6-2012

Modulation Classification and Parameter Estimation in Wireless Networks

Ruoyu Li
Syracuse University, liruoyuxgd2006@gmail.com

ABSTRACT

Modulation classification is a crucial step between data recognition and demodulation in Cognitive Radio systems. In this thesis, automatic blind modulation classification approaches designed for multiple modulation classification in distributed sensor network are discussed. Optimal Bayesian Approach and Likelihood Based Detection are the main mathematical foundations. First, we build a channel and signal model under the assumption of Additive White Gaussian Noise (AWGN) and Rayleigh Fading Channel. For the distributed detection scheme, we compare the performance of Chair-Varshney Fusion Rule with Majority Fusion Rule in a coherent communication environment. For a more general scenario with embedded unknown channel parameters, Hybrid Likelihood Ratio Test method with different types of estimations are studied. Treating transmitted symbols as hidden variables, we propose an Expectation Maximization Algorithm based approach to determine maximum likelihood estimates (MLE) so that we can have closed form expressions of MLEs. Centralized data fusion based classifier equipped with EM algorithm based maximum likelihood estimates is then evaluated via comparison to the classifier equipped with Method of Moments (MoM) estimates. From computer simulation experiments, we conclude that EM algorithm based MLE can efficiently obtain better classification performance than Moments-Based estimation especially in low Signal-to-Noise Ratio environments.

MODULATION CLASSIFICATION AND PARAMETER ESTIMATION IN WIRELESS NETWORKS

By

Ruoyu Li
B.E. Northwestern Polytechnical University, 2010

THESIS

Submitted in partial fulfillment of the requirements for the degree of
Master of Science in Electrical Engineering

Syracuse University
June 2012

Advised by:
Prof. Pramod K. Varshney

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1 Motivation

Because of the development of modern communication technology and requirements from both commercial and military communication service users, digital communication designers need to build new communication systems with more flexible and robust information transmitting and receiving schemes for the purpose of reacting to the diversity of communication environments and for making full use of channel resources [10], [3]. In order to fulfill such kind of service requirements provided by users, the radio receivers are required to have the ability of sensing the state of communication channel as well as the electromagnetic environment and of extracting useful information from received noisy data. Modulation Classification is one of the techniques that could allow the receiver to become aware of the current status of data transmitter and communication channel. If the receiver can successfully classify the modulation scheme being used in the received data, the software defined radio (SDR) used as the receiver could modify the demodulator part of the receiver to react to such changes and to avoid the increasing error rate caused by usage of the inappropriate demodulation algorithm. Also, it can help in identifying and locating the transmitter.

In practical applications, modulation classification could be realized by inserting the information regarding the modulation in use in the data sequence and transmitting it to the receiver [21], [20]. This straightforward modulation awareness method is based on the assumption of adequate bandwidth resources as well as the availability of hardware/software resources at the receivers. In many applications, extra bandwidth or additional hardware resources may not be acceptable. Also in non-cooperating environments, the transmitter may not be willing to add extra information for modulation classification. Thus, it is imperative that a blind solution be found which does not require any additional bandwidth or receiver complexity and is able to operate well in non-cooperating environments.

1.2 Technical Background

The implementation of blind modulation classification techniques is based on a number of advanced communication and signal processing techniques. These advances as well as networked sensing have made the proposed automatic modulation classification techniques physically achievable. In the next subsections, we will discuss some of the underlying techniques.

A. Cognitive and Software-Defined Radios

The concept of cognitive radio (CR) was first presented by Joseph Mitola III [19] in 1998. Cognitive radio technology is a wireless communication technology which can sense radio resources and users' needs. Various wireless devices such as personal digital assistants (PDAs) and the related networks are sufficiently intelligent about the radio resources and the related computer-to-computer communications to detect users' needs as a function of users' context, and to provide radio resources and wireless services most appropriate to those needs. Cognitive radio technology is meant as a new architecture of large-scale communication system and networks. Instead of centralized traditional cellular wireless communication systems constructed by optical cables and host/base stations, a more flat

ad-hoc communication system is developed and studied which has been made possible by the increasing computation capability of mobile electronic devices. Every single device can at least partially undertake some data processing work which was usually done by the computation center that collects all the data from the nodes.

This has been made possible by the Software-Defined-Radio (SDR) technology that has emerged to make the communication more flexible and intelligent. The SDR is a radio communication system where components that have been typically implemented in hardware (e.g. mixers, filters, amplifiers, modulators/demodulators, detectors, etc.) [1] [25] are instead implemented by means of software on embedded computing devices. Since the cost of adjustment of software coding is much less than that of changing hardware circuits and antennas, the advantage of software-defined radio network over traditional hardware-defined radio is obvious. There are already some well developed platforms available to business customers and academic researchers. Based on a SDR platform, the receiver can easily modify itself to react to the varying channel conditions and even transfer the channel information back to the transmitter. Then the transmitter could also adapt its working model including coding scheme and modulation format to improve communication quality. For the facilication of the long-term development of cognitive radios, all technology developments and industrial applications of CR and SDR techniques need to follow the rules made by the Federal Committee of Communications [5].

B. Wireless Sensor Networks

A wireless sensor network (WSN) consists of spatially distributed autonomous sensors to monitor physical or environmental conditions, such as temperature, sound, vibration, pressure, motion and to cooperatively pass their data through the network to a main location or a fusion center for data processing. In WSNs, the fusion center and distributed sensors have a bi-directional data exchange and fusion center does not only collect data from the sensors but also broadcasts information to the whole network or some specific nodes.

Even though the concept of a wireless sensor network and its associated technology

was developed by military users for military applications such as battlefield surveillance, nowadays the principles and architecture of WSNs are being applied in many industrial and commercial applications for process monitoring and control, machine health monitoring, and so on. The sensor network is built of "nodes" which are the topological representations of physical sensors. Each such wireless sensor network node typically has several parts: a radio transceiver with an internal antenna or connection to an external antenna, a microcontroller, an electronic circuit for interfacing with the sensors and an energy source, usually a battery or an embedded form of energy harvesting. The cost of sensor nodes varies, ranging from a few to hundreds of dollars, depending on the complexity of the individual sensor nodes. Size and cost constraints on sensor nodes result in corresponding constraints on resources such as energy, memory, computational speed and communications bandwidth. Due to the constraints on implementation of physical sensors, sensors are allowed to undertake reasonable-scale data processing tasks according to sensor capabilities.

The topology of the WSNs can vary from a simple star network to an advanced multi-hop wireless mesh network. The data propagation techniques between the hops of the network are based on some routing protocols. If a centralized architecture is used in a wireless sensor network and the central node fails, then the entire network will collapse. However, the reliability of the sensor network can be increased by using a distributed control architecture. In wireless sensor networks, information fusion, also known as data fusion, has been developed for processing sensor data by filtering, aggregating, and making inferences about the gathered data. Information fusion deals with the combination of data from multiple sources to obtain improved information that is of greater quality or greater relevance. In this thesis, we assume that all the nodes in the wireless sensor network are identical to simplify analysis [25].

1.3 Literature Review

Automatic modulation classification (AMC) is widely used in various commercial and military applications. With decades of research, many classification algorithms have been proposed and studied. They can be classified into two main classes of modulation classification algorithms: likelihood function based algorithms and features based algorithms. As to the first class of algorithms, all the methods which belong to this class are based on the likelihood function which is usually a function of transmitted symbols as well as signal and channel parameters. The likelihood function is calculated under each hypothesis (modulation format). Under the assumption of equally likely priors, the decision is made by selecting the modulation format that maximizes the likelihood function. The features based method needs designers to select or design features that could uniquely represent each modulation format. The decision is based on the observation of the values of the proposed set of features. The likelihood function based algorithms are optimal in the Bayesian sense because they minimize the probability of error of classification. However, because of channel conditions or the absence of knowledge of parameters, sometimes the evaluation of likelihood function could become very messy, which makes the likelihood based method hard to implement. Compared to the optimal one, a suboptimal method with reasonable computation complexity becomes a good choice. Features-based method [6], [11] and [13] could also obtain near-optimal performance with much simpler implementation if it is designed well.

In this thesis, we will focus on the likelihood based (LB) algorithm. For LB algorithms, channel and signal parameters are usually unknown to the receivers. In such a scenario, the receivers are further required to be capable of recovering these unknown parameters in order to compute the likelihood function. We propose a new optimization method based on expectation-maximization algorithm that can increase the efficiency of estimation. Automatic modulation classification becomes a challenging task especially in non-cooperative environments where there is no prior knowledge of the system model available to the re-

ceivers in addition to the time-varying, multipath propagation and frequency-selectivity nature of the channel.

However, the fact that there is a finite pool of candidate modulation formats available makes the problem tractable. Here, we introduce several LB algorithms that are commonly used and are also discussed in this thesis. Depending on the model chosen for unknown parameters, the algorithms are classified and proposed as follows: averaged likelihood ratio test (ALRT), generalized likelihood ratio test (GLRT), hybrid likelihood ratio test (HLRT) and also quasi hybrid likelihood ratio test (QHLRT) [7]. As to ALRT, it treats every unknown parameter as a random variable with a certain known probability density function (PDF). The likelihood function $LF(R|H_i)$ under each hypothesis is averaged with respect to each parameter using its PDF. It is worth noting that these parameters are assumed to be independent from each other. The averaged likelihood functions under each hypothesis are employed for classification which makes it a composite hypothesis testing problem with multiple hypotheses. The decision *D* could be obtained through, $D = arg\max_i E[LF(R|H_i)]$, where H_i is the *i*th hypothesis and *R* is the received data, $E[\star]$ is the expectation operator. We should notice that ALRT requires a multi-dimensional integration which may be difficult to calculate. In GLRT, all unknown quantities including transmitted data symbols are obtained via a multi-dimensional maximization scheme. Regardless of the computational complexity problem, the maximization over the unknown data symbols in GLRT can lead to the same value of the LF for nested signal constellations [7], which in turn leads to incorrect classification. To overcome this problem, an algorithm called hybrid likelihood ratio test (HLRT) has been introduced [14] which combines ALRT with GLRT. Through averaging of data symbols with assumed PDF of constellation points, HLRT does not suffer from the nested constellations problem. Other unknown parameters still need to be replaced by their maximum likelihood estimates (MLEs). As we mentioned above, the multi-dimensional optimization approach is often difficult to implement with a reasonable efficiency. So some non-MLE estimation methods have been

proposed by researchers to avoid heavy computational workload forming the so-called Quasi-HLRT (QHLRT) scheme [14]. Moments based estimators are very commonly used when researchers are looking for efficient estimation schemes for unknown parameters in the QHLRT scheme. The Cramer Rao Lower Bound (CRLB) of the unknown parameters could be used as a lower bound for the variances of proposed estimators. A good unbiased estimator should have a variance that approaches the CRLB as SNR or the size of data increases. To replace the unknown parameters with random variables with means that are equal to their true values and variances that are equal to CRLBs, the performance based on such a likelihood function could be treated as an upper bound for classification schemes with estimates including GLRT, HLRT as well as QHLRT [1], [7].

One of the main contributions of this thesis is that we apply a likelihood function based estimation and detection method for automatic modulation classification in a multi-sensor communication environment. In Chapter 3, we compare the Chair-Varshney fusion rule with the Majority rule in sensor networks with different sizes to see if it is always necessary to use the Chair-Varshney rule. Further, the papers mentioned above [21] - [25] do not sufficiently consider the non-coherent scenario in which the sensors or the signal receivers do not have knowledge of the channel and signal parameters which are crucial in calculation of likelihood functions. In the non-coherent case, the preprocessing of the incoming data is a necessary and challenging task for receiver designers since the accuracy of the parameter estimators directly determines the performances of classifiers. We note that some researchers have studied the possibility of applying existing estimation methods in non-coherent modulation classification. However, their research was only based on single sensor scenario [7], [9]. Compared to the previous work [1], [21], this thesis focuses on both coherent case and non-coherent modulation classification through the study of multiple estimation methods in a multi-sensor network.

1.4 Outline

In Chapter 2, we formulate the problem and describe some preliminary concepts. In Chapter 3, we introduce the ALRT based decision making scheme and distributed decision fusion with the extended Chair-Varshney fusion rule. Also, we compare the performance of the Chair-Varshney fusion rule with the Majority fusion rule in the coherent multi-sensor multi-hypothesis scenario. In Chapter 4, we consider the parameter estimation problem for the centralized data fusion scheme. We first introduce the Maximum Likelihood Estimation approach. After that, we propose an Expectation-Maximization Algorithm based MLE method that has high computational efficiency. Considering both the synchronous and the asynchronous sampling cases, the Cramer-Rao Lower Bounds (CRLBs) for both the single and the multiple sensor scenarios are calculated numerically. The performance of classifiers with the proposed EM estimators and the existing moments based estimators are evaluated and further compared with the classifier with perfect sensing of parameters in a Rayleigh fading channel environment. Finally, the main contributions of the thesis are summarized in Chapter 5.

CHAPTER 2
PROBLEM FORMULATION AND DESCRIPTION

2.1 Signal Model

In this thesis, we consider the modulation classification problem that employs extracted information from the baseband signal. The baseband time-domain signal with additive white Gaussian noise (AWGN) could be modeled as the summation of the signal sequence and the corresponding noise sequence. The mathematical expression of the baseband noisy signal is given by [7] :

$$r(t) = s(t) + n(t), \tag{2.1}$$

where $s(t)$ is the noise-free baseband complex envelope of the modulated data symbol sequence. The channel noise function is denoted as $n(t)$. A comprehensive form of the noiseless signal $s(t)$ is modeled as follows [8] :

$$s(t, \vec{u_i}) = \alpha \sum_{n=0}^{N-1} s_n^i e^{j(\omega_c t + \theta_c)} e^{j\varphi} g(t - nT_s - \epsilon T_s) \quad 0 \leq t \leq NT_s \tag{2.2}$$

In the above equation, some parameters are deterministic unknown variables, while others are random variables with unknown or known distributions. α is the quantity that represents channel gain. In this thesis, the transmission channel is assumed to be a Rayleigh Fading Channel. Instead of a fixed value, the channel gain α follows Rayleigh distribution. s_n^i is the transmitted data symbol selected from symbol constellation i during the nth time interval. The function $g(t) = P_{tx}(t) \bigotimes h(t)$, where $P_{tx}(t)$ is the transmitter pulse shape function, and $h(t)$ is the channel impulse response, and $\bigotimes$ denotes convolution [8]. In a multi-sensor scenario, for different sensors in the network, the channel gain α could vary for different sensors due to the fading channel properties. So we index α by subscript *l*, so that α_l represents the channel gain of the signals received by sensor *l*. As to the carrier waveform, ω_c and θ_c are respectively the residual carrier frequency and the carrier phase. ϵ is the timing offset caused by channel asynchronism. φ is a random variable that describes the carrier phase offset. Similarly, in multi-sensor scenario,φ also differs from sensor to sensor. Further, in this thesis, we assume that the phase offset φ_l is a uniformly distributed random variable with range $[-\pi, \pi]$. Here, we assume there is no phase jitter corresponding to each transmitted symbol. On the receiver side, the simplified baseband signal sequence at the output port of the matched filter is defined as below:

$$\vec{r} = \alpha e^{j\varphi} \vec{s}^{(i)} + \vec{N}, \tag{2.3}$$

which is the sample sequence generated by the matched filter at the given sampling rate $1/T_s$. In the thesis, the averaged power of transmitted symbol s_n is normalized to be unity. $\vec{N}$ is the sequence of complex noise samples. Noise component N_n is defined to be an independent zero-mean complex Gaussian random variable with variance that is equal to N_0. Both the real and the imaginary parts of the noise variable N_n are defined to be i.i.d Gaussian random variable that have the variance equal to $N_0/2$.

In the multiple sensor scenario, the received data are filtered by sensors seperately (2.3).

The noise $\vec{N}$ at each sensor is assumed to be independent and identically distributed complex Gaussian. The noise power N_0 is assumed to be fixed but unknown to the classifiers. Maximum Likelihood Estimation (MLE) and Method of Moments (MoM) estimators are employed for the estimation of N_0 and other unknown quantities. If each sensor is receiving the identical symbols, the *l*th sensor 's baseband sample during the *n*th time interval r_n^l is written as follows [14] :

$$r_n^l = \alpha_l e^{j\varphi_l} s_n^{(i)} + N_n. \tag{2.4}$$

The superscript *l* of r_n^l is the index of the sensor, and α_l is the channel gain of the *l*th sensor, which is a random variable due to channel fading property. Carrier phase offset of the *l*th sensor φ_l is a uniformly distributed random variable. s_n^i is the transmitted symbol during the *n*th time interval under hypothesis *i*. Typically, the per-symbol Signal-to-Noise Ratio (SNR) is defined to be the ratio of signal power and noise power. Given the complex noise sample $N_n \sim N(0, N_0)$, the SNR can be written as α_l^2/N_0 corresponding to the particular sensor *l*. If sensors are receiving different transmitted symbols, $s_n^{l,(i)}$ needs to replace $s_n^{(i)}$ in (2.4).

2.2 Coherent and Non-coherent Environment

When the classifiers work in the coherent environment, the transmitters and the receivers employ handshake with each other. Information about the signal and transmission channel is shared with all the nodes in the sensor network [21]. In the coherent case, modulation classification is easier because the receivers know the nature of the channel and signals including signal amplitudes α_l, noise variance N_0 as well as other needed carrier and channel parameters. The only unknown quantity is the sequence of transmitted symbols under each hypothesis *i*, $\vec{s}^{(i)} = s_1, s_2, \ldots, s_N$. If the detection is not decision-directed or data-aided, we could also average each symbol over the symbol point constellation *i* to obtain the averaged likelihood function [7]. When the classifiers operate in the non-

coherent communication environment, classification becomes complicated for there is no information sharing between the transmitters and the distributed sensors. Before the modulation classification procedure starts, we must have likelihood functions available. The necessary preprocessing of the received data is removing the uncertainty of the unknown quantities embedded in likelihood functions. The applicable data processing methodologies vary. Given a single sensor, we have the sample sequence as R that consists of observations r_n that were sampled during continuous *N* time intervals. Each discrete sample r_n comes from the output port of the pulse-shaping matched filter. The conditional probability density function (PDF) of r_n as one component of PDF of R under hypothesis H_i with unknown signal parameters $\vec{u}_i$ is explicitly defined as :

$$p(r_n|H_i,\vec{u}_i) = \frac{1}{\pi N_0} exp(-\frac{|r_n|^2 + |s_n(\vec{u}_i)|^2 - 2Re(r_n^* s_n(\vec{u}_i))}{N_0}), \tag{2.5}$$

where the operator $Re()$ computes the real part of a complex variable and r_n^* is the conjugate transformation of r_n. The unknown parameters in above equation (2.5) are shown as : $\{\alpha, \varphi, \vec{s}^i, N_0\}$. α is the signal amplitude (averaged transmitting power of symbol is unity) and $\vec{s}^i$ is the transmitted data symbols from the constellation *i*. Besides, the consistent variance of the complex noise sample N_0 is also unknown to the classifier. The difference between the coherent and the non-coherent scenario is reflected in the different unknown parameter vectors $\vec{u}_i$:

$$\text{coherent :} \quad \vec{u}_i = \{\vec{s}_i\}; \tag{2.6}$$

$$\text{non-coherent :} \quad \vec{u}_i = \{\alpha, \varphi, N_0, \vec{s}_i\}. \tag{2.7}$$

Classification under the assumption of coherent environment will be discussed in Chapter 3 and the estimation methods designed for the non-coherent environment will be proposed and evaluated in Chapter 4.

2.3 Likelihood Based Classification Approach

In the thesis, we only focus on the likelihood based (LB) classification and estimation approaches. First, the likelihood based classification approaches are introduced separately under the assumptions of coherent environment and non-coherent environment.

A. Coherent Environment

Under the assumption of coherent environment, the unknown quantities are assumed to be already given except the transmitted data symbols as indicated in (2.7). $\vec{s}_i$ consists of $s_n^{(i)}$ that is the transmitted symbol point from the constellation *i* during the *n*th time interval. In each particular constellation, all points of this constellation are assumed to be equaly likely to be transmitted. Through averaging the probability density function given by (2.5) over constellation *i*, we obtain the averaged single sensor likelihood function $LF(r_n|H_i)$ of each received sample r_n under hypothesis H_i as follows :

$$LF(r_n|H_i) = \frac{1}{M_i}\sum_{k=1}^{M_i}\frac{1}{\pi N_0}exp(-\frac{|r_n|^2 + |\alpha e^{j\varphi}s_n^{k,i}|^2 - 2Re(r_n^*\alpha e^{j\varphi}s_n^{k,i})}{N_0}), \tag{2.8}$$

where $s_n^{k,i}$ is any symbol point from the constellation *i*. Signal amplitude α, phase offset φ, noise power N_0 are all known. The numerical evaluation of the performance of the classifer in the coherent scenario is of importance because it plays the role of the upper bound for the performances of all Likelihood-Based classifications.

B. Noncoherent Environment

In most of the cases, the classifiers work without any prior information provided by the transmitters and the transmitting channel. It is usually called "blind" classification. Under the non-coherent environment, signal amplitude and noise power are both assumed to be fixed but unknown in each observation. There are several methods that we can apply to deal with this problem such as the Averaged likelihood Ratio Test (ALRT) and the Generalized Likelihood Ratio Test (GLRT) as well as HLRT and QHLRT. The difference between

them is how the classifier processes unknown parameters. The mathematical approach to compute ALRT is displayed as below:

$$LF(r_n|H_i) = \int_S \frac{1}{M_i} \sum_{k=1}^{M_i} \frac{1}{\pi N_0} exp(-\frac{|r_n|^2 + |\alpha e^{j\varphi} s_n^{k,i}|^2 - 2Re(r_n^* \alpha e^{j\varphi} s_n^{k,i})}{N_0}) P(\vec{u}|H_i) \mathrm{d}\vec{u}.$$

In the above equation, $\vec{u}$ represents the unknown parameter vector. Because the unknown transmitted data symbols $s_n^{(i)}$ have been processed by averaging over symbol constellation *i*, the vector $\vec{u}_i$ becomes $\vec{u} = \{\alpha, N_0, \varphi\}$. S is the three dimensional value space for the three parameters in $\vec{u}$. Also, it is worth noting that we assume that all the parameters mentioned are uncorrelated with each other. $P(\vec{u}|H_i)$ is the joint conditional probability of $\vec{u}$. In GLRT, all of the unknown quantities are treated as unknown variables with deterministic values and replaced by their maximum likelihood estimates (MLE).

$$LF(r_n|H_i) = \frac{1}{\pi \hat{\sigma}^2} exp(-\frac{|r_n|^2 + |\hat{\alpha} e^{j\hat{\varphi}} \hat{s_n}^{(i)}|^2 - 2Re(r_n^* \hat{\alpha} e^{j\hat{\varphi}} \hat{s_n}^{(i)})}{\hat{N_0}}), \tag{2.9}$$

where $\hat{\alpha}$ $\hat{\varphi}$ and $\hat{N_0}$ are the Maximum Likelihood Estimates (MLE) of α, φ and N_0. Also, $\hat{s_n}^{(i)}$ denotes the estimates of symbols being transmitted during the nth time interval under hypothesis *i*. Considering the accuracy of estimation, maximum likelihood estimate is asymptotically optimal. But the multi-dimensional optimization of likelihood function always brings a heavy computation workload and sometimes it turns out that no closed-form solution is available. The Method of Moments (MoM) estimator is a simpler Non-Data-Aided estimator with an acceptable accuracy. In some papers, Quasi-Hybrid Likelihood Ratio Test (QHLRT) takes place of ALRT and GLRT by using MoM estimators. Because the samples are independent from each other, likelihood function of the sample sequence $LF(R|H_i)$ could be written as the product of likelihood functions of each sample $LF(r_n|H_i)$:

$$LF(R|H_i) = \prod_{n=0}^{N-1} LF(r_n|H_i). \tag{2.10}$$

As a summary, the advantage of using estimators is the reduced computation complexity. But the disadvantage is that the performance of the classifiers is largely dependent on the accuracy of the estimators.

C. Likelihood Ratio Test

After likelihood functions (2.10) of received data R under all hypotheses are available, we move to the decision making step. According to the optimal Bayesian detection scheme, the Likelihood Ratio Test scheme is designed for binary hypothesis testing problem in which we simply choose the hypothesis associated with the larger likelihood. Because of the Bayesian rule, the threshold is the ratio of the two prior probabilities of hypotheses H_i and H_j :

$$\frac{LF(R|H_i)}{LF(R|H_j)} \lessgtr_{H_i}^{H_j} \frac{P_j}{P_i}, \tag{2.11}$$

where P_i and P_j are the prior probability of hypotheses H_i and H_j. Then, taking the logarithm on both sides of (2.11), we have:

$$log\frac{LF(R|H_i)}{LF(R|H_j)} \lessgtr_{H_i}^{H_j} log\frac{P_j}{P_i}. \tag{2.12}$$

If $P_i = P_j$, the threshold $log\frac{P_j}{P_i}$ becomes zero. In the case of multi-hypotheses testing, likelihood ratio test also works via a decision tree. But the maximum likelihood criterion is a more straightforward way:

$$i_{\text{guess}} = arg \max_{i\in\{1,2,\ldots,I\}} LF(R|H_i), \tag{2.13}$$

where *I* is the cardinality of the set of candidate modulation schemes.

CHAPTER 3
LIKELIHOOD BASED DECISION MAKING AND DISTRIBUTED DECISION FUSION FOR MODULATION CLASSIFICATION

In modern sensing and communication systems design, the advantages of the implementation of a large scale sensor network are being more and more emphasized by researchers [4]. In distributed decision making and decision fusion schemes, each sensor is capable of generating its own local decision on the actual hypothesis based on its own observations. The decisions are then fused at a fusion center. In this chapter, we employ such a decision fusion scheme for modulation classification.

3.1 The Classic Chair-Varshney Fusion Rule

As we discussed above, when we have a distributed sensor network with multiple active sensors available, we can expect multiple decisions from distributed sensors. In such

scenario, we assume that each sensor is able to process its own data independently. In other words, they do not have to transfer all their information to the fusion center, which consequently increases the consumption of limited channel bandwidth. Instead, sensors only transmit their decisions to the fusion center. We denote the decision made by the sensor *l* as d_l. The decision vector $\vec{d}$ consists of all of the local sensor decisions so that $\vec{d} = \{d_1, d_2, \ldots, d_L\}$, where *L* is the cardinality of the sensor set. For the binary hypothesis testing case, after the fusion center has received $\vec{d}$, the likelihood functions of decision sequence could be written as [4] :

$$P(H_1|\vec{d}) = \frac{P_1}{P(\vec{d})} \prod_{l \in S_+} (1 - P_{M_l}) \prod_{l \in S_-} P_{M_l} \tag{3.1}$$

$$P(H_0|\vec{d}) = \frac{P_0}{P(\vec{d})} \prod_{l \in S_-} (1 - P_{F_l}) \prod_{l \in S_+} P_{F_l}. \tag{3.2}$$

Then, the optimal fusion rule could be expressed in terms of the decision statistics :

$$log\frac{P(H_1|\vec{d})}{P(H_0|\vec{d})} = log\frac{P_1}{P_0} + \sum_{S_+} log\frac{1 - P_{M_l}}{P_{F_l}} + \sum_{S_-} log\frac{P_{M_l}}{1 - P_{F_l}} \tag{3.3}$$

where P_{M_l} is the probability of a miss (given 1, guess 0) associated with sensor *l*, and P_{F_l} is the probability of a false alarm associated with sensor *l*. These two probability vectors are assumed to be one of the characteristics of the sensors. P_0 and P_1 are the prior probabilities of two hypotheses H_1 and H_0 which are usually assumed to be $1/2$. The sensor set S_+ consists of all the sensors whose decisions d_l = 1 and S_- is the sensor set in which all d_l = 0. Therefore, the Chair-Varshney fusion rule [4] may be expressed as :

$$f(u_1, u_2, \ldots, u_n) = \begin{cases} 1 & log\frac{P(H_1|\vec{d})}{P(H_0|\vec{d})} > 0 \\ 0 & otherwise \end{cases} \tag{3.4}$$

3.2 The Chair-Varshney Fusion Rule for Multiple Hypotheses

The classical Chair-Varshney fusion rule was derived for the binary hypothesis testing case. Actually, the Chair-Varshney rule is still applicable for the multiple hypotheses detection case. The decision fusion algorithm is still based on the likelihood functions of received decision sequence under each hypothesis. In multiple hypotheses case, the likelihood functions have very little difference from (3.2). By using Bayesian Rule, the likelihood function is shown as below :

$$P(H_j|\vec{d}) = \frac{P(H_j,\vec{d})}{P(\vec{d})} \tag{3.5}$$

$$= \frac{P(H_j)}{P(\vec{d})}\prod_{l\in s_1} P(d_l=1|H_j)\ldots\prod_{l\in s_I} P(d_l=I|H_j) \tag{3.6}$$

where $\vec{d}$ is the vector of distributed decisions, and L is the cardinality of the set of involved sensors, and d_l could take values from 1 to *I* where *I* is the cardinality of candidate hypotheses set, and s_i is the set of sensors with decision that are equal to *i*. The expression $P(d_l = i|H_j)$ represents the probability of the l th sensor's making decision d_l being hypothesis H_i given the actual hypothesis H_j. This probability is one element of the $I \times I$ confusion matrix corresponding to sensor l which is shown as follows :

$$\mathbf{P_l} = \begin{pmatrix} P(d_l=1|H_1)\ldots P(d_l=I|H_1) \\ P(d_l=1|H_2)\ldots P(d_l=I|H_2) \\ \vdots \quad \ddots \quad \vdots \\ P(d_l=1|H_I)\ldots P(d_l=I|H_I) \end{pmatrix}. \tag{3.7}$$

The confusion matrix P_l is only associated with sensor l. Further, it is also dependent on Signal-to-Noise Ratio (SNR). So in practice the confusion matrix should be written as the function of sensor index and SNR, as $P(l, SNR)$. After confusion matrix (3.7) is

calculated, we can easily compute $P(H_j|\vec{d})$ via (3.6) for every candidate hypothesis H_j. Next, we finally determine which one is the most likely to be true by using:

$$j_{\text{decision fusion}} = arg \max_{j\in\{1,2,\ldots,I\}} P(H_j|u), \tag{3.8}$$

where $j_{\text{decision fusion}}$ is the data fusion results by using the extended Chair-Varshney fusion rule for multiple hypotheses detection. In the simulation section, we show that the decision fusion approach improves the performance of classification as more sensors become active in the network.

3.3 Majority Fusion Rule

In this section, we introduce another decision fusion rule called the Majority rule. The Majority rule comes from a simple idea that the fusion center goes through every received decision and determines which hypothesis occurs most often among the individual decisions. The fusion center then chooses that hypothesis as the final decision. Assume that we have *L* decisions $\{d_1, d_2, \ldots, d_L\}$, and there are *I* candidate hypotheses to be classified. The Majority rule has two steps. Step *1* is to classify the received decisions into *I* sets :

$$\begin{cases} number_1 = number_1 + 1, & if \quad d_l = 1 \\ number_2 = number_2 + 1, & if \quad d_l = 2 \\ \vdots \qquad\qquad\qquad\qquad \vdots & \\ number_I = number_I + 1, & if \quad d_l = I \end{cases} \tag{3.9}$$

, where $\{number_1, number_2, \ldots, number_I\}$ are counters that we use to record the appearances of each hypothesis. They are all initialized to be zero before the counting begins.

Step *2* is to choose the hypothesis that occures most often as the decision fusion result:

$$j_{\text{final decision}} = arg \max_{i \in 1,2,\ldots,I} number_i \tag{3.10}$$

Due to its simple implementation, Majority fusion rule sometimes becomes a good choice especially for the scenarios that require fast processing speed and have a severe limitation on computational complexity.

3.4 Simulation Results

In the following simulation, the performance of the two given distributed decision fusion rules are evaluated in the coherent communication environment which means there are no unkown parameters and estimations. The modulation classification approach is repeated for 10000 Monte Carlo trials for each value of SNR. The scale of the network varies from 2 sensors to 20 sensors. The equally likely candidate modulation formats are 16-PSK, 16-QAM, 8-PSK and 32-QAM. Each sensor in the network, we assume, is receiving identical observations. So in the simulations, the decision fusion rule is the only characteristic that makes a difference. It could be concluded from the simulation result of the 2-sensor scenario that the performance of Majority fusion rule is pretty close to that of the Chair-Varshney fusion rule. However, when network scales up to 10 sensors and 20 sensors, the degradation of classification performance caused by the use of the suboptimal Majority fusion rule becomes more and more obvious. We then conclude that the Chair-Varshney fusion rule is especially necessary for a large scale sensor network because the difference in detection performance is not negligible.

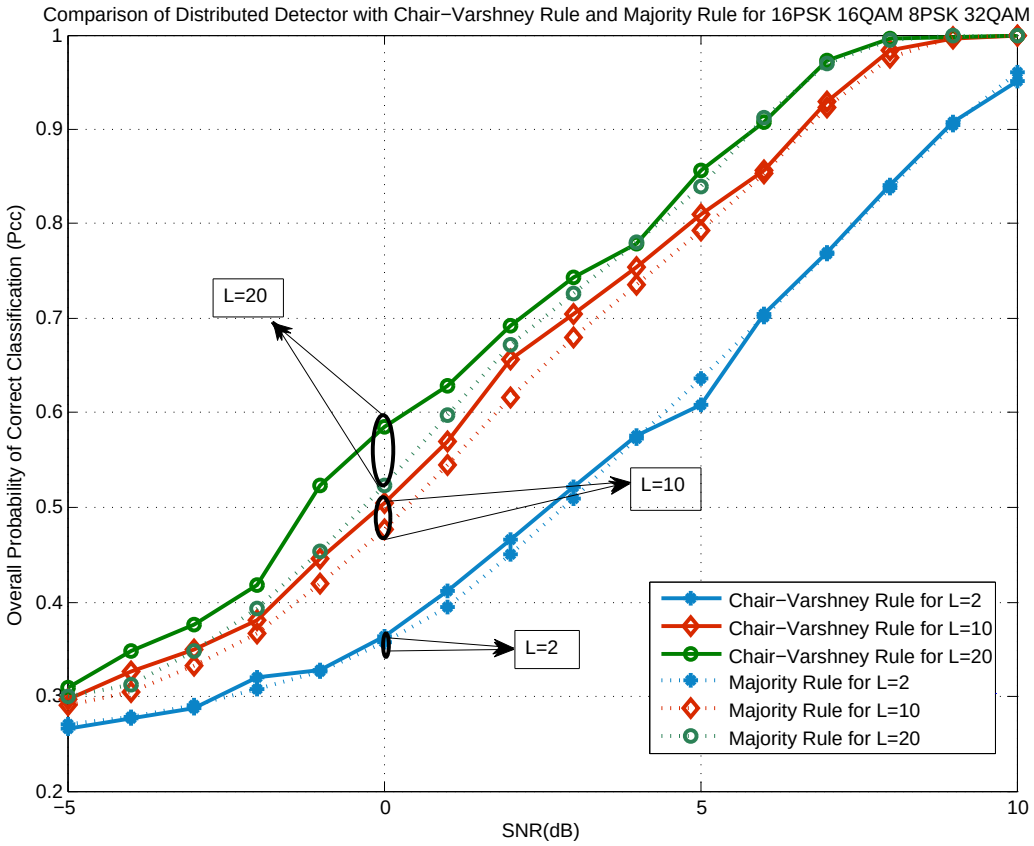

Fig. 3.1: Coherent Distributed Detection and Decision Fusion by using Chair-Varshney rule and Majority rule for the classification of 16PSK, 16QAM, 8PSK, 32QAM

CHAPTER 4
CENTRALIZED DATA FUSION AND EM ALGORITHM BASED ESTIMATION

In a distributed sensor network, besides the distributed detection and decision fusion scheme discussed in Chapter 3, we could also fuse the observations themselves instead of distributed decisions. In centralized data fusion, observations from all the sensors are sent to the fusion center. The final decision is directly dependent on the whole observations set. The disadvantage of centralized data fusion is the greatly increased transmission bandwidth requirement due to the need for transmission of complete observations from the sensors to the fusion center. The formulation of centralized data fusion is exactly identical to that of the single sensor multiple hypotheses testing

$$i_{\text{data fusion}} = arg \max_{i \in \{1,2,\ldots,I\}} LF(R|H_i, \hat{\vec{u}}), \tag{4.1}$$

where the vector $\hat{\vec{u}}$ is the estimate of unknown parameters which does not include transmitted symbols $\vec{s}^{(i)}$. Data symbols need to be averaged over symbol constellation corresponding to the hypothesis which the likelihood function is based on. The observations *R* represent the entire set of observations received from all *L* sensors where *L* is the number

of active sensors in the distributed sensor network.

4.1 Moments Based Estimation

Beside maximum likelihood estimation, we could also accomplish the estimation task by applying other methodologies such as moments-based estimation [18] which is sub-optimal but usually more computationally efficient than maximum likelihood estimation. Moreover, moments-based estimators are Non-Data-Aided (NDA) estimators [12]. So it is naturally a blind estimation technique [23], [24]. Sometimes, moment-based estimation is the only possible solution because of the complexity of the likelihood function. However, in some cases, MLE is still useful if we could simplify the optimization approach. There are some existing moments based estimators we are going to use in the rest of the thesis. Given the second and the fourth order absolute moments of observation r_n [1] :

$$\hat{M}_2 = \frac{1}{N}\sum_{n=1}^{N}|r_n|^2 \tag{4.2}$$

$$\hat{M}_4 = \frac{1}{N}\sum_{n=1}^{N}|r_n|^4 \tag{4.3}$$

We can further write the following moments-based estimators of signal amplitude α and noise power N_0 as follows [1] :

$$\hat{\alpha} = \Big(\frac{2\hat{M}_2^{\,2} - \hat{M}_4}{g^{i^2}T_0^4(2 - \frac{E[|s_n^i|^4]}{g^{i^2}})}\Big)^{\frac{1}{4}} \tag{4.4}$$

$$\hat{N_0} = \hat{M}_2 - \hat{\alpha}^2 \tag{4.5}$$

where $g^{i^2} = E[|s_k^i|^2]$ is the average power of symbol constellation *i* which is normalized to be unity in this thesis. $1/T_0$ is the sampling rate which is also assumed to be unity. For the multiple sensor scenario, given Rayleigh fading transmitting channel, the channel gain *G*

is a random variable with probability density function (PDF) as below:

$$f(G;\sigma) = \frac{G}{N_0} e^{-\frac{G^2}{2N_0}}, \quad where \quad \sigma > 0 \tag{4.6}$$

In multi-sensor scenario, the channal gains corresponding to different sensors are different due to the channel fading property. Assuming that the transmitted symbol is with unit power, the channel gain actually becomes the signal amplitude of the received signal. Because the signal amplitude of sensor l, α_l, is only dependent on lth sensor's observation r_n^l, we have [1] :

$$\hat{\alpha_l} = (\frac{2\hat{M_2^l}^2 - \hat{M_4^l}}{2 - E[|s_n^i|^4]})^{\frac{1}{4}} \tag{4.7}$$

$$\hat{N_{0l}} = \hat{M_2^l} - \hat{\alpha_l}^2, \tag{4.8}$$

where $\hat{M_2^l} = \frac{1}{N}\sum_{n=1}^{N} |r_n^l|^2$ and $\hat{M_4^l} = \frac{1}{N}\sum_{n=1}^{N} |r_n^l|^4$. The final estimate of noise power of the channel is a weighted summation of $\hat{N_{0l}}$.

$$\hat{N_0} = \sum_{l=1}^{L} \frac{\hat{\alpha_l}}{\sum_{k=1}^{L} \hat{\alpha_k}} \times \hat{N_{0l}}. \tag{4.9}$$

For the estimation of the carrier phase offsets, we have different estimators for different modulation formats [15]. As for M-PSK signal, M is the order of the PSK modulation. The moments-based phase offset estimator is [14] :

$$\hat{phi}_{\text{i=M-PSK}} = \frac{1}{M}\angle(\sum_{n=1}^{N} r_n^M). \tag{4.10}$$

For M-QAM signal, moments-based estimate of the carrier phase offset is given as [1] :

$$\hat{phi}_{\text{i=M-QAM}} = \frac{1}{4}\angle(\sum_{n=1}^{N} r_n^4). \tag{4.11}$$

Simulation results for the proposed moments based estimator are provided in the last section of this chapter.

4.2 MLE based Parameter Estimation

4.2.1 Single Sensor MLE

In order to discuss the MLE in this case, let us go back to likelihood function of r_n and display it as below:

$$LF(r_n|H_i,\vec{u}) = \frac{1}{M_i}\sum_{k=1}^{M_i}\frac{1}{\pi N_0}exp(-\frac{|r_n|^2+|\alpha e^{j\varphi}s_n^{k,i}|^2-2Re(r_n^*\alpha e^{j\varphi}s_n^{k,i})}{N_0}). \quad (4.12)$$

Recall equation (2.10) :

$$LF(R|H_i,\vec{u}) = \prod_{n=1}^{N}\frac{1}{M_i}\sum_{k=1}^{M_i}\frac{1}{\pi N_0}exp(-\frac{|r_n|^2+|\alpha e^{j\varphi}s_n^{k,i}|^2-2Re(r_n^*\alpha e^{j\varphi}s_n^{k,i})}{N_0}). \quad (4.13)$$

According to the definition of maximum likelihood estimate, the MLE estimate vector $\hat{\vec{u}} = \{\hat{N}_0,\hat{\alpha},\hat{\varphi}\}$ is the value of the parameter vector $\vec{u}$ that maximizes the likelihood function (4.13) :

$$\hat{\vec{u}} = arg\max_{(N_0,\alpha,\varphi)}\prod_{n=1}^{N}\frac{1}{M_i}\sum_{k=1}^{M_i}\frac{1}{\pi N_0}exp(-\frac{|r_n|^2+|\alpha e^{j\varphi}s_n^{k,i}|^2-2Re(r_n^*\alpha e^{j\varphi}s_n^{k,i})}{N_0}). \quad (4.14)$$

MLE is the solution to this multi-dimensional and (generally) non-convex optimization problem which could become very messy to be solved due to the high computation complexity.

4.2.2 Multiple Sensor MLE

As more sensors become available for the estimators, we expect a performance improvement based on multiple data sources. However, depending on different structures of the network, multi-sensor estimation could be divided into two scenarios. One is centralized estimation and the other one is the decentralized estimation. Intuitively, centralized estimation requires all sensors to transmit the received data to the fusion center where the decision is made. The other is decentralized scheme where all the sensors only transmit their estimation results to the fusion center instead of observations. In this thesis, we mainly focus on centralized estimation. In the centralized structure, all received baseband data is transmitted to the fusion center for processing without any decision making procedure done at the sensors. In this section, we consider the synchronization of sampling between different sensors because the synchronization influences the expression of the likelihood function based on multiple sensor's observations.

A. Synchronous Sampling

If the sensors are synchronously sampling the received data, the received symbol sequences from different sensors are probably overlapping and correlated with each other. In order to simplify the problem, we assume that all baseband data sequence at the sensors are simply identical. Under that assumption, the signal model of *l* th sensor could be shown as:

$$r_n^l = \alpha_l e^{\varphi_l} s_n^{(i)} + n \tag{4.15}$$

where sensor is indexed by *l* and *i* is the index of hypothesis and real/imaginary part of $n \sim N(0, N_0/2)$. The overall likelihood function of multi-sensor observations in *N* time intervals is given as :

$$LF(R|H_i, \vec{u}) = \prod_{n=1}^{N} \frac{1}{M_i} \sum_{k=1}^{M_i} (\frac{1}{\pi N_0})^L exp(-\frac{\sum_{l=1}^{L} |r_n^l - \alpha_l e^{j\varphi_l} s_n^{(i)}|^2}{N_0}) \tag{4.16}$$

In the above equation, α_l is channel gain associated with sensor *l*. Assuming that the averaged power of transmitted symbol is unity, α_l^2 is the averaged signal power of data at sensor *l*. φ_l is the *l*th sensor's corresponding phase offset. N_0 is the variance or averaged power of background noise which is assumed to be identical at different sensors. ML estimate for the synchronous sampling case is based on Equation (4.16). The likelihood function (4.16) is dependent on the unknown parameter vector $\vec{u} = \{\alpha_1, \ldots, \alpha_L, \varphi_1, \ldots, \varphi_L, N_0\}$ which contains 2L+1 unknown parameters. So, the MLE $\hat{\vec{u}}$ becomes the vector which maximizes the likelihood function (4.16). $\hat{\vec{u}}$ is obtained as

$$\hat{\vec{u}} = arg \max_{(\alpha_1,\ldots,\alpha_L,\varphi_1,\ldots,\varphi_L,N_0)} LF(R|H_i, \vec{u}). \tag{4.17}$$

As we can observe, it becomes a 2L+1 dimensional optimization problem for which it is hard to obtain the exact optimal point in the 2L+1 dimensional space. A suboptimal estimator with acceptable computation workload is needed especially in a large-scale sensor network scenario where a large number of unknown quantities are to be estimated.

B. Asynchronous Sampling

When sensors work in the asynchronous mode, the observations of each sensor are likely to depend on totally different transmitted symbols. It is possible that the observations from one sensor are independent from that of other sensor. This property is useful to modify the logarithm of the likelihood function into the form of summation. The likelihood function of the asynchronously sampled data is given as

$$LF(R|H_i, \vec{u}) = \prod_{n=1}^{N} \prod_{l=1}^{L} \frac{1}{M_i} \sum_{k=1}^{M_i} \frac{1}{\pi N_0} exp(-\frac{|r_n^l - \alpha_l e^{j\varphi_l} s_n^{l,(i)}|^2}{N_0}). \tag{4.18}$$

If we take the logarithm on both sides, it would become

$$logLF(R|H_i, \vec{u}) = \sum_{n=1}^{N} \sum_{l=1}^{L} log \frac{1}{M_i} \sum_{k=1}^{M_i} \frac{1}{\pi N_0} exp(-\frac{|r_n^l - \alpha_l e^{j\varphi_l} s_n^{l,(i)}|^2}{N_0}). \tag{4.19}$$

Once again a (2L+1) dimensional optimization would give us the global optimal estimate of parameters $\vec{u}$.

4.3 Expectation Maximization Algorithm based MLE

4.3.1 Expectation Maximization Algorithm

Given the likelihood function $LF(X|\theta)$, the maximum likelihood estimate (MLE) of unknown parameters θ is obtained by :

$$\hat{\theta}_{MLE} = arg \max_{\theta} LF(R|\theta) \tag{4.20}$$

In many cases, the optimization of likelihood function $LF(X|\theta)$ could be very tedious because it usually becomes a multiple dimensional optimization problem. The general grid search method requires high computation workload. Luckily, with some assumptions, the MLE could be simplified by assuming the existence of hidden variables $\vec{Y}$. The likelihood function with hidden variables $\vec{Y}$ involved is developed to be the complete-data likelihood $LF(X, \vec{Y}|\theta)$ and then the former likelihood $LF(X|\theta)$ becomes the incomplete-data likelihood. Instead of maximizing the incomplete-data likelihood which usually is very difficult, we focus on optimization of the complete-data likelihood. That is often easier to compute.

Under the assumption of hidden variable, the incomplete-data likelihood $LF(X, \vec{Y}|\theta)$ could be written as the expectation of the complete-data likelihood. In each realization of complete-data likelihood, the hidden variables are assumed to be known in the sense of posterior probabilities of hidden variables conditioned on received observations X and other unknown quantities. Here we have the expression of the complete-data likelihood after averaging over the hidden variables :

$$LF(X|\theta) = \int_{\vec{Y} \in S} LF(X, \vec{Y}|\theta) P(\vec{Y}|X, \theta^{(r)}) d\vec{Y}, \tag{4.21}$$

in which *S* is the value space of $\vec{Y}$. The hidden variables are unobservable from the received data. But we know the prior probability density distribution of Y no matter what the physical meaning that $\vec{Y}$ have in practice. The prior PDF is given in the form of $P(Y_i = y_i)$. However, considering the independence between hidden variable and channel parameters, we can write the conditional probability of Y_i as $P(Y_i = y_i|\theta) = P(Y_i = y_i)$. In the above expression of $P(y|x, \theta^{(r)})$, $\theta^{(r)}$ means the estimates of parameters after the *r*th round iteration. After we take expectation of the complete-data likelihood, we need to maximize the expectation only with respect to θ regardless of $\theta^{(r)}$. Hopefully, we could successfully obtain the closed-form solutions to the optimization problem. Let us define it as $\hat{\theta}^{(r+1)}$.

$$\hat{\theta}^{(r+1)} = arg\max_{\theta} \int_{\vec{Y}\in S} LF(X, \vec{Y}|\theta) P(\vec{Y}|X, \theta^{(r)}) d\vec{Y} \tag{4.22}$$

It will replace $\theta^{(r)}$ in the posterior probability $P(y|x, \theta^{(r)})$ for the next round of the iteration. The iteration keeps running in this way until we find out that the iteration no longer increases the likelihood function. ϵ which is defined as $\epsilon = LF(X|\hat{\theta}^{(r+1)}) - LF(X|\hat{\theta}^{(r)})$ becomes the threshold that we can adjust to determine when to terminate the iterations. The first round starts from an appropriate initial value set as $\theta^{(0)}$.

4.3.2 EM in Modulation Classification

In this section, we will employ EM algorithm in likelihood based modulation classification [2], [16], [17] and [22]. During parameter estimation in the likelihood function, the hidden variables $\vec{Y}$ consists of Y_n that is the index of the transmitted symbol s_n during the *n*th time interval. The values that could be assigned to Y_n are the symbol's indexes in the symbol constellation corresponding to the hypotheses 16 PSK, 16 QAM, 8 PSK and 32 QAM. Based on (4.21), we have the incomplete-data logarithmic likelihood function as

follows:

$$LLF(R|\theta) = \int_{\vec{Y}\in S} \sum_{n=1}^{N} LF(R_n, Y_n = y|\theta) P(Y_n = y|R_n, \theta^{(r)}) d\vec{Y} \tag{4.23}$$

where *N* is the size of observations we process simultaneously. $\vec{Y}$ is the hidden variable vector which consists of Y_n which represents the index of the symbol transmitted in *n*th time interval. We rewrite (4.23) as follow:

$$LLF(R|\theta) = \sum_{n=1}^{N} \int_{Y_n} LF(R_n, Y_n = y|\theta) P(Y_n = y|R_n, \theta^{(r)}) \mathrm{d}Y_n \tag{4.24}$$

As we know the symbol indexed by Y_n is randomly selected from a symbol constellation corresponding to the specific modulation scheme. Normally we deal with multiple modulation schemes, so the estimation approach is redone under each hypothesis(modulation scheme). Generally we define the order of the candidate modulation format under hypothesis *i* as M_i :

$$LLF(R|\theta) = \sum_{n=1}^{N} \sum_{y=1}^{M_i} LF(R_n = n, Y_n = y|\theta) P(Y_n = y|R_n, \theta^{(r)}). \tag{4.25}$$

Just as we mentioned above, in Expectation-Maximum algorithm based Estimation, there are two steps we need to follow. The first step is the expectation of the likelihood over the hidden variables; the second is the maximization of the expectation of the likelihood. There is one crucial thing we need emphasize that the parameter vector $\theta^{(r)}$ of $P(Y_n = y|R_n, \theta^{(r)})$ is actually the previous estimate (after *r*th round) of these parameters which is treated as a constant in the maximization step of the current round. The θ in complete-data likelihood $LLF(R_n, Y_n = y|\theta)$ is the one that we are going to manipulate to maximize the target likelihood function (4.16).

Here, we just want to inform readers of the motivation of EM algorithm based MLE that we could successfully have a closed- form expression of the MLE given the complete-data

likelihood. Assuming that the estimator knows the transmitted symbol s_n, which means there is no hidden variable $\vec{Y}$, so we can skip the expectation part and directly maximize the complete-data logarithmic likelihood function :

$$\begin{aligned} LLF(R|\theta, H_i, S) &= log \prod_{n=1}^{N} (\frac{1}{\pi N_0})^L exp(-\frac{\sum_{l=1}^{L} |r_n^l - \alpha_l e^{j\varphi_l} s_n|^2}{N_0}) \\ &= -NLlog(\pi N_0) - \frac{\sum_{n=1}^{N} \sum_{l=1}^{L} |r_n^l - \alpha_l e^{j\varphi_l} s_n|^2}{N_0} \end{aligned} \tag{4.26}$$

Then we take the first order derivative of the right hand side of the above expression separately with respect to each element of the unknown parameter vector$[N_0, \alpha_{[1,2,...L]}, \varphi_{[1,2,...L]}]$ and set them to zero. After some simplification, we have the following closed form expressions of ML estimators:

$$\hat{\varphi}_l = arctan(\frac{Im(R^l)Re(S)^T - Re(R^l)Im(S)^T}{Re(R^l)Re(S)^T + Im(R^l)Im(S)^T}), \tag{4.27}$$

$$\hat{\alpha}_l = \frac{Re(R^l)Re(e^{j\hat{\varphi}^l}S)^T + Im(R^l)Im(e^{j\hat{\varphi}^l}S)^T}{\sum_{n=1}^{N} |s_n|^2}, \tag{4.28}$$

$$\hat{N}_0 = \frac{1}{NL} \sum_{n=1}^{N} \sum_{l=1}^{L} |r_n^l - \hat{\alpha}_l e^{j\hat{\varphi}_l} s_n|^2, \tag{4.29}$$

where $Re(\star)$ is the operator that extracts the real part of a complex value. If $\star$ is a vector, $Re(\star)$ is also a vector. Operator $Im(\star)$ obtains the imaginary part. Further, *S* is the symbol vector, and R^l is the received data vector at sensor *l*, and *T* is the transpose sign.

Let us now get back to the scenario where transmitted symbols are unknown and treated as hidden variables in the EM model. Recall (4.25), the maximization comes as a regular Maximum Likelihood Estimation method, then we have the estimate based on EM algo-

rithm, under hypothesis *i*, the EM based MLE $\hat{\theta}_{EM}$ are obtained via :

$$\hat{\theta}_{EM} = arg\max_{\theta} LLF(R|\theta) = arg\max_{\theta} \sum_{n=1}^{N}\sum_{y=1}^{M_i} LLF(R_n, Y_n = y|\theta)P(Y_n = y|R_n, \theta^{(r)}) \tag{4.30}$$

where

$$LLF(R_n, Y_n = y|\theta) = logP(R_n, Y_n = y|\theta) = logP(R_n|Y_n = y, \theta)P(Y_n = y|\theta) \tag{4.31}$$

In this section, we assume that the network only has two sensors active with synchronized sampling scheme. So the conditional probability of the observation R_n during the *n*th time interval under hypothesis *i* is given as below:

$$P(R_n|Y_n = y, \theta) = \frac{1}{(\pi N_0)^2} exp(-\frac{\sum_{l=1}^{2} |r_n^l - \alpha_l e^{j\varphi_l} s(y)^{(i)}|^2}{N_0}) \tag{4.32}$$

where $s(y)^{(i)}$ represents the symbol transmitted under hypothesis *i* during the *n* th time interval as the mapping of the hidden variable Y_n. Further, given $P(Y_n = y|\theta) = \frac{1}{M_i}$, we have the following:

$$LF(R_n, Y_n = y|\theta) = log\frac{1}{(\pi N_0)^2 M_i} exp(-\frac{\sum_{l=1}^{2} |r_n^l - \alpha_l e^{j\varphi_l} s(y)^{(i)}|^2}{N_0}) \tag{4.33}$$

The posterior probability of hidden variable Y_n is defined as :

$$P(Y_n = y|R_n, \theta^{(r)}) = \frac{P(Y_n = y, R_n|\theta^{(r)})}{P(R_n|\theta^{(r)})} = \frac{P(R_n|Y_n = y, \theta^{(r)})P(Y_n = y|\theta^{(r)})}{\sum_{y=1}^{M_i} P(R_n|Y_n = y, \theta^{(r)})P(Y_n = y|\theta^{(r)})} \tag{4.34}$$

Next, we finally have the detailed expression of the posterior probability of hidden variables with knowing that $\theta^{(r)} = \{N_0^{(r)}\alpha_1^{(r)}\alpha_2^{(r)}\varphi_1^{(r)}\varphi_2^{(r)}\}$ are the estimates after the *r*th round

iteration.

$$P(Y_n = y|R_i, \theta^{(r)}) = \frac{\frac{1}{(\pi N_0^{(r)})^2 M_i} exp(-\frac{\sum_{l=1}^{2} |r_n^l - \alpha_l^{(r)} e^{j\varphi_l^{(r)}} s(y)^{(i)}|^2}{N_0^{(r)}})}{\sum_{y=1}^{M_i} \frac{1}{(\pi N_0^{(r)})^2 M_i} exp(-\frac{\sum_{l=1}^{2} |r_n^l - \alpha_l^{(r)} e^{j\varphi_l^{(r)}} s(y)^{(i)}|^2}{N_0^{(r)}})}. \tag{4.35}$$

Then, we can plug (4.33) into (4.25). After that, we have the expression of expectation of complete data likelihood which is also the target function for the maximization step as follows:

$$LF(R|\theta) = \sum_{n=1}^{N} \sum_{y=1}^{M_i} log \frac{1}{(\pi N_0)^2 M_i} exp(-\frac{\sum_{l=1}^{2} |r_n^l - \alpha_l e^{j\varphi_l} s(y)^{(i)}|^2}{N_0}) P(Y_n = y|R_n, \theta^{(r)}). \tag{4.36}$$

Using (4.35), we take the first order derivative with respect to each unknown parameter in the above expression. Then we have the following closed-form expressions of MLE under hypothesis *i*. First, the EM based estimator of the carrier phase offset φ_l after round $r+1$, $\hat{\varphi_l}^{(r+1)}$ =

$$arctan\left(\frac{\sum_{n=1}^{N} \sum_{y=1}^{M_i} P(Y_n = y|R_n, \theta^{(r)})(Im(R_n^l) Re(s(y)^{(i)}) - Re(R_n^l) Im(s(y)^{(i)}))}{\sum_{n=1}^{N} \sum_{y=1}^{M_i} P(Y_n = y|R_n, \theta^{(r)})(Re(R_n^l) Re(s(y)^{(i)}) + Im(R_n^l) Im(s(y)^{(i)}))}\right), \tag{4.37}$$

where $Re(\star)$ is the operator that extracts the real part of a complex number and similarly $Im(\star)$ represents the imaginary part. R_n^l is the observation received by sensor *l* during the *n* th time interval. Further, $\hat{\alpha_l}^{(r+1)}$ =

$$\frac{\sum_{n=1}^{N} \sum_{y=1}^{M_i} P(Y_n = y|R_n, \theta^{(r)})(Re(R_n^l) Re(e^{j\hat{\varphi_l}^{(r+1)}} s(y)^{(i)}) + Im(R_n^l) Im(e^{j\hat{\varphi_l}^{(r+1)}} s(y)^{(i)}))}{\sum_{n=1}^{N} \sum_{y=1}^{M_i} P(Y_n = y|R_n, \theta^{(r)}) |s(y)^{(i)}|^2} \tag{4.38}$$

And

$$\hat{N_0}^{(r+1)} = \frac{1}{2N} \sum_{n=1}^{N} \sum_{y=1}^{M_i} P(Y_n = y|R_n, \theta^{(r)}) \sum_{l=1}^{2} |R_n^l - \hat{\alpha_l}^{(r+1)} e^{j\hat{\varphi_l}^{(r+1)}} s(y)^{(i)}|^2. \tag{4.39}$$

First, $\hat{\varphi_l}^{(r+1)}$ is the MLE of the carrier phase offset associated with sensor *l* after *r+1*th round. $\hat{\alpha_l}^{(r+1)}$ is the MLE of the signal amplitude of sensor *l* after *r+1*th round. We can also see that $\hat{\alpha_l}^{(r+1)}$ depends on the corresponding newest phase offset estimate $\hat{\varphi_l}^{(r+1)}$. Similarly, the *r+1* round estimation of the noise power $\hat{N_0}^{(r+1)}$ is dependent on both $\hat{\varphi_l}^{(r+1)}$ and $\hat{\alpha_l}^{(r+1)}$. Given the above closed-form expressions of EM based MLE, we can start the iteration with reasonable initial values assigned to the estimation vector denoted as $\theta^{(0)}$. So the iteration runs as follows:

1. Plug $\theta^{(0)}$ into (4.37), (4.38) and (4.39) to calculate the first round estimate $\theta^{(1)}$;

2. Sequentially apply (4.37), (4.38) and (4.39) using estimates $\theta^{(r)}$ from the last round– the *r*th round. $\theta^{(r+1)} = \{\hat{N_0}, \hat{\alpha_l}, \hat{\varphi_l}\}_{[l=1,2]}$. Next, update the estimate vector $\hat{\theta}_{EM}$ with the newest results $\theta^{(r+1)}$.

3. If the increase of the likelihood after *r+1* th round the $\epsilon = LF(R|\theta^{(r+1)}) - LF(R|\theta^{(r)})$ $< \epsilon^{\star}$ where $\epsilon^{\star}$ is the threshold we set, the iteration stops; Or keep repeating step 1 and 2.

The final estimates $\hat{\theta}_{EM}$ are actually the local optima of the likelihood function $LF(R|\theta)$. Some bad initial values may lead to wrong estimation results especially when SNR is rather low. However, one obvious advantage of the EM algorithm based estimation is the much lower computational workload compared to other numerical methods aimed at obtaining global optimization.

4.4 Cramer Rao Lower Bounds of Unknown Parameters

As we have discussed in the previous sections, the maximum likelihood estimate usually becomes very difficult to obtain. However, it is still essential to know the performance

of the estimators which could serve as a benchmark. There is a well-known lower bound– Cramer-Rao Lower Bound (CRLB) which evaluates the performance of the estimator in terms of a lower bound on its variance [9]. Moreover, if CRLB is attained by the estimator, it means the estimator is an efficient estimator.

In the likelihood function (4.13) of the single sensor scenario, we have three unknown parameters $\{\alpha, \varphi, N_0\}$ which could also be represented by $\vec{u} = \{\theta_1, \theta_2, \theta_3\}$. For a three-parameter case, the 3×3 Fisher Information Matrix consists of the elements $[J_i]_{p,q}$ which are obtained via :

$$[J_i]_{p,q} = E\left[-\frac{\partial^2 logLF(R|H_i, \vec{u})}{\partial\theta_p \partial\theta_q}\right], \tag{4.40}$$

which is the [p,q] element of the FIM of $\vec{u}$. Once the FIM is obtained, the diagonal elements of its inverse matrix are the CRLBs for the parameter estimation $\hat{\vec{u}}$. The Cramer Rao Lower Bound (CRLB) of *p*th parameter θ_p in vector $\vec{u}$ is shown as below :

$$C_p = [J_i^{-1}]_{p,p}. \tag{4.41}$$

It is worth noting that CRLBs are calculated by using the likelihood function based on hypothesis *i*. That is to say, for different hypotheses, the performance of the same estimator may change, which will directly reflect on the performance of modulation classification.

In the multiple sensor case, we introduce CRLB derivation for two scenarios. In the thesis, we mainly focus on the synchronous sampling scenario in the derivation of CRLBs. We recall the likelihood function of the synchronous sampling scenario as below :

$$LF(R|H_i, \vec{u}) = \prod_{n=1}^{N} \frac{1}{M_i} \sum_{k=1}^{M_i} (\frac{1}{\pi N_0})^L exp(-\frac{\sum_{l=1}^{L} |r_n^l - \alpha_l e^{j\varphi_l} s_n^{(i)}|^2}{N_0}). \tag{4.42}$$

Then we take the logarithm on both sides :

$$logLF(R|H_i, \vec{u}) = \sum_{n=1}^{N} log\frac{1}{M_i} \sum_{k=1}^{M_i} (\frac{1}{\pi N_0})^L exp(-\frac{\sum_{l=1}^{L} |r_n^l - \alpha_l e^{j\varphi_l} s_n^{(i)}|^2}{N_0}) \tag{4.43}$$

After that, we take the second order differential with respect to any two of the unknown parameters in $\vec{u}$, which will be used as elements of the Fisher Information Matrix of $\vec{u}$, $[J_i]_{p,q}$:

$$[J_i]_{p,q} = E\left[-\frac{\partial logLF(R|H_i,\vec{u})}{\partial\theta_p\partial\theta_q}\right] \tag{4.44}$$

$$= \sum_{n=1}^{N} E\left[-\frac{\partial logLF(R_n|H_i,\vec{u})}{\partial\theta_p\partial\theta_q}\right] \tag{4.45}$$

$$= \sum_{n=1}^{N} E\left[-\frac{\partial}{\partial\theta_p\partial\theta_q} log\frac{1}{M_i}\sum_{k=1}^{M_i}(\frac{1}{\pi N_0})^L exp(-\frac{\sum_{l=1}^{L}|r_n^l - \alpha_l e^{j\varphi_l}s_n^{(i)}|^2}{N_0})\right], \tag{4.46}$$

where *i* is the index of hypothesis on which the derivation is based, and θ_p, θ_q are the *p* th and the *q* th parameters of $\vec{u}$. Further details of the derivations and calculations about the Fisher Information Matrix J_i are provided in Appendix A. Through the above derivation, we can see that the Fisher Information Matrix is the summation of the FIMs based on each independent time interval :

$$J_i = \sum_{n=1}^{N} J_i^n. \tag{4.47}$$

Because of averaging over each transmitted symbol in $logLF(R|H_i,\vec{u})$, the FIMs based on the observations of different time intervals are actually identical. In the synchronous sampling case, the FIM could be further rewritten as :

$$J_i = N \times J_i^n. \tag{4.48}$$

When we are dealing with the asynchronous sampling case, independent observations of each time interval give us more flexibility. So recall the likelihood function of asynchronous sampling case, we then take the logarithm on both sides :

$$logLF(R|H_i,\vec{u}) = \sum_{n=1}^{N}\sum_{l=1}^{L} log\frac{1}{M_i}\sum_{k=1}^{M_i}\frac{1}{\pi N_0}exp(-\frac{|r_n^l - \alpha_l e^{j\varphi_l}s_n^{l,k,(i)}|^2}{N_0}). \tag{4.49}$$

If θ_p represents *p*th parameter in $\vec{u}$, the matrix element $[J_i]_{p,q}$ of Fisher Information Matrix could be obtained as :

$$[J_i]_{p,q} = \sum_{n=1}^{N}\sum_{l=1}^{L} E\left[-\frac{\partial}{\partial\theta_p\partial\theta_q} log \frac{1}{M_i}\sum_{k=1}^{M_i}\frac{1}{\pi N_0} exp(-\frac{|r_n^l - \alpha_l e^{j\varphi_l} s_n^{l,k,(i)}|^2}{N_0})\right]. \quad (4.50)$$

It is easy to observe that the FIM under hypothesis *i* could be written as the summation of the FIMs based on the independent observations received by each sensor during each time interval.

$$J_i = \sum_{n=1}^{N}\sum_{l=1}^{L} J_i^{n,l}, \quad (4.51)$$

where $J_i^{n,l}$ is the matrix that consists of elements which are shown as follows :

$$[J_i^{n,l}]_{p,q} = E\left[-\frac{\partial}{\partial\theta_p\partial\theta_q} log \frac{1}{M_i}\sum_{k=1}^{M_i}\frac{1}{\pi N_0} exp(-\frac{|r_n^l - \alpha_l e^{j\varphi_l} s_n^{l,k,(i)}|^2}{N_0})\right]. \quad (4.52)$$

In the asynchronous sampling model, during the same time interval, the observations of sensors are totally independent from each other. For a single sensor, its observations during different time intervals are also independent. Further, the averaging method makes the likelihood function of each single observation sample statistically independent from the exact data symbol that the sensor received during that time interval. In other words, the averaged likelihood functions of each intervals are identical, and they are only dependent on the 2L+1 parameters. Considering the independence between time intervals and between sensors, in the asynchronous sampling case, we rewrite the FIM finally as follows :

$$J_i = N \times L \times J_i^{n,l}, \quad (4.53)$$

, where *i* is the index of the hypothesis on which it is based, and *N* is the number of the time intervals, and *L* is the number of the sensors we have. Both in the synchronous and the asynchronous cases, under hypothesis *i*, by using (4.41), the CRLB of the particular

parameter θ_p which is denoted as C_p could be obtained by extracting the pth diagonal element from the inverse of J_i :

$$C_p = [(J_i)^{-1}]_{p,p}. \tag{4.54}$$

4.5 Simulation Results

In the thesis, we provide the computer simulation results to illustrate the performance of different estimators and modulation classifiers. The calculations of the CRLBs corresponding to the estimation of the five unknown parameters in a two-sensor scenario are based on the 5×5 Fisher Information Matrix. In order to obtain the FIM, we need to take expectation of the second order derivatives of the logarithm of the likelihood functions, which is computationally prohibitive. Due to the heavy computations, we employed numerical approximation of CRLBs to obtain their values. Through the theoretical derivation of CRLB (4.46) and (4.41), the CRLBs are directly dependent on the actual values of these unknown quantities. In the simulation, we assign different values to different sensor's channel gain to show the Rayleigh fading property of the transmission channel.

In Figure 4.1, we did not consider carrier phase offsets and so we only had 3 parameters N_0, A_1, A_2 to estimate. The logarithmic CRLBs of these three parameters are plotted versus the ratio of symbol power and noise power E^2/N_0 which increases from -5 dB to 10 dB. The CRLB of the noise power N_0 is lower than the others because, in our channel model, noise at different sensors has an identical distribution $N(0, N_0)$. This assumption regarding the channel reflects on the estimation by giving more information about the noise parameter such as N_0, which helps the estimation of N_0 perform better. The CRLB of the signal amplitude of different sensor varies because the actual values we assigned to these two amplitude parameters A_1 and A_2 were distinct. The true value of A_1 was 3 and that of A_2 was 1.5 which was only half of A_1. So the sensor with higher signal power can expect a better estimation of its signal amplitude. The logarithmic CRLBs were also plotted versus

the data size N which varies from 50 to 500. We can conclude that when N or SNR is large enough, the CRLBs of all parameters will converge to zero, which means the classification performance will approch that of the coherent case with known SNR.

We further calculated the CRLBs of the two-sensor scenario with carrier phase offsets included. In Figures 4.2 and 4.3, we have 5 parameters $\{N_0, A_1, A_2, \varphi_1, \varphi_2\}$ to estimate. The actual values of these five parameters are $\{1, 3, 1.5, 0.2\pi, -0.4\pi\}$. We compared the Mean Square Error (MSE) with the corresponding CRLB in the scenario of increasing data size N and fixed Signal-to-Noise Ratio (SNR = 0 dB). We can conclude from the figures that for given data size N and SNR, the estimation performances of EM algorithm based maximum likelihood estimators are always better than that of the moments-based estimators when N increases from 100 to 500. Even though we fix the threshold for stopping the iterations of EM based estimation, we could still obtain an improvement of performance when N gets larger regardless of the fact that the absolute value of the likelihood is also increasing with N. Especially for the estimation of phase offsets φ_1, φ_2, the difference between these two estimation approaches are large. It probably makes great difference to plug in estimates from the EM based estimators or the Moments estimators. Because of the definition of CRLB, CRLB of N_0 may not only depend on N_0 itself but also on other parameters that are embedded in the likelihood function. So the CRLBs of the same parameter in 3-parameter and 5-parameter cases are probably different. The simulation results shown by Figures 4.2 and 4.1 verify this characteristic of Cramer Rao Lower Bound. In these figures, we use logarithmic MSE (log MSE) and logarithmic CRLB (log CRLB),and each curve of Figs 4.2 and 4.3 describes the estimation performance of one parameter of $N_0, A_1, A_2, \varphi_1, \varphi_2$.

From the numerical simulation shown by Figure 4.4, we can conclude that expectation maximization algorithm is a very powerful tool for finding local optima especially when the target function is too complicated to maximize directly. The assumption of the existence of hidden variables depends on the particular problem. For this thesis, we selected the

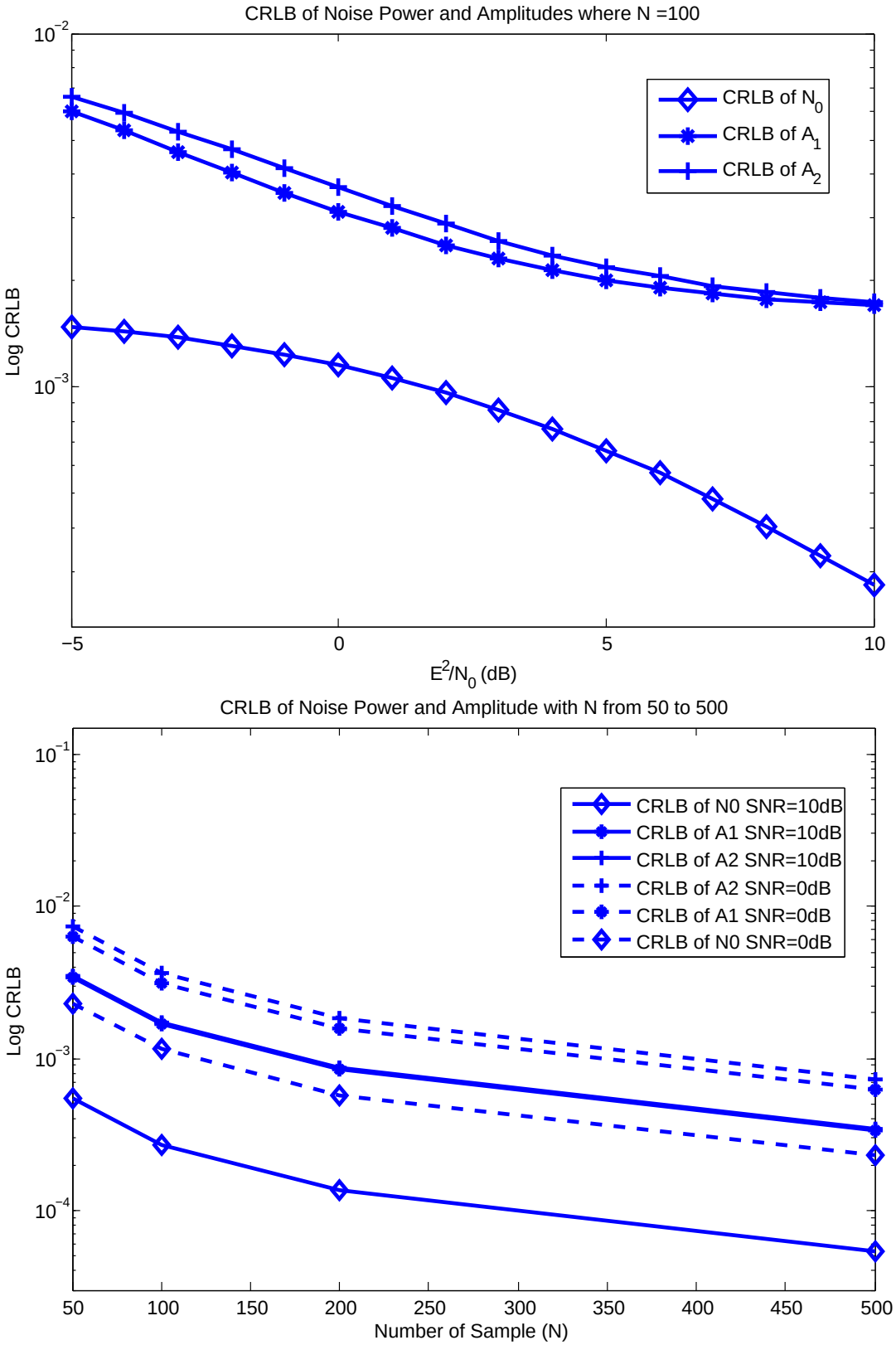

Fig. 4.1: The logarithmic CRLBs of the estimations of noise variance $\hat{N0}$, signal amplitudes $\hat{A}_1$ and $\hat{A}_2$; the actual values of amplitudes are deterministic: A_1 = 3, A_2 =1.5; In the top figure,data size N=100 and the ratio of symbol power and noise power (E^2/N_0) varies from -5dB to 10dB; In the bottom figure, E^2/N_0 = 0dB or 10dB, data size N increases from 50 to 500.

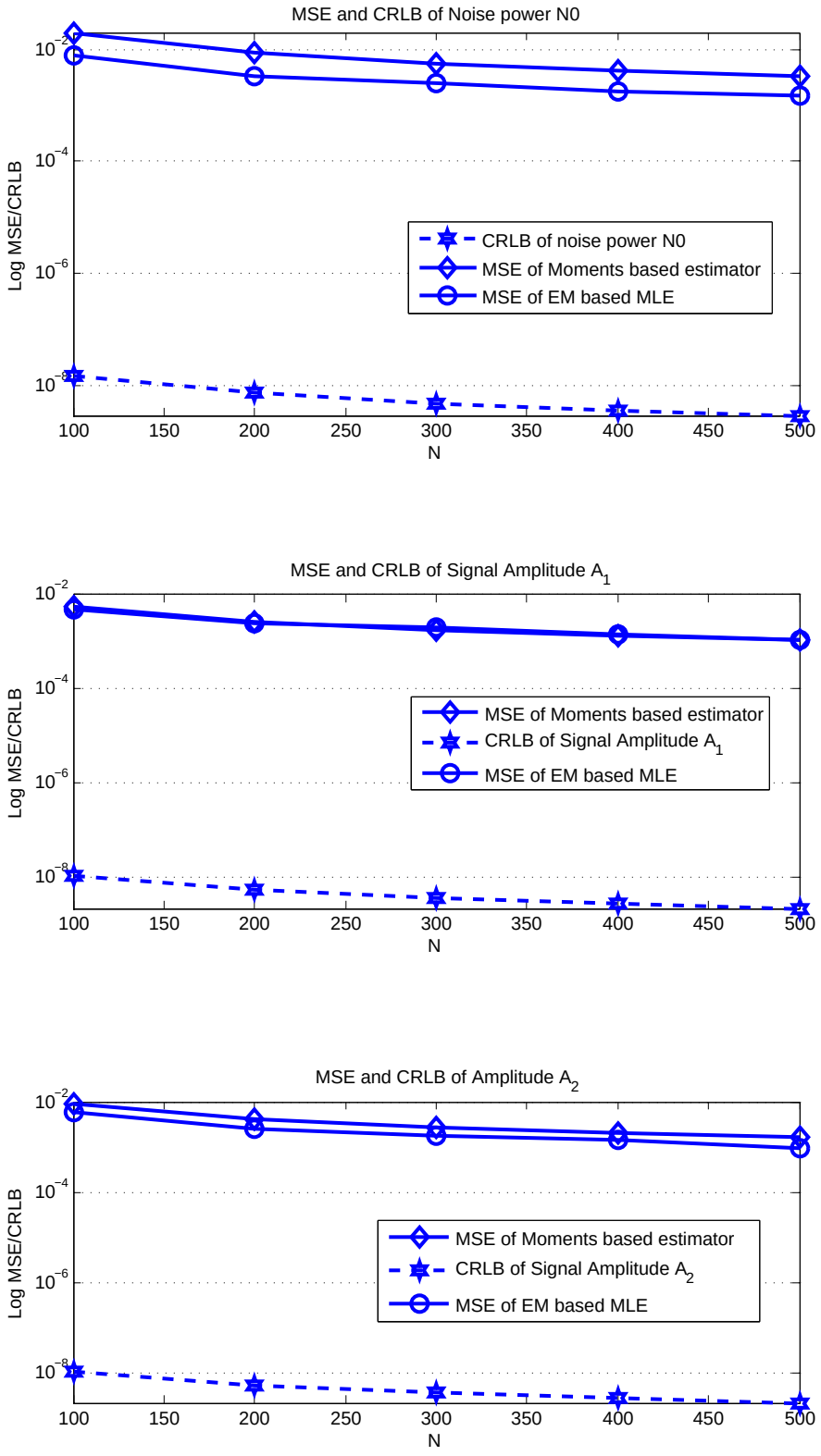

Fig. 4.2: The logarithmic MSE and CRLB of N_0,A_1,A_2,φ_1,φ_2: SNR=0 dB; Data size N increases from 100 to 500; The above three figures are about the logarithmic MSEs and CRLBs of N_0,A_1,A_2.

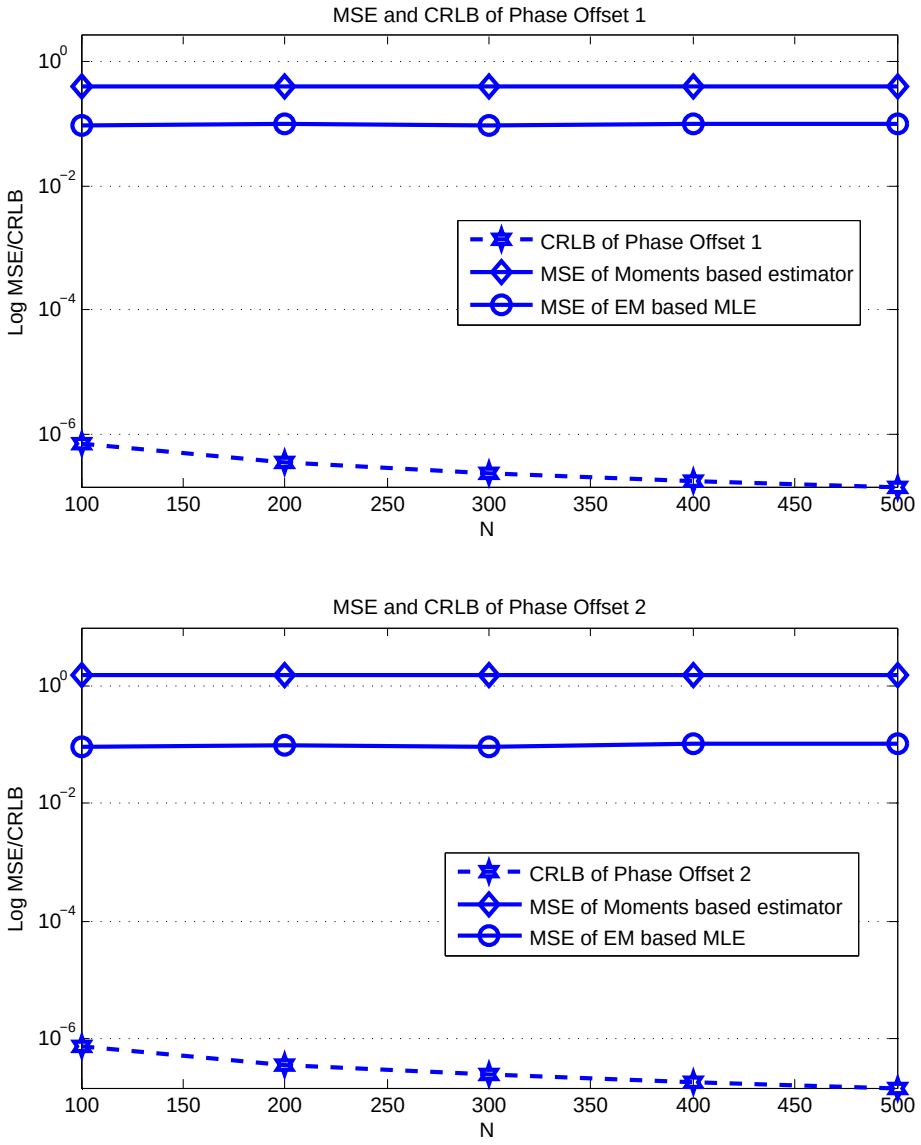

Fig. 4.3: The logarithmic MSE and CRLB of N_0,A_1,A_2,φ_1,φ_2: SNR=0 dB; Data size N increases from 100 to 500; The above two figures are about the logarithmic MSEs and CRLBs of φ_1 and φ_2.

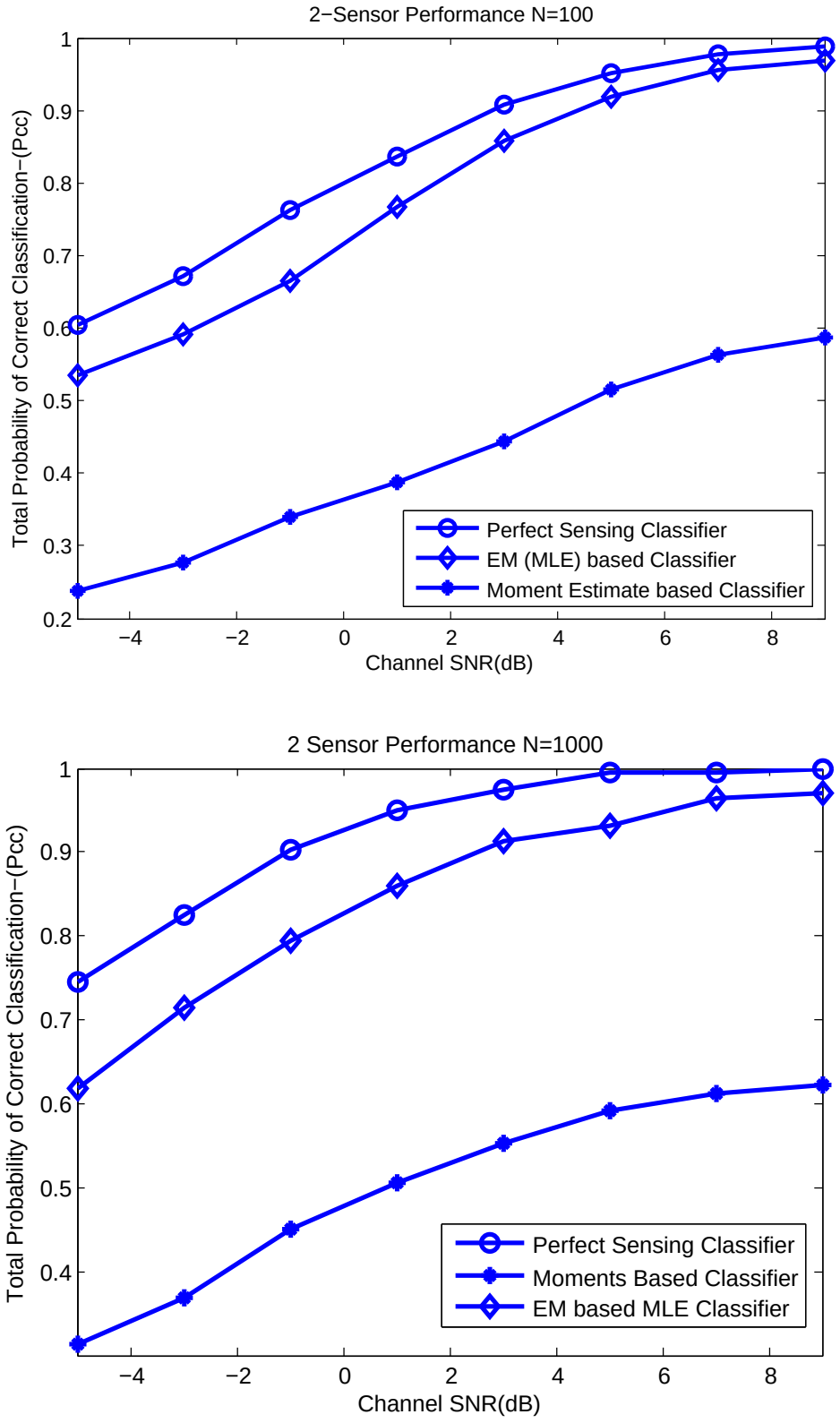

Fig. 4.4: Performance of classifiers equipped with EM based MLE and Moments based Estimates compared with that of perfect sensing case; 16PSK v.s 16QAM; Rayleigh Fading channel with σ =1; Uniformly random carrier phaseoffset; In the top figure,data size N=100; In the bottom figure, N=1000.

transmitted symbols during each time interval T_s as the hidden variable because it fits the definitions of hidden variables and it also has a known probability density function (PDF) corresponding to the current hypothesis. This property makes it convenient to construct the posterior PDF and to average the likelihood function with respect to the hidden variables. In the simulation, the channel gains of sensors are assumed to be Rayleigh distributed random variables and σ equals to one. Carrier phase offsets are also included and assumed to be uniformly distributed in the interval $[-\pi/2, \pi/2]$. In this simulation, the equal-likely candidate modulation formats are only 16PSK and 16QAM.

According to the derivation in Section 4.3, we can clearly see that the setting of the initial values and termination strategy play important roles in determining the efficiency and the accuracy of this method. A bad initial value may mislead the estimator and make it generate wrong estimates because the EM algorithm can only guarantee the solution is optimal locally. In this simulation, we assume that all initial valuses are generated randomly, which means we have no prior knowledge of these unknown quantities. But we have a range for some parameters such as phase offsets. Phase offsets are assumed to be uniformly distributed in $[-\pi/2, \pi/2]$. From the proof of EM algorithm, we know that the iteration should always increase the likelihood function. The likelihood would converge to its maximum value as more iteration rounds we run. But we do not decide how many rounds the simulation is going to run, because we do not know exactly when the estimation will become good enough until we find out the iteration can not increase the likelihood obviously. So we can claim the current estimation of these parameters is final results and we plug in the likelihood function. Next, in the simulation, the threshold for terminating the iteration is designed as: $\epsilon = |LF(R|\theta^{(r)})|/10^6$. The iteration stops as $LF(R|\theta^{(r+1)}) - LF(R|\theta^{(r)}) < \epsilon$, which means when the iteration can no longer increase the likelihood by 0.0001% or larger, the estimations become good enough. It seems a rather tough condition to satisfy, but according to our observations, the iteration usually increases the likelihood in the first several rounds. After that, the improvement of the estimation becomes relatively small and the

curve of the likelihood remains almost flat. What's more, it is worth noting that the channel gain α_l is defined to be a Rayleigh random variable (4.6). Further the channel Signal-to-Noise Ratio is defined by SNR = $2\sigma^2/N_0$ where $2\sigma^2$ is the mean of the Rayleigh random variable α_l. For simplicity, the average noise power N_0 is fixed and equal to one.

Besides the evaluation of the classifier equipped with EM algorithm based MLE, we also presented the performance of the classifier working with moments-based estimators. The moments based estimators that we use in the thesis were introcuded in (4.8), (4.9) and (4.10). In the thesis, all the simulations are done for a multi-sensor scenario. In the particular simulation of centralized data fusion, there are two sensors active for this task. So the fusion center has two incoming sets of complex observations, R_1 and R_2. The received sample sequences are dependent on each other because they were sampled synchronously at sensors. Also, these two classifiers which collaborate with estimations are compared with the one which has perfect sensing ability. It means the system works in the coherent mode (2.7). The performance of such a classifier could serve as an upper bound for performance of modulation classification with unknown channel quantities. Through comparing the distance from the upper bound, we can conclude from Figure 4.4 that the classifier with EM estimators perform much better than the one with moments based estimators and the EM algorithm allows us to make the performance of the classifier as close as possible to the bound if we could employ more data or let more iterations run for the estimation.

CHAPTER 5
SUMMARY AND CONCLUSION

In the thesis, we discussed the following topics: 1) distributed decision making and decision fusion with multiple fusion rules; 2) centralized data fusion with unknown channel quantities and EM algorithm based maximum likelihood estimation.

The distributed modulation classification scheme with parallel decision making and centralized decision fusion was introduced in Chapter 3. In the thesis, we used two fusion rules to accomplish the decision fusion task. One was the Chair-Varshney fusion rule and the other was the Majority Rule. Our simulation results showed the degradation of performance of the classifier equipped with Majority Rule becomes larger as the number of sensors L increased. In other words, when we are working with a larger sensor network, the optimality of the Chair-Varshney fusion rule becomes more obvious. But if the sensor network is relatively small, the Majority fusion rule becomes a good choice due to its simpler implementation.

In Chapter 4, we discussed the scenario that the likelihood functions have unknown parameters involved besides the transmitted symbols. For estimating unknown parameters, we studied several Non-Data-Aided (NDA) estimation methods based on multiple sensor scenario including normal Maximum Likelihood Estimation and Method-of-Moments Estimation. To avoid the computational difficulty of MLE and the theoretical suboptimality

of moments-based estimators, we proposed an expectation maximization algorithm based MLE method. The greatest merit of the proposed approach is that we could have a closed-form expression of the MLE of particular parameters. Even a large group of unknown parameters does not obviously influence the computational efficiency of this method. The demerit of this iteration based estimation is that EM algorithm can only guarantee to find the local optima, which implies that the obtained results may be away from the actual values especially when noise inference is relatively large. In Figure 4.4, we presented the advantages of EM algorithm based MLE.

As an extension to the current work, a sequential detector with dynamic data size is expected to accomplish the modulation classification task with less requirement on the data transmission. Moreover, the SPRT based classifier will make it possible to achieve real-time classification result. Besides, there are a large amount of unknown quantities embedded in the signal model which may also influence the classification or have special physical meanings which are worth estimating. For example, the time delays between sensors may influence the synchronization of sensors, and it will reflect on the form of the likelihood functions we use for classification. Multi-sensor network and data fusion techniques should not only improve sensors' performance of modulation classification but the whole sensing abilities of the network. In this thesis, we only considered the classification with at most four candidate modulation formats. This LB algorithm still may become messy when we have more possible candidates. A new detection approach need to be developed. If we treat the modulation format as one of the unknown quantities embedded in the likelihood function, the EM algorithm could also be considered as one of the candidate solutions.

APPENDIX A
DERIVATION OF MULTIPLE SENSOR CRLB

First let us recall the FIM element $[J_i]_{p,q}$ based on likelihood function of L sensor's distributed observations in N time intervals in the synchronous sampling scenario.

$$[J_i]_{p,q} = \sum_{n=1}^{N} E\left[-\frac{\partial^2}{\partial\theta_p \partial\theta_q} log \frac{1}{M_i} \sum_{k=1}^{M_i} (\frac{1}{\pi N_0})^L exp(-\frac{\sum_{l=1}^{L} |r_n^l - \alpha_l e^{j\varphi_l} s_n^{k,(i)}|^2}{N_0}) \right] \quad \text{(A.1)}$$

which could be rewritten as

$$[J_i]_{p,q} = N \times E\left[-\frac{\partial^2}{\partial\theta_p \partial\theta_q} log \frac{1}{M_i} \sum_{k=1}^{M_i} (\frac{1}{\pi N_0})^L exp(-\frac{\sum_{l=1}^{L} |r_n^l - \alpha_l e^{j\varphi_l} s_n^{k,(i)}|^2}{N_0}) \right] \quad \text{(A.2)}$$

where the expectation could be replaced by a multiple integral with respect to received data sample in *n*th time interval. The (A.2) could be rewritten as

$$\begin{aligned} [J_i]_{p,q} &= N \times [J_i^n]_{p,q} && \text{(A.3)} \\ &= N \int_{\omega} -\frac{\partial^2}{\partial\theta_p \partial\theta_q} log \frac{1}{M_i} \sum_{k=1}^{M_i} (\frac{1}{\pi N_0})^L e^{-\frac{\sum_{l=1}^{L} |r_n^l - \alpha_l e^{j\varphi_l} s_n^{k,(i)}|^2}{N_0}} \mathrm{d}\vec{R_n} && \text{(A.4)} \end{aligned}$$

where $\mathrm{d}\vec{R}_n = \mathrm{d}r_n^1 \mathrm{d}r_n^2 \ldots \mathrm{d}r_n^L$. ω is the *2L* dimensional space. Because, each received noisy data sample consists of the real part and the imaginary part. We need to calculate *2L* integral in order to obtain the value of each FIM element in the *L* sensor scenario. Next let us focus on the differentials inside the above expression(A.4). Recall the parameter vector $\vec{u}$

$$\vec{u} = \{\alpha_1, \alpha_2, \ldots, \alpha_L, \varphi_1, \varphi_2, \ldots, \varphi_L, N_0\}. \tag{A.5}$$

A.1 Derivation of CRLBs in Figure 4.1

If we ignore φ_l parameters at this section and assume that there are only 2 sensors in the network, we will have 3 parameters to estimate. They are $\{\alpha_1, \alpha_2, N_0\}$. So the logarithmic likelihood function is displayed as follows:

$$LLF(r_n|H_i) = log\frac{1}{M_i}\sum_{k=1}^{M_i}(\frac{1}{\pi N_0})^2 exp(-\frac{\sum_{l=1}^{2}|r_n^l - \alpha_l s_n^{k,(i)}|^2}{N_0}) \tag{A.6}$$

Next we derive and display the detailed expression of FIM J_i elements. We start from the element of $[J_i]_{3,3}$

$$[J_i]_{3,3} = N \times [J_i^n]_{3,3} \tag{A.7}$$

$$= N \times \int_\omega -\frac{\partial^2}{\partial N_0^2} LLF(r_n|H_i)\mathrm{d}\vec{R}_n \tag{A.8}$$

$$= N \times \int_\omega -\frac{component_1 \times component_2 - component_3^2}{component_2^2}\mathrm{d}\vec{R}_n, \tag{A.9}$$

Here, we difine a notation that we will often use in the following equations:

$$RA = \sum_{l=1}^{2}|r_n^l - \alpha_l s_n^{k,(i)}|^2 \tag{A.10}$$

where

$$component_1 = \sum_{k=1}^{M_i}(\frac{6}{\pi^2 N_0^4} - \frac{6RA}{\pi^2 N_0^5} + \frac{RA^2}{\pi^2 N_0^6}) \times exp(-\frac{RA}{N_0})$$

$$component_2 = \sum_{k=1}^{M_i}(\frac{1}{\pi N_0})^2 exp(-\frac{RA}{N_0})$$

$$component_3 = \sum_{k=1}^{M_i}(\frac{RA}{\pi^2 N_0^4} - \frac{2}{\pi^2 N_0^3}) \times exp(-\frac{RA}{N_0})$$

For simplicity, we define another notation $partial_\alpha$ as follows:

$$partial_\alpha = \frac{\partial |r_n^l - \alpha_l s_n^{k,(i)}|^2}{\partial \alpha_l} \quad (A.11)$$

$$= 2\alpha_l |s_n^{k,(i)}|^2 - 2(Re(r_n^l)Re(s_n^{k,(i)}) + Im(r_n^l)Im(s_n^{k,(i)})) \quad (A.12)$$

Next, we compute $[J_i]_{3,1}$ and $[J_i]_{3,2}$.

$$[J_i]_{3,l} = N \times [J_i^n]_{3,l} \quad (A.13)$$

$$= N \times \int_\omega -\frac{\partial^2}{\partial N_0 \partial \alpha_l} LF(r_n|H_i) \mathrm{d}\vec{R_n} \quad (A.14)$$

$$= N \times \int_\omega -\frac{component_1 \times component_2 - component_3 \times component_4}{component_2^2} \mathrm{d}\vec{R_n} \quad (A.15)$$

where index *l* = 1 or 2. And: where

$$component_1 = \sum_{k=1}^{M_i}(\frac{2partial_\alpha}{\pi^2 N_0^4} + \frac{RA \times partial_\alpha}{\pi^2 N_0^5} + \frac{partial_\alpha}{\pi^2 N_0^4}) \times exp(-\frac{RA}{N_0})$$

$$component_2 = \sum_{k=1}^{M_i}(\frac{1}{\pi N_0})^2 exp(-\frac{RA}{N_0})$$

$$component_3 = \sum_{k=1}^{M_i}(\frac{RA}{\pi^2 N_0^4} - \frac{2}{\pi^2 N_0^3}) \times exp(-\frac{RA}{N_0})$$

$$component_4 = \sum_{k=1}^{M_i} -\frac{1}{\pi^2 N_0^2} exp(-\frac{RA}{N_0}) \frac{partial_\alpha}{N_0}$$

Then, we calculate $[J_i]_{1,3}$ and $[J_i]_{2,3}$.

$$[J_i]_{l,3} = N \times [J_i^n]_{l,3} \tag{A.16}$$

$$= N \times \int_\omega -\frac{\partial^2}{\partial \alpha_l \partial N_0} LF(R|H_i) \mathrm{d}\vec{R_n} \tag{A.17}$$

$$= N \times \int_\omega -\frac{component_1 \times component_2 - component_3 \times component_4}{component_2^2} \mathrm{d}\vec{R_n} \tag{A.18}$$

,where index *l* equals 1 or 2. And:

$$component_1 = \sum_{k=1}^{M_i}(\frac{2partial_\alpha}{\pi^2 N_0^4} - \frac{RA \times partial_\alpha}{\pi^2 N_0^5} + \frac{partial_\alpha}{\pi^2 N_0^4}) \times exp(-\frac{RA}{N_0})$$

$$component_2 = \sum_{k=1}^{M_i}(\frac{1}{\pi N_0})^2 exp(-\frac{RA}{N_0})$$

$$component_3 = \sum_{k=1}^{M_i} -\frac{1}{\pi^2 N_0^2} exp(-\frac{RA}{N_0}) \frac{partial_\alpha}{N_0}$$

$$component_4 = \sum_{k=1}^{M_i}(\frac{RA}{\pi^2 N_0^4} - \frac{2}{\pi^2 N_0^3}) \times exp(-\frac{RA}{N_0})$$

Next, we calculate $[J_i]_{1,2}$ and $[J_i]_{2,1}$.

$$[J_i]_{l_1,l_2} = N \times [J_i^n]_{l_1,l_2} \tag{A.19}$$

$$= N \times \int_\omega -\frac{\partial^2}{\partial\alpha_{l_1}\partial\alpha_{l_2}} LF(R|H_i)\mathrm{d}\vec{R_n} \tag{A.20}$$

$$= N \times \int_\omega -\frac{component_1 \times component_2 - component_3 \times component_4}{component_2^2}\mathrm{d}\vec{R_n} \tag{A.21}$$

where $[l_1, l_2]$ could be either [1,2] or [2,1]. And:

$$component_1 = \sum_{k=1}^{M_i} \frac{1}{\pi^2 N_0^4} exp(-\frac{RA}{N_0})(2\alpha_{l_2}|s_n^{k,(i)}|^2 - 2(Re(r_n^{l_2})Re(s_n^{k,(i)}) + Im(r_n^{l_2})Im(s_n^{k,(i)}))) \times (2\alpha_{l_1}|s_n^{k,(i)}|^2 - 2(Re(r_n^{l_1})Re(s_n^{k,(i)}) + Im(r_n^{l_1})Im(s_n^{k,(i)})))$$

$$component_2 = \sum_{k=1}^{M_i} (\frac{1}{\pi N_0})^2 exp(-\frac{RA}{N_0})$$

$$component_3 = \sum_{k=1}^{M_i} -\frac{1}{\pi^2 N_0^3} exp(-\frac{RA}{N_0})(2\alpha_{l_1}|s_n^{k,(i)}|^2 - 2(Re(r_n^{l_1})Re(s_n^{k,(i)}) + Im(r_n^{l_1})Im(s_n^{k,(i)})))$$

$$component_4 = \sum_{k=1}^{M_i} -\frac{1}{\pi^2 N_0^3} exp(-\frac{RA}{N_0})(2\alpha_{l_2}|s_n^{k,(i)}|^2 - 2(Re(r_n^{l_2})Re(s_n^{k,(i)}) + Im(r_n^{l_2})Im(s_n^{k,(i)})))$$

Next calculate $[J_i]_{1,1}$ and $[J_i]_{2,2}$.

$$[J_i]_{l,l} = N \times [J_i^n]_{l,l} \tag{A.22}$$

$$= N \times \int_{\omega} -\frac{\partial^2}{\partial \alpha_l^2} LF(R|H_i) \mathrm{d}\vec{R_n} \tag{A.23}$$

$$= N \times \int_{\omega} -\frac{component_1 \times component_2 - component_3^2}{component_2^2} \mathrm{d}\vec{R_n} \tag{A.24}$$

,where index *l* equals 1 or 2.

$$component_1 = \sum_{k=1}^{M_i} \frac{1}{\pi^2 N_0^2} (\frac{partial_\alpha^2}{N_0^2} - \frac{2|s_n^{k,(i)}|^2}{N_0}) exp(-\frac{RA}{N_0})$$

$$component_2 = \sum_{k=1}^{M_i} (\frac{1}{\pi N_0})^2 exp(-\frac{RA}{N_0})$$

$$component_3 = \sum_{k=1}^{M_i} -\frac{1}{\pi^2 N_0^2} exp(-\frac{RA}{N_0}) \frac{partial_\alpha}{N_0}$$

After all elements $[J_i]_{p,q}$ are obtained, we have Fisher Information Matrix *J*. Then we take inverse of *J*– J^{-1}. Cramer Rao Lower Bounds of estimators of Parameter $\{\alpha_1, \alpha_2, N_0\}$ are below:

$$C_{\alpha_1} = [J^{-1}]_{1,1} \tag{A.25}$$

$$C_{\alpha_2} = [J^{-1}]_{2,2} \tag{A.26}$$

$$C_{N_0} = [J^{-1}]_{3,3} \tag{A.27}$$

A.2 Derivation of CRLBs in Figure 4.2 and 4.3

Recall the logarithm likelihood function we are going to deal with :

$$LLF(R|H_i) = \sum_{n=1}^{N} log \frac{1}{M_i} \sum_{k=1}^{M_i} (\frac{1}{\pi N_0})^2 exp(-\frac{\sum_{l=1}^{L} |r_n^l - \alpha_l e^{j\varphi_l} s_n^{k,(i)}|^2}{N_0}). \tag{A.28}$$

where we have the following unknown quantities to determine $\{\alpha_1, ...\alpha_L, \varphi_1, ..., \varphi_L, N_0\}$. For a two-sensor case (L=2), we have five parameters to estimate in the modulation classification. They are $\{N_0, \alpha_1, \alpha_2, \varphi_1, \varphi_2\}$. Reuse equations (A.2) and (A.4), we can obtain the CRLBs. Now we focus on the second order derivatives of the log likelihood function. Matrix J_i is the Fisher Information Matrix under hypothesis *i*. $[J_i]_{(p,q)}$ is the *p*th row and *q*th column element of the FIM.

$$\sum_{l=1}^{L} |r_n^l - \alpha_l e^{j\varphi_l} s_n^{k,(i)}|^2 = \sum_{l=1}^{L} |Re(r_n^l) + jIm(r_n^l) - \alpha_l |s_n^{k,(i)}| e^{j(\theta_s+\varphi_l)}|^2 \tag{A.29}$$

$$= \sum_{l=1}^{L} Re(r_n^l)^2 + Im(r_n^l)^2 + \alpha_l^2 |s_n^{k,(i)}|^2 - 2Re(r_n^l)\alpha_l |s_n^{k,(i)}| cos(\theta_s + \varphi_l)$$

$$= \sum_{l=1}^{L} \big(|r_n^l|^2 + \alpha_l^2 |s_n^{k,(i)}|^2 - 2Re(r_n^l)\alpha_l |s_n^{k,(i)}| cos(\theta_s + \varphi_l) -$$

$$2Im(r_n^l)\alpha_l |s_n^{k,(i)}| sin(\theta_s + \varphi_l)\big)$$

We define some quantities that may be often used in the following derivations:

$$sensor_l = \sum_{l=1}^{L} |Re(r_n^l) + jIm(r_n^l) - \alpha_l |s_n^{k,(i)}| e^{j(\theta_s+\varphi_l)}|^2 \tag{A.30}$$

$$exp_{sensor_l} = exp(-\frac{sensor_l}{N_0}) \tag{A.31}$$

$$partial_{alpha(l)} = 2\alpha_l |s_n^{k,(i)}|^2 - 2Re(r_n^l)|s_n^{k,(i)}| cos(\theta_s + \varphi_l) - 2Im(r_n^l)|s_n^{k,(i)}| sin(\theta_s + \varphi_l)$$

$$partial_{varphi(l)} = 2Re(r_n^l)\alpha_l |s_n^{k,(i)}| sin(\theta_s + \varphi_l) - 2Im(r_n^l)\alpha_l |s_n^{k,(i)}| cos(\theta_s + \varphi_l) \tag{A.32}$$

Now we begin to derive the second order derivatives of the likelihood function. The elements of the FIM is the expectation of these derivatives over received data R.

$$[J_i]_{(1,1)} = \int_R -\frac{\partial^2 logLF(R|H_i)}{\partial N_0^2} dR, \tag{A.33}$$

where

$$\frac{\partial^2 logLF(R|H_i)}{\partial N_0^2} = \sum_{n=1}^{N} \frac{1}{(\sum_{k=1}^{M_i} exp_{sensor_l})^2} \Bigg(\sum_{k=1}^{M_i} exp_{sensor_l} (\frac{2}{N_0^2} - \frac{4sensor_l}{N_0^3} + \frac{sensor_l^2}{N_0^4}) \times \sum_{k=1}^{M_i} exp_{sensor_l} - \sum_{k=1}^{M_i} (exp_{sensor_l} (-\frac{2}{N_0} + \frac{sensor_l}{N_0^2})) \times \sum_{k=1}^{M_i} exp_{sensor_l} \frac{sensor_l}{N_0^2} \Bigg).$$

Further,

$$[J_i]_{(1,l)} = \int_R -\frac{\partial^2 logLF(R|H_i)}{\partial N_0 \partial \alpha_l} dR, \quad l = 1, 2 \tag{A.34}$$

where

$$\frac{\partial^2 logLF(R|H_i)}{\partial N_0 \partial \alpha_l} = \sum_{n=1}^{N} \frac{1}{(\sum_{k=1}^{M_i} exp_{sensor_l})^2} \Bigg(\sum_{k=1}^{M_i} exp_{sensor_l} \times partial_{alpha} \times (\frac{3}{N_0^2} - \frac{sensor_l}{N_0^3}) \times \sum_{k=1}^{M_i} exp_{sensor_l} - \sum_{k=1}^{M_i} (exp_{sensor_l} (\frac{sensor_l}{N_0^2} - \frac{2}{N_0})) \times \sum_{k=1}^{M_i} exp_{sensor_l} \frac{partial_{alpha}}{-N_0} \Bigg)$$

And,

$$[J_i]_{(1,l+3)} = \int_R -\frac{\partial^2 logLF(R|H_i)}{\partial N_0 \partial \varphi_l} dR, \quad l = 1, 2 \tag{A.35}$$

where

$$\frac{\partial^2 logLF(R|H_i)}{\partial N_0 \partial \alpha_l} = \sum_{n=1}^{N} \frac{1}{(\sum_{k=1}^{M_i} exp_{sensor_l})^2} \Bigg(\sum_{k=1}^{M_i} exp_{sensor_l} \times partial_{varphi} \times (\frac{3}{N_0^2} - \frac{sensor_l}{N_0^3}) \times \sum_{k=1}^{M_i} exp_{sensor_l} - \sum_{k=1}^{M_i} (exp_{sensor_l} (\frac{sensor_l}{N_0^2} - \frac{2}{N_0})) \times \sum_{k=1}^{M_i} exp_{sensor_l} \frac{partial_{varphi}}{-N_0} \Bigg)$$

Next

$$[J_i]_{(l+1,l+1)} = \int_R -\frac{\partial^2 logLF(R|H_i)}{\partial \alpha_l^2} dR, \quad l = 1, 2 \tag{A.36}$$

where

$$\frac{\partial^2 logLF(R|H_i)}{\partial \alpha_l^2} = \sum_{n=1}^{N} \frac{1}{(\sum_{k=1}^{M_i} exp_{sensor_l})^2} \Bigg(\sum_{k=1}^{M_i} exp_{sensor_l} (\frac{partial_{alpha}^2}{N_0^2} - \frac{2|s_n^{k,(i)}|^2}{N_0}) \times \sum_{k=1}^{M_i} exp_{sensor_l} - (\sum_{k=1}^{M_i} exp_{sensor_l} \frac{partial_{alpha}}{-N_0})^2 \Bigg)$$

And similarly,

$$[J_i]_{(l+3,l+3)} = \int_R -\frac{\partial^2 logLF(R|H_i)}{\partial \varphi_l^2} dR, \quad l = 1, 2 \tag{A.37}$$

where

$$\frac{\partial^2 logLF(R|H_i)}{\partial \varphi_l^2} = \sum_{n=1}^{N} \frac{1}{(\sum_{k=1}^{M_i} exp_{sensor_l})^2} \Bigg(\sum_{k=1}^{M_i} exp_{sensor_l} (\frac{partial_{varphi}^2}{N_0^2} - \frac{2|s_n^{k,(i)}|^2}{N_0}) \times \sum_{k=1}^{M_i} exp_{sensor_l} - (\sum_{k=1}^{M_i} exp_{sensor_l} \frac{partial_{varphi}}{-N_0})^2 \Bigg)$$

Next

$$[J_i]_{(l_1+1,l_2+1)} = \int_R -\frac{\partial^2 logLF(R|H_i)}{\partial \alpha_{l_1} \partial \alpha_{l_2}} dR, \quad (l_1 = 1, l_2 = 2) or (l_1 = 2, l_2 = 1) \tag{A.38}$$

where $\frac{\partial^2 logLF(R|H_i)}{\partial \alpha_{l_1} \partial \alpha_{l_2}}$

$$= \sum_{n=1}^{N} \frac{1}{(\sum_{k=1}^{M_i} exp_{sensor_l})^2} \Bigg(\sum_{k=1}^{M_i} \frac{exp_{sensor_l}}{N_0^2} \times partial_{alpha}(l = l_1) \times partial_{alpha}(l = l_2) \times \sum_{k=1}^{M_i} exp_{sensor_l} - \sum_{k=1}^{M_i} (exp_{sensor_l} \frac{partial_{alpha}(l = l_1)}{-N_0}) \times \sum_{k=1}^{M_i} exp_{sensor_l} \frac{partial_{alpha}(l = l_2)}{-N_0} \Bigg)$$

Next, as the φ_l

$$[J_i]_{(l_1+3,l_2+3)} = \int_R -\frac{\partial^2 logLF(R|H_i)}{\partial \varphi_{l_1} \partial \varphi_{l_2}} dR, \quad (l_1 = 1, l_2 = 2) or (l_1 = 2, l_2 = 1) \tag{A.39}$$

where $\frac{\partial^2 logLF(R|H_i)}{\partial\varphi_{l_1}\partial\varphi_{l_2}}$ is shown as follows :

$$= \sum_{n=1}^{N} \frac{1}{(\sum_{k=1}^{M_i} exp_{sensor_l})^2} \Bigg(\sum_{k=1}^{M_i} \frac{exp_{sensor_l}}{N_0^2} \times partial_{varphi i}(l = l_1) \times partial_{varphi i}(l = l_2)$$
$$\times \sum_{k=1}^{M_i} exp_{sensor_l} - \sum_{k=1}^{M_i} (exp_{sensor_l} \frac{partial_{varphi i}(l = l_1)}{-N_0}) \times \sum_{k=1}^{M_i} exp_{sensor_l} \frac{partial_{varphi i}(l = l_2)}{-N_0} \Bigg)$$

Further, we have

$$[J_i]_{(l_1+1,l_2+3)} = \int_R -\frac{\partial^2 logLF(R|H_i)}{\partial\alpha_{l_1}\partial\varphi_{l_2}} dR, \quad (l_1 = 1, l_2 = 2) or (l_1 = 2, l_2 = 1) \quad \text{(A.40)}$$

where $\frac{\partial^2 logLF(R|H_i)}{\partial\alpha_{l_1}\partial\varphi_{l_2}}$ is shown as follows :

$$\sum_{n=1}^{N} \frac{1}{(\sum_{k=1}^{M_i} exp_{sensor_l})^2} \Bigg(\sum_{k=1}^{M_i} \frac{exp_{sensor_l}}{N_0^2} \times partial_{alpha}(l = l_1) \times partial_{varphi i}(l = l_2)$$
$$\times \sum_{k=1}^{M_i} exp_{sensor_l} - \sum_{k=1}^{M_i} (exp_{sensor_l} \frac{partial_{alpha}(l = l_1)}{-N_0}) \times \sum_{k=1}^{M_i} exp_{sensor_l} \frac{partial_{varphi i}(l = l_2)}{-N_0} \Bigg)$$

What if l_1 = l_2 :

$$[J_i]_{(l+1,l+3)} = \int_R -\frac{\partial^2 logLF(R|H_i)}{\partial\alpha_l\partial\varphi_l} dR, \quad l = 1, 2 \quad \text{(A.41)}$$

where $\frac{\partial^2 logLF(R|H_i)}{\partial\alpha_l\partial\varphi_l}$ is shown as follows :

$$\sum_{n=1}^{N} \frac{1}{(\sum_{k=1}^{M_i} exp_{sensor_l})^2} \Bigg(\sum_{k=1}^{M_i} (\frac{exp_{sensor_l}}{N_0^2} \times partial_{alpha} \times partial_{varphi i} - \frac{exp_{sensor_l}}{N_0} partial_{varphi i})$$
$$\times \sum_{k=1}^{M_i} exp_{sensor_l} - \sum_{k=1}^{M_i} (\frac{exp_{sensor_l}}{-N_0} \times partial_{alpha}) \times \sum_{k=1}^{M_i} \frac{exp_{sensor_l}}{-N_0} \times partial_{varphi i} \Bigg)$$

It is worth noting that the expression of $[J_i]_{(l+3,l+1)}$ with l=1,2 is the same as (A.40) and

the expression of $[J_i]_{(l_1+3,l_2+1)}$ with l=1,2 is also the same as that of (A.41). Next,

$$[J_i]_{(l+1,1)} = \int_R -\frac{\partial^2 logLF(R|H_i)}{\partial\alpha_l \partial N_0} dR, \quad l = 1, 2 \tag{A.42}$$

where $\frac{\partial^2 logLF(R|H_i)}{\partial\alpha_l \partial N_0}$ is shown as follows :

$$\sum_{n=1}^{N} \frac{1}{(\sum_{k=1}^{M_i} exp_{sensor_l})^2} \Bigg(\sum_{k=1}^{M_i} exp_{sensor_l} (\frac{1}{N_0^2} - \frac{sensor_l}{N_0^3}) \times partial_{alpha} \times \sum_{k=1}^{M_i} exp_{sensor_l} - \sum_{k=1}^{M_i} (\frac{exp_{sensor_l}}{-N_0} \times partial_{alpha}) \times \sum_{k=1}^{M_i} exp_{sensor_l} \frac{sensor_l}{N_0^2} \Bigg)$$

For $[J_i]_{(l+3,1)}$ with l=1,2

$$[J_i]_{(l+3,1)} = \int_R -\frac{\partial^2 logLF(R|H_i)}{\partial\varphi_l \partial N_0} dR, \quad l = 1, 2 \tag{A.43}$$

where $\frac{\partial^2 logLF(R|H_i)}{\partial\varphi_l \partial N_0}$ is shown as follows :

$$\sum_{n=1}^{N} \frac{1}{(\sum_{k=1}^{M_i} exp_{sensor_l})^2} \Bigg(\sum_{k=1}^{M_i} exp_{sensor_l} (\frac{1}{N_0^2} - \frac{sensor_l}{N_0^3}) \times partial_{varphi} \times \sum_{k=1}^{M_i} exp_{sensor_l} - \sum_{k=1}^{M_i} (\frac{exp_{sensor_l}}{-N_0} \times partial_{varphi}) \times \sum_{k=1}^{M_i} exp_{sensor_l} \frac{sensor_l}{N_0^2} \Bigg)$$

After J_i is obtained, we take the inverse of the J_i. The CRLBs of parameters $\{N_0, \alpha_1, \alpha_2, \varphi_1, \varphi_2\}$ are shown as follows :

$$C_{N_0} = [J_i^{-1}]_{(1,1)} \tag{A.44}$$

$$C_{\alpha_1} = [J_i^{-1}]_{(2,2)}, \quad C_{\alpha_2} = [J_i^{-1}]_{(3,3)} \tag{A.45}$$

$$C_{\varphi_1} = [J_i^{-1}]_{(4,4)}, \quad C_{\varphi_2} = [J_i^{-1}]_{(5,5)} \tag{A.46}$$

REFERENCES

[1] A. Abdi, O. Dobre, R. Choudhry, Y. Bar-Ness, and W. Su, "Modulation classification in fading channels using antenna arrays," in *Military Communications Conference, 2004. MILCOM 2004. 2004 IEEE*, vol. 1, oct.-3 nov. 2004, pp. 211 – 217 Vol. 1.

[2] J. A.Bilmes, "A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models," April 1998.

[3] O. Azarmanesh and S. Bilen, "New results on a two-stage novel modulation classification technique for cognitive radio applications," in *MILITARY COMMUNICATIONS CONFERENCE, 2011 - MILCOM 2011*, nov. 2011, pp. 266 –271.

[4] Z. Chair and P. Varshney, "Optimal data fusion in multiple sensor detection systems," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. AES-22, no. 1, pp. 98 –101, jan. 1986.

[5] F. C. Commission, "Spectrum policy task force," Nov. 2003.

[6] K. Davidson, J. Goldschneider, L. Cazzanti, and J. Pitton, "Feature-based modulation classification using circular statistics," in *Military Communications Conference, 2004. MILCOM 2004. 2004 IEEE*, vol. 2, oct.-3 nov. 2004, pp. 765 – 771 Vol. 2.

[7] O. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Survey of automatic modulation classification techniques: classical approaches and new trends," *Communications, IET*, vol. 1, no. 2, pp. 137 –156, april 2007.

[8] O. Dobre, A. Abdi, and Y. W. S. Bar-Ness, "Blind modulation classification: a concept whose time has come," in *Advances in Wired and Wireless Communication, 2005 IEEE/Sarnoff Symposium on*, april 2005, pp. 223 –228.

[9] O. Dobre and F. Hameed, "On performance bounds for joint parameter estimation and modulation classification," in *Sarnoff Symposium, 2007 IEEE*, 30 2007-may 2 2007, pp. 1 –5.

[10] O. A. Dobre and F. Hameed, "Likelihood-based algorithms for linear digital modulation classification in fading channels," in *Electrical and Computer Engineering, 2006. CCECE '06. Canadian Conference on*, may 2006, pp. 1347 –1350.

[11] P. Forero, A. Cano, and G. Giannakis, "Distributed feature-based modulation classification using wireless sensor networks," in *Military Communications Conference, 2008. MILCOM 2008. IEEE*, nov. 2008, pp. 1 –7.

[12] W. Gappmair, R. Lopez-Valcarce, and C. Mosquera, "Joint nda estimation of carrier frequency/phase and snr for linearly modulated signals," *Signal Processing Letters, IEEE*, vol. 17, no. 5, pp. 517 –520, may 2010.

[13] F. Haitao, W. Qun, and S. Rong, "Modulation classification based on cyclic spectral features for co-channel time-frequency overlapped two-signal," in *Circuits, Communications and Systems, 2009. PACCS '09. Pacific-Asia Conference on*, may 2009, pp. 31 –34.

[14] F. Hameed, O. Dobre, and D. Popescu, "On the likelihood-based approach to modulation classification," *Wireless Communications, IEEE Transactions on*, vol. 8, no. 12, pp. 5884 –5892, december 2009.

[15] W. Headley and C. da Silva, "Asynchronous classification of digital amplitude-phase modulated signals in flat-fading channels," *Communications, IEEE Transactions on*, vol. 59, no. 1, pp. 7 –12, january 2011.

[16] L. Hong, "Maximum likelihood bpsk and qpsk classifier in fading environment using the em algorithm," in *System Theory, 2006. SSST '06. Proceeding of the Thirty-Eighth Southeastern Symposium on*, march 2006, pp. 313 –317.

[17] Q. Ma, J. Wang, D. Shasha, and C. Wu, "Dna sequence classification via an expectation maximization algorithm and neural networks: a case study," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 31, no. 4, pp. 468 –475, nov 2001.

[18] U. Mengali and A. N.D'Andrea, *Synchronization Techniques for Digital Receivers*. NY: Plenum Press, 1997.

[19] J. Mitola, "Policy languages for cognitive radio," in *Radio and Wireless Symposium, 2009. RWS '09. IEEE*, jan. 2009, p. 219.

[20] Q. Shi and Y. Karasawa, "Maximum likelihood based modulation classification for unsynchronized qams," in *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, 30 2008-dec. 4 2008, pp. 1 –5.

[21] W. Wei and J. Mendel, "Maximum-likelihood classification for digital amplitude-phase modulations," *Communications, IEEE Transactions on*, vol. 48, no. 2, pp. 189 –193, feb 2000.

[22] Y. Xie and C. Georghiades, "An em-based channel estimation algorithm for ofdm with transmitter diversity," in *Global Telecommunications Conference, 2001. GLOBECOM '01. IEEE*, vol. 2, 2001, pp. 871 –875 vol.2.

[23] H. Xu, Z. Li, and H. Zheng, "A non-data-aided snr estimation algorithm for qam signals," in *Communications, Circuits and Systems, 2004. ICCCAS 2004. 2004 International Conference on*, vol. 1, june 2004, pp. 103 – 107 Vol.1.

[24] H. Xu, D. Zhang, and J. Guo, "New results of snr estimation for high-order modulation," in *Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007. International Conference on*, sept. 2007, pp. 1212 –1215.

[25] J. Xu, W. Su, and M. Zhou, "Software-defined radio equipped with rapid modulation recognition," *Vehicular Technology, IEEE Transactions on*, vol. 59, no. 4, pp. 1659 –1667, may 2010.

VITA

NAME OF AUTHOR: Ruoyu Li

PLACE OF BIRTH: Luoyang, Henan, China

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

Syracuse University, U.S.A August 2010 - expected June 2012

Northwestern Polytechnical University, China August 2006 - July 2010

DEGREES AWARDED:

B.E, Northwestern Polytechnical University, China July 2010

PROFESSIONAL EXPERIENCE:

- Research Assistant, 2011-2012
- Intern Avionics Engineer, 2009-2010