Modulation Classification in Fading Channels using Expectation Maximization

A Project Report

submitted by

CHANUMOLU LENIN KUMAR

in partial fulfilment of the requirements

for the award of the degree of

MASTER OF TECHNOLOGY

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY MADRAS

JUNE 2021

CERTIFICATE

This is to certify that the thesis titled **Modulation Classification in Fading Channels using Expectation Maximization**, submitted by **CHANUMOLU LENIN KUMAR (EE19M017)**, to the Indian Institute of Technology Madras, for the award of the degree of **Master of Technology**, is a bona fide record of the research work done by him under my supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Place: Chennai

Date: June 2021

Dr. Sheetal Kalyani
Project Guide,
Associate Professor,
Dept. of Electrical Engineering,
IIT Madras, 600 036

ACKNOWLEDGEMENTS

I am greatly indebted to Dr. Sheetal Kalyani for guiding me through the entire course of my M.Tech project. She always took the time and effort to discuss the problem and to suggest different methods to experiment. Her valuable remarks always gave new directions to my project.

I am thankful to all the professors whose courses helped me improve my knowledge in the areas of Communication and Signal Processing through the course of the two years of my M.Tech program. Their classes always inspired me to think beyond class rooms into more practical scenarios.

A special thanks to Nancy Nayak who helped me a lot in doing this project. This project would not have been possible without her contributions and insightful observations. A special word of thanks to all the wonderful people I met at IITM without whom life would not have been the same.

ABSTRACT

KEYWORDS: Modulation Classification, Maximum Likelihood Estimation, Expectation Maximization, Fading Channel, k-means Clustering, Pattern Recognition,

Modulation Classification is the key step in adaptive communications, where the transmitter changes the modulation according to the channel conditions. In such scenarios the receiver needs to identify the modulation present in the incoming signal. In a simple AWGN channel the task of Modulation Classification may be easy but in fading channels with carrier phase and frequency offsets and other channel impairments, this task becomes difficult. In this thesis we tried to do the Modulation Classification task in fading channels using Expectation Maximization Algorithm. First we tried to do the MC in flat fading channels where we use k-means clustering algorithm to initialize the unknown channel parameters for a better estimation using EM algorithm. Then we extend similar procedure to the frequency selective fading channel conditions. Simulations were done for both the cases. In flat fading channel model 4 modulations (BPSK,QPSK,16QAM and 64 QAM) are considered. In frequency selective fading channel model 3 modulations (BPSK, QPSK,8PSK) are considered and the classification accuracy plots against SNR are obtained.

TABLE OF CONTENTS

Table of Contents (continued) **Page**

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

MC	Modulation Classification
PR	Pattern Recognition
MIMO	Multiple Input Multiple Output
EM	Expectation Maximization
AWGN	Additive White Gaussian Noise
BPSK	Binary Phase Shift Keying
QPSK	Quadrature Phase Shift Keying
QAM	Quadrature Amplitude Modulation
ALRT	Average Likelihood Ratio Test
GLRT	Generalized Likelihood Ratio Test
HLRT	Hybrid Likelihood Ratio Test
SNR	Signal to Noise Ratio

NOTATION

$\|x\|$	Absolute value of x
x^*	Complex conjugate of x
$\underline{a}^T$	Transpose of $\underline{a}$
$\underline{a}^H$	Transposed conjugate (Hermitian) of $\underline{a}$
$E[x]$	Expectation of x
c_m	m^{th} cluster center
h	fading coefficient
$\underline{h}$	channel vector
σ_v^2	Noise variance

CHAPTER 1

INTRODUCTION

1.1 Introduction

Modulation classification is the key step in many military and civilian applications. It is used in Software Defined Radios, adaptive modems and cognitive radios. In such systems depending on the environmental (channel) conditions the transmitter will change the modulation scheme used. Then, the receiver needs to identify the modulation type by itself and need to apply the corresponding demodulation scheme. For example, when the channel is good the transmitter can use a modulation with higher constellations. But to achieve the same BER(Bit Error Rate) at worst channel conditions one can not use higher constellations and hence it is required to use a modulation of lower constellations. In such scenarios the receiver has to identify the modulation type present in the coming signal and use the corresponding demodulation. Having said that we understand that Modulation Classification (MC) is an intermediate step between signal detection and demodulation. There has been a lot of work being done in the area of MC for several years. Several techniques are proposed for MC and some of them were compared in Dobre *et al.* (2007) and Hazza *et al.* (2013). In Hazza *et al.* (2013) feature based MC techniques were compared. In Dobre *et al.* (2007) likelihood based MC techniques such as ALRT, GLRT, HLRT and many other techniques were compared. With the advancements in deep learning and machine learning, several techniques were proposed for doing MC which use supervised learning. Many of these techniques use Support Vector Machines (SVM), Neural Networks(NN) and some Pattern Recognition (PR) techniques for doing MC. Some use a combination of these techniques also. In Huynh-The *et al.* (2020) and Hermawan *et al.* (2020) neural networks are proposed which use convolution layers, max pooling layers and ReLu activation functions. They used regularization with dropout and Gaussian noise layers also. However all these techniques require training data. In this work we perform MC in digitally modulated signals without the knowledge of channel parameters.

1.2 Related work done

There are some papers which use Expectation Maximization for doing modulation classification. Soltanmohammadi and Naraghi-Pour (2013) use EM algorithm for doing channel parameter estimation and then use a Pattern Matching technique to find the modulation. Some part of this work is based on this paper with some modifications. In Zhu and Nandi (2015) modulation classification in flat fading MIMO systems is performed using EM algorithm(Moon (1996)) and Maximum Likelihood classification. Chavali and da Silva (2011) also used EM algorithm for doing Modulation Classification. Here the noise added is not Gaussian but is from a mixture of Gaussians.

1.3 Organization of this thesis

This thesis is organized as follows:

$Chapter$ 2 explains the modulation classification in flat fading channels, where maximum likelihood estimation of the channel parameters is done via Expectation Maximization algorithm. For a good initialization of the parameters k-means clustering algorithm is used. Then Maximum Likelihood Classification is done using the maximum likelihood estimates obtained using the EM algorithm. Lastly, simulation results are included.

$Chapter$ 3 extends the modulation classification method discussed in $chapter$ 2 to frequency selective fading channels. First, signal model is discussed and then the parameter estimation using EM algorithm is presented. Finally Maximum Likelihood Classification and simulation results are presented.

$Chapter$ 4 summarizes the work done and provides some concluding remarks and observations.

CHAPTER 2

MODULATION CLASSIFICATION IN FLAT FADING CHANNELS USING EM

2.1 Introduction

In this chapter we will see how MC can be done in flat fading channels without any channel information such as channel coefficient(h) and noise variance. The signal model and parameter estimation using Expectation Maximization (EM) are explained in the subsequent sections. The whole MC technique goes as follows: For a pool of modulation types we will assume each modulation at a time and do the parameter estimation and then find the likelihood of that modulation with the estimated channel parameters. Finally a maximum likelihood classifier is used for doing MC. That is whichever modulation has the maximum likelihood it is assumed to be present in the incoming signal. In this entire process the noise and channel parameters are unknown. Part of this procedure is the main theme of the work done by Soltanmohammadi and Naraghi-Pour (2013), but there they don't use Maximum Likelihood Classifier. Instead they use some pattern recognition techniques for doing MC after finding the likelihoods.Also they are initializing the noise variance to a random number. In this chapter we will see how this procedure works in flat fading channels and in the next chapter we will extend it to the case of frequency selective fading channels where channel is assumed as a multi-tap FIR filter.

2.2 Signal Model

The received signal model assumed is given by $r_n = ha_n + v_n$ where, r_n is the n^{th} received sample, a_n is the n^{th} transmitted symbol and v_n is the complex AWGN(Additive

White Gaussian Noise) present in the n^{th} received sample. Here we don't know the fading coefficient h, transmitted symbol a_n and also the noise variance σ_v^2. We need to identify the modulation type present in $\{r_n\}$ just by using $\{r_n\}$. The complex valued fading coefficient h is assumed to be a constant and does not have any probability distribution (probability density function). The fading coefficient h is assumed to include the effect of multi-path fading and path loss and also the unknown energy of the transmitted symbols. Hence the transmitted symbol a_n has an average energy of 1. That is $E[|a_n|^2] = 1$. $\{v_n\}$ is a complex valued circularly symmetric White Gaussian Noise sample. The transmitted symbol $a_n \in \mathcal{S} = \{s_1, s_2, s_3..., s_M\}$, where $\mathcal{S}$ is a set of M unit average energy constellation points belonging to a given modulation type. For example in BPSK, $\mathcal{S} = \{+1, -1\}$ and in QPSK, $\mathcal{S} = \{\frac{1+1i}{\sqrt{2}}, \frac{1-1i}{\sqrt{2}}, \frac{-1+1i}{\sqrt{2}}, \frac{-1-1i}{\sqrt{2}}\}$. Our aim is to estimate the channel coefficient h and the noise variance σ_v^2 from the received samples $\{r_n\}$ assuming a particular modulation type present in it.

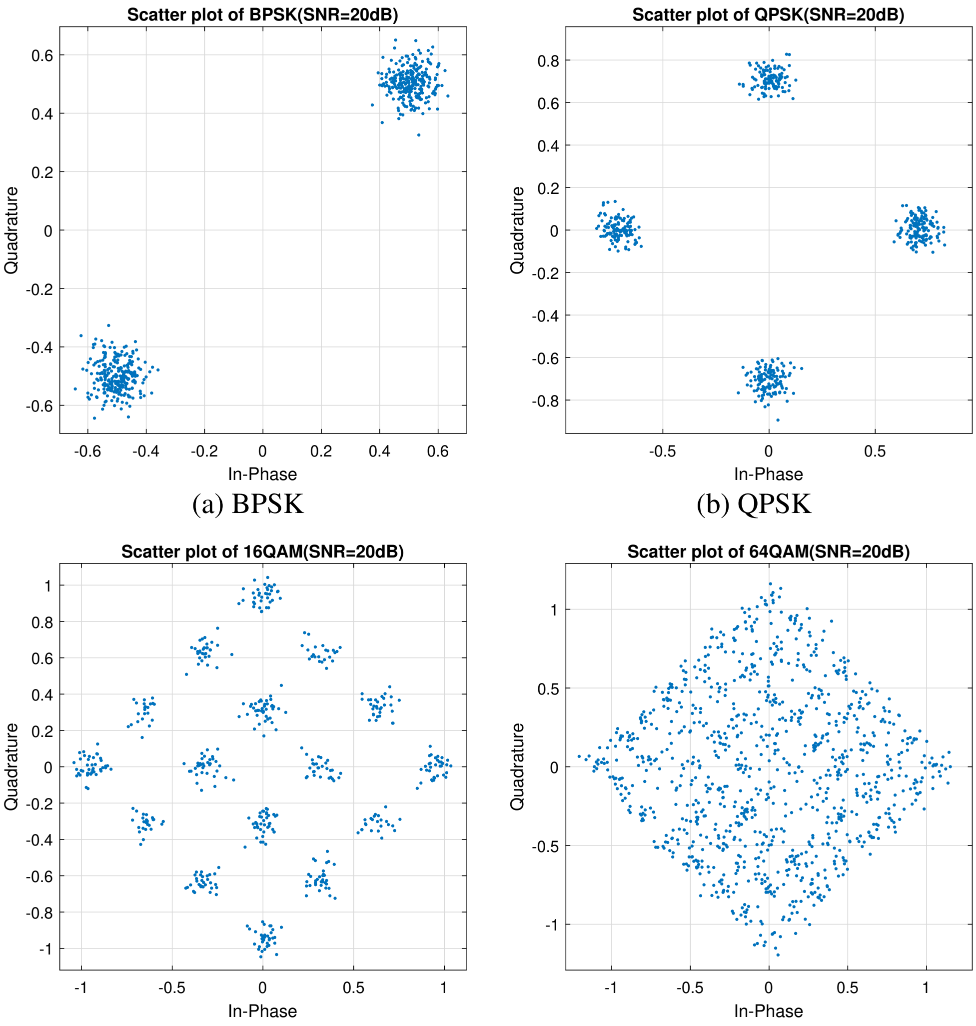
Let us denote the transmitted symbol vector by $\underline{\mathbf{a}} = [a_1 a_2 a_3 ... a_N]^T \in \mathcal{S}^N$ and the received symbol vector by $\underline{\mathbf{r}} = [r_1 r_2 r_3 ... r_N]^T$. The superscript T indicates transpose of a vector. N is the total number of received samples. The scatter plots of received samples in 4 modulations {BPSK, QPSK, 16-QAM and 64-QAM} at an SNR of 20 dB are shown in figure 2.1. Since the fading channel coefficient has a phase value of $45°$ the constellations are rotated by $45°$.

The unknown parameter vector is denoted by $\Theta \equiv (h, \sigma_v^2)$. Let us define the binary class matrix $\mathcal{Z} = \{z_{nm}\}_{N \times M}$ which is an $N \times M$ matrix and denotes the membership of the received symbol r_n to a point in the constellation. That is if $z_{nm} = 1$ then the received sample r_n is obtained from the transmitted symbol s_m. In short $z_{nm} = 1$ if $a_n = s_m$ and 0 otherwise. Only one element in any row of $\mathcal{Z}$ is unity and all the other elements of the row are zero. This structure makes it a binary class matrix. The Maximum Likelihood (ML) estimator for Θ is given by

$$\hat{\Theta} = \arg\max_{\mathcal{Z}} Pr(\underline{\mathbf{r}}/\Theta)$$

where

$$Pr(\underline{\mathbf{r}}/\Theta) = \sum_{\mathcal{Z}} Pr(\underline{\mathbf{r}}, \mathcal{Z}/\Theta)$$



(a) BPSK (b) QPSK (c) 16QAM (d) 64QAM

Fig. 2.1: The scatter plots of received samples in (a)BPSK, (b)QPSK, (c)16QAM, (d)64QAM modulations at an SNR of 20 dB with $h = 0.5 + 0.5i$.

This is a mixture model and hence it does not have an analytical solution. However it can be solved iteratively using Expectation Maximization (EM) which is presented in appendix A. But the convergence of the EM algorithm depends on the initial estimate of the parameter. To have a good initial estimate for both h and σ_v^2 we use two different techniques which are described in the next sections.

2.3 Parameter Estimation

To have a good initial estimate of the channel coefficient h for the EM algorithm to converge faster, we use k-means estimation which is described as follows.

2.3.1 k-means Estimation of Channel Coefficient

Let us define the objective function $\mathcal{J}$ as

$$\mathcal{J} = \sum_{m=1}^{M}\sum_{n=1}^{N} z_{nm}|r_n - hs_m|^2$$

Now, the task is to minimize the function $\mathcal{J}$. That is to find the values of h and $\mathcal{Z}$ that minimize $\mathcal{J}$. This can be done iteratively as follows: For a fixed value of h the value of $\mathcal{Z}$ which minimizes the function $\mathcal{J}$ is given by

$$z_{nm} = \begin{cases} 1 & \text{if } m = \arg\min_{m} |r_n - hs_m|^2 \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

Now for a given value of $\mathcal{Z}$, to minimize $\mathcal{J}$ with respect to h we find the derivative of $\mathcal{J}$ with respect to h and equate it to zero. That is we solve for $\frac{\partial \mathcal{J}}{\partial h} = 0$

$$\begin{aligned} \frac{\partial \mathcal{J}}{\partial h} &= 0 \\ \frac{\partial}{\partial h}\left\{\sum_{m=1}^{M}\sum_{n=1}^{N} z_{nm}|r_n - hs_m|^2\right\} &= 0 \\ \sum_{m=1}^{M}\sum_{n=1}^{N} z_{nm}\frac{\partial}{\partial h}\{|r_n - hs_m|^2\} &= 0 \\ \sum_{m=1}^{M}\sum_{n=1}^{N} z_{nm}\frac{\partial}{\partial h}\{(r_n - hs_m)^*(r_n - hs_m)\} &= 0 \\ \sum_{m=1}^{M}\sum_{n=1}^{N} z_{nm}\frac{\partial}{\partial h}\{|r_n|^2 - r_n^* hs_m - h^* s_m^* r_n + |h|^2|s_m|^2\} &= 0 \end{aligned}$$

$$\sum_{m=1}^{M}\sum_{n=1}^{N} z_{nm}\{-r_n^* s_m + h^*|s_m|^2\} = 0$$

$$\sum_{m=1}^{M}\sum_{n=1}^{N} z_{nm} r_n^* s_m = \sum_{m=1}^{M}\sum_{n=1}^{N} z_{nm}|s_m|^2 h^*$$

$$h^* = \frac{\sum_{m=1}^{M}\sum_{n=1}^{N} z_{nm} r_n^* s_m}{\sum_{m=1}^{M}\sum_{n=1}^{N} z_{nm}|s_m|^2}$$

where the superscript $*$ denotes complex conjugate. So the value of h that minimizes $\mathcal{J}$ is given by

$$h = \frac{\sum_{m=1}^{M}\sum_{n=1}^{N} z_{nm} r_n s_m^*}{\sum_{m=1}^{M}\sum_{n=1}^{N} z_{nm}|s_m|^2} \tag{2.2}$$

The repeated execution of equations 2.1 and 2.2 until a stopping criterion is satisfied will give a best initial estimate of the channel coefficient h for the EM parameter estimation.

2.3.2 k-means Estimation of σ_v^2

The initial estimate of σ_v^2 is also obtained by general k-means algorithm. This algorithm can be simply explained as follows: Let us say we have 4 constellation points in the modulation, then there will be four clusters in the scatter plot of the received samples (as shown in 2.1b) The centers or means of these clusters are denoted by c_m $(m = 1, 2, 3, 4)$ Now the k-means algorithm is described as follows:

Algorithm 1: k-means Clustering Algorithm

```
Initialize the cluster centers c_m to random values;
while c_m's do not change by a certain percentage do
    for n = 1 to N do
        assign r_n to cluster m if r_n is close to c_m
    end
    compute the new cluster centers c_m
end
```

The above algorithm returns final cluster centers and cluster memberships. Now take any cluster and find the variance of that cluster. This variance is used as an initial estimate for the EM parameter estimation described in the next section. The initial cluster centers need not be initialized randomly and MATLAB uses k-means++ algorithm for initialization of the cluster centers.

2.3.3 Parameter Estimation using EM

As described earlier in section 2.2 we need to find Θ that maximizes $Pr(\underline{\mathbf{r}}/\Theta)$. Which means finding

$$\hat{\Theta} = \arg\max_{\mathcal{Z}} Pr(\underline{\mathbf{r}}/\Theta)$$

Now,

$$\begin{aligned} Pr(\underline{\mathbf{r}}, \mathcal{Z}/\Theta) &= Pr(\underline{\mathbf{r}}/\mathcal{Z}, \Theta) Pr(\mathcal{Z}/\Theta) \\ &= \frac{1}{M^N} \prod_{m=1}^{M} \prod_{n=1}^{N} \left[\frac{1}{\pi\sigma_v^2} \exp\left(-\frac{|r_n - hs_m|^2}{\sigma_v^2} \right) \right]^{z_{nm}} \end{aligned}$$

Since maximizing the likelihood function and maximizing the log of the likelihood function are same, we maximize the Log Likelihood function which is given by

$$\begin{aligned} L(\Theta; \underline{\mathbf{r}}, \mathcal{Z}) &= \log \Pr(\underline{\mathbf{r}}, \mathcal{Z}/\Theta) \\ &= -N \log M - \sum_{m=1}^{M} \sum_{n=1}^{N} z_{nm} \left[\log\left(\pi\sigma_v^2\right) + \frac{|r_n - hs_m|^2}{\sigma_v^2} \right] \end{aligned}$$

Since we can not find the parameters that maximize the Log Likelihood function analytically we use the Expectation Maximization Algorithm for finding the maximum likelihood estimates of the parameters (because the binary class matrix used z_{nm} used here is not constant and it is a random variable it made the problem a mixture model). Here the latent (unknown) variables are z_{nm}. In EM algorithm we find the maximum likelihood estimates in two iterative steps namely E-Step and the M-Step.

a) E-Step: In this step the Expectation of the Log Likelihood function with respect to the latent variables z_{nm} is calculated conditioned on the previous estimates $\Theta^{(old)} = (h^{(old)}, \sigma_v^{2(old)})$ and the received samples $\underline{\mathbf{r}}$.

$$\begin{aligned} \Psi\left(\Theta; \Theta^{(\text{old})}\right) &= E_{\mathcal{Z}}\left[L(\Theta; \underline{\mathbf{r}}, \mathcal{Z}) / \underline{\mathbf{r}}, \Theta^{(\text{old})} \right] \\ &= -N \log M - \sum_{m=1}^{M} \sum_{n=1}^{N} \gamma(n,m) \left[\log\left(\pi\sigma_v^2\right) + \frac{|r_n - hs_m|^2}{\sigma_v^2} \right] \end{aligned} \tag{2.3}$$

where,

$$
\begin{aligned}
\gamma(n,m) &= E\left[z_{nm}/\underline{\mathbf{r}},\Theta^{(\text{old})}\right] \\
&= 1 \times Pr\left(z_{nm}=1/\underline{\mathbf{r}},\Theta^{(\text{old})}\right) + 0 \times Pr\left(z_{nm}=0/\underline{\mathbf{r}},\Theta^{(\text{old})}\right) \\
&= Pr\left(z_{nm}=1/\underline{\mathbf{r}},\Theta^{(\text{old})}\right) \\
&= \frac{\exp\left(-\frac{\left|r_n - h^{(\text{old})}s_m\right|^2}{\sigma_u^{2(\text{old})}}\right)}{\sum_{j=1}^{M}\exp\left(-\frac{\left|r_n - h^{(\text{old})}s_j\right|^2}{\sigma_u^{2(old)}}\right)}
\end{aligned} \tag{2.4}
$$

b) M-Step: In the M-Step we find the values of h and σ_v^2 that maximize the expectation of the Log Likelihood function $\Psi\left(\Theta;\Theta^{(\text{old})}\right)$. To do this we find the derivatives of Ψ with respect to h and σ_v^2 and equate them to zero.

$$
\begin{aligned}
\frac{\partial \Psi}{\partial h} &= 0 \\
\frac{\partial}{\partial h}\left\{-N\log M - \sum_{m=1}^{M}\sum_{n=1}^{N}\gamma(n,m)\left[\log\left(\pi\sigma_v^2\right) + \frac{|r_n - hs_m|^2}{\sigma_v^2}\right]\right\} &= 0 \\
\sum_{m=1}^{M}\sum_{n=1}^{N}\gamma(n,m)\frac{\partial}{\partial h}\{|r_n - hs_m|^2\} &= 0 \\
\sum_{m=1}^{M}\sum_{n=1}^{N}\gamma(n,m)\frac{\partial}{\partial h}\{(r_n - hs_m)^*(r_n - hs_m)\} &= 0 \\
\sum_{m=1}^{M}\sum_{n=1}^{N}\gamma(n,m)\frac{\partial}{\partial h}\{|r_n|^2 - r_n^* h s_m - h^* s_m^* r_n + |h|^2|s_m|^2\} &= 0 \\
\sum_{m=1}^{M}\sum_{n=1}^{N}\gamma(n,m)\{-r_n^* s_m + h^*|s_m|^2\} &= 0 \\
\sum_{m=1}^{M}\sum_{n=1}^{N}\gamma(n,m) r_n^* s_m &= \sum_{m=1}^{M}\sum_{n=1}^{N}\gamma(n,m)|s_m|^2 h^*
\end{aligned}
$$

which gives

$$
h^* = \frac{\sum_{m=1}^{M}\sum_{n=1}^{N}\gamma(n,m) r_n^* s_m}{\sum_{m=1}^{M}\sum_{n=1}^{N}\gamma(n,m)|s_m|^2}
$$

So the value of h that maximizes $\Psi\left(\Theta;\Theta^{(\text{old})}\right)$ is given by

$$
h = \frac{\sum_{m=1}^{M}\sum_{n=1}^{N}\gamma(n,m) r_n s_m^*}{\sum_{m=1}^{M}\sum_{n=1}^{N}\gamma(n,m)|s_m|^2} \tag{2.5}
$$

We also need to find σ_v^2 that maximizes the expectation of the Log Likelihood function $\Psi\left(\Theta;\Theta^{(\text{old})}\right)$ which obtained by solving $\frac{\partial\Psi}{\partial\sigma_v^2}=0$.

$$
\begin{aligned}
\frac{\partial\Psi}{\partial\sigma_v^2} &= 0 \\
\frac{\partial}{\partial\sigma_v^2}\left\{-N\log M-\sum_{m=1}^{M}\sum_{n=1}^{N}\gamma(n,m)\left[\log\left(\pi\sigma_v^2\right)+\frac{|r_n-hs_m|^2}{\sigma_v^2}\right]\right\} &= 0 \\
\sum_{m=1}^{M}\sum_{n=1}^{N}\gamma(n,m)\frac{\partial}{\partial\sigma_v^2}\left\{\log\left(\pi\sigma_v^2\right)+\frac{|r_n-hs_m|^2}{\sigma_v^2}\right\} &= 0 \\
\sum_{m=1}^{M}\sum_{n=1}^{N}\gamma(n,m)\left\{\frac{\pi}{\pi\sigma_v^2}-\frac{|r_n-hs_m|^2}{\sigma_v^4}\right\} &= 0
\end{aligned}
$$

which results in

$$
\begin{aligned}
\sum_{m=1}^{M}\sum_{n=1}^{N}\frac{\gamma(n,m)}{\sigma_v^2} &= \sum_{m=1}^{M}\sum_{n=1}^{N}\frac{\gamma(n,m)|r_n-hs_m|^2}{\sigma_v^4} \\
\sigma_v^2 &= \frac{\sum_{m=1}^{M}\sum_{n=1}^{N}\gamma(n,m)|r_n-hs_m|^2}{\sum_{m=1}^{M}\sum_{n=1}^{N}\gamma(n,m)}
\end{aligned}
$$

So the value of σ_v^2 that maximizes $\Psi\left(\Theta;\Theta^{(\text{old})}\right)$ is given by

$$
\sigma_v^2=\sum_{m=1}^{M}\sum_{n=1}^{N}\frac{\gamma(n,m)|r_n-hs_m|^2}{N} \tag{2.6}
$$

Equations 2.4,2.5 and 2.6 are executed repeatedly till convergence is obtained. The values of h and σ_v^2 obtained using this EM algorithm are used in doing Maximum Likelihood Classification which is described in the next section.

2.4 Maximum Likelihood Classification

In this section we will use the estimated channel parameters using EM and do the Modulation Classification. Let $\mathcal{M}=\{\mathcal{S}_1,\mathcal{S}_2,\mathcal{S}_3...,\mathcal{S}_K\}$ denote the set of possible modulations with $\mathcal{S}_k=\{s_1^{(k)},s_2^{(k)},s_3^{(k)},...s_{M^{(k)}}^{(k)}\}$ being the k^{th} modulation. k^{th} modulation contains $M^{(k)}$ symbols in its constellation. Let $\hat{\Theta}^{(k)}=(\hat{h}^{(k)},\hat{\sigma}_v^{2(k)})$ be the estimated

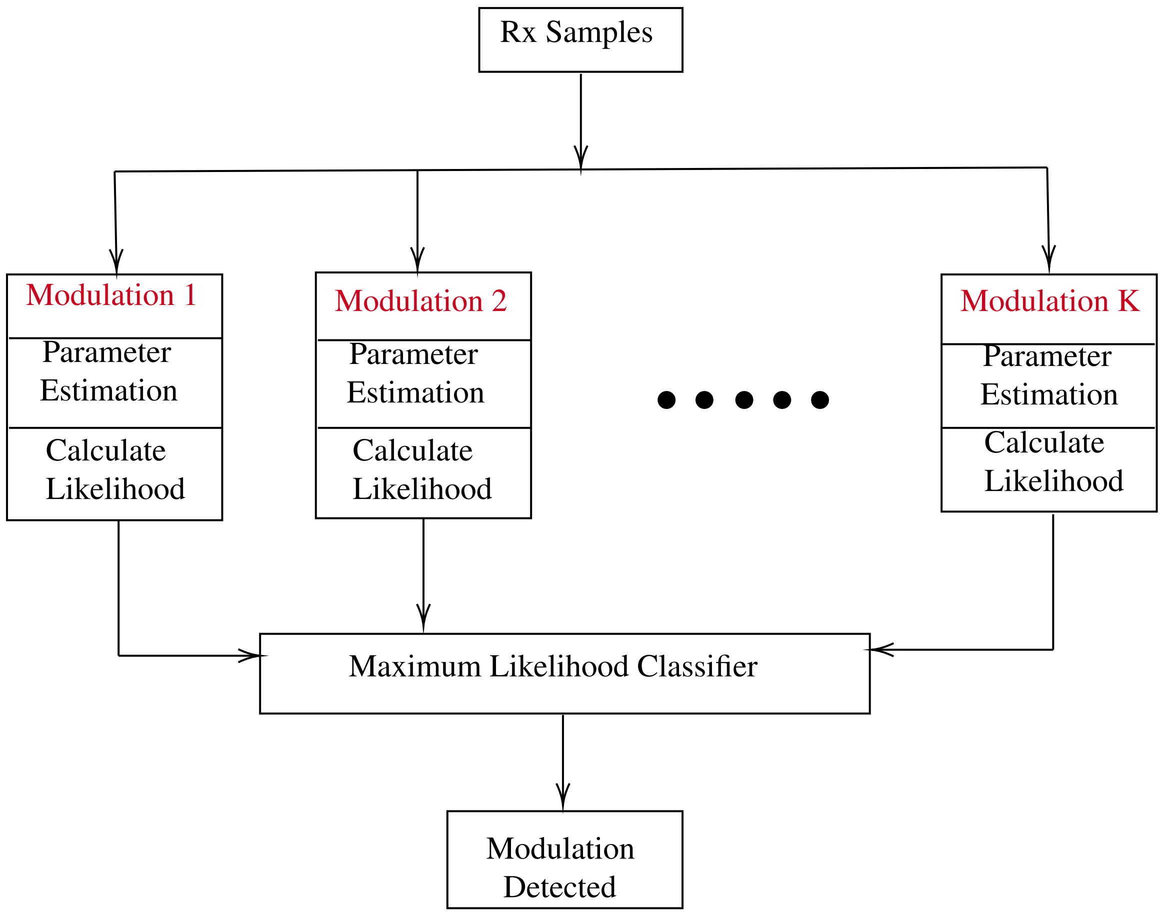


Fig. 2.2: Block Diagram of the Modulation Classification Procedure

channel parameters assuming that $\mathcal{S}_k$ is the actual modulation. After calculating these values for all modulations considered, we will calculate the Log Likelihood for each modulation which is given by

$$\begin{aligned}\mathcal{L}^{(k)}(\underline{\mathbf{r}}) &= \frac{1}{N}\log Pr(\underline{\mathbf{r}}/\mathcal{S}_k, \hat{\Theta}^{(k)}) \\ &= \frac{1}{N}\sum_{n=1}^{N}\log\left\{\frac{1}{M^{(k)}\pi\hat{\sigma}_v^{2(k)}}\sum_{m=1}^{M^{(k)}} e^{\left\{-\frac{|r_n-\hat{h}^{(k)}s_m^{(k)}|^2}{\hat{\sigma}_v^{2(k)}}\right\}}\right\}\end{aligned}$$

After calculating these likelihoods for all modulations considered, we will take the modulation with maximum likelihood as the modulation present in the received signal. That is

$$\hat{k} = \arg\max_k \mathcal{L}^{(k)}(\underline{\mathbf{r}})$$

The overall procedure can be summarized into the block diagram shown in figure 2.2.

2.5 Simulation Results

The performance of the modulation classification method discussed was examined by considering a pool of modulations containing BPSK, QPSK, 16QAM and 64QAM. Experiments were done with various lengths of the received signal. All the experiments were done for 10,000 runs over a range of Signal to Noise Ratios(SNR). SNRs ranging from 0 dB to 20 dB in steps of 2 dB are considered.

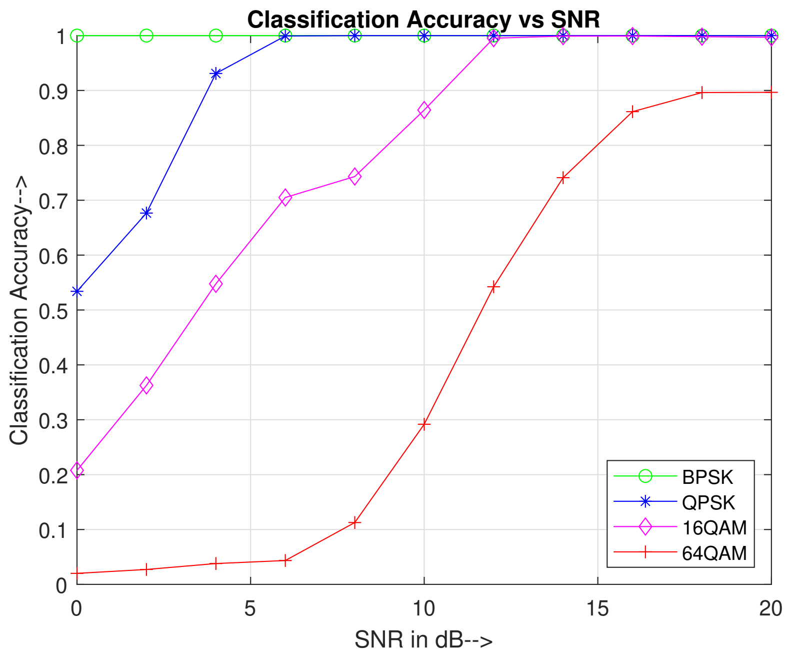


Fig. 2.3: Classification Accuracy for $N = 128$

Figure 2.3 shows the classification accuracy for all the modulations considered when the signal length is N=128. From the figure we can observe that the method is able to classify BPSK with 100% accuracy over all SNRs. It is able to classify QPSK also with almost 100% accuracy. However the method is not able to recognize higher constellations such as 16QAM and 64QAM at low SNR region. At high SNR region the method classifies all signals with 100% accuracy. When the Signal length is increased from N=128 to N=256 there is a slight improvement in the accuracy but it is not significant. for QPSK and 16QAM the accuracy improved at all SNRs but for 64QAM there is an improvement at high SNR region and degradation at low SNR region. This is shown in figure 2.4. In figure 2.5 the accuracy curves for various signal lengths are plotted.

We can observe that the accuracy is improved when signal length is increased but the improvement is not significant. This improvement can be explained by the fact that we have more information when we have more number of samples.

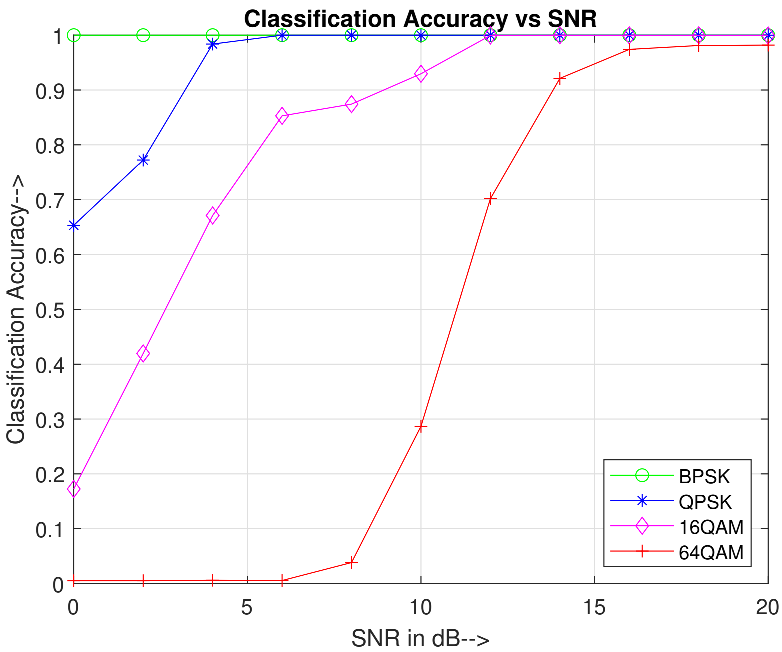


Fig. 2.4: Classification Accuracy for $N = 256$

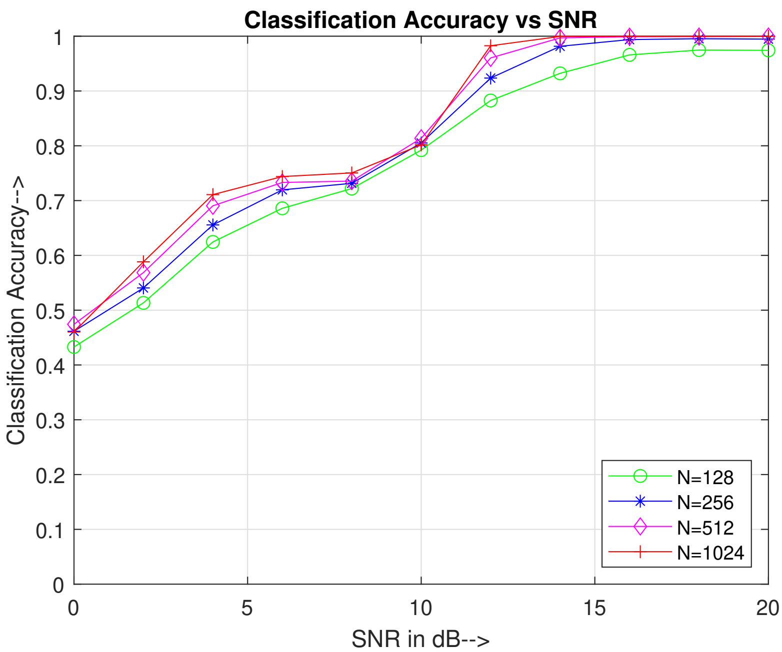


Fig. 2.5: Classification Accuracy for various signal lengths

CHAPTER 3

MODULATION CLASSIFICATION IN FREQUENCY SELECTIVE FADING CHANNELS USING EM

3.1 Introduction

In chapter 2 we have seen the modulation classification in flat fading channels where the channel coefficient h is a constant and does not follow any distribution. In this chapter we will see how the modulation classification can be done in the case of frequency selective fading channels where the channel is assumed as an FIR filter. Here also the channel taps are assumed to be constants and do not follow any distribution. The procedure we follow here is almost similar to the one we followed in flat fading case.The signal model, parameter estimation and the maximum likelihood classification are described in the following sections.

3.2 Signal Model

The frequency selective fading channel is modelled as:

$$r_n = h_0 a_n + h_1 a_{n-1} + h_2 a_{n-2} + ... + h_{L-1} a_{n-L+1} + v_n \tag{3.1}$$

where r_n is the n^{th} received sample which contains past $L-1$ transmitted symbols along with the current symbol in it. v_n is a complex-valued White Gaussian noise sequence with variance σ_v^2. a_n through a_{n-L+1} are the past L transmitted symbols from the current time instant n. The above equation 3.1 can be re-written as:

$$r_n = \begin{bmatrix} h_0 & h_1 & h_2 & ... & h_{L-1} \end{bmatrix} \begin{bmatrix} a_n \\ a_{n-1} \\ a_{n-2} \\ . \\ . \\ . \\ a_{n-L+1} \end{bmatrix} + v_n \tag{3.2}$$

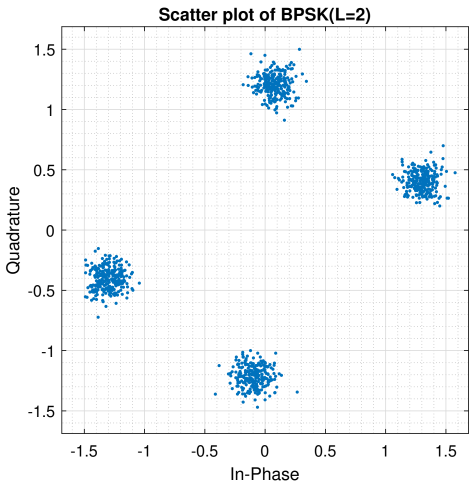
or simply

$$r_n = \underline{h}^T \underline{a} + v_n \qquad \text{for } n = 1, 2, ...N$$

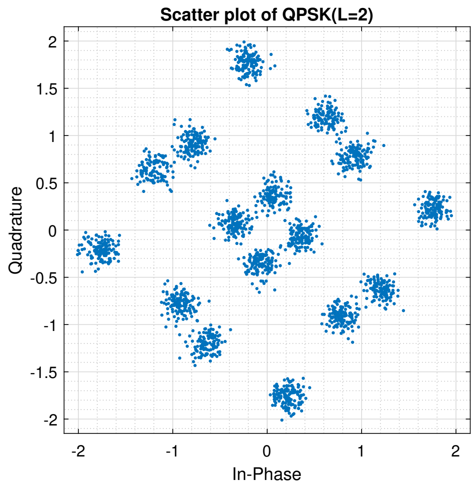
where $\underline{h} = [h_0 h_1 h_2 ... h_{L-1}]^T$ is the channel vector and the $\underline{a} = [a_n a_{n-1} a_{n-2} ... a_{n-L+1}]^T$ is the vector of past L transmitted symbols. Here T denotes the transpose of the column vector. The channel vector $\underline{h}$ is assumed as a constant and it does not follow any probability distribution. We need to identify the modulation type present in $\{r_n\}$ just by using $\{r_n\}$.

The channel vector $\underline{h}$ is assumed to include the effect of multi-path fading and path loss and also the unknown energy of the transmitted symbols. Hence the transmitted symbols a_n has an average energy of 1. That is $E[|a_n|^2] = 1$. The transmitted symbol $a_n \in \mathcal{S} = \{s_1, s_2, s_3..., s_M\}$, where $\mathcal{S}$ is a set of M unit average energy constellation points belonging to a given modulation type. For example in BPSK, $\mathcal{S} = \{+1, -1\}$ and in QPSK, $\mathcal{S} = \{\frac{1+1i}{\sqrt{2}}, \frac{1-1i}{\sqrt{2}}, \frac{-1+1i}{\sqrt{2}}, \frac{-1-1i}{\sqrt{2}}\}$. Our aim is to estimate the channel vector $\underline{h}$ and the noise variance σ_v^2 from the received samples $\{r_n\}$ assuming a particular modulation type present in it.

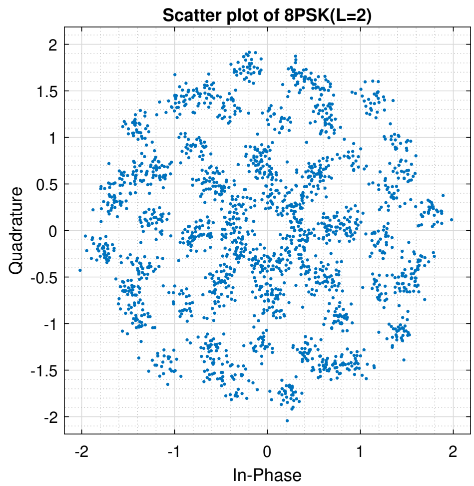
Let us denote the total number of received samples by N. Since there are M points in the constellation and L elements in the channel vector each r_n has possibility of coming from one of the M^L total possible transmit vectors. For example for $L = 2$ in BPSK($M = 2$) there will be $M^L = 2^2 = 4$ possible transmit vectors and hence there will be 4 clusters in the scatter plot of the received signal. Similarly in QPSK($M = 4$) there will be $M^L = 4^2 = 16$ clusters in the scatter plot of the received signal. The scat-



(a) BPSK(L=2)



(b) QPSK(L=2)



(c) 8PSK(L=2)

Fig. 3.1: The scatter plots of received samples in (a)BPSK, (b)QPSK and (c)8PSK modulations at an SNR of 20 dB with $\underline{h} = \{0.7 + 0.8i, 0.6 - 0.4i\}$.

ter plots of the received signals in BPSK, QPSK and 8PSK at an SNR of 20dB are shown in figure 3.1. The $L = 2$ tap channel vector assumed here is $\{0.7 + 0.8i, 0.6 - 0.4i\}$. From figure 3.1 we can observe that in 3.1a BPSK there are 4 clusters, in 3.1b QPSK there are 16 clusters and in 3.1c 8PSK there are 64 clusters in the scatter plot of the received signal.

The unknown parameter vector is denoted by $\Theta \equiv (\underline{h}, \sigma_v^2)$. The received signal vector is denoted by $\underline{\mathbf{r}} = \{r_1 r_2 r_3 ... r_N\}^T$. Let us define the binary class matrix $\mathcal{Z} = \{z_{nm}\}_{N \times M^L}$ which is an $N \times M^L$ matrix and denotes the membership of the received symbol r_n to a possible vector of past L transmitted symbols. That is if $z_{nm} = 1$ then the received sample r_n is obtained from the transmit vector $\underline{s}_m$. In short $z_{nm} = 1$

if $\underline{a} = \underline{s}_m$ and 0 otherwise. Only one element in any row of $\mathcal{Z}$ is unity and all the other elements of the row are zero. The Maximum Likelihood (ML) estimator for Θ is given by

$$\hat{\Theta} = \arg\max_{\mathcal{Z}} Pr(\underline{\mathbf{r}}/\Theta)$$

where

$$Pr(\underline{\mathbf{r}}/\Theta) = \sum_{\mathcal{Z}} Pr(\underline{\mathbf{r}}, \mathcal{Z}/\Theta)$$

Similar to chapter 2 this problem can be solved iteratively using Expectation Maximization (EM). But the convergence of the EM algorithm depends on the initial estimate of the parameter. To have a good initial estimate for $\underline{h}$ we use k-means initialization which is described in the next section.

3.3 Parameter Estimation

To have a good initial estimate of the channel vector $\underline{h}$ for the EM algorithm to converge faster (and also to a global optimum) we use k-means estimation which is described as follows.

3.3.1 K-Means Initialization of Channel Vector

We define an objective function $\mathcal{J}$ as

$$\mathcal{J} = \sum_{m=1}^{M^L} \sum_{n=1}^{N} z_{nm} \left| r_n - \underline{h}^T \underline{s}_m \right|^2$$

Here M^L is the number of possible transmitted symbol vectors and N is the total number of received samples. We need to find the $\underline{h}$ and z_{nm} that minimizes the above cost function. The values that minimize the above cost function are obtained iteratively as follows:

$$z_{nm} = \begin{cases} 1 & \text{if } m = \arg\min_{m} \left| r_n - \underline{h}^T \underline{s}_m \right|^2 \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

To minimize $\mathcal{J}$ with respect to $\underline{h}$ we find its gradient and equate it to the zero vector. The cost function is rewritten as

$$\mathcal{J} = \sum_{m=1}^{M^L} \sum_{n=1}^{N} z_{nm}[|r_n|^2 - \underline{h}^T \underline{s}_m r_n^* - r_n \underline{h}^H \underline{s}_m^* + \underline{h}^H (\underline{s}_m^* \underline{s}_m^T) \underline{h}]$$

After equating the gradient of above quantity to zero we get

$$\sum_{m=1}^{M^L} \sum_{n=1}^{N} z_{nm} \underline{h}^H (\underline{s}_m^* \underline{s}_m^T) = \sum_{m=1}^{M^L} \sum_{n=1}^{N} z_{nm} \underline{s}_m^T r_n^*$$

which gives

$$\underline{h} = \left[\sum_{m=1}^{M^L} \sum_{n=1}^{N} z_{nm} (\underline{s}_m \underline{s}_m^H) \right]^{-T} \left[\sum_{m=1}^{M^L} \sum_{n=1}^{N} z_{nm} \underline{s}_m^* r_n \right] \tag{3.4}$$

or

$$\underline{h}^H = \left[\sum_{m=1}^{M^L} \sum_{n=1}^{N} z_{nm} (\underline{s}_m^* \underline{s}_m^T) \right]^{-1} \left[\sum_{m=1}^{M^L} \sum_{n=1}^{N} z_{nm} \underline{s}_m^T r_n^* \right] \tag{3.5}$$

Equations 3.3 and 3.4 are repeatedly evaluated until the channel vector $\underline{h}$ gets converged within a specified error. Alternatively we can also find the optimal $\underline{h}$ that minimizes the cost function $\mathcal{J}$ by finding its elements individually. That means solving for $\frac{\partial \mathcal{J}}{\partial h_i} = 0$ for $i = 0, 1, ...L - 1$.

$$\frac{\partial \mathcal{J}}{\partial h_i} = 0$$

$$\frac{\partial}{\partial h_i} \left\{ \sum_{m=1}^{M^L} \sum_{n=1}^{N} z_{nm} \left| r_n - \underline{h}^T \underline{s}_m \right|^2 \right\} = 0$$

$$\sum_{m=1}^{M^L} \sum_{n=1}^{N} z_{nm} \frac{\partial}{\partial h_i} \left\{ |r_n - \sum_{l=0}^{L-1} h_l s_{ml}|^2 \right\} = 0$$

$$\sum_{m=1}^{M^L} \sum_{n=1}^{N} z_{nm} \frac{\partial}{\partial h_i} \left\{ (r_n - \sum_{l=0}^{L-1} h_l s_{ml})(r_n^* - \sum_{l=0}^{L-1} h_l^* s_{ml}^*) \right\} = 0$$

$$\sum_{m=1}^{M^L} \sum_{n=1}^{N} z_{nm} \frac{\partial}{\partial h_i} \left\{ |r_n|^2 - r_n \sum_{l=0}^{L-1} h_l^* s_{ml}^* - r_n^* \sum_{l=0}^{L-1} h_l s_{ml} + \sum_{i=0}^{L-1} \sum_{l=0}^{L-1} h_i s_{mi} h_l^* s_{ml}^* \right\} = 0$$

which results in

$$\sum_{m=1}^{M^L}\sum_{n=1}^{N} z_{nm}\left\{r_n^* s_{mi} - \sum_{l=0;l\neq i}^{L-1} h_l^* s_{mi} s_{ml}^*\right\} = \sum_{m=1}^{M^L}\sum_{n=1}^{N} z_{nm} h_i^* |s_{mi}|^2$$

$$h_i^* = \frac{\sum_{m=1}^{M^L}\sum_{n=1}^{N} z_{nm} s_{mi}\left\{r_n^* - \sum_{l=0;l\neq i}^{L-1} h_l^* s_{ml}^*\right\}}{\sum_{m=1}^{M^L}\sum_{n=1}^{N} z_{nm}|s_{mi}|^2}$$

So the value of the i^{th} channel coefficient that minimizes $\mathcal{J}$ is given by

$$h_i = \frac{\sum_{m=1}^{M^L}\sum_{n=1}^{N} z_{nm} s_{mi}^*\left\{r_n - \sum_{l=0;l\neq i}^{L-1} h_l s_{ml}\right\}}{\sum_{m=1}^{M^L}\sum_{n=1}^{N} z_{nm}|s_{mi}|^2} \tag{3.6}$$

The $\underline{h}$ obtained using above procedure is used as a starting point for the EM channel estimation discussed below.

3.3.2 EM Parameter Estimation

As described earlier in section 3.2 we need to find Θ that maximizes $Pr(\underline{\mathbf{r}}/\Theta)$. Which means finding

$$\hat{\Theta} = \arg\max_{\mathcal{Z}} Pr(\underline{\mathbf{r}}/\Theta)$$

Now,

$$\begin{aligned} Pr(\underline{\mathbf{r}}, \mathcal{Z}/\Theta) &= Pr(\underline{\mathbf{r}}/\mathcal{Z}, \Theta) Pr(\mathcal{Z}/\Theta) \\ &= \frac{1}{M^{LN}} \prod_{m=1}^{M^L}\prod_{n=1}^{N}\left[\frac{1}{\pi\sigma_v^2}\exp\left(-\frac{\left|r_n - \underline{h}^T \underline{s}_m\right|^2}{\sigma_v^2}\right)\right]^{z_{nm}} \end{aligned}$$

Since maximizing the likelihood function and maximizing the log of the likelihood function are same, we maximize the Log Likelihood function which is given by

$$\begin{aligned} L(\Theta; \underline{\mathbf{r}}, \mathcal{Z}) &= \log \Pr(\underline{\mathbf{r}}, \mathcal{Z}/\Theta) \\ &= -LN\log M - \sum_{m=1}^{M^L}\sum_{n=1}^{N} z_{nm}\left[\log\left(\pi\sigma_v^2\right) + \frac{\left|r_n - \underline{h}^T \underline{s}_m\right|^2}{\sigma_v^2}\right] \end{aligned}$$

Since we can not find the parameters that maximize the Log Likelihood function analytically we use the Expectation Maximization Algorithm for finding the maximum likelihood estimates of the parameters (because the binary class matrix used z_{nm} used here is not constant and it is a random variable which made the problem a mixture model). Here the latent (unknown) variables are z_{nm}. In EM algorithm we find the maximum likelihood estimates in two iterative steps namely E-Step and the M-Step.

a) E-Step: In this step the Expectation of the Log Likelihood function with respect to the latent variables z_{nm} is calculated conditioned on the previous estimates $\Theta^{(old)} = (h^{(old)}, \sigma_v^{2(old)})$ and the received samples $\underline{\mathbf{r}}$.

$$\begin{aligned}\Psi\left(\Theta; \Theta^{(\text{old})}\right) &= E_{\mathcal{Z}}\left[L(\Theta; \underline{\mathbf{r}}, \mathcal{Z})/\underline{\mathbf{r}}, \Theta^{(\text{old})}\right] \\ &= -LN \log M - \sum_{m=1}^{M^L}\sum_{n=1}^{N}\gamma(n,m)\left[\log\left(\pi\sigma_v^2\right) + \frac{\left|r_n - \underline{h}^T\underline{s}_m\right|^2}{\sigma_v^2}\right]\end{aligned} \tag{3.7}$$

where,

$$\begin{aligned}\gamma(n,m) &= E\left[z_{nm}/\underline{\mathbf{r}}, \Theta^{(\text{old})}\right] \\ &= 1 \times Pr\left(z_{nm} = 1/\underline{\mathbf{r}}, \Theta^{(\text{old})}\right) + 0 \times Pr\left(z_{nm} = 0/\underline{\mathbf{r}}, \Theta^{(\text{old})}\right) \\ &= Pr\left(z_{nm} = 1/\underline{\mathbf{r}}, \Theta^{(\text{old})}\right) \\ &= \frac{\exp\left(-\frac{\left|r_n - \underline{h}^{(\text{old})T}\underline{s}_m\right|^2}{\sigma_v^{2(\text{old})}}\right)}{\sum_{j=1}^{M^L}\exp\left(-\frac{\left|r_n - \underline{h}^{(\text{old})T}\underline{s}_j\right|^2}{\sigma_v^{2(old)}}\right)}\end{aligned} \tag{3.8}$$

b) M-Step: In the M-step we need to find the parameters Θ that maximize the expectation $\Psi\left(\Theta; \Theta^{(\text{old})}\right)$. Following the same procedure as in k-means estimation of $\underline{h}$ we get the value of $\underline{h}$ that maximizes $\Psi\left(\Theta; \Theta^{(\text{old})}\right)$ as

$$\underline{h} = \left[\sum_{m=1}^{M^L}\sum_{n=1}^{N}\gamma(n,m)(\underline{s}_m\underline{s}_m^H)\right]^{-T}\left[\sum_{m=1}^{M^L}\sum_{n=1}^{N}\gamma(n,m)\underline{s}_m^* r_n\right] \tag{3.9}$$

Variable	Dimension	Description
r_n	Complex scalar	n^{th} received sample
$\underline{h}$	$L \times 1$	Channel vector
$\underline{a}$	$L \times 1$	Vector of past L transmit symbols
σ_v^2	Real scalar	Noise variance
$\mathcal{Z} = \{z_{nm}\}_{N\times M^L}$	$N \times M^L$	Binary class matrix
$\underline{s}_m$	$L \times 1$	m^{th} possible transmit vector
Ψ	scalar	Expected log-likelihood given the parameters Θ
$\gamma = \gamma(n,m)_{N\times M^L}$	$N \times M^L$	Expectation of $\mathcal{Z} = \{z_{nm}\}_{N\times M^L}$
h_i	Complex scalar	i^{th} channel coefficient
L	Integer	Number of channel taps
N	Integer	Number of received samples
M	Integer	Number of symbols in any modulation

Table 3.1: Variables used in EM parameter estimation and their dimensions

or we can find the individual elements of the channel vector as given by the following equation

$$h_i = \frac{\sum_{m=1}^{M^L}\sum_{n=1}^{N}\gamma(n,m)s_{mi}^*\left\{r_n - \sum_{l=0;l\neq i}^{L-1} h_l s_{ml}\right\}}{\sum_{m=1}^{M^L}\sum_{n=1}^{N}\gamma(n,m)|s_{mi}|^2} \tag{3.10}$$

To find the variance σ_v^2 that maximizes $\Psi\left(\Theta;\Theta^{(\text{old})}\right)$ we need to solve

$$\frac{\partial\Psi}{\partial\sigma_v^2} = \sum_{m=1}^{M^L}\sum_{n=1}^{N}\gamma(n,m)\left(-\frac{1}{\sigma_v^2} + \frac{\left|r_n - \underline{h}^T\underline{s}_m\right|^2}{\sigma_v^4}\right) = 0$$

which results in

$$\sum_{m=1}^{M^L}\sum_{n=1}^{N}\gamma(n,m) = \sum_{m=1}^{M}\sum_{n=1}^{N}\frac{\gamma(n,m)}{\sigma_v^2}\left|r_n - \underline{h}^T\underline{s}_m\right|^2$$

$$\sigma_v^2 = \frac{\sum_{m=1}^{M^L}\sum_{n=1}^{N}\gamma(n,m)\left|r_n - \underline{h}^T\underline{s}_m\right|^2}{\sum_{m=1}^{M}\sum_{n=1}^{N}\gamma(n,m)}$$

So the value of σ_v^2 that maximizes $\Psi\left(\Theta;\Theta^{(\text{old})}\right)$ is given by

$$\sigma_v^2 = \sum_{m=1}^{M^L}\sum_{n=1}^{N}\frac{\gamma(n,m)}{N}\left|r_n - \underline{h}^T\underline{s}_m\right|^2 \tag{3.11}$$

Repeated execution of equations 3.8, 3.9 and 3.11 till convergence is reached will give the channel vector and noise variance estimates. The variables used in this chapter and their dimensions are tabulated in table 3.1. The values of $\underline{h}$ and σ_v^2 obtained using this EM algorithm are used in doing Maximum Likelihood Classification which is described in the next section.

3.4 Maximum Likelihood Classification

In this section we will use the estimated channel parameters using EM and do the Modulation Classification. Let $\mathcal{M} = \{\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3..., \mathcal{S}_K\}$ denote the set of possible modulations with $\mathcal{S}_k = \{s_1^{(k)}, s_2^{(k)}, s_3^{(k)}, ...s_{M^{(k)}}^{(k)}\}$ being the k^{th} modulation. k^{th} modulation contains $M^{(k)}$ symbols in its constellation. Let $\hat{\Theta}^{(k)} = (\underline{\hat{h}}^{(k)}, \hat{\sigma}_v^{2(k)})$ be the estimated channel parameters assuming that $\mathcal{S}_k$ is the actual modulation. After calculating these values for all modulations considered, we will calculate the Log Likelihood for each modulation which is given by

$$\begin{aligned}\mathcal{L}^{(k)}(\underline{\mathbf{r}}) &= \frac{1}{N} \log Pr(\underline{\mathbf{r}}/\mathcal{S}_k, \hat{\Theta}^{(k)}) \\ &= \frac{1}{N} \sum_{n=1}^{N} \log \left\{ \frac{1}{{M^{(k)}}^L \pi \hat{\sigma}_v^{2(k)}} \sum_{m=1}^{{M^{(k)}}^L} e^{\left\{ - \frac{|r_n - \underline{\hat{h}}^{(k)} \underline{s}_m^{(k)}|^2}{\hat{\sigma}_v^{2(k)}} \right\}} \right\}\end{aligned}$$

After calculating these likelihoods for all modulations considered, we will take the modulation with maximum likelihood as the modulation present in the received signal. That is

$$\hat{k} = \arg\max_k \mathcal{L}^{(k)}(\underline{\mathbf{r}})$$

3.5 Simulation Results

The performance of the modulation classification method discussed in this chapter was examined by considering a pool of modulations containing BPSK, QPSK, and 8PSK. Experiments were done with various lengths of the received signal. All the experiments were done for 1000 runs over a range of Signal to Noise Ratios(SNR).

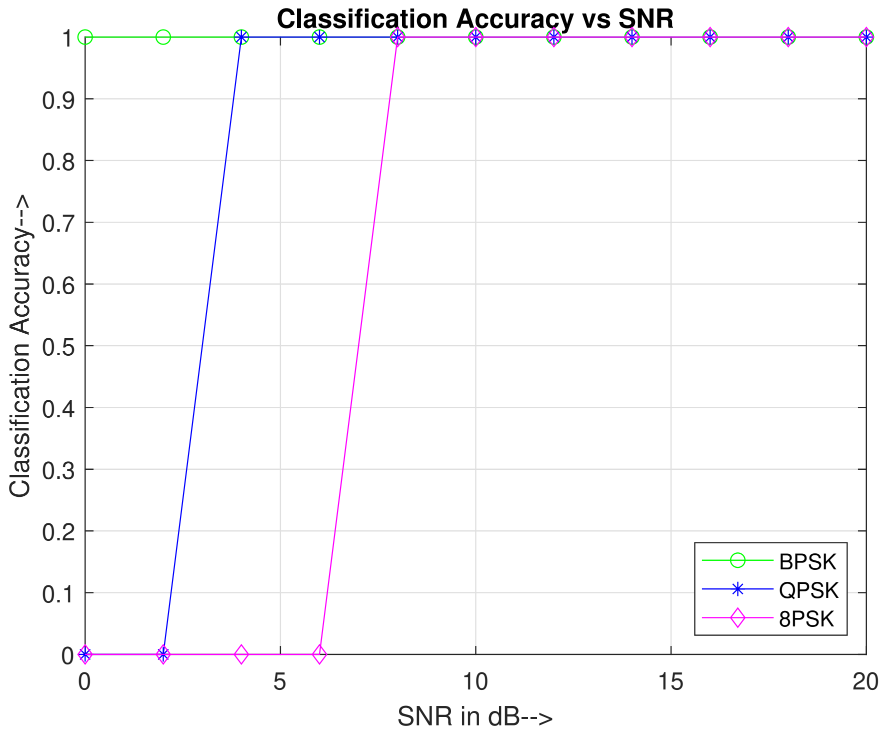


Fig. 3.2: Classification Accuracy for $N = 512$

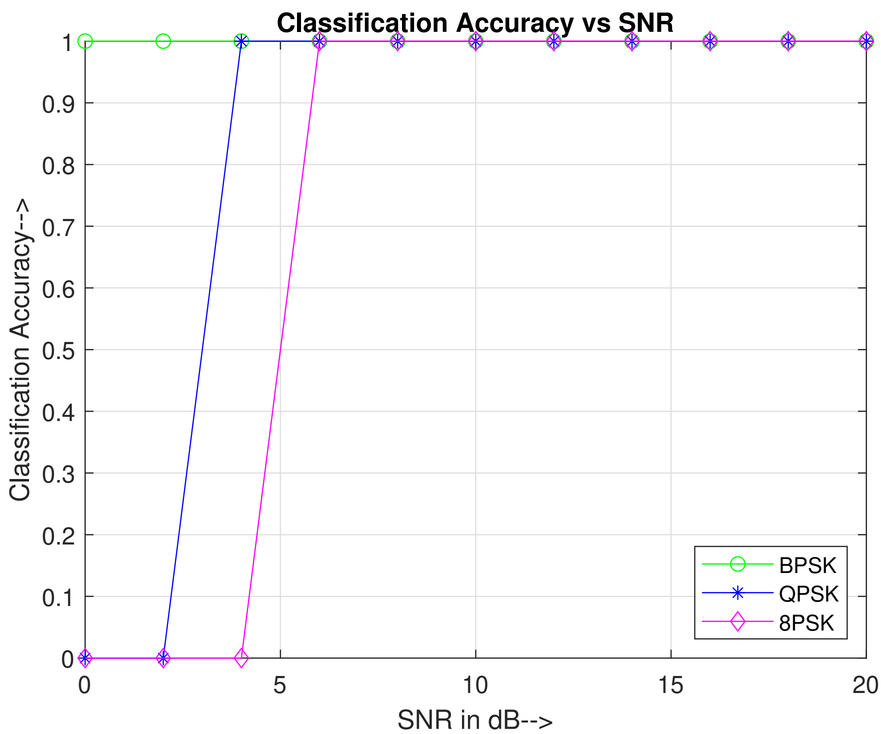


Fig. 3.3: Classification Accuracy for $N = 1024$

SNRs ranging from 0 dB to 20 dB in steps of 2 dB are considered. Figure 3.2 shows the classification accuracy for all the modulations considered when the signal length is

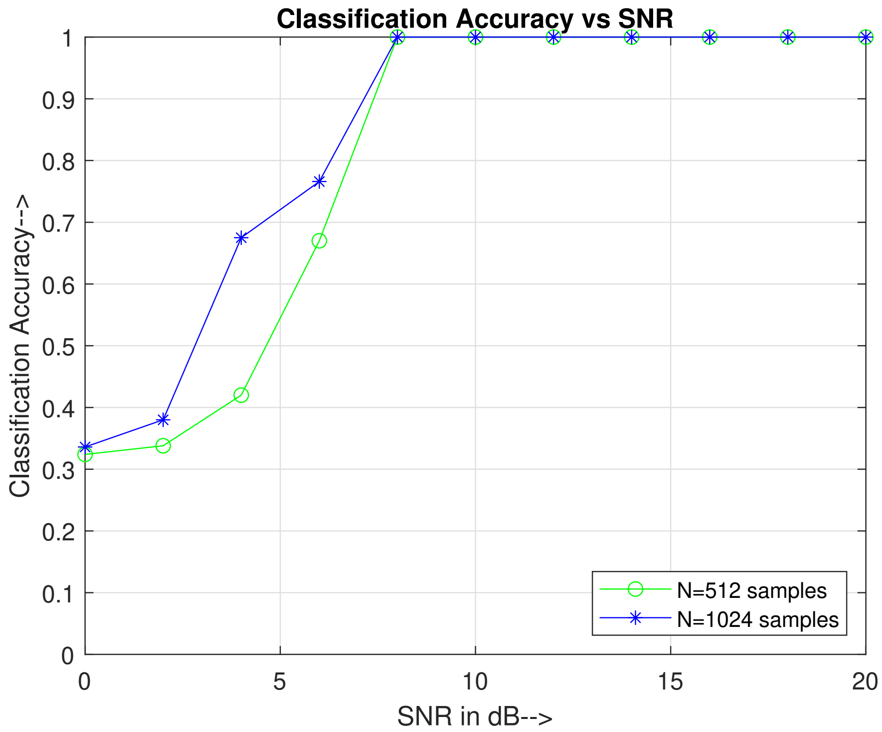


Fig. 3.4: Classification Accuracy for various signal lengths

$N = 512$. From the figure we can observe that the method is able to classify BPSK with 100% accuracy over all SNRs. It is able to classify QPSK also with almost 100% accuracy. Accuracy when 8PSK is considered is also good after an SNR of 8dB . At high SNR region the method classifies all signals with 100% accuracy. When the Signal length is increased from N=512 to N=1024 there is not much change in the performance for BPSK and QPSK but for 8PSK there is an improvement in the classification accuracy. This can be observed from figure 3.3. In figure 3.4 the accuracy curves for various signal lengths are plotted. We can observe that the accuracy is improved when signal length is increased but the improvement is not significant. This improvement can be explained by the fact that we have more information when we have more number of samples.

CHAPTER 4

CONCLUSIONS AND FUTURE WORK

The use of Modulation Classification eliminates the feedback between the transmitter and the receiver and hence the data rates can be increased. Using the method discussed BPSK modulation can be detected with 100% accuracy in both the flat fading and frequency selective fading channels. QPSK modulation also can be detected with 100% accuracy over a wide range of SNR. However, when we go for higher constellations the classification accuracy is reduced. In such cases, to improve the performance of the algorithm we can use some Pattern Matching techniques instead of Maximum Likelihood Classification as the last step.

APPENDIX A

MAXIMUM LIKELIHOOD ESTIMATION AND EXPECTATION MAXIMIZATION

A.1 Maximum Likelihood Estimation

Consider a random variable X whose probability density function $f(x;\theta)$ depends on an unknown value θ. Suppose that $X_1, ..., X_n$ are iid random variables with common *pdf* $f(x;\theta)$. The parameter θ is unknown,which we would like to estimate from the observations $X_1, ..., X_n$. The likelihood function of θ is given by the equation

$$L(\theta; \mathbf{x}) = \prod_{i=1}^{n} f(x_i; \theta) \tag{A.1}$$

where $\mathbf{x} = (x_1, ..., x_n)'$. L is treated as a function of $\mathbf{x}$ and θ and it is often written as $L(\theta)$ also. The log of this function $L(\theta)$ is usually more convenient to be used and it is denoted by

$$l(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_i; \theta) \tag{A.2}$$

Since the log is a one-to-one function, there is no loss of information in using $l(\theta)$. The point estimator of θ is $\hat{\theta} = \hat{\theta}(X_1, X_2, ...X_n)$ where $\hat{\theta}$ maximizes the function $L(\theta)$ or $l(\theta)$.

Maximum Likelihood Estimator: We say that $\hat{\theta} = \hat{\theta}(\mathbf{x})$ is a maximum likelihood estimator (mle) of θ if

$$\hat{\theta} = \arg\max L(\theta; \mathbf{x}) \tag{A.3}$$

The notation $argmax$ means that the maximum value of $L(\theta; \mathbf{x})$ is achieved at $\hat{\theta}$. To determine the mle, take the log of the likelihood and determine its critical value; that is, letting $l(\theta) = logL(\theta)$, the mle is obtained by solving the equation $\frac{\partial l(\theta)}{\partial \theta} = 0$.

A.2 The Expectation Maximization Algorithm

In practice, we can be in a situation where we do not have the complete data. In such situations the EM Algorithm can be used to get the maximum likelihood estimates. Suppose a sample of n items is considered, where n_1 of the items are observed and $n_2 = n - n_1$ items are not observable. Denote the observed data by $\mathbf{x} = (X_1, X_2, ..., X_{n_1})$ and unobserved data by $\mathbf{z} = (Z_1, Z_2, ..., Z_{n_2})$. Assume that the X_is are iid with pdf $f(\mathbf{x}|\theta)$. Assume that Z_js and the X_is are mutually independent. Let $g(\mathbf{x}|\theta)$ denote the joint pdf of $\mathbf{x}$. Let $h(\mathbf{x}, \mathbf{z}|\theta)$ denote the joint pdf of the observed and unobserved data. Let $k(\mathbf{z}|\theta, \mathbf{x})$ denote the conditional pdf of the missing data given the observed data. By the definition of a conditional pdf, we have the identity

$$k(\mathbf{z}|\theta, \mathbf{x}) = \frac{h(\mathbf{x}, \mathbf{z}|\theta)}{g(\mathbf{x}|\theta)} \tag{A.4}$$

The observed likelihood function is $L(\theta|\mathbf{x}) = g(\mathbf{x}|\theta)$. The complete likelihood function is defined by $L^c(\theta|\mathbf{x}, \mathbf{z}) = h(\mathbf{x}, \mathbf{z}|\theta)$. Our goal is maximize the likelihood function $L(\theta|\mathbf{x})$ by using the complete likelihood $L^c(\theta|\mathbf{x}, \mathbf{z})$ in this process. Using A.4, we derive the following basic identity for an arbitrary but fixed $\theta_0 \in \Omega$:

$$\begin{aligned}
\log L(\theta \mid \mathbf{x}) &= \int \log L(\theta \mid \mathbf{x}) k\left(\mathbf{z} \mid \theta_0, \mathbf{x}\right) d\mathbf{z} \\
&= \int \log g(\mathbf{x} \mid \theta) k\left(\mathbf{z} \mid \theta_0, \mathbf{x}\right) d\mathbf{z} \\
&= \int [\log h(\mathbf{x}, \mathbf{z} \mid \theta) - \log k(\mathbf{z} \mid \theta, \mathbf{x})] k\left(\mathbf{z} \mid \theta_0, \mathbf{x}\right) d\mathbf{z} \\
&= \int \log[h(\mathbf{x}, \mathbf{z} \mid \theta)] k\left(\mathbf{z} \mid \theta_0, \mathbf{x}\right) d\mathbf{z} - \int \log[k(\mathbf{z} \mid \theta, \mathbf{x})] k\left(\mathbf{z} \mid \theta_0, \mathbf{x}\right) d\mathbf{z} \\
&= E_{\theta_0}\left[\log L^c(\theta \mid \mathbf{x}, \mathbf{z}) \mid \theta_0, \mathbf{x}\right] - E_{\theta_0}\left[\log k(\mathbf{z} \mid \theta, \mathbf{x}) \mid \theta_0, \mathbf{x}\right],
\end{aligned} \tag{A.5}$$

where the expectations are taken under the conditional pdf $k(\mathbf{z}|\theta_0, \mathbf{x})$. Define the first term on the RHS to be the function

$$Q\left(\theta \mid \theta_0, \mathbf{x}\right) = E_{\theta_0}\left[\log L^c(\theta \mid \mathbf{x}, \mathbf{z}) \mid \theta_0, \mathbf{x}\right] \tag{A.6}$$

The expectation which defines the function Q is called the E-step of the EM algorithm. We want to maximize $logL(\theta|x)$. To do so, we need to maximize $Q(\theta \mid \theta_0, \mathbf{x})$ as discussed below. This maximization is called the M-step of the EM algorithm. Denote by $\hat{\theta}^{(0)}$ an initial estimate of θ, perhaps based on the observed likelihood. Let $\hat{\theta}^{(1)}$ be the argument which maximizes $Q(\theta|\hat{\theta}^{(0)}, x)$. This is the first-step estimate of θ. Proceeding this way, we obtain a sequence of estimates $\hat{\theta}^{(m)}$. This algorithm is formally defined as follows:

EM Algorithm: Let $\hat{\theta}^{(m)}$ denote the estimate on the m^{th} step. To compute the estimate on the $(m+1)^{st}$ step, do:

1 Expectation Step: Compute

$$Q\left(\theta \mid \widehat{\theta}^{(m)}, \mathbf{x}\right) = E_{\widehat{\theta}(m)}\left[\log L^c(\theta \mid \mathbf{x}, \mathbf{z}) \mid \widehat{\theta}_m, \mathbf{x}\right] \tag{A.7}$$

where the expectation is taken under the conditional *pdf* $k(z|\hat{\theta}^{(m)}, x)$.

2 Maximization Step: Let

$$\hat{\theta}^{(m+1)} = \arg\max Q(\theta|\hat{\theta}^{(m)}, x). \tag{A.8}$$

Repeated execution of the equations A.7 and A.8 will give the maximum likelihood estimate of the parameter θ. For more details on maximum likelihood estimation and the EM algorithm one can refer Hogg *et al.* (2012) and Moon (1996)

APPENDIX B

COMPLEX DERIVATIVES

B.1 Complex Derivatives

In this section some basics on the derivatives of complex valued functions are presented. z is a complex variable which is defined as $z = x + iy$. $g(z) = u(x, y) + iv(x, y)$ being any function of z its derivative exists only if

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \text{ and } \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y} \tag{B.1}$$

and is defined as

$$\frac{\partial g}{\partial z} = \frac{1}{2}\left\{\frac{\partial g}{\partial x} - i\frac{\partial g}{\partial y}\right\}$$

The above pair of equations B.1 are called CR equations. Some useful derivatives are presented in table B.1

$\mathbf{g(z)}$	$\mathbf{g^{'}(z) = \frac{\partial g}{\partial z}}$
z	1
z^2	$2z$
z^*	0
$\|z\|^2$	z^*
$\lambda + \alpha z + \beta z^* + \gamma \|z\|^2$	$\alpha + \gamma z^*$

Table B.1: Derivatives of some functions of scalar variable z

Now we will see the gradients of functions of complex vectors. $g(\underline{z})$ being a function of many complex variables, where $\underline{z} = [z_1 z_2 z_3 ... z_n]^T$ and $z_k = x_k + iy_k$, the gradient of $g(\underline{z})$ is defined as

$$\nabla_{\underline{z}}\, g(\underline{z}) = \left[\frac{\partial g}{\partial z_1}\frac{\partial g}{\partial z_2}\frac{\partial g}{\partial z_3}\cdots\frac{\partial g}{\partial z_n}\right]$$

$\mathbf{g}(\underline{\mathbf{z}})$	$\nabla_{\underline{\mathbf{z}}}\mathbf{g}(\underline{\mathbf{z}})$
$\underline{\alpha}^H \underline{z}$	$\underline{\alpha}^H$
$\underline{\alpha}^T \underline{z}$	$\underline{\alpha}^T$
$\underline{z}^H \underline{\beta}$	$\underline{0}^H$
$\underline{z}^T \underline{\beta}$	$\underline{\beta}^T$
$\underline{z}^H \underline{z}$	$\underline{z}^H$
$\lambda + \underline{\alpha}^H \underline{z} + \underline{z}^H \underline{\beta} + \underline{z}^H \Gamma \underline{z}$	$\underline{\alpha}^H + \underline{z}^H \underline{\gamma}$

Table B.2: Derivatives of some functions of vector variable $\underline{z}$

Some useful derivatives of linear functions of complex vector $\underline{z}$ are presented in table B.2

REFERENCES

1. **Chavali, V. G.** and **C. R. C. M. da Silva** (2011). Maximum-likelihood classification of digital amplitude-phase modulated signals in flat fading non-gaussian channels. *IEEE Transactions on Communications*, **59**(8), 2051–2056, doi:10.1109/TCOMM.2011.051711.100184.

2. **Dobre, O.**, **A. Abdi**, and **W. Su** (2007). Survey of automatic modulation classification techniques: Classical approaches and new trends. *Communications, IET*, **1**, 137 – 156, doi:10.1049/iet-com:20050176.

3. **Hazza, A.**, **M. Shoaib**, **S. A. Alshebeili**, and **A. Fahad** (2013). An overview of feature-based methods for digital modulation classification. *In 2013 1st International Conference on Communications, Signal Processing, and their Applications (ICCSPA)*. doi:10.1109/ICCSPA.2013.6487244.

4. **Hermawan, A. P.**, **R. R. Ginanjar**, **D.-S. Kim**, and **J.-M. Lee** (2020). Cnn-based automatic modulation classification for beyond 5g communications. *IEEE Communications Letters*, **24**(5), 1038–1041, doi:10.1109/LCOMM.2020.2970922.

5. **Hogg, R.**, **J. McKean**, and **A. Craig** (2012). *Introduction to Mathematical Statistics*. Pearson Education. ISBN 9780321849434. URL https://books.google.co.in/books?id=YdwsAAAAQBAJ.

6. **Huynh-The, T.**, **C.-H. Hua**, **Q.-V. Pham**, and **D.-S. Kim** (2020). Mcnet: An efficient cnn architecture for robust automatic modulation classification. *IEEE Communications Letters*, **24**(4), 811–815, doi:10.1109/LCOMM.2020.2968030.

7. **Moon, T.** (1996). The expectation-maximization algorithm. *IEEE Signal Processing Magazine*, **13**(6), 47–60, doi:10.1109/79.543975.

8. **Soltanmohammadi, E.** and **M. Naraghi-Pour** (2013). Blind modulation classification over fading channels using expectation-maximization. *IEEE Communications Letters*, **17**(9), 1692–1695, doi:10.1109/LCOMM.2013.070913.130752.

9. **Zhu, Z.** and **A. K. Nandi** (2015). Modulation classification in mimo fading channels via expectation maximization with non-data-aided initialization. *In 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. doi:10.1109/ICASSP.2015.7178524.