EFFICIENT BLIND SYMBOL RATE ESTIMATION AND DATA SYMBOL DETECTION ALGORITHMS FOR LINEARLY MODULATED SIGNALS

A Thesis

by

SANG WOO PARK

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

May 2008

Major Subject: Electrical Engineering

EFFICIENT BLIND SYMBOL RATE ESTIMATION AND DATA SYMBOL DETECTION ALGORITHMS FOR LINEARLY MODULATED SIGNALS

A Thesis

by

SANG WOO PARK

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee,	Erchin Serpedin
Committee Members,	Jean-François Chamberland
	Aydin Karsilayan
	Eun Jung Kim
Head of Department,	Costas Georghiades

May 2008

Major Subject: Electrical Engineering

ABSTRACT

Efficient Blind Symbol Rate Estimation and
Data Symbol Detection Algorithms
for Linearly Modulated Signals. (May 2008)
Sang Woo Park, B.S., Chung-Ang University, Seoul, Korea
Chair of Advisory Committee: Dr. Erchin Serpedin

Blind estimation of unknown channel parameters and data symbol detection represent major open problems in non-cooperative communication systems such as automatic modulation classification (AMC). This thesis focuses on estimating the symbol rate and detecting the data symbols. A blind oversampling-based signal detector under the circumstance of unknown symbol period is proposed. The thesis consists of two parts: a symbol rate estimator and a symbol detector.

First, the symbol rate is estimated using the EM algorithm. In the EM algorithm, it is difficult to obtain the closed form of the log-likelihood function and the density function. Therefore, both functions are approximated by using the Particle Filter (PF) technique. In addition, the symbol rate estimator based on cyclic correlation is proposed as an initialization estimator since the EM algorithm requires initial estimates. To take advantage of the cyclostationary property of the received signal, there is a requirement that the sampling period should be at least four times less than the symbol period on the receiver side.

Second, the blind data symbol detector based on the PF algorithm is designed. Since the signal is oversampled at the receiver side, a delayed multi-sampling PF detector is proposed to manage inter-symbol interference, which is caused by over-sampling, and to improve the demodulation performance of the data symbols. In the PF algorithm, the hybrid importance function is used to generate both data samples

and channel model coefficients, and the Mixture Kalman Filter (MKF) algorithm is used to marginalize out the fading channel coefficients. At the end, two resampling schemes are adopted.

To my family and friends

ACKNOWLEDGMENTS

Before everything, I express my gratitude to my advisor, Dr. Erchin Serpedin, who enthusiastically guided and supported me while overcoming obstacles during my graduate studies. He has known me well, and he has encouraged me academically. I also greatly appreciate my committee members, Dr. Jean-François Chamberland, Dr. Aydin Karsilayan, and Dr. Eun Jung Kim helping and supporting me.

I express my thanks to people in the Wireless Communications Laboratory (WCL). I especially thank to Kyunglae Noh, Jangsub Kim, and Jaewon Ryu, who have discussed some problems in my research. I also thank all the students of our lab, Jaehan, Huseyin, Qasim, and others.

I appreciate all of my friends in Korea: my oldest friends, Seunghwan, Byungrok, Hyunsoo, etc., my friends in college, Yoonho, Hwan, Jaejin, Yongjun, etc., and my colleague, Juhyun. They have encouraged me to start this valuable journey and to finish it.

I thank my friends in the United States. I am very grateful to Marcus, Ryan, and Eric, who have helped me improve my English as my conversation partners and have hung out together as friends. I also appreciate Chulmin, Soolyeon, and Kyoseung, who have helped me settle down here. I thank all the members of NANTA, the Korean basket ball club. They have made some unforgettable memories for me.

I owe much to my family, my grandmother, mother, father, sister, and mother-in-law. I am deeply grateful to my family. They always love, take cafe of, and support me. Without their help, I would never have overcome this challenge.

Last, I wish to show my final gratitude to my wife, who has always been with me. Her endless love and interest enabled me to complete my work.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

FIGURE Page

CHAPTER I

INTRODUCTION

Recently, non-cooperative communication systems have attracted a lot of attention. Especially in military and civilian application areas, many researchers have focused on systems such as automatic modulation classification (AMC) [1],[2],[3],[4],[5]. Before or after identifying the modulation parameters, estimation of unknown channel parameters represents a major open problem in AMC.

One of the key issues is that of symbol-rate estimation. After modulation classification, the demodulation step requires accurate symbol-rate estimation [6]. Several approaches for symbol-rate estimation have been recently proposed in the literature. A symbol-rate estimator which uses the wavelet transform is suggested in [6]. However, in reference [6], this algorithm is based on the assumption that the transmitted signal has an invariant instantaneous amplitude during each symbol period. This implies that a rectangular pulse shaping filter is used at the transmitter. However, many practical communication systems do not use a rectangular pulse since it requires large bandwidth. Another cyclic correlation (CC) based symbol-rate estimator was proposed in [7] and [8]. Even though the CC-based symbol-rate estimator is very powerful for AMC applications since no prior information is required, the performance of the estimator should be improved for efficient demodulation of data symbols and channel tracking.

In addition to symbol-rate estimation, data symbols should be also blindly detected. In many real-world applications, narrowband mobile communication channels are generally modeled as frequency flat Rayleigh fading channels. To estimate the

---

The journal model is *IEEE Transactions on Automatic Control.*

symbol rate, oversampling, which causes inter-symbol interference (ISI), is used at the receiver side. A lot of research has been reported for signal detection and channel estimation in ISI circumstances. Most of these works rely on techniques such as the maximum likelihood sequence estimation (MLSE) [9], [10]. Since these optimal solutions are based on the Viterbi algorithm and require an additional channel estimation step based on the Kalman Filter for each possible sequence, they entail huge decision delays and computational complexity. Moreover, in conventional MLSE, the metrics of trellis branches are evaluated based on the delayed channel parameter estimates which are updated according to the detected data. Since the data symbol detection is based on delayed estimates of the channel, this method is not suitable for fast fading channels.

To reduce the complexity of MLSE, suboptimal detectors were proposed such as the per-survival sequence detector [11] and [12]. This class of suboptimal detectors is more appropriate for fast fading channels since it avoids delayed channel estimates. However, it has a number of drawbacks. First, it still requires a huge computational complexity since it detects the data symbols based on trellis. Second, it requires a separate channel estimator, which needs preamble symbols to track the channel.

Recently, novel sequential Monte Carlo algorithms, which jointly estimate the channel and detect the data symbols, have been suggested in [13] and [14]. Without compromising the system model, they approximate the optimal solution using sequential Monte Carlo techniques. However, the assumption of known model coefficients is required. In practice, the model coefficients should be estimated in advance. To obtain accurate estimate, a large number of training data is required. Using a blind Particle Filter (PF) detector, Huang et al. [15] suggested an improved algorithm. In [15], the proposed algorithm employs a novel resampling algorithm, which increases the computational complexity, to prevent the error floors caused by the modeling

errors. However, this detector cannot be adopted when the symbol-rate is unknown since the symbol rate estimation generally requires to oversample the received signal. Therefore, this algorithm is not suitable for efficient demodulation of data symbols in AMC.

This thesis proposes a blind oversampling-based signal detector under the circumstance of unknown symbol period to jointly deal with two major issues, mentioned above, in AMC. The proposed algorithm consists of two major parts: a symbol rate estimator and a symbol detector.

First, the symbol rate is estimated using a combination between a CC-based approach and the Expectation Maximization (EM) algorithm, a framework which requires oversampling or fractionally sampling (sampling faster than the symbol rate). One of problems to use the EM algorithm is to obtain an appropriate initial value of the unknown parameter. To figure it out, the CC-based approach is adopted as an initial symbol-rate estimator. The difficulty of obtaining density functions is another problem in the EM algorithm. By using the Particle Filter (PF) algorithm, we not only solve the difficulty of finding the density functions but also reduce the computational complexity in the EM algorithm.

Then, data symbols are detected by using the same PF algorithm based on the oversampled received signal. The oversampling of the received signal improves the performance of data-symbol detectors. In addition, the proposed scheme only requires general resampling steps, which are much simpler than the novel resampling steps proposed in [15]. The PF algorithm employs a modified hybrid importance function [16] and the Mixture Kalman Filter algorithm [13] to reduce its computational complexity. An AR(2) process is used to model the fading channel, and both the AR coefficients and channel coefficients are estimated. Finally, two resampling techniques are adopted and compared in terms of their demodulation performance of

data symbols.

Based on the proposed algorithms for the symbol-rate estimator and the data symbol detector, we iteratively estimate a symbol-rate and detect data symbols.

The rest of this thesis is organized as follows. Basic principles which are associated with the proposed algorithms such as the PF, the cyclic correlations, and the EM are explained in Chapters II-IV. A novel blind symbol-rate estimator and a data-symbol detector are introduced in Chapter V. In Chapter VI, some simulation results are provided to show the performance of efficient demodulation of the received signal. Finally, conclusions and future directions are mentioned in Chapter VII.

CHAPTER II

PARTICLE FILTER

A. Introduction

The Particle Filter (PF) is a sequential Monte Carlo method. Its basic idea is to recursively compute relevant probability distributions using importance sampling and to approximate the probability distributions with discrete random measures [17]. There are some optimal algorithms for Bayesian state estimation such as the Kalman Filter (KF). However, the optimal solutions need some restricted requirements such as Gaussian noise and a linear model. Therefore, several suboptimal algorithms were proposed. The PF algorithms currently represent the most powerful suboptimal algorithms because of their versatility such as parallel implementation. Based on the PF algorithm, any distribution can be approximated by generating samples from the proposal distributions. Moreover, the PF algorithms can be adopted in both linear and nonlinear models.

In this chapter, the PF algorithms are introduced from fundamentals to details. We mainly focus on the general Sequential Importance Sampling (SIS) algorithm.

B. Dynamic Models

In many cases of filtering applications, sequential processing is represented by state space and observation equations as depicted by

$$\begin{aligned} \mathbf{x}_t &= f(\mathbf{x}_{t-1}, \mathbf{v}_t), \\ \mathbf{y}_t &= f(\mathbf{x}_t, \mathbf{e}_t), \end{aligned} \tag{2.1}$$

where $\mathbf{x}_t$ is a state vector, $\mathbf{v}_t$ denotes a processing noise, $\mathbf{y}_t$ stands for an observation vector, $\mathbf{e}_t$ represents a noise measurement, and $f(\cdot)$ is a system transition function. The goal of filtering is to recursively estimate the state vector $\mathbf{x}_t$ given the observation $\mathbf{y}_t$.

C. Monte Carlo Integration

First of all, the concept of Monte Carlo integration will be introduced. Then, it is expanded to the sequential Monte Carlo technique based on the state-space dynamic model (2.1).

Consider the method of approximating a multidimensional integral represented as

$$J = \int h(\mathbf{x})p(\mathbf{x})d\mathbf{x}, \tag{2.2}$$

where $\mathbf{x} \in R^n$, $p(\mathbf{x})$ is a probability density which satisfies $\int p(\mathbf{x})d\mathbf{x} = 1$, and $p(\mathbf{x}) \geq 0$. If we approximate the probability density $p(\mathbf{x})$ as

$$p(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N}\delta(\mathbf{x} - \mathbf{x}^{(i)}), \tag{2.3}$$

J_D, which is the approximation of the multidimensional integral J, is represented as

$$\begin{aligned} J_D &= \int h(\mathbf{x})(\frac{1}{N}\sum_{i=1}^{N}p(\mathbf{x})\delta(\mathbf{x} - \mathbf{x}^{(i)}))d\mathbf{x} \\ &= \frac{1}{N}\sum_{i=1}^{N}h(\mathbf{x}^{(i)}), \end{aligned} \tag{2.4}$$

where $\delta(\cdot)$ is the Dirac delta function, and N is the number of samples drawn from the probability density. Based on the assumption of independent samples, $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}$, and large N, J_D is an unbiased estimate and converges almost surely to J [18]. If the variance of $h(\mathbf{x})$ is finite, then by the central limit theorem [18], the estimation error

converges to the normal distribution,

$$\lim_{N\to\infty} \sqrt{N}(J_D - J) \sim N(0, \sigma^2), \tag{2.5}$$

where

$$\sigma^2 = \int (h(\mathbf{x}) - J)^2 p(\mathbf{x}) d\mathbf{x}. \tag{2.6}$$

In the Bayesian framework, $p(\mathbf{x})$ is chosen as a posterior density. However, a proposal density, which covers the original density, is alternatively adopted in many cases since generating samples from the posterior density is generally intractable.

Consider a known density function $q(\mathbf{x})$, which is called the proposal density function, and where $q(\mathbf{x}) > 0$. Then, the Monte Carlo integral (2.2) is rewritten as

$$\begin{aligned} J &= \int h(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \\ &= \int h(\mathbf{x}) \frac{p(\mathbf{x})}{q(\mathbf{x})} q(\mathbf{x}) d\mathbf{x}. \end{aligned} \tag{2.7}$$

If the samples, $\mathbf{x}^{(1)}$,$\mathbf{x}^{(2)}$,...,$\mathbf{x}^{(N)}$, are drawn from the proposal density $q(\mathbf{x})$, then equation (2.7) is approximated as

$$J_D = \frac{1}{N} \sum_{i=1}^{N} h(\mathbf{x}^{(i)}) \hat{w}(\mathbf{x}^{(i)}), \tag{2.8}$$

with the weights

$$\hat{w}(\mathbf{x}^{(i)}) \propto \frac{p(\mathbf{x}^{(i)})}{q(\mathbf{x}^{(i)})}, \tag{2.9}$$

for $i = 1, \ldots, N$. After the weights are normalized, the equation (2.8) is written as

$$\begin{aligned} J_D &= \frac{\frac{1}{N}\sum_{i=1}^{N} h(\mathbf{x}^{(i)})\hat{w}(\mathbf{x}^{(i)})}{\frac{1}{N}\sum_{i=1}^{N} \hat{w}(\mathbf{x}^{(i)})} \\ &= \sum_{i=1}^{N} h(\mathbf{x}^{(i)}) w(\mathbf{x}^{(i)}), \end{aligned} \tag{2.10}$$

where the normalized weight

$$w(\mathbf{x}^{(i)}) = \frac{\hat{w}(\mathbf{x}^{(i)})}{\sum_{i=1}^{N} \hat{w}(\mathbf{x}^{(i)})}. \tag{2.11}$$

This technique is generally used when generating samples from the original density function is intractable. Instead of directly generating samples from the original density function, we generate samples from the known density function $q(\mathbf{x})$, and assign the weights as in the equation (2.11).

D. Sequential Importance Sampling (SIS)

The Importance Sampling is further improved by sequentially generating samples. The sequential Monte Carlo sampling technique is referred in the literature under different names such as the Particle Filter [19], the Bootstrap Filter [20], and so on. The main idea of these algorithms is that the required posterior density function is approximated by random samples and weights generated from the proposal density function. As the number of samples increases, these algorithms become optimal Bayesian estimators.

Suppose that the posterior density function, $p(\mathbf{x}_{0:t}|\mathbf{y}_{0:t})$, is approximated by the discrete random samples $\{\mathbf{x}_{0:t}^{(i)}, \mathbf{w}_t^{(i)}\}$, where $i = 1, \ldots, N$. The approximation is sequentially updated by drawing new samples $\{\mathbf{x}_t^{(i)}, \mathbf{w}_t^{(i)}\}$ based on the previous samples

$\{\mathbf{x}_{0:t-1}^{(i)}, \mathbf{w}_{t-1}^{(i)}\}$, for $i = 1, \ldots, N$. In detail, the approximated joint posterior density is given by

$$p(\mathbf{x}_{0:t}|\mathbf{y}_{0:t}) \approx \sum_{i=1}^{N} \hat{w}_t^{(i)} \delta(\mathbf{x}_{1:t} - \mathbf{x}_{0:t}^{(i)}), \tag{2.12}$$

where

$$\hat{w}_t^{(i)} \propto \frac{p(\mathbf{x}_{0:t}^{(i)}|\mathbf{y}_{0:t})}{q(\mathbf{x}_{0:t}^{(i)}|\mathbf{y}_{0:t})}. \tag{2.13}$$

Suppose that we approximate the joint posterior density $p(\mathbf{x}_{0:t-1}^{(i)}|\mathbf{y}_{0:t-1})$ with the samples drawn at the $(t-1)^{\text{th}}$ time index. When a new measurement $\mathbf{y}_t$ is observed, we re-approximate the density as $p(\mathbf{x}_{0:t}^{(i)}|\mathbf{y}_{0:t})$. The posterior density function $p(\mathbf{x}_{0:t}|\mathbf{y}_{0:t})$ is factorized such that

$$\begin{aligned} p(\mathbf{x}_{0:t}|\mathbf{y}_{0:t}) &= \frac{p(\mathbf{y}_t|\mathbf{x}_{0:t}, \mathbf{y}_{0:t-1}) p(\mathbf{x}_{0:t}|\mathbf{y}_{0:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{0:t-1})} \\ &= \frac{p(\mathbf{y}_t|\mathbf{x}_{0:t}, \mathbf{y}_{0:t-1}) p(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{0:t-1}) p(\mathbf{x}_{0:t-1}|\mathbf{y}_{0:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{0:t-1})} \\ &\propto p(\mathbf{y}_t|\mathbf{x}_t) p(\mathbf{x}_t|\mathbf{x}_{t-1}) p(\mathbf{x}_{0:t-1}|\mathbf{y}_{0:t-1}). \end{aligned} \tag{2.14}$$

If the proposal density function is factorized as

$$q(\mathbf{x}_{0:t}|\mathbf{y}_{0:t}) = q(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{0:t}) q(\mathbf{x}_{0:t-1}|\mathbf{y}_{0:t-1}), \tag{2.15}$$

then the weights can be expressed as

$$\begin{aligned} \hat{w}_t^{(i)} &\propto \frac{p(\mathbf{x}_{0:t}^{(i)}|\mathbf{y}_{0:t})}{q(\mathbf{x}_{0:t}^{(i)}|\mathbf{y}_{0:t})} \\ &= \frac{p(\mathbf{y}_t|\mathbf{x}_t) p(\mathbf{x}_t|\mathbf{x}_{t-1}) p(\mathbf{x}_{0:t-1}|\mathbf{y}_{0:t-1})}{q(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{0:t}) q(\mathbf{x}_{0:t-1}|\mathbf{y}_{0:t-1})} \\ &= \hat{w}_{t-1}^{(i)} \frac{p(\mathbf{y}_t|\mathbf{x}_t) p(\mathbf{x}_t|\mathbf{x}_{t-1})}{q(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{0:t})} \\ &= \hat{w}_{t-1}^{(i)} \frac{p(\mathbf{y}_t|\mathbf{x}_t) p(\mathbf{x}_t|\mathbf{x}_{t-1})}{q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_t)}. \end{aligned} \tag{2.16}$$

These equations are developed using (2.14) and (2.15). Several density functions such

as a prior density, a posterior density, or a hybrid density [21], can be adopted as the proposal density function. Based on the generated samples and weights from the proposal density function, the joint posterior density function is approximated.

E. Resampling Step

There exists a major problem in the sequential Monte Carlo sampling techniques. The sequential Monte Carlo technique represents an approximation method using discrete random samples. All the samples with negligible assigned weights except a few cause a degeneracy problem. Whenever a significant degeneracy is observed, the resampling is required as a countermeasure. The idea of the resampling step is very simple. The samples, which are assigned small weights, are eliminated and the samples, which are assigned large weights, are duplicated. Finally, the same importance weights, $1/N$, are assigned to all samples. That is,

$$\{\mathbf{x}_t^{(i)}, \hat{w}_t^{(i)}\} \Rightarrow \{\mathbf{x}_t^{(\xi^i)}, \frac{1}{N}\}, \tag{2.17}$$

where $i = 1, \ldots, N$, and ξ^i denotes a newly resampled index.

The time when the resampling step is required is easily determined with the effective sample size N_{eff} and the threshold value N_{thr}, which are introduced in [22]. The estimation of the effective sample size is represented via

$$\hat{N}_{\text{eff}} = \frac{1}{\sum_{i=1}^{N} (w_t^{(i)})^2}, \tag{2.18}$$

where $w_t^{(i)}$ is the normalized weight. N_{eff} is determined between 1 and N since N_{eff} is equal to N when all weights are equally assigned. The pseudcode of the resampling algorithm is depicted in the Table I.

Even though the resampling step improves the performance of the Particle Filter

Table I. Resampling Algorithm

- Assume the samples $\{x_t^{(i)}, w_t^{(i)}\}$, for $i = 1, \ldots, N$
- Define the new samples $\{x_t^{(j)}, w_t^{(j)}\}$ after resampling
- Initialize the cumulative sum of the weight: $c_0 = 0$
- For $i = 1 : N$
    - $c_i = c_{i-1} + w_t^{(i)}$
- End
- Draw an initial point from the uniform distribution: $u_1 \sim U[0, \frac{1}{N}]$
- For $j = 1 : N$
    - $u_j = u_1 + (j-1)/N$
    - While $u_j > c_i$
        - $i = i + 1$
    - End
    - Assign the sample: $x_t^{(j)} = x_t^{(i)}$
    - Assign the weight: $w_t^{(j)} = \frac{1}{N}$
- End

(PF) by removing a degeneracy problem, it increases the correlation among samples and the computational complexity. Therefore, the appropriate employment of the resampling step is required.

F. SIR Filter (Bootstrap Filter)

The Sequential Importance Resampling (SIR) Filter (Bootstrap Filter) was introduced by Gordon, Salmond, and Smith [20]. This algorithm is derived from the SIS Filter by generating samples from a prior density and resampling every time index.

Therefore, the resulting algorithm increases the correlation among samples and the computational complexity while the importance weights and the density are easily calculated.

Consider samples, $\mathbf{x}_t^{(1)}, \ldots, \mathbf{x}_t^{(N)}$, drawn from the prior density $p(\mathbf{x}_t|\mathbf{x}_{t-1}^{(i)})$. Then, the weights are calculated as

$$w_t^{(i)} \propto w_{t-1}^{(i)} p(\mathbf{y}_t|\mathbf{x}_t^{(i)}). \tag{2.19}$$

Since the resampling is carried during every step, the previous weights are always $1/N$. At time index t, the weight is updated by

$$\begin{aligned} w_t^{(i)} &\propto \frac{1}{N} p(\mathbf{y}_t|\mathbf{x}_t^{(i)}) \\ &\propto p(\mathbf{y}_t|\mathbf{x}_t^{(i)}). \end{aligned} \tag{2.20}$$

G. Auxiliary Particle Filter

The Auxiliary Particle Filter (APF) was proposed by Pitt and Shephard [23]. The key idea is that the APF reverses the order of drawing samples and resampling steps. By considering an auxiliary variable κ, where $\kappa \in \{1, \ldots, N\}$, the joint density $p(\mathbf{x}_t, \kappa = i|\mathbf{y}_{1:t})$ is expressed by

$$\begin{aligned} p(\mathbf{x}_t, \kappa = i|\mathbf{y}_{0:t}) &\propto p(\mathbf{y}_t|\mathbf{x}_t) p(\mathbf{x}_t, \kappa = i|\mathbf{y}_{0:t-1}) \\ &= p(\mathbf{y}_t|\mathbf{x}_t) p(\mathbf{x}_t|\kappa = i, \mathbf{y}_{0:t-1}) p(i|\mathbf{y}_{0:t-1}) \\ &= p(\mathbf{y}_t|\mathbf{x}_t) p(\mathbf{x}_t|\mathbf{x}_{t-1}^{(i)}) \hat{w}_{t-1}^{(i)}. \end{aligned} \tag{2.21}$$

The samples $\{\mathbf{x}_t, \kappa\}$ generated from the joint density $p(\mathbf{x}_t^{(i)}, \kappa^{(i)}|\mathbf{y}_{1:t})$ produce a sample $\{\mathbf{x}_t^{(\kappa)}\}$ from the marginalized density $p(\mathbf{x}_t|\mathbf{y}_{1:t})$. If we draw samples $\{\mathbf{x}_t^{(i)}, \kappa^{(i)}\}$

from the proposal density function depicted by

$$q(\mathbf{x}_t, \kappa|\mathbf{y}_{1:t}) \propto p(\mathbf{y}_t|\boldsymbol{\phi}_t^{(i)})p(\mathbf{x}_t|\mathbf{x}_{t-1}^{(i)})\hat{w}_{t-1}^{(i)}, \tag{2.22}$$

where $\boldsymbol{\phi}_t^{(i)}$ is the value related to the density $p(\mathbf{x}_t|\mathbf{x}_{t-1}^{(i)})$ such as a drawn sample, a mean, or a mode. Applying Baye's theorem, the joint proposal density can be expressed as

$$q(\mathbf{x}_t, \kappa|\mathbf{y}_{1:t}) = q(\mathbf{x}_t|\kappa, \mathbf{y}_{1:t})q(\kappa|\mathbf{y}_{1:t}). \tag{2.23}$$

If we define

$$q(\mathbf{x}_t, \kappa|\mathbf{y}_{1:t}) \triangleq p(\mathbf{x}_t|\mathbf{x}_{t-1}^{(i)}), \tag{2.24}$$

then, according to the equations (2.22), (2.23), and (2.24),

$$q(\kappa|\mathbf{y}_{1:t}) \propto p(\mathbf{y}_t|\boldsymbol{\phi}_t^{(i)})\hat{w}_{t-1}^{(i)}. \tag{2.25}$$

According to equation (2.16), the weight is updated by

$$\begin{aligned} \hat{w}_t^{(i)} &\propto \hat{w}_{t-1}^{(\kappa^i)} \frac{p(\mathbf{y}_t|\mathbf{x}_t^{(i)})p(\mathbf{x}_t^{(i)}|\mathbf{x}_{t-1}^{(\kappa^i)})}{q(\mathbf{x}_t^{(i)}, \kappa^{(i)}|\mathbf{y}_t)} \\ &= \frac{p(\mathbf{y}_t|\mathbf{x}_t^{(i)})}{p(\mathbf{y}_t|\boldsymbol{\phi}_t^{(\kappa^i)})}. \end{aligned} \tag{2.26}$$

The original APF presents an additional resampling step [23]. However, we remove the additional resampling step since [24] shows that the last step is unnecessary.

Compared to the conventional PF algorithm, the APF exhibits the advantage of more accurate estimation by resampling one step ahead since the resampling step reflects higher likelihood function which does not include current step's random variables.

H. Kernel Smoothing

The Kernel Smoothing (KS) method was developed and extended by West in [25] and [26]. The smoothing kernel density is given via

$$p(\mathbf{a}|y_{0:t}) \approx \sum_{i=1}^{N} w_t^{(i)} N(\mathbf{a}; \tilde{\mathbf{a}}_t^{(i)}, h^2 \mathbf{V}_t), \tag{2.27}$$

where the smoothing parameter $h > 0$, $N(\cdot; \tilde{\mathbf{a}}, \mathbf{V})$ denotes a multivariate normal density with the mean $\tilde{\mathbf{a}}$ and the covariance matrix $\mathbf{V}$. In the conventional kernel methods, h is chosen as a slowly decreasing function of N. Therefore, the kernel components are more concentrated near $\tilde{\mathbf{a}}_t^{(i)}$ for large N. However, this result causes an over-disperse problem relative to the posterior sample in the sense that the variance matrix of the mixture normals is $(1+h^2)\mathbf{V}_t$, which is larger than the variance matrix of the posterior samples $\mathbf{V}_t$. To correct this problem, West suggested the new idea of shrinkage of kernel location as

$$\tilde{\mathbf{a}}_t^{(i)} = \epsilon \mathbf{a}_t^{(i)} + (1-\epsilon)\bar{\mathbf{a}}_t, \tag{2.28}$$

where $\epsilon = \sqrt{1-h^2}$. Based on the proposed kernel location, the normal mixture density has mean $\bar{\mathbf{a}_t}$ and the correct variance matrix $\mathbf{V}_t$. Therefore, the over-dispersed problem is resolved [25] and [26].

CHAPTER III

CYCLOSTATIONARITY

A. Definitions

A signal having statistical properties which are periodic with time is called a cyclostationary process. This section briefly introduces some definitions of cyclostationarity such as a cyclic mean and cyclic correlations.

Giannakis defined cyclostationarity as follows [27]. First, a mean and a covariance are defined as $\mu(m) \triangleq E[y(m)]$ and $c(m,\tau) \triangleq E[(y(m)-\mu(m))(y(m+\tau)-\mu(m+\tau)]$. Then, the discrete random process $y(m)$ is cyclostationary (CS) if and only if the mean and the covariance have an integer period M. In other words, $\mu(m) = \mu(m+kM)$, and $c(m,\tau) = c(m+kM,\tau)$, $\forall m$, $k \in Z$ where Z is the set of integers. Since they are periodic, they can be represented by Fourier Series expansions over complex harmonic cycles with the set of cycles defined as $F^c \triangleq \{f_k = 2\pi k/M, k = 0,\ldots,M-1\}$. For example, the covariance and its Fourier coefficients, called cyclic correlations, are

$$c\,(m,\tau) = \sum_{k=0}^{M-1} C\left(\frac{2\pi k}{M},\tau\right) e^{j\frac{2\pi kn}{M}}, \tag{3.1}$$

$$C\left(\frac{2\pi k}{M},\tau\right) = \frac{1}{M}\sum_{m=0}^{M-1} c(m,\tau)e^{-j\frac{2\pi km}{M}}. \tag{3.2}$$

In engineering applications, almost periodicity is more common. Therefore, we rather focus on this notion. If the mean and correlations of discrete random process $y(m)$ are almost periodic sequences, then $y(m)$ is defined as almost cyclostationary (ACS). Similar to equations (3.1) and (3.2), the time-varying and cyclic correlations

are defined as

$$c(m,\tau) = \sum_{f_k \in F^c} C(f_k,\tau) e^{j2\pi f_k m}, \tag{3.3}$$

$$C(f_k,\tau) = \lim_{M\to\infty} \frac{1}{M} \sum_{m=0}^{M-1} c(m,\tau) e^{-j2\pi f_k m}, \tag{3.4}$$

respectively, where the set of cycles, $F^c(\tau) = \{f_k : C(f_k,\tau) \neq 0, -\frac{1}{2} < f_k \leq \frac{1}{2})\}$, must be finite, and the limit is assumed to exist at least in the mean-square sense [28].

B. Estimation of Cyclic Statistics

Consider the ACS process with the known cycles f_k. If $y(n)$ has a nonzero mean, then cyclic mean can be estimated as

$$\hat{U}(f_k) = \frac{1}{M} \sum_{m=0}^{M-1} y(m) e^{-j2\pi f_k m}. \tag{3.5}$$

If the set of cycles is finite, we also estimate the time-varying mean as

$$\hat{\mu}(m) = \sum_{f_k} \hat{U}(f_k) e^{j2\pi f_k m}. \tag{3.6}$$

Similarly, for zero-mean CS processes, cyclic correlations and time-varying correlations are estimated, respectively, via

$$\hat{C}(f_k,\tau) = \frac{1}{M} \sum_{m=0}^{M-1} y(m) y(m+\tau) e^{-j2\pi f_k m}, \tag{3.7}$$

$$\hat{c}(m,\tau) = \sum_{f_k \in F^c} \hat{C}(f_k,\tau) e^{j2\pi f_k m}. \tag{3.8}$$

C. Cyclic Correlation Based Symbol Rate Estimators

Cyclic correlation based symbol rate estimator is very simple and powerful to blindly estimate the symbol rates. The classical estimator based on cyclic correlation was

proposed in [29]. However, its performance depends on the excess bandwidth. Therefore, an improved estimator was suggested in [7] and [8]. The key concept of the estimator is to choose the cyclic frequency which maximizes the sum of the square moduli of cyclic correlations.

Assume that $y_c(t)$ denotes the continuous received signal at the output of a fading channel. Signal $y_c(t)$ can be expressed as

$$y_c(t) = \sum_{k=0}^{K-1} b_k h_c(t - kT) + w_c(t), \tag{3.9}$$

where K is the number of the transmitted data symbols, b_k denotes a zero-mean and unit-variance independent and identically distributed (i.i.d.) sequence of symbols, T denotes the symbol period, $h_c(t)$ represents the convolution of the shaping filter with the unknown fading channel, $w_c(t)$ is an additive Gaussian noise. The sampled received signal $y(m)$ is $y_c(mT_s)$, where the sampling period T_s is sufficiently small to satisfy $T_s < T/4$ [8]. If the parameter p_0 is defined by

$$p_0 = \frac{T_s}{T}, \tag{3.10}$$

estimating the symbol period T is equivalent to estimate the parameter p_0. In additions, cyclic frequencies only exist at $-1/T$, 0, and $1/T$ since the signal $y_c(t)$ is bandlimited and cyclostationary. According to equation (3.3),

$$c(m, \tau) = C(0, \tau) + C(p_0, \tau)e^{j2\pi p_0 m} + C(-p_0, \tau)e^{-j2\pi p_0 m}. \tag{3.11}$$

It is clear that $C(f_k, \tau) = 0$ when f_k is not equal to 0 and $\pm p_0$.

Then, the symbol rate estimator is defined as

$$\hat{p}_0 = \arg \max_{f_k \in I} \hat{\mathbf{C}}(f_k)^* \hat{\mathbf{C}}(f_k), \tag{3.12}$$

where I is a searching interval. According to equation (3.7), the matrix $\hat{\mathbf{C}}(f_k)$ is

defined as

$$\hat{\mathbf{C}}(f_k) = \left[\hat{C}(f_k, -\Lambda), \cdots, \hat{C}(f_k, \Lambda)\right]^T. \tag{3.13}$$

This estimator will be used as an initial symbol rate estimator in later chapters.

CHAPTER IV

EXPECTATION MAXIMIZATION

A. Principles

The Expectation Maximization (EM) algorithm represents a powerful algorithm that was applied successfully in numerous applications. Dempster proved the convergence of the EM algorithm in [30]. The EM algorithm produces maximum likelihood (ML) estimates of the parameters under many to one mapping [31]. The main idea of the EM algorithm is that of the Maximization step (M-step), similar to ML approach followed by the Expectation step (E-step), which marginalizes out unknown variables.

Consider two sample spaces, $\mathbb{X}$ and $\mathbb{Y}$, and a many to one mapping from $\mathbb{X}$ to $\mathbb{Y}$. The observation data $\mathbf{y}$ are directly observed from the sample space $\mathbb{Y}$. However, the corresponding $\mathbf{x}$ in the sample space $\mathbb{X}$ are not observed directly but indirectly observed through the observation data $\mathbf{y}$. Dempster called the observation $\mathbf{y}$ the incomplete data, and $\mathbf{x}$ the complete data even though $\mathbf{x}$ includes parameters in certain cases [30].

Assume that the ML estimate of the parameter α is in the sample space $\mathbb{A}$. The probability density of the incomplete data $\mathbf{y}$ is depicted via

$$p(\mathbf{y}|\alpha) = \int_{\mathbb{X}(\mathbf{y})} p(\mathbf{x}|\alpha) d\mathbf{x}. \tag{4.1}$$

The EM algorithm finds the parameter α_0 which maximizes the density $p(\mathbf{y}|\alpha)$. Equivalently, if we assume that

$$\mathbf{B}(\alpha) \triangleq \log\{p(\mathbf{y}|\alpha)\}, \tag{4.2}$$

which denotes the log-likelihood function, the EM algorithm finds the value which maximizes $\mathbf{B}(\alpha)$.

The EM algorithm assumes two steps: E-step and M-step.

E-step: Generate the Q function via

$$Q\left(\alpha|\alpha^{(n-1)}\right) = E\left[\log p(\mathbf{x}|\alpha)|\mathbf{y}, \alpha^{(n-1)}\right], \tag{4.3}$$

where n denotes the index of the iterations.

M-step: Find $\alpha^{(n)}$ which maximizes the Q function,

$$\alpha^{(n)} = \arg\max_{\alpha} Q\left(\alpha|\alpha^{(n-1)}\right). \tag{4.4}$$

After several iterations, $\alpha^{(n)}$ converges to a local maximum value.

B. Convergence

The convergence problem is very significant in every iterative algorithm. The convergence of the EM algorithm is proved by the following procedures. Let

$$g(\mathbf{x}|\mathbf{y}, \alpha) = \frac{p(\mathbf{x}|\alpha)}{p(\mathbf{y}|\alpha)}. \tag{4.5}$$

According to equations (4.2) and (4.5), the log-likelihood function

$$\begin{aligned} \mathbf{L}(\alpha) &= \log p(\mathbf{y}|\alpha) \\ &= \log p(\mathbf{x}|\alpha) - \log g(\mathbf{x}|\mathbf{y}, \alpha). \end{aligned} \tag{4.6}$$

We also define

$$\begin{aligned} \mathbf{D}\left(\hat{\alpha}|\alpha\right) &= E\left[\log g\left(\mathbf{x}|\mathbf{y}, \hat{\alpha}\right)|\mathbf{y}, \alpha\right] \\ &= E\left[\log p\left(\mathbf{x}|\hat{\alpha}\right) - \mathbf{L}\left(\hat{\alpha}\right)|\mathbf{y}, \alpha\right] \\ &= Q\left(\hat{\alpha}|\alpha\right) - \mathbf{L}\left(\hat{\alpha}\right), \end{aligned} \tag{4.7}$$

and

$$f : \alpha^{(n-1)} \to \alpha^{(n)}. \tag{4.8}$$

Then,

$$\mathbf{L}\left(f\left(\alpha^{(n)}\right)\right) \geq \mathbf{L}(\alpha), \tag{4.9}$$

where the equality holds if and only if

$$\begin{aligned} Q(f(\alpha)|\alpha) &= Q(\alpha|\alpha), \\ g(\mathbf{x}|\mathbf{t}, f(\alpha)) &= g(\mathbf{x}|\mathbf{y}, \alpha). \end{aligned} \tag{4.10}$$

Proof.

$$\mathbf{L}(f(\alpha)) - \mathbf{L}(\alpha) = Q(f(\alpha)|\alpha) - \mathbf{D}(f(\alpha)|\alpha) - Q(\alpha|\alpha) + \mathbf{D}(\alpha|\alpha). \tag{4.11}$$

By the M step (4.4),

$$Q(f(\alpha)|\alpha) \geq Q(\alpha|\alpha), \tag{4.12}$$

$\forall \alpha \in \mathbb{A}$. Also,

$$\mathbf{D}\left(\hat{\alpha}|\alpha\right) \leq \mathbf{D}(\alpha|\alpha), \tag{4.13}$$

$\forall(\hat{\alpha}, \alpha) \in \mathbb{A} \times \mathbb{A}$. Based on Jensen's inequality, equation (4.13) is satisfied by

$$\begin{aligned}
\mathbf{D}\left(\hat{\alpha}|\alpha\right) - \mathbf{D}(\alpha|\alpha) &= E\left[\log g\left(\mathbf{x}|\mathbf{y}, \hat{\alpha}\right)\Big|\mathbf{y}, \alpha\right] - E\left[\log g(\mathbf{x}|\mathbf{y}, \alpha)|\mathbf{y}, \alpha\right] \\
&= E\left[\log \frac{g\left(\mathbf{x}|\mathbf{y}, \hat{\alpha}\right)}{g\left(\mathbf{x}|\mathbf{y}, \alpha\right)}|\mathbf{y}, \alpha\right] \\
&\leq \log E\left[\frac{g\left(\mathbf{x}|\mathbf{y}, \hat{\alpha}\right)}{g(\mathbf{x}|\mathbf{y}, \alpha)}\Big|\mathbf{y}, \alpha\right] \\
&= \log \int_{\mathbb{X}} \frac{g\left(\mathbf{x}|\mathbf{y}, \hat{\alpha}\right)}{g(\mathbf{x}|\mathbf{y}, \alpha)} g(\mathbf{x}|\mathbf{y}, \alpha) d\mathbf{x} \\
&= \log \int_{\mathbb{X}} g\left(\mathbf{x}|\mathbf{y}, \hat{\alpha}\right) d\mathbf{x} \\
&= \log(1) \\
&= 0. \qquad (4.14)
\end{aligned}$$

By equation (4.13) and the conditions for equality in Jensen's inequality, the equality condition is only hold for above statement (4.10) [31], [32]. □

As mentioned above, the likelihood function increases at each iteration until the equality condition is satisfied. If α_0 is the Maximum Likelihood (ML) estimate, $\mathbf{L}\left(f\left(\alpha_0\right)\right) \geq \mathbf{L}(\alpha)$, $\forall \alpha \in \mathbb{A}$. Then, $\mathbf{L}\left(f\left(\alpha_0\right)\right) = \mathbf{L}\left(\alpha_0\right)$. Since the likelihood function is bounded, the parameter estimates $\alpha^{(0)}, \ldots, \alpha^{(n)}$ yield a bounded nondecreasing sequence $\mathbf{L}\left(\alpha^{(0)}\right) \leq \cdots \leq \mathbf{L}\left(\alpha^{(n)}\right)$, and the sequence must converge as $n \to \infty$. Moreover, by equations (4.9) and (4.10), the parameter estimate also converges as n goes infinity [31].

C. Discrete EM

Wymeersch proposed a modified EM algorithm for discrete parameters [33]. Since it is not guaranteed that $\alpha^{(n)}$ converges to the ML estimates when the parameter α is discrete [34], the EM algorithm cannot be used for the discrete parameters.

However, using the convergence state of the EM algorithm, it can be modified

for the discrete parameters. Based on the equations (4.9) and (4.10), $\alpha^{(n)} = \alpha^{(n-1)}$ when the sequence $\alpha^{(n)}$ converges. Therefore, equations (4.5) and (4.4) are rewritten, respectively, as

$$Q(\alpha|\alpha) = E[\log p(\mathbf{x}|\alpha)|\mathbf{y}, \alpha]. \tag{4.15}$$

$$\begin{aligned} \alpha_0 &= \arg\max_{\alpha \in \mathbb{A}} Q(\alpha|\alpha) \\ &= \arg\max_{\alpha \in \mathbb{A}} Q(\alpha). \end{aligned} \tag{4.16}$$

Even though this algorithm requires a discrete set of possible values for the parameter α, it overcomes the difficulty of choosing initial values. In additions, when the finite set $\mathbb{A}$ is known, equation (4.16) can be solved by a search algorithm without any convergence problems. Therefore, this algorithm is adopted to estimate the symbol rate in this thesis.

CHAPTER V

SIGNAL MODELS

A. Dynamic Signal Models

We consider a blind data symbol detection problem assuming a wireless channel environment modeled in terms of a Rayleigh flat fading channel [35]. The Rayleigh flat fading channel is modeled using Jakes' model. Because it is not feasible to directly apply Jakes' model into dynamic state-space models, alternatively, an AR process is used to approximate Jakes' model [36]. An AR(2) process was depicted as

$$h_t = -a_1 h_{t-1} - a_2 h_{t-2} + v_t, \tag{5.1}$$

where h_t denotes the fading channel coefficient, a_1 and a_2 are the model coefficients, and v_t is normally distributed noise with zero mean and σ_v^2 variance [37]. Based on the assumption of unit power fading process, the noise variance σ_v^2 can be calculated via

$$\sigma_v^2 = \frac{(1-a_2)((1+a_2)^2 - a_1^2)}{(1+a_2)}. \tag{5.2}$$

Herein, we only consider linearly modulated signals. We assume M-ary PSK modulated signals, and that a square-root raised cosine filter is used as a shaping filter.

1. One Sample per Symbol Period

Consider one sample per symbol period. Based on the known symbol period, the received data signal is sampled at every symbol period. It not only prevents inter-symbol interference (ISI) but also reduces computational complexity. Based on the

state-space dynamic equations (2.1),

$$\begin{aligned} \text{state equations}: \quad \mathbf{a}_t &= \mathbf{a}_{t-1}, \\ \mathbf{h}_t &= \mathbf{A}\mathbf{h}_{t-1} + \mathbf{v}_t, \\ \text{observation equation}: \; y_t &= \mathbf{g}^{\mathrm{T}}\mathbf{h}_t b_t + e_t, \end{aligned} \tag{5.3}$$

where, $\mathbf{h}_t = [h_t, h_{t-1}]^{\mathrm{T}}$, $\mathbf{a}_t = [a_{t,1}, a_{t,2}]^{\mathrm{T}}$, $\mathbf{v}_t = [v_t, 0]^{\mathrm{T}}$, $\mathbf{g} = [1, 0]^{\mathrm{T}}$,and

$$\mathbf{A} = \begin{bmatrix} -a_{t,1} & -a_{t,2} \\ 1 & 0 \end{bmatrix}.$$

Fading channel taps are represented by h_t, and AR coefficients are denoted by $a_{t,1}$ and $a_{t,2}$. The process noise v_t is assumed to be normally distributed with zero mean and σ_v^2 variance. In the observation equation, y_t denotes the received signal, b_t stands for data symbol, and e_t is an additive Gaussian noise (AWGN) with zero mean and σ^2 variance. Since the channel is assumed to be stationary, AR coefficients $a_{t,1}$ and $a_{t,2}$ are considered as static parameters [15].

2. Multiple Samples per Symbol Period

In many references, e.g., [13], [15], the dynamic state-space model (5.3)of one sample per symbol period was adopted. Based on the known symbol period, the received data signal is sampled at every symbol period. Such an approach not only prevents inter-symbol interference (ISI) but also reduces computational complexity. However, it cannot be adopted when the symbol rate is unknown since the symbol rate estimation generally requires oversampling of the received signal. Therefore, we suggest novel state-space dynamic equations to capture the effects of oversampling.

To estimate the symbol period T, it is necessary to oversample the received signal. If we assume that the sampling period is sufficiently small relative to the

symbol period and satisfies $T_s < T/4$, where T and T_s denote the symbol period and the sampling period, respectively, a dynamic state-space channel model can be designed assuming multiple samples per symbol period. The dynamic state-space model is depicted by

$$\begin{aligned} \text{state equations}: \quad \mathbf{a}_t &= \mathbf{a}_{t-1}, \\ \mathbf{h}_t &= \mathbf{A}_t \mathbf{h}_{t-1} + \mathbf{v}_t, \\ \text{observation equation}: \, y_t &= \mathbf{g}^{\mathrm{T}} \mathbf{h}_t s_t + e_t, \end{aligned} \tag{5.4}$$

where

$$s_t = \sum_{l=0}^{L-1} b_{\lfloor \frac{t}{\alpha} \rfloor - l} p_{t,l}, \tag{5.5}$$

where L denotes the number of past symbols correlated with the tth samples, $p_{t,l}$ denotes the pulse shaping filter tap, and $\lfloor \gamma \rfloor$ denotes an integer less than or equal to γ. Other parameters are defined in equation (5.3).

The Particle Filter algorithm is next applied to blindly detect the data symbols based on this dynamic state-space model.

B. Blind Symbol Detection

Due to oversampling, inter-symbol interference is present. To exploit the information contained in the received signal, the delayed PF algorithm is adopted.

First, consider the joint posterior density of transmitted symbols, $b_0, \ldots, b_{\lfloor \frac{t}{\alpha} \rfloor}$, and AR coefficients, $\mathbf{a}_0, \ldots, \mathbf{a}_t$. Using Bayes' rule, the joint posterior density can be

expressed as

$$
\begin{aligned}
& p\Big(b_{0:\lfloor \frac{t}{\alpha} \rfloor}, \mathbf{a}_{0:t+\Delta_1} | y_{0:t+\Delta_1}\Big) \\
\propto \;& p\left(b_{\lfloor \frac{t}{\alpha} \rfloor} | b_{0:\lfloor \frac{t}{\alpha} \rfloor - 1}, \mathbf{a}_{0:t+\Delta_1}, y_{0:t+\Delta_1}\right) p\left(b_{0:\lfloor \frac{t}{\alpha} \rfloor - 1}, \mathbf{a}_{0:t+\Delta_1} | y_{0:t+\Delta_1}\right) \\
\propto \;& p\left(b_{\lfloor \frac{t}{\alpha} \rfloor} | b_{0:\lfloor \frac{t}{\alpha} \rfloor - 1}, \mathbf{a}_{0:t+\Delta_1}, y_{0:t+\Delta_1}\right) p\left(b_{0:\lfloor \frac{t}{\alpha} \rfloor - 1}, \mathbf{a}_{0:t+\Delta_1}, y_{t+\Delta_1-\Delta_2+1:t+\Delta_1} | y_{0:t+\Delta_1-\Delta_2}\right) \\
\propto \;& p\left(b_{\lfloor \frac{t}{\alpha} \rfloor} | b_{0:\lfloor \frac{t}{\alpha} \rfloor - 1}, \mathbf{a}_{0:t+\Delta_1}, y_{0:t+\Delta_1}\right) p\left(\mathbf{a}_{t+\Delta_1}, y_{t+\Delta_1} | b_{0:\lfloor \frac{t}{\alpha} \rfloor - 1}, \mathbf{a}_{0:t+\Delta_1-1}, y_{0:t+\Delta_1-1}\right) \cdots \\
& \times p\left(\mathbf{a}_{t+\Delta_1-\Delta_2+1}, y_{t+\Delta_1-\Delta_2+1} | b_{0:\lfloor \frac{t}{\alpha} \rfloor - 1}, \mathbf{a}_{0:t+\Delta_1-\Delta_2}, y_{0:t+\Delta_1-\Delta_2}\right) \\
& \times p\left(b_{0:\lfloor \frac{t}{\alpha} \rfloor - 1}, \mathbf{a}_{0:t-\lceil \alpha \rceil} | y_{0:t-\lceil \alpha \rceil}\right) \\
\propto \;& p\left(b_{\lfloor \frac{t}{\alpha} \rfloor} | b_{0:\lfloor \frac{t}{\alpha} \rfloor - 1}, \mathbf{a}_{0:t+\Delta_1}, y_{0:t+\Delta_1}\right) \\
& \times \left\{ \prod_{j=0}^{\Delta_2 - 1} p\left(y_{t+\Delta_1-j} | b_{0:\lfloor \frac{t}{\alpha} \rfloor - 1}, \mathbf{a}_{0:t+\Delta_1-j}, y_{0:t+\Delta_1-j-1}\right) p\left(\mathbf{a}_{t+\Delta_1-j} | \mathbf{a}_{t+\Delta_1-j-1}\right) \right\} \\
& \times p\left(b_{0:\lfloor \frac{t}{\alpha} \rfloor - 1}, \mathbf{a}_{0:t+\Delta_1-\Delta_2} | \mathbf{y}_{0:t+\Delta_1-\Delta_2}\right) \\
\propto \;& p\left(b_{\lfloor \frac{t}{\alpha} \rfloor} | b_{0:\lfloor \frac{t}{\alpha} \rfloor - 1}, \mathbf{a}_{0:t+\Delta_1}, y_{0:t+\Delta_1}\right) \prod_{j=0}^{\Delta_2 - 1} p\left(\mathbf{a}_{t+\Delta_1-j} | \mathbf{a}_{t+\Delta_1-j-1}\right) \\
& \times \prod_{j=0}^{\Delta_2 - 1} p\left(y_{t-j} | b_{0:\lfloor \frac{t}{\alpha} \rfloor - 1}, \mathbf{a}_{0:t-j}, y_{t-j-1}\right) p\left(b_{0:\lfloor \frac{t}{\alpha} \rfloor - 1}, \mathbf{a}_{0:t-\Delta_2} | \mathbf{y}_{0:t-\Delta_2}\right),
\end{aligned}
\tag{5.6}
$$

where Δ_1 denotes the number of samples delayed, and Δ_2 stands for the number of samples per symbol period, and $\Delta_3 = \Delta_1 - \Delta_2$.

The samples are generated from the right hand side of equation (5.6) which is referred to as a hybrid importance function,

$$
p\left(b_{\lfloor \frac{t}{\alpha} \rfloor} | b_{0:\lfloor \frac{t}{\alpha} \rfloor - 1}, \mathbf{a}_{0:t+\Delta_1}, y_{0:t+\Delta_1}\right) \prod_{j=0}^{\Delta_2 - 1} p\left(\mathbf{a}_{t+\Delta_1-j} | \mathbf{a}_{t+\Delta_1-j-1}\right), \tag{5.7}
$$

where $p(\mathbf{a}_{t+\Delta_1-j} | \mathbf{a}_{t+\Delta_1-j-1}) = \delta(\mathbf{a}_{t+\Delta_1-j} - \mathbf{a}_{t+\Delta_1-j-1})$, $\delta(\cdot)$ is the Dirac delta function,

and $p(b_{\lfloor \frac{t}{\alpha} \rfloor} | b_{0:\lfloor \frac{t}{\alpha} \rfloor - 1}, \mathbf{a}_{0:t+\Delta_1}, y_{0:t+\Delta_1})$ is shown in Table II. The weight of the function is updated by

$$\hat{w}_{\lfloor \frac{t}{\alpha} \rfloor} \propto w_{\lfloor \frac{t}{\alpha} \rfloor - 1} \prod_{j=0}^{\Delta_2 - 1} p\left(y_{t+\Delta_1 - j} | b_{0:\lfloor \frac{t}{\alpha} - 1 \rfloor - 1}, \mathbf{a}_{0:t+\Delta_1 - j}, y_{0:t+\Delta_1 - j - 1} \right). \tag{5.8}$$

The proposal density function (5.7) does not include any vector related to the channel taps. Therefore the channel vector must be marginalized out. This is implemented using the predictive and update steps of the Kalman filter. The details are shown in Table III.

Table II. Posterior Density Function

$$
\begin{aligned}
& p\left(b_{\lfloor \frac{t}{\alpha} \rfloor} = b_l | b_{0:\lfloor \frac{t}{\alpha} \rfloor - 1}, \mathbf{a}_{0:t+\Delta_1}, y_{0:t+\Delta_1} \right) \\
& \propto \prod_{j=0}^{\Delta_1 + \Delta_2 - 1} p\left(y_{t+\Delta_1 - j} | b^{(i)}_{\lfloor \frac{t}{\alpha} \rfloor} = b_l, b^{(i)}_{1:\lfloor \frac{t}{\alpha} \rfloor - 1}, y_{1:t+\Delta_1 - j - 1} \right) \\
& \propto \prod_{j=0}^{\Delta_1 + \Delta_2 - 1} p\left(y_{t+\Delta_1 - j} | b_f, b^{(i)}_{\lfloor \frac{t}{\alpha} \rfloor} = b_l, b^{(i)}_{1:\lfloor \frac{t}{\alpha} \rfloor - 1}, y_{1:t+\Delta_1 - j - 1} \right) \\
& \propto \prod_{j=0}^{\Delta_1 + \Delta_2 - 1} \sum_{\mathbf{b}_f} \int p\left(y_{t+\Delta_1 - j}, h_{t+\Delta_1 - j} | \mathbf{b}_f, b^{(i)}_{\lfloor \frac{t}{\alpha} \rfloor} = b_l, b^{(i)}_{1:\lfloor \frac{t}{\alpha} \rfloor - 1}, y_{1:t+\Delta_1 - j - 1} \right) dh_{t+\Delta_1 - j} \\
& \propto \prod_{j=0}^{\Delta_1 + \Delta_2 - 1} \sum_{\mathbf{b}_f} \int N\left(h_{t+\Delta_1 - j} s^{(i)}_{t+\Delta_1 - j, f, l}, \sigma^2 \right) N\left(\mu^{(i)}_{t+\Delta_1 - j, f, l}, \Sigma^{(i)}_{t+\Delta_1 - j, f, l} \right) dh_{t+\Delta_1 - j} \\
& \propto \prod_{j=0}^{\Delta_1 + \Delta_2 - 1} \sum_{\mathbf{b}_f} N\left(\mu^{(i)}_{t+\Delta_1 - j, f, l} s^{(i)}_{t+\Delta_1 - j, f, l}, c^{(i)}_{t+\Delta_1 - j, f, l} \right),
\end{aligned}
$$

where $\mathbf{b}_f$ denotes future symbols, $\mathbf{b}_f = [b_{\lfloor \frac{t}{\alpha} \rfloor + 1}, b_{\lfloor \frac{t}{\alpha} \rfloor + 2}, \ldots, b_{\lfloor \frac{t+\Delta_1}{\alpha} \rfloor - 1}, b_{\lfloor \frac{t+\Delta_1}{\alpha} \rfloor}]$, $c^{(i)}_{t+\Delta_1 - j, f, l} = \mathbf{g}^T \Sigma^{(i)}_{t+\Delta_1 - j, f, l} \mathbf{g} + \sigma^2 s^{(i)2}_{t+\Delta_1 - j, f, l}$, and all other parameters are obtained by the Kalman filter in the Table III.

To prevent phase ambiguity, initial AR coefficients are generated via

$$
\begin{aligned}
a_1 &= -2 r_d \cos\left(\frac{2\pi f_d T}{\sqrt{2}} \right), \\
a_2 &= r_d^2,
\end{aligned} \tag{5.9}
$$

Table III. Kalman Filter

1. Time update the channel vector

$$\mu^{(i)}_{t+\Delta_1,f,l} = \mathbf{g}^{\mathrm{T}} A^{(i)}_{t+\Delta_1} \boldsymbol{\gamma}^{(i)}_{t+\Delta_1-1,f,l}$$

$$\Sigma^{(i)}_{t+\Delta_1,f,l} = A^{(i)}_{t+\Delta_1} \Sigma^{(i)}_{t+\Delta_1-1,f,l} A^{(i)\mathrm{T}}_{t+\Delta_1} + \sigma^{2(i)}_{v,t+\Delta_1} \mathbf{g}\mathbf{g}^{\mathrm{T}}.$$

2. Measurement update the channel vector

$$K^{(i)}_{t+\Delta_1,f,l} = \Sigma^{(i)}_{t+\Delta_1,f,l} \mathbf{g} c^{(i)-1}_{t+\Delta_1,f,l} s^{(i)}_{t+\Delta_1,f,l}$$

$$\boldsymbol{\gamma}^{(i)}_{t+\Delta_1,f,l} = A^{(i)}_{t+\Delta_1} \boldsymbol{\gamma}^{(i)}_{t+\Delta_1-1,f,l} + K^{(i)}_{t+\Delta_1,f,l} \left(y_{t+\Delta_1} - \mu^{(i)}_{t+\Delta_1,f,l} s^{(i)}_{t+\Delta_1,f,l} \right)$$

$$C^{(i)}_{t+\Delta_1,f,l} = \left(I - K^{(i)}_{t+\Delta_1,f,l} \mathbf{g}^{\mathrm{T}} s^{(i)}_{t+\Delta_1,f,l} \right) \Sigma^{(i)}_{t+\Delta_1,f,l}.$$

and

$$f_d = \frac{v}{\lambda}, \tag{5.10}$$

where v denotes the speed of the vehicle, λ stands for the carrier wavelength, and r_d is the pole radius of the AR model and f_d is the maximum Doppler frequency, which are drawn from the regions $[0.9, 0.999]$ and $[0, 0.1]$, respectively. The region of f_dT is decided by considering real-world communication systems. For examples, f_dT must be less than 0.062 if a system assumes 2 GHz carrier frequency, symbol rates are greater than 3600 Hz, and the vehicle speeds are less than 75 miles/h [15].

Having introduced all elements required for the implementation of the PF algorithm, the resulting weighted samples, $b^{(i)}_{\lfloor \frac{t}{\alpha} \rfloor}$ and $w^{(i)}_{\lfloor \frac{t}{\alpha} \rfloor}$, $i = 1, \ldots, N$, approximate $p(b_{\lfloor \frac{t}{\alpha} \rfloor} | y_{0:t})$, and the minimum mean square error (MMSE) estimate is calculated via

$$\hat{b}_{\lfloor \frac{t}{\alpha} \rfloor} = \sum_{i=1}^{N} b^{(i)}_{\lfloor \frac{t}{\alpha} \rfloor} w^{(i)}_{\lfloor \frac{t}{\alpha} \rfloor}. \tag{5.11}$$

The resampling step should be added at the end. However, the general resampling step does not prevent AR coefficients, $\mathbf{a}_t$, from degenerating and assume very few

different values. Huang and Djurić proposed a novel resampling step [15] based on the Auxiliary Particle Filter (APF) and smoothing kernel approach, which was originally proposed by Liu and West in [38].

Whenever the resampling step is required, instead of the general resampling step, the following procedure is performed. First, the sampled mean and covariance matrix are computed via

$$\begin{aligned} \bar{\mathbf{a}}_{t-1} &= \sum_{i=1}^{N} w_{t-1}^{(i)} \mathbf{a}_{t-1}^{(i)}, \\ V_{t-1} &= \sum_{i=1}^{N} w_{t-1}^{(i)} \left(\mathbf{a}_{t-1}^{(i)} - \bar{\mathbf{a}}_{t-1} \right)^2. \end{aligned} \tag{5.12}$$

A new mean vector is defined as $\tilde{\mathbf{a}}_t^{(i)} = \epsilon \mathbf{a}_{t-1}^{(i)} + (1-\epsilon)\bar{\mathbf{a}}_{t-1}$. An auxiliary variable is generated from the index set $\{1, \ldots, N\}$ with the probability proportional to

$$\begin{aligned} & q(i|y_{0:t+\Delta_1}) \\ & \propto w_{\lfloor \frac{t}{\alpha} \rfloor - 1} \prod_{j=0}^{\Delta_2 - 1} p\left(y_{t+\Delta_1-j} | b_{0:\lfloor \frac{t}{\alpha} \rfloor - 1}^{(i)}, \tilde{\mathbf{a}}_{t+\Delta_1-\Delta_2+1:t+\Delta_1-j}^{(i)}, \mathbf{a}_{0:t+\Delta_1-\Delta_2}^{(i)}, y_{0:t+\Delta_1-j-1} \right). \end{aligned} \tag{5.13}$$

Consider the generated sample index as a new index ξ, and draw the channel model coefficients $\mathbf{a}_{t+\Delta_1-\Delta_2+1:t+\Delta_1}^{(i)}$ from the density represented by

$$\begin{aligned} & q\left(\mathbf{a}_{t+\Delta_1-\Delta_2+1:t+\Delta_1} | \mathbf{a}_{0:t+\Delta_1-\Delta_2}^{(\xi)} \right) \\ & = p\left(\mathbf{a}_{t+\Delta_1} | \mathbf{a}_{t+\Delta_1-1}^{(i)} \right) p\left(\mathbf{a}_{t+\Delta_1-1}^{(i)} | \mathbf{a}_{t+\Delta_1-2}^{(i)} \right) \cdots \\ & \times p\left(\mathbf{a}_{t+\Delta_1-\Delta_2+2}^{(i)} | \mathbf{a}_{t+\Delta_1-\Delta_2+1}^{(i)} \right) p\left(\mathbf{a}_{t+\Delta_1-\Delta_2+1}^{(i)} | \mathbf{a}_{t+\Delta_1-\Delta_2}^{(i)} \right) \\ & = \delta\left(\mathbf{a}_{t+\Delta_1} - \mathbf{a}_{t+\Delta_1-1}^{(i)} \right) \delta\left(\mathbf{a}_{t+\Delta_1-1}^{(i)} - \mathbf{a}_{t+\Delta_1-2}^{(i)} \right) \cdots \\ & \times TN\left(\mathbf{a}_{t+\Delta_1-\Delta_2+1}; \tilde{\mathbf{a}}_{t+\Delta_1-\Delta_2+1}^{(\xi)}, h^2 V_{t+\Delta_1-\Delta_2+1} | \left[a_{l1}, a_{u1} \right], \left[a_{l2}, a_{u2} \right] \right), \end{aligned} \tag{5.14}$$

where $TN(\beta;\gamma_1,\Delta|[a_{l1},a_{u1}],[a_{l2},a_{u2}])$ denotes a truncated multivariate normal distribution with the mean γ_1, covariance matrix Δ, and boundaries, $[a_{l1},a_{u1}]$ and $[a_{l2},a_{u2}]$. Since the channel is assumed to be stationary, the Dirac delta function can be used as the prior function of the channel model coefficients. However, the Dirac delta function makes the algorithm depend on the initial sample values since the previous samples are transferred without any changes. Therefore, at each data symbol $b_t^{(i)}$ drawing, the first Dirac delta function is replaced by the truncated normal distribution to vary the samples. Based on the generated samples $\mathbf{a}^{(i)}_{t+\Delta_1-\Delta_2+1:t+\Delta_1}$, other samples $b_{\lfloor\frac{t}{\alpha}\rfloor}$ are drawn from the hybrid importance function (5.7). The new updated weight is also evaluated via

$$\hat{w}^{(i)}_{\lfloor\frac{t}{\alpha}\rfloor} \propto \frac{\prod_{j=0}^{\Delta_2-1} p\left(y_{t+\Delta_1-j}|b^{(i)}_{0:\lfloor\frac{t}{\alpha}-1\rfloor-1},\mathbf{a}^{(i)}_{0:t+\Delta_1-j},y_{0:t+\Delta_1-j+1}\right)}{\prod_{j=0}^{\Delta_2-1} p\left(y_{t+\Delta_1-j}|b^{(\xi)}_{0:\lfloor\frac{t}{\alpha}-1\rfloor-1},\mathbf{a}^{(\xi)}_{0:t+\Delta_1-j},y_{0:t+\Delta_1-j+1}\right)}. \tag{5.15}$$

C. Symbol Period Estimation

The Expectation Maximization (EM) algorithm is adopted to estimate the symbol rate. Based on the channel model, define the vectors

$$\begin{aligned}\mathbf{b} &= \left[b_0,b_1,\cdots,b_{\lfloor\frac{t}{\alpha}\rfloor-1},b_{\lfloor\frac{t}{\alpha}\rfloor}\right],\\ \mathbf{y} &= \left[y_0,\cdots,y_{t+\Delta_1}\right],\\ \mathcal{A} &= \left[\mathbf{a}_0,\cdots,\mathbf{a}_{t+\Delta_1}\right].\end{aligned} \tag{5.16}$$

Based on the vectors in (5.16), the E-step in the discrete EM (D-EM) method is implemented through

$$Q(\alpha)=\int_{\mathcal{A}}\int_{\mathbf{b}} p(\mathbf{b},\mathcal{A}|\mathbf{y},\alpha)\log p(\mathbf{y}|\mathbf{b},\mathcal{A},\alpha)d\mathbf{b}d\mathcal{A}. \tag{5.17}$$

To simplify the Q-function in equation (5.17), we approximate both the probability density function and log-likelihood function using the Particle Filter (PF) algorithm. The joint probability density and the log-likelihood function are rewritten as

$$p(\mathbf{b}, \mathcal{A}|\mathbf{y}, \alpha) = p\left(b_{0:\lfloor \frac{t}{\alpha} \rfloor}, \mathbf{a}_{0:t+\Delta_1} | y_{0:t+\Delta_1}, \alpha\right), \tag{5.18}$$

$$\log p(\mathbf{y}|\mathbf{b}, \mathcal{A}, \alpha) = \log p\left(y_{0:t+\Delta_1} | b_{0:\lfloor \frac{t}{\alpha} \rfloor}, \mathbf{a}_{0:t+\Delta_1}, \alpha\right). \tag{5.19}$$

Based on the Table II, we generate samples $b^{(i)}_{0:\lfloor \frac{t}{\alpha} \rfloor}$ and $\mathbf{a}^{(i)}_{0:t+\Delta_1}$ from the (5.18). The Q-function of the D-EM can be approximated as

$$\begin{aligned}
Q(\alpha) &\approx \int_{\mathcal{A}} \int_{\mathbf{b}} \sum_{i=1}^{N} w^{(i)}_{\lfloor \frac{t}{\alpha} \rfloor} \delta\left(b_{0:\lfloor \frac{t}{\alpha} \rfloor} - b^{(i)}_{0:\lfloor \frac{t}{\alpha} \rfloor}\right) \delta\left(\mathbf{a}_{0:t+\Delta_1} - \mathbf{a}^{(i)}_{0:t+\Delta_1}\right) \\
&\quad \times \log p\left(y_{0:t+\Delta_1} | b_{0:\lfloor \frac{t}{\alpha} \rfloor}, \mathbf{a}_{0:t+\Delta_1}, \alpha\right) d\mathbf{b} d\mathcal{A} \\
&= \sum_{i=1}^{N} w^{(i)}_{\lfloor \frac{t}{\alpha} \rfloor} \log p\left(y_{0:t+\Delta_1} | b^{(i)}_{0:\lfloor \frac{t}{\alpha} \rfloor}, \mathbf{a}^{(i)}_{0:t+\Delta_1}, \alpha\right) \\
&= \sum_{i=1}^{N} w^{(i)}_{\lfloor \frac{t}{\alpha} \rfloor} \log \prod_{j=0}^{t+\Delta_1} p\left(y_j | b^{(i)}_{0:\frac{t}{\alpha}}, \mathbf{a}^{(i)}_{0:t+\Delta_1}, y_{0:j-1}, \alpha\right) \\
&= \sum_{i=1}^{N} w^{(i)}_{\lfloor \frac{t}{\alpha} \rfloor} \sum_{j=0}^{t+\Delta_1} \log p\left(y_j | b^{(i)}_{0:\frac{t}{\alpha}}, \mathbf{a}^{(i)}_{0:t+\Delta_1}, y_{0:j-1}, \alpha\right) \\
&= \sum_{i=1}^{N} w^{(i)}_{\lfloor \frac{t}{\alpha} \rfloor} \sum_{j=0}^{t+\Delta_1} \log N\left(\mu^{(i)}_j s^{(i)}_j, c^{(i)}_j\right).
\end{aligned} \tag{5.20}$$

Then, the M-step of the D-EM takes the form

$$\alpha_0^{(n)} = \arg \max_{\alpha \in \mathbb{A}^{(n)}} Q(\alpha), \tag{5.21}$$

where n denotes the number of iterations of the EM and $\mathbb{A}$ represents a discrete set of possible values for α. As the number of iteration increases, we should shrink the

range of the discrete set $\mathbb{A}^{(n)}$. The procedure is diagramed as

$$\begin{array}{c} \mathbb{A}^{(0)} \\ \downarrow \\ \alpha_0^{(1)}, \quad \varepsilon_{(1)} \quad , \eta_{(1)} \\ \downarrow \\ \mathbb{A}^{(1)} = \{\alpha_0^{(1)} - \eta_{(1)}, \alpha_0^{(1)} - \eta_{(1)} + \varepsilon_{(1)}, \cdots, \quad \alpha_0^{(1)} \quad , \cdots, \alpha_0^{(1)} + \eta_{(1)} - \varepsilon_{(1)}, \alpha_0^{(1)} + \eta_{(1)}\} \\ \downarrow \\ \vdots \\ \downarrow \\ \alpha_0^{(n)}. \end{array} \tag{5.22}$$

where ε and η are small values which satisfy the conditions $\varepsilon_{(1)} > \cdots > \varepsilon_{(n-1)}$ and $\eta_{(1)} > \cdots > \eta_{(n-1)}$, respectively. Given the $(n-1)^{\text{th}}$ discrete set $\mathbb{A}^{(n-1)}$, the estimated oversampling factor $\alpha^{(n)}$ is estimated by the D-EM. The n^{th} discrete set $\mathbb{A}^{(n)}$ consists of the number of $\lfloor 2\eta_{(n)}/\varepsilon_{(n)} \rfloor$ elements neighboring $\alpha^{(n)}$. For example, when $\alpha^{(n)} = 5.2$, $\eta_{(n)} = 0.07$, and $\varepsilon_{(n)} = 0.01$, then

$$\mathbb{A}^{(n)} = \{5.13, 5.14, \ldots, 5.26, 5.27\}. \tag{5.23}$$

To represent the $(n+1)^{\text{th}}$ discrete set $\mathbb{A}^{(n+1)}$, we choose values for $\eta_{(n+1)}$ and $\varepsilon_{(n+1)}$ smaller than $\eta_{(n)}$ and $\varepsilon_{(n)}$, respectively, and repeat the process until the convergence is achieved. After certain iterations, we finally obtain an accurate estimate $\hat{\alpha}$.

D. Initial Symbol Period Estimation

In the previous section, we have discussed the symbol rate estimator using the discrete EM (D-EM) algorithm. The D-EM algorithm requires an initial finite set that will be obtained by using the cyclic correlation based symbol-rate estimator. The cyclic correlation based symbol rate estimator is suitable as an initialization estimator because it only requires a sufficiently small sampling period so that $T_s < T/4$. According to [7] and [8], the initial estimate can be obtained via

$$\hat{p}_0 = \arg\max_{f_k \in I} \hat{\mathbf{C}}(f_k)^* \hat{\mathbf{C}}(f_k), \tag{5.24}$$

where $\hat{\mathbf{C}}(f_k)$ stands for the vector of cyclic correlations (see [7], [8] for more details). There is a reciprocal relation between the oversampling parameter α_0 and the cyclic frequency p_0. Therefore, the estimate of oversampling factor can be represented by

$$\hat{\alpha_0} \;=\; \frac{1}{\hat{p}_0}. \tag{5.25}$$

For more efficient estimation, based on equation (5.24), the symbol-rate estimator is reformulated as

$$\begin{aligned} \hat{p}_{0,j} &= \arg\max_{f_k \in I_j} \hat{\mathbf{C}}(f_k)^* \hat{\mathbf{C}}(f_k), \\ \hat{\alpha}_{0,j} &= \frac{1}{\hat{p}_{0,j}}. \end{aligned} \tag{5.26}$$

where $j = 1, \ldots, J$, and J is the number of searching sub-intervals. The searching interval I should be divided into several sub-intervals, $I_1, \ldots, I_J$, and each local maximum value should be selected from the sub-intervals. The selected local maximum values consist of the initial finite set $\mathbb{A}^{(0)}$, i.e.,

$$\mathbb{A}^{(0)} = \left\{ \hat{\alpha}_{0,1}, \ldots, \hat{\alpha}_{0,J} \right\}. \tag{5.27}$$

CHAPTER VI

SIMULATION RESULTS

In this chapter, the performance of the proposed algorithm is illustrated through computer simulations. In all computer simulations, a Rayleigh flat fading channel, BPSK modulation with unit power, and a square-root raised cosine pulse shaping filter are assumed. In addition, all transmitted data symbols are differentially encoded to prevent phase ambiguities. The signal to noise ratio (SNR) is calculated as the averaged received SNR.

In the first computer simulation, we compared the BER performance of the multiple samples per symbol period signal data detector (MSSD) to the single sample per symbol period signal data detector (SSSD). As shown in Fig. 1, the MSSD improves the BER performance much more than the SSSD. Based on the PF with general resampling, MSSD eliminates the visible error floor which is shown with the SSSD. The performance gain is much larger at high SNR since the MSSD tracks the channel much better, and the overall performance is limited by the channel fading.

In the Fig. 2, the BER performances of each method, namely Mixture Kalman Filter (MKF) with known channel model coefficients, Particle Filter with general Resampling (PF-RS), and Particle Filter with Smoothing Kernel (PF-SK), are plotted. When we oversample the received signal, the gain due to the smoothing kernel is negligible. Therefore, using PF-RS, the complexity caused by smoothing kernel method can be reduced. Both PF-RS and PF-SK show better performance than the Dual Kalman Filter (DKF) method. To show the lower bound, the performance of the MKF with known channel model coefficients is also presented.

In the Fig. 3, according to the number of particles that are considered, the

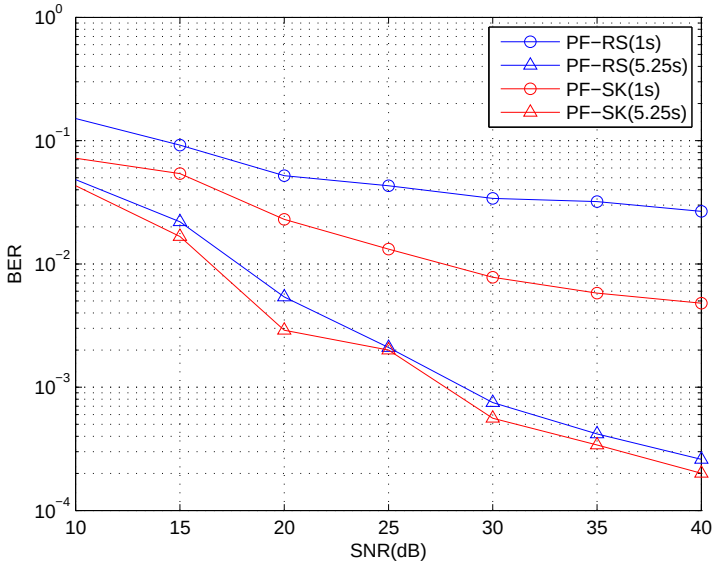

Fig. 1. BERs of PF-SK and PF-RS with one sample per symbol period and 5.25 samples per symbol period (f_dT=0.05, $\alpha = 5.25$).

BER performances are compared. As the number of particles increases, the BER performance is improved. Moreover, the PF-RS algorithm shows better performance than the DKF.

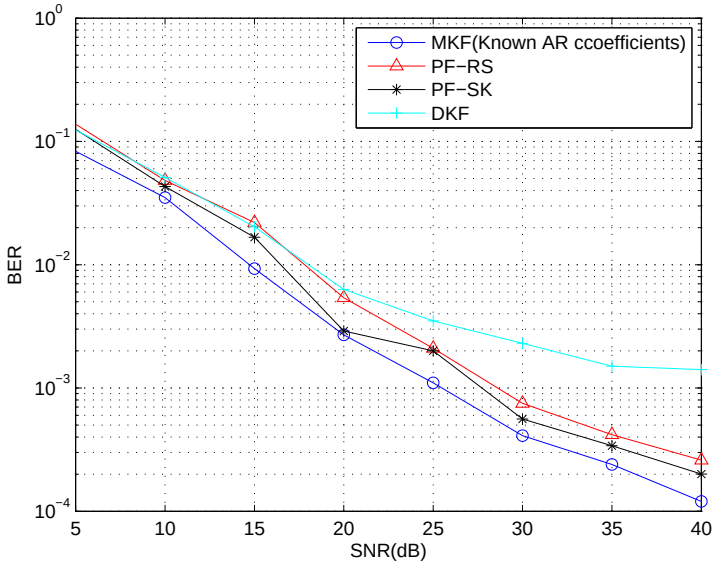

Fig. 2. BERs of PF-SK, PF-RS, and MKF with known AR coefficients with 50 particles, f_dT=0.05, and $\alpha = 5.25$.

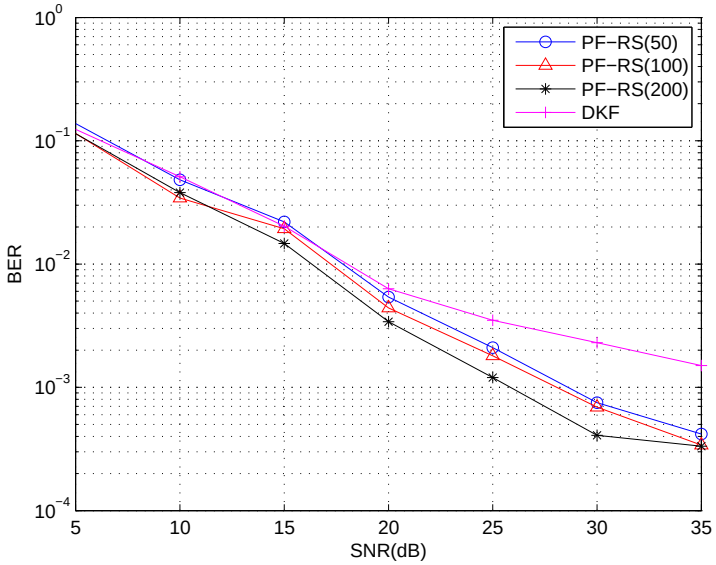

Fig. 3. BERs of the PF-RS with 50, 100, and 200 particles (f_dT=0.05, $\alpha = 5.25$).

CHAPTER VII

CONCLUSIONS AND FUTURE WORKS

A. Conclusions

Recently, non-cooperative communication systems have attracted a lot of attention. Numerous researchers have focused on systems such as automatic modulation classification in the military and civilian application areas. The importance of blind estimation of the channel parameters and blind detection of the data symbols also comes from the increasing attention given to AMC applications.

In this thesis, novel symbol rate estimators with improved performance compared to the estimator based on cyclic correlation were proposed. The EM algorithm, which is used in the symbol rate estimator, is simplified and made tractable by using the PF algorithm. A delayed oversampling based data symbol detector is also proposed under the modeling framework of Rayleigh flat fading channels. Using the delayed oversampling data symbol detector, the performance of the data symbol detector is improved compared to the classical blind PF detector. Moreover, the general resampling technique, which is very simple, can be adopted since this detector reduces the effect of the AR coefficient estimation errors. Finally, since both the symbol rate estimator and data symbol detector rely on the same PF algorithm, the resulting algorithm presents low computational complexity.

B. Future Works

There are numerous directions for future research work. First, other parameters such as phase offset and timing offset can be jointly estimated. Combining the symbol rate estimator and data symbol detector reduces the overall computational complex-

ity. Finally, this algorithm could be expanded in several different directions such as frequency flat Rician fading channels, frequency selective Rayleigh fading channels, and frequency selective Rician fading channels.

REFERENCES

[1] C. M. Spooner, "Classification of cochannel communication signals using cyclic cumulants," in *Proc. of the 29th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, 1995, vol. 2, pp. 531-536.

[2] C. M. Spooner, W. A. Brown, and G. K. Young, "Automatic radio-frequency environment analysis," in *Proc. of the 29th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, 2000, vol. 2, pp. 1181-1186.

[3] C. M. Spooner, "On the utility of sixth-order cyclic cumulants for RF signal classification," in *Proc. of the 29th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, 2001, vol. 1, pp. 890-897.

[4] O. A. Dobre, A. Abdi, Y. Bar-Ness, S. Wei, "Blind modulation classification: a concept whose time has come," in *Proc. of Advances in Wired and Wireless Communication, 2005 IEEE/Sarnoff Symposium*, Apr. 2005, pp. 223-228.

[5] O. A. Dobre, A. Abdi, Y. Bar-Ness, S. Wei, "Cyclostationarity-based blind classification of analog and digital modulations," in *Proc. of Military Communications Conference*, Oct. 2006, Washington, DC, pp. 1-7.

[6] Y. T. Chan, J. W. Plews, K. C. Ho, "Symbol rate estimation by the wavelet transform," in *Proc. of IEEE ISCAS'97*, Hong Kong, Jun. 1997, vol. 1, pp. 177-180.

[7] L. Mazet, P. Loubaton, "Cyclic correlation based symbol rate estimation," in *Proc. of the 33th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Oct. 1999, vol. 2, pp. 1008-1012.

[8] P. Ciblat, P. Loubaton, E. Serpedin, and G. B. Giannakis, "Asymptotic analysis of blind cyclic correlation-based symbol-rate estimators," *IEEE Trans. Information Theory*, vol. 48, no. 7, Jul. 2002.

[9] J. H. Lodge, and M. L. Moher, "Maximum likelihood sequence estimation of CPM signals transmitted over Rayleigh flat-fading channels," *IEEE Trans. Commun.*, vol. 38, no. 6, pp. 787-794, Jun. 1990.

[10] D. Makrakis, P. T. Mathiopoulos, and D. P. Bouras, "Optimal decoding of coded PSK and QAM signals in correlated fast fading channels and AWGN: a combined envelope, multiple differential and coherent detection approach," *IEEE Trans. Commun.*, vol. 42, no. 1, pp. 63-75, Jan. 1994.

[11] X. Yu, and S. Pasupathy, "Innovations-based MLSE for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4 pp. 1534-1544, Feb./Mar./Apr. 1995.

[12] G. M. Vitetta, and D. P. Taylor, "Maximum likelihood decoding of uncoded and coded PSK signal sequences transmitted over Rayleigh flat fading channels," *IEEE Trans. Commun.*, vol. 43, no. 11, pp. 2750-2758, Nov. 1995.

[13] R. Chen, X. Wang, and J. S. Liu, "Adaptive joint detection and decoding in flat-fading channels via mixture Kalman filtering," *IEEE Trans. Information Theory*, vol. 46, no. 6, pp. 2079-2094, Sep. 2000.

[14] E. Punskaya, C. Andrieu, A. Doucet, and W. J. Fitzgerald, "Particle filtering for demodulation in fading channels with non-Gaussian additive noise," *IEEE Trans. Commun.*, vol. 49, no. 4, pp. 579-582, Apr. 2001.

[15] Y. Huang, and P. M. Djurić, "A blind particle filtering detector of signals transmitted over flat fading channels," *IEEE Trans. Signal Processing*, vol. 52, no. 7, Jul. 2004.

[16] Y. Huang, and P. M. Djurić, "A hybrid importance function for particle filtering," *IEEE Signal Processing Letters*, vol. 11, no. 3, pp. 404-406, Mar. 2004.

[17] P. M. Djurić, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, J. Miguez, "Particle filtering," *IEEE Signal Processing Magazine*, vol. 20, no. 5, pp. 19-38, Sep. 2003.

[18] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*, Boston: Artech House, 2004.

[19] J. Carpenter, P. Clifford, and P. Fearnhead, "Improved particle filter for nonlinear problems," *IEE Proc. Radar and Sonar Navigation*, vol. 146, no. 1, pp.2-7. Feb. 1999.

[20] N. J. Gordon, D. J. Salmon, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proc.-F Radar and Signal Processing*, vol. 140, no. 2, pp. 107-113, Apr. 1993.

[21] Y. Huang, and P. M. Djurić, "A new importance function for particle filtering and its application to blind detection in flat fading channels," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Orlando, FL, 2002, vol. 2, pp. 1617-1620.

[22] A. Kong, J. S. Liu, and W. H. Wong, "Sequential imputations and Bayesian missing data problems," *Journal of the American Statistical Association*, vol. 89, no. 425, pp. 278-288, Mar. 1994.

[23] M. Pitt, and N. Shephard, "Filtering via simulation: auxiliary particle filters," *Journal of the American Statistical Association*, vol. 94, no. 446, pp. 590-599, Jun. 1999.

[24] P. Fearnhead, "Sequential Monte Carlo methods in filter theory," Ph.D. dissertation, University of Oxford, 1998.

[25] M. West, "Approximating posterior distributions by mixtures," *Journal of the Royal Statistical Society*, vol. 55, no. 2, pp. 409-422, 1993.

[26] M. West, "Mixture models, Monte Carlo, Bayesian updating and dynamic models," in *J. H. Newton, Computing Science and Statistics: Proceedings of the 24th Symposium on the Interface, Interface Foundation of North America*, Fairfax Station, VA, 1993, pp. 325-333.

[27] G. B. Giannakis, "Cyclostationary signal analysis," in *the Digital Signal Processing Handbook*, Florida: CRC Press, 1999.

[28] C. Corduneanu, *Almost Periodic Functions*, New York: John Wiley & Sons, 1968.

[29] W. A. Gardner, "Signal interception: a unifying theoretical framework for feature detection," *IEEE Trans. Commun.*, vol. 36, no. 8, pp. 897-906, Aug. 1988.

[30] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society*, Series B, vol. 39, no. 1, pp. 1-38, 1977.

[31] T. K. Moon, and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*, Upper Saddle River: Prentice Hall, 2000.

[32] G. J. McLachlan, and T. Krishnan, *The EM Algorithm and Extensions*, New York: John Wiley & Sons, 1997.

[33] H. Wymeersch, and M. Moeneclaey, "Code-aided phase and timing ambiguity resolution for AWGN channels," presented at *the IASTED Int. Conf. Acoustics, Signal, Image Processing (SIP03)*, Honolulu, HI, Aug. 2003.

[34] P. Spasojevic, and C. N. Georghiades, "On the (non) convergence of the EM algorithm for discrete parameter estimation," in *Proc. of the 38th Annual Allerton Conference*, Urbana, IL, Oct, 2000.

[35] D. Tse, and P. Viswanath, *Fundamentals of Wireless Communication*, New york: Cambridge Univ. Press, 2005.

[36] L. Lindbom, A. Ahlen, M. Sternad, and M. Falkenstrom, "Tracking of time-varying mobile radio channels-part II: a case study," *IEEE Trans. Commun.*, vol. 50, no. 1, pp. 156-167, Jan. 2002.

[37] H. Y. Wu and A. Duel-Hallen, "Multiuser detection and channel estimation for flat Rayleigh fading CDMA," [Online]. Available: citeseer.nj.nec.com/406 501.html.

[38] J. Liu and M. West, "Combined parameter and state estimation in simulation-based filtering," in *Sequential Monte Carlo Methods in Practice*, A. Doucet, J. F. G. De Freitas, and N. J. Gordon, Eds. New York: Springer-Verlag, 2000.

VITA

Sang Woo Park received the B.S. degree in electrical and electronics engineering in 2004, from Chung-Ang University, Seoul, Korea. From 2004 to 2005, he worked as an assistant engineer in Mobile Communication Division, Telecommunication Network Business at Samsung Electronics, Suwon, Korea. In 2005, he joined Texas A&M University, College Station, to pursue his M.S. degree. He received his M.S. degree in electrical engineering in May 2008. He may be reached at the following e-mail addresses or mailing address: sangwoo78@gmail.com, swpark78@neo.tamu.edu, or 2503 Samsungomni tower, 395-62 Shindaebang 2dong, Donjakgu, Seoul, Korea

The typist for this thesis was Sang Woo Park.