

Rochester Institute of Technology

**RIT Digital Institutional Repository**

---

Theses

---

2012

## **Signal to noise ratio estimation using the Expectation Maximization Algorithm**

Ahmed Almradi

Follow this and additional works at: <https://repository.rit.edu/theses>

---

### **Recommended Citation**

Almradi, Ahmed, "Signal to noise ratio estimation using the Expectation Maximization Algorithm" (2012). Thesis. Rochester Institute of Technology. Accessed from

This Thesis is brought to you for free and open access by the RIT Libraries. For more information, please contact [repository@rit.edu](mailto:repository@rit.edu).

***Signal to Noise Ratio Estimation using the Expectation Maximization Algorithm***

By

**AHMED M ALMRADI**

A Thesis is submitted in partial fulfillment of the requirements for the degree of

**MASTER OF SCIENCE in Electrical Engineering**

APPROVED BY:

.....

Dr. Sohail Dianat, Thesis advisor

.....

Dr. Ernest Fokoué, Thesis Committee member

.....

Dr. Chance M. Glenn, Thesis Committee member

.....

Dr. Vincent Amuso, Thesis Committee member

.....

Dr. Sohail Dianat, Department head

DEPARTMENT OF ELECTRICAL AND MICROELECTRONIC ENGINEERING

KATE GLEASON COLLEGE OF ENGINEERING

ROCHESTER INSTITUTE OF TECHNOLOGY

ROCHESTER, NEW YORK

MARCH 2012

# ***Signal to Noise Ratio Estimation using the Expectation Maximization algorithm***

By:

**AHMED ALMRADI**

**Thesis advisor:**

**Dr. SOHAIL DIANAT**

## **Abstract**

Signal to Noise Ratio (SNR) estimation when the transmitted symbols are unknown is a common problem in many communication systems, especially those which require an accurate SNR estimation. For instance, modern wireless communication systems usually require accurate estimate of SNR without knowledge of the transmitted symbols. In addition, SNR estimation is required in order to perform efficient signal detection, power control, and adaptive modulation. In this study, Non data Aided (NDA) SNR estimation for Binary Phase Shift Keying (BPSK) and Quadrature Phase Shift Keying (QPSK) using the Expectation Maximization (EM) algorithm is developed. The assumption here is that the received data samples are drawn from a mixture of Gaussians distribution and they are independent and identically distributed (i. i. d.). The quality of the proposed estimator is examined via the Cramer-Rao Lower Bound (CRLB) of NDA SNR estimator. It is also assumed that the channel gain is constant during each symbol interval, and the noise is Additive White Gaussian (AWGN). Maximum Likelihood estimator is being used if we have access to the complete data, in this case the problem would be much easier since we get the exact closed form solution, but when the observed data are incomplete or partially available, the EM algorithm will be used. This approach is an iterative method to get an approximated result which is either an approximated global maximum or local maximum. However, in the NDA SNR estimation, we only have a global maximum since our assumption is that the distribution is a mixture of Gaussians. This is being investigated for the cases of Single Input Single Output (SISO) and Single Input Multiple Output (SIMO). The main concern about the receive diversity is the cost, size, and power, that is why we resort to the transmit diversity such as Multiple Input single Output (MISO) with space time block codes (STBC). The base station usually serves hundreds to thousands of remote units which is the sole reason of using transmit diversity at the base station instead of at every remote unit covered by the base station. It is more economical in this case to add equipment to the base station instead of the remote units. Alamouti used a simple transmit diversity technique and assumed in his paper that the receiver has perfect knowledge of the channel transition matrix. However, this assumption may seem highly unrealistic. One of our contributions is to estimate the channel information, as well as the noise variance which would be used in estimating the SNR and deriving the CRLB for both DA and NDA case. The performance of our estimator would be empirically assessed using Monte-Carlo simulations, with CRLB as a performance metric.

## TABLE OF CONTENTS

|  |      |
|--|------|
| List of figures . . . . .  | iv   |
| List of symbols and abbreviations . . . . .                                      | v    |
| Acknowledgments. . . . .   | viii |
| Abstract . . . . .   | ix   |
| 1. Introduction . . . . .  | 1    |
| 1.1 Estimation. . . . .  | 2    |
| 1.2 Multivariate Gaussian distribution. . . . .                                  | 5    |
| 1.3 Minimum Variance Unbiased Estimator (MVUE) . . . . .                         | 7    |
| 1.4 Cramer Rao Lower Bound (CRLB) . . . . .                                      | 8    |
| 2. Mixtures of Gaussians (MoG) . . . . .   | 10   |
| 2.1 The Basic definition of MoG. . . . .   | 10   |
| 2.2 The Likelihood function. . . . .   | 12   |
| 3. The EM algorithm. . . . .   | 15   |
| 3.1 Introduction. . . . .  | 15   |
| 3.2 Jensen's Equality. . . . .   | 16   |
| 3.3 Kullback-Leibler (KL) Divergence. . . . .                                    | 17   |
| 3.4 The EM Derivation. . . . .   | 18   |
| 3.5 Mixture of Gaussians and the EM algorithm. . . . .                           | 22   |
| 4. SISO NDA SNR estimation using the EM algorithm in the Literature . . . . .    | 27   |
| 4.1 The BPSK case. . . . .   | 27   |
| 4.1.1 The system model. . . . .  | 27   |
| 4.1.2 The CRLB for BPSK . . . . .  | 28   |
| 4.1.3 SISO NDA SNR estimation using the EM algorithm for the BPSK case . . . . . | 31   |

|   |    |
|---|----|
| 4.2 The QPSK case. . . . .  | 38 |
| 4.2.1 The CRLB for QPSK. . . . .  | 41 |
| 4.2.2 SISO NDA SNR estimation using the EM algorithm for the QPSK case . . . . .            | 43 |
| 5. MISO with STBC NDA SNR estimation using the Expectation Maximization algorithm . . . . . | 51 |
| 5.1 Conventional Receive diversity (SIMO) model. . . . .                                    | 51 |
| 5.2 Transmit diversity (MISO) model. . . . .  | 52 |
| 5.3 MIMO model. . . . .   | 53 |
| 5.4 The transmit diversity for BPSK case. . . . .   | 54 |
| 5.5 The CRLB for the transmit diversity in the BPSK case. . . . .                           | 58 |
| 5.6 STBC NDA SNR estimation using the EM algorithm. . . . .                                 | 65 |
| 6. Conclusions . . . . .  | 70 |
| 7. Future work . . . . .  | 70 |
| 8. References. . . . .  | 71 |

## LISTOF FIGURES

|   |    |
|---|----|
| Figure 1.1 Normal distribution with $\mu = 0$ and $\sigma^2 = 1$ . . . . .                    | 3  |
| Figure 2.1 Mixtures of two Gaussians . . . . .  | 11 |
| Figure 2.2 Samples and their contour of the previous distribution . . . . .                   | 11 |
| Figure 3.1 mixtures of four Gaussians data and their fit when $C = 0.351$ . . . . .           | 26 |
| Figure 3.2 mixtures of four Gaussians data and their fit when $C = 0.051$ . . . . .           | 26 |
| Figure 4.1 The MSE & CRLB in $\text{dB}^2$ vs. SNR in dB for the BPSK when $N=128$ . . . . .  | 35 |
| Figure 4.2 The MSE & CRLB in $\text{dB}^2$ vs. SNR in dB for the BPSK when $N=512$ . . . . .  | 36 |
| Figure 4.3 The MSE & CRLB in $\text{dB}^2$ vs. SNR in dB for the BPSK when $N=1024$ . . . . . | 37 |
| Figure 4.4 The Error in $\text{dB}^2$ vs. SNR in dB for BPSK at different $N_s$ . . . . .     | 37 |
| Figure 4.5 The MSE in $\text{dB}^2$ vs. $N$ . . . . .   | 38 |
| Figure 4.6 The MSE & CRLB in $\text{dB}^2$ vs. SNR in dB for the QPSK when $N=128$ . . . . .  | 47 |
| Figure 4.7 The MSE & CRLB in $\text{dB}^2$ vs. SNR in dB for the QPSK when $N=512$ . . . . .  | 48 |
| Figure 4.8 The MSE & CRLB in $\text{dB}^2$ vs. SNR in dB for the QPSK when $N=1024$ . . . . . | 49 |
| Figure 4.9 The Error in $\text{dB}^2$ vs. SNR in dB for QPSK at different $N_s$ . . . . .     | 49 |
| Figure 5.1 SIMO/ Receive diversity system model . . . . .                                     | 51 |
| Figure 5.1 MISO/ Transmit diversity system model . . . . .                                    | 52 |
| Figure 5.3 MIMO system model . . . . .  | 53 |
| Figure 5.4. The MSE & CRLB in $\text{dB}^2$ vs. SNR in dB for STBC when $N=128$ . . . . .     | 67 |
| Figure 5.5. The MSE & CRLB in $\text{dB}^2$ vs. SNR in dB for STBC when $N=512$ . . . . .     | 67 |
| Figure 5.6. The MSE & CRLB in $\text{dB}^2$ vs. SNR in dB for STBC when $N=1024$ . . . . .    | 68 |
| Figure 5.7. The MISO with STBC BPSK case . . . . .  | 69 |

## LIST OF SYMBOLS AND ABBREVIATIONS

### Symbols

|                 |   |
|-----------------|---|
| $SNR$           | Signal to Noise Ratio in dB             |
| $\widehat{SNR}$ | Estimated Signal to Noise Ratio in dB   |
| $NDA$           | Non Data Aided                          |
| $DA$            | Data Aided                              |
| $BPSK$          | Binary Phase Shift Keying               |
| $QPSK$          | Quadrature Phase Shift Keying           |
| $EM$            | Expectation Maximization                |
| $i. i. d.$      | Independent and Identically Distributed |
| $CRLB$          | Cramer Rao Lower Bound                  |
| $AWGN$          | Additive White Gaussian Noise           |
| $ML$            | Maximum Likelihood                      |
| $SISO$          | Single Input Single Output              |
| $SIMO$          | Single Input Multiple Output            |
| $MISO$          | Multiple Input Single Output            |
| $MIMO$          | Multiple Input Multiple Output          |
| $STBC$          | Space- Time Block Codes                 |
| $MVUE$          | Minimum Variance Unbiased Estimator     |
| $MoG$           | Mixtures of Gaussians                   |
| $GMM$           | Gaussian Mixture Model                  |
| $MLE$           | Maximum Likelihood Estimate             |

|                             |   |
|-----------------------------|---|
| $MC$                        | Monte Carlo   |
| $PDF$                       | Probability Density Function  |
| $\theta$                    | Parameters which need to be estimated                                 |
| $\hat{\theta}$              | The Estimated parameter Values  |
| $\mathcal{N}(\mu, \Sigma)$  | Real Gaussian distribution with mean $\mu$ and Covariance $\Sigma$    |
| $\mathcal{C}(\mu, \Sigma)$  | Complex Gaussian distribution with mean $\mu$ and Covariance $\Sigma$ |
| $L(\theta)$                 | Likelihood function   |
| $\mathcal{L}(\theta)$       | Log Likelihood function   |
| $\nabla_{\theta}$           | Gradient with respect to the parameters                               |
| $\Re$                       | Real of   |
| $\Im$                       | Imaginary of  |
| $\mu$                       | Mean  |
| $\sigma^2$                  | Variance  |
| $\hat{\mu}$                 | Estimated mean  |
| $\hat{\sigma}^2$            | Estimated variance  |
| $\mathbb{E}$                | The Expected value  |
| $I(\theta)$                 | Fisher Information Matrix   |
| $[I(\theta)]_{ij}$          | Element $ij$ of the Fisher Information Matrix                         |
| $MSE$                       | Mean Square Error   |
| $b(\theta)$                 | Bias  |
| $\pi$                       | Mixing proportion   |
| $Q(\theta^k, \theta^{k+1})$ | Auxiliary function  |



|           |                       |
|-----------|-----------------------|
| $KL$      | Kullback-Leibler      |
| $\rho$    | Signal to Noise ratio |
| $\otimes$ | Kroncker Product      |

## **Acknowledgments**

I introduce my deep gratitude to my advisor, Dr. Sohail Dianat, for his constant guidance and support till we made this work visible and valuable. His vision and ideas make mine grow gradually during my work in finishing this project. I am thankful for the opportunity in being under his guidance and support.

I also would like to send my thanks to Dr. Ernest Fokoue, from the center for quality and applied statistics, for his appreciated and valuable discussion which helped me shape my project in a better way. His knowledge filled in the gap which I had in some statistical concepts regarding density estimation when incomplete data is present.

Finally, special thanks go to my family; my mother (Mabrooka) and my father (Mohammed) who were always supportive and helpful. I also thank my wife (Afaf) and my daughter (Shahad) for their patience while I am busy working on this project. Their support and smile has helped me in completing this work properly.

## Abstract

Signal to Noise Ratio (SNR) estimation when the transmitted symbols are unknown is a common problem in many communication systems, especially those which require an accurate SNR estimation. In this study, Non data Aided (NDA) SNR estimation for Binary Phase Shift Keying (BPSK) and Quadrature Phase Shift Keying (QPSK) using the Expectation Maximization (EM) algorithm is developed. The assumption here is that the received data samples are drawn from a mixture of Gaussians distribution and they are independent and identically distributed (*i. i. d.*). The quality of the proposed estimator is examined via the Cramer-Rao Lower Bound (CRLB) of NDA SNR estimator. It is also assumed that the channel gain is constant during each symbol interval, and the noise is Additive White Gaussian (AWGN). Maximum Likelihood estimator is being used if we have access to the complete data, in this case the problem would be much easier since we get the exact closed form solution, but when the observed data are incomplete or partially available, the EM algorithm will be used. This approach is an iterative method to get an approximated result which is either an approximated global maximum or local maximum. However, in the NDA SNR estimation, we only have a global maximum since our assumption is that the distribution is a mixture of Gaussians. This is being investigated for the cases of Single Input Single Output (SISO); [3], [4], [10], and Single Input Multiple Output (SIMO); [5], [9]. The main concern about the receive diversity is the cost, size, and power, that is why we resort to the transmit diversity such as Multiple Input single Output(MISO) with space time block codes (STBC). The base station usually serves hundreds to thousands of remote units which is the sole reason of using transmit diversity at the base station instead of at every remote unit covered by the base station. It is more economical in this case to add equipments to the base station instead of the remote units. Alamouti [12] used a simple transmit diversity technique and assumed in his paper that the receiver has perfect knowledge of the channel transition matrix. However, this assumption may seem highly unrealistic. One of our contributions is to estimate the channel information, as well as the noise variance which would be used in estimating the SNR and deriving the CRLB for both DA and NDA case. The performance of our estimator would be empirically assessed using Monte-Carlo simulations, with CRLB as a performance metric.

## 1 Introduction

Many modern wireless communication systems require the Signal to Noise ratio (SNR) estimation without the knowledge of the transmitted data symbols in Non-Data Aided (NDA) manner. SNR estimators may be divided to two categories, Data Aided (DA) and Non Data Aided (NDA). In DA estimators, the transmitted symbols are used or known at the receiver and used in the estimation process, in NDA estimators; the estimation process is being done without the knowledge of the transmitted symbols, just based on the received samples. The Cramer-Rao Lower Bound (CRLB) for NDA SNR estimation, which gives the minimum variance of unbiased estimators, will be used in evaluating our estimator [5]. The CRLB for NDA SNR estimation for the BPSK and QPSK will be derived and plotted against the EM-ML NDA SNR estimator. Here, the transmitted symbols are treated as unknown parameters, but nuisance parameter; unknown and unwanted. This estimation procedure would be close to the CRLB for the SNR of interest; at sufficiently high SNR. The main purpose of this study is to derive a closed form approximation for the NDA CRLB in the BPSK and QPSK case. Besides, the derivations of the EM algorithm for NDA SNR estimation which will iteratively maximize the likelihood function till reach approximately the global maximum.

The structure of the rest of this research is as follows, in this first chapter, we introduce the problem of estimation especially for a Gaussian distributed data, followed by the Minimum Variance Unbiased Estimator (MVUE) which will then lead us to the Cramer-Rao Lower Bound (CRLB). In chapter two, Mixtures of Gaussians (MoG) will be introduced followed by the Maximum Likelihood estimate (MLE) of the MoG model. In chapter three, we introduce the Expectation Maximization (EM) algorithm, and then we solve one of its applications which is mixtures of Gaussians parameter estimation. In chapter four, we introduce SISO NDA SNR estimation using the EM algorithm in the literature, besides solving this model for the BPSK and QPSK case. Also CRLB for both BPSK and/or QPSK will be derived and plotted. In chapter five, MISO with space and time block codes (STBC) NDA SNR estimation using the EM algorithm will be addressed. NDA SNR estimator will be derived and plotted in both cases and for different sample size  $N$  using Monte Carlo simulations. Improvements in SNR estimation accuracy will be proven when exploiting diversity method due to the use of mutual information. The last chapter will talk about the conclusion and summary of the research.

## 1.1 Estimation

The problem of parameter estimation is a common problem that is widely faced in many areas such as Radar, speech, image analysis, biomedicine, communications, control, and much more. The problem we are faced with is estimating some parameters based on some seen or observed samples due to the use of digital communication systems or digital computers [2]. Mathematically, we observe  $N$  data points at the receiver side  $\{y(1), y(2), \dots, y(N)\}$  which sampled from an unknown probability density function  $p_Y(y; \theta)$  parameterized by an unknown parameter  $\theta$ , which is the parameter of interest. Unfortunately, we don't really get to see the distribution of the received sample  $Y$ . We only get that sample,  $Y$ , which we then use to estimate our parameter  $\theta$ . As a result of that, our estimator won't give us the exact  $\theta$ , but an estimated version of it. Our estimator will depend highly through some function of the received sample  $Y$  as,

$$\hat{\theta} = f(y(1), y(2), \dots, y(N)) \quad (1.1)$$

This estimator might be used to estimate the carrier frequency, channel phase, signal power, noise power,....etc. Since the received data are random, we can describe its behavior according to its PDF function,  $p(y(1), y(2), \dots, y(N); \theta)$ . The semicolon is used to indicate that the distribution of PDF is parameterized by the parameter  $\theta$ . And it is constant. This kind of estimation is called classical parameter estimation because the parameters of interest are assumed to be deterministic and needs to be estimated. Since we are assuming the noise which corrupted our received signal is additive white Gaussian noise, AWGN, the distribution of the PDF of the received samples will be Gaussian distribution. For example, if we received one sample and the parameter which needs to be estimated is the mean and variance. Figure 1.1 shows a normal distribution of zero mean and unite variance. This is just an illustrative example which shows us the parameters which we will be estimating. Assume that we have a signal which happen to be constant corrupted by AWGN,  $w(n)$ , our model will be as follows,

$$y(n) = \mu_y + w(n)$$

The parameters which need to be estimated here are two, the DC component or average, and the variance of our Gaussian noise.

If one sample has been observed,

$$y(1) = \mu_y + w(1)$$

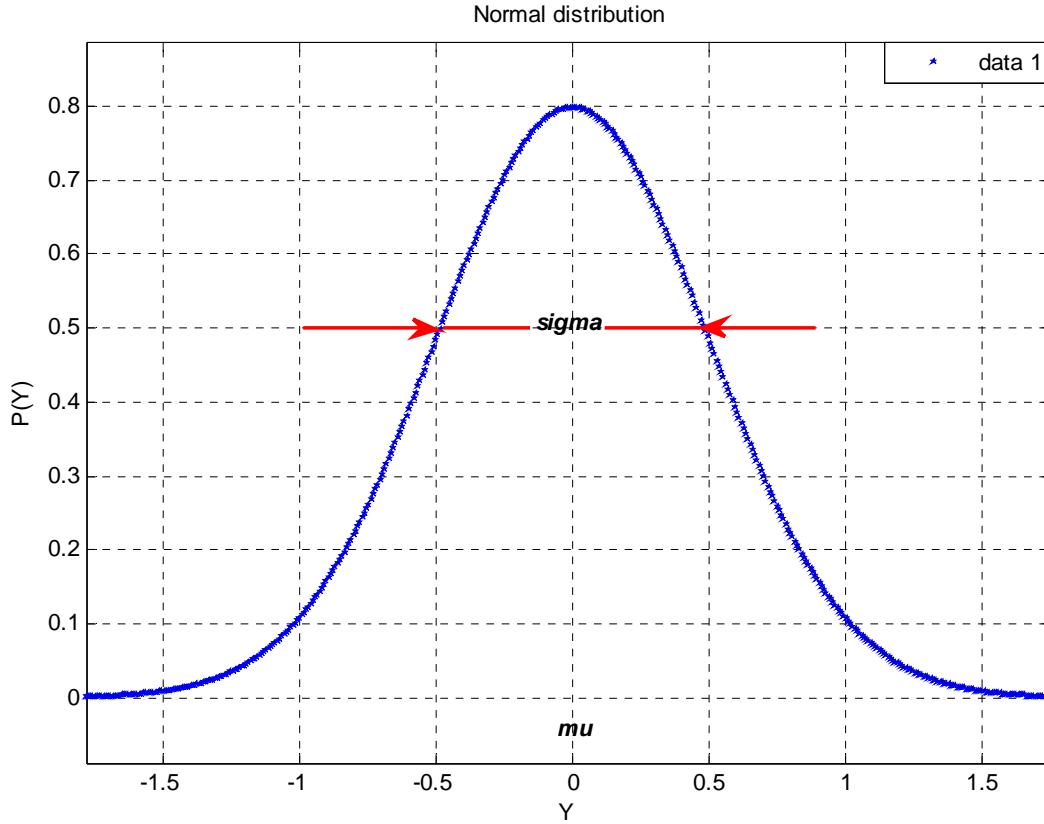


Figure 1.1 Normal distribution with  $\mu=0$ ,  $\sigma=1$

$$\theta = [\mu_y \ \sigma^2]^T$$

then,

$$p(y(1); \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y(1) - \mu_y)^2}{2\sigma^2}\right) \quad (1.2)$$

The assumption is that the noise  $w(1) \sim \mathcal{N}(0, \sigma^2)$ . Or in general, each sample of  $w(n)$  has the same PDF as  $w(1)$ . Since these samples are uncorrelated with each other, uncorrelated means independent if the distribution is Gaussian, we can now generalize the PDF of the received samples  $\mathbf{Y}$  of size  $N$  as follows,

$$p(y(1), y(2), \dots, y(N); \theta) = \prod_{n=1}^N p(y(n); \theta) = p(\mathbf{y}; \theta) = \mathbf{L}(\theta/\mathbf{y})$$

$$\begin{aligned}
&= \prod_{n=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y(n) - \mu_y)^2}{2\sigma^2}\right\} \\
&= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{n=1}^N (y(n) - \mu_y)^2\right\} \tag{1.3.1}
\end{aligned}$$

Equation (1.3) is called the likelihood function of the received data, which would be used to find the Maximum Likelihood estimate (MLE) of the mean and variance in case of our received data were generated by single Gaussian. I just wanted to make this clear because in the next chapter we will be talking about received sample which has been generated by more than one Gaussian; Mixture of Gaussians. To find the MLE of equation (1.3), we need to find the values of  $\mu_y$  and  $\sigma^2$  which maximizes the likelihood function  $\mathcal{L}(\theta)$ . We can instead maximize the log likelihood function since the logarithm is a monotonic function. MLE has the asymptotic property of being unbiased. It is also proved that MLE achieves the CRLB for high SNR.

$$\mathcal{L}(\theta) = -\frac{N}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=1}^N (y(n) - \mu_y)^2 \tag{1.3.2}$$

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \mathcal{L}(\theta)$$

which can be found by setting the gradient of  $\mathcal{L}(\theta)$  with respect to  $\theta$  equal to zero;

$$\nabla_{\theta} \mathcal{L}(\theta) = 0.$$

The ML solution is given as,

$$\hat{\mu}_y = \frac{1}{N} \sum_{n=1}^N y(n) \tag{1.4}$$

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{n=1}^N (y(n) - \hat{\mu}_y)^2 \tag{1.5}$$

To check the unbiasedness, we find the expected value of the estimated parameters and compare them with the true parameters. It can be easily shown as,

$$\mathbb{E}[\hat{\mu}_y] = \frac{1}{N} \sum_{n=1}^N \mathbb{E}\{y(n)\}$$

$$\mathbb{E}[\hat{\mu}_y] = \mu_y \tag{1.6}$$

$$\mathbb{E}[\widehat{\sigma^2}] = \frac{1}{N} \sum_{n=1}^N \mathbb{E}\{(y(n) - \hat{\mu}_y)^2\}$$

$$\mathbb{E}[\widehat{\sigma^2}] = \frac{1}{N} \sum_{n=1}^N \mathbb{E}\left\{y^2(n) - \frac{2}{N} y(n) \sum_{m=1}^N y(m) + \frac{1}{N^2} \sum_{m=1}^N \sum_{k=1}^N y(m)y(k)\right\}$$

$$\mathbb{E}[\widehat{\sigma^2}] = \frac{N-1}{N} \sigma^2 \tag{1.7}$$

So, we see that the MLE of the variance is biased, while the MLE of the mean is unbiased. The MLE distribution may be approximated as,

$$\hat{\Theta} \stackrel{a}{\sim} \mathcal{N}(\Theta, I^{-1}(\Theta)) \tag{1.8}$$

Where,  $\stackrel{a}{\sim}$  means asymptotically distributed according to,  $I(\Theta)$  is the fisher information which is given by,

$$I(\Theta) = -\mathbb{E}\left[\frac{\partial^2 \log p(Y; \Theta)}{\partial \Theta^2}\right] \tag{1.9}$$

A computer simulation could be performed to define how large the data length  $N$  had to be in order to get this asymptotic result. Monte Carlo method could be used for various realizations (say  $M$ ) of our estimate  $\hat{\Theta}$ , The mean and variance of the estimator could then be calculated as,

$$\hat{\mathbb{E}}[\hat{\Theta}] = \frac{1}{M} \sum_{i=1}^M \hat{\Theta}_i \tag{1.10}$$

$$\widehat{var}(\hat{\Theta}) = \frac{1}{M} \sum_{i=1}^M (\hat{\Theta}_i - \hat{\mathbb{E}}[\hat{\Theta}])^2 \tag{1.11}$$

The choice of a better estimator could be done as; among all the estimators we have, we choose the one which gives a closer estimate (on average) to the global true value, and has higher precision (lower variance) [2].



## 1.2 Multivariate Gaussian distribution

Since the focus on this research will be on Gaussian distribution, it is worth introducing the general multivariate Gaussian distribution, its behavior and properties.

Suppose we have a random variable that follows a multivariate Gaussian distribution. This random variable may be signal corrupted by noise.

$$Y = [y_1(n) \ y_2(n) \ y_3(n) \ \dots \ y_D(n)]^T$$

$$\text{Its mean is; } \mu_Y = \mathbb{E}[Y] = [\mathbb{E}[y_1(n)] \ \mathbb{E}[y_2(n)] \ \mathbb{E}[y_3(n)] \ \dots \ \mathbb{E}[y_D(n)]]^T$$

$$\text{Its covariance matrix is; } \Sigma_Y = \mathbb{E}[(Y - \mu_Y)(Y - \mu_Y)^T]$$

Then, the distribution of this random variable will be given by,

$$p_Y(Y) = \frac{1}{|2\pi\Sigma_Y|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(Y - \mu_Y)^T \Sigma_Y^{-1} (Y - \mu_Y)\right) \quad (1.12)$$

If the Covariance matrix  $\Sigma_Y = \sigma^2 \mathbf{I}$ , independent random variables with equal variances, each of the received samples dimension in equation (1.12) will return to equation (1.2). Let us mention an important property which is the effect of linear transformation on Gaussian distribution, let,

$$Y = AX + b$$

where, A and b are constants matrix and vector respectively, and X is a vector of random variable.

$$\text{then, } \mu_Y = \mathbb{E}[Y] = A \mu_X + b$$

$$\text{and, } \Sigma_Y = A \Sigma_X A^T$$

The MLE in the multivariate case would be as follows, the model parameters which need to be estimated is,

$$\Theta = [\mu_Y \ \text{vec}(\Sigma_Y)]^T, \text{ assuming i.i.d samples,}$$

$$\begin{aligned} p_Y(Y; \Theta) &= \prod_{n=1}^N \frac{1}{|2\pi\Sigma_Y|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(Y(n) - \mu_Y)^T \Sigma_Y^{-1} (Y(n) - \mu_Y)\right) \\ &= |2\pi\Sigma_Y|^{-\frac{N}{2}} \exp\left(-\frac{1}{2} \sum_{n=1}^N (Y(n) - \mu_Y)^T \Sigma_Y^{-1} (Y(n) - \mu_Y)\right) \end{aligned}$$

$$\mathcal{L}(\theta) = \log p_Y(\mathbf{Y}; \theta) = -\frac{N}{2} \log |2\pi \Sigma_Y| - \frac{1}{2} \sum_{n=1}^N (Y(n) - \mu_Y)^T \Sigma_Y^{-1} (Y(n) - \mu_Y)$$

To find the MLE of this parameter, we need to go through the same process of differentiating the likelihood of equation (1.12) and equating it to zero we get the following,

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta} = \mathbf{0}, \text{ first we differentiate with respect to } \mu_Y \text{ as } \frac{\partial \mathcal{L}(\theta)}{\partial \mu_Y} = \mathbf{0}, \text{ we get,}$$

$$\hat{\mu}_Y = \frac{1}{N} \sum_{n=1}^N Y(n) \quad (1.13)$$

and differentiating with respect to the covariance matrix  $\frac{\partial \mathcal{L}(\theta)}{\partial \Sigma_Y} = \mathbf{0}$ , we get,

$$\hat{\Sigma}_Y = \frac{1}{N} \sum_{n=1}^N (Y(n) - \hat{\mu}_Y)(Y(n) - \hat{\mu}_Y)^T \quad (1.14)$$

The parameter estimate in this case is straight forward computed in a closed form since the assumption that  $Y$ 's samples are being drawn from only one Gaussian distribution. However, this won't be the case when we have a mixture of Gaussians. The proposed algorithm will be a mixture of multivariate Gaussian distribution since we are dealing with PBSK and QPSK constellation systems and the transmitted data symbols are unknown.

### 1.3 Minimum Variance Unbiased Estimator (MVUE)

The search for good estimators of unknown deterministic parameters isn't always an easy task. Our focus will be on estimators which on average yield the true parameter value, and then among all those estimators, we look for the one which has the least variance. For an estimator to be unbiased,

$$\mathbb{E}[\hat{\theta}] = \theta \quad (1.15)$$

The bias of an estimator may be defined as,

$$b(\theta) = \mathbb{E}[\hat{\theta}] - \theta$$

To find an optimal estimator, we need to find a criterion like the mean square error (MSE),

$$MSE(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2] \quad (1.16)$$

$$MSE(\hat{\theta}) = var(\hat{\theta}) + b^2(\theta) \quad (1.17)$$

Good estimators have small MSE or small variance if the bias is zero. Equations (1.16) and (1.17) are unrealizable estimators since they are function of the true parameter value, so we cannot assess the performance of the estimator in this case.

MSE is composed to error due to the bias and variance. MVUE does not always exist, and if it did, sometimes it is not easy to be found [2]. One way of finding it is using the Cramer Rao Lower Bound (CRLB). When CRLB satisfies with equality for all  $\theta$ , the  $CRLB = MVUE$ .

#### 1.4 Cramer Rao Lower Bound (CRLB)

Finding a lower bound on the variance of an unbiased estimator will be extremely important in practice. This will serve as a benchmark in comparing the goodness of an estimator. CRLB allows us to assure that an estimator is the minimum variance unbiased estimator (MVUE), which will be the case when the estimator attains the CRLB for all values of the unknown parameter.

If  $p(y; \theta)$  satisfies regularity conditions, the variance of unbiased estimator is defined as,

$$var(\hat{\theta}) \geq \frac{1}{-\mathbb{E}\left[\frac{\partial^2 \log p(y; \theta)}{\partial \theta^2}\right]} \quad (1.18)$$

An estimator may be found that attains the bound for all  $\theta$  if and only if the score function can be written as,

$$\frac{\partial \log p(y; \theta)}{\partial \theta} = I(\theta)(g(y) - \theta)$$

For some functions  $g$  and  $I$ , that estimator which is the MVUE is given by,

$$\hat{\theta} = g(y)$$

And the minimum variance is

$$var(\hat{\theta}) = \frac{1}{I(\theta)}$$

Then this estimator is said to be efficient.

Let us extend the CRLB to a vector parameter,

$$\theta = [\theta_1 \theta_2 \theta_3 \dots \theta_p]^T$$

$$Var(\hat{\theta}_i) \geq [I^{-1}(\theta)]_{ii} \quad (1.19)$$

where  $I(\theta)$  is the  $p \times p$  Fisher information matrix.

$$[I(\theta)]_{ij} = -\mathbb{E}\left[\frac{\partial^2 \log p(Y; \theta)}{\partial \theta_i \partial \theta_j}\right] \quad (1.20)$$

For example if

$$\theta = [\mu_y \ \sigma^2]^T$$

Then,

$$I(\theta) = \begin{bmatrix} -\mathbb{E}\left[\frac{\partial^2 \log p(Y; \theta)}{\partial \mu_y^2}\right] & -\mathbb{E}\left[\frac{\partial^2 \log p(Y; \theta)}{\partial \mu_y \partial \sigma^2}\right] \\ -\mathbb{E}\left[\frac{\partial^2 \log p(Y; \theta)}{\partial \sigma^2 \partial \mu_y}\right] & -\mathbb{E}\left[\frac{\partial^2 \log p(Y; \theta)}{\partial \sigma^2^2}\right] \end{bmatrix}$$

This matrix is symmetric and positive definite.

If we are estimating a function of the parameter as in the research study, if

$$\alpha = f(\theta)$$

Then,

$$Cov(\hat{\alpha}) \geq \frac{\partial f(\theta)}{\partial \theta} I^{-1}(\theta) \frac{\partial f(\theta)^T}{\partial \theta}$$

## 2 Mixtures of Gaussians (MoG)

When it is not possible for one Gaussian distribution to represent our model, we resort the mixture of Gaussians method. It is a combined (mixture) of some known distributions that are combined to form our model. In our estimation problem, our assumption is that we are given a received samples  $Y$  of length  $N$ , which are drawn from a mixture of Gaussians distribution. The number of Gaussians is determined by the constellation order we use. For example, in PPSK, we have two mixtures of Gaussians per antenna with means  $\{-1, +1\}$ , and in QPSK, we have four mixtures of Gaussians per antenna with means  $\{\pm \frac{1}{\sqrt{2}}, \pm j \frac{1}{\sqrt{2}}\}$ . Gaussian Mixture Model (GMM) can be used in modeling complicated probability density functions (PDFs) which have complicated shapes that cannot be modeled using a single Gaussian.

### 2.1 The basic definition of MoG

The general form of a Gaussian mixture models is

$$\begin{aligned} p(y/\theta) &= \sum_{m=1}^M p(y, \omega_m/\theta_m) \\ &= \sum_{m=1}^M \pi_m p(y/\omega_m, \theta_m) \\ &= \sum_{m=1}^M \pi_m \mathcal{N}(y; \mu_m, \Sigma_m) \end{aligned} \tag{2.1}$$

where  $\pi_m$  is the component prior of each Gaussian component. There is a necessary condition for  $\pi_m$ s to be a valid probability density function that,  $\sum_{m=1}^M \pi_m = 1$  and  $\pi_m \geq 0$ . The parameters which need to be estimated are  $\pi_1, \pi_2, \dots, \pi_M, \theta_1, \theta_2, \dots, \theta_M$ . Increasing the number of components would result in more parameters which need to be estimated, which leads to a more accurate class conditional probability distribution function. We need to be careful about finding the correct number of mixture components to avoid singularities, when using the ML estimation to estimate the parameters [1]. Figure 2.1 below shows the probability distribution function of a mixture of two Gaussians, and figure 2.2 shows the contours of those two Gaussians with the samples drawn from the same distribution. In fact this exact distribution will not be given in our case of study. What is given is just a received sample of size  $N$  which is drawn in figure 2.2.

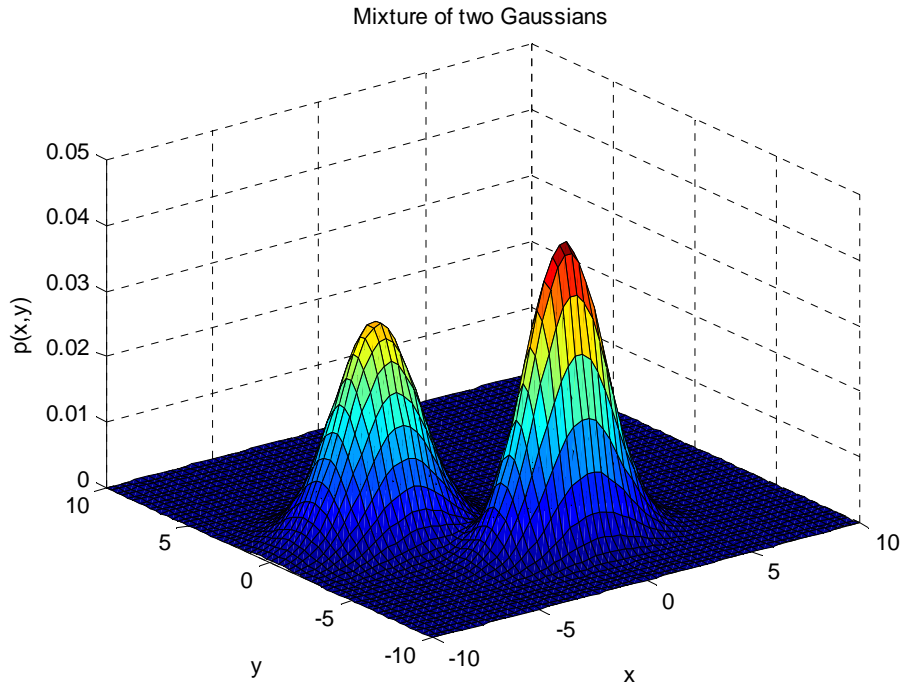


Figure 2.1. Mixtures of two Gaussians.

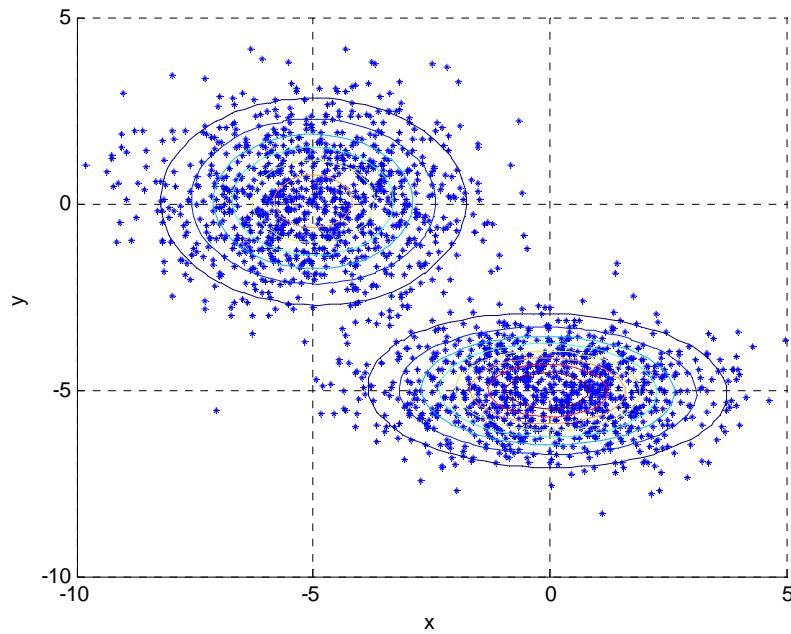


Figure 2.2. Samples and Contours of the previous distribution.

## 2.2 The Likelihood function

The likelihood function is a function of the parameters of a statistical model given the observation data. The maximum likelihood solution selects values for the model parameters that give a distribution which gives the observed data the highest probability. Let us introduce the Maximum Likelihood estimate (MLE) of a Gaussian Mixture Model (GMM). Maximizing the likelihood function is the same as maximizing the Log Likelihood function since the likelihood function is a monotonic function. This, in the case of one Gaussian or when there is no hidden variables would simplify the math a lot since the likelihood function is exponential function, so the log function will cancel the exponential and we would be left with the exponent. The Log likelihood function for the data is given by,

$$\begin{aligned}
 \mathcal{L}(\theta) &= \sum_{n=1}^N \log p(y_n/\theta) \\
 &= \sum_{n=1}^N \log \sum_{m=1}^M \pi_m p(y_n/\omega_m, \theta_m) \\
 &= \sum_{n=1}^N \log \sum_{m=1}^M \pi_m \mathcal{N}(y_n; \mu_m, \Sigma_m)
 \end{aligned} \tag{2.2}$$

Where, it is clear here that the likelihood function depends on the PDF at a given mixture component. In the estimation problem we've assumed that the noise is AWGN, so  $\Sigma_m = \sigma_m^2 \mathbf{I}$ , but  $\Sigma_m$  in general may take any covariance matrix form. Therefore the log Likelihood function becomes,

$$\mathcal{L}(\theta) = \sum_{n=1}^N \log \sum_{m=1}^M \frac{\pi_m}{(2\pi\sigma_m^2)^{\frac{d}{2}}} \exp\left(-\frac{\|y_n - \mu_m\|^2}{2\sigma_m^2}\right) \tag{2.3}$$

$$\hat{\theta}_{ML} = \underset{\theta}{\operatorname{argmax}} \mathcal{L}(\theta) \tag{2.4}$$

This may be solved by differentiating the log Likelihood function and equate it to zero as follows,

$$\nabla_{\theta} \mathcal{L}(\theta) = \mathbf{0},$$

The maximum likelihood solution for this likelihood function is given by setting the partial derivative of the log likelihood function,  $\mathcal{L}(\theta)$  with respect to the parameter of interest to zero and solving for the parameter of interest. Considering a particular parameter,  $\theta_m$ , which is related to the  $m^{\text{th}}$  mixture component.

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta_m} = \sum_{n=1}^N \left\{ \frac{\pi_m}{p(y_n/\theta)} \frac{\partial p(y_n/\omega_m, \theta_m)}{\partial \theta_m} \right\} = 0$$

Where,  $\theta_m = [\mu_m, \Sigma_m]^T$  in the Gaussian distribution case, the distribution of interest,

Let us first let  $\theta_m = \mu_m$  then,

$$\frac{\partial \mathcal{L}(\theta)}{\partial \mu_m} = \sum_{n=1}^N \frac{\pi_m \mathcal{N}(y_n/\mu_m, \Sigma_m)}{\sum_{k=1}^M \pi_k \mathcal{N}(y_n/\mu_k, \Sigma_k)} \Sigma_m^{-1} (y_n - \mu_m) = 0$$

$$\text{we know that, } P(\omega_m/y_n, \theta) = \frac{\pi_m \mathcal{N}(y_n/\mu_m, \Sigma_m)}{\sum_{k=1}^M \pi_k \mathcal{N}(y_n/\mu_k, \Sigma_k)}$$

$$\text{Then, } \hat{\mu}_m = \frac{\sum_{n=1}^N P(\omega_m/y_n, \theta) y_n}{\sum_{n=1}^N P(\omega_m/y_n, \theta)} \quad (2.5)$$

Similarly, let  $\theta_m = \Sigma_m$  then,

$$\frac{\partial \mathcal{L}(\theta)}{\partial \Sigma_m} = \sum_{n=1}^N \frac{\pi_m}{\sum_{k=1}^M \pi_k \mathcal{N}(y_n/\mu_k, \Sigma_k)} \frac{\partial \mathcal{N}(y_n/\mu_m, \Sigma_m)}{\partial \Sigma_m} = 0$$

$$\frac{\partial \mathcal{L}(\theta)}{\partial \Sigma_m} = \sum_{n=1}^N \frac{\pi_m}{\sum_{k=1}^M \pi_k \mathcal{N}(y_n/\mu_k, \Sigma_k)} \frac{1}{2} \mathcal{N}(y_n/\mu_m, \Sigma_m) ((\Sigma_m^{-1} (y_n - \mu_m)) (\Sigma_m^{-1} (y_n - \mu_m))^T - \Sigma_m^{-1})$$

$$\text{Then, } \hat{\Sigma}_m = \frac{\sum_{n=1}^N P(\omega_m/y_n, \theta) (y_n - \mu_m) (y_n - \mu_m)^T}{\sum_{n=1}^N P(\omega_m/y_n, \theta)} \quad (2.6)$$

We conclude that  $\hat{\mu}_m, \hat{\Sigma}_m$  estimates will only be correct when using the correct posterior distribution  $P(\omega_m/y_n, \theta)$ , but the posterior distribution  $P(\omega_m/y_n, \theta)$  depends on them both. So, there is no compact form solution for these parameters. We then turn to an iterative solution to finding the maximum likelihood estimate which will be our target in this research where we will be using the Expectation Maximization as an iterative approach in finding the Maximum Likelihood solution.

Let us now maximize the log likelihood function with respect to the mixing coefficients,  $\pi_m$ . This is a constraint optimization problem, so Lagrange multiplier will be used to solve this problem.



$$\mathcal{L}(\theta, \lambda) = \sum_{n=1}^N \log p(y_n/\theta) + \lambda \left( \sum_{m=1}^M \pi_m - 1 \right)$$

Then now differentiate this function,  $\mathcal{L}(\theta, \lambda)$ , with respect to both, the mixing coefficients,  $\pi_m$ , and the Lagrange multiplier,  $\lambda$ , we get the estimated mixture component as below,

$$\hat{\pi}_m = \frac{1}{N} \sum_{n=1}^N P(\omega_m/y_n, \theta) \tag{2.7}$$

Finally, we realize that without the knowledge of the class or component of each drawn sample  $P(\omega_m/y_n, \theta)$ , we cannot find the maximum likelihood estimate of the parameters of interest. The estimates  $\hat{\mu}_m$  and  $\hat{\Sigma}_m$  will be one iteration in the direction of the maximum likelihood estimate of the EM algorithm which would be clear in the next chapter.

### 3 The Expectation Maximization (EM) Algorithm

#### 3.1 Introduction

The EM algorithm was invented and implemented by several researchers till Dempster [6] collected their ideas together, assured convergence, and stated the term “EM algorithm. “ which has been used since then. One of the main application areas of the EM algorithm is estimating parameters of mixture distribution which is the aim in this research. The EM algorithm has broad areas of applications some of which is in genetics, econometric, clinical, and sociological studies. In signal processing areas like Maximum Likelihood tomographic image reconstruction, training of hidden Markov models in speech recognition [7]. The advent of the EM algorithm has come to solving the problem of latent (hidden) (unobserved) variables which MLE couldn't afford. If we introduce a joint distribution over both the observed and hidden variables, then the corresponding distribution (the distribution of the observed variables alone) is given by marginalization. This would help putting complex distributions over observed variables to be in a more tractable form of distribution formed by both observed and latent variables. The main focus in this study will be on EM algorithm in Mixtures of Gaussians (MoG). The latent variable in our case would be the class label. We would be given a bunch of data which are not labeled, and need to estimate some parameters out of it. Here we have two problems, the first is to estimate the latent or missing data, and then we come to our goal of estimating the desired parameters which would not be estimated without estimating the missing (latent) classes. The latent variable will be discrete and constant. For example, in the case of Mixture of two Gaussians, we have two latent variables. It will eventually tend to a classification problem. The class will then be belonging to either class one or two depending on the observed data. This would give which of the Gaussian mixture components is associated with each vector (sample) in the observed data. The EM algorithm would be the method of choice when direct maximum likelihood (ML) parameter estimation is not possible without the knowledge of the latent variables. The goal of the EM algorithm is to find the Maximum Likelihood (ML) solutions for models which have latent variables. Let us denote the set of observed data by  $Y$ , and denote the latent variables by  $Z$ . The set of all model parameter is denoted by  $\theta$ , and then the likelihood function would be given by,

$$\log p(Y/\theta) = \log \sum_Z p(Y, Z/\theta) \quad (3.1)$$

This is for discrete random variables, the same is true when working with continuous random variables except switching the summation in equation (3.1) in to integration over the latent variable  $Z$ . The bad news here in equation (3.1) is that the summation over the latent variable  $Z$  is inside the logarithm which made it clear that the solution of this equation won't be an easy task. We can see that since the logarithm and exponentials won't be canceled with each other as we've seen in the previous chapter. In the previous chapter the summation was outside the logarithm which made it easy to be cancelled with the exponential of the Gaussian distribution. Let us assume that for every observed sample in  $Y$ , we know the corresponding latent variable  $Z$ . Then we can call  $\{Y, Z\}$  as the complete data set. We will refer to  $Y$  as incomplete data. Maximization of the complete data log likelihood function  $\log p(Y, Z)$  is straight forward done by MLE. However, we don't really get to see the latent variables  $Z$ , but only the incomplete data  $Y$ . we know  $Z$  by the posterior distribution  $p(Z/Y, \theta)$ . Since we cannot use the complete data log likelihood, we instead take its expected value under the posterior distribution of the latent variable. This step corresponds to the E step of the EM algorithm. The M step then would be maximizing this expectation with respect to the parameters of interest. If we denote the current estimate as  $\theta^k$ , then the E and M steps would result in a better estimate  $\theta^{k+1}$ . These two steps will then be repeated till we get some precision degree [1]. We start our EM algorithm by choosing an initial point  $\theta^0$  for the parameters to start with. In the E step, we use the current parameter value  $\theta^k$  to find the posterior distribution of the latent variables  $p(Z/Y, \theta^k)$ . We then use this posterior distribution in finding the expectation of the complete data log likelihood computed at  $\theta^{k+1}$ , we denote this expectation as  $Q(\theta^k, \theta^{k+1})$ , which called auxiliary function, is given by,

$$Q(\theta^k, \theta^{k+1}) = \sum_Z p(Z/Y, \theta^k) \log p(Y, Z/\theta^{k+1}) \quad (3.2)$$

$\theta^{k+1}$  is then computed in the M step as,

$$\theta^{k+1} = \underset{\theta}{\operatorname{argmax}} Q(\theta^k, \theta^{k+1}) \quad (3.3)$$

Let us prove those equations alongside with the convergence of the EM algorithm regardless of any initial starting point. To do that, we need to introduce some definitions.

### 3.2 Jensen's Inequality

This inequality will be useful in the derivation of the update formulas in the mixture models. It says that,

$$f\left(\sum_{m=1}^M \lambda_m y_m\right) \geq \sum_{m=1}^M \lambda_m f(y_m) \quad (3.4)$$

Where  $f$  is any concave function and  $\sum_{m=1}^M \lambda_m = 1$ , and  $0 \leq \lambda_m \leq 1$ .

Also,

$$f(\mathbb{E}(y)) \geq \mathbb{E}[f(y)] \text{ or,}$$

$$f\left(\frac{1}{M} \sum_{m=1}^M y_m\right) \geq \frac{1}{M} \sum_{m=1}^M f(y_m)$$

### 3.3 Kullback-Leibler (KL) Divergence

This helps us find the divergence between two probability distribution functions. Let us take two distributions,  $p(y)$  and  $q(y)$ . Then the KL divergence between them is given by,

$$\mathcal{D}(p(y), q(y)) = \int p(y) \log \frac{p(y)}{q(y)} dy = - \int p(y) \log \frac{q(y)}{p(y)} dy$$

Using the fact that  $\log(y) \leq y - 1$ , we then get,

$$\begin{aligned} \int p(y) \log \frac{q(y)}{p(y)} dy &\leq \int p(y) \left[ \frac{q(y)}{p(y)} - 1 \right] dy \\ &\leq \int (q(y) - p(y)) dy \\ &\leq 0 \end{aligned}$$

Then we see that  $\mathcal{D}(p(y), q(y)) \geq 0$ .

This gives the following,

$$\int p(y) \log p(y) dy \geq \int p(y) \log q(y) dy$$

Or in discrete version,

$$\sum_Y p(y) \log p(y) \geq \sum_Y p(y) \log q(y) \quad (3.5)$$

Let us try to write the KL divergence in the case of two Gaussians,

Let,  $p(y) = \mathcal{N}(y; \mu_1, \Sigma_1)$  and,  $q(y) = \mathcal{N}(y; \mu_2, \Sigma_2)$  then,

$$\mathcal{D}(p(y), q(y)) = \frac{1}{2} \text{tr}(\Sigma_2^{-1}\Sigma_1 - I) + (\mu_1 - \mu_2)^T \Sigma_2^{-1}(\mu_1 - \mu_2) + \log \frac{|\Sigma_2|}{|\Sigma_1|} \quad (3.6)$$

### 3.4 The EM Derivation

If we consider a mixture of Gaussians distribution, the parameters values which need to be estimated are the means,  $\mu_1, \mu_2, \dots, \mu_M$  and covariances,  $\Sigma_1, \Sigma_2, \dots, \Sigma_M$  and component priors,  $\pi_1, \pi_2, \dots, \pi_M$ . Those values will change from iteration to iteration,  $\theta^k$  to  $\theta^{k+1}$  till they reach the optimum value or stabilize at some values. Once the  $\theta^k$  goes to  $\theta^{k+1}$ , the PDF will also iterate from  $p(Y/\theta^k)$  to  $p(Y/\theta^{k+1})$ . The increase in the log likelihood will be,

$$\begin{aligned} \mathcal{L}(\theta^{k+1}) - \mathcal{L}(\theta^k) &= \sum_{n=1}^N \log p(y_n/\theta^{k+1}) - \log p(y_n/\theta^k) \\ &= \sum_{n=1}^N \log \frac{p(y_n/\theta^{k+1})}{p(y_n/\theta^k)} \end{aligned} \quad (3.7)$$

In the case of mixture model, we need to introduce the latent variables as being one of the M mixtures.

$$\begin{aligned} \mathcal{L}(\theta^{k+1}) - \mathcal{L}(\theta^k) &= \sum_{n=1}^N \log \frac{1}{p(y_n/\theta^k)} \sum_{m=1}^M p(y_n, \omega_m/\theta^{k+1}) \\ &= \sum_{n=1}^N \log \frac{1}{p(y_n/\theta^k)} \sum_{m=1}^M \frac{p(\omega_m/y_n, \theta^k) p(y_n, \omega_m/\theta^{k+1})}{p(\omega_m/y_n, \theta^k)} \end{aligned}$$

Since the log function is strictly concave, we can apply Jensen's Inequality as follows,

$$\mathcal{L}(\theta^{k+1}) - \mathcal{L}(\theta^k) \geq \sum_{n=1}^N \sum_{m=1}^M p(\omega_m/y_n, \theta^k) \log \frac{p(y_n, \omega_m/\theta^{k+1})}{p(y_n/\theta^k) p(\omega_m/y_n, \theta^k)} \quad (3.8)$$

We've used  $\lambda_m$  as  $p(\omega_m/y_n, \theta^k)$  in Jensen's inequality.

Also, we've seen from (3.2) that the auxiliary function may be written as,

$$Q(\theta^k, \theta^{k+1}) = \sum_{n=1}^N \sum_{m=1}^M p(\omega_m/y_n, \theta^k) \log p(y_n, \omega_m/\theta^{k+1})$$

We can then rewrite equation (3.7) as,

$$\mathcal{L}(\theta^{k+1}) - \mathcal{L}(\theta^k) \geq Q(\theta^k, \theta^{k+1}) - Q(\theta^k, \theta^k)$$

The difference in the auxiliary function gives a lower bound on the increase in the log likelihood. We can see that  $Q(\theta^k, \theta^k)$  depends on the current parameters, so if we just maximize the auxiliary function  $Q(\theta^k, \theta^{k+1})$ , the likelihood function will also be maximized as a result of that. Now our aim is to maximize the function  $Q(\theta^k, \theta^{k+1})$  to get to the parameters of interest. We can do that by differentiating the auxiliary function  $Q(\theta^k, \theta^{k+1})$  with respect to the new parameters  $\theta^{k+1}$  and equate the result to zero. We need to make sure that when maximizing the function with respect to the mixing proportion, we need to use Lagrange multiplier.

$$\nabla_{\theta^{k+1}} Q(\theta^k, \theta^{k+1}) = 0.$$

What we need to guarantee is that as we iterate from  $\theta^k$  to  $\theta^{k+1}$ , we need to make sure that we are in fact increasing the log likelihood function or else we are not improving or going towards the optimum solution. We need,

$$\mathcal{L}(\theta^{k+1}) \geq \mathcal{L}(\theta^k) \text{ or,}$$

$$\mathcal{L}(\theta^{k+1}) - \mathcal{L}(\theta^k) \geq 0.$$

Let us introduce the posterior distribution of the latent variable  $p(Z/y, \theta^k)$ ,

$$\log p(y/\theta^{k+1}) - \log p(y/\theta^k) = \sum_Z p(Z/y, \theta^k) [\log p(y/\theta^{k+1}) - \log p(y/\theta^k)]$$

Since,

$$\sum_Z p(Z/y, \theta^k) = 1$$

From the definition of conditional probability,

$$p(y/\theta^{k+1}) = \frac{p(Z, y/\theta^{k+1})}{p(Z/y, \theta^{k+1})}$$

We can then write the following,

$$\sum_Z p(Z/y, \theta^k) \log p(y/\theta^{k+1}) = \sum_Z p(Z/y, \theta^k) \log \frac{p(Z, y/\theta^{k+1})}{p(Z/y, \theta^{k+1})}$$

Similarly for the second term,

$$\sum_Z p(Z/y, \theta^k) \log p(y/\theta^k) = \sum_Z p(Z/y, \theta^k) \log \frac{p(Z, y/\theta^k)}{p(Z/y, \theta^k)}$$

Combining the last two equations we get,

$$\begin{aligned} \mathcal{L}(\theta^{k+1}) - \mathcal{L}(\theta^k) &= \sum_Z p(Z/y, \theta^k) \log p(y, Z/\theta^{k+1}) - \sum_Z p(Z/y, \theta^k) \log p(Z/y, \theta^{k+1}) \\ &\quad - \sum_Z p(Z/y, \theta^k) \log p(y, Z/\theta^k) + \sum_Z p(Z/y, \theta^k) \log p(Z/y, \theta^k) \end{aligned}$$

From the definition of the KL divergence,

$$\mathcal{D}(p(Z/y, \theta^k), p(Z/y, \theta^{k+1})) = \sum_Z p(Z/y, \theta^k) \log \frac{p(Z/y, \theta^k)}{p(Z/y, \theta^{k+1})} \geq 0$$

Using this KL divergence equation in the last equation we get,

$$\mathcal{L}(\theta^{k+1}) - \mathcal{L}(\theta^k) \geq \sum_Z p(Z/y, \theta^k) \log p(y, Z/\theta^{k+1}) - \sum_Z p(Z/y, \theta^k) \log p(y, Z/\theta^k)$$

Where, the difference between the left and right hand sides is the KL divergence. If we can assure that the right hand side is positive, then the left hand side will be also positive.

Suffices to show that,

$$\sum_Z p(Z/y, \theta^k) \log p(y, Z/\theta^{k+1}) \geq \sum_Z p(Z/y, \theta^k) \log p(y, Z/\theta^k)$$

Which will assure that,

$$\mathcal{L}(\theta^{k+1}) \geq \mathcal{L}(\theta^k).$$

Let us recall the auxiliary function,

$$Q(\theta^k, \theta^{k+1}) = \sum_Z p(Z/y, \theta^k) \log p(y, Z/\theta^{k+1})$$

If the auxiliary function increases, then the likelihood will also increase. If,

$$Q(\theta^k, \theta^{k+1}) \geq Q(\theta^k, \theta^k)$$

Then,

$$\mathcal{L}(\theta^{k+1}) \geq \mathcal{L}(\theta^k).$$

We can see the difference between the likelihoods and the auxiliary functions would be the KL divergence, as follows,

$$[\mathcal{L}(\theta^{k+1}) - \mathcal{L}(\theta^k)] - [Q(\theta^k, \theta^{k+1}) - Q(\theta^k, \theta^k)] = \mathcal{D}(p(Z/y, \theta^k), p(Z/y, \theta^{k+1})) \quad (3.9)$$

The increase in the auxiliary function would be a lower bound on the increase of the log likelihood. We noted that maximizing the auxiliary function once, doesn't result in the optimum parameter values, we need to iterate till convergence is being achieved. We can now summarize the EM algorithm in two steps:

⇒ The Expectation step which is done after guessing a starting point  $\theta^0$  and calculating the posterior distribution of the latent variable,  $p(Z/y, \theta^k)$ , by taking the expected value of the log likelihood of the complete data in terms of the new parameter values  $\theta^{k+1}$ ,  $\log(p(y, Z/\theta^{k+1})|y, \theta^k)$  with respect to the posterior distribution of the latent variables,  $p(Z/y, \theta^k)$ . As follows,

$$Q(\theta^k, \theta^{k+1}) = \mathbb{E}_{p(Z/y, \theta^k)}\{\mathbf{log} p(y, Z/\theta^{k+1})|y, \theta^k\} \quad (3.10)$$

⇒ The Maximization step which maximizes the auxiliary function,  $Q(\theta^k, \theta^{k+1})$ , with respect to the new parameters  $\theta^{k+1}$ .

$$\theta^{k+1} = \underset{\theta}{\operatorname{argmax}} Q(\theta^k, \theta^{k+1})$$

The major problem with the initial parameter values that if there is at least one local maxima besides the global maxima, then the EM algorithm might only get a local maxima instead of getting



the global maxima. One way of trying to overcome this problem is by starting at many different initial points and then chooses the one which gives the highest likelihood.

### 3.5 Mixture of Gaussians and the EM algorithm

One of the applications of the EM algorithm is to find the MLE of the Mixture of Gaussians parameters in the existing of unobserved variables or data. We've seen that the parameters couldn't be found by a direct MLE. An iterative procedure needs to be done in order for us to get to the optimum results for the parameters. The major problem which we encounter here is that which Gaussian component was responsible for generating that specific sample. We know that the latent variables are discrete which determines the number of mixtures which have been used in generating the observed data. If we have prior knowledge about the components which were responsible for generating the data, then the problem would become easier and could be solved using a direct MLE. So now our major problem is estimating the component which was responsible for generating each data point in my observed data. Once we know the posterior distribution of the hidden variables  $p(Z/Y, \Theta)$ , the problem then become like a direct MLE problem in the auxiliary function sense.

Assume that,  $z_{nk} = \begin{cases} 1 & \text{if the observed data } y_n \text{ was generated by component } \omega_k \\ 0 & \text{otherwise} \end{cases}$

Let us have a look at a single received sample which we suppose that it was generated by component  $\omega_k$ . We can write,

$$\begin{aligned} p(z_n, y_n / \Theta) &= p(y_n / \omega_k, \theta_k) p(\omega_k) \\ &= \prod_{m=1}^M [p(y_n / \omega_m, \theta_m) p(\omega_m)]^{z_{nm}} \end{aligned} \quad (3.11)$$

In addition of Y being *i. i. d.* samples, Z also are *i. i. d.* samples,

$$p(Z, \mathbf{y} / \Theta) = \prod_{n=1}^N p(z_n, y_n / \Theta)$$

In the Expectation step of the EM algorithm we need to compute iteratively the posterior distribution of the latent variable  $p(Z/\mathbf{y}, \Theta^k)$  in order to find the auxiliary function. Then we maximize the auxiliary function.

$$\begin{aligned}
Q(\theta^k, \theta^{k+1}) &= \sum_Z p(Z/\mathbf{y}, \theta^k) \mathbf{log} p(Z, \mathbf{y}/\theta^{k+1}) \\
&= \sum_Z p(Z/\mathbf{y}, \theta^k) \sum_{n=1}^N \mathbf{log} p(y_n, z_n/\theta^{k+1}) \\
&= \sum_{n=1}^N \sum_Z p(Z/\mathbf{y}, \theta^k) \mathbf{log} p(y_n, z_n/\theta^{k+1}) \\
&= \sum_{n=1}^N \sum_{z_n} p(z_n/y_n, \theta^k) \mathbf{log} p(y_n, z_n/\theta^{k+1}) \\
&= \sum_{n=1}^N \sum_{z_n} p(z_n/y_n, \theta^k) \sum_{m=1}^M z_{nm} \mathbf{log} (p(y_n/\omega_m, \theta_m^{k+1}) p^{k+1}(\omega_m)) \\
&= \sum_{n=1}^N \sum_{z_n} p(z_n/y_n, \theta^k) \sum_{m=1}^M z_{nm} \mathbf{log} p(y_n/\omega_m, \theta_m^{k+1}) \\
&\quad + \sum_{n=1}^N \sum_{z_n} p(z_n/y_n, \theta^k) \sum_{m=1}^M z_{nm} \mathbf{log} p^{k+1}(\omega_m) \\
Q(\theta^k, \theta^{k+1}) &= \sum_{m=1}^M \left\{ \sum_{n=1}^N p(\omega_m/y_n, \theta^k) \mathbf{log} p(y_n/\omega_m, \theta_m^{k+1}) \right. \\
&\quad \left. + \sum_{n=1}^N p(\omega_m/y_n, \theta^k) \mathbf{log} p^{k+1}(\omega_m) \right\} \tag{3.12}
\end{aligned}$$

We know that the log likelihood function for a component mixture is,

$$\mathbf{log} p(\mathbf{y}; \mu_m, \Sigma_m) = -\frac{1}{2} [\mathbf{log} (2\pi)^D |\Sigma_m| + (\mathbf{y} - \mu_m)^T \Sigma_m^{-1} (\mathbf{y} - \mu_m)]$$

This is then substituted in the auxiliary function as follows,

$$\begin{aligned}
\mathcal{Q}(\theta^k, \theta^{k+1}) &= \sum_{m=1}^M \sum_{n=1}^N p(\omega_m/y_n, \theta^k) \left\{ -\frac{1}{2} (y_n - \hat{\mu}_m)^T \hat{\Sigma}_m^{-1} (y_n - \hat{\mu}_m) \right\} \\
&+ \sum_{m=1}^M \sum_{n=1}^N p(\omega_m/y_n, \theta^k) \left\{ -\frac{1}{2} \log \{ (2\pi)^D |\hat{\Sigma}_m| \} \right\} \\
&+ \sum_{m=1}^M \sum_{n=1}^N p(\omega_m/y_n, \theta^k) \log p^{k+1}(\omega_m)
\end{aligned} \tag{3.13}$$

We now need to estimate the parameters of interest of component  $\omega_m$  at iteration  $k + 1$  by differentiating the auxiliary function with respect to one of the parameters and equating to zero,

$$\nabla_{\theta} \mathcal{Q}(\theta^k, \theta^{k+1}) = 0$$

In the case of estimating the mean of the  $m^{th}$  mixture component,

$$\nabla_{\mu_m} \mathcal{Q}(\theta^k, \theta^{k+1}) = 0$$

This will result in the updating mean formula,

$$\hat{\mu}_m = \frac{\sum_{n=1}^N p(\omega_m/y_n, \theta^k) y_n}{\sum_{n=1}^N p(\omega_m/y_n, \theta^k)} \tag{3.14}$$

$$\nabla_{\Sigma_m} \mathcal{Q}(\theta^k, \theta^{k+1}) = 0$$

This will result in the updating Covariance formula,

$$\hat{\Sigma}_m = \frac{\sum_{n=1}^N p(\omega_m/y_n, \theta^k) (y_n - \hat{\mu}_m)(y_n - \hat{\mu}_m)^T}{\sum_{n=1}^N p(\omega_m/y_n, \theta^k)} \tag{3.15}$$

Equations (3.13) to (3.15) will be iterated till convergence is achieved.

To sum up, the EM algorithm for Mixture of Gaussians may be summarized as follows,

1. Initialize the parameters mean  $\mu_m^0$ , Covariance  $\Sigma_m^0$ , and mixing coefficient  $\pi_m^0 = p^0(\omega_m)$ , then calculate the initial value for the log likelihood function.
2. In the E step, we need to calculate the responsibilities,  $p(\omega_m/y_n, \theta^k)$ ,
$$p(\omega_m/y_n, \theta^k) = \frac{\pi_m \mathcal{N}(y_n/\mu_m, \Sigma_m)}{\sum_{j=1}^M \pi_j \mathcal{N}(y_n/\mu_j, \Sigma_j)}$$
3. In the M step, we need to re estimate the parameters, using the current responsibility,

$$\mu_m^{k+1} = \frac{\sum_{n=1}^N p(\omega_m/y_n, \theta^k) y_n}{\sum_{n=1}^N p(\omega_m/y_n, \theta^k)}$$

$$\Sigma_m^{k+1} = \frac{\sum_{n=1}^N p(\omega_m/y_n, \theta^k) (y_n - \mu_m^{k+1})(y_n - \mu_m^{k+1})^T}{\sum_{n=1}^N p(\omega_m/y_n, \theta^k)}$$

$$\pi_m^{k+1} = \frac{\sum_{n=1}^N p(\omega_m/y_n, \theta^k)}{N}$$

4. Evaluate the log likelihood checking for increasing the log likelihood and stopping the iteration process if a given precision has been achieved.

$$\log p(\mathbf{y}/\pi, \mu, \Sigma) = \sum_{n=1}^N \log \sum_{m=1}^M \pi_m \mathcal{N}(y_n/\mu_m, \Sigma_m)$$

If the precision hasn't been achieved, we go back to point number 2 and iterate once again. We do the same procedure till we achieve the precision condition.

For instance, let us consider an example showing a mixture of four Gaussians as an illustration example. I assumed that the means as the values of the four sent symbols in the QPSK case,  $a_n \in \{\pm 1, \pm 1\}$ , and have equal covariance of  $\Sigma = 0.35 I$ , where  $I$  is a  $2 \times 2$  matrix. Figure 3.1 shown shows these generated data with its Gaussian fit using the previous algorithm, my assumption model was like,

$$\mathbf{y} = a_n + \mathbf{w}$$

Where  $\mathbf{y}$  is the received data which corrupted by additive white Gaussian noise AWGN,  $\mathbf{w}$  of covariance,  $\Sigma$ , both of size  $N$ , and  $a_n = \{\pm 1, \pm 1\}$ . This is done by generating 800 samples drawn from each mixture of four Gaussians and then pretending like we don't know the parameters from which those samples were drawn and trying to estimate those parameters by finding the best contour match and estimating the parameters using the EM algorithm.

As we can see from the coming figures, the more we increase the variances, the more the inference become difficult and we might not get the true parameters as a result. As we decrease the variance, the mixtures would become more circular and separated and easy to be classified. Figure 3.2 shows the mixture of four Gaussians exactly as in the previous figure except that we decreased the variances to 0.05. We clearly see the difference in estimating the contours of those mixtures. In this figure classification could be done easily without encountering too much error. K-means plot is also drawn to see that it is just a special case of the EM algorithm when the covariance approaches zero.

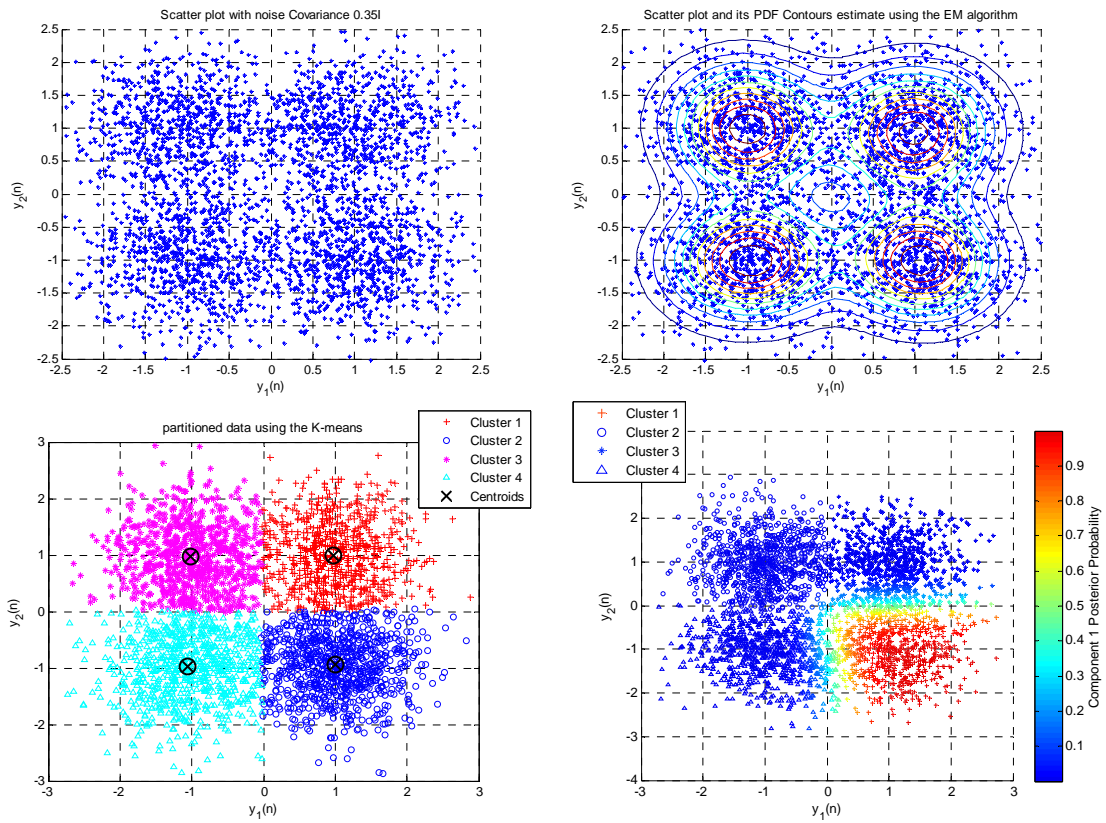


Figure 3.1. mixtures of four Gaussians data and their fit when  $C = 0.35I$

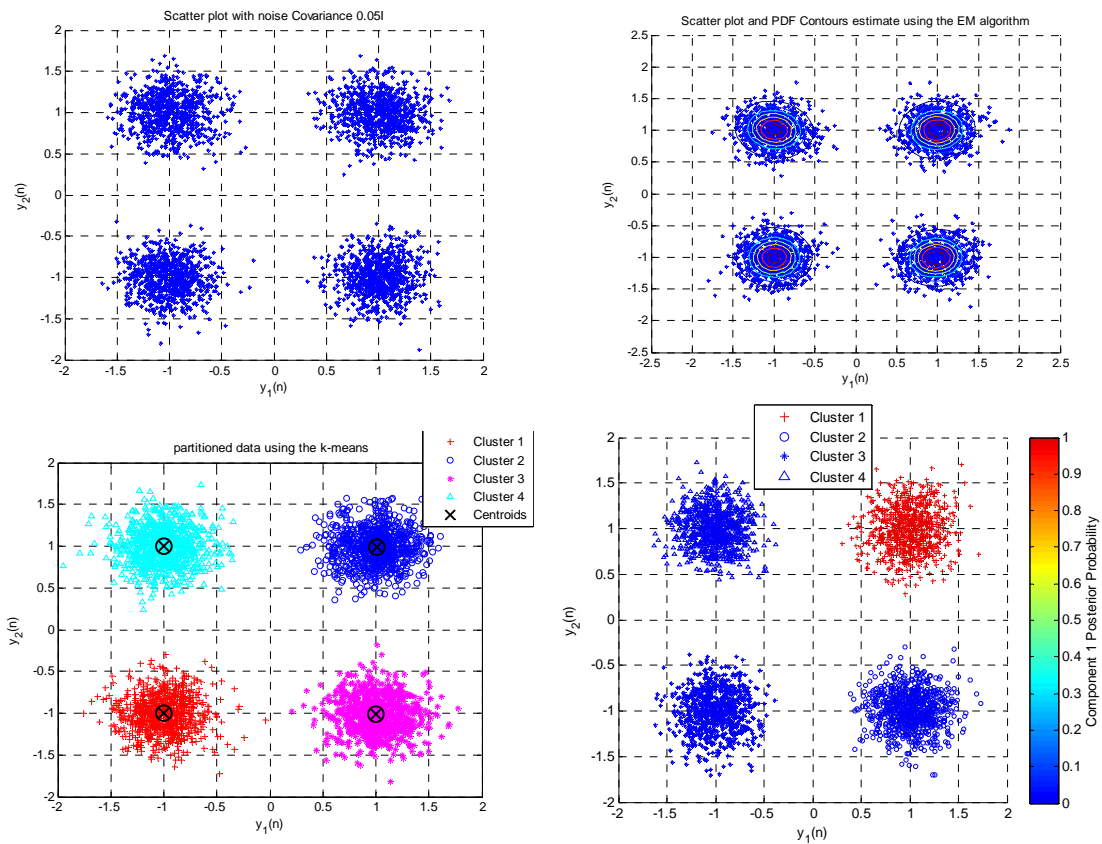


Figure 3.2. mixtures of four Gaussians data and their fit when  $C = 0.05I$

## 4 SISO NDA SNR estimation using the EM algorithm in the Literature

### 4.1 BPSK case

In the binary phase shift keying, we have a mixture of two classes,  $a_n \in \{+1, -1\}$ . Those are the transmitted symbols.

#### 4.1.1 The system model

We are assuming that there is no error in timing, phase deviation, or frequency deviation. The matched filter output may be given by,

$$y_n = a_n h + w_n, \quad n = 1, \dots, N \quad (4.1)$$

where,  $y_n$  is the received sample,  $a_n$  is the transmitted symbol,  $h$  is the channel gain, and  $w_n$  is a zero mean AWGN with variance  $\sigma^2$ .

We can write equation (4.1) in a vector form as,

$$\mathbf{y} = \mathbf{a}h + \mathbf{w} \quad (4.2)$$

where,  $\mathbf{y} = [y_1 \ y_2 \ y_3 \ \dots \ y_N]^T$ ,  $\mathbf{a} = [a_1 \ a_2 \ a_3 \ \dots \ a_N]^T$ ,  $\mathbf{w} = [w_1 \ w_2 \ w_3 \ \dots \ w_N]^T$ . The transmitted symbols may be modeled as realizations of *i.i.d.* binary random variables. The probability distribution function may be written as,

$$p(y_n/a_n, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(y_n - a_n h)^2}{2\sigma^2} \right\} \quad (4.3)$$

Assuming equally likely symbols, we have,

$$\begin{aligned} p(y_n; \theta) &= \sum_{a_n} p(y_n/a_n, \theta) p(a_n/\theta) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \left\{ \frac{1}{2} \exp \left\{ -\frac{(y_n - h)^2}{2\sigma^2} \right\} + \frac{1}{2} \exp \left\{ -\frac{(y_n + h)^2}{2\sigma^2} \right\} \right\} \end{aligned}$$

where,  $\theta = [h \ \sigma^2]^T$ . Now for N received *i.i.d.* samples we have,

$$p(\mathbf{y}; \theta) = \prod_{n=1}^N p(y_n; \theta)$$

$$p(\mathbf{y}; \theta) = \prod_{n=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \left(\frac{1}{2}\right) \left\{ \exp\left\{-\frac{(y_n - h)^2}{2\sigma^2}\right\} + \exp\left\{-\frac{(y_n + h)^2}{2\sigma^2}\right\} \right\} \quad (4.4)$$

This is the likelihood function of the received data. Now given the observation data, estimate the parameter values  $\hat{\theta} = [\hat{h} \ \widehat{\sigma^2}]^T$  and then substitute to find the SNR,

$$\widehat{SNR} = 10 \log_{10} \left( \frac{\hat{h}^2}{\widehat{\sigma^2}} \right)$$

#### 4.1.2 The CRLB for BPSK

The Cramer-Rao Lower Bound (CRLB) is a lower bound for the Mean Square Error (MSE) of any unbiased estimator which satisfies the regularity condition [2]. In our case of SNR estimation, the CRLB can be written as,

$$CRLB(SNR) = \frac{\partial SNR^T}{\partial \theta} \mathbf{I}^{-1}(\theta) \frac{\partial SNR}{\partial \theta} \quad (4.5)$$

Where  $\mathbf{I}(\theta)$  is the fisher information matrix given by,

$$\mathbf{I}(\theta) = \begin{bmatrix} -\mathbb{E} \left[ \frac{\partial^2 \log p(Y; \theta)}{\partial h^2} \right] & -\mathbb{E} \left[ \frac{\partial^2 \log p(Y; \theta)}{\partial h \partial \sigma^2} \right] \\ -\mathbb{E} \left[ \frac{\partial^2 \log p(Y; \theta)}{\partial \sigma^2 \partial h} \right] & -\mathbb{E} \left[ \frac{\partial^2 \log p(Y; \theta)}{\partial \sigma^2^2} \right] \end{bmatrix} \quad (4.6)$$

And,

$$\frac{\partial SNR}{\partial \theta} = \left[ \frac{20}{h \log(10)} \quad -\frac{10}{\sigma^2 \log(10)} \right]^T \quad (4.7)$$

**For the NDA case,**

The log likelihood function is given as,

$$\begin{aligned} \mathcal{L}(\theta) &= \sum_{n=1}^N \log p(y_n/\theta) \\ &= -\frac{N}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=1}^N [(y_n)^2 + h^2] + \sum_{n=1}^N \log \cosh\left(\frac{y_n h}{\sigma^2}\right) \end{aligned} \quad (4.8)$$

After differentiating this function with respect to the channel gain and the variance we get the following,

$$\begin{aligned}\frac{\partial^2 \log p(\mathbf{y}; \Theta)}{\partial h^2} &= -\frac{N}{\sigma^2} + \sum_{n=1}^N \operatorname{sech}^2\left(\frac{y_n h}{\sigma^2}\right) \left(\frac{y_n}{\sigma^2}\right)^2 \\ \frac{\partial^2 \log p(\mathbf{y}; \Theta)}{\partial h \partial \sigma^2} &= \frac{N h}{\sigma^4} - \sum_{n=1}^N \operatorname{sech}^2\left(\frac{y_n h}{\sigma^2}\right) \frac{y_n^2 h}{\sigma^6} - \sum_{n=1}^N \tanh\left(\frac{y_n h}{\sigma^2}\right) \frac{y_n}{\sigma^4} \\ \frac{\partial^2 \log p(\mathbf{y}; \Theta)}{\partial \sigma^2 \partial h} &= \frac{\partial^2 \log p(\mathbf{y}; \Theta)}{\partial h \partial \sigma^2} \\ \frac{\partial^2 \log p(\mathbf{y}; \Theta)}{\partial \sigma^2^2} &= \frac{N}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{n=1}^N [(y_n)^2 + h^2] + \sum_{n=1}^N \tanh\left(\frac{y_n h}{\sigma^2}\right) \frac{2y_n h}{\sigma^6} + \sum_{n=1}^N \operatorname{sech}^2\left(\frac{y_n h}{\sigma^2}\right) \left(\frac{y_n h}{\sigma^4}\right)^2 \\ \mathbf{I}(\Theta) &= \begin{bmatrix} \frac{N(1-f(\rho))}{\sigma^2} & \frac{N h}{\sigma^4} f(\rho) \\ \frac{N h}{\sigma^4} f(\rho) & \frac{N}{\sigma^4} \left\{ \frac{1}{2} - \frac{h^2 f(\rho)}{\sigma^2} \right\} \end{bmatrix}\end{aligned}\quad (4.9)$$

where,

$$f(\rho) = \frac{\exp(-\frac{\rho}{2})}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \frac{x^2 \exp(-\frac{x^2}{2})}{\cosh(x\sqrt{\rho})} dx \quad (4.10)$$

We can now substitute this result into equation (4.5) and get the following,

$$\begin{aligned}CRLB(SNR) &= \begin{bmatrix} 20 & 10 \\ h \log(10) & \sigma^2 \log(10) \end{bmatrix} \begin{bmatrix} \frac{N(1-f(\rho))}{\sigma^2} & \frac{N h}{\sigma^4} f(\rho) \\ \frac{N h}{\sigma^4} f(\rho) & \frac{N}{\sigma^4} \left\{ \frac{1}{2} - \frac{h^2 f(\rho)}{\sigma^2} \right\} \end{bmatrix} \begin{bmatrix} 20 \\ h \log(10) \\ 10 \\ -\sigma^2 \log(10) \end{bmatrix} \\ CRLB(SNR) &= \frac{200}{N(\log(10))^2} \frac{\frac{2}{\rho} - f(\rho) + 1}{1 - f(\rho) - 2\rho f(\rho)} \quad dB^2\end{aligned}\quad (4.11)$$



For the DA case,

$$\begin{aligned}\mathcal{L}(\theta) &= \sum_{n=1}^N \log p(y_n/a_n, \theta) \\ &= \sum_{n=1}^N \left\{ -\frac{1}{2} \log(2\pi\sigma^2) - \frac{(y_n - a_n h)^2}{2\sigma^2} \right\}\end{aligned}$$

where  $a_n$  is given to be either +1 or -1.

After differentiating this function with respect to the channel gain and the variance we get the following,

$$\begin{aligned}\frac{\partial^2 \log p(\mathbf{y}; \theta)}{\partial h^2} &= -\frac{N a_n^2}{\sigma^2} \\ \frac{\partial^2 \log p(\mathbf{y}; \theta)}{\partial h \partial \sigma^2} &= -\frac{N(a_n y_n - h)}{\sigma^4} \\ \frac{\partial^2 \log p(\mathbf{y}; \theta)}{\partial \sigma^2 \partial h} &= \frac{\partial^2 \log p(\mathbf{y}; \theta)}{\partial h \partial \sigma^2} \\ \frac{\partial^2 \log p(\mathbf{y}; \theta)}{\partial \sigma^2^2} &= \frac{N}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{n=1}^N (y_n - a_n h)^2\end{aligned}$$

$$\mathbf{I}(\theta) = \begin{bmatrix} \frac{N}{\sigma^2} & 0 \\ 0 & \frac{N}{2\sigma^4} \end{bmatrix} \quad 4.12$$

We may now substitute this result into equation (4.5) and get the following,

$$\begin{aligned}CRLB(SNR) &= \begin{bmatrix} \frac{20}{h \log(10)} & -\frac{10}{\sigma^2 \log(10)} \\ 0 & \frac{N}{2\sigma^4} \end{bmatrix} \begin{bmatrix} \frac{N}{\sigma^2} & 0 \\ 0 & \frac{N}{2\sigma^4} \end{bmatrix} \begin{bmatrix} \frac{20}{h \log(10)} \\ \frac{10}{\sigma^2 \log(10)} \end{bmatrix} \\ CRLB(SNR) &= \frac{200}{N(\log(10))^2} \left( \frac{2}{\rho} + 1 \right) \quad dB^2 \quad (4.13)\end{aligned}$$

### 4.1.3 NDA SNR estimation using the EM algorithm

We now introduce the method of EM algorithm in estimating the SNR. Our assumptions are that the data was drawn from a mixture of two Gaussians, and are uncorrelated with each other. The samples are drawn in an *i. i. d.* manner. The parameters which we need to estimate now are,

$$\begin{aligned}\hat{\Theta} &= [\hat{h} \quad \hat{\sigma}^2]^T \\ &= \underset{\Theta}{\operatorname{argmax}} p(\mathbf{y}; \Theta)\end{aligned}$$

$$\hat{\rho} = \frac{\hat{h}^2}{\hat{\sigma}^2}$$

$$\widehat{SNR} = 10 \log_{10}(\hat{\rho}) \quad (4.14)$$

The EM algorithm will be used to iteratively estimate the channel gain and the noise variance then use them to estimate the signal to noise ratio. The received data will be seen as incomplete data set since we don't know its class (transmitted symbol). The auxiliary function in this case may be given as,

$$\begin{aligned}Q(\Theta^k, \Theta^{k+1}) &= \mathbb{E}_{p(Z/y, \Theta^k)} \{ \log(p(\mathbf{y}, Z/\Theta^{k+1})) | \mathbf{y}, \Theta^k \} \\ Q(\Theta^k, \Theta^{k+1}) &= \sum_{n=1}^N \sum_{m=1}^M p(\omega_m/y_n, \Theta^k) \log p(y_n/\omega_m, \Theta^{k+1}) \\ &\quad + \sum_{n=1}^N \sum_{m=1}^M p(\omega_m/y_n, \Theta^k) \log p^{k+1}(\omega_m)\end{aligned} \quad (4.15)$$

$$\Theta^{k+1} = \underset{\Theta}{\operatorname{argmax}} Q(\Theta^k, \Theta^{k+1}) \quad (4.16)$$

Now solving those two equations will give us the E & M steps. Since we know the mixing proportion  $p(\omega_m)$ , we don't need to estimate it.

The auxiliary function may be given as,

$$Q(\Theta^k, \Theta^{k+1}) = \sum_{n=1}^N \sum_{m=1}^M p(\omega_m/y_n, \Theta^k) \log p(y_n/\omega_m, \Theta^{k+1})$$

$$\begin{aligned} Q(\theta^k, \theta^{k+1}) &= \sum_{m=1}^M \sum_{n=1}^N p(\omega_m/y_n, \theta^k) \left\{ -\frac{1}{2} \left( (y_n - \mu_m^{k+1})^T \Sigma_m^{k+1^{-1}} (y_n - \mu_m^{k+1}) \right) \right\} \\ &\quad + \sum_{m=1}^M \sum_{n=1}^N p(\omega_m/y_n, \theta^k) \left\{ -\frac{1}{2} (\log [(2\pi)^D |\Sigma_m^{k+1}|]) \right\} + \sum_{m=1}^M \sum_{n=1}^N p(\omega_m/y_n, \theta^k) \log \{C\} \end{aligned}$$

$$\log p(y_n/\omega_m, \theta^{k+1}) = -\frac{1}{2} \log(2\pi\sigma^{2^{k+1}}) - \frac{(y_n - a_m h^{k+1})^2}{2\sigma^{2^{k+1}}}$$

$$\frac{\partial Q(\theta^k, \theta^{k+1})}{\partial h^{k+1}} = \frac{\partial}{\partial h^{k+1}} \left\{ \sum_{m=1}^M \sum_{n=1}^N p(\omega_m/y_n, \theta^k) \left\{ -\frac{(y_n - a_m h^{k+1})^2}{2\sigma^{2^{k+1}}} \right\} \right\} = 0$$

$$\sum_{m=1}^M \sum_{n=1}^N p(\omega_m/y_n, \theta^k) \left\{ \frac{a_m (y_n - a_m h^{k+1})}{\sigma^{2^{k+1}}} \right\} = 0$$

$$\sum_{m=1}^M \sum_{n=1}^N p(\omega_m/y_n, \theta^k) \{a_m y_n - h^{k+1}\} = 0$$

$$\sum_{m=1}^M \sum_{n=1}^N p(\omega_m/y_n, \theta^k) \{a_m y_n\} = N h^{k+1}$$

$$\sum_{n=1}^N y_n p(\omega_1/y_n, \theta^k) - \sum_{n=1}^N y_n p(\omega_2/y_n, \theta^k) = N h^{k+1}$$

since,  $p(\omega_1/y_n, \theta^k) + p(\omega_2/y_n, \theta^k) = 1$

$$\sum_{n=1}^N y_n p(\omega_1/y_n, \theta^k) - \sum_{n=1}^N y_n [1 - p(\omega_1/y_n, \theta^k)] = N h^{k+1}$$

$$N h^{k+1} = \sum_{n=1}^N y_n \{2 p(\omega_1/y_n, \theta^k) - 1\}$$

$$h^{k+1} = \frac{1}{N} \sum_{n=1}^N y_n (2 p(\omega_1/y_n, \theta^k) - 1)$$

$$\begin{aligned}
\{2p(\omega_1/y_n, \theta^k) - 1\} &= 2 \left\{ \frac{\exp(-\frac{(y_n - h^k)^2}{2\sigma^{2k}})}{\exp(-\frac{(y_n - h^k)^2}{2\sigma^{2k}}) + \exp(-\frac{(y_n + h^k)^2}{2\sigma^{2k}})} \right\} - 1 \\
&= \frac{2\exp(-\frac{(y_n - h^k)^2}{2\sigma^{2k}}) - \exp(-\frac{(y_n - h^k)^2}{2\sigma^{2k}}) - \exp(-\frac{(y_n + h^k)^2}{2\sigma^{2k}})}{\exp(-\frac{(y_n - h^k)^2}{2\sigma^{2k}}) + \exp(-\frac{(y_n + h^k)^2}{2\sigma^{2k}})} \\
&= \frac{\exp(-\frac{(y_n - h^k)^2}{2\sigma^{2k}}) - \exp(-\frac{(y_n + h^k)^2}{2\sigma^{2k}})}{\exp(-\frac{(y_n - h^k)^2}{2\sigma^{2k}}) + \exp(-\frac{(y_n + h^k)^2}{2\sigma^{2k}})} \\
&= \tanh \left\{ \frac{y_n h^k}{\sigma^{2k}} \right\}
\end{aligned}$$

$$h^{k+1} = \frac{1}{N} \sum_{n=1}^N y_n \tanh \left\{ \frac{y_n h^k}{\sigma^{2k}} \right\} \quad (4.17)$$

Solving for the variance, we have,

$$\begin{aligned}
\frac{\partial Q(\theta^k, \theta^{k+1})}{\partial \sigma^{2k+1}} &= \\
\frac{\partial}{\partial \sigma^{2k+1}} \left\{ \sum_{n=1}^N p(\omega_1/y_n, \theta^k) \log p(y_n/\omega_1, \theta^{k+1}) + \sum_{n=1}^N p(\omega_2/y_n, \theta^k) \log p(y_n/\omega_2, \theta^{k+1}) \right\} &= 0 \\
\sum_{n=1}^N p(\omega_1/y_n, \theta^k) \left\{ -\frac{1}{2\sigma^{2k+1}} + \frac{(y_n - h^{k+1})^2}{2\sigma^{4k+1}} \right\} + \sum_{n=1}^N p(\omega_2/y_n, \theta^k) \left\{ -\frac{1}{2\sigma^{2k+1}} + \frac{(y_n + h^{k+1})^2}{2\sigma^{4k+1}} \right\} &= 0 \\
\sum_{n=1}^N \left\{ p(\omega_1/y_n, \theta^k) \left\{ -\frac{1}{2\sigma^{2k+1}} + \frac{(y_n - h^{k+1})^2}{2\sigma^{4k+1}} \right\} + [1 - p(\omega_1/y_n, \theta^k)] \left\{ -\frac{1}{2\sigma^{2k+1}} + \frac{(y_n + h^{k+1})^2}{2\sigma^{4k+1}} \right\} \right\} &= 0 \\
\sum_{n=1}^N \left\{ p(\omega_1/y_n, \theta^k) \left\{ \frac{(y_n - h^{k+1})^2}{2\sigma^{4k+1}} \right\} + \left\{ -\frac{1}{2\sigma^{2k+1}} + \frac{(y_n + h^{k+1})^2}{2\sigma^{4k+1}} \right\} - p(\omega_1/y_n, \theta^k) \left\{ \frac{(y_n + h^{k+1})^2}{2\sigma^{4k+1}} \right\} \right\} &= 0
\end{aligned}$$

$$\sum_{n=1}^N \left\{ -2 p(\omega_1/y_n, \theta^k) \left\{ \frac{y_n h^{k+1}}{\sigma^{4k+1}} \right\} + \left\{ -\frac{1}{2\sigma^{2k+1}} + \frac{(y_n + h^{k+1})^2}{2\sigma^{4k+1}} \right\} \right\} = 0$$

$$\sum_{n=1}^N \left\{ 4 p(\omega_1/y_n, \theta^k) y_n h^{k+1} + \sigma^{2k+1} - (y_n)^2 - 2y_n h^{k+1} - (h^{k+1})^2 \right\} = 0$$

$$\sum_{n=1}^N \left\{ 2y_n h^{k+1} (2 p(\omega_1/y_n, \theta^k) - 1) + \sigma^{2k+1} - (y_n)^2 - (h^{k+1})^2 \right\} = 0$$

$$\sum_{n=1}^N \left\{ 2y_n h^{k+1} \tanh \left\{ \frac{y_n h^k}{\sigma^{2k}} \right\} + \sigma^{2k+1} - (y_n)^2 - (h^{k+1})^2 \right\} = 0$$

$$2N(h^{k+1})^2 + N\sigma^{2k+1} - \sum_{n=1}^N (y_n)^2 - N(h^{k+1})^2 = 0$$

$$N\sigma^{2k+1} = \sum_{n=1}^N y_n^2 - N(h^{k+1})^2$$

$$\sigma^{2k+1} = \frac{1}{N} \sum_{n=1}^N y_n^2 - (h^{k+1})^2 \tag{4.18}$$

$$\hat{\rho} = \frac{\hat{h}^2}{\sigma^2}$$

$$\widehat{SNR} = 10 \log_{10}(\hat{\rho})$$

$$\mathbb{E}[\widehat{SNR}] = \frac{1}{M} \sum_{i=1}^M \widehat{SNR}_i \tag{4.19}$$

$$MSE(\widehat{SNR}) = \mathbb{E}(\widehat{SNR} - SNR)^2 = var(\widehat{SNR}) = \frac{1}{M} \sum_{i=1}^M (\widehat{SNR}_i - \mathbb{E}[\widehat{SNR}])^2 \tag{4.20}$$

We don't need to worry about the mixing coefficient since we know that they are equally likely. The hyperbolic tangent in updating the channel gain will serve as a soft threshold. When the hyperbolic tangent argument is large, this threshold would become more like a signum function threshold. After Starting with initial values for the channel gain and noise variance, equations (4.17) and (4.18) will be used to iteratively find the estimates to the parameters which would be then used to

estimate the signal to noise ratio,  $\hat{\rho}$ ,  $\widehat{SNR}$  as given above. The *MSE* will be found for different estimated values of SNR. Then this MSE will be assessed using the NDA CRLB in units of dB<sup>2</sup>. The estimation process will be done for a range of SNR from 1 dB to 20 dB, and for different values of received sample size, N, 128, 512 and 1024. Figure 4.1 shown explains three things when BPSK is used as the transmitted symbols. The red curve is the CRLB for NDA signal, the blue curve is the CRLB for DA signal, and the black curve shows the MSE estimated using the EM algorithm. The EM algorithm estimate is repeated 10000 times using Monte Carlo simulations to get 10000 realizations of the SNR estimates, then averaging them to get the estimated SNR, then finding the variance to get the MSE for that specific SNR estimate value. We then repeat this process for different values of SNR, say from 0 dB to 20 dB and their resultant MSE. Those estimates then assessed with the CRLB as shown below. We see that for low SNR estimate values, the MSE is large compared with the NDA CRLB. The higher the SNR become, the lower the MSE we get till they (the CRLB and MSE) both match. This was for sample size of N=128. The next step we would do is increase N and see what is going to happen to our estimation accuracy.

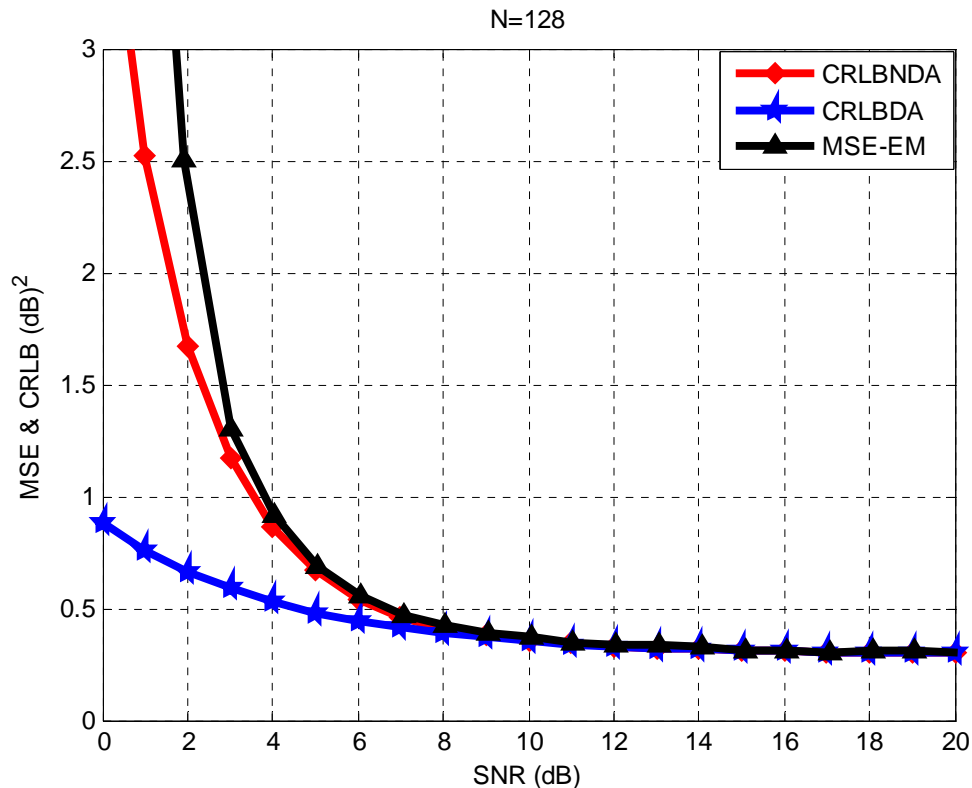


Figure 4.1 The MSE & CRLB in dB<sup>2</sup> vs. SNR in dB when N=128 for the BPSK case

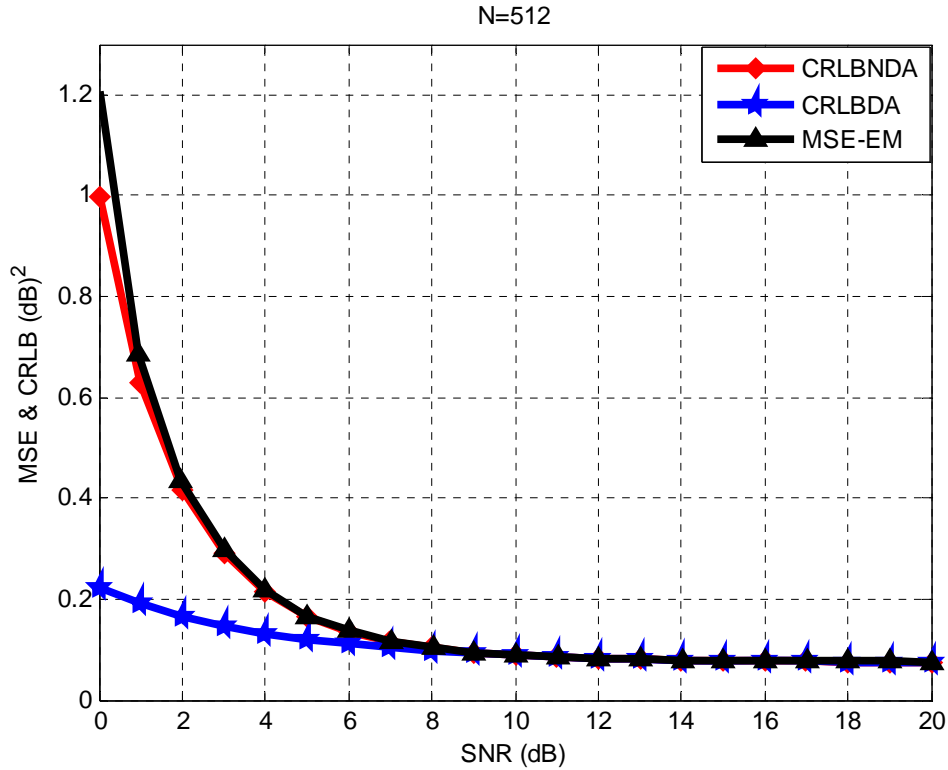


Figure 4.2 The MSE & CRLB in dB<sup>2</sup> vs. SNR in dB when N=512 for the BPSK case

Figure 4.2 shows the relation between MSE and CRLB vs. SNR for N=512. What we realize here is that the MSE & CRLB are decreased compared with the case when N=128. CRLB itself decreased from 0.3 dB<sup>2</sup> for high SNR when N was 128, to less than 0.1 dB<sup>2</sup> for high SNR when N is 512. Figure 4.3 shows the same thing but for N=1024. Figure 4.3 shows that the precision of the MSE and CRLB is in the order of less than 0.05 dB<sup>2</sup> for SNRs greater than 8 dB. Figure 4.4 shows the error which is the difference between the MSE and CRLB in dB<sup>2</sup> and the SNRs for the previous three cases of different Ns, 128, 512, and 1024. This figure proves that as N increases, the error decreases which means that the MSE become closer and closer to the CRLB until they both match at some SNR value and beyond. From the SNR equation we see that the SNR is inversely proportional to the noise variance, which means as we increase the noise variance, the SNR will decrease as a result of that, and the two mixtures then will intervene with each other which makes it difficult for classification. On the other hand however, as we decrease the noise variance, the SNR will increase as a result of that and the two mixtures then will be far away from each other which makes classification become easier and results in almost no error..

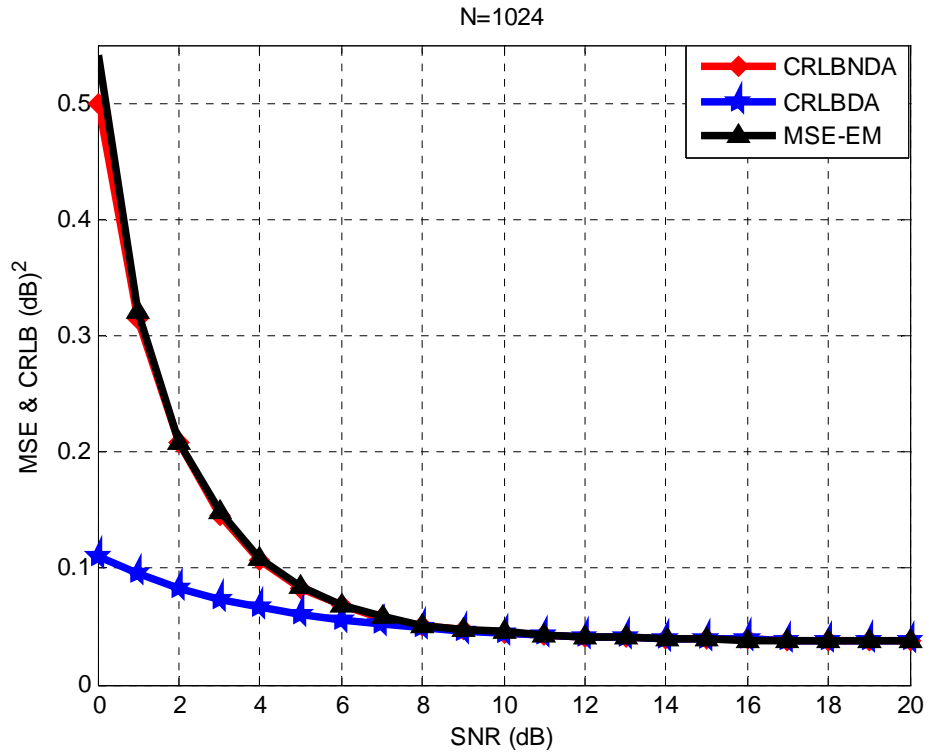


Figure 4.3 The MSE & CRLB in dB<sup>2</sup> vs. SNR in dB when N=1024 for the BPSK case

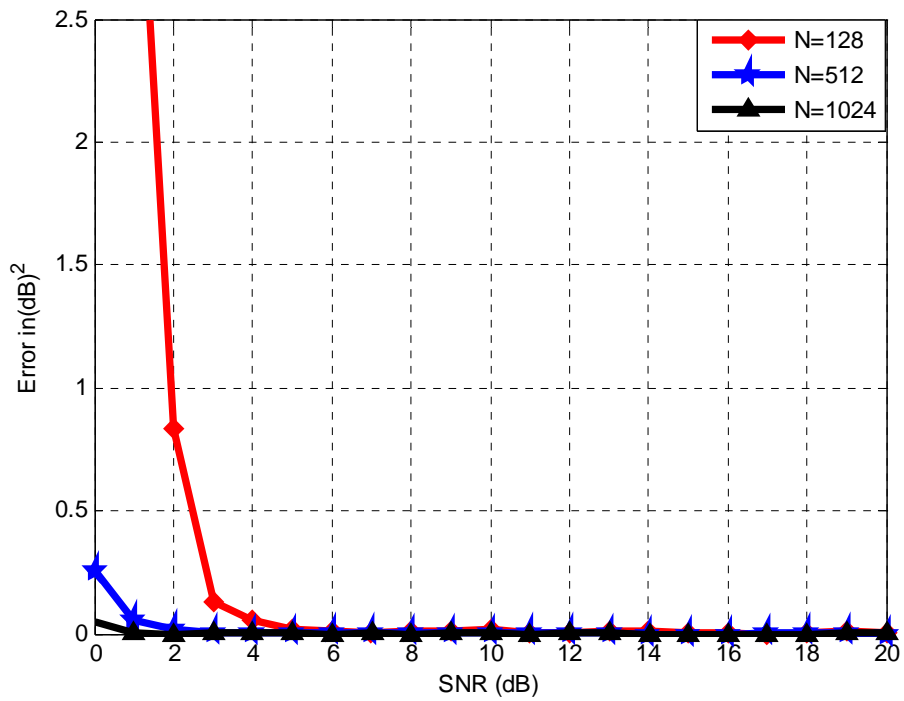


Figure 4.4 The Error in dB<sup>2</sup> vs. SNR in dB for different Ns.



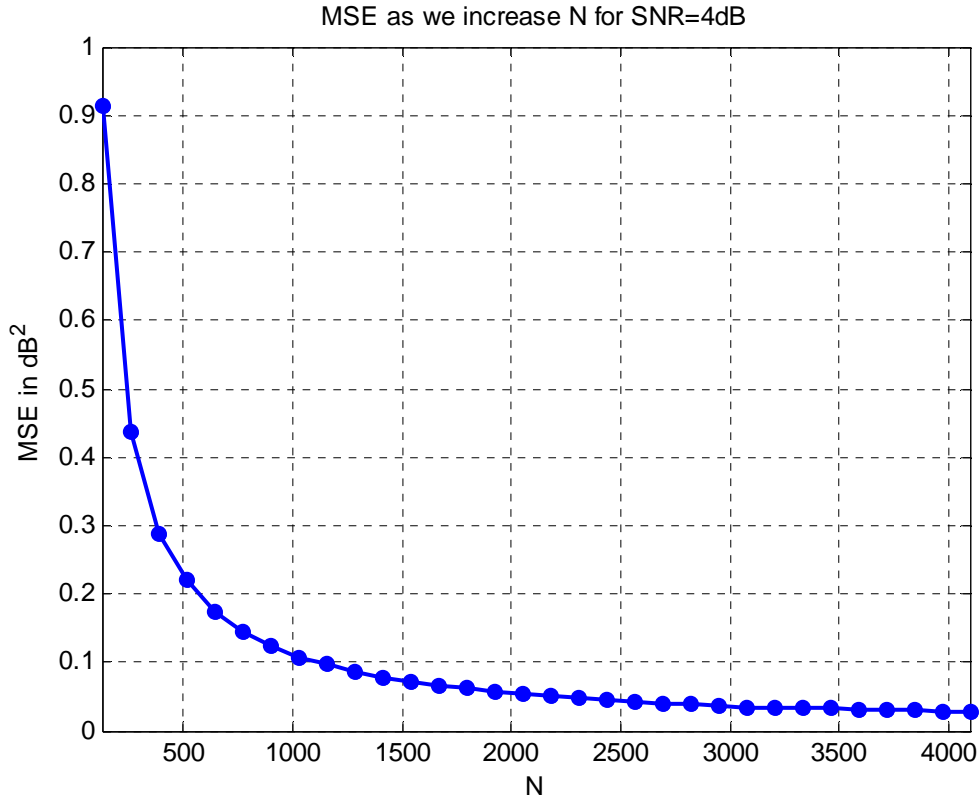


Figure 4.5 The MSE in dB<sup>2</sup> vs. N.

The last figure 4.5 which shows the changes in the MSE as we increase N. this is easily could be inferred from the previous figures, but we want it to be clear that as we increase N, MSE will decrease as a result. We vary N from 128 to 4096 and see the resultant MSE when SNR has been chosen constant in this case to be 4 dB. The MSE decreased from 0.45 dB<sup>2</sup> when N was 128 to about 0.025 dB<sup>2</sup> when N is 4096. Our expectation is that as N tends to infinity, MSE will tend to zero as a result

#### 4.2 The QPSK case

In the case of QPSK case, it is assumed that the received samples were drawn independently and identically from a mixture of four circular symmetric complex Gaussian distribution. Our model in this case would be exactly the same except for the transmitted symbols now four instead of two. We are assuming that there is no error in timing, phase deviation, or frequency deviation. The matched filter output may be given as,

$$y_n = a_n h + w_n, \quad n = 1, \dots, N \quad (4.21)$$

where,  $y_n$  is the received sample,  $a_n \in \left\{ \pm \frac{1}{\sqrt{2}} \pm j \frac{1}{\sqrt{2}} \right\}$  is the transmitted symbol,  $h$  is the channel gain, and  $w_n$  is a zero mean AWGN with variance  $\sigma^2$ .

We can write equation (4.21) in a vector form as,

$$\mathbf{y} = \mathbf{a}h + \mathbf{w} \quad (4.22)$$

where,  $\mathbf{y} = [y_1 \ y_2 \ y_3 \ \dots \ y_N]^T$ ,  $\mathbf{a} = [a_1 \ a_2 \ a_3 \ \dots \ a_N]^T$ ,  $\mathbf{w} = [w_1 \ w_2 \ w_3 \ \dots \ w_N]^T$ . The transmitted symbols may be modeled as realizations of *i. i. d.* circular symmetric complex Gaussian distribution.

The probability distribution function may be written as,

$$p(y_n/a_n, \Theta) = \frac{1}{\pi\sigma^2} \exp\left\{-\frac{1}{\sigma^2} |y_n - a_n h|^2\right\} \quad (4.23)$$

Assuming equally likely symbols,

$$\begin{aligned} p(y_n; \Theta) &= \sum_{a_n} p(y_n/a_n, \Theta) p(a_n/\Theta) \\ &= \frac{1}{\pi\sigma^2} \left\{ \frac{1}{4} \exp\left\{-\frac{1}{\sigma^2} \left|y_n - h\left(\frac{1}{\sqrt{2}} + j\frac{1}{\sqrt{2}}\right)\right|^2\right\} + \frac{1}{4} \exp\left\{-\frac{1}{\sigma^2} \left|y_n - h\left(\frac{1}{\sqrt{2}} - j\frac{1}{\sqrt{2}}\right)\right|^2\right\} \right. \\ &\quad \left. + \frac{1}{4} \exp\left\{-\frac{1}{\sigma^2} \left|y_n - h\left(-\frac{1}{\sqrt{2}} + j\frac{1}{\sqrt{2}}\right)\right|^2\right\} + \frac{1}{4} \exp\left\{-\frac{1}{\sigma^2} \left|y_n - h\left(-\frac{1}{\sqrt{2}} - j\frac{1}{\sqrt{2}}\right)\right|^2\right\} \right\} \\ &= \frac{1}{\pi\sigma^2} \frac{1}{4} \left\{ \exp\left\{-\frac{1}{\sigma^2} \left|y_n - \frac{h}{\sqrt{2}}(1+j)\right|^2\right\} + \exp\left\{-\frac{1}{\sigma^2} \left|y_n - \frac{h}{\sqrt{2}}(1-j)\right|^2\right\} \right. \\ &\quad \left. + \exp\left\{-\frac{1}{\sigma^2} \left|y_n - \frac{h}{\sqrt{2}}(-1+j)\right|^2\right\} + \exp\left\{-\frac{1}{\sigma^2} \left|y_n - \frac{h}{\sqrt{2}}(-1-j)\right|^2\right\} \right\} \end{aligned}$$

$$|y_n - a_n h|^2 = (y_n - a_n h)^* (y_n - a_n h)$$

$$= |y_n|^2 + h^2 |a_n|^2 - 2\Re\{y_n a_n^* h\}$$

$$p(y_n; \Theta) = \frac{1}{\pi\sigma^2} \frac{1}{4} \left\{ \exp \left\{ -\frac{1}{\sigma^2} \left| y_n - h \exp(j\frac{\pi}{4}) \right|^2 \right\} + \exp \left\{ -\frac{1}{\sigma^2} \left| y_n - h \exp(-j\frac{\pi}{4}) \right|^2 \right\} \right. \\ \left. + \exp \left\{ -\frac{1}{\sigma^2} \left| y_n + h \exp(-j\frac{\pi}{4}) \right|^2 \right\} + \exp \left\{ -\frac{1}{\sigma^2} \left| y_n + h \exp(j\frac{\pi}{4}) \right|^2 \right\} \right\}$$

$$p(y_n; \Theta) = \frac{1}{\pi\sigma^2} \frac{1}{4} \exp \left\{ -\frac{|y_n|^2}{\sigma^2} \right\} \exp \left\{ -\frac{h^2}{\sigma^2} \right\} \left\{ \exp \left\{ \frac{2}{\sigma^2} \Re \left\{ y_n h \exp(-j\frac{\pi}{4}) \right\} \right\} \right. \\ \left. + \exp \left\{ \frac{2}{\sigma^2} \Re \left\{ y_n h \exp(j\frac{\pi}{4}) \right\} \right\} + \exp \left\{ -\frac{2}{\sigma^2} \Re \left\{ y_n h \exp(j\frac{\pi}{4}) \right\} \right\} \right. \\ \left. + \exp \left\{ -\frac{2}{\sigma^2} \Re \left\{ y_n h \exp(-j\frac{\pi}{4}) \right\} \right\} \right\}$$

$$p(y_n; \Theta) = \frac{1}{\pi\sigma^2} \frac{1}{4} \exp \left\{ -\frac{|y_n|^2}{\sigma^2} \right\} \exp \left\{ -\frac{h^2}{\sigma^2} \right\} \left\{ 2 \cosh \left\{ \frac{2}{\sigma^2} \Re \left\{ y_n h \exp(-j\frac{\pi}{4}) \right\} \right\} \right. \\ \left. + 2 \cosh \left\{ \frac{2}{\sigma^2} \Re \left\{ y_n h \exp(j\frac{\pi}{4}) \right\} \right\} \right\}$$

$$p(y_n; \Theta) = \frac{1}{\pi\sigma^2} \frac{1}{2} \exp \left\{ -\frac{|y_n|^2}{\sigma^2} \right\} \exp \left\{ -\frac{h^2}{\sigma^2} \right\} \left\{ \cosh \left\{ \frac{2}{\sigma^2} \Re \left\{ y_n h \exp(-j\frac{\pi}{4}) \right\} \right\} \right. \\ \left. + \cosh \left\{ \frac{2}{\sigma^2} \Re \left\{ y_n h \exp(j\frac{\pi}{4}) \right\} \right\} \right\}$$

$$p(y_n; \Theta) = \frac{1}{\pi\sigma^2} \exp \left\{ -\frac{|y_n|^2}{\sigma^2} \right\} \exp \left\{ -\frac{h^2}{\sigma^2} \right\} \cosh \left\{ \frac{\sqrt{2}h}{\sigma^2} \Re \{ y_n \} \right\} \cosh \left\{ \frac{\sqrt{2}h}{\sigma^2} \Im \{ y_n \} \right\}$$

The probability distribution function of  $N$  i. i. d. received samples of  $\mathbf{y}$  will be as follows,

$$p(\mathbf{y}; \Theta) = \prod_{n=1}^N \frac{1}{\pi\sigma^2} \exp \left\{ -\frac{|y_n|^2 + h^2}{\sigma^2} \right\} \cosh \left\{ \frac{\sqrt{2}h}{\sigma^2} \Re \{ y_n \} \right\} \cosh \left\{ \frac{\sqrt{2}h}{\sigma^2} \Im \{ y_n \} \right\} \quad (4.24)$$

The Log Likelihood function is then defined as follows,

$$\mathcal{L}(\Theta) = \sum_{n=1}^N \log p(y_n/\Theta)$$

$$\begin{aligned}
\mathcal{L}(\theta) &= \sum_{n=1}^N \left\{ \log \left\{ \frac{1}{\pi\sigma^2} \exp \left\{ -\frac{|y_n|^2}{\sigma^2} \right\} \exp \left\{ -\frac{h^2}{\sigma^2} \right\} \cosh \left\{ \frac{\sqrt{2}h}{\sigma^2} \Re\{y_n\} \right\} \cosh \left\{ \frac{\sqrt{2}h}{\sigma^2} \Im\{y_n\} \right\} \right\} \right\} \\
&= \sum_{n=1}^N \left\{ \log \left\{ \frac{1}{\pi\sigma^2} \right\} - \frac{|y_n|^2}{\sigma^2} - \frac{h^2}{\sigma^2} + \log \cosh \left\{ \frac{\sqrt{2}h}{\sigma^2} \Re\{y_n\} \right\} + \log \cosh \left\{ \frac{\sqrt{2}h}{\sigma^2} \Im\{y_n\} \right\} \right\} \quad (4.25)
\end{aligned}$$

This is the Log Likelihood function of the received data of  $\mathbf{y}$ , now the problem of estimation would be as follows, given the observation vector,  $\mathbf{y}$ , estimate the parameters,  $\hat{\theta} = [\hat{h} \ \widehat{\sigma^2}]^T$ . Then use them to find the SNR estimate,

$$\widehat{SNR} = 10 \log_{10} \left( \frac{\hat{h}^2}{\widehat{\sigma^2}} \right)$$

#### 4.2.1 The CRLB for QPSK

The Cramer-Rao Lower Bound (CRLB) is a lower bound for the Mean Square Error (MSE) of any unbiased estimator which satisfies the regularity condition [2]. In our case of SNR estimation, the CRLB can be given as,

$$CRLB(SNR) = \frac{\partial SNR^T}{\partial \theta} I^{-1}(\theta) \frac{\partial SNR}{\partial \theta}$$

Where  $I(\theta)$  is the fisher information matrix given by,

$$I(\theta) = \begin{bmatrix} -\mathbb{E} \left[ \frac{\partial^2 \log p(Y; \theta)}{\partial h^2} \right] & -\mathbb{E} \left[ \frac{\partial^2 \log p(Y; \theta)}{\partial h \partial \sigma^2} \right] \\ -\mathbb{E} \left[ \frac{\partial^2 \log p(Y; \theta)}{\partial \sigma^2 \partial h} \right] & -\mathbb{E} \left[ \frac{\partial^2 \log p(Y; \theta)}{\partial \sigma^2^2} \right] \end{bmatrix}$$

And,

$$\frac{\partial SNR}{\partial \theta} = \left[ \frac{20}{h \log(10)} \quad -\frac{10}{\sigma^2 \log(10)} \right]^T$$

**For the NDA case,**

The log likelihood function is given by,

$$\mathcal{L}(\Theta) = \sum_{n=1}^N \left\{ \log \left\{ \frac{1}{\pi\sigma^2} \right\} - \frac{|y_n|^2}{\sigma^2} - \frac{h^2}{\sigma^2} + \log \cosh \left\{ \frac{\sqrt{2}h}{\sigma^2} \Re\{y_n\} \right\} + \log \cosh \left\{ \frac{\sqrt{2}h}{\sigma^2} \Im\{y_n\} \right\} \right\}$$

This procedure is similar to the one which being done by Alagha [8],

$$I(\Theta) = \begin{bmatrix} \frac{2N(1-2f(\rho))}{\sigma^2} & \frac{4Nh}{\sigma^4} f(\rho) \\ \frac{4Nh}{\sigma^4} f(\rho) & \frac{N}{\sigma^4} (1-4\rho f(\rho)) \end{bmatrix} \quad (4.26)$$

where,

$$f(\rho) = \frac{\exp(-\frac{\rho}{2})}{\sqrt{\pi}} \int_{-\infty}^{+\infty} \frac{x^2 \exp(-x^2)}{\cosh(x\sqrt{2\rho})} dx \quad (4.27)$$

We can now substitute this result in equation (4.5) and get the following,

$$CRLB(SNR) = \begin{bmatrix} \frac{20}{h \log(10)} & -\frac{10}{\sigma^2 \log(10)} \end{bmatrix} \begin{bmatrix} \frac{2N(1-2f(\rho))}{\sigma^2} & \frac{4Nh}{\sigma^4} f(\rho) \\ \frac{4Nh}{\sigma^4} f(\rho) & \frac{N}{\sigma^4} (1-4\rho f(\rho)) \end{bmatrix} \begin{bmatrix} \frac{20}{h \log(10)} \\ -\frac{10}{\sigma^2 \log(10)} \end{bmatrix}$$

$$CRLB(SNR) = \frac{100}{N(\log(10))^2} \frac{1 + \frac{2}{\rho} - 2f(\rho)}{1 - 2f(\rho) - 4\rho f(\rho)} \quad dB^2 \quad (4.28)$$

**For the DA case,**

The log likelihood function is given by,

$$\begin{aligned} \mathcal{L}(\Theta) &= \sum_{n=1}^N \log p(y_n/a_n, \Theta) \\ \mathcal{L}(\Theta) &= \sum_{n=1}^N \log \left\{ \frac{1}{\pi\sigma^2} \exp \left\{ -\frac{|y_n - a_n h|^2}{\sigma^2} \right\} \right\} \\ &= \sum_{n=1}^N -\log(\pi\sigma^2) - \frac{|y_n - a_n h|^2}{\sigma^2} \end{aligned}$$

Then substituting to find the fisher information matrix,

$$I(\theta) = \begin{bmatrix} \frac{2N}{\sigma^2} & 0 \\ 0 & \frac{N}{\sigma^4} \end{bmatrix}$$

$$CRLB(SNR) = \begin{bmatrix} \frac{20}{h \log(10)} & -\frac{10}{\sigma^2 \log(10)} \\ 0 & \frac{N}{\sigma^4} \end{bmatrix} \begin{bmatrix} \frac{2N}{\sigma^2} & 0 \\ 0 & \frac{N}{\sigma^4} \end{bmatrix} \begin{bmatrix} \frac{20}{h \log(10)} \\ \frac{10}{\sigma^2 \log(10)} \end{bmatrix}$$

$$CRLB(SNR) = \frac{100}{N(\log(10))^2} \left( \frac{2}{\rho} + 1 \right) \quad dB^2 \quad (4.29)$$

#### 4.2.2 NDA SNR estimation using the EM algorithm

We use the EM algorithm now in estimating the SNR, its MSE and compare our estimated values with the CRLB values. Our assumptions are that the data was drawn from a mixture of four Gaussians, and are uncorrelated with each other. The samples are drawn in an *i. i. d.* manner. The parameters which we need to estimate now are,

$$\begin{aligned} \hat{\Theta} &= [\hat{h} \quad \hat{\sigma}^2]^T \\ &= \underset{\theta}{\operatorname{argmax}} p(\mathbf{y}; \theta) \end{aligned}$$

$$\hat{\rho} = \frac{\hat{h}^2}{\hat{\sigma}^2}$$

$$\widehat{SNR} = 10 \log_{10}(\hat{\rho}) \quad (4.30)$$

The EM algorithm will be used to iteratively estimate the channel gain and the noise variance then use them to estimate the signal to noise ratio. The received data will be seen as incomplete data set since we don't know the class of each transmitted symbol. The auxiliary function in this case may be given by,

$$Q(\theta^k, \theta^{k+1}) = \mathbb{E}_{p(Z/Y, \theta^k)} \{ \log(p(\mathbf{y}, Z/\theta^{k+1})) | \mathbf{y}, \theta^k \}$$

$$\begin{aligned}
Q(\theta^k, \theta^{k+1}) &= \sum_{n=1}^N \sum_{m=1}^M p(\omega_m/y_n, \theta^k) \log p(y_n/\omega_m, \theta^{k+1}) \\
&\quad + \sum_{n=1}^N \sum_{m=1}^M p(\omega_m/y_n, \theta^k) \log p^{k+1}(\omega_m) \\
Q(\theta^k, \theta^{k+1}) &= \sum_{m=1}^M \sum_{n=1}^N p(\omega_m/y_n, \theta^k) \left\{ -\frac{1}{2} \{ (y_n - \mu_m^{k+1})^T \Sigma_m^{k+1-1} (y_n - \mu_m^{k+1}) \} \right\} \\
&\quad + \sum_{m=1}^M \sum_{n=1}^N p(\omega_m/y_n, \theta^k) \left\{ -\frac{1}{2} \{ \log \{ (2\pi)^D |\Sigma_m^{k+1}| \} \} \right\} \\
&\quad + \sum_{m=1}^M \sum_{n=1}^N p(\omega_m/y_n, \theta^k) \log \{ C \}
\end{aligned}$$

$$\log p(y_n/\omega_m, \theta^{k+1}) = -\log \{ \pi \sigma^{2k+1} \} - \frac{|y_n - a_m h^{k+1}|^2}{\sigma^{2k+1}}$$

$$Q(\theta^k, \theta^{k+1}) = \sum_{m=1}^M \sum_{n=1}^N p(\omega_m/y_n, \theta^k) \left\{ -\log \{ \pi \sigma^{2k+1} \} - \frac{|y_n - a_m h^{k+1}|^2}{\sigma^{2k+1}} \right\}$$

$$\theta^{k+1} = \underset{\theta}{\operatorname{argmax}} Q(\theta^k, \theta^{k+1}) \quad (4.31)$$

Now solving those two equations will give us the E & M steps.

When solving for the channel gain we get,

$$\frac{\partial Q(\theta^k, \theta^{k+1})}{\partial h^{k+1}} = \frac{\partial}{\partial h^{k+1}} \left\{ \sum_{m=1}^M \sum_{n=1}^N p(\omega_m/y_n, \theta^k) \left\{ -\log \{ \pi \sigma^{2k+1} \} - \frac{|y_n - a_m h^{k+1}|^2}{\sigma^{2k+1}} \right\} \right\} = 0$$

$$\frac{\partial Q(\theta^k, \theta^{k+1})}{\partial h^{k+1}} = \frac{\partial}{\partial h^{k+1}} \left\{ \sum_{m=1}^M \sum_{n=1}^N p(\omega_m/y_n, \theta^k) \left\{ -\frac{|y_n - a_m h^{k+1}|^2}{\sigma^{2k+1}} \right\} \right\} = 0$$

$$\sum_{m=1}^M \sum_{n=1}^N p(\omega_m/y_n, \theta^k) \{ 2h^{k+1} - 2\Re\{a_m^* y_n\} \} = 0$$

$$\sum_{m=1}^M \sum_{n=1}^N p(\omega_m/y_n, \theta^k) h^{k+1} = \sum_{m=1}^M \sum_{n=1}^N p(\omega_m/y_n, \theta^k) \Re\{a_m^* y_n\}$$

$$\begin{aligned}
Nh^{k+1} &= \sum_{m=1}^4 \sum_{n=1}^N p(\omega_m/y_n, \theta^k) \Re\{\exp[-i(2m-1)\pi/4]y_n\} \\
Nh^{k+1} &= \sum_{n=1}^N \left\{ p(\omega_1/y_n, \theta^k) \Re\{\exp[-i\pi/4]y_n\} + p(\omega_2/y_n, \theta^k) \Re\{\exp[-i3\pi/4]y_n\} \right. \\
&\quad \left. + p(\omega_3/y_n, \theta^k) \Re\{\exp[-i5\pi/4]y_n\} + p(\omega_4/y_n, \theta^k) \Re\{\exp[-i7\pi/4]y_n\} \right\} \\
p(\omega_4/y_n, \theta^k) &= 1 - \{p(\omega_1/y_n, \theta^k) + p(\omega_2/y_n, \theta^k) + p(\omega_3/y_n, \theta^k)\} \\
Nh^{k+1} &= \sum_{n=1}^N \left\{ p(\omega_1/y_n, \theta^k) \Re\{\exp[-i\pi/4]y_n\} + p(\omega_2/y_n, \theta^k) \Re\{\exp[-i3\pi/4]y_n\} \right. \\
&\quad \left. + p(\omega_3/y_n, \theta^k) \Re\{\exp[-i5\pi/4]y_n\} \right. \\
&\quad \left. + \left\{ 1 - \{p(\omega_1/y_n, \theta^k) + p(\omega_2/y_n, \theta^k) + p(\omega_3/y_n, \theta^k)\} \right\} \Re\{\exp[-i7\pi/4]y_n\} \right\} \\
Nh^{k+1} &= \sum_{n=1}^N \left\{ p(\omega_1/y_n, \theta^k) \{ \Re\{\exp[-i\pi/4]y_n\} - \Re\{\exp[-i7\pi/4]y_n\} \} \right. \\
&\quad \left. + p(\omega_2/y_n, \theta^k) \{ \Re\{\exp[-i3\pi/4]y_n\} - \Re\{\exp[-i7\pi/4]y_n\} \} \right. \\
&\quad \left. + p(\omega_3/y_n, \theta^k) \{ \Re\{\exp[-i5\pi/4]y_n\} - \Re\{\exp[-i7\pi/4]y_n\} \} + \Re\{\exp[-i7\pi/4]y_n\} \right\} \\
h^{k+1} &= -\frac{\sqrt{2}}{N} \sum_{n=1}^N \left\{ p(\omega_1/y_n, \theta^k) \Re\{iy_n\} + p(\omega_2/y_n, \theta^k) \Re\{(1+i)y_n\} + p(\omega_3/y_n, \theta^k) \Re\{y_n\} \right. \\
&\quad \left. - \frac{1}{2} \Re\{[1+i]y_n\} \right\} \\
h^{k+1} &= -\frac{\sqrt{2}}{N} \sum_{n=1}^N \left\{ \left\{ p(\omega_1/y_n, \theta^k) + p(\omega_2/y_n, \theta^k) - \frac{1}{2} \right\} \Re\{iy_n\} \right. \\
&\quad \left. + \left\{ p(\omega_2/y_n, \theta^k) + p(\omega_3/y_n, \theta^k) - \frac{1}{2} \right\} \Re\{y_n\} \right\} \tag{4.32}
\end{aligned}$$

When solving for the variance we get,

$$\frac{\partial Q(\theta^k, \theta^{k+1})}{\partial \sigma^{2k+1}} = \frac{\partial}{\partial \sigma^{2k+1}} \left\{ \sum_{m=1}^M \sum_{n=1}^N p(\omega_m/y_n, \theta^k) \left\{ -\log\{\pi\sigma^{2k+1}\} - \frac{|y_n - a_m h^{k+1}|^2}{\sigma^{2k+1}} \right\} \right\} = 0$$



$$\sum_{m=1}^M \sum_{n=1}^N p(\omega_m/y_n, \Theta^k) \left\{ -\frac{1}{\sigma^{2^{k+1}}} + \frac{|y_n - a_m h^{k+1}|^2}{\sigma^{4^{k+1}}} \right\} = 0$$

$$\sum_{m=1}^M \sum_{n=1}^N p(\omega_m/y_n, \Theta^k) \{|y_n - a_m h^{k+1}|^2\} = N\sigma^{2^{k+1}}$$

$$\begin{aligned} N\sigma^{2^{k+1}} &= \sum_{n=1}^N \left\{ p(\omega_1/y_n, \Theta^k) \{|y_n - a_1 h^{k+1}|^2\} + p(\omega_2/y_n, \Theta^k) \{|y_n - a_2 h^{k+1}|^2\} \right. \\ &\quad + p(\omega_3/y_n, \Theta^k) \{|y_n - a_3 h^{k+1}|^2\} \\ &\quad \left. + \{1 - \{p(\omega_1/y_n, \Theta^k) + p(\omega_2/y_n, \Theta^k) + p(\omega_3/y_n, \Theta^k)\}\} \{|y_n - a_4 h^{k+1}|^2\} \right\} \end{aligned}$$

$$\begin{aligned} N\sigma^{2^{k+1}} &= \sum_{n=1}^N \left\{ p(\omega_1/y_n, \Theta^k) \{-2\Re\{a_1^* y_n h^{k+1}\} + 2\Re\{a_4^* y_n h^{k+1}\}\} \right. \\ &\quad + p(\omega_2/y_n, \Theta^k) \{-2\Re\{a_2^* y_n h^{k+1}\} + 2\Re\{a_4^* y_n h^{k+1}\}\} \\ &\quad \left. + p(\omega_3/y_n, \Theta^k) \{-2\Re\{a_3^* y_n h^{k+1}\} + 2\Re\{a_4^* y_n h^{k+1}\}\} + \{|y_n - a_4 h^{k+1}|^2\} \right\} \end{aligned}$$

$$\begin{aligned} N\sigma^{2^{k+1}} &= \sum_{n=1}^N \left\{ -2p(\omega_1/y_n, \Theta^k) h^{k+1} \{\Re\{\exp[-i\pi/4]y_n\} - \Re\{\exp[-i7\pi/4]y_n\}\} \right. \\ &\quad - 2p(\omega_2/y_n, \Theta^k) h^{k+1} \{\Re\{\exp[-i3\pi/4]y_n\} - \Re\{\exp[-i7\pi/4]y_n\}\} \\ &\quad - 2p(\omega_3/y_n, \Theta^k) h^{k+1} \{\Re\{\exp[-i5\pi/4]y_n\} - \Re\{\exp[-i7\pi/4]y_n\}\} \\ &\quad \left. + \{|y_n - a_4 h^{k+1}|^2\} \right\} \end{aligned}$$

$$\begin{aligned} N\sigma^{2^{k+1}} &= \sum_{n=1}^N \left\{ -2p(\omega_1/y_n, \Theta^k) h^{k+1} \Re\{-\sqrt{2}iy_n\} - 2p(\omega_2/y_n, \Theta^k) h^{k+1} \Re\{-\sqrt{2}(1+i)y_n\} \right. \\ &\quad \left. - 2p(\omega_3/y_n, \Theta^k) h^{k+1} \Re\{-\sqrt{2}y_n\} + \{|y_n - a_4 h^{k+1}|^2\} \right\} \end{aligned}$$

$$\begin{aligned} \sigma^{2^{k+1}} &= \frac{2\sqrt{2}}{N} \sum_{n=1}^N \left\{ p(\omega_1/y_n, \Theta^k) h^{k+1} \Re\{iy_n\} + p(\omega_2/y_n, \Theta^k) h^{k+1} \Re\{(1+i)y_n\} \right. \\ &\quad \left. + p(\omega_3/y_n, \Theta^k) h^{k+1} \Re\{y_n\} + \frac{1}{2\sqrt{2}} \{|y_n - \exp[i7\pi/4]h^{k+1}|^2\} \right\} \end{aligned}$$

$$\sigma^{2^{k+1}} = \frac{2\sqrt{2}}{N} \sum_{n=1}^N \left\{ \left\{ p(\omega_1/y_n, \theta^k) + p(\omega_2/y_n, \theta^k) - \frac{1}{2} \right\} h^{k+1} \Re\{iy_n\} \right. \\ \left. + \left\{ p(\omega_2/y_n, \theta^k) + p(\omega_3/y_n, \theta^k) - \frac{1}{2} \right\} h^{k+1} \Re\{y_n\} + \frac{1}{2\sqrt{2}} |y_n|^2 + \frac{1}{2\sqrt{2}} h^{k+1} \right\}^2 \\ \sigma^{2^{k+1}} = \frac{1}{N} \sum_{n=1}^N |y_n|^2 - h^{k+1}^2 \quad (4.33)$$

We don't need to worry about the mixing coefficient since we know that they are equally likely. After Starting with initial values for the channel gain and noise variance, equations (4.27) and (4.28) will be used to iteratively find the estimates to the parameters which would be then used to estimate the  $\widehat{SNR}$  as given above. The *MSE* will be found for different estimated values of different SNR values and will be assessed using the CRLB.

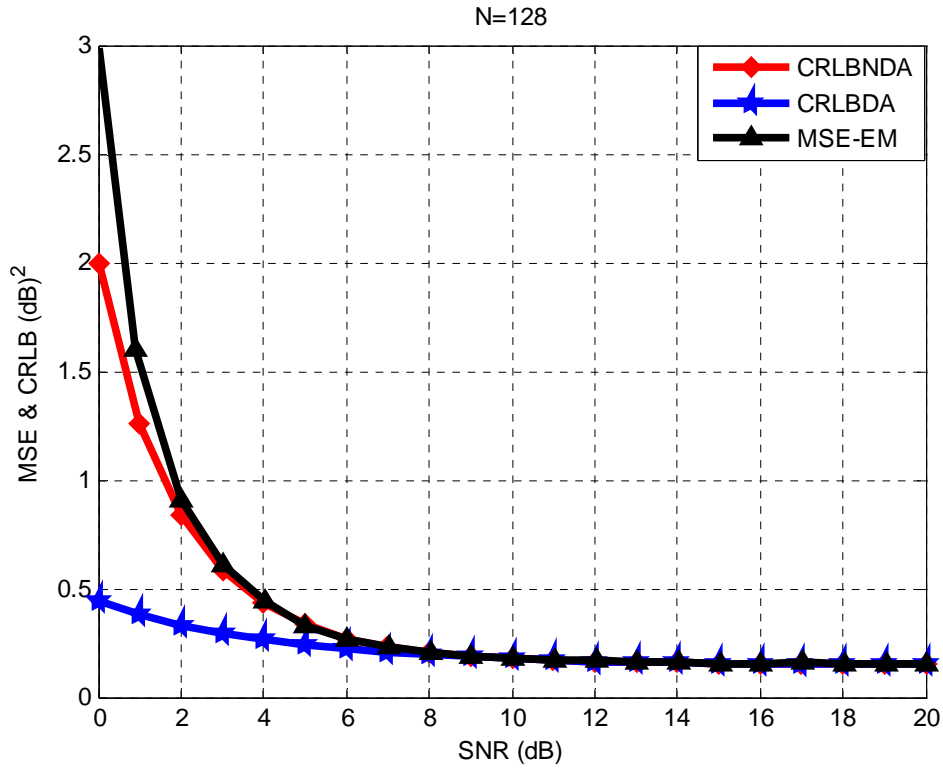


Figure 4.6 The MSE & CRLB in dB<sup>2</sup> vs. SNR in dB when N=128

Then this MSE performance will be compared to the fundamental limit of the CRLB in units of dB<sup>2</sup> at each estimated SNR value. The estimation process will be done for a range of SNR from 0 dB to 20

dB, and for different values of received sample size  $N$ , 128, 512, and 1024. Figure 4.6 shown explains three things when QPSK is used as the transmitted symbols. The red curve is the CRLB for NDA signal, the blue curve is the CRLB for DA signal, and the black curve which shows the MSE estimated using the EM algorithm. The EM algorithm SNR estimate is repeated 10000 times using Monte Carlo simulations to get 10000 realizations of the SNR estimates, then averaging them to get the estimated SNR, and find their variance to get the MSE for that specific SNR estimate value.

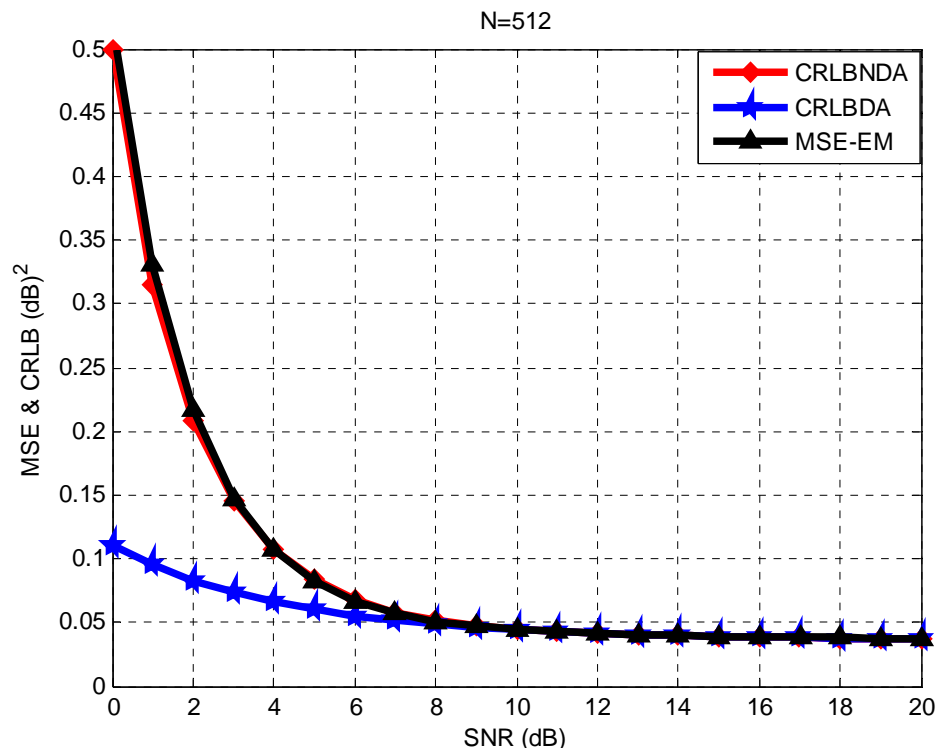


Figure 4.7 The MSE & CRLB in  $\text{dB}^2$  vs. SNR in dB when  $N=512$

We then repeat this process for different values of SNR, say from 0 dB to 20 dB and their resultant MSE. Those estimates then compared with the CRLB as shown above. We see that for low SNR estimate values; the MSE is large compared with the NDA CRLB. The higher the SNR become, the lower the MSE we get till they (the CRLB and MSE) both match. This was for sample size of  $N=128$  in figure 4.6 in the previous page. The next step we would do is increase  $N$  and see what is going to happen to our estimation accuracy. Figure 4.7 shows the relation between MSE and CRLB vs. SNR for  $N=512$ . What we realize here is that the MSE is decreased, also CRLB is decreased from what it was before when  $N=128$ . CRLB itself decreased from 2 when  $N$  was 128 and SNR of 0dB, to less than 0.5 when  $N$  is 512 at the same SNR.

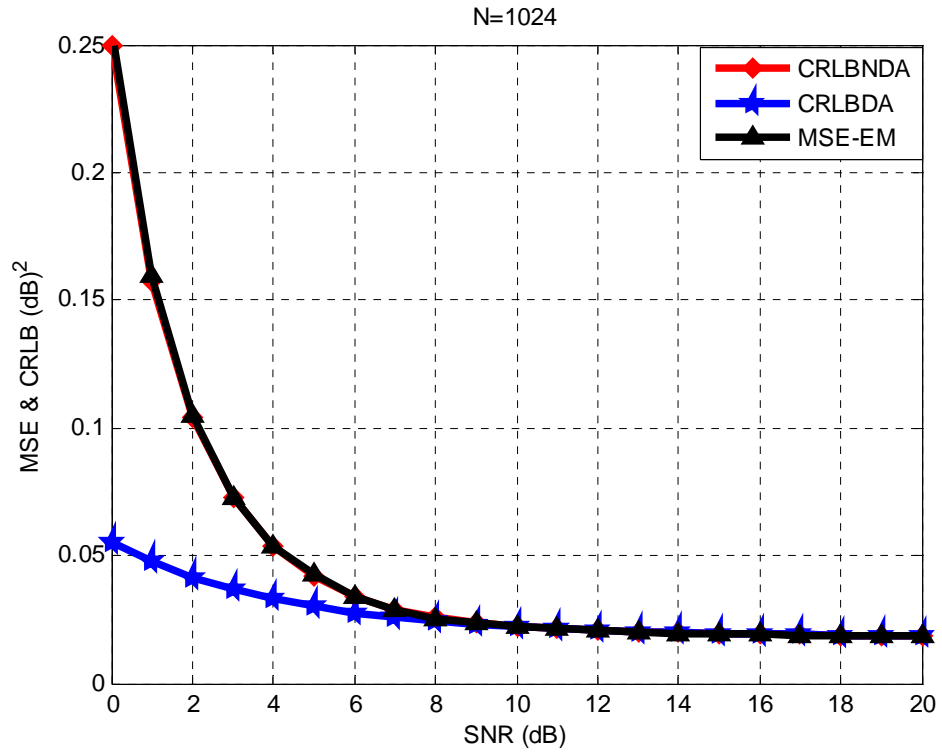


Figure 4.8. The MSE & CRLB in  $\text{dB}^2$  vs. SNR in dB when  $N=1024$

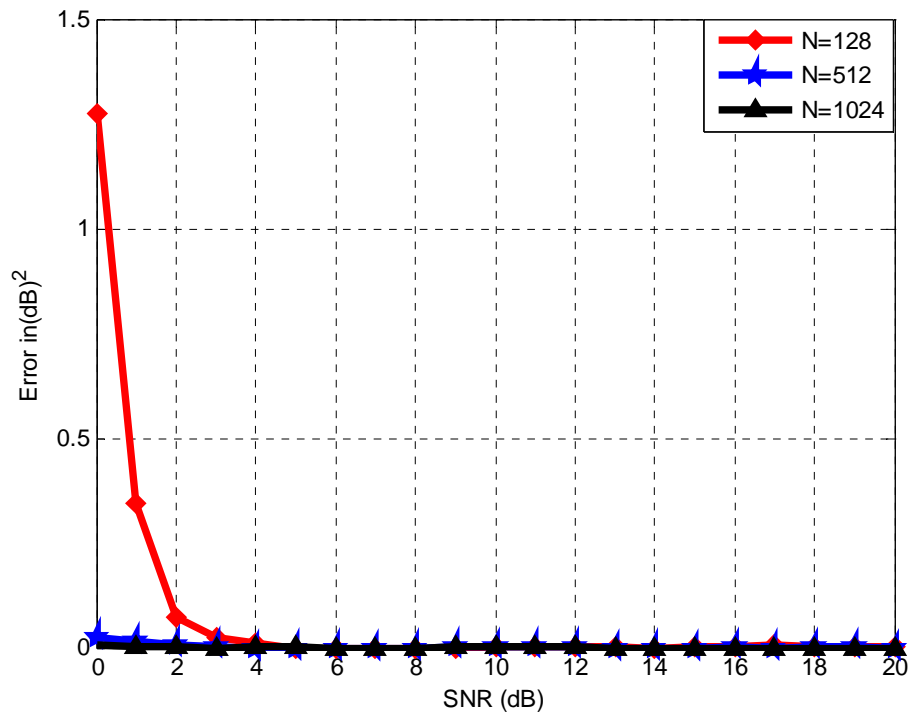


Figure 4.9. The Error in  $\text{dB}^2$  vs. SNR in dB for different  $N_s$ .

Figure 4.8 shows the same thing Figure 4.7 shows, but for  $N=1024$ . The precision of the MSE and CRLB is in the order of less than  $0.025 \text{ dB}^2$  for SNRs greater than 6 dB. Figure 4.9 shows the error which is the difference between the MSE and NDA CRLB in  $\text{dB}^2$  at different values of SNRs shown for the previous three cases of different  $N_s$ , 128, 512, and 1024. This figure proves that as  $N$  increases, the error decreases which means that the MSE become closer and closer to the NDA CRLB until they both match at some SNR value and beyond. From the SNR equation we see that the SNR is inversely proportional to the noise variance, which means as we increase the noise variance, the SNR will decrease as a result of that and the four mixtures then will intervene with each other which makes it difficult for us to classify each class. On the other hand however, as we decrease the noise variance, the SNR will increase as a result of that and the four mixtures then will be far away from each other (classes are separated from each other) which makes classification become easier and results in almost no error.

## 5. MISO with STBC NDA SNR estimation using the Expectation Maximization algorithm

Modern communication systems require accurate estimation of the Signal to Noise Ratio SNR for optimal usage of radio resources [5], [9]. The knowledge of the SNR is a requirement in most applications in order to make optimal signal detection, power control. MIMO systems gives a significant enhancement to the accuracy of estimating the SNR besides many other improvements in data rate, channel capacity, and being able to reduce Multipath fading at the remote units. Different diversity modes are being used, time diversity, frequency diversity, spatial diversity. A MIMO system usually consists of M transmit and N receive antennas. Attention will be given to the transmit diversity for some reasons, which then will be compared with SISO and SIMO. Transmit diversity or MISO would improve our system's performance without the cost, size, and power that SIMO will have in order to improve the quality of the received signal [12]. For that reason, antenna diversity techniques are utilized at the Base station. A simple transmit diversity technique will be explained here which been done by Alamouti in [12]. Space-time codes are used to generate a redundant signal. The signal copy is transmitted not only from a different antenna, but at a different time. The signal  $s_1$  and  $s_2$  are sent in the first symbol, then a signal replication of  $-s_2^*$  and  $s_1^*$  is added to create the Alamouti space-time block code.

### 5.1 Conventional Receive diversity (SIMO) model:

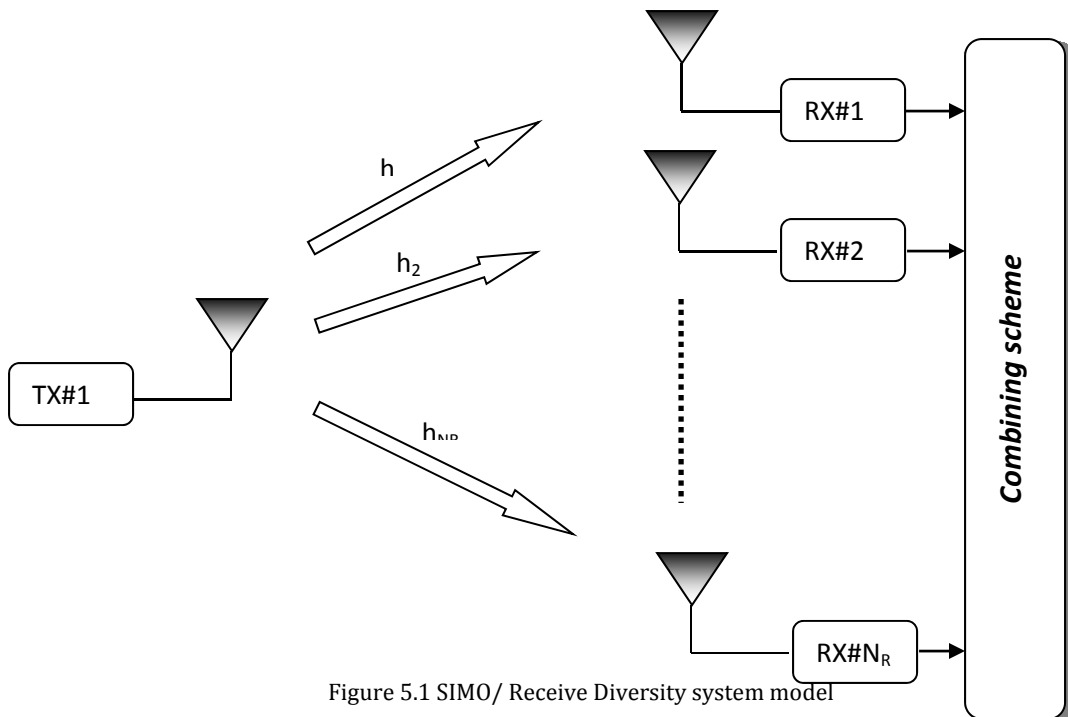


Figure 5.1 SIMO/ Receive Diversity system model

$$y_i(n) = h_i s(n) + w_i(n)$$

where  $i=1, 2, \dots, N_R$ . and  $n=1, 2, \dots, N$ .

$$p(y_i(n)/s(n), \Theta) = \frac{1}{\pi^2 \sigma^4} \exp \left\{ -\frac{\|y_i(n) - h_i s(n)\|^2}{\sigma^2} \right\}$$

$$p(Y/s, \Theta) = \prod_{n=1}^N \prod_{i=1}^{N_R} \frac{1}{\pi^2 \sigma^4} \exp \left\{ -\frac{\|y_i(n) - h_i s(n)\|^2}{\sigma^2} \right\}$$

## 5.2 Transmit diversity (MISO) model:

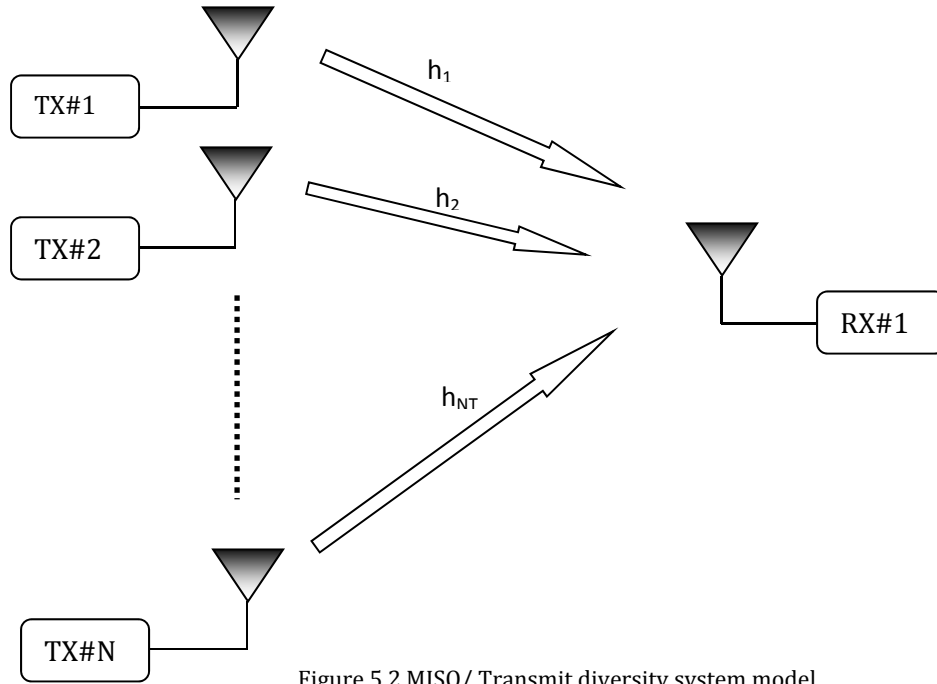


Figure 5.2 MISO/ Transmit diversity system model

$$y(n) = \sum_{i=1}^{N_T} h_i s(n) + w(n)$$

where  $i=1, 2, \dots, N_T$ . and  $n=1, 2, \dots, N$ .

$$p(y(n)/s(n), \Theta) = \frac{1}{\pi^2 \sigma^4} \exp \left\{ -\frac{\|y(n) - \sum_{i=1}^{N_T} h_i s(n)\|^2}{\sigma^2} \right\}$$

$$p(y/s, \Theta) = \prod_{n=1}^N \frac{1}{\pi^2 \sigma^4} \exp \left\{ -\frac{\|y(n) - \sum_{i=1}^{N_T} h_i s(n)\|^2}{\sigma^2} \right\}$$

### 5.3 MIMO model:

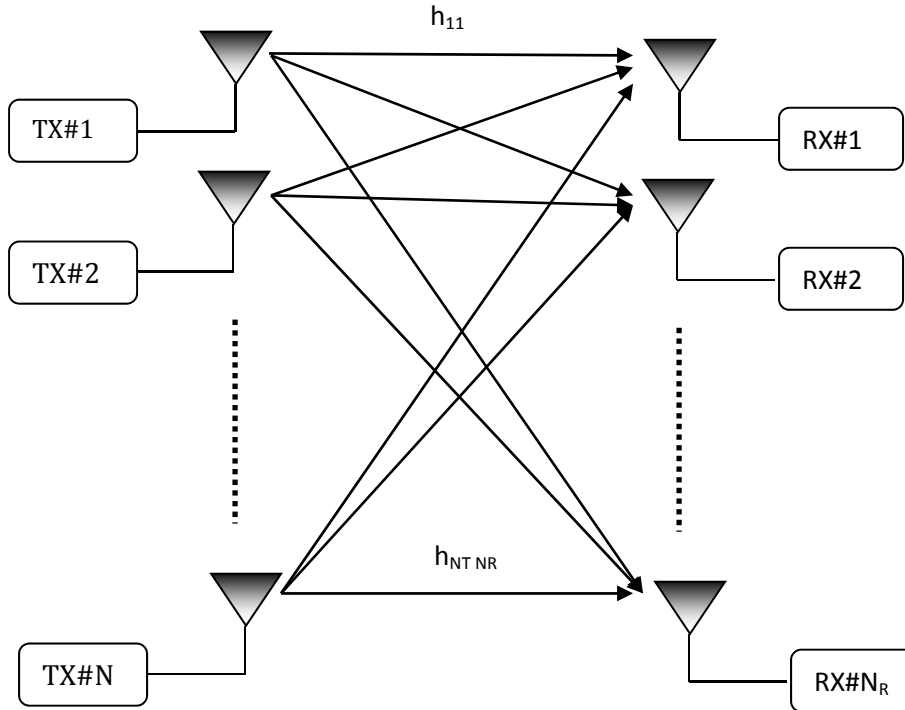


Figure 5.3 MIMO system model

$$y_j(n) = h_{ij}s(n) + w_j(n)$$

where  $i=1, 2, \dots, N_T$ ,  $j=1, 2, \dots, N_R$ , and  $n=1, 2, \dots, N$ , and  $h_{ij}$  is the channel gain between the transmit antenna  $i$  and the receive antenna  $j$ .

$$p(y_j(n)/s(n), \Theta) = \prod_{i=1}^{N_T} \frac{1}{\pi^2 \sigma^4} \exp \left\{ -\frac{\|y_j(n) - h_{ij}s(n)\|^2}{\sigma^2} \right\}$$

$$p(\mathbf{Y}/\mathbf{s}, \Theta) = \prod_{n=1}^N \prod_{j=1}^{N_R} \prod_{i=1}^{N_T} \frac{1}{\pi^2 \sigma^4} \exp \left\{ -\frac{\|y_j(n) - h_{ij}s(n)\|^2}{\sigma^2} \right\}$$



#### 5.4 The transmit diversity for BPSK case:

This approach is being done by Alamouti [12] where two transmit antennas and one receive antenna has been used. At a given time,  $n$ , two signals are simultaneously sent from the two antennas. Denoting the signal sent by antenna one by,  $S_1$ , and the signal sent by antenna two by,  $S_2$ . Then we repeat sending  $-S_2^*$  from the first antenna, then  $S_1^*$  from the second antenna, this ordering will be done to all of our data. This is called space and time coding. Our model in this case will look like,

$$r_{2n-1} = h_1 s_{2n-1} + h_2 s_{2n} + w_{2n-1}$$

$$r_{2n} = -h_1 s_{2n}^* + h_2 s_{2n-1}^* + w_{2n}$$

We then may write it in a matrix form as,

$$\begin{bmatrix} r_{2n-1} \\ r_{2n}^* \end{bmatrix} = \begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix} \begin{bmatrix} s_{2n-1} \\ s_{2n} \end{bmatrix} + \begin{bmatrix} w_{2n-1} \\ w_{2n} \end{bmatrix}, \quad n = 1, 2, \dots, N/2. \quad (5.1)$$

$$R(n) = HS(n) + W(n)$$

$$\mathbf{R} = (\mathbf{I} \otimes H)\mathbf{S} + \mathbf{W}$$

where  $R(n)$  is the received samples from both antennas at time  $n$ ,  $H$  is the complex channels gains matrix,  $S(n)$  are the sent symbols at time  $n$  from the two antennas,  $W(n)$  are the Circularly symmetric complex Gaussian noise samples at time  $n$ ,  $\mathbf{R}$  is the vector of received samples,  $\mathbf{I}$  is  $\frac{N}{2} \times \frac{N}{2}$  Identity matrix where  $N$  is the total number of received samples,  $\mathbf{S}$  is the sent vector of symbols,  $\mathbf{W}$  is the noise vector at the receiver,  $\otimes$  is the kroncker product, and  $H$  is as follows;

$$H = \begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix}$$

The probability distribution function for the received data at a given time  $n$ , given the transmitted data, given by,

$$p(R(n)/S(n), \Theta) = \frac{1}{\pi^2 \sigma^4} \exp \left\{ -\frac{\|R(n) - HS(n)\|^2}{\sigma^2} \right\} \quad (5.2)$$

$$p(R(n); \Theta) = \sum_{m=1}^4 p(R(n)/S(m), \Theta) p(S(m)/\Theta)$$

Suppose an equally likely data symbols.

$$p(R(n); \Theta) = \frac{1}{4\pi^2 \sigma^4} \sum_{m=1}^4 \exp \left\{ -\frac{\|R(n) - HS(m)\|^2}{\sigma^2} \right\}$$

$$p(R(n); \Theta) = \frac{1}{4\pi^2\sigma^4} \sum_{m=1}^4 \exp \left\{ -\frac{|r_{2n-1} - (h_1 s_{2m-1} + h_2 s_{2m})|^2 + |r_{2n} - (h_2 s_{2m-1} - h_1 s_{2m})|^2}{\sigma^2} \right\}$$

For the BPSK case,  $s_{2m-1} = \{-1, +1\}$ ,  $s_{2m} = \{-1, +1\}$ .

where  $\begin{bmatrix} s_{2m-1} \\ s_{2m} \end{bmatrix} \in \left\{ \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ +1 \end{bmatrix}, \begin{bmatrix} +1 \\ -1 \end{bmatrix}, \begin{bmatrix} +1 \\ +1 \end{bmatrix} \right\}$ .

We can then rewrite  $p(R(n); \Theta)$  as follows,

$$\begin{aligned} p(R(n); \Theta) &= \frac{1}{4\pi^2\sigma^4} \left\{ \exp \left\{ -\frac{|r_{2n-1} + (h_1 + h_2)|^2 + |r_{2n} + (h_2 - h_1)|^2}{\sigma^2} \right\} \right. \\ &\quad + \exp \left\{ -\frac{|r_{2n-1} + (h_1 - h_2)|^2 + |r_{2n} + (h_2 + h_1)|^2}{\sigma^2} \right\} \\ &\quad + \exp \left\{ -\frac{|r_{2n-1} - (h_1 - h_2)|^2 + |r_{2n} - (h_2 + h_1)|^2}{\sigma^2} \right\} \\ &\quad \left. + \exp \left\{ -\frac{|r_{2n-1} - (h_1 + h_2)|^2 + |r_{2n} - (h_2 - h_1)|^2}{\sigma^2} \right\} \right\} \end{aligned}$$

$p(R(n); \Theta)$

$$\begin{aligned} &= \frac{1}{4\pi^2\sigma^4} \left\{ \exp \left\{ -\frac{|r_{2n-1}|^2 + |r_{2n}|^2 + 2\Re(r_{2n-1}(h_1 + h_2)^*) + 2\Re(r_{2n}^*(h_2 - h_1)) + |h_1 + h_2|^2 + |h_2 - h_1|^2}{\sigma^2} \right\} \right. \\ &\quad + \exp \left\{ -\frac{|r_{2n-1}|^2 + |r_{2n}|^2 + 2\Re(r_{2n-1}(h_1 - h_2)^*) + 2\Re(r_{2n}^*(h_2 + h_1)) + |h_1 - h_2|^2 + |h_2 + h_1|^2}{\sigma^2} \right\} \\ &\quad + \exp \left\{ -\frac{|r_{2n-1}|^2 + |r_{2n}|^2 - 2\Re(r_{2n-1}(h_1 + h_2)^*) - 2\Re(r_{2n}^*(h_2 - h_1)) + |h_1 + h_2|^2 + |h_2 - h_1|^2}{\sigma^2} \right\} \\ &\quad \left. + \exp \left\{ -\frac{|r_{2n-1}|^2 + |r_{2n}|^2 - 2\Re(r_{2n-1}(h_1 - h_2)^*) - 2\Re(r_{2n}^*(h_2 + h_1)) + |h_1 - h_2|^2 + |h_2 + h_1|^2}{\sigma^2} \right\} \right\} \end{aligned}$$

$$p(R(n); \Theta) = \frac{1}{4\pi^2\sigma^4} \exp \left\{ -\frac{|r_{2n-1}|^2 + |r_{2n}|^2 + |h_1 + h_2|^2 + |h_2 - h_1|^2}{\sigma^2} \right\} *$$

$$\begin{aligned}
& \left\{ \exp \left\{ -\frac{2\Re(r_{2n-1}(h_1 + h_2)^*) + 2\Re(r_{2n}^*(h_2 - h_1))}{\sigma^2} \right\} \right. \\
& + \exp \left\{ -\frac{2\Re(r_{2n-1}(h_1 - h_2)^*) + 2\Re(r_{2n}^*(h_2 + h_1))}{\sigma^2} \right\} \\
& + \exp \left\{ -\frac{-2\Re(r_{2n-1}(h_1 + h_2)^*) - 2\Re(r_{2n}^*(h_2 - h_1))}{\sigma^2} \right\} \\
& \left. + \exp \left\{ -\frac{-2\Re(r_{2n-1}(h_1 + h_2)^*) - 2\Re(r_{2n}^*(h_2 - h_1))}{\sigma^2} \right\} \right\} \\
p(R(n); \Theta) &= \frac{1}{4\pi^2\sigma^4} \exp \left\{ -\frac{|r_{2n-1}|^2 + |r_{2n}|^2 + |h_1 + h_2|^2 + |h_2 - h_1|^2}{\sigma^2} \right\} \\
& * \left\{ 2 \cosh \left( \frac{2\Re(r_{2n-1}(h_1 + h_2)^*) + 2\Re(r_{2n}^*(h_2 - h_1))}{\sigma^2} \right) \right. \\
& \left. + 2 \cosh \left( \frac{2\Re(r_{2n-1}(h_1 - h_2)^*) + 2\Re(r_{2n}^*(h_2 + h_1))}{\sigma^2} \right) \right\}
\end{aligned}$$

We know that,

$$\cosh(x) + \cosh(y) = 2\cosh\left(\frac{x+y}{2}\right)\cosh\left(\frac{x-y}{2}\right)$$

$$\begin{aligned}
p(R(n); \Theta) &= \frac{1}{2\pi^2\sigma^4} \exp \left\{ -\frac{|r_{2n-1}|^2 + |r_{2n}|^2 + |h_1 + h_2|^2 + |h_2 - h_1|^2}{\sigma^2} \right\} \\
& * \left\{ 2 \cosh \left( \frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^2} \right) \cosh \left( \frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^2} \right) \right\} \\
p(R(n); \Theta) &= \frac{1}{\pi^2\sigma^4} \exp \left\{ -\frac{|r_{2n-1}|^2 + |r_{2n}|^2 + 2|h_1|^2 + 2|h_2|^2}{\sigma^2} \right\} \\
& * \left\{ \cosh \left( \frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^2} \right) \cosh \left( \frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^2} \right) \right\} \quad (5.3)
\end{aligned}$$

For *N i. i. d.* received samples of  $R(n)$ , the Joint PDF distribution will be written as follows,

$$\begin{aligned}
p(\mathbf{R}; \Theta) &= \prod_{n=1}^{N/2} \frac{1}{\pi^2\sigma^4} \exp \left\{ -\frac{|r_{2n-1}|^2 + |r_{2n}|^2 + 2|h_1|^2 + 2|h_2|^2}{\sigma^2} \right\} \\
& * \left\{ \cosh \left( \frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^2} \right) \cosh \left( \frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^2} \right) \right\} \quad (5.4)
\end{aligned}$$

The Log Likelihood function will then be given by,

$$\begin{aligned}
\mathcal{L}(\theta) &= \sum_{n=1}^{N/2} \log p(R(n); \theta) \\
\mathcal{L}(\theta) &= \sum_{n=1}^{N/2} \log \left\{ \frac{1}{\pi^2 \sigma^4} \exp \left\{ -\frac{|r_{2n-1}|^2 + |r_{2n}|^2 + 2|h_1|^2 + 2|h_2|^2}{\sigma^2} \right\} \right. \\
&\quad \left. * \left\{ \cosh \left( \frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^2} \right) \cosh \left( \frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^2} \right) \right\} \right\} \\
\mathcal{L}(\theta) &= \sum_{n=1}^{N/2} -\log(\pi^2 \sigma^4) - \frac{|r_{2n-1}|^2 + |r_{2n}|^2 + 2|h_1|^2 + 2|h_2|^2}{\sigma^2} \\
&\quad + \log \cosh \left( \frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^2} \right) \\
&\quad + \log \cosh \left( \frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^2} \right) \tag{5.5}
\end{aligned}$$

Now given the received sample  $\mathbf{R}$ , estimate the parameters  $h_1$ ,  $h_2$ , and  $\sigma^2$ . However,  $h_1$ ,  $h_2$  are complex variables, so we need to estimate their real and imaginary parts separately.

Let

$$h_1 = a + ib \text{ and } h_2 = c + id$$

Now our parameters are as follows,

$$\hat{\theta} = [\hat{a} \ \hat{b} \ \hat{c} \ \hat{d} \ \hat{\sigma}^2]^T$$

$$= \underset{\theta}{\operatorname{argmax}} p(\mathbf{R}; \theta)$$

$$\hat{\rho} = \frac{|\hat{h}_1|^2 + |\hat{h}_2|^2}{\hat{\sigma}^2}$$

$$\widehat{SNR} = 10 \log_{10}(\hat{\rho}) \tag{5.6}$$

## 5.5 The CRLB for the transmit diversity in the BPSK case.

**For the DA case:**

In this case we are assuming that we know the transmitted data  $\mathbf{S}$  and would like to find a lower bound on the variance in this case. We know that CRLB is given by,

$$CRLB(SNR) = \frac{\partial SNR^T}{\partial \theta} I^{-1}(\theta) \frac{\partial SNR}{\partial \theta}$$

where  $I(\theta)$  is the fisher information matrix given by,

$$[I(\theta)]_{ij} = -\mathbb{E} \left[ \frac{\partial^2 \log p(\mathbf{R}; \theta)}{\partial \theta_i \partial \theta_j} \right]$$

And,

$$\frac{\partial SNR}{\partial \theta} = \left[ \frac{20a}{M \log(10)} \quad \frac{20b}{M \log(10)} \quad \frac{20c}{M \log(10)} \quad \frac{20d}{M \log(10)} \quad -\frac{10}{\sigma^2 \log(10)} \right]^T$$

where  $M = |h_1|^2 + |h_2|^2$ .

$$p(R(n)/S(n), \theta) = \frac{1}{\pi^2 \sigma^4} \exp \left\{ -\frac{\|R(n) - HS(n)\|^2}{\sigma^2} \right\}$$

$$p(\mathbf{R}/S(n), \theta) = \prod_{n=1}^{N/2} \frac{1}{\pi^2 \sigma^4} \exp \left\{ -\frac{\|R(n) - HS(n)\|^2}{\sigma^2} \right\}$$

$$\mathcal{L}(\theta) = \sum_{n=1}^{N/2} -\log(\pi^2 \sigma^4) - \frac{\|R(n) - HS(n)\|^2}{\sigma^2}$$

$$= \sum_{n=1}^{N/2} -\log(\pi^2 \sigma^4) - \frac{|r_{2n-1} - (h_1 s_{2n-1} + h_2 s_{2n})|^2 + |r_{2n} - (h_2 s_{2n-1} - h_1 s_{2n})|^2}{\sigma^2}$$

$$= \sum_{n=1}^{N/2} -\log(\pi^2 \sigma^4)$$

$$- \frac{|r_{2n-1} - ((a + ib)s_{2n-1} + (c + id)s_{2n})|^2 + |r_{2n} - ((c + id)s_{2n-1} - (a + ib)s_{2n})|^2}{\sigma^2}$$

$$\begin{aligned}
&= \sum_{n=1}^{N/2} -\log(\pi^2 \sigma^4) \\
&\quad - \frac{|r_{2n-1} - (as_{2n-1} + ib_{2n-1} + cs_{2n} + ids_{2n})|^2 + |r_{2n}^* - (cs_{2n-1} - ids_{2n-1} - as_{2n} + ib_{2n})|^2}{\sigma^2} \\
&= \sum_{n=1}^{N/2} -\log(\pi^2 \sigma^4) - \frac{(\Re(r_{2n-1}) - as_{2n-1} - cs_{2n})^2 + (\Im(r_{2n-1}) - bs_{2n-1} - ds_{2n})^2}{\sigma^2} \\
&\quad - \frac{(\Re(r_{2n}^*) - cs_{2n-1} + as_{2n})^2 + (\Im(r_{2n}^*) - bs_{2n} + ds_{2n-1})^2}{\sigma^2}
\end{aligned}$$

$$\frac{\partial \mathcal{L}(\theta)}{\partial a} = \sum_{n=1}^{N/2} -\frac{1}{\sigma^2} (-2s_{2n-1}(\Re(r_{2n-1}) - as_{2n-1} - cs_{2n}) + 2s_{2n}(\Re(r_{2n}^*) - cs_{2n-1} + as_{2n}))$$

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}(\theta)}{\partial a^2} &= \sum_{n=1}^{N/2} -\frac{1}{\sigma^2} (-2s_{2n-1}(-s_{2n-1}) + 2s_{2n}(s_{2n})) \\
&= \sum_{n=1}^{N/2} -\frac{1}{\sigma^2} (2s_{2n-1}^2 + 2s_{2n}^2)
\end{aligned}$$

Since  $s_{2n-1}^2 = s_{2n}^2 = 1$

$$\frac{\partial^2 \mathcal{L}(\theta)}{\partial a^2} = \sum_{n=1}^{N/2} -\frac{1}{\sigma^2} (4) = -\frac{2N}{\sigma^2}$$

$$\begin{aligned}
[\mathbf{I}(\theta)]_{11} &= -\mathbb{E} \left[ \frac{\partial^2 \mathcal{L}(\theta)}{\partial a^2} \right] = [\mathbf{I}(\theta)]_{22} = -\mathbb{E} \left[ \frac{\partial^2 \mathcal{L}(\theta)}{\partial b^2} \right] = [\mathbf{I}(\theta)]_{33} = -\mathbb{E} \left[ \frac{\partial^2 \mathcal{L}(\theta)}{\partial c^2} \right] = [\mathbf{I}(\theta)]_{44} \\
&= -\mathbb{E} \left[ \frac{\partial^2 \mathcal{L}(\theta)}{\partial d^2} \right] = \frac{2N}{\sigma^2}, \quad [\mathbf{I}(\theta)]_{55} = -\mathbb{E} \left[ \frac{\partial^2 \mathcal{L}(\theta)}{\partial \sigma^2} \right] = \frac{N}{\sigma^4}
\end{aligned}$$

$$[\mathbf{I}(\theta)]_{ij} = -\mathbb{E} \left[ \frac{\partial^2 \mathcal{L}(\theta)}{\partial \theta_i \partial \theta_j} \right] = 0 \text{ for } i \neq j$$

$$\mathbf{I}(\Theta) = Ni(\Theta) = N \begin{bmatrix} \frac{2}{\sigma^2} & 0 & 0 & 0 & 0 \\ 0 & \frac{2}{\sigma^2} & 0 & 0 & 0 \\ 0 & 0 & \frac{2}{\sigma^2} & 0 & 0 \\ 0 & 0 & 0 & \frac{2}{\sigma^2} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\sigma^4} \end{bmatrix}$$

$$CRLB(SNR) = \frac{\partial SNR^T}{\partial \Theta} \mathbf{I}^{-1}(\Theta) \frac{\partial SNR}{\partial \Theta}$$

$$= \begin{bmatrix} \frac{20a}{M \log(10)} & \frac{20b}{M \log(10)} & \frac{20c}{M \log(10)} & \frac{20d}{M \log(10)} & \frac{-10}{\sigma^2 \log(10)} \end{bmatrix} \begin{bmatrix} \frac{2N}{\sigma^2} & 0 & 0 & 0 & 0 \\ 0 & \frac{2N}{\sigma^2} & 0 & 0 & 0 \\ 0 & 0 & \frac{2N}{\sigma^2} & 0 & 0 \\ 0 & 0 & 0 & \frac{2N}{\sigma^2} & 0 \\ 0 & 0 & 0 & 0 & \frac{N}{\sigma^4} \end{bmatrix} \begin{bmatrix} \frac{20a}{M \log(10)} \\ \frac{20b}{M \log(10)} \\ \frac{20c}{M \log(10)} \\ \frac{20d}{M \log(10)} \\ \frac{-10}{\sigma^2 \log(10)} \end{bmatrix}$$

$$= \frac{100}{N(\log(10))^2} \left( \frac{2}{\rho} + 1 \right) \quad (5.7)$$

**For the NDA case:**

$$\mathcal{L}(\Theta) = \sum_{n=1}^{N/2} -\log(\pi^2 \sigma^4) - \frac{|r_{2n-1}|^2 + |r_{2n}|^2 + 2|h_1|^2 + 2|h_2|^2}{\sigma^2} + \log \cosh \left( \frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^2} \right) + \log \cosh \left( \frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^2} \right)$$

We need to find the fisher information matrix for this Likelihood function in order to find the CRLB for the NDA case.

$$[\mathbf{I}(\Theta)]_{ij} = -\mathbb{E} \left[ \frac{\partial^2 \mathcal{L}(\Theta)}{\partial \theta_i \partial \theta_j} \right], \quad \text{and} \quad [\mathbf{I}(\Theta)]_{ij} = [\mathbf{I}(\Theta)]_{ji} \text{ for } i \neq j,$$

$$[\mathbf{I}(\Theta)]_{11} = -\mathbb{E} \left[ \frac{\partial^2 \mathcal{L}(\Theta)}{\partial a^2} \right], \quad [\mathbf{I}(\Theta)]_{12} = -\mathbb{E} \left[ \frac{\partial^2 \mathcal{L}(\Theta)}{\partial a \partial b} \right], \quad [\mathbf{I}(\Theta)]_{13} = -\mathbb{E} \left[ \frac{\partial^2 \mathcal{L}(\Theta)}{\partial a \partial c} \right],$$

$$[\mathbf{I}(\Theta)]_{14} = -\mathbb{E} \left[ \frac{\partial^2 \mathcal{L}(\Theta)}{\partial a \partial d} \right], \quad [\mathbf{I}(\Theta)]_{15} = -\mathbb{E} \left[ \frac{\partial^2 \mathcal{L}(\Theta)}{\partial a \partial \sigma^2} \right], \quad [\mathbf{I}(\Theta)]_{22} = -\mathbb{E} \left[ \frac{\partial^2 \mathcal{L}(\Theta)}{\partial b^2} \right],$$

$$[\mathbf{I}(\Theta)]_{23} = -\mathbb{E} \left[ \frac{\partial^2 \mathcal{L}(\Theta)}{\partial b \partial c} \right], \quad [\mathbf{I}(\Theta)]_{24} = -\mathbb{E} \left[ \frac{\partial^2 \mathcal{L}(\Theta)}{\partial b \partial d} \right], \quad [\mathbf{I}(\Theta)]_{25} = -\mathbb{E} \left[ \frac{\partial^2 \mathcal{L}(\Theta)}{\partial b \partial \sigma^2} \right],$$

$$[\mathbf{I}(\Theta)]_{33} = -\mathbb{E} \left[ \frac{\partial^2 \mathcal{L}(\Theta)}{\partial c^2} \right], \quad [\mathbf{I}(\Theta)]_{34} = -\mathbb{E} \left[ \frac{\partial^2 \mathcal{L}(\Theta)}{\partial c \partial d} \right], \quad [\mathbf{I}(\Theta)]_{35} = -\mathbb{E} \left[ \frac{\partial^2 \mathcal{L}(\Theta)}{\partial c \partial \sigma^2} \right]$$

$$[\mathbf{I}(\Theta)]_{44} = -\mathbb{E} \left[ \frac{\partial^2 \mathcal{L}(\Theta)}{\partial d^2} \right], \quad [\mathbf{I}(\Theta)]_{45} = -\mathbb{E} \left[ \frac{\partial^2 \mathcal{L}(\Theta)}{\partial d \partial \sigma^2} \right], \quad [\mathbf{I}(\Theta)]_{55} = -\mathbb{E} \left[ \frac{\partial^2 \mathcal{L}(\Theta)}{\partial \sigma^2} \right].$$

$$\begin{aligned} \frac{\partial^2 \mathcal{L}(\Theta)}{\partial a^2} &= \sum_{n=1}^{N/2} -\frac{4}{\sigma^2} + \operatorname{sech}^2 \left( \frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^2} \right) \left( -\frac{2\Re(r_{2n}^*)}{\sigma^2} \right)^2 \\ &\quad + \operatorname{sech}^2 \left( \frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^2} \right) \left( \frac{2\Re(r_{2n-1})}{\sigma^2} \right)^2 \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \mathcal{L}(\Theta)}{\partial a \partial b} &= \sum_{n=1}^{N/2} \operatorname{sech}^2 \left( \frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^2} \right) \left( -\frac{2\Re(r_{2n}^*)}{\sigma^2} \right) \left( \frac{2\Im(r_{2n}^*)}{\sigma^2} \right) \\ &\quad + \operatorname{sech}^2 \left( \frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^2} \right) \left( \frac{2\Re(r_{2n-1})}{\sigma^2} \right) \left( \frac{2\Im(r_{2n-1})}{\sigma^2} \right) \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \mathcal{L}(\Theta)}{\partial a \partial c} &= \sum_{n=1}^{N/2} \operatorname{sech}^2 \left( \frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^2} \right) \left( -\frac{2\Re(r_{2n}^*)}{\sigma^2} \right) \left( \frac{2\Re(r_{2n-1})}{\sigma^2} \right) \\ &\quad + \operatorname{sech}^2 \left( \frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^2} \right) \left( \frac{2\Re(r_{2n-1})}{\sigma^2} \right) \left( \frac{2\Re(r_{2n}^*)}{\sigma^2} \right) \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \mathcal{L}(\Theta)}{\partial a \partial d} &= \sum_{n=1}^{N/2} \operatorname{sech}^2 \left( \frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^2} \right) \left( -\frac{2\Re(r_{2n}^*)}{\sigma^2} \right) \left( \frac{2\Im(r_{2n-1})}{\sigma^2} \right) \\ &\quad + \operatorname{sech}^2 \left( \frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^2} \right) \left( \frac{2\Re(r_{2n-1})}{\sigma^2} \right) \left( -\frac{2\Im(r_{2n}^*)}{\sigma^2} \right) \end{aligned}$$



$$\begin{aligned}
\frac{\partial^2 \mathcal{L}(\Theta)}{\partial a \partial \sigma^2} &= \sum_{n=1}^{N/2} \frac{4a}{\sigma^4} + \tanh\left(\frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^2}\right) \left(\frac{2\Re(r_{2n}^*)}{\sigma^4}\right) \\
&\quad + \operatorname{sech}^2\left(\frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^2}\right) \left(-\frac{2\Re(r_{2n}^*)}{\sigma^2}\right) \left(-\frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^4}\right) \\
&\quad + \tanh\left(\frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^2}\right) \left(-\frac{2\Re(r_{2n-1})}{\sigma^4}\right) \\
&\quad + \operatorname{sech}^2\left(\frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^2}\right) \left(\frac{2\Re(r_{2n-1})}{\sigma^2}\right) \left(-\frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^4}\right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}(\Theta)}{\partial b^2} &= \sum_{n=1}^{N/2} -\frac{4}{\sigma^2} + \operatorname{sech}^2\left(\frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^2}\right) \left(\frac{2\Im(r_{2n}^*)}{\sigma^2}\right)^2 \\
&\quad + \operatorname{sech}^2\left(\frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^2}\right) \left(\frac{2\Im(r_{2n-1})}{\sigma^2}\right)^2
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}(\Theta)}{\partial b \partial c} &= \sum_{n=1}^{N/2} \operatorname{sech}^2\left(\frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^2}\right) \left(\frac{2\Im(r_{2n}^*)}{\sigma^2}\right) \left(\frac{2\Re(r_{2n-1})}{\sigma^2}\right) \\
&\quad + \operatorname{sech}^2\left(\frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^2}\right) \left(\frac{2\Im(r_{2n-1})}{\sigma^2}\right) \left(\frac{2\Re(r_{2n}^*)}{\sigma^2}\right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}(\Theta)}{\partial b \partial d} &= \sum_{n=1}^{N/2} \operatorname{sech}^2\left(\frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^2}\right) \left(\frac{2\Im(r_{2n}^*)}{\sigma^2}\right) \left(\frac{2\Im(r_{2n-1})}{\sigma^2}\right) \\
&\quad + \operatorname{sech}^2\left(\frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^2}\right) \left(\frac{2\Im(r_{2n-1})}{\sigma^2}\right) \left(-\frac{2\Im(r_{2n}^*)}{\sigma^2}\right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}(\Theta)}{\partial b \partial \sigma^2} &= \sum_{n=1}^{N/2} \frac{4b}{\sigma^4} + \tanh\left(\frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^2}\right) \left(-\frac{2\Im(r_{2n}^*)}{\sigma^4}\right) \\
&\quad + \operatorname{sech}^2\left(\frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^2}\right) \left(\frac{2\Im(r_{2n}^*)}{\sigma^2}\right) \left(-\frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^4}\right) \\
&\quad + \tanh\left(\frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^2}\right) \left(-\frac{2\Im(r_{2n-1})}{\sigma^4}\right) \\
&\quad + \operatorname{sech}^2\left(\frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^2}\right) \left(\frac{2\Im(r_{2n-1})}{\sigma^2}\right) \left(-\frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^4}\right)
\end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \mathcal{L}(\theta)}{\partial c^2} &= \sum_{n=1}^{N/2} -\frac{4}{\sigma^2} + \operatorname{sech}^2 \left( \frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^2} \right) \left( \frac{2\Re(r_{2n-1})}{\sigma^2} \right)^2 \\ &\quad + \operatorname{sech}^2 \left( \frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^2} \right) \left( \frac{2\Re(r_{2n}^*)}{\sigma^2} \right)^2 \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \mathcal{L}(\theta)}{\partial c \partial d} &= \sum_{n=1}^{N/2} \operatorname{sech}^2 \left( \frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^2} \right) \left( \frac{2\Re(r_{2n-1})}{\sigma^2} \right) \left( \frac{2\Im(r_{2n-1})}{\sigma^2} \right) \\ &\quad + \operatorname{sech}^2 \left( \frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^2} \right) \left( \frac{2\Re(r_{2n}^*)}{\sigma^2} \right) \left( \frac{-2\Im(r_{2n}^*)}{\sigma^2} \right) \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \mathcal{L}(\theta)}{\partial c \partial \sigma^2} &= \sum_{n=1}^{N/2} \frac{4c}{\sigma^4} + \tanh \left( \frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^2} \right) \left( -\frac{2\Re(r_{2n-1})}{\sigma^4} \right) \\ &\quad + \operatorname{sech}^2 \left( \frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^2} \right) \left( \frac{2\Re(r_{2n-1})}{\sigma^2} \right) \left( -\frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^4} \right) \\ &\quad + \tanh \left( \frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^2} \right) \left( -\frac{2\Re(r_{2n}^*)}{\sigma^4} \right) \\ &\quad + \operatorname{sech}^2 \left( \frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^2} \right) \left( \frac{2\Re(r_{2n}^*)}{\sigma^2} \right) \left( -\frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^4} \right) \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \mathcal{L}(\theta)}{\partial d^2} &= \sum_{n=1}^{N/2} -\frac{4}{\sigma^2} + \operatorname{sech}^2 \left( \frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^2} \right) \left( \frac{2\Im(r_{2n-1})}{\sigma^2} \right)^2 \\ &\quad + \operatorname{sech}^2 \left( \frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^2} \right) \left( \frac{-2\Im(r_{2n}^*)}{\sigma^2} \right)^2 \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \mathcal{L}(\theta)}{\partial d \partial \sigma^2} &= \sum_{n=1}^{N/2} \frac{4d}{\sigma^4} + \tanh \left( \frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^2} \right) \left( -\frac{2\Im(r_{2n-1})}{\sigma^4} \right) \\ &\quad + \operatorname{sech}^2 \left( \frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^2} \right) \left( \frac{2\Im(r_{2n-1})}{\sigma^2} \right) \left( -\frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^4} \right) \\ &\quad + \tanh \left( \frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^2} \right) \left( \frac{2\Im(r_{2n}^*)}{\sigma^4} \right) \\ &\quad + \operatorname{sech}^2 \left( \frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^2} \right) \left( \frac{-2\Im(r_{2n}^*)}{\sigma^2} \right) \left( -\frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^4} \right) \end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}(\theta)}{\partial \sigma^{2^2}} &= \sum_{n=1}^{N/2} \frac{2}{\sigma^4} - 2 \frac{|r_{2n-1}|^2 + |r_{2n}^*|^2 + 2|h_1|^2 + 2|h_2|^2}{\sigma^6} \\
&+ \tanh\left(\frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^2}\right) \left(2 \frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^6}\right) \\
&+ \operatorname{sech}^2\left(\frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^2}\right) \left(-\frac{2\Re(r_{2n-1}h_2^*) - 2\Re(r_{2n}^*h_1)}{\sigma^4}\right)^2 \\
&+ \tanh\left(\frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^2}\right) \left(2 \frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^6}\right) \\
&+ \operatorname{sech}^2\left(\frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^2}\right) \left(-\frac{2\Re(r_{2n-1}h_1^*) + 2\Re(r_{2n}^*h_2)}{\sigma^4}\right)^2
\end{aligned}$$

Numerical integration is used to find the statistical expectation of each entry in the following fisher information matrix.

$$\mathbf{I}(\theta) = \begin{bmatrix} -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial a^2}\right] & -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial a \partial b}\right] & -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial a \partial c}\right] & -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial a \partial d}\right] & -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial a \partial \sigma^2}\right] \\ -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial b \partial a}\right] & -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial b^2}\right] & -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial b \partial c}\right] & -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial b \partial d}\right] & -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial b \partial \sigma^2}\right] \\ -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial c \partial a}\right] & -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial c \partial b}\right] & -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial c^2}\right] & -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial c \partial d}\right] & -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial c \partial \sigma^2}\right] \\ -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial d \partial a}\right] & -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial d \partial b}\right] & -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial d \partial c}\right] & -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial d^2}\right] & -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial b \partial \sigma^2}\right] \\ -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial \sigma^2 \partial a}\right] & -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial \sigma^2 \partial b}\right] & -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial \sigma^2 \partial c}\right] & -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial \sigma^2 \partial d}\right] & -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial \sigma^2}\right] \end{bmatrix}$$

This is then used to find the CRLB for the SNR as follows,

$$\text{CRLB}(\text{SNR}) = \frac{\partial \text{SNR}^T}{\partial \theta} \mathbf{I}^{-1}(\theta) \frac{\partial \text{SNR}}{\partial \theta}$$

## 5.6 STBC NDA SNR estimation using the EM algorithm.

In the coming SNR estimation, BPSK data will be used, but one may choose any signaling scheme. The EM algorithm will be used to estimate the SNR. The EM algorithm will iteratively estimate the channels and the noise variance which then be used to estimate the SNR. The problem is, given the received samples  $\mathbf{R}$  without the knowledge of the transmitted symbols, estimate the unknown parameters. At each time, we receive a sample, we don't know to which of the four mixtures does this sample belongs to. Our received data will be incomplete which won't be easy for the MLE to be

used to estimate its parameters; we resort to an iterative method like the EM algorithm. We start by writing the auxiliary function,

$$Q(\theta^k, \theta^{k+1}) = \mathbb{E}_{p(Z/Y, \theta^k)} \{ \mathbf{log}(p(\mathbf{R}, Z/\theta^{k+1})) | \mathbf{R}, \theta^k \}$$

$$Q(\theta^k, \theta^{k+1}) = \sum_{n=1}^{N/2} \sum_{m=1}^M p(\omega_m/R(n), \theta^k) \mathbf{log} p(R(n)/\omega_m, \theta^{k+1}) \\ + \sum_{n=1}^{N/2} \sum_{m=1}^M p(\omega_m/R(n), \theta^k) \mathbf{log} p^{k+1}(\omega_m)$$

$$Q(\theta^k, \theta^{k+1}) = \sum_{m=1}^M \sum_{n=1}^{N/2} p(\omega_m/R(n), \theta^k) \left\{ - \left\{ (R(n) - \mu_m^{k+1})^H \Sigma_m^{k+1-1} (R(n) - \mu_m^{k+1}) \right\} \right. \\ \left. + \sum_{n=1}^M \sum_{n=1}^{N/2} p(\omega_m/R(n), \theta^k) \left\{ - \left\{ \mathbf{log} \{ (2\pi)^D |\Sigma_m^{k+1}| \} \right\} \right\} \right. \\ \left. + \sum_{m=1}^M \sum_{n=1}^{N/2} p(\omega_m/R(n), \theta^k) \mathbf{log} \{ C \} \right.$$

$$\mathbf{log} p(R(n)/\omega_m, \theta^{k+1}) = -\mathbf{log} \{ \pi^2 \sigma^{4k+1} \} - \frac{\|R(n) - H^{k+1} S_m\|^2}{\sigma^{2k+1}}$$

$$Q(\theta^k, \theta^{k+1}) = \sum_{m=1}^M \sum_{n=1}^{N/2} p(\omega_m/R(n), \theta^k) \left\{ -\mathbf{log} \{ \pi^2 \sigma^{4k+1} \} - \frac{\|R(n) - H^{k+1} S_m\|^2}{\sigma^{2k+1}} \right\}$$

$$\theta^{k+1} = \underset{\theta}{\operatorname{argmax}} Q(\theta^k, \theta^{k+1})$$

$$\frac{\partial Q(\theta^k, \theta^{k+1})}{\partial a^{k+1}} = - \sum_{n=1}^{N/2} \sum_{m=1}^M p(\omega_m/R(n), \theta^k) \frac{\partial}{\partial a^{k+1}} \frac{\|R(n) - H^{k+1} S_m\|^2}{\sigma^{2k+1}}$$

$$a^{k+1} = \frac{1}{N} \sum_{n=1}^{N/2} \sum_{m=1}^M p(\omega_m/R(n), \theta^k) \Re(r_{2n-1} s_{2m-1} - r_{2n}^* s_{2m}) \quad (5.8)$$

$$\frac{\partial Q(\theta^k, \theta^{k+1})}{\partial b^{k+1}} = - \sum_{n=1}^{N/2} \sum_{m=1}^M p(\omega_m/R(n), \theta^k) \frac{\partial}{\partial b^{k+1}} \frac{\|R(n) - H^{k+1} S_m\|^2}{\sigma^{2k+1}}$$

$$b^{k+1} = \frac{1}{N} \sum_{n=1}^{N/2} \sum_{m=1}^M p(\omega_m/R(n), \theta^k) \Im(r_{2n-1} s_{2m-1} + r_{2n}^* s_{2m}) \quad (5.9)$$

$$\frac{\partial Q(\theta^k, \theta^{k+1})}{\partial c^{k+1}} = - \sum_{n=1}^{N/2} \sum_{m=1}^M p(\omega_m/R(n), \theta^k) \frac{\partial}{\partial c^{k+1}} \frac{\|R(n) - H^{k+1} S_m\|^2}{\sigma^{2k+1}}$$

$$c^{k+1} = \frac{1}{N} \sum_{n=1}^{N/2} \sum_{m=1}^M p(\omega_m/R(n), \theta^k) \Re(r_{2n-1} s_{2m} + r_{2n}^* s_{2m-1}) \quad (5.10)$$

$$\frac{\partial Q(\theta^k, \theta^{k+1})}{\partial d^{k+1}} = - \sum_{n=1}^{N/2} \sum_{m=1}^M p(\omega_m/R(n), \theta^k) \frac{\partial}{\partial d^{k+1}} \frac{\|R(n) - H^{k+1} S_m\|^2}{\sigma^{2k+1}}$$

$$d^{k+1} = \frac{1}{N} \sum_{n=1}^{N/2} \sum_{m=1}^M p(\omega_m/R(n), \theta^k) \Im(r_{2n-1} s_{2m} - r_{2n}^* s_{2m-1}) \quad (5.11)$$

$$\frac{\partial Q(\theta^k, \theta^{k+1})}{\partial \sigma^{2k+1}} = - \sum_{n=1}^{N/2} \sum_{m=1}^M p(\omega_m/R(n), \theta^k) \frac{\partial}{\partial \sigma^{2k+1}} \left\{ -\log\{\pi^2 \sigma^{4k+1}\} - \frac{\|R(n) - H^{k+1} S_m\|^2}{\sigma^{2k+1}} \right\}$$

$$\sigma^{2k+1} = \frac{1}{N} \sum_{n=1}^{N/2} \sum_{m=1}^M p(\omega_m/R(n), \theta^k) \|R(n) - H^{k+1} S_m\|^2 \quad (5.12)$$

where  $M=4$  in our PBSK case, and  $N$  is the received sample size. Also, the mixtures classes are assumed equally likely. Equations (5.8)-(5.12) will be used to update the parameters iteratively starting from initial values for those parameters. Luckily that our model has one global optimal parameters, which keep us away from thinking that the values which we estimated might be local maxima. The MSE & CRLB will be drawn for different values of  $N$ .

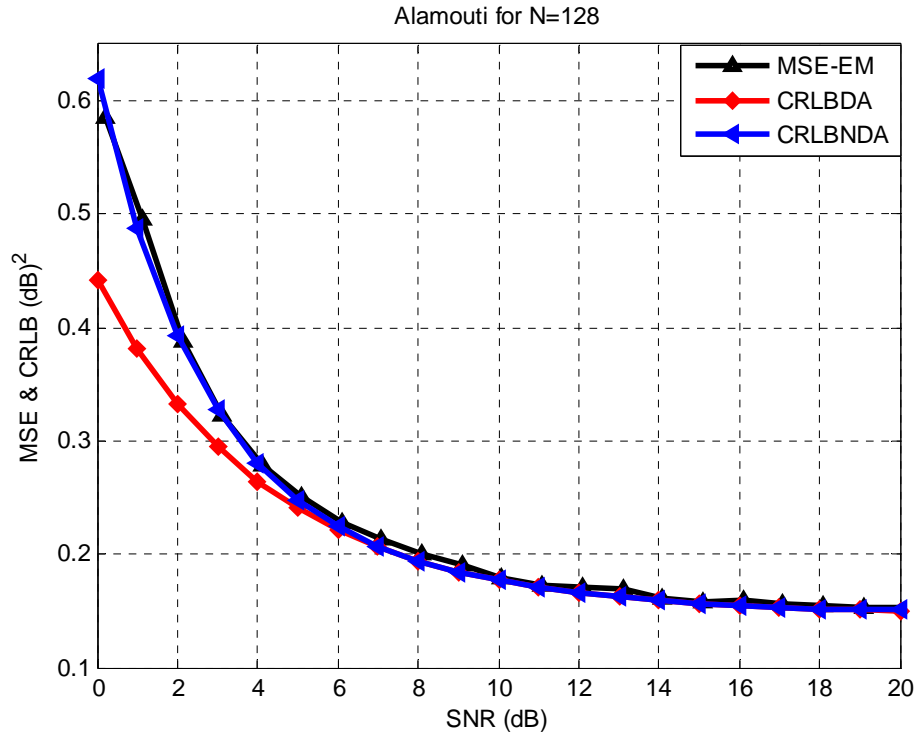


Figure 5.4. The MSE & CRLB in dB<sup>2</sup> vs. SNR in dB when N=128.

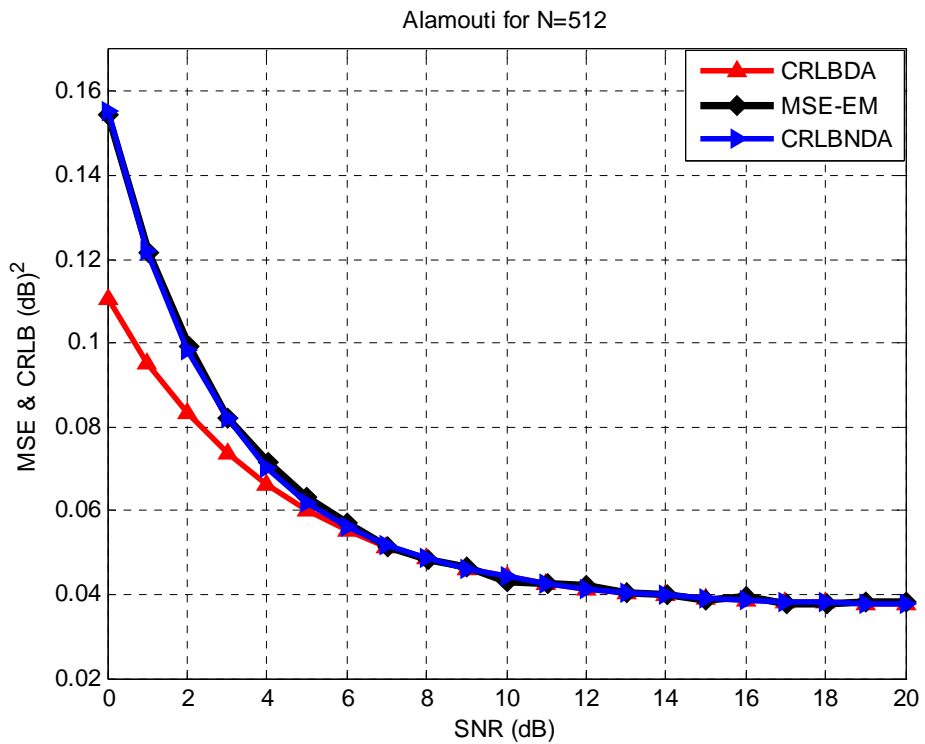


Figure 5.5. The MSE & CRLB in dB<sup>2</sup> vs. SNR in dB when N=512.

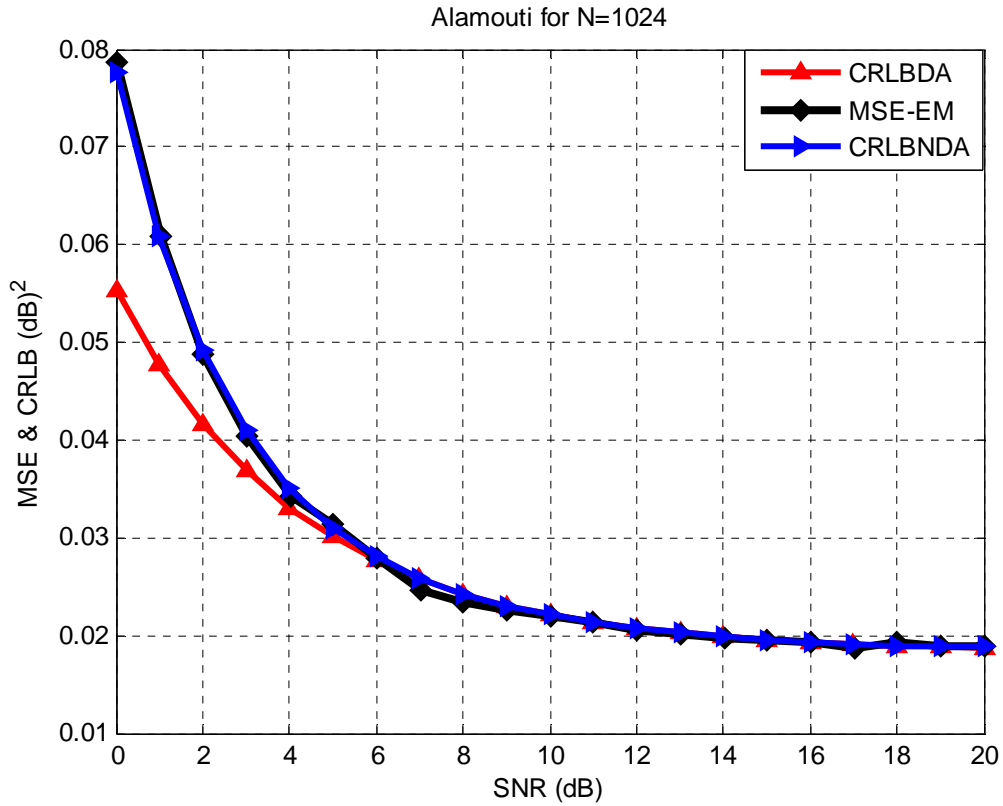


Figure 5.6. The MSE & CRLB in dB<sup>2</sup> vs. SNR in dB when N=1024.

**Convergence time,  $\tau$  of the EM Algorithm. ( SNR=8dB):**

| Data size, N | Convergence time, $\tau$ in seconds for one run | Convergence time, $\tau$ in seconds for 2000 Monte Carlo simulations | Convergence time, $\tau$ in seconds for 10000 Monte Carlo simulations |
|--------------|---|--|---|
|              | $\tau_1$  | $\tau_{2000}$  | $\tau_{10,000}$   |
| N=128        | 0.131456  | 6.881052   | 34.264368   |
| N=512        | 0.140788  | 27.668970  | 139.303882  |
| N=1024       | 0.173774  | 91.254040  | 456.952999  |

The simulations were performed using MATLAB version 7.13.0.564 (R2011b), with processor Intel(R) Core(TM) i7 CPU 860 @ 2.80 GHz with RAM 4 GB.

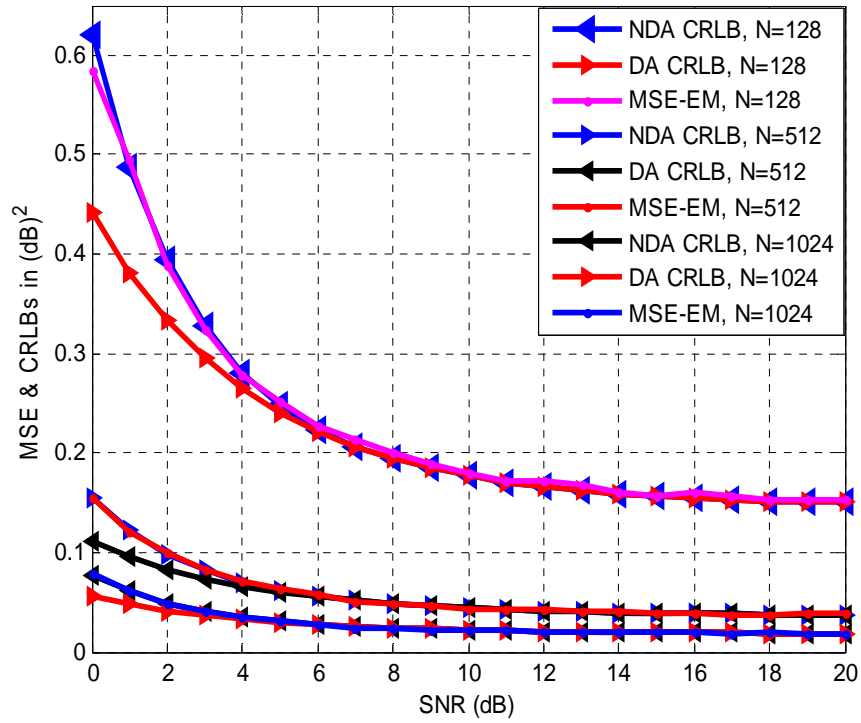


Fig. 5.7. The MISO with STBC BPSK case.

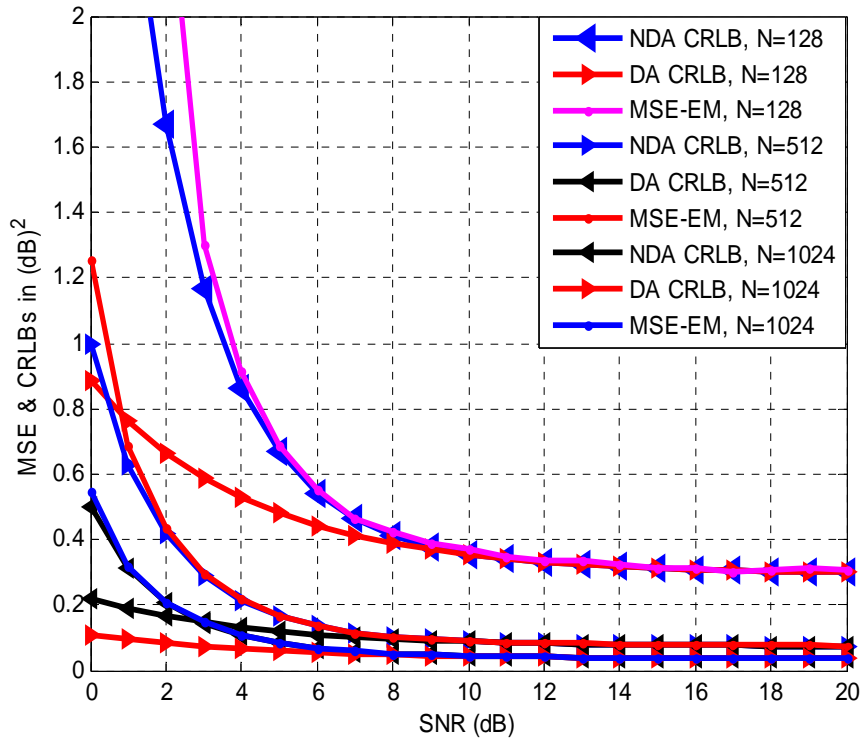


Fig. 5.8. The SISO BPSK case.



## 6. Conclusions

In this research, transmit diversity is used as a way of improving the SNR estimation accuracy. The SNR was estimated using the EM algorithm in a NDA manner using space-time block coding (STBC) which was done by Alamouti [12]. Also, the CRLB for the DA and NDA case were derived and plotted which then be used to assess the performance of our estimator. This was done for the BPSK constellation system. It is being done for different sample sizes,  $N$ ; 128, 512, and 1024. We've seen that the CRLB and MSE are inversely proportional to the sample size,  $N$ . In addition, as we increase the SNR, CRLB and MSE become closer and closer to each other till they match at some high SNR point. In general, the detection problem depends highly on the SNR and  $N$ . It is a tradeoff between these two factors. For example, if the SNR is low, then we need to increase  $N$  in order to decrease the error probability and the MSE estimate. On the other hand however, if the SNR is high, then small  $N$  will work for us. This technique was then compared with the SISO [3], [4], [8], [10] and SIMO [5], [9] models. SIMO systems would have some drawbacks especially when being applied for mobile phones like, size and cost of the phone, and power consumption. Instead of having the diversity done at each remote unit, instead, it is being done at the base station. For example, having two antennas at the base station covering hundreds of remote units (having one receive antenna) would be more economical than having hundreds of remote units (each having two antennas) and one transmit antenna at the base station. This is the main reason of choosing MISO with STBC systems over SIMO systems.

## 7. Future work

1. SNR estimation using the EM algorithm with other linearly modulated signals other than BPSK like M-QAM and M-PSK.
2. Bayesian SNR estimation.
3. MIMO SNR estimation using the EM algorithm.
4. SNR estimation in non-Gaussian noise.

## 8. References:

- [1] C. M. Bishop, Pattern recognition and machine learning. Springer 2006.
- [2] S. M. Kay, Fundamentals of statistical signal processing. Prentice Hall, 1993.
- [3] A. Wiesel, J. Goldberg, and H. Messer, "Non-data-aided signal-to-noise-ratio estimation," Communications, 2002. ICC 2002. IEEE International Conference on, vol. 1, pp. 197-201, 2002.
- [4] A. Das, "NDA SNR estimation: CRLBs and EM based estimators," in proc. TENCON 2008, Hyderabad, India, Nov. 2008, pp. 1-6.
- [5] M. A. Boujelben, F. Bellili, S. Affes, and A. Stephenne, "EM algorithm for non-data-aided SNR estimation of linearly-modulated signals over SIMO channels," in proc. Communications IEEE GLOBECOM 2009, pp. 1-6.
- [6] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," Journal of the Royal Statistical Society. Series B, no. 1, pp. 1-38, 1977.
- [7] T. K. Moon, "The expectation-maximization algorithm," IEEE Signal process. Mag., vol. 13, pp. 47-60, Nov. 1996.
- [8] N. S. Alagha, "Cramer-rao bounds of SNR estimates for BPSK and QPSK modulated signals," IEEE Communications letters, vol. 5, no. 1, pp. 10-12, January 2001.
- [9] M. A. Boujelben, F. Bellili, S. Affes, and A. Stephenne, "SNR estimation over SIMO channels from linearly modulated signals," in proc. Signal processing IEEE British Crown, vol. 58, n0. 12, December 2010, pp. 6017-6028.
- [10] A. Das and M. Miller, "Remote non-linearity detection via burst power dithering and EM based SNR estimation," International workshop on Satellite and space communications. IWSSC, sept. 2007.
- [11] F. Rice, B. Cowley, and M. Rice, "Cramer-rao lower bounds for QAM phase and frequency estimation," IEEE Trans. Communications, vol. 49, pp. 1582-1591, sept. 2001.
- [12] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," IEEE Trans. Communications, vol. 16, pp. 1451-1458, sept. 1998.

- [13 ] N. Celandroni, E. Ferro, and F. Potorti, "Quality estimation of PSK modulated signals," IEEE Comm. Mag., pp. 50-55. July 1997.
- [14] K. Balachandran, S. R. Kadaba, and S. Nanda, "Channel Quality estimation and rate adaptation for cellular mobile radio," IEEE J. Select Areas Comm., vol. 17, pp. 1244-1256, July 1999.
- [15] F. Bellili, S. Affes, and A. Stephenne, "Cramer-Rao Bound for NDA SNR estimates of Square QAM Modulated Signals," IEEE in WCNC 2009, pp. 1-5.
- [16] F. Bellili, S. Affes, and A. Stephenne, "Cramer-Rao Bounds for NDA SNR estimates of Square QAM Modulated Transmissions," IEEE in Comm. 2010, pp. 3211-3218.
- [17] S. Caiyao, Y. Hongyi, and Z. Tianzhong, "Cramer-Rao bounds of SNR Estimation for BPSK and QPSK signals over SIMO Channels," IEEE ICSPS 2010, pp. 647-650.
- [18] W. Gappmair, R. Lopez-Valcare, and C. Mosquera, "ML and EM algorithm for non-data-aided SNR estimation of linearly modulated signals," in Proc. IEEE symp. Commun. Syst., Graz, Austria, Jul. 2008, pp. 530-534.
- [19] R. Lopez-Valcare, C. Mosquera, and W. Gappmair, "Iterative envelop based SNR estimation for nonconstant modulus constellation," in Proc. IEEE SPAWC'07, Helsinki, Finland, Jun. 2007, pp. 1-5.
- [20] W. Gappmair, R. Lopez-Valcare, and C. Mosquera, "Cramer-Rao lower bound and EM algorithm for envelope-based SNR estimation of nonconstant modulus constellation," IEEE Trans. Commun. Vol. 57, no. 6, June 2009, pp. 1622-1627.