW algorithm

was studied to provide an alternative method of estimation. The BW plug-in classifier estimates recursively the posterior probabilities directly and determine the most likely modulation type from a known dictionary. The modified LMS-type update algorithm was applied to enhance the convergence and computation speed. Although, the BW algorithm does not estimate the unknown parameters directly, they can be derived as side results. Classification performance of the BW plug-in classifier was assessed by simulation for linear modulations in unknown ISI channels. It was shown that using the HOS initialization improved the classification performance compared to random initialization.

The BW plug-in classifier does not take into account the frequency offset. To solve this drawback, the frequency compensation is obtained using PLLs. The most important part of PLL for our classification application is the phase detector. The polarity-type decision-feedback phase detector was studied because it was suitable for all modulation schemes under study. Due to long acquisition time of PLL, it is suggested that this approach is appropriate for long packet communication.

The classifier using the MCMC parameter estimation technique relies on finding the average minimum distance between the received symbols and each constellation point in a known catalogue. Therefore, it cannot distinguish different modulation types having the same constellation shapes such as QPSK and OQPSK, or similar such as nonlinear GMSK modulations ($BT = 0.25$ and $BT = 0.5$). The problem was solved using the BW plug-in classifier to recognize QPSK from OQPSK. The classifier can also identify two GMSK modulations ($BT = 0.25$ and $BT = 0.5$). The same strategy can be applied to classify linear and nonlinear modulations.

An interesting perspective is to recognize modulations in new satellite communication standards such as digital video broadcasting satellite handheld (DVB-SH). This standard uses QPSK, 8PSK, 16APSK modulations and orthogonal frequency division multiplexing (OFDM) whose automatic recognition is a challenging problem.

Appendix A

Existing Modulation Classification Techniques

A.1 Decision-theoretic classifiers

There are three major techniques depending on the model selected for the unknown quantities proposed in the literature: average likelihood ratio test (ALRT), generalized likelihood ratio test (GLRT), and hybrid likelihood ratio test (HLRT). ALRT considers the unknown parameters as random variables with known pdf and averages the likelihood function over this pdf [WM00; Sil99; SM96; KP88; PK90; LCP94; HH03; BW96; BW98]. However, ALRT classifier is not robust with respect to modeling errors and classification of QAM signals is more sensitive than PSK signals [Sil99]. Unlike ALRT, the GLRT regards the unknown parameters as deterministic unknown variables and maximizes the likelihood function with respect to them [HH00; LP95; PAP00]. The GLRT estimates the unknown parameters assuming the i^{th} hypothesis is true, then uses these estimates in a likelihood ratio test as if they were correct. Some implementation advantages over ALRT and HLRT are achieved as it avoids the computation of exponential function and does not require any knowledge of noise power to calculate the likelihood function. ALRT demands a multidimensional integration while GLRT needs a multidimensional maximization. The multidimensional integration complexity and the requirement of knowing the pdf of random variables may cause the ALRT impractical, particularly for a large number of unknown parameters. In GLRT, the multivariate maximization over the unknown variables can lead to the same value of

the likelihood function for the nested signal constellations, e.g., BPSK, QPSK, 16QAM, etc., which results in incorrect classification. HLRT is a combination of the two techniques where vectors of unknown quantities are modeled as random and deterministic variables [CLP95; HH01; HH02]. Averaging over the unknown variables in HLRT vanishes the nested constellation problem of GLRT.

The three approaches are briefly discussed below by assuming the baseband received complex envelope is given by

$$r(t) = s(t) + n(t), \tag{A.1}$$

where $s(t)$ is the baseband complex envelope of the modulated received signal and $n(t)$ is the baseband complex noise.

•**For ALRT**, the likelihood function under hypothesis H_i, representing the i^{th} modulation is given by

$$\Lambda_A^i[r(t)] = \int \Lambda[r(t)|\mathbf{v}_i, H_i] p(\mathbf{v}_i|H_i) d\mathbf{v}_i \tag{A.2}$$

where $\Lambda[r(t)|\mathbf{v}_i, H_i]$ is the conditional likelihood function , $\mathbf{v}_i$ is the unknown vector of the likelihood under hypothesis H_i, and $p(\mathbf{v}_i|H_i)$ is the prior information regarding these parameters.

•**For GLRT**, the likelihood function is given by

$$\Lambda_G^i[r(t)] = \max_{\mathbf{v}_i} \Lambda[r(t)|\mathbf{v}_i, H_i] = \Lambda[r(t)|\hat{\mathbf{v}}_{ML}, H_i] \tag{A.3}$$

where $\hat{\mathbf{v}}_{ML}$ is the maximum likelihood estimator of $\mathbf{v}_i$ under hypothesis H_i.

•**For HLRT**, the likelihood function is given by

$$\Lambda_H^i[r(t)] = \max_{\mathbf{v}_{i_1}} \int \Lambda[r(t)|\mathbf{v}_{i_1}, \mathbf{v}_{i_2}, H_i] p(\mathbf{v}_{i_2}|H_i) d\mathbf{v}_{i_2} \tag{A.4}$$

where $\mathbf{v}_{i_1}$ is a vector gathering deterministic variables and $\mathbf{v}_{i_2}$ is a vector of random variables.

In a two-hypothesis classification problem, the decision is made according to

$$\frac{\Lambda^{(1)}[r(t)]}{\Lambda^{(2)}[r(t)]} \underset{H_2}{\overset{H_1}{\gtrless}} \Gamma(PFA) \tag{A.5}$$

where $\Gamma(PFA)$ is an appropriate threshold depending on the probability of false alarm (PFA).

A.2 Pattern recognition approach

The high complexity of ML methods motivates the search for statistical features leading to pattern recognition approach that will yield good results with low complexity. The design of a pattern recognition algorithm is basically based on feature extraction (to represent the received data), followed by pattern recognizer or decision making. Different methods were proposed for decision making, such as density-based [SS00], the Hellinger distance and unsupervised clustering techniques [DH97]. Examples of features such as the variance of the centered normalized signal amplitude, phase and frequency [AN96a], the variance of the magnitude of the signal wavelet transform (WT) after peak removal [HPC95], moments, and cumulants of the signal [SS00; DWW02; LB97] are explained in this appendix.

A.2.1 Instantaneous amplitude, phase, and frequency

The fact that the information is hidden in the instantaneous amplitude, phase, and frequency of the signal can be exploited. For spectral based features, the standard deviation of the normalized-centered instantaneous amplitude defined by [AN96a; AN96b] is

$$\sigma_{aa} = \sqrt{\frac{1}{L}\left[\sum_{A_n(i)>a_t} A_{cn}^2(i)\right] - \left[\frac{1}{L}\sum_{A_n(i)>a_t} A_{cn}(i)\right]^2} \tag{A.6}$$

where $A_{cn}(i)$ is the value of the normalized-centered instantaneous amplitude at time instant $t = i/f_s$, $(i = 1, 2, ..., N_s)$, f_s is the sampling rate, $A_n(i)$ is the normalized instantaneous amplitude at time instant $t = i/f_s$, L is the number of samples in $A_{cn}(i)$ for which $A_n(i) > a_t$, and a_t is a threshold value.

The standard deviation of the normalized-centered instantaneous phase is

$$\sigma_{ap} = \sqrt{\frac{1}{D}\left[\sum_{A_n(i)>a_t} \phi^2(i)\right] - \left[\frac{1}{D}\sum_{A_n(i)>a_t} |\phi(i)|\right]^2} \tag{A.7}$$

where D is the number of samples in $\phi(i)$ for which $A_n(i) > a_t$.

The standard deviation of the normalized-centered instantaneous frequency is

$$\sigma_{af} = \sqrt{\frac{1}{D}\left[\sum_{A_n(i)>a_t} f^2(i)\right] - \left[\frac{1}{D}\sum_{A_n(i)>a_t} |f(i)|\right]^2}. \tag{A.8}$$

The maximum value of the spectral power density of the normalized-centered instantaneous amplitude is defined as

$$\gamma_{max} = \frac{\max |DFT(A_{cn}(i))|^2}{N_s}. \tag{A.9}$$

A.2.2 Wavelet transform

Digital modulated waveform is a cyclostationary signal that contains transients in amplitude, frequency or phase. Different modulation schemes have different transients, and the differences can be exploited for modulation classification. For instance, FSK changes frequency, whereas PSK changes phase. Another example is that an M-ary PSK signal has M possible phase changes. The WT provides a constant-Q analysis that is suitable for transient detection and characterization.

For digital implementation, the magnitude discrete Haar WT of the MPSK received signal were used in [HPC95; HPC00] for modulation classification:

$$|\overline{WT}_p(a,n)| = 2\sqrt{\frac{S}{a}}\left|\frac{\sin(\omega_c a/4)\sin(\omega_c a/4+\alpha/2)}{\sin(\omega_c/2)}\right| \tag{A.10}$$
$$(i-1)T_s + a/2 \le n \le iT_s - a/2$$

where S is the signal power, a is the scale, and $\alpha \in \{(m-1)2\pi/M\}_{m=1}^{M}$.

The magnitude discrete Haar WT of the MFSK signal can be obtained from

$$|WT_F(a,n)| = 2\sqrt{\frac{S}{a}}\frac{\sin^2((\omega_c+\omega_i)a/4)}{|\sin(\omega_c+\omega_i/2)|} \tag{A.11}$$

where $\omega_i \in \{\omega_1, \omega_2, ..., \omega_M\}$ is the frequency deviation and can have a negative value.

The WT magnitude of PSK has one DC level and many levels of peaks, whereas that of FSK has several levels for DC and peaks. After filtering, the FSK WT magnitude still contains different DC levels, whereas there is only one in PSK. Thus the median filter output for FSK will have a higher variance, and the variance test is a simple method to separate the two. Identification of M-ary PSK is achieved by observing the number of possible peak values at the times of phase changes.

A.2.3 Phase PDF and statistical moments

The received signal in (A.1) has the general MPSK signal component given by

$$s(t) = \frac{\sqrt{2E}}{T_s}\cos(\omega_c t + b(t) + \theta_c). \tag{A.12}$$

The information bearing signal is written as

$$b(t) = \sum_n \varphi_n u(t - nT_s) \tag{A.13}$$

where $u(t)$ is the standard unit pulse of duration T_s. $\varphi_n = a_n(\pi/M)$ is the phase transmitted at time $nT_s \leq t < (n+1)T_s$. The phase component extracted from $r(t)$ at the i^{th} sample $\psi(i)$ can be expressed as

$$\psi(i) = \theta_M(i) + \xi(i), -\pi < \psi(i) \leq \pi \tag{A.14}$$

where $\theta_M(i)$ is the sampled phase component of $s(t)$, and $\xi(i)$ is the random phase component due the noise.

The pdfs of phase ψ were derived in [YS91a]:

$$p(\psi; M) = \frac{1}{2\pi}\left\{1 + 2\sum_{m=1}^{\infty} c_m \cos(m\psi)\right\}, M = 1(CW) \tag{A.15}$$

$$p(\psi; M) = \frac{1}{2\pi}\left\{1 + 2\sum_{n=1}^{\infty} c_{nM}(-1)^n \cos(nM\psi)\right\}, M = 2, 4, 8, ... \tag{A.16}$$

where $c_m = e^{-\gamma}\sum_{k=0}^{\infty}\frac{\Gamma(\frac{m}{2}+1+k)\gamma^{m/2+k}}{k!(m+k)!}$ and γ is the SNR.

The optimum classifier then determines which one of the possible transmitted signals has the maximum a posteriori probability according to Bayes rule

$$p(H_i|\psi_{N_s}) = \frac{p(H_i)p(\psi_{N_s}|H_i)}{p(\psi_{N_s})} \tag{A.17}$$

where $p(\psi_{N_s}|H_i) = \prod_{k=1}^{N_s} p(\psi(k)|H_i)$.

The pdf of phase can be approximated by the Tikhonov function and thus can be expressed as

$$p(\psi; M) = \frac{1}{M}\sum_{k=1}^{M}\frac{\exp[2\gamma cos(\psi - \eta_k(M))]}{2\pi I_0[2\gamma]} \tag{A.18}$$

where $I_0[.]$ is the zero order modified Bessel function of the first kind and $\eta_k(M) = (2k - M - 1)/M$.

The n^{th} moment of the phase is defined as

$$m_n(M) = \int_{-\pi}^{\pi} \psi^n p(\psi; M) d\psi. \tag{A.19}$$

Since $p(\psi; M)$ is an even function of ψ, $m_n(M) = 0$ for odd n. For classification purposes, the even moments from the sampled phases is needed. Thus the sampled moments are defined as

$$\overline{m}_n(M) = \frac{1}{N_s} \sum_{i}^{N_s} \psi^n(i; M) \tag{A.20}$$

where N_s is the number of samples. By the central limit theorem, the pdf of $\overline{m}_n(M)$, $p(\overline{m}_n(M))$, approaches a Gaussian density as N_s increases. That is

$$p(\overline{m}_n(M)) = N(\mu_n(M), \sigma_z^2(M)) \tag{A.21}$$

where $\mu_n(M) = m_n(M)$ and $\sigma_z^2(M) = (m_{2n}(M) - m_n^2(M))/N_s$.

A.2.4 Cyclic-cumulants

The received signal $x(t)$ can be modeled as a cyclostationary process. As a result, time-dependency must be taken into account when expressing the temporal cumulants of $x(t)$.

Let $C_{x,p+q,p}(t; \tau)$ be the $(p+q)^{th}$-order cumulant-based correlation of the process $x(t)$, defined with p non-conjugated terms and q conjugated terms:

$$C_{x,p+q,p}(t; \tau) = cum(x(t), x(t+\tau_1), ..., x(t+\tau_{p-1}), x^*(t-\tau_p), ..., x^*(t-\tau_{p+q-1})). \tag{A.22}$$

Since $x(t)$ is almost-cyclostationary, there are at most countably values of α for which the so-called $(p+q)^{th}$-order cyclic correlation defined as

$$C^{\alpha}_{x,p+q,p}(\tau) = \lim_{T \to +\infty} \frac{1}{T} \sum_{t=0}^{T-1} C_{x,p+q,p}(t; \tau) \exp(-j\alpha t). \tag{A.23}$$

Therefore, the modulus of the 2^{th}-order cyclic-cumulant is given by [Spo95]:

$$|C^{\alpha}_{x,2,1}(\tau)| = \left| \frac{C_{21}}{T_s} \sum_{t=-\infty}^{+\infty} g(t) g(t-\tau) \exp(-j\alpha t) \right| \tag{A.24}$$

where $C_{21} = cum(x_k, x_k^*)$ and $g(t)$ is the real-valued pulse function. Similarly, the modulus of the cyclic tricorrelation (also called fourth-order temporal cumulant) is

$$|C^{\alpha}_{x,4,2}(\tau)| = \left| \frac{C_{42}}{T_s} \sum_{t=-\infty}^{+\infty} g(t)g(t+\tau_1)g(t-\tau_2)g(t-\tau_3)\exp(-j\alpha t) \right|. \tag{A.25}$$

A.2.5 Algorithms for linearly modulated signals

Pattern recognition algorithms depend on features that can be extracted from modulated received signals. For linear modulation recognition, features extracted from the instantaneous amplitude and phase of the received signal were exploited . The variance of the absolute value of the normalized-centered instantaneous amplitude was used to differentiate between 2ASK and 4ASK [AN96a; AN96b; NA97; NA98; WN01]. For PSK signals, the phase PDF is multimodal, thus the number of modes provides information for PSK order identification [YS91a; SH92; YS91b; YS95; YS97; YL98]. In the high-SNR region, MPSK exhibits M distinct modes. However with the decrease of SNR or the increase of M , the peaks smear off and the PDF converges to a uniform PDF [SH92]. An approximation using the Tikhonov PDF and a Fourier series expansion of the phase PDF with a log-likelihood ratio test were employed [YS91a; YS97; YL98]. By using these methods to compute the phase PDF, closed-form expressions for the phase statistical moments were derived, and pdfs of the sample estimates of the moments were used for decision making [YS91a; SH92; YS95]. The histogram of phase difference between two adjacent symbols was compared against a particular pattern, for PSK signal identification in [Lie84; HS89; HS90]. The DFT was applied to the phase histogram of the received symbols to analyze the periodic components of the phase PDF or the empirical characteristic function of the phase [SMH95; SR97]. By exploiting an additional information about the magnitude of the received signal, the algorithm in [SR97] was extended to QAM signal classification. Other features such as the kurtosis of the amplitude extracted from the instantaneous amplitude and phase were investigated for PSK and QAM identification in [TM99; DBG91; UIK00]. Different PSK signals give rise to different sets of peak values in the magnitude of the Haar wavelet transform, and hence, the input was classified as PSK of order M if the histogram of the peak magnitudes had M/2 to M-1 distinct modes [HPC95; HPC00].

Cumulant-based features were also proposed in [SS00] to recognize the modulation order of ASK, PSK and QAM signals: the normalized cumulant of fourth-order with two

conjugations for ASK, the magnitude of the normalized cumulant of fourth-order with zero conjugations for PSK (M > 2) and the normalized cumulant of fourth-order with zero conjugations for QAM. This technique was modified to signals in frequency-selective channels [SBS00]. Artificial neural network (ANN) using spectral based feature set for modulation recognition was addressed in [AN96a; AN96b; NA97; NA98] for PSK signals. In addition to the spectral based feature set, multi-layer perceptrons (MLP) recognizer was implemented in [WN01] with better generalization by addition of a new statistical cumulant-based features set to include QAM signals.

The cyclostationarity property of linear modulation signal was also exploited by two main techniques: spectral line generation and periodic fluctuations with time of cumulants up to the n th-order. The first method analyzed the set of spectral lines generated by various m-th law devices using a decision-tree provides good classification results in an AWGN environment [Rei92]. To improve the robustness of the classifier in unknown environments, the authors in [SKKR97] explored a Hidden-Markov-Model (HMM) based classifier. These techniques can also be used to classify ASK, BPSK, QPSK, 2FSK, MSK and CW. The second method computed cyclic-cumulants (CCs) of different orders at the cycle frequency equal to the symbol rate. A CC-based feature proposed in [MLL98] was used to recognize QAM signals. Cyclic-cumulants up to the sixth-order were investigated in [Spo95; SBY00; Spo01] for one and multiple incoming signals. Multiple signals, which overlap in time and frequency but have distinct symbol rates and hence, different cycle frequencies, can also be distinguished using cyclic-cumulants [GS94]. Eight-order cyclic-cumulants were studied in [DBNS03] for classifying real- and complex-valued constellations, respectively. CC-based features which are robust to carrier frequency offset and phase jitter were proposed for QAM classification in [DBNS04].

Signal moments similar to the CC-based feature proposed in [MLL98] were applied to distinguish between QPSK and 16QAM in [**?**]. Specifically, a linear combination of the fourth-order moment with two conjugations and the squared second-order moment with one conjugation were employed, with the coefficients and the delay vector optimized, to maximize the probability of correct classification. A set of features was chosen for certain values of the delay vector, and classification was made based on the correlation between the observed and theoretical feature vectors. The relationships between the second- and higher moments of received signal was employed to discriminate BPSK, 4ASK, 8PSK, and 16QAM

signals in [DWW02].

A.2.6 Algorithms for nonlinearly modulated signals

FSK signals are characterized by constant instantaneous amplitude, whereas ASK signals have amplitude fluctuations, and PSK signals have information in the phase. Thus, the instantaneous frequency information can be exploited for identifying FSK signals. In [AN96a; AN96b; NA97; NA98; WN01], the variance of the absolute value of the normalized centered instantaneous frequency was used to distinguish between 2FSK and 4FSK. In [HS89; HS90], the variance of the zero-crossing interval was used as a feature to distinguish between unmodulated waveform (UW) and PSK, and FSK classes. This zero-crossing interval is a measure of the instantaneous frequency and hence, for FSK signals it is a staircase function, whereas for UW and PSK signals it is a constant. The variance of the instantaneous frequency was also employed in [AFM92; FM93] to distinguish between single-tone (UW and PSK) and multiple tones (FSK). As in PSK classification, the number of modes in the instantaneous frequency histogram was employed in [HS89; HS90] for FSK classification. The instantaneous frequency derivative was also used to distinguish between 2FSK and 4FSK [AFM92; FM93], while, the number of modes in the pdf of the Haar wavelet transform magnitude was investigated for FSK signal identification [HPC95; HPC00]. Finally, spectral properties of FSK signals were explored for classification in [YSS03].

List of Publications

1. Anchalee Puengnim, Nathalie Thomas, Jean-Yves Tourneret, and Josep Vidal, "Classification of linear and nonlinear modulations using the Baum-Welch algorithm", *Proceeding of the 16th European Signal Processing Conference (Eusipco2008)*, Lausanne, Switzerland, 25-29 August 2008.

2. Anchalee Puengnim, Nathalie Thomas, Jean-Yves Tourneret, and Herve Guillon, "Classification of GMSK Signals with Different Bandwidths", *Proceeding of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2008)*, Las Vegas, USA, 30 March- 4 April 2008.

3. Anchalee Puengnim, Thierry Robert, Nathalie Thomas, and Josep Vidal, "Hidden Markov Models for Digital Modulation Classification in Unknown ISI Channels", *Proceeding of the 15th European Signal Processing Conference (Eusipco2007)*, Poland, 3 - 7 September 2007.

4. Anchalee Puengnim, Nathalie Thomas, and Jean-Yves Tourneret, "Digital Modulation Classification in Flat-Fading Channels", *Proceeding of the 14th European Signal Processing Conference (Eusipco2006)*, Italy, 4 - 8 September 2006.

Bibliography

[AAS86] J. Anderson, T. Aulin, and C-E. Sundberg. *Digital Phase Modulation.* New York:Plenum Publishing Company, 1986.

[ADC+04] A. Abdi, O. A. Dobre, R. Chauchy, Y. Bar-Ness, and W. Su. Modulation classification in fading channels using antenna arrays. In *Proc. IEEE MILCOM*, volume 1, pages 211–217, November 2004.

[AFM92] K. Assaleh, K. R. Farrell, and R. J. Mammone. A new method of modulation classification for digitally modulated signals. In *Proc. IEEE MILCOM*, volume 2, pages 712–716, October 1992.

[AN96a] E. E. Azzouz and A. K. Nandi. *Automatic Modulation Recognition of Communication Signals.* Kluwer Academic, 1996.

[AN96b] E. E. Azzouz and A. K. Nandi. Procedure for automatic recognition of analogue and digital modulations. *IEE Proc. Commun*, 143, October 1996.

[Bou97] M-L. Boucheret. *Récepteur numérique pour systèmes à base de modulations codées en treillis: algorithmes et architecture.* PhD thesis, Ecole Nationale Supérieure des Télécommunications, Paris, France, July 1997.

[BSSS00] S. Barbarossa, A. Swami, B. Sadler, and G. Spadafora. Classification of digital constellations under unknown multipath propagation conditions. In *Proc. SPIE*, volume 79, pages 278–305, April 2000.

[BW96] B. F. Beidas and C. L. Weber. Higher-order correlation-based classification of asynchronous MFSK signal. In *Proc. IEEE MILCOM*, volume 3, pages 1003–1009, October 1996.

[BW98] B. F. Beidas and C. L. Weber. Asynchronous classification of MFSK signals using the higher order correlation domain. *IEEE Journal on Sel. Areas in Commun*, 46:480–493, April 1998.

[CLP95] K. M. Chugg, C. S. Long, and A. Polydoros. Combined likelihood power estimation and multiple hypothesis modulation classification. In *Proc. ASILOMAR*, volume 2, pages 1137–1141, November 1995.

[Con01] Consulative Committee for Space Data Systems (CCSDS). *Radio Frequency and Modulation Systems*. Number 401.0-B in Blue Book. CCSDS, June 2001.

[CP94] C-D. Chung and A. Polydoros. Envelope-based classification schemes for continuous-phase binary frequency-shift-keyed modulations. In *Proc. ASILOMAR*, volume 3, pages 796–800, October 1994.

[DBG91] L. V. Dominguez, J. M. P. Borrallo, and J. P. Garcia. A general approach to the automatic classification of radio communication signals. *Signal Processing*, 22(3):239–250, March 1991.

[DBNS03] O. A. Dobre, Y. Bar-Ness, and W. Su. Higher-order cyclic cumulants for high order modulation classification. In *Proc. IEEE MILCOM*, volume 1, pages 112–117, October 2003.

[DBNS04] O. A. Dobre, Y. Bar-Ness, and W. Su. Robust QAM modulation classification algorithm based on cyclic cumulants. In *Proc. WCNC*, volume 2, pages 745–748, March 2004.

[DH97] D. L. Donoho and X. Huo. Large-sample modulation classification using hellinger representation. In *Proc. Signal Processing Advances in Wireless Communications*, pages 133–136, April 1997.

[DWW02] W. Dai, Y. Wang, and J. Wang. Joint power and modulation classification using second- and higher statistics. In *Proc. WCNC*, volume 1, pages 155–158, March 2002.

[EMN02] A. E. El-Mahdy and N. M. Namazi. Classification of multiple M-ary frequency-shift keying over a Rayleigh fading. *IEEE Trans. Commun.*, 50, June 2002.

[FM93] K. R. Farrell and R. J. Mammone. Modulation classification using a neural tree network. In *Proc. IEEE MILCOM*, volume 3, pages 1028–1032, October 1993.

[FV94] J. A. R. Fonollosa and J. Vidal. Application of Hidden Markov Models to blind channel characterisation and data detection. In *Proc. IEEE ICASSP-94*, volume 4, pages 185–188, April 1994.

[GRS96] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Introducing Markov Chain Monte Carlo*. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, Eds.. London, U.K.: Chapman & Hall, 1996.

[GS94] W. A. Gardner and C. M. Spooner. The cumulant theory of cyclostationary time-series, part I: foundation. *IEEE Trans. Signal Processing*, 42:3387–3408, December 1994.

[HH00] L. Hong and K. C Ho. BPSK and QPSK modulation classification with unknown signal level. In *Proc. IEEE MILCOM*, volume 2, pages 976–980, October 2000.

[HH01] L. Hong and K. C Ho. Modulation classification of BPSK and QPSK signal using a two element antenna array receiver. In *Proc. IEEE MILCOM*, volume 1, pages 118–122, October 2001.

[HH02] L. Hong and K. C Ho. An antenna array likelihood modulation classifier for BPSK and QPSK signals. In *Proc. IEEE MILCOM*, volume 1, pages 647–651, October 2002.

[HH03] L. Hong and K. C Ho. Classification of BPSK and QPSK signals with unknown signal level using the Bayes technique. In *Proc. IEEE ISCAS*, volume 4, pages 1–4, May 2003.

[HP92] C. Y. Huang and A. Polydoros. Two small-snr classification rules for CPM. In *Proc. IEEE MILCOM*, volume 3, pages 1236–1240, October 1992.

[HPC95] K. C. Ho, W. Prokopiw, and Y. T. Chan. Modulation identification by the wavelet transform. In *Proc. IEEE MILCOM*, volume 2, pages 886–890, November 1995.

[HPC00] K. C. Ho, W. Prokopiw, and Y. T. Chan. Modulation identification of digital signals by the wavelet transform. *IEE Proc. Radar, Sonar and Navig.*, 147:169–176, August 2000.

[HS89] S. Z. Hsue and S. S. Soliman. Automatic modulation recognition of digitally modulated signals. In *Proc. IEEE MILCOM*, volume 3, pages 645–649, October 1989.

[HS90] S. Z. Hsue and S. S. Soliman. Automatic modulation classification using zero crossing. *IEE Radar and Signal Processing*, 137:459–464, December 1990.

[KM93] V. Krishnamurthy and J. Moore. On-line estimation of Hidden Markov Model parameters based on the Kullback-Leibler information measure. *IEEE Trans. Signal Processing*, 41(8):2557 – 2573, August 1993.

[KP88] K. Kim and A. Polydoros. Digital modulation classification: The BPSK versus QPSK case. In *Proc. IEEE MILCOM*, volume 2, pages 431–436, October 1988.

[Lau86] P.A. Laurent. Exact and approximate construction of digital phase modulations by superposition of amplitude modulated pulses. *IEEE Trans. Commun.*, 34:150–160, February 1986.

[LB97] C. LeMartret and D. M. Boiteau. Modulation classification by means of different order statistical moments. In *Proc. IEEE MILCOM*, volume 3, pages 1387–1391, November 1997.

[LC81] W. C. Lindsey and C. M. Chie. A survey of digital phase-locked loops. *Proc. IEEE*, 69(4):410–431, April 1981.

[LCP94] C. Long, K. Chugg, and A. Polydoros. Further results in likelihood classification of QAM signals. In *Proc. IEEE MILCOM*, volume 1, pages 57–61, October 1994.

[Lie84] F. F. Liedtke. Computer simulation of an automatic classification procedure for digitally modulated communication signals with unknown parameters. *Signal Processing*, 6(4):311–323, August 1984.

[LP95] N. Lay and A. Polydoros. Modulation classification of signals in unknown ISI environments. In *Proc. IEEE MILCOM*, volume 1, pages 170–174, November 1995.

[LTD01] S. Lesage, J-Y. Tourneret, and P. M. Djuric. Classification of digital modulation by MCMC sampling. In *Proc. IEEE ICASSP*, volume 4, pages 2553–2555, May 2001.

[LV83] A. Leclert and P. Vandamme. Universal carrier recovery loop for QASK and PSK signal sets. *IEEE Trans. Commun.*, 31(1):130–136, January 1983.

[Men91] J. M. Mendel. Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications. In *Proc. IEEE*, volume 79, pages 278–305, March 1991.

[MH81] K. Mutota and K. Hirade. GMSK modulation for digital mobile telephony. *IEEE Trans. Commun.*, 29:1044–1050, July 1981.

[MLL98] P. Marchand, J. L. Lacoume, and C. LeMartret. Multiple hypothesis classification based on cyclic cumulants of different orders. In *Proc. IEEE ICASSP*, volume 4, pages 2157–2160, May 1998.

[NA97] A. K. Nandi and E. E. Azzouz. Modulation recognition using artificial neural networks. *Signal Processing*, 56:165–175, January 1997.

[NA98] A. K. Nandi and E. E. Azzouz. Algorithms for automatic recognition of communication signals. *IEEE Trans. Commun.*, 46(4):431–436, April 1998.

[PAP00] P. Panagiotou, A. Anastasoupoulos, and A. Polydoros. Likelihood ratio tests for modulation classification. In *Proc. IEEE MILCOM*, volume 2, October 2000.

[PK90] A. Polydoros and K. Kim. On the detection and classification of quadrature digital modulation in broad-band noise. *IEEE Trans. Commun.*, 38:1199–1211, August 1990.

[Pro01] J. G. Proakis. *Digital Communications.* McGraw Hill, 2001.

[Rab89] L. Rabiner. A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proc. IEEE*, 77(2):257 – 286, February 1989.

[Rei92] J. Reichert. Automatic classification of communication signals using higher order statistics. In *Proc. IEEE ICASSP*, volume 5, pages 221–224, March 1992.

[Rob98] C. P. Robert. *Discretization and MCMC Convergence Assessment.* Berlin: Springer-Verlag, 1998.

[SBS00] A. Swami, S. Barbarossa, and B. Sadler. Blind source separation and signal classification. In *Proc. ASILOMAR*, volume 2, pages 1187–1191, November 2000.

[SBY00] C. M. Spooner, W. A. Brown, and G. K. Yeung. Automatic radio-frequency environment analysis. In *Proc. ASILOMAR*, volume 2, pages 1181–1186, November 2000.

[SH92] S. S. Soliman and S. Z. Hsue. Signal classification using statistical moments. *IEEE Trans. Commun.*, 40:908–916, May 1992.

[Sil99] J. A. Sills. Maximum likelihood modulation classification for PSK/QAM. In *Proc. IEEE MILCOM*, volume 1, pages 217–220, November 1999.

[SKKR97] C. Schreyogg, C. Kittel, U. Kressel, and J. Reichert. Robust classification of modulation types using spectral features applied to HMM. In *Proc. IEEE MILCOM*, volume 3, pages 1377–1381, November 1997.

[SM96] P. C. Sapiano and J. D. Martin. Maximum likelihood PSK classifier. In *Proc. IEEE MILCOM*, volume 3, pages 1010–1014, October 1996.

[SMH95] P. C. Sapiano, J. Martin, and R. Holbeche. Classification of PSK signals using the DFT of phase histogram. In *Proc. IEEE ICASSP*, volume 3, pages 1868–1871, May 1995.

[Spo95] C. M. Spooner. Classification of co-channel communication signals using cyclic cumulants. In *Proc. ASILOMAR*, volume 1, pages 531–536, November 1995.

[Spo01] C. M. Spooner. On the utility of sixth-order cyclic cumulants for RF signal classification. In *Proc. ASILOMAR*, volume 2, pages 890–897, November 2001.

[SR97] C. Schreyogg and J. Reichert. Modulation classification of QAM schemes using the DFT of phase histogram combined with modulus information. In *Proc. IEEE MILCOM*, volume 3, pages 1372–1376, November 1997.

[SS00] A. Swami and B. Sadler. Hierarchical digital modulation classification using cumulants. *IEEE Trans. Commun.*, 48(3):416–429, March 2000.

[TM99] S. Taira and E. Murakami. Automatic classification of analogue modulation signals by statistical parameters. In *Proc. IEEE MILCOM*, volume 1, pages 202–207, 1999.

[UIK00] K. Umebayashi, S. Ishii, and R. Kohno. Blind adaptive estimation of modulation scheme for software defined radio. In *Proc. PIMRC*, volume 1, pages 43–47, September 2000.

[VMP+02] M. Visintin, M. Mondin, M. Pent, F. Dovis, M.Falletti, and F. Sellone. End-to-end study of GMSK modulation. Technical Report 14295/00/DCS, ESA, 2002.

[VTK02] J. Venalainen, L. Terho, and V. Koivunen. Modulation classification in fading multipath channel. In *Proc. ASILOMAR*, pages 1890–1894, Pacific Grove, CA, November 2002.

[VV02] E. Vassallo and M. Visintin. Carrier phase synchronization for GMSK signals. *Int. J. Satell. Commun.*, 20(6):391–415, November 2002.

[WM00] W. Wei and J. M. Mendel. Maximum likelihood classification for digital amplitude-phase modulation. *IEEE Trans. Commun.*, 48:189–193, February 2000.

[WN01] M. L. D. Wong and A. K. Nandi. Automatic digital modulation recognition using spectral and statistical features with multi-layer perceptrons. In *Proc. Int. Symp. Signal Processing and Its Applications*, volume 2, pages 390–393, August 2001.

[YL98] Y. Yang and C. H. Liu. An asymptotic optimal algorithm for modulation classification. *IEEE Comm. Letters*, 2:117–119, May 1998.

[YS91a] Y. Yang and S. S. Soliman. Optimum classifier for M-ary PSK signals. In *Proc. ICC*, volume 3, pages 1693–1697, June 1991.

[YS91b] Y. Yang and S. S. Soliman. Statistical moments based classifier for MPSK signals. In *Proc. GLOBECOM*, volume 1, pages 72–76, December 1991.

[YS95] Y. Yang and S. S. Soliman. An improved moment-based algorithm for signal classification. *Signal Processing*, 43:231–244, May 1995.

[YS97] Y. Yang and S. S. Soliman. A suboptimal algorithm for modulation classification. *IEEE Trans. Aerosp. Electron. Syst.*, 33:38–45, January 1997.

[YSS03] Z. Yu, Y. Q. Shi, and W. Su. M-ary frequency shift keying signal classification based on discrete Fourier transform. In *Proc. IEEE MILCOM*, volume 2, pages 1167–1172, October 2003.

AUTEUR : Anchalee PUENGNIM

TITRE : Classification of linear and nonlinear modulations using Bayesian methods

---

RESUME en anglais

This thesis studies classification of digital linear and nonlinear modulations using Bayesian methods. Modulation recognition consists of identifying, at the receiver, the type of modulation signals used by the transmitter. It is important in many communication scenarios, for example, to secure transmissions by detecting unauthorized users, or to determine which transmitter interferes the others.

The received signal is generally affected by a number of impairments. We propose several classification methods that can mitigate the effects related to imperfections in transmission channels. More specifically, we study three techniques to estimate the posterior probabilities of the received signals conditionally to each modulation. The first technique estimates the unknown parameters associated with various imperfections using a Bayesian approach coupled with Markov Chain Monte Carlo (MCMC) methods. A second technique uses the Baum Welch (BW) algorithm to estimate recursively the posterior probabilities and determine the most likely modulation type from a catalogue. The last method studied in this thesis corrects synchronization errors (phase and frequency offsets) with a phase-locked loop (PLL).

The classification algorithms considered in this thesis can recognize a number of linear modulations such as Quadrature Amplitude Modulation (QAM), Phase Shift Keying (PSK), and nonlinear modulations such as Gaussian Minimum Shift Keying (GMSK).

AUTEUR : Anchalee PUENGNIM

TITRE : Classification de modulations linéaires et non-linéaires à l'aide de méthodes Bayésiennes

DIRECTEUR DE THESE : Jean-Yves TOURNERET

LIEU ET DATE DE SOUTENANCE : Salle des théses de l'ENSEEIHT, le 26 septembre 2008

RESUME en français

La reconnaissance de modulations numériques consiste à identifier, au niveau du récepteur d'une chaîne de transmission, l'alphabet auquel appartiennent les symboles du message transmis. Cette reconnaissance est nécessaire dans de nombreux scénarios de communication, afin, par exemple, de sécuriser les transmissions pour détecter d'éventuels utilisateurs non autorisés ou bien encore de déterminer quel terminal brouille les autres.

Le signal observé en réception est généralement affecté d'un certain nombre d'imperfections, dues à une synchronisation imparfaite de l'émetteur et du récepteur, une égalisation imparfaite du canal de transmission. Nous proposons plusieurs méthodes de classification qui permettent d'annuler les effets liés aux imperfections de la chaîne de transmission. Les symboles reçus sont alors corrigés puis comparés à ceux du dictionnaire des symboles transmis. Les algorithmes considérés ont permis de reconnaître un certain nombre de modulations linéaires de types QAM et PSK mais aussi des modulations non linéaires de type GMSK.

TITRE et résumé en anglais au recto

MOTS-CLES : DIGITAL MODULATION CLASSIFICATION, LINEAR AND NON-LINEAR MODULATIONS, BAYESIAN METHODS, PARAMETER ESTIMATION, BAUM-WELCH ALGORITHM, MARKOV CHAIN MONTE CARLO METHODS, GAUSSIAN MINIMUM SHIFT KEYING

DISCIPLINE ADMINISTRATIVE : Signal, Image, Acoustique et Optimisation

---

INTITULE ET ADRESSE DU LABORATOIRE : IRIT - Site ENSEEIHT, 2 rue Charles Camichel, BP 7122, 31071 Toulouse, CEDEX 7