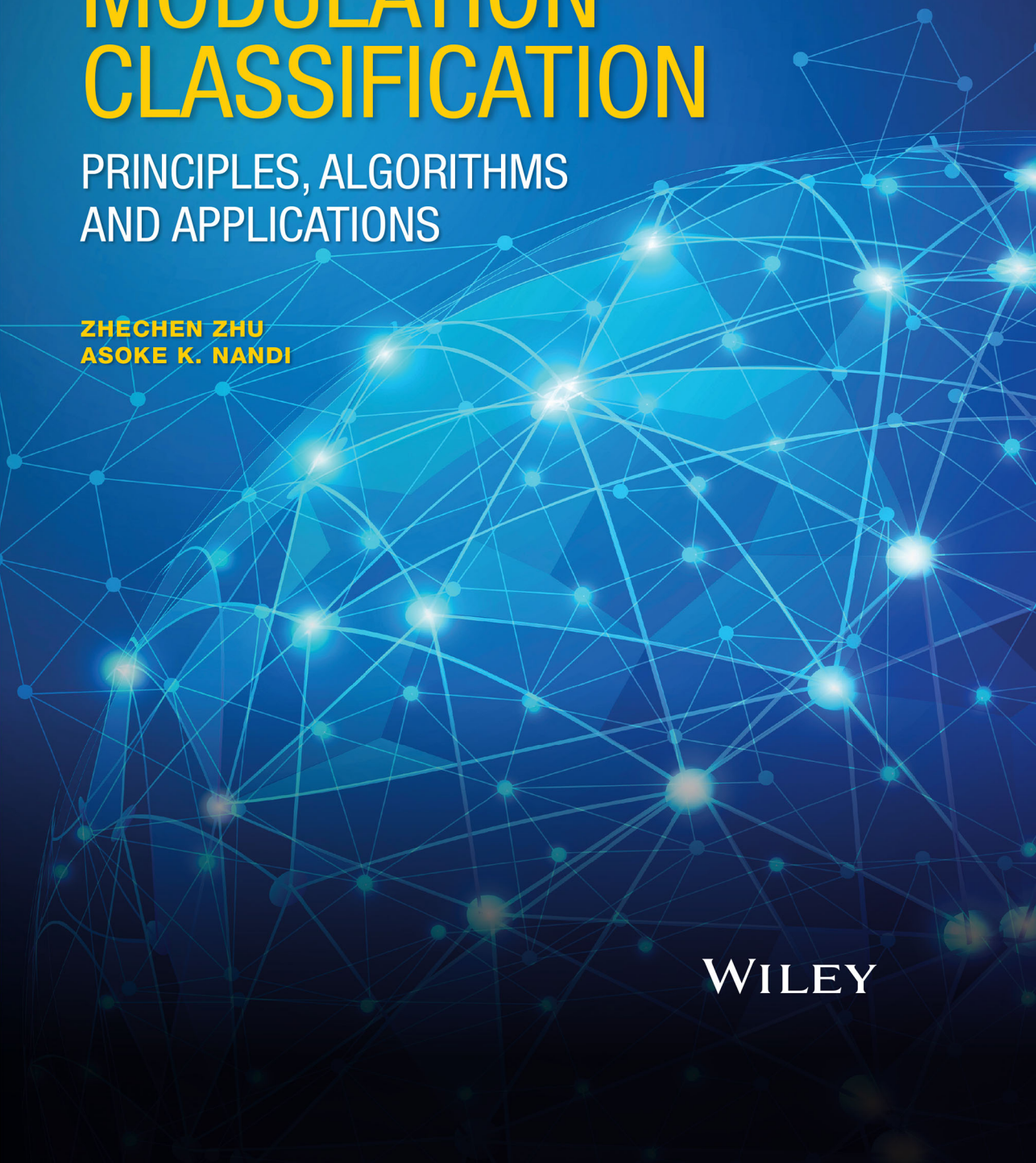
# AUTOMATIC MODULATION CLASSIFICATION

## PRINCIPLES, ALGORITHMS AND APPLICATIONS

ZHECHEN ZHU
ASOKE K. NANDI



WILEY

# Automatic Modulation Classification

# Automatic Modulation Classification

## Principles, Algorithms and Applications

**Zhechen Zhu and Asoke K. Nandi**

*Brunel University London,* UK

WILEY

This edition first published 2015


*Registered Office*
John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom



*Library of Congress Cataloging-in-Publication Data*

Zhu, Zhechen.
Automatic modulation classification : principles, algorithms, and applications / Zhechen Zhu and Asoke K. Nandi.
pages cm
Includes bibliographical references and index.
ISBN 978-1-118-90649-1 (cloth)
1. Modulation (Electronics) I. Nandi, Asoke Kumar. II. Title.
TK5102.9.Z477 2015
621.3815′36–dc23
2014032270

A catalogue record for this book is available from the British Library.

Set in 10/12.5pt Palatino by SPi Publisher Services, Pondicherry, India

1 2015

*To*
*Xiaoyan and Qiaonan Zhu*
*Marion, Robin, David, and Anita Nandi*

# Contents

# About the Authors

Zhechen Zhu received his B.Eng. degree from the Department of Electrical Engineering and Electronics at the University of Liverpool, Liverpool, UK, in 2010. Before graduating from the University of Liverpool, he also studied in Xi'an Jiaotong-Liverpool University, People's Republic of China for two years. He recently submitted his thesis for the degree of PhD to the Department of Electronic and Computer Engineering at Brunel University London, UK. Since 2009, he has been working closely with Professor Asoke K. Nandi on the subject of automatic modulation classification. Their collaboration has made an important contribution to the advancement of automatic modulation classification in complex channels using modern machine learning techniques. His work has since been published in three key journal papers and reported in several high quality international conferences.

Asoke K. Nandi joined Brunel University London in April 2013 as the Head of Electronic and Computer Engineering. He received a PhD from the University of Cambridge, UK, and since then has worked in many institutions, including CERN, Geneva; University of Oxford, UK; Imperial College London, UK; University of Strathclyde, UK; and University of Liverpool, UK. His research spans many different topics, including automatic modulation recognition in radio communications for which he received the Mountbatten Premium of the Institution of Electrical Engineers in 1998, machine learning, and blind equalization for which he received the 2012 IEEE Communications Society Heinrich Hertz Award from the Institute of Electrical and Electronics Engineers (USA).

In 1983 Professor Nandi was a member of the UA1 team at CERN that discovered the three fundamental particles known as $W^+$, $W^-$ and $Z^0$, providing the evidence necessary for the unification of the electromagnetic and weak forces, which was recognized by the Nobel Committee for Physics in 1984. He has been honoured with the Fellowship of the Royal Academy of Engineering (UK) and the Institute of Electrical and Electronics Engineers (USA). He is a Fellow of five other professional institutions, including the Institute of Physics (UK), the Institute of Mathematics and its Applications (UK), and the British Computer Society. His publications have been cited well over 16 000 times and his *h*-index is 60 (Google Scholar).

# Preface

Automatic modulation classification detects the modulation type of received signals to guarantee that the signals can be correctly demodulated and that the transmitted message can be accurately recovered. It has found significant roles in military, civil, intelligence, and security applications.

Analogue Modulations (e.g., AM and FM) and Digital Modulations (e.g., PSK and QAM) transform baseband message signals (of lower frequency) into modulated bandpass signals (of higher frequency) using a carrier signal for the purpose of enhancing the signal's immunity against noise and extending the transmission range. Different modulations require different hardware configurations and bandwidth allocations. Meanwhile, they provide different levels of noise immunity, data rate, and robustness in various transmission channels. In order to demodulate the modulated signals and to recover the transmitted message, the receiving end of the system must be equipped with the knowledge of the modulation type.

In military applications, modulations can serve as another level of encryption, preventing receivers from recovering the message without knowledge of the modulation type. On the other hand, if one hopes to recover the message from a piece of intercepted and possibly adversary communication signal, a modulation classifier is needed to determine the modulation type used by the transmitter. Apart from retrieving the transmitted message, modulation classification is also useful for identifying the transmitting unit and to generate jamming signals with matching modulations. The process is initially implemented manually with experienced signal engineers and later automated with automatic modulation classification systems to extend the range of operable modulations and to improve the overall classification performance.

In modern civilian applications, unlike in much earlier communication systems, multiple modulation types can be employed by a signal transmitter to control the data rate, to control the bandwidth usage, and to guarantee the integrity of the message. Though the pool of modulation types is known both to transmitting and receiving ends, the selection of the modulation type is adaptive and may not be known at the receiving end. Therefore, an automatic modulation classification mechanism is

required for the receiving end to select the correct demodulation approach in order to guarantee that the message can be successfully recovered.

This research monograph covers different algorithms developed for the automatic classification of communications signal modulation types. The theoretical signal models are explained in the first two chapters to provide the principles on which the analyses are based. An important step is to unify various signal models proposed in different studies and to provide a common framework for analysis of different automatic modulation classification algorithms.

This book includes the majority of the methods developed over the last two decades. The algorithms are systematically classified to five major categories: likelihood-based classifiers, distribution test-based classifiers, feature-based classifiers, machine learning-assisted classifiers, and blind modulation classifiers. For each type of automatic modulation classifier, the assumptions and system requirements are listed, and the design and implementation are explained through mathematical expressions, graphical illustrations and programming pseudo codes. Performance comparisons among several automatic modulation classifiers from each category are presented with both theoretical analysis and simulated numerical experiments. MATLAB® source code of selected methods will be available on https://code.google.com/p/amc-toolbox/.

The accumulated knowledge on the principle of automatic modulation classification and the characteristics of different automatic modulation classification algorithms is used to suggest the detailed implementation of modulation classifiers in specific civilian and military applications.

As the field is still developing, such a book cannot be definitive or complete. Nonetheless it is hoped that graduate students should be able to learn enough basics before studying journal papers; researchers in related fields should be able to get a broad perspective on what has been achieved; and current researchers as well as engineers in this field should be able to use it as a reference.

A work of this magnitude will unfortunately contain errors and omissions. We would like to take this opportunity to apologise unreservedly for all such indiscretions in advance. We welcome any comments or corrections; please send them by email to a.k.nandi@ieee.org or by any other means.

Zhechen Zhu and Asoke K. Nandi
London, UK

# List of Abbreviations

| | |
|---|---|
| AD | Anderson–Darling |
| ALRT | Average likelihood ratio test |
| AM | Amplitude modulation |
| AMC | Automatic modulation classification |
| AM&C | Adaptive modulation and coding |
| ANN | Artificial neural network |
| ASK | Amplitude-shift keying |
| AWGN | Additive white Gaussian noise |
| BMC | Blind modulation classification |
| BP | Back propagation |
| BPL | Broadband over power line |
| BPSK | Binary phase-shift keying modulation |
| CDF | Cumulative distribution function |
| CDP | Cyclic domain profile |
| CSI | Channel state information |
| CvM | Cramer–von Mises |
| CWT | Continuous wavelet transform |
| DFT | Discrete Fourier transform |
| DLRT | Discrete likelihood ratio test |
| DSB | Double-sideband modulation |
| DSSS | Direct sequence spread frequency |
| EA | Electronic attack |
| ECDF | Empirical cumulative distribution function |
| ECM | Expectation/condition maximization |
| EM | Expectation maximization |
| EP | Electronic protect |
| ES | Electronic support |
| EW | Electronic warfare |
| FB | Feature-based |
| FHSS | Frequency-hopping spread spectrum |

| | |
|---|---|
| FM | Frequency modulation |
| FSK | Frequency-shift keying |
| GA | Genetic algorithm |
| GLRT | Generalized likelihood ratio test |
| GMM | Gaussian mixture model |
| GoF | Goodness of fit |
| GP | Genetic programming |
| HLRT | Hybrid likelihood ratio test |
| HoS | High-order statistics |
| ICA | Independent component analysis |
| I-Q | In-phase and quadrature |
| KNN | K-nearest neighbour |
| KS | Kolmogorov–Smirnov |
| LA | Link adaptation |
| LB | Likelihood-based |
| LF | Likelihood function |
| LPD | Low probability of detection |
| LSB | Lower sideband modulation |
| LUT | Lookup table |
| MAP | Maximum a posteriori |
| MDLF | Minimum distance likelihood function |
| MIMO | Multiple-input and multiple-output |
| ML | Maximum likelihood |
| MLP | Multi-layer perceptron |
| MSE | Mean squared error |
| M-ASK | M-ary amplitude shift keying modulation |
| M-FSK | M-ary frequency shift keying modulation |
| M-PAM | M-ary pulse amplitude modulation |
| M-PSK | M-ary phase-shift keying modulation |
| M-QAM | M-ary quadrature amplitude modulation |
| ML-M | Magnitude-based maximum likelihood classifier |
| ML-P | Phase-based maximum likelihood classifier |
| NPLF | Non-parametric likelihood function |
| ODST | Optimized distribution sampling test |
| PAM | Pulse amplitude modulation |
| PD | Phase difference |
| PDF | Probability density function |
| PM | Phase modulation |
| PSK | Phase-shift keying modulation |
| QAM | Quadrature amplitude modulation |
| QPSK | Quadrature phase-shift keying modulation |

| | |
|---|---|
| SC | Spectral coherence |
| SCF | Spectral correlation function |
| SISO | Single-input and single-output |
| SM | Spatial multiplexing |
| SNR | Signal-to-noise ratio |
| SSB | Single-sideband modulation |
| STC | Space-time coding |
| SVM | Support vector machine |
| SαS | Symmetric alpha stable |
| USB | Upper sideband modulation |
| VSB | Vestigial sideband modulation |

# List of Symbols

| | |
|---|---|
| $A$ | Modulation alphabet |
| $H$ | Channel matrix |
| $h$ | Channel coefficient |
| $I$ | In-phase segment |
| $L$ | Number of signal realizations |
| $N$ | Number of samples / signal length |
| $Q$ | Quadrature segment |
| $P_{cc}$ | Classification accuracy |
| $\boldsymbol{r}$ | Received signal |
| $\boldsymbol{s}$ | Transmitted signal |
| $W$ | Bandwidth |
| $w$ | Weight |
| $\mathfrak{I}$ | Imaginary component |
| $\mathfrak{M}$ | Candidate modulation pool |
| $\mathfrak{R}$ | Real component |
| $\boldsymbol{\Theta}$ | Channel parameter set |
| $\Lambda$ | Likelihood ratio |
| $\sigma^2$ | Signal variance |
| $\omega$ | Additive noise |
| $\mathcal{H}$ | Hypothesis model |
| $\mathcal{L}$ | Likelihood |
| $\log \mathcal{L}$ | Log-likelihood |
| $\mathcal{M}$ | Modulation |
| $\mathbb{F}$ | Modulation classification feature |
| $\mathbf{F}$ | Modulation classification feature set |

# 1

# Introduction

## 1.1 Background

Automatic modulation classification (AMC) was first motivated by its application in military scenarios where electronic warfare, surveillance and threat analysis requires the recognition of signal modulations in order to identify adversary transmitting units, to prepare jamming signals, and to recover the intercepted signal. The term 'automatic' is used as opposed to the initial implementation of manual modulation classification where signals are processed by engineers with the aid of signal observation and processing equipment. Most modulation classifiers developed in the past 20 years are implemented through electronic processors. During the 1980s and 1990s there were considerable numbers of researchers in the field of signal processing and communications who dedicated their work to the problem of automatic modulation classification. This led to the publication of the first well received book on the subject by Azzouz and Nandi (1996). The interest in AMC for military purposes is sustained to this very day.

The beginning of twenty-first century saw a large number of innovations in communications technology. Among them are few that made essential contributions to the staggering increase of transmission throughput in various communication systems. Link adaptation (LA), also known as adaptive modulation and coding (AM&C), creates an adaptive modulation scheme where a pool of multiple modulations are employed by the same system (Goldsmith and Chua, 1998). It enables the optimization of the transmission reliability and data rate through the adaptive selection of modulation schemes according to channel conditions. While the transmitter has the freedom to choose how the signals are modulated, the receiver must have the knowledge of the modulation type to demodulation the signal so that the transmission can be successful. An easy way to achieve that is to include the modulation information in each signal frame so that the receivers would be notified about the change in modulation scheme,

*Automatic Modulation Classification: Principles, Algorithms and Applications*, First Edition. Zhechen Zhu and Asoke K. Nandi.

and react accordingly. However, this strategy affects the spectrum efficiency due to the extra modulation information in each signal frame. In the current situation where the wireless spectrum is extremely limited and valuable, the aforementioned strategy is simply not efficient enough. For this reason, AMC becomes an attractive solution to the problem. Thanks to the development in microprocessors, receivers nowadays are much enabled in terms of their computational power. Thus, the signal processing required by AMC algorithms becomes feasible. By automatically identifying the modulation type of the received signal, the receiver does not need to be notified about the modulation type and the demodulation can still be successfully achieved. In the end, spectrum efficiency is improved as no modulation information is needed in the transmitted signal frame. AMC has become an integral part of intelligent radio systems, including cognitive radio and software-defined radio.

Over the years, there have been many terms used to describe the same problem: modulation recognition, automatic modulation recognition, modulation identification, modulation classification, and automatic modulation classification. There are other names for the problem, such as PSK (phase-shift keying modulation) classification and M-QAM (M-ary quadrature amplitude modulation) classification that have a more specific target but which still operate under the same principle of automatic modulation classification. In this book, we have decided to use automatic modulation classification and AMC as a consistent reference to the same problem.

## 1.2 Applications of AMC

Having discussed the possible use of AMC in both military and civilian scenarios, in this section we take a close look at how AMC is incorporated in different military and civilian systems.

### *1.2.1 Military Applications*

AMC has an essential role in many military strategies. Modern electronic warfare (EW) consists of three major components: electronic support (ES), electronic attack (EA) and electronic protect (EP) (Poisel, 2008). In ES, the goal is to gather information from radio frequency emissions. This is often where AMC is employed after the signal detection is successfully achieved. The resulting modulation information could have several uses extending into all the components in EW. An illustration of how a modulation classifier is incorporated in the military EW systems is given in Figure 1.1.

To further the process of ES, the modulation information can be used for demodulating the intercepted signal in order to recover the transmitted message among adversary units. This is of course completed with the aid of signal decryption and translation. Meanwhile, the modulation information alone can also provide vital information to the electronic mapping system where it could be used to identify the adversary units and their possible locations.

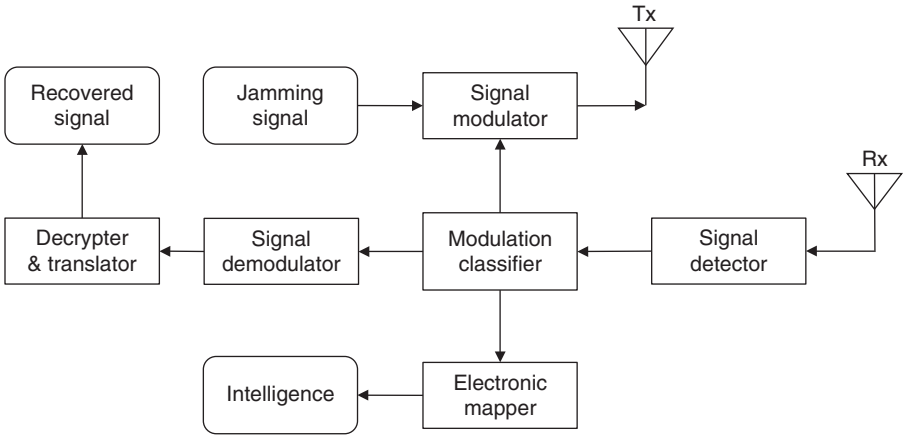



**Figure 1.1** Military signal intelligence system.

In EA, jamming is the primary measure to prevent the communication between adversary units. There are many jamming techniques available. However, the most common one relies on deploying jammers in the communication channel between adversary units and also transmitting noise signals or made-up signals using the matching modulation type. To override the adversary communication, the jamming signal must occupy the same frequency band as the adversary signal. This information is available from the signal detector. The power of the jamming signal must be significantly high, which is achieved by using an amplifier before transmitting the jamming signal. More importantly, the jamming signal must be modulated using the modulation scheme detected by the modulation classifier.

In EP, the objective is to protect friendly communications from adversary EA measures. As mentioned above, jammers transmit higher power signals to override adversary communication in the same frequency band. The key is to have the same signal modulation. An effective strategy to prevent friendly communication being jammed is to have awareness of the EA effort from adversary jammers and to dodge the jamming effort. More specifically, the friendly transmitter could monitor the jamming signal's modulation and switch the friendly unit to a different modulation scheme to avoid jamming.

### *1.2.2 Civilian Applications*

In the civilian scene, AMC is most important for the application of LA. As demonstrated in Figure 1.2, the signal modulator in the LA transmitter is replaced by an adaptive modulation unit. The role of the adaptive modulator is to select the modulation from a predefined candidate pool and to complete the modulation process. The selection of modulation from the candidate pool is determined by the system

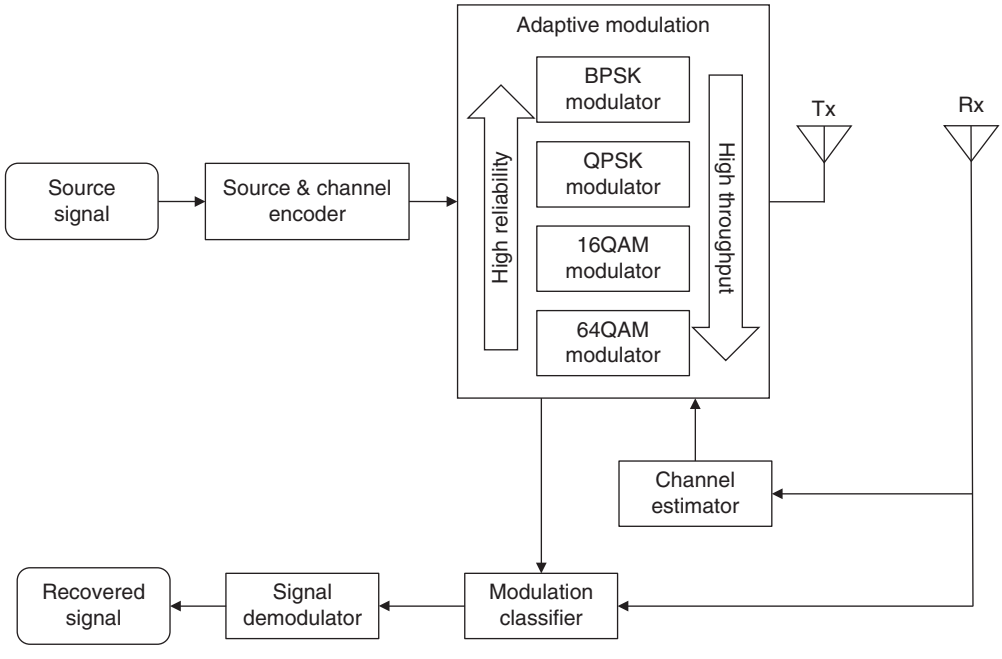



**Figure 1.2** AMC in link adaptation system.

specification and channel conditions. The lower-order and more robust modulations such as BPSK (binary phase-shift keying modulation) and QPSK (quadrature phase-shift keying modulation) are often selected when the channel is noisy and complex, given that the system requires high link reliability. The higher-order and more efficient modulations such as 16-QAM (16-quadrature amplitude modulation) and 64-QAM (64-quadrature amplitude modulation) are often selected to satisfy the demand for high-speed transmission in clear channels. The only communication between adaptive modulation module and the receiver is completed at system initialization where the information of the modulation candidate pool is notified to the receiver. During normal transmission the adaptive modulator embeds no extra information in the communication stream. At the receiving end of the LA system, channel estimation is performed prior to other tasks. If the channel is static, the estimation is only performed at the initial stage. If the channel is time variant, the channel state information (CSI) could be estimated regularly throughout the transmission. The estimated CSI and other information would then be fed back to the transmitter where the CSI will be used for the selection of modulation schemes. More importantly, the CSI is required to assist the modulation classifier. Depending on the AMC algorithm, different channel parameters are needed to complete the modulation classification. Normally, the accuracy of channel estimation has a significant impact on the

performance of the modulation classifier. The resulting modulation classification decision is then fed to the reconfigurable signal demodulator for appropriate demodulation. If the modulation classification is accurate, the correct demodulation method would capture the message and complete the successful transmission. If the modulation classification is incorrect, the entire transmission fails as the message cannot be recovered from the demodulator. It is not difficult to see the importance of AMC in LA systems.

## 1.3 Field Overview and Book Scope

Given the importance of AMC in various military and civilian communication applications, there has been a large amount of research work dedicated to the problem of AMC in a wide variety of settings. The nature of the problem creates multiple dimensions in its solutions and inspires continuous contribution from generations of researchers.

First, the modulation classifier needs to be accurate. The accuracy is measured by the percentage of errors made in a number of signal frames being transmitted. The lower the error the better the classifier is perceived to be. The likelihood-based classifiers first introduced by Polydoros and Kim (1990) provides optimal classification accuracy given matching signal model and perfect CSI knowledge. The approach has since been indulged by many researchers and led to many likelihood classifiers with various traits (Wei and Mendel, 2000; Hameed, Dobre and Popescu, 2009; Chavali and Da Silva, 2011; Xu, Su and Zhou, 2011; Shi and Karasawa, 2012).

Second, the modulation classifier needs to be robust. Since the communication channel can be unpredictable, especially in wireless channels, the classifier needs to have consistent classification accuracy in various channel conditions. In reality, conditions like multi-path fading, shadowing, Doppler effect, and additive noise have significant impact on the classification accuracy. Most of the works on AMC consider additive white Gaussian noise (AWGN) as a standard channel condition when evaluating their algorithms (Gardner and Spooner, 1988; Nandi and Azzouz, 1998; Swami and Sadler, 2000; Wei and Mendel, 2000). However, the consideration of fading channel and impulsive noises has become necessary for practically application of an AMC algorithm and has since been featured in many recent publications (Headley and Da Silva, 2011; Chavali and Da Silva, 2013).

Third, the classifier needs to be computationally efficient. The computational cost is reflected in two aspects of system performance. A complex AMC algorithm requires more powerful hardware to support it. In addition, a complex algorithm requires a longer time to complete the classification process, which may render certain applications unsuited if real-time decisions are needed. With some of the most fundamental AMC algorithms established, we are seeing more and more works which contribute to improve the computational complexity of some state-of-the-art algorithms

(Wong and Nandi, 2008; Wang and Wang, 2010; Xu, Su and Zhou, 2011). With the popularity of mobile communication, computational efficiency will remain a major consideration in the development of AMC algorithms.

Fourth, the classifier needs to be versatile. The versatility of an AMC classifier consists of many aspects. The classifier needs to handle as many modulation types as possible. The classifier needs to be operable in scenarios where limited knowledge of the channel or the communication system is available. The classifier needs to provide information other than the modulation type as by-product in real time. The classifier needs to be applicable in various communication systems such as single-input and single-output systems (SISO) and multiple-input and multiple-output systems (MIMO). The classification of both analogue and digital modulation can be effectively achieved by multiple signal features suggested by Azzouz and Nandi (1996). The current focus of versatility of an AMC algorithm lies in the classification of multi-ordered digital modulations in MIMO systems (Choqueuse *et al.*, 2009; Hassan *et al.*, 2012; Mühlhaus *et al.*, 2013).

In this book, we focus on the more related issues in the current AMC development environment. We will revisit most of the existing AMC algorithms and sketch their implementations under a unified signal model (Chapter 2). The classifiers will be classified into five categories and presented in five chapters (from Chapters 3 to 7). As these classifiers all have their strengths and weaknesses, we will exam some of the key algorithms from each category and assess their performance in simulated environments (Chapter 8). The simulation focuses on digital modulations that are most relevant to the current communication systems. As we develop a comprehensive understanding of all the algorithms and their characteristics, in Chapters 9 and 10, we attempt to suggest designs of AMC algorithms that are tailored to some of the specific applications in civilian and military scenarios.

## 1.4 Modulation and Communication System Basics

To familiarize the readers with the technical concepts that are used in this book, we dedicate this section to the basics of communications theory.

### *1.4.1 Analogue Systems and Modulations*

We assume the source signal $x(t)$ is analogue, non-negative and continuous at time $t$. In analogue systems, the signal is modulated before transmission using analogue modulations. Depending on the modulation type, the modulator is preconfigured and not subject to future change. Here we consider three types of analogue modulation, namely amplitude modulation (AM), frequency modulation (FM), and phase modulation (PM). An illustration of the analogue radio communication system is given in Figure 1.3.

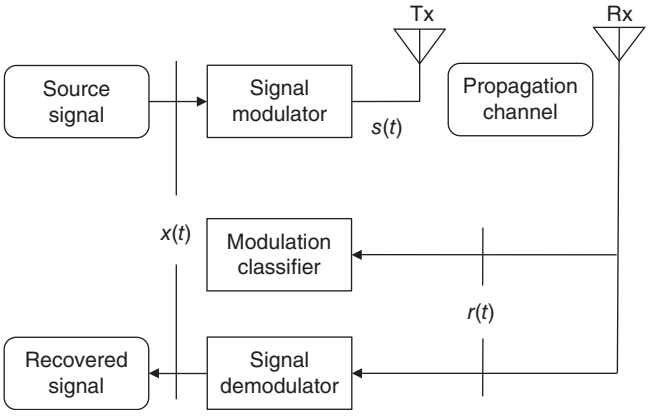



**Figure 1.3** Analogue communication system.

For AM modulation, the signal is modulated with a carrier signal $c(t) = A \cdot \cos(2\pi f_c t)$ with a carrier frequency $f_c$. The source signal is multiplied by the carrier signal to create the transmitted modulation signal $s(t)$ given by equation (1.1).

$$s(t) = x(t) \cdot A \cdot \cos(2\pi f_c t) \tag{1.1}$$

The resulting transmitted signal is called the bandpass signal where the source signal is embedded in the signal amplitude envelope. Figure 1.4 gives examples of signal waveform of the carrier signal, the source signal, and the modulation signal using AM.

For FM modulations, the same carrier can be used. However, the source signal is added to modify the frequency component of the carrier signal. The time series FM modulation signal is given by equation (1.2),

$$s(t) = A \cdot \cos\left(2\pi f_c t + 2\pi \Delta f \int_0^t x(\tau) d\tau\right) \tag{1.2}$$

where $\Delta f$ is the frequency deviation that controls the variation of the modulated signal frequency. The waveform of the FM signal is given in Figure 1.5.

For PM modulation, the source signal is added to the carrier signal by modifying the signal phase. The expression of the PM modulation signal is found as shown in equation (1.3).

$$s(t) = A \cdot \cos(2\pi f_c t + x(t)) \tag{1.3}$$

The waveform of a PM signal is given in Figure 1.6.

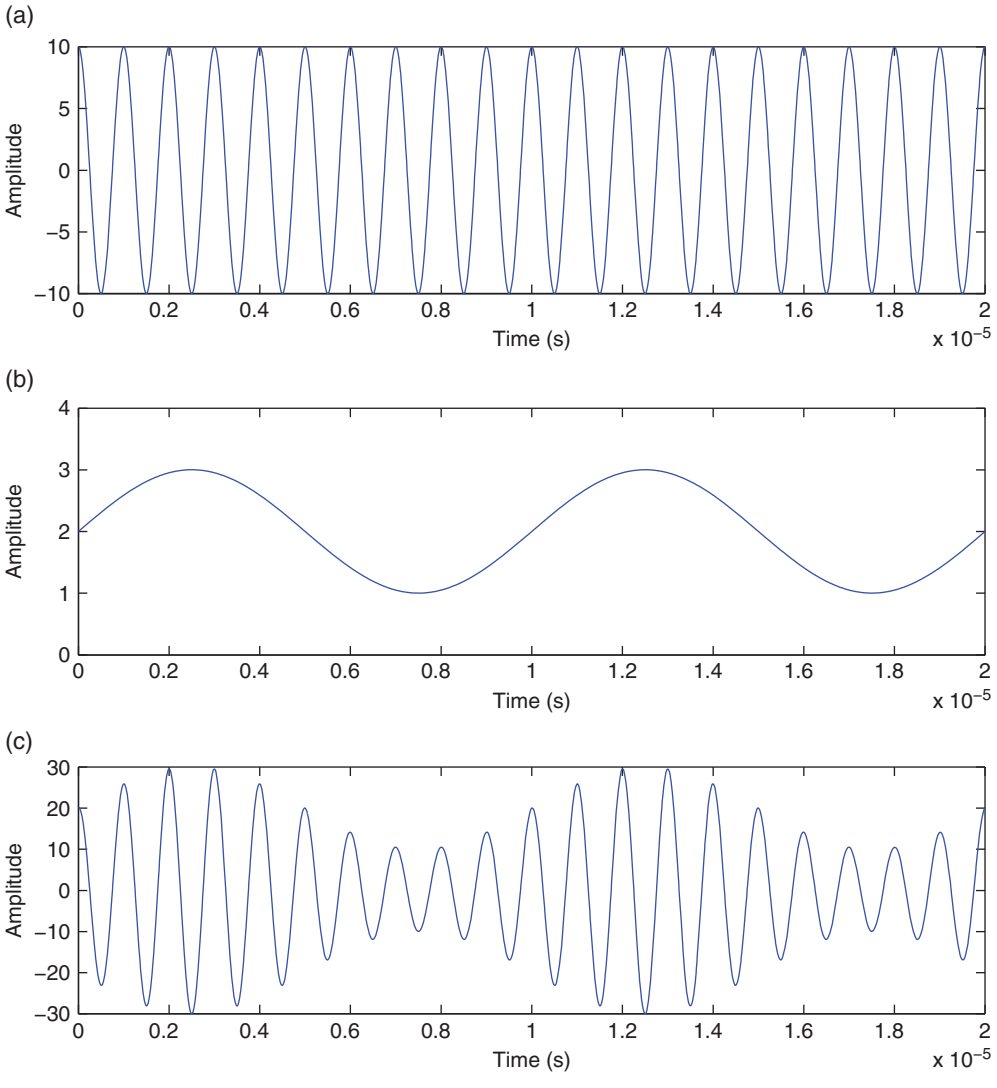



**Figure 1.4** (a) Carrier signal, (b) source signal and (c) AM signal.

### *1.4.2 Digital Systems and Modulations*

Modern communication systems make less use of analogue modulations. Instead, digital modulations are well favoured thanks to a better match with digital data and stronger immunity against interference. Notable digital modulation types include amplitude-shift keying (ASK), frequency-shift keying (FSK), phase-shift keying (PSK),

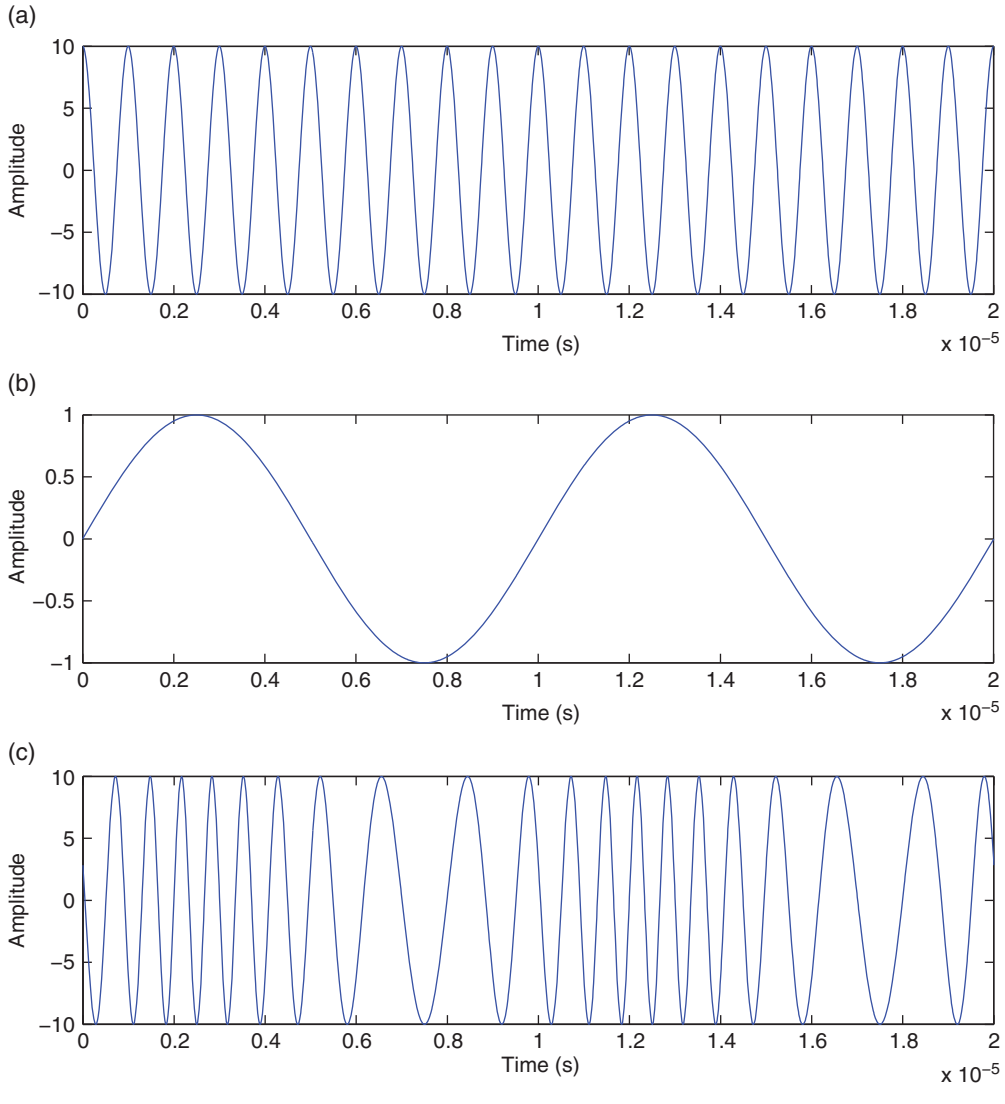



**Figure 1.5** (a) Carrier signal, (b) source signal and (c) FM signal.

pulse amplitude modulation (PAM), and quadrature amplitude modulation (QAM). To meet the demand for higher transmission throughput, digital modulations of higher orders including M-ary ASK (M-ASK), M-ary FSK (M-FSK), M-ary PSK (M-PSK), M-ary PAM (M-PAM), and M-ary QAM (M-QAM) are often used. The label "M" indicates the number of samples in the modulation alphabet set.

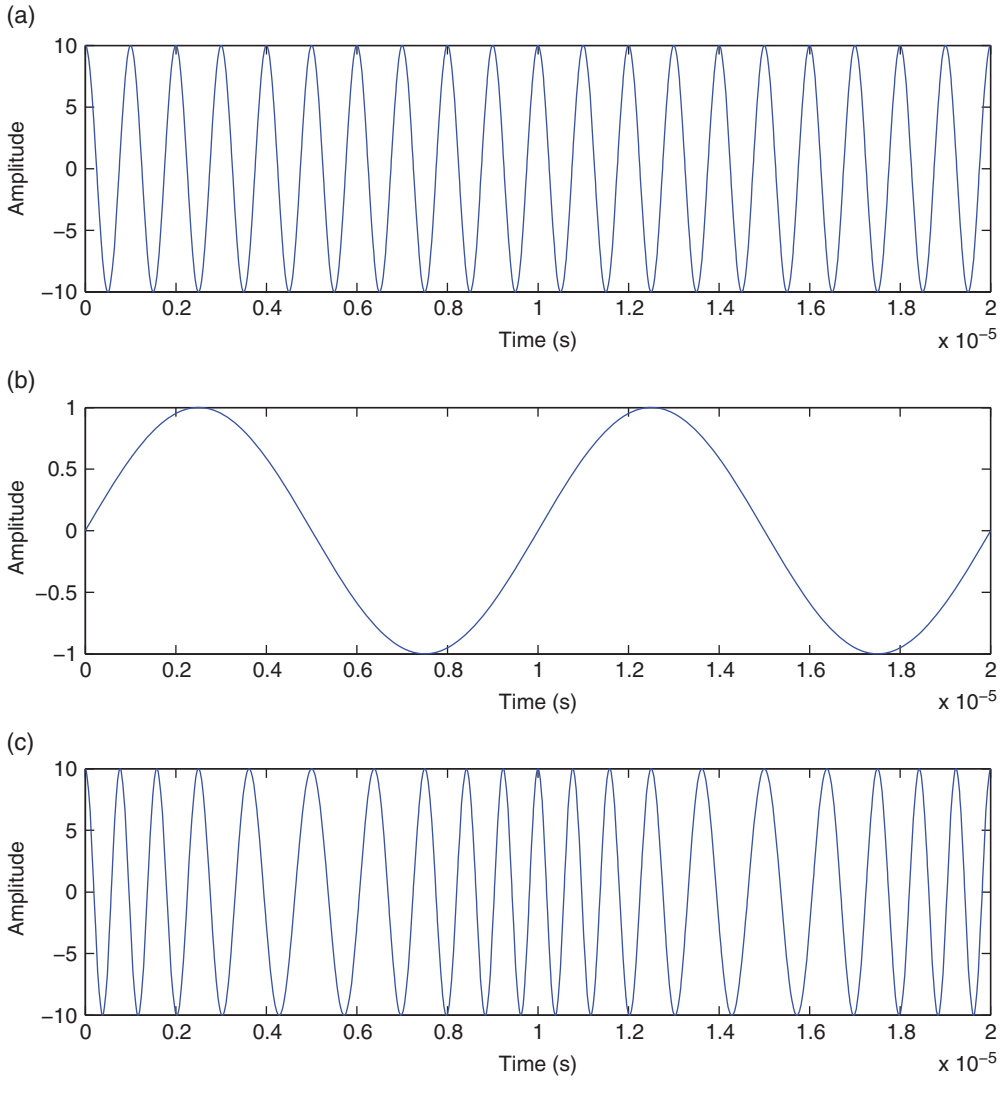



**Figure 1.6** (a) Carrier signal, (b) source signal and (c) PM signal.

Like the analogue systems, we have a source signal $x(t)$. The source signal is digitized by sampling and quantization, the resulting digital signal is then coded by various means for the purpose of data security and limiting transmission errors. In the context of AMC, the digitization process and coding theory makes little impact on the classifier design and classification performance. Therefore, we neglect the details of these processes and assume the end product to be $u[n] \sim x(t)$, $(n-1)T < t < nT$, where $u[n]$

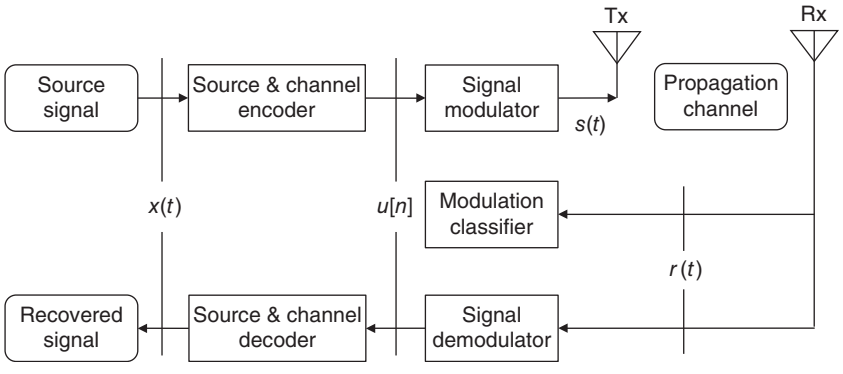



**Figure 1.7** Digital communication system.

is the $n$th source signal symbol and $T$ is the symbol timing for the digitized source signal. Depending on the modulation type, the modulated signal is generated in different ways. An illustration of the digital radio communication system is given in Figure 1.7.

For ASK modulations, the expression for the modulated signal given in equation (1.4) is very similar to that for AM, the only difference being that the source signal is digital instead of analogue.

$$s(t) = u(t) \cdot A \cdot \cos(2\pi f_c t) \tag{1.4}$$

The waveform of an ASK signal is given in Figure 1.8.

For FSK modulations, the expression also matches the one from FM modulation, as shown in equation (1.5).

$$s(t) = A \cdot \cos\left(2\pi f_c t + 2\pi \Delta f \int_0^t u(\tau) d\tau\right) \tag{1.5}$$

The waveform of an FSK signal is given in Figure 1.9.

For PSK modulation, the expression is similar to the one for the PM modulation, as shown in equation (1.6).

$$s(t) = A \cdot \cos(2\pi f_c t + \pi u(t)) \tag{1.6}$$

The waveform of FSK signal is given in Figure 1.10.

PAM modulation shares a similar principle of embedding the source signal in the amplitude of the carrier signal. However, the expression of the modulated signal, as shown in equation (1.7), differs due to the added pulse-shaping factor $g(t)$.

$$s(t) = u(t) \cdot g(t) \cdot A \cdot \cos(2\pi f_c t) \tag{1.7}$$

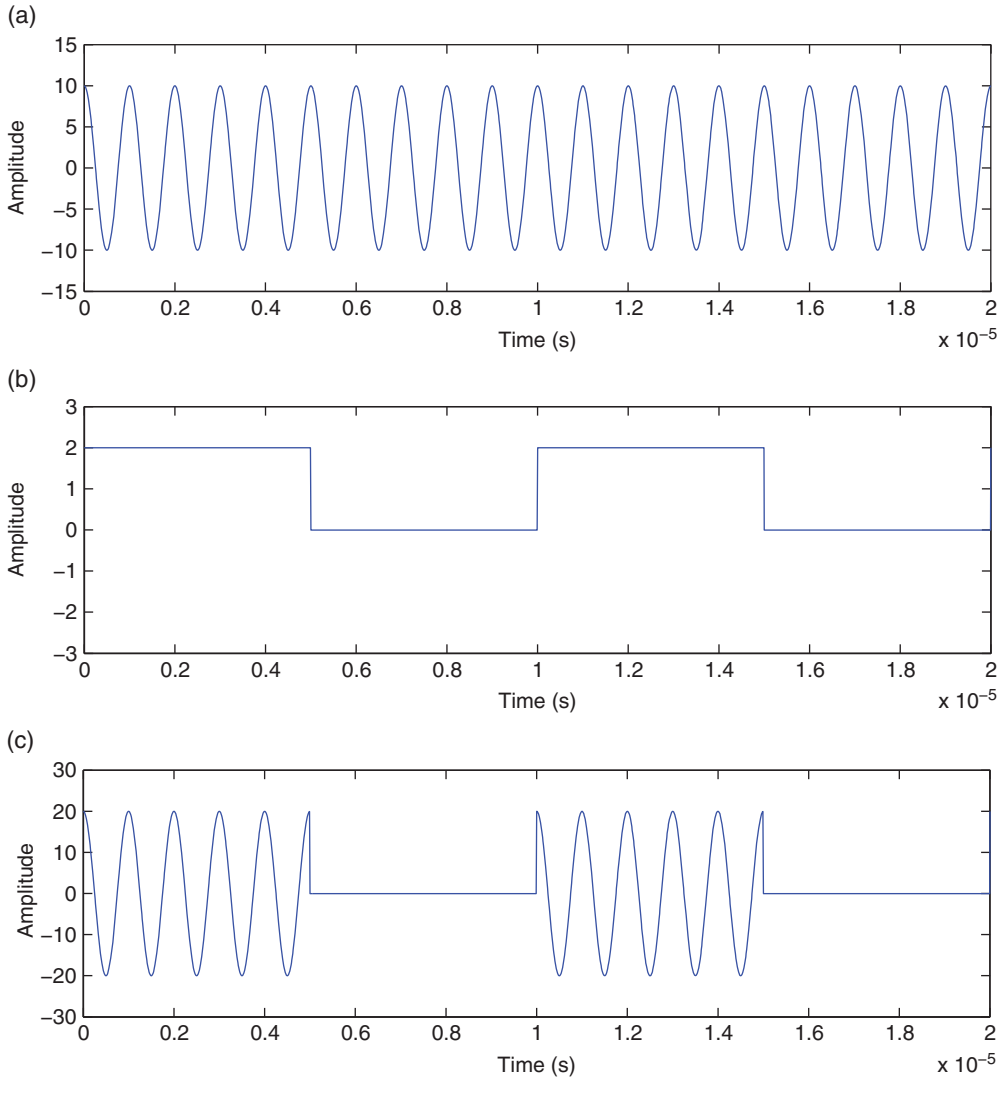



**Figure 1.8** (a) Carrier signal, (b) source signal and (c) ASK signal.

The factor $g(t)$ defines the shape of the pulse. A common example is the raised cosine frequency shaping filter, which is defined by equation (1.8), where α is the roll-off factor between 0 and 1.

$$g(t) = \mathrm{sinc}\left(\frac{t}{T}\right)\frac{\cos(\pi\alpha t/T)}{1-4\alpha^2 t^2/T^2} \tag{1.8}$$

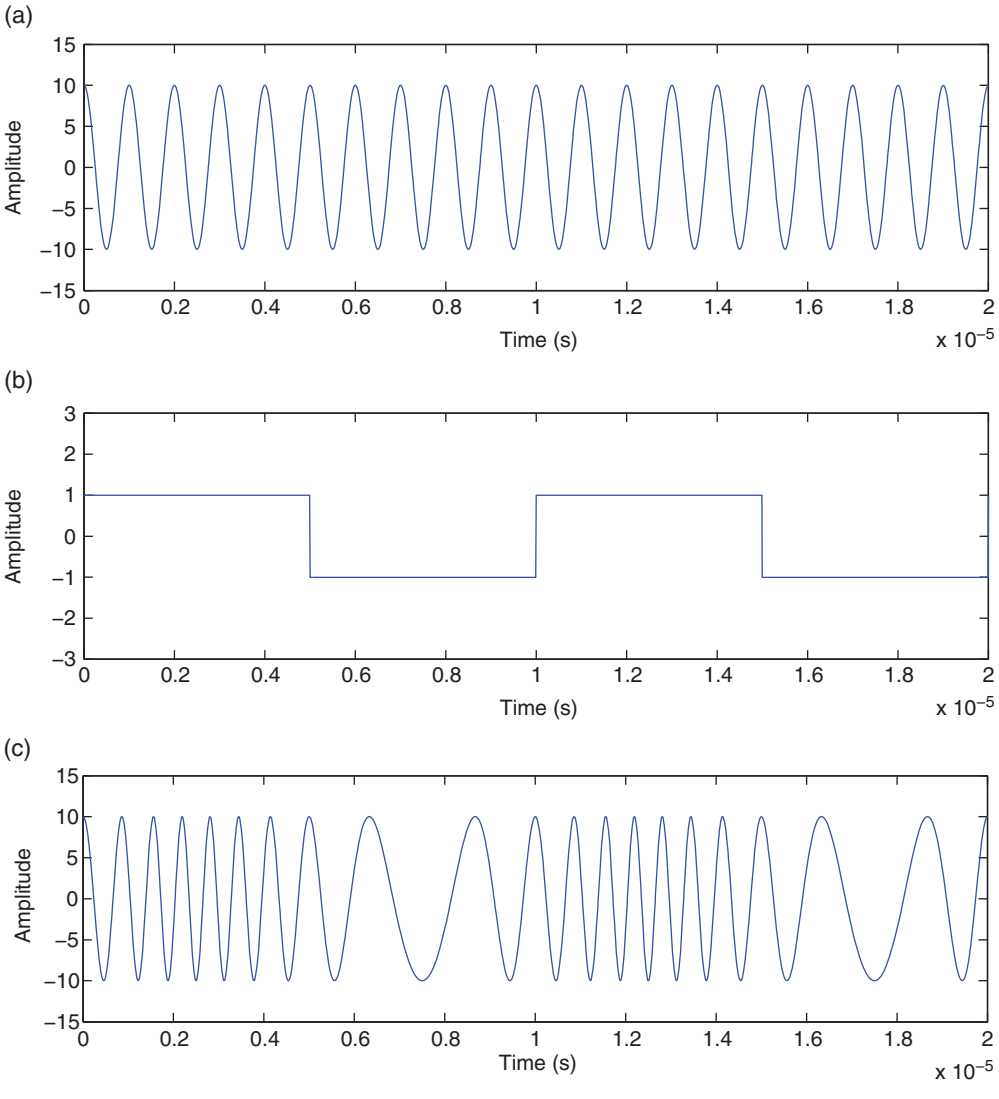



**Figure 1.9** (a) Carrier signal, (b) source signal and (c) FSK signal.

QAM modulations are similar to PAM modulations and are often considered as a combination of PSK and ASK modulations. Instead of a real-valued source signal, the source signal is mapped to a complex baseband waveform, equation (1.9).

$$u(t) = \Re\{u(t)\} + j\Im(u(t)) \tag{1.9}$$

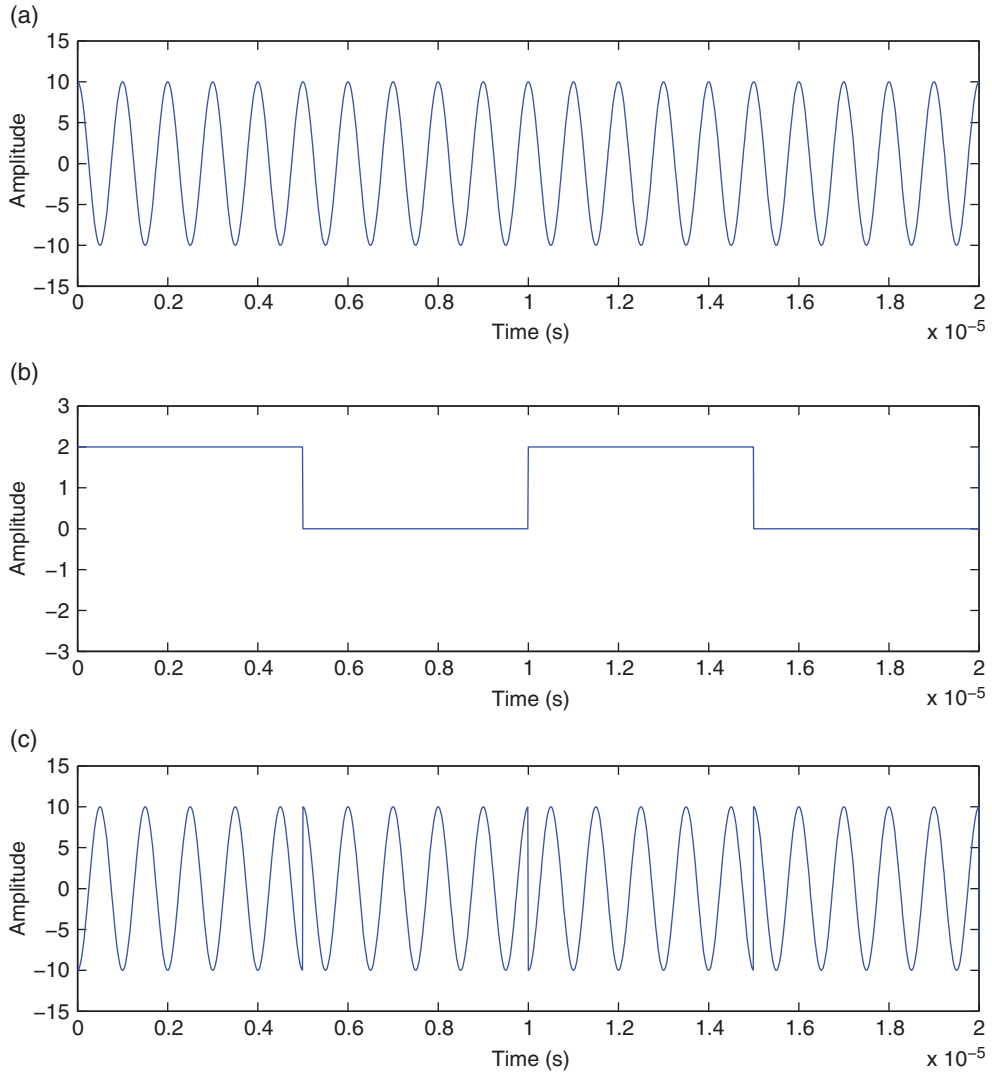



**Figure 1.10** (a) Carrier signal (b) source signal and (c) PSK signal.

The QAM modulated signal is composed as shown in equation (1.10), where $|u(t)|$ is the magnitude of the complex baseband signal and $\arg\{u(t)\}$ is the phase of the complex baseband signal.

$$s(t) = |u(t)|\cos\left(\arg\{u(t)\}\right)\cos(2\pi f_c t) - |u(t)|\sin\left(\arg\{u(t)\}\right)\sin(2\pi f_c t) \tag{1.10}$$

For digital modulations, the signal is often visualized using a constellation plot where the in-phase and quadrature (I-Q) components of a signal are used to provide

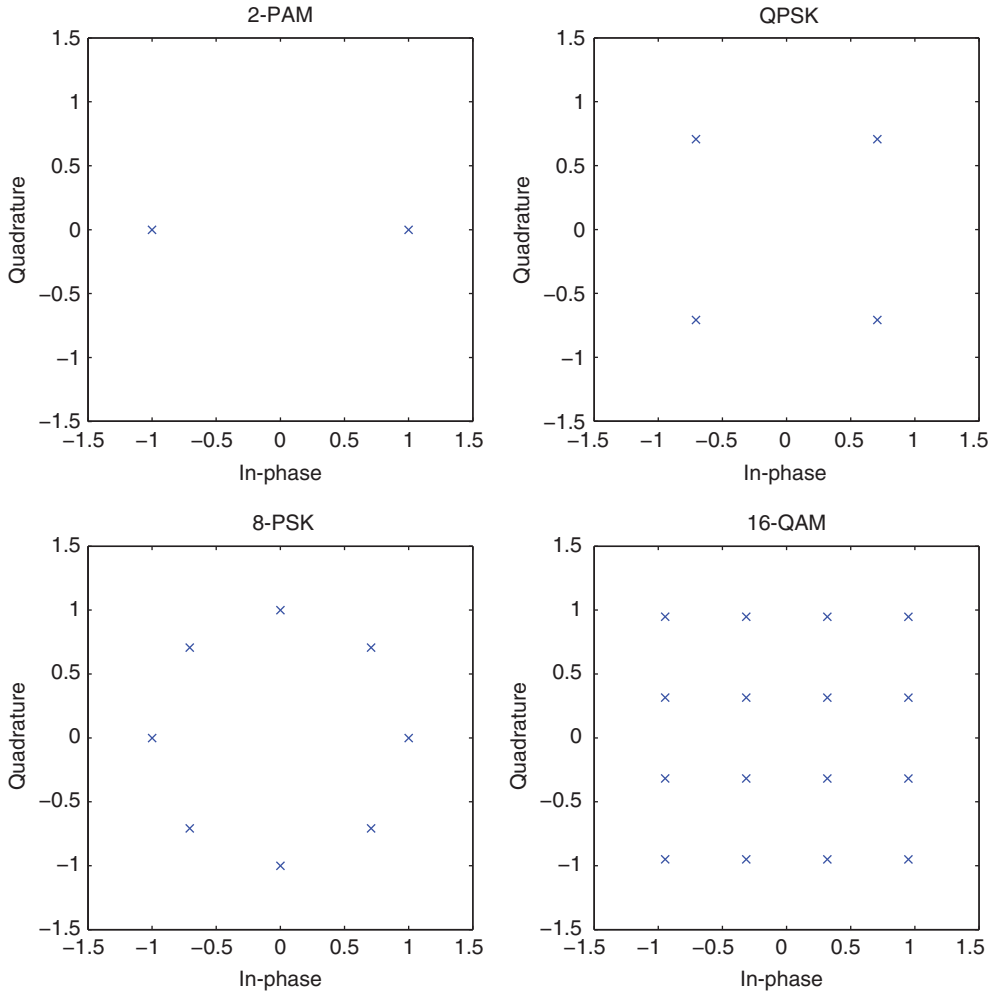



**Figure 1.11** Constellation plots of 2-PAM, QPSK, 8-PSK and 16-QAM.

coordinates. Constellation plots of 2-PAM, QPSK, 8-PSK, and 16-QAM modulations are given in Figure 1.11 as examples.

### 1.4.3 Received Signal with Channel Effects

Regardless of the transmitter setting and modulation selection, the transmitted signals are subject to the same channel conditions. Here we give a signal model that includes a majority of the channel effect a single wireless radio frequency may encounter. The received signal is given by equation (1.11), where $\alpha$ is the channel gain, $f_o$ and $\theta_o$ are carrier frequency and phase offsets, $s(\tau)$ is the transmitted signal sample at time $\tau$,

$p(\cdot)$ is the pulse shaping, $h(\cdot)$ is the channel response, $\epsilon_T$ is the symbol timing error, and $\omega(\cdot)$ is the additive noise. The noise level associated with $\omega(\cdot)$ is often measured in signal-to-noise ratio (SNR). SNR is the ratio between the power of the transmitted symbols and the additive noises.

$$r(t) = \alpha e^{j(2\pi f_o t + \theta_o)} \int_{\tau=-\infty}^{\tau=\infty} s(\tau)p(\tau - t)h(t - \tau + \epsilon_T)d\tau + \omega(t) \tag{1.11}$$

The $n$th discrete signal sample at time $t = nT$ is given by equation (1.12),

$$r[n] = \alpha e^{j(2\pi f_o t + \theta_o)} \sum_{l=-\infty}^{\infty} s[n]p(lT - nT)h(nT - lT + \epsilon_T) + \omega[n] \tag{1.12}$$

where $s[n]$ is the $n$th transmitted signal sample and $\omega[n]$ is the additive noise component at $t = nT$.

Under common circumstances, assumptions are made that the pulse shaping and channel response is known to the receiver. Therefore, the signal model after matching the filter could be simplified as equation (1.13).

$$r[n] = \alpha e^{j(2\pi f_o t + \theta_o)} s[n] + \omega[n] \tag{1.13}$$

The majority of our analysis will be based on the sampled discrete signal using the above equation. More details on signal models in different channels are given in Chapter 2.

## 1.5 Conclusion

In this chapter, we introduce the historical background of AMC where it originated from military electronic warfare as an alternative to manual modulation classification. It is also highlighted that AMC holds an important position in modern civilian communications systems. A brief showcase of the implementation of AMC in military applications and civilian applications is given. An overview of the field of AMC study is given which leads to the scope of this book. The chapter is finished with some basic knowledge of communication systems that are considered in the remainder of the book.

## References

Azzouz, E.E. and Nandi, A.K. (1996) *Automatic Modulation Recognition of Communication Signals*, Kluwer, Boston.

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

Chavali, V.G. and Da Silva, C.R.C.M. (2013) Classification of digital amplitude-phase modulated signals in time-correlated non-gaussian channels. *IEEE Transactions on Communications*, **61** (6), 2408–2419.

Choqueuse, V., Azou, S., Yao, K. *et al.* (2009) Blind modulation recognition for MIMO systems. *Military Technical Academy Review*, **XIX** (2), 183–196.

Gardner, W.A. and Spooner, C.M. (1988) Cyclic Spectral Analysis for Signal Detection and Modulation Recognition. Military Communications Conference,San Diego, CA, USA, 23–26 October 1988, pp. 419–424.

Goldsmith, A.J. and Chua, S. (1998) Adaptive coded modulation for fading channels. *IEEE Transactions on Communications*, **46** (5), 595–602.

Hameed, F., Dobre, O.A. and Popescu, D. (2009) On the likelihood-based approach to modulation classification. *IEEE Transactions on Wireless Communications*, **8** (12), 5884–5892.

Hassan, K., Dayoub, I., Hamouda, W. *et al.* (2012) Blind digital modulation identification for spatially-correlated MIMO systems. *IEEE Transactions on Wireless Communications*, **11** (2), 683–693.

Headley, W.C. and Da Silva, C.R.C.M. (2011) Asynchronous classification of digital amplitude-phase modulated signals in flat-fading channels. *IEEE Transactions on Communications*, **59** (1), 7–12.

Mühlhaus, M.S., Jondral, F.K., Dobre, O.A. and Oner, M. (2013) A low complexity modulation classification algorithm for MIMO systems. *IEEE Communications Letters*, **17** (10), 1881–1884.

Nandi, A.K. and Azzouz, E.E. (1998) Algorithms for automatic modulation recognition of communication signals. *IEEE Transactions on Communications*, **46** (4), 431–436.

Poisel, A.R. (2008) *Introduction to Communication Electronic Warfare Systems*, Artech House, Norwood, MA.

Polydoros, A. and Kim, K. (1990) On the detection and classification of quadrature digital modulations in broad-band noise. *IEEE Transactions on Communications*, **38** (8), 1199–1211.

Shi, Q. and Karasawa, Y. (2012) Automatic modulation identification based on the probability density function of signal phase. *IEEE Transactions on Communications*, **60** (4), 1–5.

Swami, A. and Sadler, B.M. (2000) Hierarchical digital modulation classification using cumulants. *IEEE Transactions on Communications*, **48** (3), 416–429.

Wang, F. and Wang, X. (2010) Fast and robust modulation classification via Kolmogorov-Smirnov test. *IEEE Transactions on Communications*, **58** (8), 2324–2332.

Wei, W. and Mendel, J.M. (2000) Maximum-likelihood classification for digital amplitude-phase modulations. *IEEE Transactions on Communications*, **48** (2), 189–193.

Wong, M.L.D. and Nandi, A.K. (2008) Semi-blind algorithms for automatic classification of digital modulation schemes. *Digital Signal Processing*, **18** (2), 209–227.

Xu, J.L., Su, W. and Zhou, M. (2011) Likelihood-ratio approaches to automatic modulation classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, **41** (4), 455–469.

# 2

# Signal Models for Modulation Classification

## 2.1 Introduction

Signal models are the starting point of every meaningful modulation classification strategy. Algorithms such as likelihood-based (Huang and Polydoros, 1995; Wei and Mendel, 2000; Shi and Karasawa 2012), distribution test-based (Wang and Wang, 2010; Urriza *et al.*, 2011; Zhu, Aslam and Nandi, 2014) and feature-based classifiers (Azzouz and Nandi, 1996; Spooner, 1996; Swami and Sadler, 2000) all require an established signal model to derive the corresponding rules for classification decision making. While some unsupervised machine learning algorithms could function without a reference signal model, the optimization of such algorithms still relies on the knowledge of a known signal model. Meanwhile, as the validation of modulation classifiers are often realized by computer-aided simulation, accurate signal modelling provides meaningful scenarios for evaluating the performance of various modulation classifiers.

The objective of this chapter is to establish some unified signal models for the development of all modulation classifiers from Chapters 3 to 7, and to provide a level ground for the validation of each modulation classifier in Chapter 8. Through the process, the accuracy of the models will be the first priority as it provides credible evidence to aid the design of specific modulation classification strategies for real world applications in Chapters 9 and 10. That, however, is with a fine balance of simplicity in the models to enable theoretical analysis and to provide computationally efficient implementations.

*Automatic Modulation Classification: Principles, Algorithms and Applications*, First Edition. Zhechen Zhu and Asoke K. Nandi.

In this chapter, mathematical models of communication signals are instituted for three popular communication channels, namely additive white Gaussian noise channel, fading channel, and non-Gaussian noise channel.

Additive white Gaussian noise is one of the most widely used noise models in many signal processing problems. It is of much relevance to the transmission of signals in both wired and wireless communication media where wideband Gaussian noises are produced by thermal vibration in conductors and radiation from various sources. The popularity of additive white Gaussian noise is evidential in most literature on modulation classification where the noise model is considered the fundamental limitation to accurate modulation classification.

Fading channel is largely concerned with wireless communication, where signals are received as delayed and attenuated copies after being absorbed, reflected and diffracted by different objects. Fading, especially deep fading, drastically changes the property of the transmitted signal and imposes a tough challenge on the robustness of a modulation classifier. Though early literature on modulation classifier focused on the validation of algorithms in AWGN channel, the current standard requires the robustness in fading channel as an important trait. In this chapter, a unified model of a fading channel is presented with flexible representation of different fading scenarios. It is worth noting that AWGN noise will also be considered in the fading channel as to approach a more realistic real world channel condition.

Non-Gaussian noises are often used to model impulsive noises which are a step further to model the noises in a real radio communication channel. Impulsive noise, unlike Gaussian noise, has a heavy-tailed probability density function, meaning higher probability for high power noise components. Such noises are often the result of incidental electromagnetic radiation from man-made sources. While not featured in most modulation classification literature, impulsive noises have received an increasing amount of attention in recent years. Despite the complexity in the modelling of impulsive noise, it is worth the effort to try and accommodate the signal model for a more practical approximation of the real world radio channels. In this chapter, three non-Gaussian noise models will be presented for modelling the impulsive noise. However, such noises will be considered solely without extra AWGN noise or fading effects.

## 2.2 Signal Model in AWGN Channel

Additive white Gaussian noise is characterized with constant spectral density and a Gaussian amplitude distribution of zero mean. Giving the additive noise a complex representation $\omega = I(\omega) + jQ(\omega)$, the complex probability density function (PDF) of the complex noise can be found as equation (2.1), where $\Sigma$ is the covariance matrix of the complex noise, $|\Sigma|$ is the determinant of $\Sigma$, $|x|$ is the Euclidean norm of the complex noise and the noise mean is zero.

$$f_{\omega}(x) = \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-\frac{|x|^2}{2\sqrt{|\Sigma|}}} \tag{2.1}$$

Since many algorithms are interested in the in-phase and quadrature segments of the signal, it is important to derive the corresponding PDF of the in-phase and quadrature segments of the additive noise. Fortunately, when AWGN noises are projected onto any orthonormal segments the resulting projection has independent and identical Gaussian distributions. The resulting covariance matrix can be found as equation (2.2),

$$\Sigma = \begin{bmatrix} \sigma_I^2 & \rho\sigma_I\sigma_Q \\ \rho\sigma_I\sigma_Q & \sigma_Q^2 \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \tag{2.2}$$

where variance for the in-phase segment $\sigma_I^2$ and the quadrature segment $\sigma_Q^2$ and are replaced with a shared identical variance $\sigma^2$, and the correlation between two segments is zero. Thus, the desired PDFs of each segment can be easily derived as shown in equation (2.3).

$$f_{I(\omega)}(x) = f_{Q(\omega)}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{|x|^2}{2\sigma^2}} \tag{2.3}$$

As suggested by the term "additive", the AWGN noise is added to the transmitted signal to give the signal model in AWGN channel, as shown in equation (2.4).

$$r(t) = s(t) + \omega(t) \tag{2.4}$$

An illustration of the received 4-QAM signal in an AWGN channel with SNR of 10 dB is given in Figure 2.1.

In the following subsections, the PDFs of received signals in their I-Q segments, phase and magnitude distributions are derived.

### 2.2.1 Signal Distribution of I-Q Segments

Assuming the signal modulation $\mathcal{M}$ has an alphabet $A$ of $M$ symbols and the symbol $A_m$ having the equal probability to be transmitted, with overall distribution being considered as M number of AWGN noise distributions shifted to different modulation symbols, the complex PDF of the received signal is given by equation (2.5), where $1/M$ is the probability of $A_m$ being transmitted. Figure 2.2 provides an illustration of the distribution of 4-QAM signals in an AWGN channel with SNR of 10 dB on the I-Q plane.

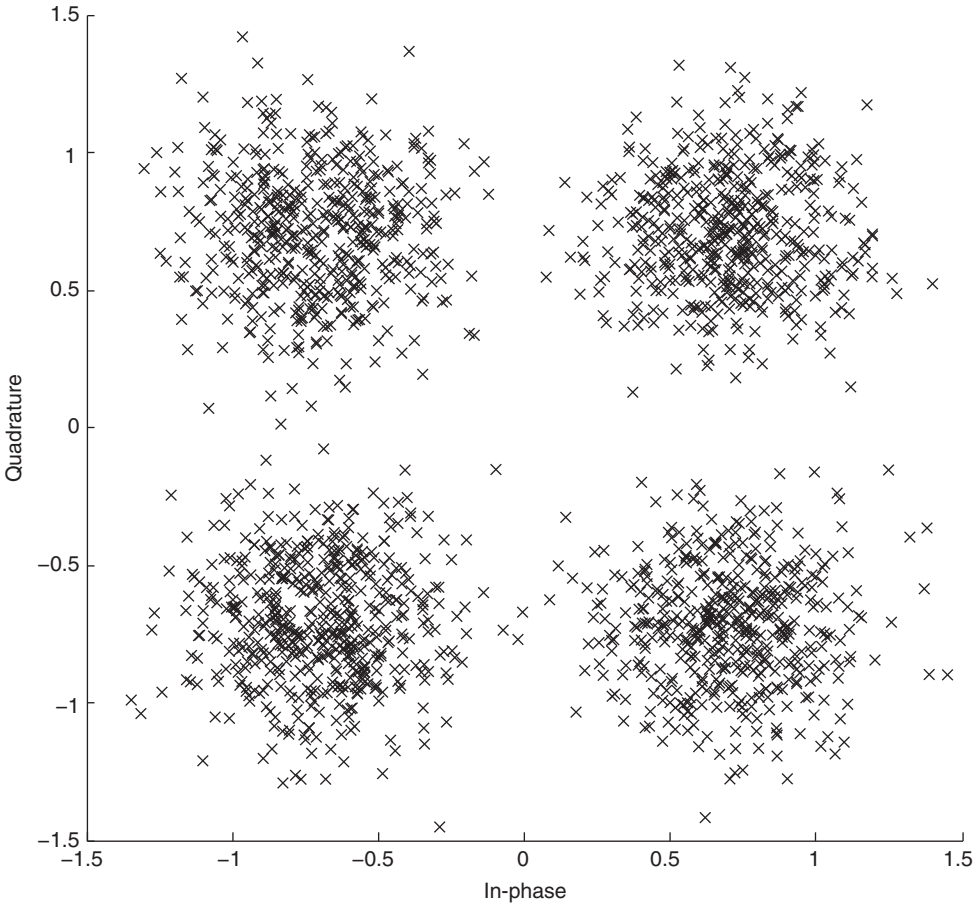



**Figure 2.1** Constellation of 4-QAM signal in AWGN with SNR = 10 dB.

$$f_r(x) = \sum_{m=1}^{M} \frac{1}{M} f_\omega(x|A_m, \Sigma) = \sum_{m=1}^{M} \frac{1}{M} \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-\frac{|x-A_m|^2}{2\sqrt{|\Sigma|}}} \tag{2.5}$$

Following the same logic of the derivation of the complex PDF, the distribution of received signals on their in-phase and quadrature segments can be found by replacing the variance by half of the noise variance and the mean of the noise distribution with in-phase and quadrature segments of the modulation symbols [equation (2.6)].

$$f_{I(r)}(x) = \sum_{m=1}^{M} \frac{1}{M} f_{I(\omega)}(x|A_m, \sigma) = \sum_{m=1}^{M} \frac{1}{M} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{|x-I(A_m)|^2}{2\sigma^2}} \tag{2.6}$$

An illustration of the PDF is given in Figure 2.3.

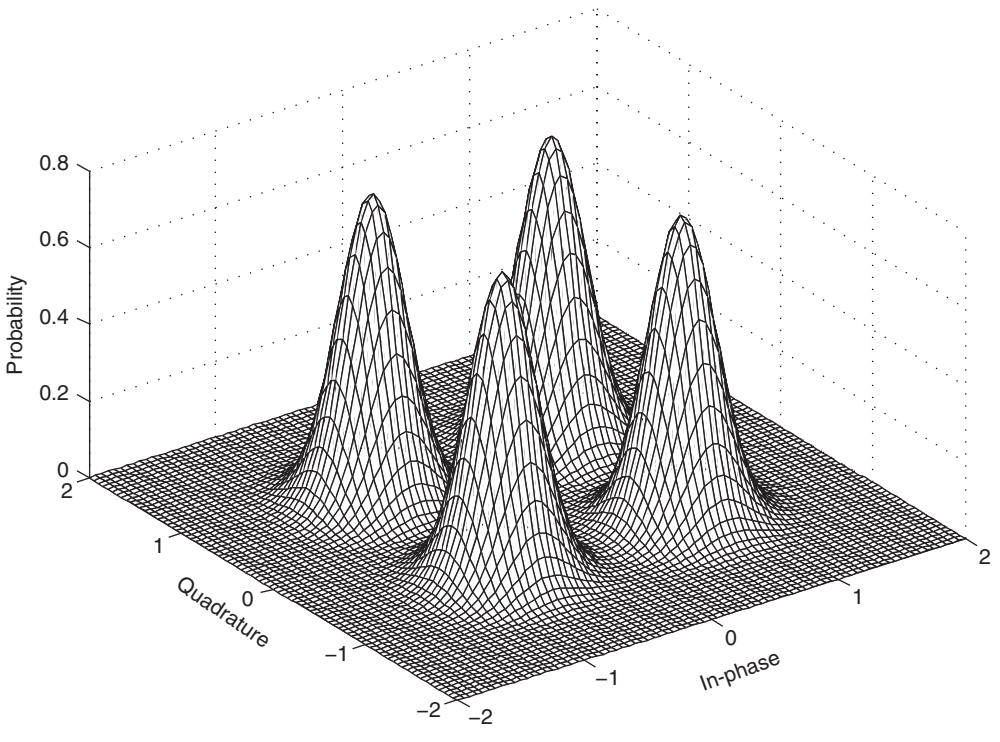



**Figure 2.2** PDF of 4-QAM signals in AWGN channel with SNR = 10 dB.

### 2.2.2 Signal Distribution of Signal Phase

Signal phase is easily recognized as an intuitive object for analysis when the classification of PSK modulations is concerned. That is without mentioning the robustness of signal phase in channel with attenuation. According to Bennett (1956), with added noise, received signal samples from the same modulation symbol $A_m$ have a phase PDF given by equation (2.7),

$$f_{\theta(r)}(x|A_m) = \frac{e^{-|A_m|^2/2\sigma^2}}{2\pi} + \frac{|A_m|\cos(x)}{2\sigma\sqrt{2\pi}}\left[1 + erf\left(\frac{|A_m|\cos(x)}{\sqrt{2}\sigma}\right)\right] e^{-\frac{|A_m|^2}{2\sigma^2}\sin^2(\theta(A_m))} \tag{2.7}$$

where $|A_m|$ is the magnitude of the symbol $A_m$, $\theta(A_m)$ is its phase, and $erf(\cdot)$ is the error function. Given the complex form of the PDF, the von Mises distribution is often considered as a close approximation [equation (2.8)], where $I_0(\cdot)$ denotes the modified Bessel function of order zero. The von Mises distribution is used primarily in this book for analytical content. However, it is worth noting that the von Mises distribution deviates from the accurate phase PDF at low SNR. Therefore, validations are provided in cases where the von Mises approximation is used.

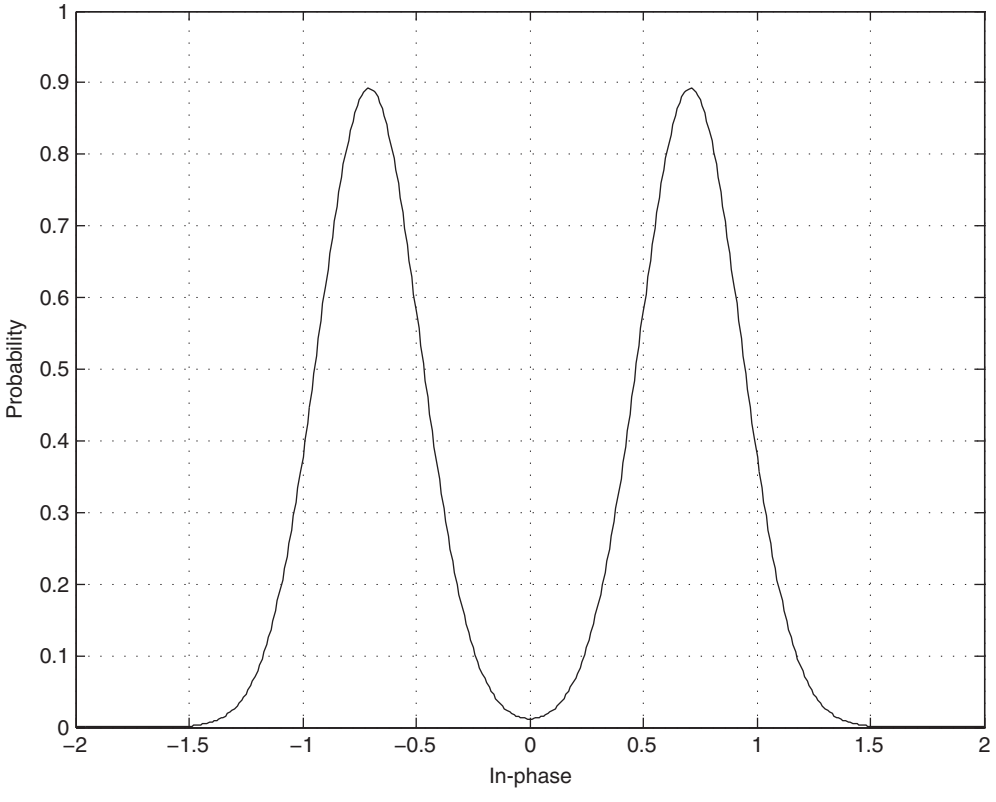



**Figure 2.3** PDF of 4-QAM signals I-Q segments in AWGN channel with SNR = 10 dB.

$$f_{\theta(r)}(x|A_m) = \frac{e^{\left(|A_m|^2/2\sigma^2\right)\cos(x-\theta(A_m))}}{2\pi I_0\left(|A_m|^2/\sigma^2\right)} \tag{2.8}$$

For any modulation, the overall distribution of a received signal in AWGN channel can be found as a combination of distributions from each modulation symbol [equation (2.9)].

$$f_{\theta(r)}(x) = \sum_{m=1}^{M} \frac{1}{M} \frac{e^{\left(|A_m|^2/2\sigma^2\right)\cos(x-\theta(A_m))}}{2\pi I_0\left(|A_m|^2/\sigma^2\right)} \tag{2.9}$$

With an identical symbol magnitude $|A|$ and the phase for symbol $A_m$ expressed as $\theta(A_m) = 2m\pi/M$, the PDF of a M-PSK modulation can be found as shown in equation (2.10).

$$f_{\theta(r)}(x) = \sum_{m=1}^{M} \frac{1}{M} \frac{e^{\left(|A|^2/2\sigma^2\right)\cos(x-2m\pi/M)}}{2\pi I_0\left(|A|^2/\sigma^2\right)} \tag{2.10}$$

### 2.2.3 Signal Distribution of Signal Magnitude

The signal magnitude has a unique feature of being robust against phase offset, frequency offset or any kind of rotational change to the transmitted signal. In this book, we approximate the PDF of the signal magnitude distribution to a Rice distribution. Given a single modulation symbol $A_m$ and AWGN noise of variance $\sigma^2$, the PDF of the received signal from this single symbol can be found as shown in equation (2.11), which leads to the overall signal magnitude distribution as a sum of Rice distributions from each modulation symbol, equation (2.12).

$$f_{|r|}(x|A_m,\sigma) = \frac{|A_m|}{\sigma^2} e^{-\left(x^2+|A_m|^2\right)/2\sigma^2} I_0\left(\frac{x|A_m|}{\sigma^2}\right) \tag{2.11}$$

$$f_{|r|}(x) = \sum_{m=1}^{M} \frac{1}{M} f_{|r|}(x|A_m,\sigma) = \sum_{m=1}^{M} \frac{1}{M} \frac{|A_m|}{\sigma^2} e^{-\left(x^2+|A_m|^2\right)/2\sigma^2} I_0\left(\frac{x|A_m|}{\sigma^2}\right) \tag{2.12}$$

## 2.3 Signal Models in Fading Channel

Instead of modelling each fading type, we characterize the joint effect of them into three categories: attenuation, phase offset and frequency offset. Depending on the nature of the fading channel, two types of fading scenarios are generally considered for signal phase offset: slow fading and fast fading. Slow fading is normally caused by shadowing (or shadow fading) when the signal is obscured by a large object from a line-of-sight communication (Goldsmith, 2005). As the coherent time of the shadow fading channel is significantly longer than the signal period, the effect of attenuation and phase offset remains constant. Therefore, a constant channel gain $\alpha$ and phase offset $\theta_0$ can be used to model the received signal after slow fading [equation (2.13)].

$$r(t) = \alpha e^{j\theta_o} s(t) + \omega(t) \tag{2.13}$$

The resulting effect can be observed from a simulated 4-QAM signal in slow fading channel with channel gain at 0.5, phase offset at 10° and AWGN noise of 10 dB in Figure 2.4.

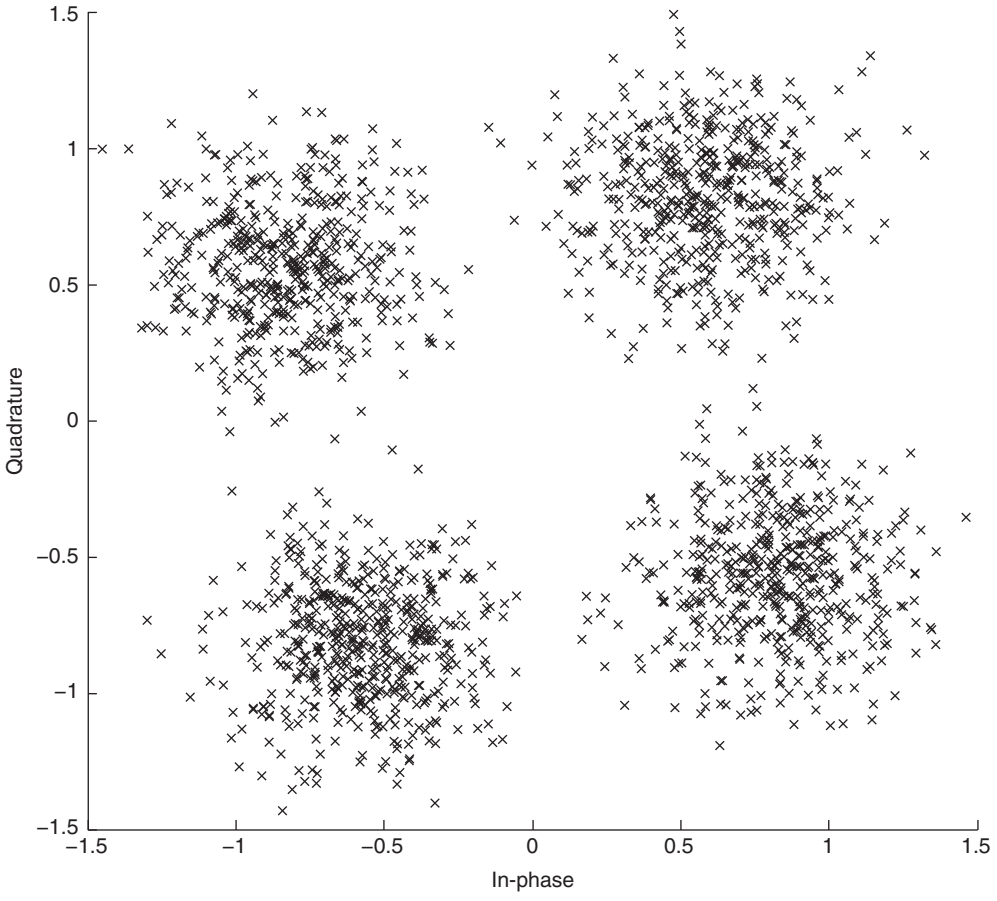



**Figure 2.4** Constellation of 4-QAM signal in slow fading channel.

For the development of likelihood-based classifiers, it is useful to derive the updated signal PDFs in the slow fading channel to compensate the effect of attenuation and phase offset. As the fading channel does not modify the additive noise, only the terms related to transmitted symbols need to be changed. Given that the constant channel gain $\alpha$ and phase offset $\theta_0$, the transmitted symbol $A_m$ is shifted to the new position given by $\alpha e^{-j\theta_o} A_m$. The updated PDF of the complex signal after a slow fading channel can then be found by substituting the shifted symbols into equation (2.5), whence equation (2.14) is derived.

$$f_r(x) = \sum_{m=1}^{M} \frac{1}{M} f_\omega\left(x | \alpha e^{-j\theta_o} A_m, \Sigma\right) = \sum_{m=1}^{M} \frac{1}{M} \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-\frac{\left|x-\alpha e^{-j\theta_0} A_m\right|^2}{2\sqrt{|\Sigma|}}} \tag{2.14}$$

Similarly, the PDFs for signal I-Q segments, phase and magnitude can be derived accordingly as equations (2.15)–(2.17).

$$f_{I(r)}(x) = \sum_{m=1}^{M} \frac{1}{M} f_{I(\omega)}\left(x|\alpha e^{-j\theta_o} A_m, \sigma\right) = \sum_{m=1}^{M} \frac{1}{M} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\left|x-I\left(\alpha e^{-j\theta_o} A_m\right)\right|^2}{2\sigma^2}} \tag{2.15}$$

$$f_{\theta(r)}(x) = \sum_{m=1}^{M} \frac{1}{M} \frac{e^{\left(\left|\alpha e^{-j\theta_o} A_m\right|^2/2\sigma^2\right)\cos\left(x-\theta\left(\alpha e^{-j\theta_o} A_m\right)\right)}}{2\pi I_0\left(\left|\alpha e^{-j\theta_o} A_m\right|^2/\sigma^2\right)} \tag{2.16}$$

$$f_{|r|}(x) = \sum_{m=1}^{M} \frac{1}{M} f_{|r|}\left(x|\alpha e^{-j\theta_o} A_m, \sigma\right) = \sum_{m=1}^{M} \frac{1}{M} \frac{\left|\alpha e^{-j\theta_o} A_m\right|}{\sigma^2} e^{-\left(x^2+\left|\alpha e^{-j\theta_o} A_m\right|^2\right)/2\sigma^2} I_0\left(\frac{x\left|\alpha e^{-j\theta_o} A_m\right|}{\sigma^2}\right) \tag{2.17}$$

It is worth noting that while the PDF of the signal phase is much less sensitive to channel attenuation, it is much more vulnerable to phase offset. Meanwhile, the signal magnitude PDF is not affected by the phase offset at all as the PDF can be rewritten as equation (2.18).

$$f_{|r|}(x) = \sum_{m=1}^{M} \frac{1}{M} f_{|r|}\left(x|\alpha e^{-j\theta_o} A_m, \sigma\right) = \sum_{m=1}^{M} \frac{1}{M} \frac{|\alpha A_m|}{\sigma^2} e^{-\left(x^2+|\alpha A_m|^2\right)/2\sigma^2} I_0\left(\frac{x|\alpha A_m|}{\sigma^2}\right) \tag{2.18}$$

Fast fading, caused by multipath fading where signals are reflected by objects of different properties in the radio channel, imposes a much different effect on the transmitted signal, as the coherent channel time in a fast fading channel is considered small. The effects of attenuation and phase offset vary with time. In this book, we assume that both the attenuation and phase offset are random processes with Gaussian distributions. The attenuation is given by equation (2.19),

$$\alpha(t) \sim \mathcal{N}(\alpha, \sigma_\alpha) \tag{2.19}$$

where $\alpha(t)$ is the channel gain at time $t$, $\alpha$ is the mean attenuation and $\sigma_\alpha^2$ is the variance of the channel gain. The phase offset is given by equation (2.20), where

$$\theta_o(t) \sim \mathcal{N}(\theta_o, \sigma_{\theta_o}) \tag{2.20}$$

$\theta_o(t)$ is the channel gain at time $t$, $\theta_o$ is the mean attenuation and $\sigma_{\theta_o}^2$ is the variance of the channel gain. Both expressions give a combined effect of slow and fast fading. When $\alpha$ and $\theta_o$ are both zero, the fading consists of only fast attenuation and fast phase offset.

When $\sigma_{\alpha}^2$ and $\sigma_{\theta_o}^2$ are both zero, the model reverts back to the case of slow fading. The resulting channel model becomes as shown in equation (2.21).

$$r(t) = \alpha(t)e^{j\theta_o(t)}s(t) + \omega(t) \tag{2.21}$$

Apart from the channel attenuation and phase offset, frequency offset is another important effect in a fading channel that is worth investigating. The shift in frequency of a received signal is mostly caused by moving antennas in mobile communication devices. Given the carrier frequency of a modulated signal as $f$, when the antenna is moving at a speed $v$ the resulting frequency offset caused by Doppler shift can be found as $fv/c$ where c is the speed of travelling light in the channel medium (Gallager, 2008). As we are only interested in the amount of frequency offset, the expression is simplified by denoting the frequency offset set as $f_o$ and the resulting signal model with frequency offset set given by equation (2.22).

$$r(t) = e^{j2\pi tf_o}s(t) + \omega(t) \tag{2.22}$$

The effect of frequency offset is illustrated in Figure 2.5 with a set of 4-QAM signals. Combining the attenuation, phase offset and frequency offset, we can derive a signal model of fading channel of all the previously mentioned effects [equation (2.23)].

$$r(t) = \alpha(t)e^{j(2\pi tf_o + \theta_o(t))}s(t) + \omega(t) \tag{2.23}$$

## 2.4 Signal Models in Non-Gaussian Channel

In this section, we start with Middleton's class A non-Gaussian noise model as a complex but accurate modelling of impulsive noises. The symmetric alpha stable model is suggested as a simplified alternative to Middleton's class A model. In addition, the Gaussian mixture model is also established for analytical convenience in some of the complex modulation-classification algorithms. The subject of non-Gaussian noise in AMC has been studied by Chavali and Da Silva extensively (2011, 2013).

### *2.4.1 Middleton's Class A Model*

Middleton proposed a series of noise models (Middleton, 1999) to approximate the impulsive noises generated by different natural and man-made electromagnetic activities in physical environments. The models have become popular in many fields, including wireless communication, thanks to the canonical nature of the model which is invariant of the noise source, noise waveform and propagation environments. The versatility of the model is enhanced by the model parameters which provide us with

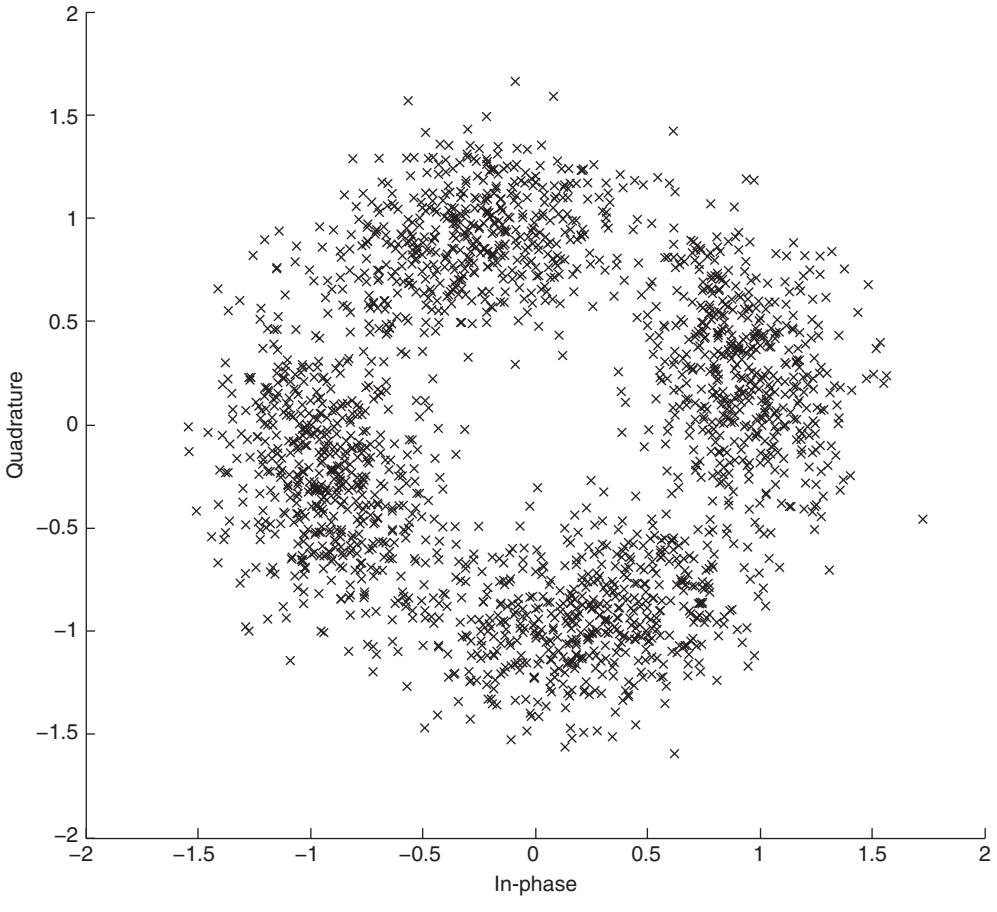



**Figure 2.5** Constellation of 4-QAM signal with frequency offset.

the possibility to specify the source distribution, propagation properties and beam patterns.

The class A model is defined for the non-Gaussian noises with bandwidth narrower than the receiver bandwidth, while the class B model is defined for the non-Gaussian noises with a wider spectrum than the receiver. In the meantime, the class C model provides a combination of the class A and class B models. In this book, the class A model is adopted. According to Middleton (1999), the PDF of the class A noise is derived as shown in equation (2.24),

$$f_{\omega}(x) = e^{-A_A} \sum_{k=0}^{\infty} \frac{A_A^k}{k! \sqrt{4\pi\sigma_{kA}^2}} e^{-\frac{x^2}{4\pi\sigma_{kA}^2}} \tag{2.24}$$

where $A_A$ is the overlap index, which defines the number of noise emissions per second multiplied by the mean duration of the typical emission. The variance of the $k$th emission element is given by equation (2.25),

$$2\sigma_{kA}^2 = \frac{\frac{k}{A_A} + \Gamma_A}{1 + \Gamma_A} \tag{2.25}$$

where $\Gamma_A$ is the Gaussian factor defined by the ratio of the average power of the Gaussian component to the average power of the non-Gaussian components. To approximate the desired impulsive nature in this section, the small overlap index and Gaussian factor are suggested to provide a heavy-tailed distribution for the noise simulation.

### 2.4.2 Symmetric Alpha Stable Model

Middleton's models are derived from details like the structure of the noise waveform and source distribution. In contrast, the Symmetric Alpha Stable (SαS) model assumes the beam patterns of the receiver antenna and noise sources to be non-directional and the noise source to be isotropically distributed in space (Nikias and Shao, 1995). Therefore, SαS has a simpler and more tractable model. The characteristic function of the SαS model is given by equation (2.26),

$$\phi(t) = e^{j\delta t - \gamma|t|^\alpha} \tag{2.26}$$

where $0 < \alpha \leq 2$ is the characteristic exponent, $\delta$ is the location parameter, and $\gamma$ is the scale parameter, also known as the dispersion. The PDF of the SαS can be expressed using the characteristic function as equation (2.27).

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi(t) e^{-jxt} \tag{2.27}$$

### 2.4.3 Gaussian Mixture Model

In the meantime, Vastola proposed to approximate the Middleton's class A model through a mixture of Gaussian noises (Vastola, 1984). The conclusion was drawn that the Gaussian mixture model (GMM) provides a close approximation to Millerton's class A model while being computationally much more efficient. The PDF of the GMM mode is given by equation (2.28),

$$f_\omega(x) = \sum_{k=1}^{K} \frac{\lambda_k}{2\pi\sigma_k^2} e^{-\frac{|x|^2}{2\sigma_k^2}} \tag{2.28}$$

where $0 < \alpha \leq 2$ is the characteristic exponent, $\delta$ is the total number of Gaussian components, $\lambda_k$ is the probability of the noise $\omega$ being chosen from the $k$th component, and $\sigma_k^2$ is the variance of the $k$th component.

As the GMM will be used as the primary model for impulsive noise, therefore we derive the PDFs of received signals in the non-Gaussian channel with a GMM noise model. Assume the GMM uses $K$ components where the probability $\lambda_k$ and variance $\sigma_k^2$ for each component are either known or estimated. The PDF of the complex signal in the non-Gaussian channel can be derived as equation (2.29),

$$f_r(x) = \sum_{m=1}^{M} \frac{1}{M} \sum_{k=1}^{K} f_r(x|A_m, \lambda_k, \sigma_k) = \sum_{m=1}^{M} \frac{1}{M} \sum_{k=1}^{K} \frac{\lambda_k}{2\pi\sigma_k^2} e^{-\frac{|x-A_m|^2}{2\sigma^2}} \tag{2.29}$$

with the corresponding variation for signal I-Q segments, signal phase and signal magnitude given in equations (2.30), (2.31) and (2.32), respectively.

$$f_{I(r)}(x) = \sum_{m=1}^{M} \frac{1}{M} \sum_{k=1}^{K} f_{I(r)}(x|A_m, \lambda_k, \sigma_k) = \sum_{m=1}^{M} \frac{1}{M} \sum_{k=1}^{K} \frac{\lambda_k}{\sigma_k\sqrt{2\pi}} e^{-\frac{|x-I(A_m)|^2}{2\sigma^2}} \tag{2.30}$$

$$f_{\theta(r)}(x) = \sum_{m=1}^{M} \frac{1}{M} \sum_{k=1}^{K} f_{\theta(r)}(x|A_m, \lambda_k, \sigma_k) = \sum_{m=1}^{M} \frac{1}{M} \sum_{k=1}^{K} \frac{\lambda_k e^{\left(|A_m|^2/2\sigma_k^2\right)\cos(x-\theta(A_m))}}{2\pi I_0\left(|A_m|^2/\sigma_k^2\right)} \tag{2.31}$$

$$f_{|r|}(x) = \sum_{m=1}^{M} \frac{1}{M} \sum_{k=1}^{K} f_{|r|}(x|A_m, \lambda_k, \sigma_k) = \sum_{m=1}^{M} \frac{1}{M} \sum_{k=1}^{K} \frac{\lambda_k |A_m|}{\sigma_k^2} e^{-\left(x^2+|A_m|^2\right)/2\sigma_k^2} I_0\left(\frac{x|A_m|}{\sigma_k^2}\right) \tag{2.32}$$

## 2.5 Conclusion

In this chapter, we establish the signal models required for constructing modulation classifiers in the following chapters. The signal models are also needed in the computer-aided simulations for the generation of signals for classification.

Three common communication channels, namely AWGN channel, fading channel, and non-Gaussian channel, are considered. The PDFs of received signals in the AWGN channel are derived for the complex signal, in-phase and quadrature segments, signal phase and signal magnitude in equations (2.5), (2.6), (2.7) and (2.12). The approximation of the PDF of signal phase in equation (2.7) is introduced into equation (2.9) as an alternative for easier theoretical analysis. An example of the approximate for M-PSK modulations is given in equation (2.10).

The fading channel has been divided into two scenarios of slow fading channel and fast fading channel. In the slow fading channel the PDF of received signal-experienced

constant attenuation and phase offset are given in equation (2.14), while its variation for signal I-Q segments, signal phase and signal magnitude are listed in equations (2.15), (2.16) and (2.17). The signal model in fast fading channel is defined in equation (2.21) with attenuation and phase offset as normally distributed random processes. The additional signal model with frequency offset is given in equation (2.22).

Three non-Gaussian models are presented for modelling the impulsive noise. The PDF of Middleton's Class A noise is given in equation (2.24), while the characteristic function of the alpha stable noise is given in equation (2.26). The Gaussian mixture model is suggested to approximate the Middleton's Class A model and the alpha stable model. Its manageable PDF in equation (2.27) makes the derivation of the complex signal, signal I-Q segments, signal phase and signal magnitude distribution PDFs in equations (2.28), (2.29), (2.30) and (2.31) much more intuitive and any further analysis relatively effortless.

## References

Azzouz, E.E. and Nandi, A.K. (1996) *Automatic Modulation Recognition of Communication Signals*, Kluwer, Boston.

Bennett, W.R. (1956) Methods of solving noise problems. *Proceedings of the IRE*, **44**, 609–638.

Chavali, V.G. and Da Silva, C.R.C.M. (2011) Maximum-likelihood classification of digital amplitude-phase modulated signals in flat fading non-Gaussian channels. *IEEE Transactions on Communications*, **59** (8), 2051–2056.

Chavali, V.G. and Da Silva, C.R.C.M. (2013) Classification of digital amplitude-phase modulated signals in time-correlated non-Gaussian channels. *IEEE Transactions on Communications*, **61** (6), 2408–2419.

Gallager, R.G. (2008) *Principles of Digital Communication*, Cambridge University Press, Cambridge.

Goldsmith, A. (2005) *Wireless Communications*, Cambridge University Press, Cambridge.

Huang, C.Y. and Polydoros, A. (1995) Likelihood methods for MPSK modulation classification. *IEEE Transactions on Communications*, **43** (2), 1493–1504.

Middleton, D. (1999) Non-Gaussian noise models in signal processing for telecommunications: new methods and results for class A and class B noise models. *IEEE Transactions on Information Theory*, **45** (4), 1129–1149.

Nikias, C.L. and Shao, M. (1995) *Signal Processing with Alpha-Stable Distributions and Applications*, John Wiley & Sons, Inc., New York.

Shi, Q. and Karasawa, Y. (2012) Automatic modulation identification based on the probability density function of signal phase. *IEEE Transactions on Communications*, **60** (4), 1–5.

Spooner, C.M. (1996) Classification of Co-channel Communication Signals Using Cyclic Cumulants. Conference Record of the Twenty-Ninth Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 30 October 1995, pp. 531–536.

Swami, A. and Sadler, B.M. (2000) Hierarchical digital modulation classification using cumulants. *IEEE Transactions on Communications*, **48** (3), 416–429.

Urriza, P., Rebeiz, E., Pawełczak, P. and Čabrić, D. (2011) Computationally efficient modulation level classification based on probability distribution distance functions. *IEEE Communications Letters*, **15** (5), 476–478.

Vastola, K.S. (1984) Threshold detection in narrow-band non-Gaussian noise. *IEEE Transactions on Communications, C* (**2**), 134–139.

Wang, F. and Wang, X. (2010) Fast and robust modulation classification via Kolmogorov-Smirnov test. *IEEE Transactions on Communications*, **58** (8), 2324–2332.

Wei, W. and Mendel, J.M. (2000) Maximum-likelihood classification for digital amplitude-phase modulations. *IEEE Transactions on Communications*, **48** (2), 189–193.

Zhu, Z., Aslam, M.W. and Nandi, A.K. (2014) Genetic algorithm optimized distribution sampling test for M-QAMmodulation classification. *Signal Processing*, **94**, 264–277.

# 3

# Likelihood-based Classifiers

## 3.1 Introduction

Likelihood-based (LB) modulation classifiers are by far the most popular modulation classification approaches. The interest in LB classifiers is motivated by the optimality of its classification accuracy when perfect channel model and channel parameters are known to the classifiers (Huang and Polydoros, 1995; Sills, 1999; Wei and Mendel, 2000; Hameed, Dobre and Popescu, 2009; Ramezani-Kebrya *et al.*, 2013).

The common approach of an LB modulation classifier consists of two steps. In the first step, the likelihood is evaluated for each modulation hypothesis with observed signal samples. The likelihood functions are derived from the selected signal model and can be modified to fulfil the need of reduced computational complexity or to be applicable in non-cooperative environments. In the second step, the likelihood of different modulation hypothesizes are compared to conclude the classification decision. Earlier methods of decision making are enabled with a ratio test between two hypothesizes. The requirement of a threshold provides another level of optimization which may provide improved classification performance but also requires more tentative effort to select thresholds. The more intuitive approach of decision making would be to find the maximum likelihood among all candidates. It is much easier to implement and does not require carefully designed thresholds.

In reality, much effort has been made to modify the likelihood approach for lower computational complexity and versatility in non-cooperative environments. In this chapter, we will first present the maximum likelihood (ML) classifier. The alternatives of average likelihood ratio test (ALRT), generalized likelihood ratio test (GLRT) and hybrid likelihood ratio test (HLRT) will be discussed. The last section will be dedicated to the complexity reduction of the likelihood-based classifiers.

*Automatic Modulation Classification: Principles, Algorithms and Applications*, First Edition. Zhechen Zhu and Asoke K. Nandi.

## 3.2 Maximum Likelihood Classifiers

Likelihood evaluation is equivalent to the calculation of probabilities of observed signal samples belonging to the models with given parameters. In a maximum likelihood classifier, with perfect channel knowledge, all parameters are known except the signal modulation. Therefore, the classification process can also be perceived as a maximum likelihood estimation of the modulation type where the modulation type is found in a finite set of candidates.

We will focus on deriving the likelihood function in the AWGN channel while modification of the likelihood function in fading channels and non-Gaussian channels will also be mentioned briefly.

### 3.2.1 Likelihood Function in AWGN Channels

Given that the likelihood of the observed signal sample $r[n]$ belonging to the modulation M is equal to the probability of the signal sample $r$ being observed in the AWGN channel modulated with $\mathcal{M}$, then equation (3.1) holds.

$$\mathcal{L}(r[n]|\mathcal{M},\sigma) = p(r[n]|\mathcal{M},\sigma) \tag{3.1}$$

As we recall the complex form PDF of received signal in AWGN channel, the likelihood function can be found as shown in equation (3.2).

$$\mathcal{L}(r[n]|\mathcal{M},\sigma) = \sum_{m=1}^{M} \frac{1}{M} \frac{1}{2\pi\sigma^2} e^{-\frac{|r[n]-A_m|^2}{2\sigma^2}} \tag{3.2}$$

Without knowing which modulation symbol the signal sample $r[n]$ belongs to, the likelihood is calculated using the average of the likelihood value between the observed signal sample and each modulation symbol $A_m$. The joint likelihood given multiple observed samples is calculated with the multiplication of all likelihoods of individual samples, as given in equation (3.3).

$$\mathcal{L}(r|\mathcal{M},\sigma) = \prod_{n=1}^{N} \sum_{m=1}^{M} \frac{1}{M} \frac{1}{2\pi\sigma^2} e^{-\frac{|r[n]-A_m|^2}{2\sigma^2}} \tag{3.3}$$

For analytical convenience in many cases, the natural logarithm of the likelihood $\mathcal{L}$ is used as the likelihood value to be compared in a maximum likelihood classifier [equation (3.4)].

$$\log\mathcal{L}(r|\mathcal{M},\sigma) = \log\left(\prod_{n=1}^{N} \sum_{m=1}^{M} \frac{1}{M} \frac{1}{2\pi\sigma^2} e^{-\frac{|r[n]-A_m|^2}{2\sigma^2}}\right) = \sum_{n=1}^{N} \log\left(\sum_{m=1}^{M} \frac{1}{M} \frac{1}{2\pi\sigma^2} e^{-\frac{|r[n]-A_m|^2}{2\sigma^2}}\right) \tag{3.4}$$

The likelihood, in the meantime, can be derived from probabilities of different aspects of sampled signals. As we have derived the PDF for in-phase segments of the received signal in AWGN channel, the corresponding likelihood function of the in-phase segments of a signal can be found as given in equation (3.5).

$$\mathcal{L}_{I(r)}(r|\mathcal{M},\sigma) = \prod_{n=1}^{N}\sum_{m=1}^{M}\frac{1}{M}\frac{1}{\sigma\sqrt{\pi}}e^{-\frac{|I(r[n])-I(A_m)|^2}{\sigma^2}} \tag{3.5}$$

The signal phase is another natural subject of study when applying maximum likelihood classification on M-PSK modulations while M-QAM modulations can also be classified (Shi and Karasawa, 2011). The advantage of using the signal phase solely for likelihood evaluation is highlighted by its robustness in channels where amplitude distortions are exhibited. However, the method is overshadowed by its vulnerability against phase and frequency offsets. Using the PDF of the signal phase in AWGN channel from equation (2.7), the phase likelihood function can be derived as given in equation (3.6).

$$\begin{aligned}\mathcal{L}_{\theta(r)}(r[n]|A_m) = {} & \frac{e^{-\frac{|A_m|^2}{2\sigma^2}}}{2\pi} + \frac{|A_m|\cos\left(\theta(r[n])-\theta(A_m)\right)}{2\sigma\sqrt{2\pi}} \\ & \left[1+erf\left(\frac{|A_m|\cos\left(\theta(r[n]-\theta(A_m))\right)}{\sqrt{2}\sigma}\right)\right]e^{-\frac{|A_m|^2}{2\sigma^2}\sin^2(\theta(r[n]-\theta(A_m)))}\end{aligned} \tag{3.6}$$

Owing to the complex form of the likelihood function, different alternative PDFs have been proposed to simplify the phase likelihood function. The von Mises distribution PDF is a much lighter alternative to the PDF given in equation (2.7). The phase likelihood function based on the von Mises PDF is given by equation (3.7).

$$\mathcal{L}_{\theta(r)}(r|\mathcal{M},\sigma) = \prod_{n=1}^{N}\sum_{m=1}^{M}\frac{1}{M}\frac{e^{\left(|A_m|^2/2\sigma^2\right)\cos(\theta(r[n])-\theta(A_m))}}{2\pi I_0\left(|A_m|^2/\sigma^2\right)} \tag{3.7}$$

As mentioned in Chapter 2, the von Mises distribution requires a certain SNR level to be valid. At low SNR level the approximation will deviate from the accurate PDF and a systematic error in likelihood evaluation will lead to inaccurate classification. For this reason, Shi and Karasawa (2012) proposed an approximation to the accurate PDF in equation (2.7) using Gaussian Legendre quadrature rules to replace the error function in equation (3.6) with a finite-range integral (Abramowitz and Stegun, 1964), as shown in equation (3.8),

$$erf\left(\frac{|A_m|\cos(x)}{\sqrt{2}\sigma}\right) \approx \frac{|A_m|}{\sqrt{2\pi}\sigma}\cos(x)\sum_{l=1}^{L}w_l e^{-\frac{A_m^2}{8\sigma^2}(x_l+1)^2\cos^2(x)} \tag{3.8}$$

where $L$ is the number of points, $\{w_l\}_{l=1}^{L}$ are the weights, and $\{x_l\}_{l=1}^{L}$ are the abscissas of the semi-infinite Gauss–Hermite quadrature rule (Steen, Byrne and Gelbard, 1969). Therefore, the corresponding phase likelihood function using the approximation can be derived by substituting equation (3.8) into equation (3.6) to give equation (3.9).

$$\mathcal{L}_{\theta(r)}(r[n]|A_m) = \frac{e^{-\frac{|A_m|^2}{2\sigma^2}}}{2\pi} + \frac{|A_m|\cos(\theta(r[n]) - \theta(A_m))}{2\sigma\sqrt{2\pi}} \left[1 + \frac{|A_m|}{\sqrt{2\pi}\sigma}\cos(\theta(r[n]) - \theta(A_m))\sum_{l=1}^{L} w_l e^{-\frac{A_m^2}{8\sigma^2}(x_l+1)^2\cos^2(x)}\right] e^{-\frac{|A_m|^2}{2\sigma^2}\sin^2(\theta(r[n])-\theta(A_m))} \tag{3.9}$$

The magnitude likelihood function can be implemented for the classification of M-QAM modulations and M-PAM modulations. The unique advantage of the magnitude likelihood appears when a phase or frequency offset is observed in the transmission channel. As the PDF of signal magnitude in AWGN or fading channel only consists of the magnitude of the transmitted symbols, any rotational shift would not alter the resulting probability evaluation. Using the PDF of signal magnitude in AWGN channel given in equation (2.12), the magnitude likelihood can be found as shown in equation (3.10).

$$\mathcal{L}_{|r|}(r) = \prod_{n=1}^{N}\sum_{m=1}^{M}\frac{1}{M}\frac{|A_m|}{\sigma^2}e^{-\left(r^2+|A_m|^2\right)/2\sigma^2} I_0\left(\frac{r|A_m|}{\sigma^2}\right) \tag{3.10}$$

### 3.2.2 Likelihood Function in Fading Channels

Though the likelihood functions in AWGN channels are the primary subject of research, it would be interesting to derive their variant in fading channels. In this section, channel attenuation and phase shift in both slow and fast fading scenarios are introduced to the ML classifier. The information is used to modify the likelihood function derived for AWGN channels to compensate the added effect from fading channels.

In slow fading channel where amplitude attenuation and phase shift are both considered constant, the modification to the likelihood function is restricted to the transmitted signal symbols. Assume the known or estimated attenuation is $\alpha$ and $\theta_o$, the symbol $A_m$ would endure a shift and result in a new position for $\alpha e^{-j\theta_o}A_m$. Substitute the new symbol positions into the likelihood function in equation (3.3), and the likelihood function in slow fading channel with constant amplitude attenuation and phase shift can be found as given in equation (3.11).

$$\mathcal{L}(r|\mathcal{M},\sigma)=\prod_{n=1}^{N}\sum_{m=1}^{M}\frac{1}{M}\frac{1}{2\pi\sigma^2}e^{-\frac{\left|r[n]-\alpha e^{-j\theta_0}A_m\right|^2}{2\sigma^2}} \tag{3.11}$$

The likelihood function in fast fading channel is negated here as the collection of parameters needed for the accurate likelihood evaluation is unlikely to be all known to the classifier. The estimation of all the parameters also seems ambitious, which renders the exact likelihood function for the fast fading channel impractical.

### 3.2.3 Likelihood Function in Non-Gaussian Noise Channels

As impulsive noises are common in communication channels, it is important to derive the likelihood function for signals in non-Gaussian noise channels with impulsive noises. Middleton's Class A model and the SαS model, though being relatively accurate, have many parameters in their PDFs and characteristic functions. Considering the extra mismatch which may be introduced in the process of estimation these parameters, the GMM model is adopted as the bases for the likelihood function in non-Gaussian noise channels. From equation (2.28), assuming the probability $\lambda_k$ and variance $\sigma_k^2$ for total number of $K$ Gaussian components are known or estimated, the likelihood function can be found as given in equation (3.12).

$$\mathcal{L}(r)=\prod_{n=1}^{N}\sum_{m=1}^{M}\frac{1}{M}\sum_{k=1}^{K}\frac{\lambda_k}{2\pi\sigma_k^2}e^{-\frac{|r[n]-A_m|^2}{2\sigma^2}} \tag{3.12}$$

### 3.2.4 Maximum Likelihood Classification Decision Making

Having established the likelihood functions in different channel scenarios from different signal segments, the decision making in an ML classifier becomes rather straightforward. Assuming a pool with finite number $I$ modulation candidates $\mathfrak{M}$, among which hypothesis $\mathcal{H}_{\mathcal{M}(i)}$ of each modulation $\mathcal{M}(i)$ is evaluated using estimated channel parameters $\hat{\boldsymbol{\Theta}}_{\mathcal{M}(i)}$ and a suitable likelihood function to obtain its likelihood evaluation $\mathcal{L}(r\,|\,\mathcal{H}_{\mathcal{M}(i)})$. With the all the likelihood values collected the decision is made simply by finding the hypothesis with the highest likelihood [equation (3.13)].

$$\hat{\mathcal{M}}=\underset{\mathcal{M}(i)\in\mathfrak{M}}{\arg\max}\,\mathcal{L}\left(r\,|\,\mathcal{H}_{\mathcal{M}(i)}\right) \tag{3.13}$$

The entire process of the ML classification is illustrated in Figure 3.1.

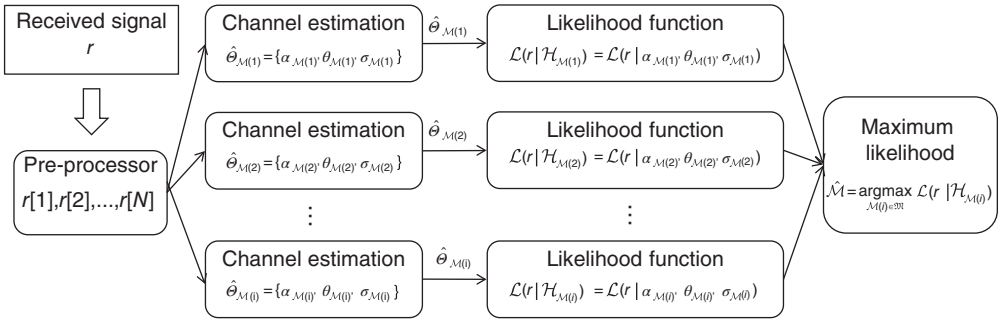



**Figure 3.1** Maximum likelihood classifier in fading channel with AWGN noise.

## 3.3 Likelihood Ratio Test for Unknown Channel Parameters

### 3.3.1 Average Likelihood Ratio Test

The issue of an unknown parameter in an ML classifier is pivotal as the likelihood function is unable to handle any missing parameter. The average likelihood ratio test (ALRT) is one way to overcome such limitation of an ML classifier. Polydoros and Kim were the first to apply ALRT to modulation classification (Polydoros and Kim, 1990), which was later adopted by Huang and Polydoros (1995), Beidas and Weber (1995), Sills (1999), Hong and Ho (2000). Different from the ML likelihood function, the ALRT likelihood function replaces unknown parameters with the integral of all their possible values and their corresponding probabilities. An example of ALRT likelihood function with unknown constant carrier phase offset is given by equation (3.14),

$$\mathcal{L}_{ALRT}(r) = \int_{\theta_o} \mathcal{L}(r|\theta_o) f(\theta_o|\mathcal{H}) d\theta = \int_{\theta_o} \prod_{n=1}^{N} \sum_{m=1}^{M} \frac{1}{M} \frac{1}{2\pi\sigma^2} e^{-\frac{\left|r[n]-\alpha e^{-j\theta_o} A_m\right|^2}{2\sigma^2}} f(\theta_o|\mathcal{H}) d\theta_o \qquad (3.14)$$

where $\mathcal{L}_{ALRT}(r)$ is the updated ALRT likelihood, $\mathcal{L}(r \mid \theta_o)$ is the likelihood given phase offset of $\theta_o$, and $f(\theta_o \mid \mathcal{H})$ is the probability of constant phase offset $\theta_o$ under modulation hypothesis $\mathcal{H}$. Other channel parameters such as noise variance $\sigma^2$ and channel gain $\alpha$ can also be treated individually as unknown parameters with their PDF or as a group of unknown parameters with their joint probability, as shown in equation (3.15),

$$\mathcal{L}_{ALRT}(r) = \int_{\Theta} \mathcal{L}(r|\Theta) f(\Theta|\mathcal{H}) d\theta = \int_{\Theta} \prod_{n=1}^{N} \sum_{m=1}^{M} \frac{1}{M} \frac{1}{2\pi\sigma^2} e^{-\frac{\left|r[n]-\alpha e^{-j\theta_o} A_m\right|^2}{2\sigma^2}} f(\alpha,\sigma,\theta_o|\mathcal{H}) d\Theta \qquad (3.15)$$

where $\Theta$ is the collection of unknown parameters. While $\mathcal{L}(r \mid \theta_o)$ is known to the classifier, $f(\Theta \mid \mathcal{H})$ depends on the definition of prior probability of unknown parameters.

The common assumption of prior PDFs of different parameters is shown in equations (3.16)–(3.18),

$$f(\alpha|\mathcal{H}) \sim \mathcal{N}(\alpha|\mu_\alpha, \sigma_\alpha) \tag{3.16}$$

$$f(\sigma^2|\mathcal{H}) \sim \text{Gamma}(\sigma^2|a_\sigma, b_\sigma) \tag{3.17}$$

$$f(\theta_o|\mathcal{H}) \sim \mathcal{N}(\theta_o|\mu_{\theta_o}, \sigma_{\theta_o}) \tag{3.18}$$

where channel gain $\alpha$ is given a normal distribution with mean $\mu_\alpha$, variance $\sigma_\alpha^2$, noise variance is given a Gamma distribution with shape parameter $a_\sigma$ and scale parameter $b_\sigma$, and phase offset is given a normal distribution with mean $\mu_{\theta_o}$ and variance $\sigma_{\theta_o}^2$. All the additional parameters associated with PDF of channel parameters are often called hyperparameters. The estimation of hyperparameters is not discussed in this book. Suitable schemes have been proposed by Roberts and Penny using a variational Bayes estimator (Roberts and Penny, 2002).

The likelihood ratio test required for the classification decision making is conducted with the assistance of a threshold $\gamma_A$. The actual likelihood ratio is calculated as given in equation (3.19), where the classification result is given using the conditional equation (3.20).

$$\Lambda_A(i,j) = \frac{\int_{\Theta} \mathcal{L}(r|\theta) f(\theta|\mathcal{H}_i) d\theta}{\int_{\Theta} \mathcal{L}(r|\theta) f(\theta|\mathcal{H}_j) d\theta} \tag{3.19}$$

$$\hat{\mathcal{M}} = \begin{cases} \mathcal{M}_i & \text{if } \Lambda_A(i,j) \geq \gamma_A \\ \mathcal{M}_j & \text{if } \Lambda_A(i,j) < \gamma_A \end{cases} \tag{3.20}$$

An easy assignment of the ratio test threshold is to define all thresholds to be one. The decision making becomes a simple process of comparing the average likelihood of two hypotheses [equations (3.21)].

$$\hat{\mathcal{M}} = \begin{cases} \mathcal{M}_i & \text{if } \mathcal{L}_{ALRT}(r|\mathcal{H}_i) \geq \mathcal{L}_{ALRT}(r|\mathcal{H}_j) \\ \mathcal{M}_j & \text{if } \mathcal{L}_{ALRT}(r|\mathcal{H}_i) < \mathcal{L}_{ALRT}(r|\mathcal{H}_j) \end{cases} \tag{3.21}$$

Using the same assignment, the maximum likelihood decision making can also be applied using equation (3.13) with the likelihood function with average likelihood.

### 3.3.2 Generalized Likelihood Ratio Test

It is not difficult to see that the ALRT likelihood function has a much more complex form when unknown parameters are introduced. The requirement of underlining models for unknown parameters confirms that successful classification depends on

the accuracy of the models. Consequently, if an accurate model is not available, the method becomes suboptimal and only an approximation to the optimal ALRT classifier. The additional requirement of estimation hyperparameters adds another level of complexity and inaccuracy to the overall performance of the ALRT classifier. It is also the fact that the likelihood function is more complex with added integration operations.

For the above reasons, Panagiotou, Anastasopoulos and Polydoros proposed the generalized likelihood ratio test (GLRT) as an alternative (Panagiotou, Anastasopoulos and Polydoros, 2000). The GLRT in essence is a combination of a maximum likelihood estimator and a maximum likelihood classifier. The likelihood function, unlike the ALRT, replaces the integration of unknown parameters with a maximization of the likelihood over a possible range for the unknown parameters. The likelihood function of the GLRT method is given by equation (3.22).

$$\mathcal{L}_{GLRT}(r) = \max_{\Theta} \mathcal{L}(r|\alpha,\sigma,\theta_o) = \max_{\Theta} \prod_{n=1}^{N} \sum_{m=1}^{M} \frac{1}{M} \frac{1}{2\pi\sigma^2} e^{-\frac{\left|r[n]-\alpha e^{-j\theta_0} A_m\right|^2}{2\sigma^2}} \tag{3.22}$$

When multiple unknown channel parameters are presented, the maximum become a process over multiple parameters. There is a favourable order of maximization when channel gain, noise variance and phase offset are all unknown. As recommended in our previous research (Zhu, Nandi and Aslam, 2013), it is easier to obtain an unbiased ML estimation of the phase offset before channel gain and noise variance. Noise variance is normally best to be estimated when all the rest of channel parameters are accurately estimated.

Panagiotou *et al.* (2000) also included another maximization step to eliminate the effect of averaging the likelihood of the signal samples belonging to different modulation symbols, as shown in equation (3.23).

$$\mathcal{L}_{GLRT}(r) = \max_{\Theta} \mathcal{L}(r|\alpha,\sigma,\theta_o) = \max_{\Theta} \prod_{n=1}^{N} \max_{A_m \in A} \frac{1}{M} \frac{1}{2\pi\sigma^2} e^{-\frac{\left|r[n]-\alpha e^{-j\theta_0} A_m\right|^2}{2\sigma^2}} \tag{3.23}$$

The complexity is notably further reduced. However, the classifier based on the modified GLRT likelihood function now becomes biased in both low SNR and high SNR scenarios. Assume the modified GLRT likelihood function is used to classify among 4-QAM and 16-QAM signals. At low SNR, when signals are well spread, a 4-QAM modulated signal is always more likely to produce a higher likelihood if using 16-QAM as hypothesis, because the 16-QAM has more symbols and they are more densely populated under the assumption of unit power. At high SNR, when signals are tight around the transmitted symbol, the maximization of the likelihood through channel gain is likely to scale the 16-QAM alphabet such that four central symbols in the alphabet will be overlapping with the alphabet of the 4-QAM modulation.

Such a phenomenon observed in nested modulations produces an equal likelihood between low-order modulations and high-order modulations when low-order modulations are being classified. Therefore, the method is clearly biased for high-order modulations in most scenarios.

The actual likelihood ratio is calculated as follows [equation (3.24)],

$$\Lambda_G(i,j) = \frac{\max_{\boldsymbol{\Theta}_i, Z_i} \mathcal{L}(r|\boldsymbol{\Theta}_i, Z_i)}{\max_{\boldsymbol{\Theta}_j, Z_j} \mathcal{L}(r|\boldsymbol{\Theta}_j, Z_j)} \tag{3.24}$$

where $Z_i$ and $Z_j$ define the membership of each observed sample with respect to modulation symbols.

### 3.3.3 Hybrid Likelihood Ratio Test

While the GLRT likelihood function provides alternative to ALRT, the fact that it is a biased classifier, as discussed in the previous section, makes it unsuitable for modulation with nested modulations (e.g. QPSK, 8-PSK; 16-QAM, 64-QAM). For this reason, Panagiotou *et al.* (2000) proposed another likelihood ratio test named hybrid likelihood ratio test (HLRT). In the original publication, the HLRT was suggested as an LB classifier for unknown carrier phase offset. The likelihood in HLRT is calculated by averaging over the transmitted symbols and then maximizing the resulting likelihood function (LF) with respect to the carrier phase. The likelihood function is thus derived as shown in equation (3.25).

$$\mathcal{L}_{HLRT}(r) = \max_{\theta_o \in [0, 2\pi]} \mathcal{L}(r|\theta_o) = \max_{\theta_o \in [0, 2\pi]} \prod_{n=1}^{N} \sum_{m=1}^{M} \frac{1}{M} \frac{1}{2\pi\sigma^2} e^{-\frac{\left|r[n] - \alpha e^{-j\theta_o} A_m\right|^2}{2\sigma^2}} \tag{3.25}$$

It is clear that the HLRT LF calculates the likelihood of each signal sample belonging to each alphabet symbol. Therefore, the case where a nested constellation creates biased classification does not existence. In addition, the maximization process replaces the integral of the unknown parameters and their PDFs for much lower analytical and computational complexity.

The decision making of the HLRT approach follows the same rules as the ALRT and GLRT approach where a threshold $\gamma_H$ is required for the ratio test to optimize the classification accuracy. The ratio test becomes the maximum likelihood classifier when the threshold is set to one.

The actual likelihood ratio is calculated as follows [equation (3.26)].

$$\Lambda_H(i,j) = \frac{\max_{\boldsymbol{\Theta}_i} \mathcal{L}(r|\boldsymbol{\Theta}_i)}{\max_{\boldsymbol{\Theta}_j} \mathcal{L}(r|\boldsymbol{\Theta}_j)} \tag{3.26}$$

## 3.4 Complexity Reduction

It is known that the ML classifier has the drawback of high computational complexity. This is mostly due to the need for calculation of natural logarithms in the likelihood function and the increased demand for additional signal samples. Many researchers have recognized the challenge and provided different approaches to reduce the complexity of the ML classifier.

### 3.4.1 Discrete Likelihood Ratio Test and Lookup Table

Xu, Su and Zhou proposed a fast likelihood function through offline computation (Xu, Su and Zhou, 2011). As pointed out in their work, most of the complexity in an ALRT classifier is contributed by the massive integration, multiplication and exponent arithmetic. Aiming to not degrade the performance of the ALRT classifier, they developed the discrete likelihood ratio test (DLRT) with the addition of a lookup table (LUT). The first step of the classification framework is to build a storage table containing the likelihood of each modulation quantized to a finite set. The quantization of the storage table is done by dividing the continuous complex plane into a $P \times Q$ grid with a uniform partition strategy. The other dimension of quantization is done for the noise variance $\sigma^2$ with $U$ intervals. When a signal sample $r[n]$ is sampled, the complex sample is first mapped to its in-phase component $x_p$ and quadrature component $y_q$ indexed by $p$ and $q$. The log-likelihood of this sampling belonging to the hypothesis modulation is found from a cell $T_i(p, q, u)$ from the LUT for modulation $i$. The table cell itself is calculated using the following equation (3.27),

$$T_i(p,q,u) = \int_{x_p}^{x_{p+1}} \int_{y_p}^{y_{p+1}} \ln\left\{ \int_0^{2\pi} \frac{1}{2\pi} \frac{1}{M_i} \sum_{m=1}^{M_i} e^{-\frac{\left|x+jy-e^{-j\theta_o}A_m\right|^2}{2\sigma_u^2}} d\theta_o \right\} dxdy \tag{3.27}$$

where $M_i$ is the size of the alphabet set of modulation $i$, and $\theta_o$ is the unknown carrier phase. The subsequent steps of the DLRT and LUT classifier are the same as those in an ALRT classifier.

The complexity and accuracy analysis provided suggests that the classification performance is largely associated with the level of quantization. With higher levels of quantization resolution, the error introduced by mismatching between received signal sample and mapped signal sample along with estimated noise variance and mapped noise variance can be minimized. The price for the increased quantization resolution is solely inflicted as the demand for much larger memory to store the recalculated storage tables.

### 3.4.2 Minimum Distance Likelihood Function

Wong and Nandi proposed to remove the exponential arithmetic in ML likelihood function with a simple Euclidean distance between a received signal sample and its nearest modulation symbol (Wong and Nandi, 2008). The minimum distance likelihood function (MDLF) is easily given by equation (3.28).

$$\mathcal{D}_{MDLF}(r) = \sum_{n=1}^{N} \min_{A_m \in A} \left| r[n] - \alpha e^{-j\theta_o} A_m \right| \tag{3.28}$$

Unlike the LB methods, the decision making is based on finding the modulation hypothesis with the minimum distance.

The obvious advantage of the minimum distance classifier is its low complexity. However, the main drawback is likewise easy to spot. Given a group of well spread signal samples, it is always easier to find a smaller distance when there is a greater number of densely populated centroids. Therefore higher-order modulations are always favoured by the classifier.

### 3.4.3 Non-Parametric Likelihood Function

Zhu and Nandi recently proposed a non-parametric likelihood function (NPLF) (Zhu and Nandi, 2014). The initial goal was to obtain the likelihood evaluation without a specific signal distribution as well as without the knowledge of the noise variance. In addition, the NPLF also achieves a very low computation complexity. The NPLF function is defined as shown in equation (3.29),

$$\mathcal{L}_{NPLF}(r) = \sum_{n=1}^{N} \sum_{m=1}^{M} \mathbb{I}\left\{ \left| r[n] - \alpha e^{-j\theta_o} A_m \right| < \mathcal{R}_{\mathcal{M}(i)} \right\} \tag{3.29}$$

which evaluates the cumulative distribution in a set of regions defined by the modulation symbol and a radius factor $\mathcal{R}_{\mathcal{M}(i)}$. The calculation of the likelihood is enabled with an indicator function $\mathbb{I}(\cdot)$ which outputs a value of 1 if the input is true and 0 if the input is false. More details of the NPLF function are given in Chapter 7, where it is employed for blind modulation classification.

## 3.5 Conclusion

In this chapter, we have examined different likelihood-based modulation classifiers. The maximum likelihood method is first presented as the optimum classifier with the requirement of known channel state information. Its likelihood function is given

in equation (3.3) and the decision making is established in equation (3.13). The assumption of perfect channel knowledge is relaxed by the subsequent ALRT, GLRT and HLRT classifiers. They all consider one or two channel parameters as being unknown. Among these classifiers, the likelihood function of ALRT classifier [equation (3.14)] is the most complex one where multiple integral and exponential operations are need. The GLRT likelihood function has a much simpler form but a biased classification performance. The HLRT presents an option where the complexity and classification performance has a better balance. Other approaches to reduce the complexity of a maximum likelihood classifier are presented in the last part of the chapter.

## References

Abramowitz, M. and Stegun, I. (1964) *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*. Dover Publications, Mineola, New York.

Beidas, B.F. and Weber, C. (1995) Higher-order correlation-based approach to modulation classification of digitally frequency-modulated signals. *IEEE Journal on Selected Areas in Communications*, **13** (1), 89–101.

Hameed, F., Dobre, O.A. and Popescu, D. (2009) On the likelihood-based approach to modulation classification. *IEEE Transactions on Wireless Communications*, **8** (12), 5884–5892.

Hong, L. and Ho, K.C. (2000) BPSK and QPSK Modulation Classification with Unknown Signal Level. Military Communications Conference, Los Angeles, CA, USA, 22 October 2000, pp. 976–980.

Huang, C.Y. and Polydoros, A. (1995) Likelihood methods for MPSK modulation classification. *IEEE Transactions on Communications*, **43** (2), 1493–1504.

Panagiotou, P., Anastasopoulos, A. and Polydoros, A. (2000) Likelihood Ratio Tests for Modulation Classification. Military Communications Conference, Los Angeles, CA, USA, 22 October 2000, pp. 670–674.

Polydoros, A. and Kim, K. (1990) On the detection and classification of quadrature digital modulations in broad-band noise. *IEEE Transactions on Communications*, **38** (8), 1199–1211.

Ramezani-Kebrya, A., Kim, I.-M., Kim, D.I. *et al.* (2013) Likelihood-based modulation classification for multiple-antenna receiver. *IEEE Transactions on Communications*, **61** (9), 3816–3829.

Roberts, S.J. and Penny, W.D. (2002) Variational bayes for generalized autoregressive models. *IEEE Transactions on Signal Processing*, **50** (9), 2245–2257.

Shi, Q. and Karasawa, Y. (2011) Noncoherent maximum likelihood classification of quadrature amplitude modulation constellations: simplification, analysis, and extension. *IEEE Transactions on Wireless Communications*, **10** (4), 1312–1322.

Shi, Q. and Karasawa, Y. (2012) Automatic modulation identification based on the probability density function of signal phase. *IEEE Transactions on Communications*, **60** (4), 1–5.

Sills, J.A. (1999) Maximum-likelihood Modulation Classification for PSK/QAM. Military Communications Conference, Atlantic City, NJ, 31 October 1999, pp. 217–220.

Steen, N.M., Byrne, G.D. and Gelbard, E.M. (1969) Gaussian quadratures for the integrals. *Mathematics of Computation*, **23** (107), 661–671.

Wei, W. and Mendel, J.M. (2000) Maximum-likelihood classification for digital amplitude-phase modulations. *IEEE Transactions on Communications*, **48** (2), 189–193.

Wong, M.L.D. and Nandi, A.K. (2008) Semi-blind algorithms for automatic classification of digital modulation schemes. *Digital Signal Processing*, **18** (2), 209–227.

Xu, J.L., Su, W. and Zhou, M. (2011) Likelihood-ratio approaches to automatic modulation classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, **41** (4), 455–469.

Zhu, Z. and Nandi, A.K. (2014) Blind Digital Modulation Classification using Minimum Distance Centroid Estimator and Non-parametric Likelihood Function. *IEEE Transactions on Wireless Communications*, doi: 10.1109/TWC.2014.2320724, 1–12.

Zhu, Z., Nandi, A.K. and Aslam, M.W. (2013) Approximate Centroid Estimation with Constellation Grid Segmentation for Blind M-QAMClassification. Military Communications Conference, San Diego, CA, USA, 18 November 2013, pp. 46–51.

# 4

# Distribution Test-based Classifier

## 4.1 Introduction

When the observed signal is of sufficient length, the empirical distribution of the modulated signal becomes an interesting subject to study for modulation classification. In Chapter 2, the signal distributions in various channels are given. It is clear that the signal distributions are mostly determined by two factors, namely modulation symbol mapping and channel parameters. Assuming that the channel parameters are pre-estimated and available, the only variable in the signal distribution becomes the symbol mapping, which is directly associated with the modulation scheme. In Figure 4.1, signal cumulative distributions of 4-QAM, 16-QAM and 64-QAM are given in the same AWGN channel.

By reconstructing the signal distribution using the empirical distribution, the observed signals can be analyzed through their signal distributions. If the theoretical distribution of different modulation candidates is available, there exists one which best matches the underlying distribution of the signal to be classified. The evaluation of equality between difference distributions is also known as goodness of fit (GoF), which indicates how the sampled data fit the reference distribution. Ultimately, the classification is completed by finding the hypothesized signal distribution that has the best goodness of fit.

There exist many different distribution tests which have been designed to evaluate the goodness of fit. Among them, we have selected three state-of-the-art distribution tests that have been adopted for modulation classification, and one customized distribution test created by the authors.

*Automatic Modulation Classification: Principles, Algorithms and Applications*, First Edition. Zhechen Zhu and Asoke K. Nandi.

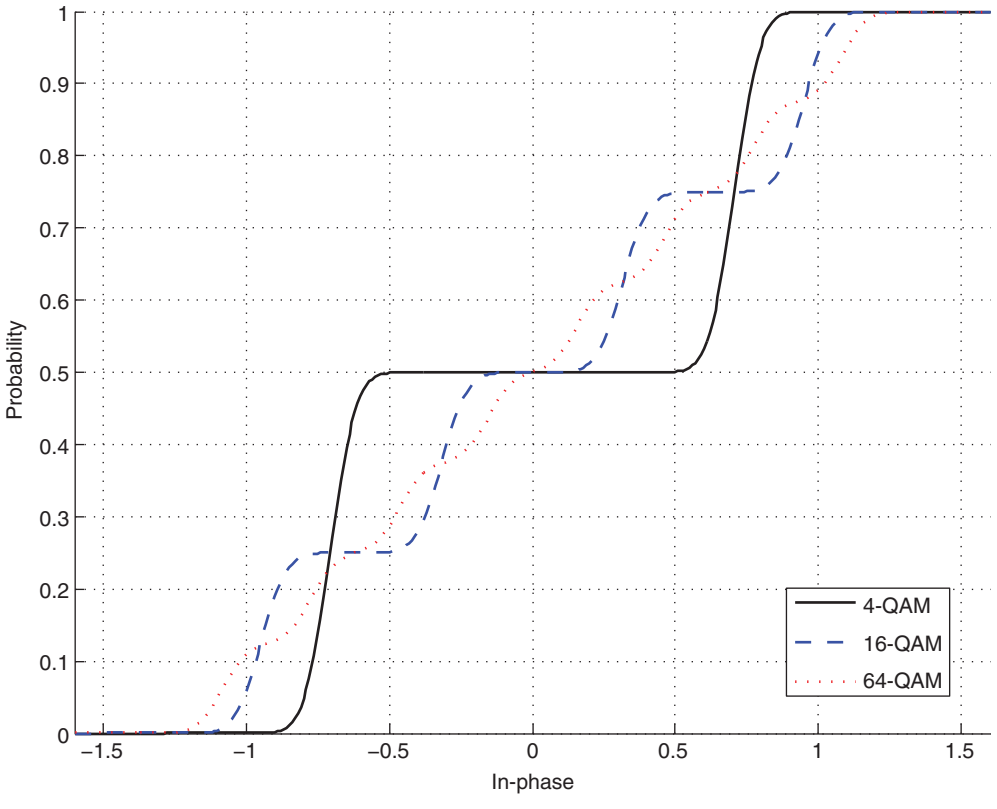



**Figure 4.1** Cumulative distribution probability of 4-QAM, 16-QAM and 64-QAM modulation signals in AWGN channel.

## 4.2 Kolmogorov–Smirnov Test Classifier

The Kolmogorov-Smirnov test (KS test) is a goodness-of-fit test which evaluates the equality of two probability distributions (Conover, 1980). The probability distributions can be either sampled empirical cumulative distribution functions (ECDF) or theoretical cumulative distribution functions (CDF). Massey first introduced the KS test (Massey, 1951) building on theories developed by Kolmogorov (1933) and Smirnov (1939). The KS test has since been applied in many signal processing problems.

Wang and Wang (2010) first adopted the KS test for modulation classification highlighting its low complexity against likelihood-based classifiers (Wei and Mendel, 2000) and high robustness versus cumulant-based classifiers (Swami and Sadler, 2000). Urriza *et al.* modified F. Wang and X. Wang's method for improved computational efficiency (Urriza *et al.*, 2011).

In this section, we first explain the basic theories of the KS test. The implementation of the KS test for modulation classification is presented subsequently.

### 4.2.1 The KS Test for Goodness of Fit

The KS test can be applied in two scenarios which are often referred to as the one-sample test and the two-sample test. In the one-sample test, a set of observed independent variables $x_1$, $x_2$, …, $x_n$ with an underlying cumulative distribution $F_1(x)$ and a hypothesized cumulative distribution function $F_0(x)$ is considered. The null hypothesis of the KS test is shown in equation (4.1),

$$\mathcal{H}_0: \quad F_1 = F_0 \tag{4.1}$$

where $F_1(x)$ is replace by the empirical cumulative distribution function $\hat{F}_1(x)$ from the observed data, as defined in equation (4.2).

$$\hat{F}_1(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(x_i \le x) \tag{4.2}$$

Given the definition of the goodness of fit, statistics is used to find the maximum difference between the underlying cumulative distribution function and the hypothesized cumulative distribution function, as shown in equation (4.3), where

$$D = \sup_{-\infty < x < \infty} |F_1(x) - F_0(x)| \tag{4.3}$$

the numerical calculation of the statistics is replaced by the maximum difference between the empirical cumulative distribution and the hypothesized cumulative distribution function, as given in equation (4.4).

$$D = \max_{1 \le i \le n} \left|\hat{F}_1(x_i) - F_0(x_i)\right| \tag{4.4}$$

According to the Glivenko–Cantelli lemma, the value of $D$ is smaller if the null hypothesis is true, and the value is bigger if the underlying distribution and the hypothesized distribution are different (DeGroot and Schervish, 2010). Therefore, it is reasonable to reject the null hypothesis when the decision statistics $\sqrt{n}D$ is higher than a constant $C$ [equation (4.5)].

$$\sqrt{n}D > C \tag{4.5}$$

The above distribution testing method is called the one-sample Kolmogorov–Smirnov test and the selection of the decision threshold constant is not presented as it is not needed in the application of modulation classification.

In the second scenario, there are two sets of observed independent variables $x_1$, $x_2, \ldots, x_m$ and $y_1, y_2, \ldots, y_n$. Each of these data sets has an underlying distribution $F(x)$ and $G(x)$, respectively. To test if these two sets of data come from the same underlying cumulative distribution, the following Kolmogorov–Smirnov test null hypothesis $\mathcal{H}_0$ can be constructed [equation (4.6)].

$$\mathcal{H}_0 : \quad F(x) = G(x) \tag{4.6}$$

Following the same rule as in equation (4.1), the Kolmogorov–Smirnov test statistics in this scenario can be derived as the supremum of the distance between the two underlying cumulative distributions [equation (4.7)].

$$D = \sup_{-\infty < x < \infty} |F(x) - G(x)| \tag{4.7}$$

Practically, the underlying cumulative distribution function is replaced by the empirical cumulative distribution function sampled from the observations. The sampling of the empirical cumulative distribution function follows the same rule as in equation (4.2).

$$\hat{F}(x) = \frac{1}{m}\sum_{i=1}^{m} \mathbb{I}(x_i \le x) \tag{4.8}$$

$$\hat{G}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(y_i \le x) \tag{4.9}$$

Equations (4.8) and (4.9) then lead to the updated test statistics representing the maximum distance between the empirical cumulative distribution functions from the two data sets [equation (4.10)].

$$D = \max_{-\infty < x < \infty} \left| \hat{F}(x) - \hat{G}(x) \right| \tag{4.10}$$

The null hypothesis is then rejected if the test statistic $\sqrt{mn}D$ is larger than a constant $C$, equation (4.11).

$$\sqrt{mn}D > C \tag{4.11}$$

The aforementioned method is named the two-sample Kolmogorov–Smirnov test. Both of the two applications of the Kolmogorov–Smirnov test can be used for modulation classification problems with different settings, and they achieve different performance characters. The implementation is presented in the following subsections.

### 4.2.2 One-sample KS Test Classifier

In the context of modulation classification, we assume there are a number $N$ of received signal samples $r[1]$, $r[2],\ldots,r[N]$ in the AWGN channel. The signal samples are first normalized to zero mean and unit power. The normalization is implemented on the in-phase and quadrature segments of the signal samples separately, as shown by

$$r_I[n] = \frac{\Re(r[n]) - \overline{\Re(r)}}{\sigma(\Re(r))} \tag{4.12}$$

$$r_Q[n] = \frac{\Im(r[n]) - \overline{\Im(r)}}{\sigma(\Im(r))} \tag{4.13}$$

equations (4.12) and (4.13), where $\overline{\Re(r)}$ and $\overline{\Im(r)}$ are the mean of the real and imaginary part of the complex signal, with $\sigma(\Re(r))$ and $\sigma(\Im(r))$ being the standard deviation of the real and imaginary part of the complex signal. In the case of non-blind modulation classification, the effective channel gain and noise variance after normalization is assumed to be known. The assumption is demanding, while alternatives can be found where these parameters are estimated as part of a blind modulation classification. More discussion on blind modulation classification will be given in Chapter 7.

For the hypothesis modulation $\mathcal{M}(i)$ (with alphabet $A_m \in A, m = 1,\ldots,M$) in the AWGN channel with effective gain $\alpha$ and noise variance $\sigma^2$, the hypothesis cumulative distribution function can be derived from the PDF of signal I-Q segments in equation (2.6), as given in equations (4.14) and (4.15).

$$F_i^I(x) = \int_{-\infty}^{x} \sum_{m=1}^{M} \frac{1}{M} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{|x-\Re(\alpha A_m)|^2}{2\sigma^2}} dx \tag{4.14}$$

$$F_i^Q(x) = \int_{-\infty}^{x} \sum_{m=1}^{M} \frac{1}{M} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{|x-\Im(\alpha A_m)|^2}{2\sigma^2}} dx \tag{4.15}$$

As only the cumulative distributions at the signal samples are need, the cumulative distribution values are calculated for $F_i^I(\Re(r[1])), F_i^I(\Re(r[2])),\ldots,F_i^I(\Re(r[N]))$ and $F_i^Q(\Im(r[1])), F_i^Q(\Im(r[2])),\ldots,F_i^Q(\Im(r[N]))$. These values are calculated during the classification process and therefore the computational complexity should be included as part of the classifier. The empirical cumulative distribution function is calculated following equation (4.2), as is shown in equations (4.16) and (4.17).

$$\hat{F}^I(x) = \frac{1}{N}\sum_{n=1}^{N} \mathbb{I}(\Re(r[n]) \le x) \tag{4.16}$$

$$\hat{F}^Q(x) = \frac{1}{N}\sum_{n=1}^{N} \mathbb{I}(\Im(r[n]) \le x) \tag{4.17}$$

It is worth noting that the empirical cumulative distribution is independent of the test hypothesis. Therefore the collected values can be reused for all modulation hypotheses.

With both the hypothesized cumulative distribution function and empirical cumulative distribution function ready, the test statistics of the one-sample Kolmogorov–Smirnov test can be found for each signal I-Q segments, as set out in equations (4.18) and (4.19).

$$D_i^I = \max_{1 \le n \le N} \left| \hat{F}^I(\Re(r[n])) - F_i^I(\Re(r[n])) \right| \tag{4.18}$$

$$D_i^Q = \max_{1 \le n \le N} \left| \hat{F}^Q(\Im(r[n])) - F_i^Q(\Im(r[n])) \right| \tag{4.19}$$

To accommodation the multiple test statistics calculated from multiple signal segments, they are simply averaged to create a single test statistics for the modulation decision making [equation (4.20)].

$$D_i = \frac{1}{2}\left( \max_{1 \le n \le N} \left| \hat{F}^I(\Re(r[n])) - F_i^I(\Re(r[n])) \right| + \max_{1 \le n \le N} \left| \hat{F}^Q(\Im(r[n])) - F_i^Q(\Im(r[n])) \right| \right) \tag{4.20}$$

In some cases when the modulation candidates have identical distribution (e.g., M-PSK, M-QAM) on their in-phase and quadrature segments their empirical cumulative distribution can be combined to form an empirical cumulative distribution function with larger statistics [equation (4.21)].

$$\hat{F}(x) = \frac{1}{2N}\sum_{n=1}^{N} \mathbb{I}(\Re(r[n]) \le x) + \mathbb{I}(\Im(r[n]) \le x) \tag{4.21}$$

Since the signal samples are complex, the multidimensional version of the KS test has been discussed in Peacock (1983) and Fasano and Franceschini (1987). We suggest that the corresponding test statistics can be modified to give equation (4.22),

$$D_i = \max_{1 \le n \le 2N} \left| \hat{F}(z[n]) - F_i^I(z[n]) \right| \tag{4.22}$$

where the test sampling locations are a collection of the in-phase and quadrature segments of the signal samples, as given in equation (4.23).

$$z_{2n-1} = \Re(r[n]),\ \ z_{2n} = \Im(r[n]) \tag{4.23}$$

Regardless of the format of test statistics the classification decision is based on the comparison of the test statistics from all modulation hypotheses. The modulation decision is assigned to the hypothesis with the smallest test statistics [equation (4.24)].

$$\hat{\mathcal{M}} = \underset{\mathcal{M}_i \in \mathfrak{M}}{\arg\min}\, D_i \tag{4.24}$$

### 4.2.3 Two-sample KS Test Classifier

When the channel is relatively complex and the hypothesis cumulative distribution function is difficult to be modelled accurately, the two sample Kolmogorov–Smirnov test may be much easier to implement. However, training/pilot samples are needed to construct the reference empirical cumulative distribution functions. Without any prior assumption on the channel state, $K$ training samples $x[1], x[2], \ldots, x[K]$ are transmitted using modulation $\mathcal{M}(i)$. The empirical cumulative distribution functions can be found following equations (4.16) and (4.17), and are given in equations (4.25) and (4.26).

$$\hat{F}_i^I(x) = \frac{1}{N}\sum_{n=1}^{N} \mathbb{I}(\Re(x[n]) \le x) \tag{4.25}$$

$$\hat{F}_i^Q(x) = \frac{1}{N}\sum_{n=1}^{N} \mathbb{I}(\Im(x[n]) \le x) \tag{4.26}$$

The empirical cumulative distribution function of the $N$ testing signal samples $r[1]$, $r[2]$,… $r[N]$ are formulated in the same way as in equations (4.16) and (4.17). Using the two-sample test statistic in equations (4.10) and (4.19), the two-sample test statistics for modulation classification can be found as shown in equation (4.27).

$$D_i = \frac{1}{2}\left( \max_{-\infty < x < \infty} \left| \hat{F}^I(x) - \hat{F}_i^I(x) \right| + \max_{-\infty < x < \infty} \left| \hat{F}^Q(x) - \hat{F}_i^Q(x) \right| \right) \tag{4.27}$$

In practical implementations, it is easier to quantize the testing range of $x$ into a set of evenly distributed sampling locations.

The classification rule is the same as for the one-sample Kolmogorov–Smirnov test where the modulation hypothesis with the smallest test statistics is assigned as the classification decision.

### 4.2.4 Phase Difference Classifier

So far we have only used the signal in-phase and quadrature segments for the implementation of distribution test. However, there is an interesting signal feature that could be incorporated in the distribution test for exploiting the distinction of modulation signal in channels with phase or frequency offset. As discussed in Chapter 2, a fading channel, especially one with frequency offset, introduces severe distortion in the received signal in the form of a progressive rotational shift in its constellations. By measuring the phase difference between the adjacent signal samples, the phase difference error is reduced to a small constant. Thus, it has much less impact on the robustness of the modulation classifier.

In the AWGN channel where additive noises have Gaussian distribution, the theoretical CDF of the phase difference between adjacent signal sample vectors is derived by Pawula, Rice and Roberts (1982) as given in equation (4.28),

$$F_{\Delta\theta}(x) = \frac{\sin(\mu_{\Delta\theta} - x)}{4\pi} \int_{-\pi/2}^{\pi/2} \frac{e^{\frac{\alpha^2\left[1-\cos\left(\mu_{\Delta\theta}-x\right)\cos t\right]}{\sigma^2}}}{1-\cos(\mu_{\Delta\theta}-x)\cos t} dt \tag{4.28}$$

where $\mu_{\Delta\theta}$ is the mean of the phase difference. The corresponding CDF for an M-ary PSK modulation can be found as shown in equation (4.29).

$$F_{\Delta\theta}(x) = \frac{1}{M}\sum_{m=1}^{M} \frac{\sin\left(\frac{2(m-1)}{M}\pi - x\right)}{4\pi} \int_{-\pi/2}^{\pi/2} \frac{e^{\frac{\alpha^2\left[1-\cos\left(\frac{2(m-1)}{M}\pi-x\right)\cos t\right]}{\sigma^2}}}{1-\cos\left(\frac{2(m-1)}{M}\pi - x\right)\cos t} dt \tag{4.29}$$

To implement the phase difference classifier, the only remaining task is to evaluate the empirical phase distribution by using equation (4.30).

$$\hat{F}^{I}_{\Delta\theta}(x) = \frac{1}{N}\sum_{n=1}^{N} \mathbb{I}\left(\arg(r[n]) \le x\right) \tag{4.30}$$

Using the one-sample KS test, the test statistics of the phase difference test can be calculated by using equation (4.31).

$$D_i = \max_{1 \le n \le N} \left|\hat{F}_{\Delta\theta}(r[n]) - F_{\Delta\theta}(r[n])\right| \tag{4.31}$$

The classification decision making is the same as the KS test in which the hypothesis with the smallest test statistics is selected as the classification decision.

$$\hat{\mathcal{M}} = \arg\min_{\mathcal{M}_i \in \mathfrak{M}} D_i \tag{4.32}$$

## 4.3 Cramer–Von Mises Test Classifier

The Cramer–von Mises test (CvM test), also known as the Cramer–von Mises criterion, is an alternative to the KS test for goodness of fit evaluation (Honda, Oka and Ata, 2012). It is named after Harald Cramer and Richard Elder von Mises, who first introduced the method. Using the same one-sample test scenario as for the KS test, the test statistics is defined as the integral of the squared difference between the empirical CDF and the hypothesized CDF (Anderson, 1962), equation (4.33).

$$D = \int_{-\infty}^{\infty} [F_1(x) - F_0(x)]^2 dF_0(x) \tag{4.33}$$

Anderson generalized the two-sample variety of the Cramer–von Mises test that evaluates the goodness of fit between two sets of observed data. The two-sample test statistics is given by equation (4.34),

$$D = \frac{MN}{M+N} \int_{-\infty}^{\infty} [F(x) - G(x)]^2 dH_{M+N}(x) \tag{4.34}$$

where $H_{M+N}(x)$ is the empirical CDF of the combination of two sets of samples [equation (4.35)].

$$(M+N)H_{M+N} = MF(x) + NG(x) \tag{4.35}$$

In practice the test statistic is evaluated in the form of summation instead [equation (4.36)].

$$D = \frac{MN}{(M+N)^2} \left\{ \sum_{m=1}^{M} [F(x_n) - G(x_n)]^2 + \sum_{n=1}^{N} [F(y_n) - G(y_n)]^2 \right\} \tag{4.36}$$

Like the KS test classifier, the modulation classification decision using the Cramer–von Mises test is also assigned by the hypothesis with the smallest test statistics.

$$\hat{\mathcal{M}} = \underset{\mathcal{M}_i \in \mathfrak{M}}{\arg\min}\, D_i \tag{4.37}$$

## 4.4 Anderson–Darling Test Classifier

Both the KS test and Cramer–von Mises tests are relatively less sensitive when the difference between distributions is at the tails of the distributions. This is because

the cumulative distribution converges to zero and one at the tails of the distribution. To overcome this problem, Anderson and Darling proposed a weighted version of the Cramer–von Mises test called the Anderson–Darling (AD) test which gives more weight to tails of the distribution (Anderson and Darling, 1954). The test statistics is the Cramer–von Mises test with an added weight function $w(x)$, equation (4.38).

$$D = \int_{-\infty}^{\infty} [F_1(x) - F_0(x)]^2 w(x) dF_0(x) \tag{4.38}$$

The weight function is defined by equation (4.39),

$$w(x) = \frac{1}{F(x)(1-F(x))} \tag{4.39}$$

which accentuates the distribution mismatch at the tails of the distribution when the CDF is close to zero or one. The classification is achieved by comparing the test statistics where the modulation candidate associated with the smallest test statistics is assigned as the modulation classification decision. The decision making is the same as in the KS test classifier and the Cramer–von Mises test classifier, equation (4.40).

$$\hat{\mathcal{M}} = \underset{\mathcal{M}_i \in \mathfrak{M}}{\arg\min}\, D_i \tag{4.40}$$

## 4.5 Optimized Distribution Sampling Test Classifier

After the Kolmogorov–Smirnov test was proposed for modulation classification, Urriza *et al.* recognized the possibility to reduce its complexity further and potentially improve its classification performance (Urriza *et al.*, 2011). The modification is enabled by establishing the sampling point prior to the distribution test using the theoretical cumulative distribution functions of modulation hypotheses. Assuming that there are two modulation candidates $\mathcal{M}(i)$ and $\mathcal{M}(j)$, the cumulative distribution functions of both modulations in a channel with known channel state information $F_i(x)$ and $F_j(x)$ are established prior to the distribution test. The sampling point $l_{ij}$ in the test is found where the maximum distance between $F_i(x)$ and $F_j(x)$ is observed [equation (4.41)].

$$l_{ij} = \arg\max_{x} \left| F_i(x) - F_j(x) \right| \tag{4.41}$$

The motivation for the optimized sampling location is first for the complexity reduction in calculating the empirical cumulative distribution function using the test

data. Instead of calculating $N$ values for the empirical cumulative distribution function using the optimized sampling location, the number is reduced to one. Additionally, assuming that the channel model is accurate, the optimized sampling location is a better utilization of the prior channel knowledge where the potential location for the maximum distance between the empirical cumulative distribution function and the theoretical cumulative distribution function is selected prior to the distribution test. It is especially effective when the number of samples available for constructing the empirical cumulative distribution function is limited and the ECDF may be misrepresenting the underlying CDF of the transmitted signal. In such cases, the maximum distance could occur where outliners are observed and the test results may be inaccurate for the classification decision making.

The test statistic is collected differently without the complexity of reconstructing the complete empirical cumulative distribution function and maximizing its difference with the hypothesized cumulative distribution. Instead the test statistic is only calculated by sampling the single location defined by equation (4.27), as shown in equation (4.42),

$$D_i = \left|\hat{F}\left(l_{ij}\right) - F_i\left(l_{ij}\right)\right| \tag{4.42}$$

where $\hat{F}_0\left(l_{ij}\right)$ is the empirical cumulative distribution probability at location $l_{ij}$, equation (4.43).

$$\hat{F}_0\left(l_{ij}\right) = \frac{1}{N}\sum_{n=1}^{N} \mathbb{I}\left(\mathfrak{R}(x[n]) \le l_{ij}\right) \tag{4.43}$$

The classification decision making is the same as in the Kolmogorov–Smirnov test classifier in the case of two-class classification.

Being an improvement of the Kolmogorov–Smirnov test classifier, the aforementioned modification was further improved and formalized by Zhu, Aslam and Nandi (2014). The rest of this section will be dedicated to their proposed version of the optimized distribution sampling test (ODST) classification.

### 4.5.1 Sampling Location Optimization

To further enhance the robustness of the Kolmogorov–Smirnov test-based classifier, Zhu *et al.* (2014) suggested using multiple pre-optimized sampling locations instead of a signal location suggested by Urriza *et al.* (2011). When using multiple sampling locations, there are two factors that need to be considered: the number of sampling locations and how they are found or optimized. Though using more sampling locations always provides more information when the distribution test is conducted, some locations may contribute much more than others. For example, two adjacent sampling locations would provide similar information of the ECDF/CDF, while having just one of them would be enough.

After analysis of the signal distribution and investigation of these distributions in various channel conditions, we have concluded that the local maximum of the difference between two CDFs is a good option. Here we define $\mathcal{L}_{ij}$ as a set of sampling locations with $K$ individual sampling locations $l_1^{ij}, l_2^{ij}, \ldots, l_K^{ij}$. Following the rule for sampling location optimization, the sample locations should enable the distance between the CDFs $F_i(x)$ and $F_j(x)$ of modulation $\mathcal{M}(i)$ and $\mathcal{M}(j)$ to be a local maximum. A similar idea has been proposed by Wang and Chan (2012). The rule could then easily be converted to give the gradient of the function $F_i(x) - F_j(x)$, which should be zero at the locations defined by $\mathcal{L}$ [equation (4.44)].

$$\frac{d}{dx}\left(F_i\left(\mathcal{L}_{ij}\right) - F_j\left(\mathcal{L}_{ij}\right)\right) = 0 \tag{4.44}$$

For the convenience of optimization of the sampling location using the theoretical distribution function, the CDFs in equation (4.28) can be replaced with their corresponding PDFs, resulting in the following updated rule given by equation (4.45).

$$F_i\left(\mathcal{L}_{ij}\right) = F_j\left(\mathcal{L}_{ij}\right) \tag{4.45}$$

It is worth noting that the sampling location is only optimized for the discrimination between modulation $\mathcal{M}(i)$ and $\mathcal{M}(j)$. If other modulation sets are considered, a different set of sampling locations should be optimized for the new modulation set.

### 4.5.2 Distribution Sampling

Having found the optimized sampling locations, the distributions are sampled to collect the necessary test statistics. Following the scenario in the previous section, we assume there is a piece of signal $r[1], r[2], \ldots, r[N]$ with $N$ samples. The test statistics $t_1^{ij}, t_2^{ij}, \ldots, t_K^{ij}$ are sampled empirical cumulative distribution functions at the optimized sampling locations [equation (4.46)].

$$t_k^{ij} = \frac{1}{2N}\left[\sum_{n=1}^{N} \mathbb{I}\left(\Re(r[n]) < l_k^{ij}\right) + \sum_{n=1}^{N} \mathbb{I}\left(\Im(r[n]) < l_k^{ij}\right)\right] \tag{4.46}$$

It is worth noting that these sampled values have also been suggested as distribution features by Zhu, Nandi and Aslam (2013).

Before classification decision making can proceed, the reference CDF for each hypothesized models has to be calculated as well. Among the two candidate modulations, the reference CDF values need to be sampled separately from each of them. For

modulation $\mathcal{M}(i)$, the reference CDF values $T_1^{ij}, T_2^{ij}, \ldots, T_K^{ij}$ are calculated using their theoretical CDF and the optimized sampling locations [equation (4.47)].

$$T_k^{ij} = F_i\left(l_k^{ij}\right) \tag{4.47}$$

The reference CDF values for modulation $\mathcal{M}(j)$ are denoted with a slight difference. The order of $i$ and $j$ in the superscript is inverted so that the letter in front could represent where the reference value is gathered from. The actual computation follows the same method with only the CDF values being replaced by the those from modulation $\mathcal{M}(j)$ [equation (4.48)].

$$T_k^{ji} = F_j\left(l_k^{ij}\right) \tag{4.48}$$

### 4.5.3 Classification Decision Metrics

Having calculated the sampled CDF values and the reference CDF values from modulation hypotheses, the next step would be to evaluate the GoF between observed signals and hypothesized models. Unlike the KS test and other aforementioned distribution test classifiers, there are multiple sampling locations in this method. Therefore, the test statistics need be formulated as a combination of the sampled and reference CDF values. Before we present the test statistics formula, we first introduce the difference statistics that marks the difference between sampled values and reference values at individual sampling locations. The different statistics for hypothesized modulation $\mathcal{M}(i)$ at the $k$th sampling location is defined as the difference between the sampled CDF at $k$th sampling location and the reference CDF value of modulation $\mathcal{M}$(i) at the same sampling location [equation (4.49)].

$$\Delta t_k^{ij} = \left| t_k^{ij} - T_k^{ij} \right| \tag{4.49}$$

The corresponding expression for the alternative hypothesis of modulation $\mathcal{M}(j)$ is given by equation (4.50).

$$\Delta t_k^{ji} = \left| t_k^{ij} - T_k^{ji} \right| \tag{4.50}$$

The above calculation results in $k$ difference statistics for each modulation hypothesis. To combine these difference statistics into a signal test statistics, we have proposed two approaches. First, the test statistics $D_{ij}$ for hypothesized modulation $\mathcal{M}(i)$ is formulated as a non-negative uniform linear combination of the difference statistics, as shown in equation (4.51).

$$D_{ij} = \sum_{k=1}^{K} \Delta t_k^{ij} \tag{4.51}$$

The corresponding expression for the alternative hypothesized modulation $\mathcal{M}(j)$ is given by equation (4.52).

$$D_{ji} = \sum_{k=1}^{K} \Delta t_k^{ji} \tag{4.52}$$

Second, the test statistics $D_{ij}$ for hypothesized modulation $\mathcal{M}(i)$ is formulated as a non-negative weighted linear combination of the difference statistics [equation (4.53)],

$$D_{ij} = \sum_{k=1}^{K} w_k^{ij} \Delta t_k^{ij} \tag{4.53}$$

where $w_k^{ij}$ is the weight for the difference statistics $\Delta t_k^{ij}$. The corresponding expression for the alternative hypothesized modulation $\mathcal{M}(j)$ is given by equation (4.54).

$$D_{ji} = \sum_{k=1}^{K} w_k^{ij} \Delta t_k^{ji} \tag{4.54}$$

There are different ways of optimizing the weights. Among which, Zhu *et al.* used a genetic algorithm to train the weights with training signals (Zhu, Aslam and Nandi, 2014).

### 4.5.4 Modulation Classification Decision Making

The decision making in the optimized distribution sample test is different from that in the other distribution test-based classifiers. As the distribution sampling location is optimized between two modulation candidates, it could be easily implemented for a two-class classification problem. The decision is given by the modulation hypothesis which returns the smallest tests statistics [equations (4.55)].

$$\hat{\mathcal{M}} = \begin{cases} \mathcal{M}(i), \text{ if } D_{ij} \le D_{ji} \\ \mathcal{M}(j), \text{ if } D_{ij} > D_{ji} \end{cases} \tag{4.55}$$

If multi-class classification is required, the distribution test needs to be performed for every pair of two candidate modulations. The modulation candidate assigned the most time among all the paired distribution tests will be returned as the final classification decision.

## 4.6 Conclusion

In this chapter we have listed some of the distribution test-based classifiers. The distribution test-based classifiers are selected mostly for their computational complexity and suboptimal performance. Given that the channel parameters are known to the classifier, a one-sample KS test classifier or a two-sample KS test classifier achieves modulation classification by comparing the empirical distribution of the observed signal and the theoretical distribution of each modulation hypothesis. The Cramer–von Mises test classifier provides a better evaluation of the mismatch between the empirical distribution and hypothesis distribution, while the Anderson–Darling test classifier suggests a weighted evaluation of the goodness of it over the entirety of the signal distribution. The optimized distribution sample test is presented as an improved version of the KS test classifier with enhanced classification accuracy and robustness.

## References

Anderson, T. (1962) On the distribution of the two-sample Cramer–von Mises criterion. *The Annals of Mathematical Statistics*, **33** (3), 1148–1159.

Anderson, T.W. and Darling, D.A. (1954) A test of goodness of fit. *Journal of the American Statistical Association*, **49** (268), 765–769.

Conover, W. (1980) *Practical Nonparametric Statistics*, John Wiley & Sons, Inc., New York, NY.

DeGroot, M.H. and Schervish, M.J. (2010) *Probability and Statistics*, 4th edn, Pearson, Boston, MA.

Fasano, G. and Franceschini, A. (1987) A multidimensional version of the Kolmogorov-Smirnov test. *Monthly Notices of the Royal Astronomical Society*, **225**, 155–170.

Honda, C., Oka, I. and Ata, S. (2012) Signal Detection and Modulation Classification Using a Goodness of Fit Test. International Symposium on Information Theory and its Applications, Honolulu, HI, 28–31 October 2012, pp. 180–183.

Kolmogorov, A.N. (1933) Sulla Determinazione Empirica di una Legge di Distribuzione. *Giornale dell' Istituto Italiano degli Attuari*, **4**, 83–91.

Massey, F. (1951) The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, **46** (253), 68–78.

Pawula, R.F., Rice, S.O. and Roberts, J.H. (1982) Distribution of the phase angle between two vectors perturbed by gaussian noise. *IEEE Transactions on Communications*, **30** (8), 1828–1841.

Peacock, J.A. (1983) Two-dimensional goodness-of-fit testing astronomy. *Monthly Notices of the Royal Astronomical Society*, **202**, 615–627.

Smirnov, H. (1939) Sur les Ecarts de la Courbe de Distribution Empirique. *Recueil Mathematique (Matematiceskii Sbornik)*, **6**, 3–26.

Swami, A. and Sadler, B.M. (2000) Hierarchical digital modulation classification using cumulants. *IEEE Transactions on Communications*, **48** (3), 416–429.

Urriza, P., Rebeiz, E., Pawełczak, P. and Čabrić, D. (2011) Computationally efficient modulation level classification based on probability distribution distance functions. *IEEE Communications Letters*, **15** (5), 476–478.

Wang, F. and Chan, C. (2012) Variational-Distance-Based Modulation Classifier. IEEE International Conference on Communications (ICC), Ottawa, ON, 10–15 June 2012, pp. 5635–5639.

Wang, F. and Wang, X. (2010) Fast and robust modulation classification via Kolmogorov-Smirnov test. *IEEE Transactions on Communications*, **58** (8), 2324–2332.

Wei, W. and Mendel, J.M. (2000) Maximum-likelihood classification for digital amplitude-phase modulations. *IEEE Transactions on Communications*, **48** (2), 189–193.

Zhu, Z., Aslam, M.W. and Nandi, A.K. (2014) Genetic algorithm optimized distribution sampling test for M-QAM modulation classification. *Signal Processing*, **94**, 264–277.

Zhu, Z., Nandi, A.K. and Aslam, M.W. (2013) Robustness Enhancement of Distribution Based Binary Discriminative Features for Modulation Classification. IEEE International Workshop on Machine Learning for Signal Processing, Southampton, 22–25 September, 2013, pp. 1–6.

# 5

# Modulation Classification Features

## 5.1 Introduction

In Chapters 3 and 4, two decision theoretic-based approaches to AMC are presented. For certain modulations such as AM and FM, it is obvious that signal distribution is not the only way to differentiate the two modulations. By analyzing the nature of the modulation technique, one can easily identify the key features of a signal modulated using a specific modulation scheme. While decision theoretic-based classifiers provide excellent classification accuracy, their high computational complexity motivates the development of feature-based classifiers which yield sub-optimal performance for much lower computational requirements.

In this chapter we list some of the well-recognized features designed for modulation classification. While the book focuses on digital modulations, features historically used for the classification of analogue modulations will also be included in this chapter. In the remainder of the chapter we first investigate the spectral-based features which exploits the spectral properties of different signal components. The wavelet-based features are given as another approach to feature-based modulation classification using the signal waveform. The high-order statistic features are examined as optioned to classifier digital modulations of different type and orders. The cyclic features based on cyclostationary analysis are presented at the end. More advanced feature-based (FB) classifiers which could utilize the features listed in this chapter are presented in Chapter 6.

*Automatic Modulation Classification: Principles, Algorithms and Applications*, First Edition. Zhechen Zhu and Asoke K. Nandi.

## 5.2 Signal Spectral-based Features

Nandi and Azzouz proposed some key signal spectral-based features in the 1990s for the classification of basic analogue and digital modulations (Azzouz and Nandi, 1995, 1996b; Nandi and Azzouz, 1995). These key features generalized and advanced the works of Fabrizi, Lopes and Lockhart (1986), Chan and Gadbois (1989), and Jovanovic, Doroslovacki and Dragosevic (1990), which suggested different feature-extraction methods. The features exploit the unique spectral characters of different signal modulations in three key signal aspects, namely the amplitude, phase and frequency. Since different signal modulations exhibit different properties in their amplitude, phase and frequency, a complete pool of modulation candidates is broken down to sets and subsets which can be discriminated with the most effective features. A decision tree, consisting of nodes of sequential tests dedicated by different features, is often employed to give a clear guideline for the classification procedure.

In the following subsection, we first list these key features. Second, the features are analyzed for their capability in distinguishing different modulation sets. The section is concluded with some proposed classification decision trees suggested by Nandi and Azzouz for the classification of analogue and digital modulation.

### 5.2.1 Signal Spectral-based Features

The first feature, $\gamma_{max}$, is the maximum value of the spectral power density of the normalized and centred instantaneous amplitude of the received signal (Azzouz and Nandi, 1996a), equation (5.1),

$$\gamma_{max} = \max|\mathrm{DFT}(A_{cn})|^2/N \tag{5.1}$$

where DFT(-) is the discrete Fourier transform (DFT), $A_{cn}$ is the normalized and centred instantaneous amplitude of the received signal $r$, and $N$ is the total number signal samples. The normalization is achieved by adopting equation (5.2),

$$A_{cn}[n] = A_n[n] - 1, \text{ where } A_n[n] = \frac{A[n]}{\mu_A} \tag{5.2}$$

where $\mu_A$ is the mean of the instantaneous amplitude of one signal segment, equation (5.3).

$$\mu_A = \frac{1}{N}\sum_{n=1}^{N} a[n] \tag{5.3}$$

The normalization of the signal amplitude is designed to compensate the unknown channel attenuation.

The second feature, $\sigma_{ap}$, is the standard deviation of the absolute value of the non-linear component of the instantaneous phase, equation (5.4).

$$\sigma_{ap} = \sqrt{\frac{1}{N_c}\left(\sum_{A_n[n] > A_t} \phi_{NL}^2[n]\right) - \left(\frac{1}{N_c}\sum_{A_n[n] > A_t} \left|\phi_{NL}[n]\right|\right)^2} \tag{5.4}$$

where $N_c$ is the number of samples that meet the condition:$A_n[n] > A_t$. The variable $A_t$ is a threshold value which filters out the low-amplitude signal samples because of their high sensitivity to noise. Term $\phi_{NL}[n]$ denotes the non-linear component of the instantaneous phase of the $n$th signal sample.

The third feature, $\sigma_{dp}$, is the standard deviation of the non-linear component of the direct instantaneous phase, given by equation (5.5),

$$\sigma_{dp} = \sqrt{\frac{1}{N_c}\left(\sum_{A_n[n] > A_t} \phi_{NL}^2[n]\right) - \left(\frac{1}{N_c}\sum_{A_n[n] > A_t} \phi_{NL}[n]\right)^2} \tag{5.5}$$

where all parameters remain the same as in the expression for $\sigma_{ap}$. However, it is noticeable that the absolute operation on the non-linear component of the instantaneous phase is removed.

The fourth feature, $P$, is an evaluation of the spectrum symmetry around the carrier frequency, evaluated by equation (5.6),

$$P = \frac{P_L - P_U}{P_L + P_U} \tag{5.6}$$

where equations (5.7) and (5.8) hold, such that when

$$P_L = \sum_{n=1}^{f_{cn}} \left|X_c[n]\right|^2 \tag{5.7}$$

$$P_U = \sum_{n=1}^{f_{cn}} \left|X_c[n + f_{cn} + 1]\right|^2 \tag{5.8}$$

$X_c[n]$ is the Fourier transform of the signal $x_c[n]$, $(f_{cn} + 1)$ is the sample number corresponding to the carrier frequency $f_c$, and $f_s$ is the sampling rate, the equation (5.9) holds.

$$f_{cn} = \frac{f_c N}{f_s} - 1 \tag{5.9}$$

The fifth feature, $\sigma_{aa}$, is the standard deviation of the absolute value of the normalized and centred instantaneous amplitude of the signal samples, given by equation (5.10).

$$\sigma_{aa} = \sqrt{\frac{1}{N}\left(\sum_{n=1}^{N} A_{cn}^2[n]\right) - \left(\frac{1}{N}\sum_{n=1}^{N} \left|A_{cn}[n]\right|\right)^2} \tag{5.10}$$

The sixth feature, $\sigma_{af}$, is the standard deviation of the absolute value of the normalized and centred instantaneous frequency, given by equation (5.11),

$$\sigma_{af} = \sqrt{\frac{1}{N_c}\left(\sum_{A_n[n] > A_t} f_N^2[n]\right) - \left(\frac{1}{N_c}\sum_{A_n[n] > A_t} \left|f_N[n]\right|\right)^2} \tag{5.11}$$

where the centred instantaneous frequency $f_m$ is normalized by the sampling frequency $f_s$ such that equation (5.12) applies.

$$f_N[n] = f_m[n]/f_s \tag{5.12}$$

The instantaneous frequency is centred using the frequency mean $\mu_f$ as shown in equations (5.13) and (5.14).

$$f_m[n] = f[n] - \mu_f \tag{5.13}$$

$$\mu_f = \frac{1}{N}\sum_{n=1}^{N} f[n] \tag{5.14}$$

The seventh feature, $\sigma_a$, is the standard deviation of the normalized and centred instantaneous amplitude [equation (5.15)].

$$\sigma_a = \sqrt{\frac{1}{N_c}\left(\sum_{A_n[n] > A_t} a_{cn}^2[n]\right) - \left(\frac{1}{N_c}\sum_{A_n[n] > A_t} a_{cn}[n]\right)^2} \tag{5.15}$$

The eighth feature, $\mu_{42}^a$, is the kurtosis of the normalized and centred instantaneous amplitude, given by equation (5.16).

$$\mu_{42}^a = \frac{E\left\{A_{cn}^4[n]\right\}}{\left\{E\left\{A_{cn}^2[n]\right\}\right\}^2} \tag{5.16}$$

The ninth feature, $\mu_{42}^{f}$, is the kurtosis of the normalized and centred instantaneous amplitude [equation (5.17)].

$$\mu_{42}^{f} = \frac{E\left\{f_N^4[n]\right\}}{\left\{E\left\{f_N^2[n]\right\}\right\}^2} \tag{5.17}$$

## 5.2.2 Spectral-based Features Specialities

Having defined all of Nandi and Azzouz's signal spectral-based features, next we will analyze every key feature and suggest their usage in modulation classification. Building on Nandi and Azzouz's work, we have added PAM and QAM modulation to enrich the analysis of the features in a more up-to-date context.

The quantity $\gamma_{max}$ measures the variance in signal instantaneous amplitude. For modulations, where information is conveyed in the signal amplitude, the value of $\gamma_{max}$ should be non-zero. One the other hand, for modulations with constant amplitude, the value of $\gamma_{max}$ should be zero. Thus, we have two sets of modulations which can be classified by $\gamma_{max}$ and a corresponding threshold $t_{\gamma_{max}}$. The first set of modulations includes AM single-sideband modulation (SSB), M-ASK, M-PAM, M-PSK and M-QAM. The second set of modulations includes FM and M-FSK.

Quantity $\sigma_{ap}$ measures the variance in the absolute instantaneous phase. The first set of modulations includes FM, SSB, M-FSK, M-PSK ($M \geq 2$), M-QAM. The second set of modulations includes AM, M-ASK, BPSK. BPSK does not have information in its absolute instantaneous phase (centred), because there are only two states for the instantaneous phase. When the instantaneous phase is centred at zero, their absolute values are the same.

Quantity $\sigma_{dp}$ measures the variance in the absolute instantaneous phase. The first set of modulations includes FM, SSB, M-FSK, M-PSK, M-QAM. The second set of modulations includes AM, M-ASK. Being similar to $\sigma_{ap}$, $\sigma_{dp}$ provides the ability to distinguish BPSK from other modulations without phase information.

Quantity $P$ provides the criterion to classify different amplitude-based modulations with different properties in the frequency domain. One set of the modulations has a symmetrical spectrum arrangement about the carrier frequency. They include AM and double-sideband modulation (DSB). The other, obviously with asymmetrical spectrum density about the carrier frequency, includes vestigial sideband (VSB), lower sideband modulation (LSB) and upper sideband modulation (USB).

Quantity $\sigma_{aa}$ measures the amount of information in the signal's instantaneous amplitude. This feature is similar to $\gamma_{max}$. However, it is gifted with the extra ability to differentiate 2-ASK from other M-ASK modulations. This is because the absolute values of 2-ASK signal's instantaneous amplitude become identical when they are centred at zero.

Quantity $\sigma_{af}$ is primarily designed for the classification of 2-FSK and 4-FSK modulations. The theoory that it can enable discrimination is same as that for $\sigma_{ap}$ and $\sigma_{aa}$, where the frequency information of the binary-state modulation is removed by taking the absolute value of the centred instantaneous frequency.

Quantity $\mu_{42}^{a}$ measures the compactness of the instantaneous amplitude distribution. AM modulation, being an analogue modulation with varying instantaneous amplitude, has a less compact distribution compared with digital modulations such as M-ASK modulations.

Quantity $\mu_{42}^{f}$ measures the compactness of the instantaneous frequency distribution. Like AM modulation, FM has less compact frequency distribution compared with digital modulations such as M-FSK, because of its analogue instantaneous frequency.

### 5.2.3 Spectral-based Features Decision Making

Azzouz and Nandi (1995, 1996a, 1996b) designed decision trees for the classification of analogue and digital modulations. The trees consist of an input node where all the features are extracted and imported. The input node is followed by a sequence of conditional or decision steps facilitated with selected individual features. In this section, we have reorganized these decision trees and created a decision tree in Figure 5.1 for the classification of the aforementioned modulations.

The diamond block in Figure 5.1 represents a conditional sub-stage classification, with $t(\cdot)$ being the suitable threshold for different features.

### 5.2.4 Decision Threshold Optimization

To establish the decision threshold for each spectral-based feature, pilot samples could be used to achieve that. Assume we have $L$ signal realizations from group A and $L$ signal realizations from group B, where the two groups of modulations can be classified using feature $\mathbb{F}$. The optimal threshold can be calculated as shown in equation (5.18),

$$t(\mathbb{F}) = \frac{\sigma(\mathbb{F}_A)\mu(\mathbb{F}_B) + \sigma(\mathbb{F}_B)\mu(\mathbb{F}_A)}{\sigma(\mathbb{F}_A) + \sigma(\mathbb{F}_B)} \tag{5.18}$$

where $\sigma(\mathbb{F}_A)$ and $\sigma(\mathbb{F}_A)$ are the standard deviations of the feature values calculated using signals from group A and group B, respectively, using equations (5.19) and (5.20),

$$\sigma(\mathbb{F}_A) = \left[\frac{1}{L}\sum_{l=1}^{L}(\mathbb{F}_A(l) - \mu(\mathbb{F}_A))^2\right]^{\frac{1}{2}} \tag{5.19}$$

$$\sigma(\mathbb{F}_B) = \left[\frac{1}{L}\sum_{l=1}^{L}(\mathbb{F}_B(l) - \mu(\mathbb{F}_B))^2\right]^{\frac{1}{2}} \tag{5.20}$$

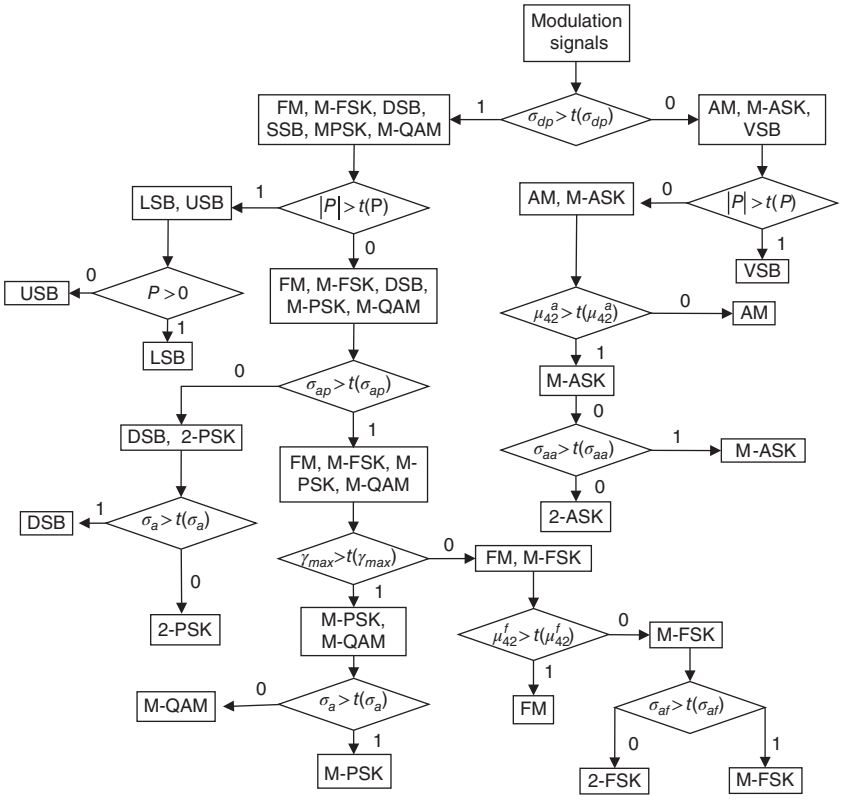



**Figure 5.1** Decision tree for modulations classification using spectral-based features.

and $\mu(\mathbb{F}_A)$ and $\mu(\mathbb{F}_B)$ are the means of the features values from each group, equations (5.21) and (5.22).

$$\mu(\mathbb{F}_A) = \frac{1}{L}\sum_{l=1}^{L} \mathbb{F}_A(l) \tag{5.21}$$

$$\mu(\mathbb{F}_B) = \frac{1}{L}\sum_{l=1}^{L} \mathbb{F}_B(l) \tag{5.22}$$

## 5.3 Wavelet Transform-based Features

The continuous wavelet transform of received signal $r$ is defined as the integral of $r(t)$ times the conjugate transpose of the wavelet function $\psi_{a,\tau}(t)$ over time, equation (5.23).

$$\mathrm{CWT}(a,\tau)=\int_{-\infty}^{\infty} r(t)\psi_{a,\tau}^{*}(t)dt \tag{5.23}$$

Among different mother wavelet functions, namely Morlet, Haar and Shannon, most researchers selected the Haar wavelet function because of its simple form and computational convenience. The Haar wavelet is given by equation (5.24),

$$\psi(t)=\begin{cases}1, & \text{if } 0\le t<\dfrac{T}{2}\\ -1, & \text{if } \dfrac{T}{2}\le t<T\\ 0, & \text{otherwise}\end{cases} \tag{5.24}$$

which can be scaled and translated to create a series of baby wavelets [equation (5.25)].

$$\psi_{a,\tau}(t)=\frac{1}{\sqrt{a}}\psi\left(\frac{t-\tau}{a}\right) \tag{5.25}$$

To enable the analysis of the continuous wavelet transform (CWT) for different modulation signals, we first established individual expressions of the transmitted signal from each modulation. For M-ASK the representation of its transmitted signal is given by equation (5.26),

$$s_{ASK}(t)=\sum_{n=1}^{N}|A_n|g_{T_s}(t-nT_s) \tag{5.26}$$

where $|A_n|\in\{|A_1|, |A_2| \ldots |A_M|\}$ is the $n$th symbol being transmitted, with $M$ being the total number of states, and $g_{T_s}$ is the pulse-shaping function of duration $T_s$. The expression for the M-FSK is given by equation (5.27),

$$s_{FSK}(t)=\sqrt{S}\sum_{n=1}^{N}e^{j(\omega_n t+\phi_0)}g_{T_s}(t-nT_s) \tag{5.27}$$

where $S$ is average signal power, $\omega_n\in\{\omega_1,\omega_2 \ldots \omega_M\}$ is the signal frequency at the $n$th symbol, and $\phi_0$ is the initial carrier phase. For PSK, the expression is given by equation (5.28),

$$s_{PSK}(t)=\sqrt{S}\sum_{n=1}^{N}e^{j\phi_n}g_{T_s}(t-nT_s) \tag{5.28}$$

where $\phi_n\in\{\phi_1,\phi_2 \ldots \phi_M\}$ is the phase of the $n$th symbol. Lastly, the expression for the QAM modulation signal is defined by equation (5.29),

$$s_{QAM}(t)=\sum_{n=1}^{N}A_n g_{T_s}(t-nT_s) \tag{5.29}$$

where $A_n \in \{A_1, A_2 \ldots A_M\}$ is the complex envelope of the $n$th symbol being transmitted.

The derivation for ASK modulations in a noise-free environment is produced by Hassan *et al.* (2010) in the form of equation (5.30).

$$|\mathrm{CWT}_{\mathrm{ASK}}(a,\tau)| = \frac{4|A_n|}{\omega_c\sqrt{a}}\sin^2\left[\omega_c\frac{aT_s}{4}\right] \tag{5.30}$$

The corresponding CWT for FSK, PSK and QAM is derived by Hong and Ho (2000), as equations (5.31)–(5.33), respectively.

$$|\mathrm{CWT}_{\mathrm{FSK}}(a,\tau)| = \frac{4\sqrt{S}}{(\omega_c+\omega_n)\sqrt{a}}\sin^2\left[\frac{(\omega_c+\omega_n)aT_s}{4}\right] \tag{5.31}$$

$$|\mathrm{CWT}_{\mathrm{PSK}}(a,\tau)| = \frac{4\sqrt{S}}{\omega_c\sqrt{a}}\sin^2\left[\omega_c\frac{aT_s}{4}\right] \tag{5.32}$$

$$\left|\mathrm{CWT}_{QAM}(a,\tau)\right| = \frac{4|A_n|}{\omega_c\sqrt{a}}\sin^2\left[\omega_c\frac{aT_s}{4}\right] \tag{5.33}$$

It is obvious that the CWT of PSK is a constant, while others modulations all have multi-step functions as their CWT. Thus PSK can be distinguished from other modulations using the CWT. To classify FSK modulation, we examine the CWT of these modulations when they are normalized by amplitude [equation (5.34)].

$$\bar{s}(t) = \frac{s(t)}{|s(t)|} \tag{5.34}$$

The updated expressions for the transmitted symbols are given below in equations (5.35)–(5.38).

$$s_{ASK}(t) = \sum_{n=1}^{N} sign(|A_n|)g_{T_s}(t-nT_s) \tag{5.35}$$

$$s_{FSK}(t) = \sum_{n=1}^{N} e^{j(\omega_n t+\phi_0)}g_{T_s}(t-nT_s) \tag{5.36}$$

$$s_{PSK}(t) = \sum_{n=1}^{N} e^{j\phi_n}g_{T_s}(t-nT_s) \tag{5.37}$$

$$s_{QAM}(t) = \sum_{n=1}^{N} e^{j\mathrm{ang}(A_n)}g_{T_s}(t-nT_s) \tag{5.38}$$

It is not difficult to see that, for normalized signals, the variance in CWT is only seen for FSK modulations. Therefore, the CWT for normalized signals can be used as a feature for the classification of FSK modulations. Having achieved the classification of PSK and FSK, the only task left is to differentiate between ASK and QAM modulations. As there is no obvious distinction between the CWTs of these two modulation types, Hassan *et al.* (2010) suggested using the high-order moments and cumulants to exploit the small differences between these two modulation types. In addition, the application of the high-order moments and cumulants also enables the intra-class classification when modulation of a different order is involved. Because there is a separate section dedicated to high-order statistical features in this chapter, the high-order moments and cumulants will be presented in the corresponding section.

## 5.4 High-order Statistics-based Features

In this section, we focus on the high-order statistics (HoS)-based features, more specifically moment-based and cumulant-based features.

### 5.4.1 High-order Moment-based Features

Hipp was the first to adopt the third-order moment of the demodulated signal amplitude as a modulation-classification feature (Hipp, 1986). Since we consider the demodulated signal as a luxury for any modulation classifier, this moment-based feature is not investigated in this book.

The usage of moments in modulation classification was later extended by Soliman and Hsue, who investigated the high-order moments of the signal phase for the classification of M-PSK modulations (Soliman and Hsue, 1992). They derived the theoretical $k$th moment of signal phase in Gaussian channel which leads to the conclusion that the moments are a monotonically increasing function with respect to the order of the M-PSK modulation. Thus, high-order M-PSK modulations have higher moment values, which provide the condition for the classification of M-PSK modulations of different orders. Meanwhile, Soliman and Hsue also made the observation that the difference in moment value between higher-order modulations is not distinct for lower-order moments. Therefore, they concluded that the effective classification of M-PSK modulation with higher order requires the moments of higher order. The calculation of the $k$th order moment of the signal phase is defined as shown in equation (5.39),

$$\mu_k(\boldsymbol{r}) = \frac{1}{N}\sum_{n=1}^{N}\phi^k(n) \tag{5.39}$$

where $\phi(n)$ is the phase of the $n$th signal sample. Azzouz and Nandi (1996a) proposed the kurtosis of the normalized-centred instantaneous amplitude $\mu_{42}^a$ and the kurtosis of

the normalized-centred instantaneous frequency $\mu_{42}^{f}$ for the classification of M-ASK and M-FSK modulations. The expressions for these two features are given in equations (5.16) and (5.17). Hero and Hadinejad-Mahram generalized the moment-based features to include the high-order moment of signal phase and frequency magnitude (Hero and Hadinejad-Mahram, 1998). Spooner employed high-order cyclic moments as features (along with cyclic moments) for the classification of modulation with identical cyclic autocorrelation functions (Spooner, 1996).

In this book we use the following expression to calculated different $k$th moment of the complex-valued signal $\boldsymbol{r} = r[1], r[2], \ldots, r[N]$ [equation (5.40)],

$$\mu_{xy}(\boldsymbol{r}) = \frac{1}{N}\sum_{n=1}^{N} r^{x}[n]\cdot r^{*y}[n] \tag{5.40}$$

where $x + y = k$ and $r^{*}[n]$ is the complex conjugate of $r[n]$.

## 5.4.2 High-order Cumulant-based Features

Swami and Sadler (2000) suggested the fourth-order cumulant of the complex-valued signal as features for the classification of M-PAM, M-PSK and M-QAM modulations. For signal $r[n]$ the second-order moments can be defined in either of two different ways [equations (5.41) and (5.42)].

$$C_{20} = E\{r^{2}[n]\} \tag{5.41}$$

$$C_{21} = E\{|r[n]|^{2}\} \tag{5.42}$$

Likewise, the fourth-order moments and cumulants can be expressed in three different ways using different placements of conjugation [equations (5.43)–(5.45)],

$$C_{40} = \mathrm{cum}(r[n], r[n], r[n], r[n]) \tag{5.43}$$

$$C_{41} = \mathrm{cum}(r[n], r[n], r[n], r^{*}[n]) \tag{5.44}$$

$$C_{42} = \mathrm{cum}(r[n], r[n], r*[n], r^{*}[n]) \tag{5.45}$$

where cum(·) is joint cumulant function defined by equation (5.46).

$$\mathrm{cum}(w,x,y,z) = E(wxyz) - E(wx)E(yz) - E(wy)E(xz) - E(wz)E(xy) \tag{5.46}$$

Meanwhile, the estimation of the second and fourth cumulants is achieved by using the following processes as shown in equations (5.47)–(5.51).

**Table 5.1** Decision tree for modulations classification using spectral-based features

| | $C_{20}$ | $C_{21}$ | $C_{40}$ | $C_{41}$ | $C_{42}$ |
|---|---|---|---|---|---|
| 2-PAM | 1.0000 | 1.0000 | −2.0000 | −2.0000 | −2.0000 |
| 4-PAM | 1.0000 | 1.0000 | −1.3600 | −1.3600 | −1.3600 |
| 8-PAM | 1.0000 | 1.0000 | −1.2381 | −1.2381 | −1.2381 |
| BPSK | 1.0000 | 1.0000 | −2.0000 | −2.0000 | −2.0000 |
| QPSK | 0.0000 | 1.0000 | 1.0000 | 0.0000 | −1.0000 |
| 8-PSK | 0.0000 | 1.0000 | 0.0000 | 0.0000 | −1.0000 |
| 4-QAM | 0.0000 | 1.0000 | 1.0000 | 0.0000 | −1.0000 |
| 16-QAM | 0.0000 | 1.0000 | −0.6800 | 0.0000 | −0.6800 |
| 64-QAM | 0.0000 | 1.0000 | −0.6191 | 0.0000 | −0.6191 |

$$\hat{C}_{20} = \frac{1}{N}\sum_{n=1}^{N} r^2[n] \tag{5.47}$$

$$\hat{C}_{21} = \frac{1}{N}\sum_{n=1}^{N} \left|r[n]\right|^2 \tag{5.48}$$

$$\hat{C}_{40} = \frac{1}{N}\sum_{n=1}^{N} r^4[n] - 3\hat{C}_{20} \tag{5.49}$$

$$\hat{C}_{41} = \frac{1}{N}\sum_{n=1}^{N} r^3[n]r*[n] - 3\hat{C}_{20}\hat{C}_{21} \tag{5.50}$$

$$\hat{C}_{42} = \frac{1}{N}\sum_{n=1}^{N} \left|r[n]\right|^4 - \left|\hat{C}_{20}\right|^2 - 2\hat{C}_{21}^2 \tag{5.51}$$

Cumulant values for some noise-free modulation signals are listed in Table 5.1. It is clear from Table 5.1 that different modulations have different cumulant values between each other. Thus the classification of these modulations can be realized. The classification decision making could be achieved with a decision where modulations are divided into subgroups for each cumulant. The decision threshold can be defined using the process suggested in Section 5.2.4.

## 5.5 Cyclostationary Analysis-based Features

Gardner pioneered the area of signal cyclostationary analysis, which exploits the periodic properties of a cyclostationary process (Gardner, 1994). Gardner and Spooner first implemented cyclostationary analysis for modulation classification problems (Gardner and Spooner, 1988), exploiting the distinct differences between the cyclic

spectrum patterns of different modulations. The implementation is well summarized in Ramkumar's tutorial on cyclic features detection (Ramkumar, 2009).

Give a sinusoidal signal $x(t)$, it exhibits cyclostationary properties or contains second-order periodicity if the cyclic autocorrelation shown in equation (5.52) exists with

$$R_x^{\alpha}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x\left(t+\frac{\tau}{2}\right) x\left(t-\frac{\tau}{2}\right) e^{-i2\pi\alpha t} \tag{5.52}$$

frequency $\alpha \neq 0$ and is not identically zero as a function of $\tau$. The Fourier transform of the cyclic autocorrelation is given by equation (5.53),

$$S_x^{\alpha}(f) \triangleq \int_{-\infty}^{\infty} R_x^{\alpha}(\tau) e^{-i2\pi f t} d\tau \tag{5.53}$$

known as the cyclic spectrum. The cyclic spectrum can be interpreted as a spectral correlation function (SCF) as defined in equation (5.54),

$$S_x^{\alpha}(f) \triangleq \lim_{T\to\infty} \lim_{\Delta t\to\infty} \frac{1}{T\Delta t} \int_{-\Delta t/2}^{\Delta t/2} X_T(t, f+\alpha/2) X_T^*(t, f-\alpha/2) dt \tag{5.54}$$

where equation (5.55) holds.

$$X_T(t,v) \triangleq \int_{t-T/2}^{t+T/2} x(u) e^{-j2\pi v u} du \tag{5.55}$$

Gardner demonstrated that the theoretical SCF plane of different modulation signals over a domain of $\alpha$ and $f$ has distinctive differences and can be used for the classification of signal modulations (Gardner, 1994). The correlation coefficient, known as spectral coherence (SC), of the SCF is defined as shown in equation (5.56).

$$C_x^{\alpha}(f) \triangleq \frac{S_x^{\alpha}(f)}{\sqrt{S_x^{\alpha}(f+\alpha/2) S_x^{\alpha}(f-\alpha/2)}} \tag{5.56}$$

To reduce the dimension of the cyclic spectrum, it is possible to construct a cyclic domain profile (CDP) by finding the maximum of SC over frequency $f$ [equation (5.57)] (Kim *et al.*, 2007).

$$I(\alpha) = \max_f \left(C_x^{\alpha}(f)\right) \tag{5.57}$$

The CDP preserves the distinction of cyclic properties between different modulation signals. It can be used for modulation classification.

While the SCF and CDP can be used for the classification of some lower-order modulations, they are not suitable for higher-order modulations. For this reason, Spooner proposed the use of cyclic cumulants for the classification of high-order modulations (Spooner, 1996). The cyclic cumulants of higher order were suggested by Dobre, Bar-Ness and Su (2003).

The $n$th moments of $x(t)$ are defined as shown in equation (5.58),

$$R_x(t,\tau;n,m)\triangleq E\prod_{j=1}^{n}x^{(*)j}\left(t+\tau_j\right) \tag{5.58}$$

where $n$ is the order, $\tau=[\tau_1 \ldots \tau_n]$ is the delay vector, $(*)j$ is an optional conjugate, and $m$ is the number of conjugate factors. The $n$th-order cumulants are defined as shown in equation (5.59),

$$C_x(t,\tau;n,m)\triangleq\sum_{P_n}\left[k(p)\prod_{j=1}^{p}R_x\left(t,\tau_{v_j};n_j,m_j\right)\right] \tag{5.59}$$

where $P_n$ is the set of partitions of $\{1,2,\ldots,n\}$, $n_j$ is the size of the set $v_j$, $m_j$ is the number of conjugated terms in the term $\left\{x^{(*)k}(t+\tau_k)\right\}_{k\in v_j}$, and $k(p)$ is the number $(-1)^{p-1}(p-1)!$. The cumulant function is also representable by a generalized Fourier series as given in equation (5.60),

$$C_x(t,\tau;n,m)=\sum_{\beta}C_x^{\beta}(\tau;n,m)e^{i2\pi\beta t} \tag{5.60}$$

where $\beta$ is called a pure $n$th-order cycle frequency, and $C_x^{\beta}(\tau;n,m)$ is the $n$th-order temporal cumulant function or cyclic cumulant.

For digital modulations such as QAM modulations, the cyclic cumulants can be defined as shown in equation (5.61),

$$C_s^{\beta}(\tau_n;k,m)=\frac{C_a(k,m)}{T_0}\int_{-\infty}^{\infty}\prod_{i=1}^{k}p^{(*)i}(t+\tau_i)e^{-j2\pi\beta t}dt\times e^{j2\pi\beta t_0}e^{j(k-2m)\phi_0}\prod_{i=1}^{k}\left[e^{j2\pi f_c\tau_n}\right]^{(*)n} \tag{5.61}$$

where $C_a(k, m)$ is the $k$th-order cumulant with $m$ conjugates. The final feature used for modulation classification is the maximum value of the function $C_x^{\beta}(\tau;n,m)$, given a set of different $n$, $m$, and $\beta$ [equation (5.62)].

$$F_s=\left\{\max\left|C_s^{\beta}(\tau_n;n,m)\right|\right\}_{\{n,m,\beta\}} \tag{5.62}$$

## 5.6 Conclusion

In this chapter, we list a collection of signal features that can be used for modulation classifiers. The spectral features are well established and suitable for the classification of both analogue and digital modulations. The high-order statistics features including moments, cumulants and cyclic cumulants focus on the classifier of high-order digital modulations. The actual classification process of a feature-based classifier requires the construction of a decision tree where the modulations must be divided into several tiers of subgroups until each modulation could be differentiated from the rest. The easiest decision-making process for a feature-based classifier requires the threshold to be set at each decision-making point on the decision tree. The threshold could be calculated using a theoretical noise-free signal or optimized with given channel conditions.

## References

Azzouz, E.E. and Nandi, A.K. (1995) Automatic identification of digital modulation types. *Signal Processing*, **47** (1), 55–69.

Azzouz, E.E. and Nandi, A.K. (1996a) *Automatic Modulation Recognition of Communication Signals*, Kluwer, Boston.

Azzouz, E.E. and Nandi, A.K. (1996b) Procedure for automatic recognition of analogue and digital modulations. *IEE Proceedings – Communications*, **143** (5), 259–266.

Chan, Y.T. and Gadbois, L.G. (1989) Identification of the modulation type of a signal. *Signal Processing*, **1** (4), 838–841.

Dobre, O.A., Bar-Ness, Y. and Su, W. (2003) Higher-Order Cyclic Cumulants for High Order Modulation Classification. Military Communications Conference, 13–16 October 2003, IEEE, pp. 112–117.

Fabrizi, P.M., Lopes, L.B. and Lockhart, G.B. (1986) Receiver Recognition of Analogue Modulation Types. Proceedings of the IERE Conference on Radio Receiver and Associated Systems, IERE, pp. 135–140.

Gardner, W.A. (1994) *Cyclostationarity in Communications and Signal Processing*, IEEE Press, New York.

Gardner, W.A. and Spooner, C.M. (1988) Cyclic spectral analysis for signal detection and modulation recognition. Military Communications Conference, 1988, MILCOM, pp. 419–424. Conference record. 21st Century Military Communications - What's Possible? 1988 IEEE, vol. 2, pp. 419, 424, 23–26 Oct. 1988. doi: 10.1109/MILCOM.1988.13425

Hassan, K., Dayoub, I., Hamouda, W. and Berbineau, M. (2010) Automatic modulation recognition using wavelet transform and neural networks in wireless systems. *EURASIP Journal on Advances in Signal Processing*, **2010**, 1–13.

Hero, A.O. and Hadinejad-Mahram, H. (1998) Digital modulation classification using power moment matrices. Proceedings of the 1998 IEEE International Conference on Acoustics, Speech, and Signal Processing, 12–15 May 1998, vol. 6, pp. 3285–3288.

Hipp, J.E. (1986) Modulation Classification Based on Statistical Moments. Military Communications Conference, - Communications-Computers: Teamed for the 90's, 1986. MILCOM 1986, 5–9 October 1986, IEEE, vol. 2, pp. 20.2.1–20.2.6 doi: 10.1109/MILCOM.1986.4805739.

Hong, L. and Ho, K.C. (2000) BPSK and QPSK Modulation classification with unknown signal level. Military Communications Conference Proceedings, vol. 2, pp. 976–980. doi: 10.1109/MILCOM.2000.904076.

Jovanovic, S.D., Doroslovacki, M.I. and Dragosevic, M.V. (1990) Recognition of Low Modulation Index AM Signals in Additive Gaussian Noise. Proceedings of the European Association for Signal Processing V Conference, EURASIP, pp. 1923–1926.

Kim, K., Akbar, I.A., Bae, K.K. *et al.* (2007) Cyclostationary Approaches to Signal Detection and Classification in Cognitive Radio. IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, pp. 212–215.

Nandi, A.K. and Azzouz, E.E. (1995) Automatic analogue modulation recognition. *Signal Processing*, **46**, 211–222.

Ramkumar, B. (2009) Automatic modulation classification for cognitive radios using cyclic feature detection. *IEEE Circuits and System Magazine*, **9** (2), 27–45.

Soliman, S.S. and Hsue, S.-Z. (1992) Signal classification using statistical moments. *IEEE Transactions on Communications*, **40** (5), 908–916.

Spooner, C.M. (1996) Classification of co-channel communication signals using cyclic cumulants. 1995 Conference Record of the Twenty-Ninth Asilomar Conference on Signals, Systems and Computers, 30 October–1 November 1995, vol. 1, pp. 531–536. doi: 10.1109/ACSSC.1995.540605.

Swami, A. and Sadler, B. M. (2000) Hierarchical Digital Modulation Classification Using Cumulants. *IEEE Transactions on Communications*, **48** (3), 416–429.

# 6

# Machine Learning for Modulation Classification

## 6.1 Introduction

In Chapter 5 we list a collection of signal features for modulation classification. Some of the classification decision making is based on multi-stage decision trees where each stage utilizes a different feature. However, the need for designing the decision tree and optimization of multiple decision thresholds is not most convenient.

To overcome these problems, various machine learning techniques have been employed to accomplish two major tasks in feature-based modulation classification. First, the machine learning techniques can provide a classification decision making process that is much easier to implement. Second, the machine learning techniques can reduce the dimension of the feature set. This is achieved by feature selection and feature generation, which enables the consideration of a more versatile feature set while maintaining the computational efficiency of the classifier.

In this chapter we first introduce two machine learning-based classifiers, namely k-nearest neighbour classifier and support vector machine classifier, for modulation classification in combination with the features listed in Chapter 5. Next, the issue of feature space dimension reduction is explored through different algorithms including linear regression, artificial neural network, genetic algorithm and genetic programming.

## 6.2 K-Nearest Neighbour Classifier

The k-nearest neighbour (KNN) classifier is a non-parametric algorithm which assigns the class to a testing signal by analyzing the number $k$ of nearest reference signals in the

*Automatic Modulation Classification: Principles, Algorithms and Applications*, First Edition. Zhechen Zhu and Asoke K. Nandi.

feature space. It has been used to solve many different classification problems (Guo and Nandi, 2006; Guo, Zhang and Nandi, 2007; Espejo, Ventura and Herrera, 2010). There are three main steps in a KNN classifier.

### 6.2.1 Reference Feature Space

To enable KNN classification, a reference feature space must be established first. The feature space should include $M$ reference values of each feature from each modulation class. The selection of $M$ depends on the problem and is normally optimized heuristically. The motivation for a larger value for $M$ is that the reference feature space provides a more accurate representation of the likely distribution of the testing signal features. On the other hand, a larger $M$-value is likely to impose a higher computational complexity in the later steps of the KNN classifier.

For modulation classification, Zhu *et al.* suggested the use of training data from the same signal source for the generation of reference feature values (Zhu, Aslam and Nandi, 2010). The advantage of this approach is that the training signal shares the same source as the testing signal. Thus the reference feature space is an accurate representation of the feature distribution of the testing signal. Meanwhile, the construction of the reference feature space is really easy as the only step required is to calculate the feature values for the training signals. However, because of the random nature of the training signal, one cannot guarantee the accuracy of the feature space to be high enough.

Synthesized reference values are more controlled over the construction of the reference feature space. Nevertheless, there needs to be a hypothesized feature distribution which is often not reliable.

### 6.2.2 Distance Definition

Since the classifier requires the evaluation of distances between the testing signal and the reference signals, a distance metric must be defined before a search of neighbouring reference signals can be achieved. There are many different metric systems that can be used for distance measurement in a KNN classifier. Euclidean distance is one of the most common distance metrics for KNN classifiers. Given a feature set $\mathbb{F} = \{\mathbb{F}_1, \mathbb{F}_2 \ldots \mathbb{F}_L\}$ with $L$ number of features, the Euclidean distance between the feature sets of signals A and B is calculated as given in equation (6.1).

$$D(\mathbb{F}(\mathrm{A}), \mathbb{F}(\mathrm{B})) = \sqrt{\sum_{l=1}^{L} [\mathbb{F}_l(\mathrm{A}) - \mathbb{F}_l(\mathrm{B})]^2} \tag{6.1}$$

Having established the distance measurement, the classification decision is achieved by finding the $k$ nearest number of reference samples and analyzing the demography with these $k$ number of samples.

### 6.2.3 K-Nearest Neighbour Decision

When the distances between the test signal and all reference signals are evaluated, a number of $k$ reference signals are recorded as the $k$ nearest neighbours. The selection of the value of $k$ should follow these rules.

1. The value should ideally be a prime number, to avoid the case where the $k$ neighbours consist of equal numbers of reference signals from different classes.
2. The value should be less than the total number of reference signals from a signal class.
3. The value of $k$ should be big enough to avoid false classification caused by outliers.

The actual optimization of the value $k$ can be heuristic because it has been shown that the classification does not vary much if the $k$ value is in a reasonable range. The end classification result is achieved by finding the majority of the $k$ nearest neighbours that share the same class. This class will be assigned to the testing signal as the classification result. A pseudo code for the KNN classifier implementation is given below.

**Algorithm 6.1 A KNN classifier pseudo code**

**Input** $M$ reference signals from every candidate modulation $\mathcal{M}(i)$, $i = 1, 2, \ldots, I$, each with a set of extracted feature sets $\mathbb{F}^i(\mathrm{m})$, an observed unknown signal with extracted features set $\mathbb{F}$, and a pre-defined $k$ value.

Step 1: The distance between $\mathbb{F}$ and every reference feature set $\mathbb{F}^i(\mathrm{m})$ is calculated using equation (6.1)
Step 2: The resulting distances $D\left(\mathbb{F}, \mathbb{F}^i(\mathrm{m})\right)$ are sorted in descending order
Step 3: The first $k$ distances are selected
Step 4: The modulation label $i$ for each distance $D\left(\mathbb{F}, \mathbb{F}^i(\mathrm{m})\right)$ is extracted
Step 5: The mode of the set extracted label set $i'$ is used to identify the modulation
Step 6: Modulation $\mathcal{M}(i')$ is returned as the classification decision $\hat{\mathcal{M}}$

**Output** $\hat{\mathcal{M}}$

The KNN is non-parametric and capable of multi-class classification. However it suffers with an increasing number of features, which raises the dimension of the feature space and the complexity of the distance calculation. Therefore, some sort of dimension reduction is needed to make this method viable. Another disadvantage of the KNN classifier is that the features contribution to the classification decision making is not weighted. Therefore, one feature with relatively sparse distribution between different modulations may come to dominate the distance evaluation. The classification of some modulations relying on other features may be affected.

## 6.3 Support Vector Machine Classifier

Support vector machine (SVM) provides another way to achieve classification in the existing multi-dimensional feature space. It has been adopted for the classification of many different data sets (Mustafa and Doroslovacki, 2004; Polat and Güneş, 2007; Akay, 2009). SVM achieves classification by finding the hyperplane that separates data from different classes. The hyperplane, meanwhile, is optimized by maximizing its distance to the signal samples on each side of the hyperplane. Depending on the nature of the signal being classified, the SVM classifiers can be divided into linear and non-linear versions.

The linear SVM classifiers have linear kernels. The kernel is defined by equation (6.2),

$$K(\boldsymbol{x},\boldsymbol{w}) = \boldsymbol{x}^T\boldsymbol{w} \tag{6.2}$$

where $\boldsymbol{x} = [x_1 \ldots x_K]$ is the input feature vector $\mathbb{F} = \{\mathbb{F}_1, \mathbb{F}_2 \ldots \mathbb{F}_K\}$ and $\boldsymbol{w} = [w_1 \ldots w_K]$ is the weight vector to be optimized. The kernel defines a linear separation hyperplane (Theodoridis, 2008), as given in equation (6.3),

$$g(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{w} + w_0 \tag{6.3}$$

where $w_0$ is a constant. The classification of a two-class (between modulation candidates $\mathcal{M}$(a) and $\mathcal{M}$(b)) problem is achieved by simply using the sign of $g(x)$, as shown in equation (6.4).

$$\hat{\mathcal{M}} = \begin{cases} \mathcal{M}(\mathrm{a}), g(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{w} + w_0 \geq 0 \\ \mathcal{M}(\mathrm{b}), g(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{w} + w_0 < 0 \end{cases} \tag{6.4}$$

To obtain the weight through training, the following optimization process [equation (6.5)] is exercised:

$$\text{maximize } J(w, w_0) = \frac{2}{\|\boldsymbol{w}\|^2} \tag{6.5}$$

$$\text{subject to } y_i\left(w^T x_i + w_0\right) \geq 1, \quad i = 1, 2, \ldots, N \tag{6.6}$$

where $y_i$ is the class indicator for the $i$th feature vector (+1 for $\mathcal{M}$(a) and −1 for $\mathcal{M}$(b)). An illustration of an SVM for a two-class problem is given in Figure 6.1.

The non-linear version of SVM classifier shares the same training and classification process, except that the kernel used for the hyperplane is replaced by a non-linear

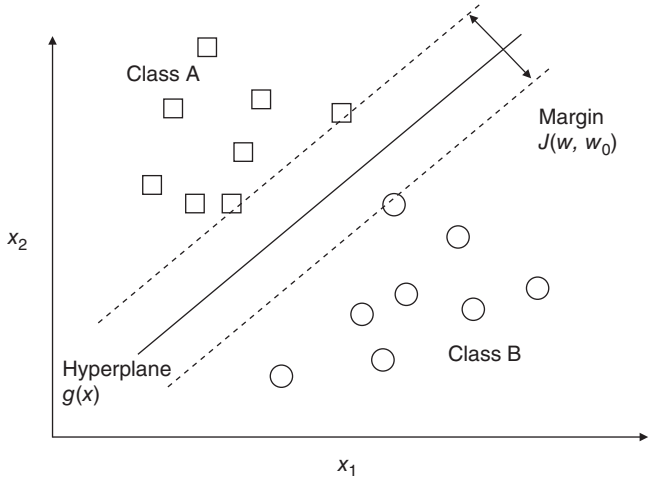



**Figure 6.1** Two-class feature space with linear support vector machine.

kernel. We have tested in the past that a polynomial kernel is enough to provide effective classification. The polynomial kernel is given by equation (6.7),

$$K(\boldsymbol{x},\boldsymbol{w}) = \left(\boldsymbol{x}^T\boldsymbol{w}\right)^d \tag{6.7}$$

where $d$ is the degree of the polynomials.

A general procedure of the SVM classifier for AMC is given below.

**Algorithm 6.2 An SVM classifier pseudo code**

**Input** $M$ reference signals from two candidate modulations $\mathcal{M}(i)$, $i = 1, 2$ each with a set of extracted feature sets $\mathbb{F}^i(\mathrm{m})$, an observed unknown signal with extracted feature set $\mathbb{F}$, and a pre-defined value $d$ if using a non-linear SVM classifier.

Step 1: **initialization** weights $\boldsymbol{w}$ and $w_0$
Step 2: **repeat**
Step 3: update the weights through (6.5) and (6.6) using $\mathbb{F}^i(\mathrm{m})$
Step 4: **end if** maximum number iteration reached
Step 5: $K(\mathbb{F},\boldsymbol{w}) + w_0$ is calculated
Step 6: **if** $K(\mathbb{F},\boldsymbol{w}) + w_0 \geq 0$, $\mathcal{M}(1)$ is given as classification decision $\hat{\mathcal{M}}$
Step 7: **if** $K(\mathbb{F},\boldsymbol{w}) + w_0 < 0$, $\mathcal{M}(2)$ is given as classification decision $\hat{\mathcal{M}}$

**Output** $\hat{\mathcal{M}}$

Compared with the KNN classifier, the SVM classifier only needs to use the training signal when establishing the separating hyperplane. Once the hyperplane is optimized, there is no need to involve the training signal in any sort of further calculation. The benefit is that the computation needed at the testing stage is relatively inexpensive compared with KNN. However, the SVM classifier is most natural for two-class classification. There are implementations of a multi-class classification using SVM; however, the implementation is much less intuitive than in the two-class case.

## 6.4 Logistic Regression for Feature Combination

For both KNN and SVM classifiers, it is always preferable to have as many features as possible for improving the classification accuracy. However, both classifiers suffer when the number of features increases. That is why reducing the feature space dimension is necessary. Using machine learning algorithms, there are two ways to do so. First, feature space dimension can be reduced by eliminating some of the features which make less or no contribution to the classification task. Second, feature space dimension can be reduced by combining the existing feature into fewer new features.

While feature selection is an effective way to reduce the complexity for a feature-based modulation classifier, the elimination of a feature can sometimes be destructive. That is without mentioning that sometimes the features are all useful in some degree and the elimination of any feature can be destructive for the classification performance. In this case, a more conservative approach is needed for dimension reduction. That is why feature combination has been considered for not just the reduction of feature dimension but also for enhancing the performance of these features.

To begin with, a linear combination of the features is the simplest but often a very effective way of the combining the features. Assuming we are combining $k$ number of existing features into a single new feature; the linear combination is given by equation (6.8),

$$\mathbb{F}_{new} = w_0 + \sum_{k=1}^{K} w_k \mathbb{F}_k \tag{6.8}$$

where $w_k$ is the weight of the $k$th feature $\mathbb{F}_k$, $w_0$ is a constant, and $K$ is the total number of features available for combination. The process to optimize these weights is called logistic regression and it aims to maximize the difference of the new feature value between different classes. It has been adopted by Zhu *et al.* in the dimension reduction for distribution-based features (Zhu, Nandi and Aslam, 2013).

There are two common logistic regression tools in the family of generalized linear regression algorithms, namely binomial logistic regression and multinomial logistic regression. The binominal logistic regression is designed to project the signal using a logistic function $p(\cdot)$ such that $p(\cdot)$ equals 1 when the signal is modulated using

$\mathcal{M}(i)$ and 0 if the signal modulation is using $\mathcal{M}(j)$. The logistic function is given in equation (6.9),

$$p(\mathbb{F}) = \frac{1}{1 + e^{-g(\mathbb{F})}} \tag{6.9}$$

where $\mathbb{F}$ is the collection of existing features and $g(\cdot)$ is the logit function, the inverse of the logistic function $p(\cdot)$, given by equation (6.10).

$$g(\mathbb{F}) = B_0 + \sum_{k=1}^{K} B_k \mathbb{F}_k \tag{6.10}$$

The estimation of each of the parameters $B_0$ and $B_k$ is often achieved using iterative processes such as the Newton–Raphson method (Hosmer and Lemeshow, 2000). The resulting estimation can be used to substitute the weights in equation (6.8).

Logistic regression provides a basic tool for feature selection and combination. However, multi-class classification is not always suited for linear regression-assisted feature selection and combination. It is sometimes better to divide the classification into multiple steps.

## 6.5 Artificial Neural Network for Feature Combination

Nandi and Azzouz pioneered in the field of machine learning for modulation classification by introducing the artificial neural network (ANN) for improved decision making (Nandi and Azzouz, 1997). They first proposed to use the ANN algorithm in conjunction with signal spectral-based features for the classification of analogue and digital modulations. It was afterwards adopted by many other researchers as a tool for feature selection and combination.

Unlike the decision tree given in Section 5.2, an ANN classifier does not separate the decision making into multiple stages. Instead, it can be used to consolidate the existing features and create a non-linear mapping of these features to afford new features of reduced dimensionality and enhanced performance.

For a single-layer perceptron network, the trained network is similar to a linear combination of the input features. The same representation can be given as equation (6.11).

$$\mathbb{F}_{\text{out}} = w_0 + \sum_{k=1}^{K} w_k \mathbb{F}_{\text{in}}(k) \tag{6.11}$$

Multi-layer perceptron (MLP) is one of the most popular mapping schemes because of its simplicity and efficient hardware implementation. MLP is a feed-forward structure

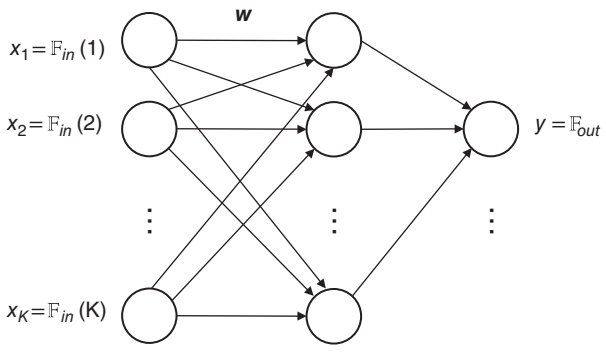



**Figure 6.2** Multilayer perceptron neural network for AMC feature combination.

of interconnection of individual non-linear parallel computing units called neurons. Inputs are propagated through the network layer by layer and MLP gives a non-linear mapping of the inputs at the output layers. Use for feature combination the MLP can be expressed as equation (6.12),

$$y_k = \phi\left(\sum_{i=1}^{q} w_{ki}\phi\left(\sum_{j=1}^{p} w_{ij}x_j\right)\right) \tag{6.12}$$

Where $y_k$ is the output of the MPL network and the feature combination $\mathbb{F}_{\text{out}}(k)$ being optimized at the $k$th output node, $\phi$ is the activation function, $w_{ij}$ is the weight value from neuron $j$ to neuron $i$, and $x_j$ is the $j$th input feature $\mathbb{F}_{\text{in}}(j)$ from the feature set. A visual illustration of the MPL network for feature combination is given in Figure 6.2.

Azzouz and Nandi (1995) adopted the back propagation (BP) to training the weights. The BP algorithm trains the weight through an iterative process by calculating the change of each weight with respect to the error function $E$. As an example, the mean squared error (MSE) function gives equation (6.13),

$$\frac{\delta E}{\delta w_{ij}} = \frac{\delta E}{\delta y_i}\frac{\delta y_i}{\delta u_i}\frac{\delta u_i}{\delta w_{ij}} \tag{6.13}$$

where $y_i$ is the output and $u_i$ is the weighted sum of the inputs of neuron $i$. The weight value can then be updated using a gradient descent approach as shown in equation (6.14),

$$w_{ij}(t+1) = w_{ij}(t) - \varepsilon\frac{\delta E}{\delta w_{ij}} \tag{6.14}$$

where $\varepsilon$ is the learning rate which dictates the speed of convergence. With a high learning rate, the convergence is faster. However, it is done with the risk of oscillation. With a low learning rate, many more iterations will be needed to reach convergence.

## 6.6 Genetic Algorithm for Feature Selection

To overcome the issue of high dimensionality in the feature space, Wong and Nandi suggested the use of a genetic algorithm (GA) as a tool for reducing the number of features (Wong and Nandi, 2004). They used binary strings to represent the selection of different features. If there are five existing features, a binary string example could be 11000, which means that the first two features are selected for classification and the last three features are neglected.

The training of such binary strings begins with a randomly generated string. According to the initial binary string, features are selected for modulation classification with some training data. The resulting classification performance achieved by these selected features is then used as a criterion for evaluating the performance of the binary string. Based on their performance, better binary strings are selected for the evolutionary process of producing new binary strings that migrate toward the optimal solution or optimal selection of features. The two genetic operators used are crossover and mutation.

For crossover, we assume there are two parent binary strings, 11011 and 01000. The crossover would randomly choose equal numbers of bits in both parents and swap their values. In the given case, if the first four digits are selected then the 'children' of the crossover operation would be 01001 and 11010, and these represent two new sets of selected features (Figure 6.3).

Meanwhile, mutation utilizes only one parent, for example, 11011. The operation is the process of selecting random digits in the parent string and generating a random value for that digit. Using the example, if the mutation operation selects the first, third, and fourth digits of the binary string, the resulting child string would become 01111. Since the new value is randomly generated it could be the same as the parent value as seen in the fourth digit or different as seen in the first and third digit (Figure 6.4).

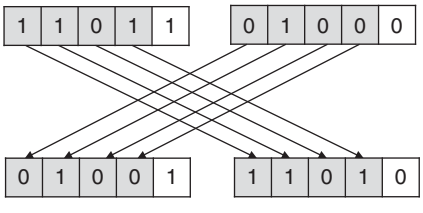



**Figure 6.3** Crossover in genetic algorithm.

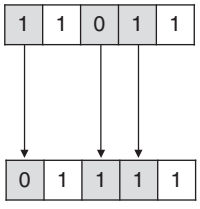



**Figure 6.4** Mutation in genetic algorithm.

The processes of fitness evaluation, parent selection, and reproduction are repeated for a predefined number of generations, after which the GA is terminated. Termination can also be triggered if the average or best fitness in the current generation reaches a predefined threshold or the improvement over the last few generations becomes lower than a predefined threshold.

In the end, the binary strings in all generations are ranked by their fitness. The string with the highest fitness is selected as the final product of the GA process. According to the nature of the binary string, the features can be selected subsequently. It is worth mentioning that the GA process can be highly random because of the random initialization and mutation operation. It is sometimes recommended to repeat the GA process several times and to produce a few sets of different feature selections from which the best feature selection can a determined by another test.

## 6.7 Genetic Programming for Feature Selection and Combination

Koza popularized the genetic programming (GP) as another evolutionary machine learning algorithm (Koza, 1992). It has since been used for classification of many different types of data and signal (Espejo, Ventura and Herrera, 2010). Zhu *et al.* first employed GP for modulation classification feature selection and combination (Zhu, Aslam and Nandi, 2010). Zhu *et al.* also extended the application of GP in modulation classification by combining GP with other machine learning algorithms to achieve improved classification performance (Zhu, Aslam and Nandi, 2011; Aslam, Zhu and Nandi, 2012).

GP belongs to the class of evolutionary algorithms which attempt to emulate a Darwinian model of natural evolution. It is a machine learning methodology that is used to optimize a population of individuals (computer programs) with the help of fitness values. GP develops the solution of a problem in the form of a mathematical formula. Each solution is a computer program and can be represented in the form of a tree. Each tree has terminal nodes (data nodes) and internal nodes (function nodes). Each individual is given a fitness value which quantifies its ability to solve the given problem. The fitness value is computed using a user-defined fitness function.

The fitness function used depends upon the nature of the problem. The advantages GP has on other machine learning methods are listed below. (i) No prior knowledge about the statistical distribution of data is needed. (ii) Preprocessing of data is not required and data can be used directly by GP in its original form. (iii) GP returns a mathematical function as output which can be used directly in the application environment. (iv) GP has the inherent capability to select useful features and to ignore others. Typically, GP implementation follows the following steps: (i) GP starts with a randomly generated population of user-defined size. (ii) Each individual is assigned a fitness value which represents the strength of the individual to solve the given problem. (iii) A genetic operator is applied on the current generation to give birth to individuals of the next generation. Genetic operators are explained in the next section. (iv) All the individuals are given fitness values and those individuals having better fitness values get transferred to the next generation. (v) Steps (iii) and (iv) are repeated till a desired solution is achieved. Otherwise GP is terminated after a certain number of generations set by the user.

### 6.7.1 Tree-structured Solution

There are different ways to represent the individuals (computer programs) in GP. One of the common representations is a tree representation and the same representation has been used here as well. A tree has terminal nodes, internal nodes and output node. Terminal nodes represent the inputs, and internal nodes represent the functions operating on inputs, while the output node gives the output of the tree. An example of a tree-structured mathematical formula $(A + B) \times C$ is given in Figure 6.5. In the case of modulation classification feature selection and combination, the input nodes are the selected raw feature. The output node represents the desired new feature combination.

### 6.7.2 Genetic Operators

Genetic operators are used for reproducing new individuals from older individuals. The operation mimics the genetic processes observed in genetic science. The traditional operators included in a standard GP are crossover and mutation. Semantically, crossover is intended for the sharing of fitter parts of two different individuals in order

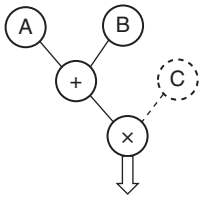



**Figure 6.5** Tree-structured individual of genetic programming.

to create a new individual which is fitter than both parents. Meanwhile, mutation generates new individuals by replacing a random branch of a parent with a randomly generated new branch in the hope that the child will have better fitness than the parent. Practically, the sematic motive of crossover and mutation is implemented with random symbolic processes. We shall use a simple example to illustrated how crossover and mutation is achieved in standard GP.

Let us assume that there are two parent trees, each representing a mathematical formula as shown in Figure 6.6.

The first step of crossover randomly selects a branch in each parent tree. The selected branch is highlighted in Figure 6.6 with dashed lines. In the second step, the selected branches are swapped between the two parents, thereby creating two new individuals as shown in Figure 6.7.

For mutation, we need select only one tree, as shown in Figure 6.8.

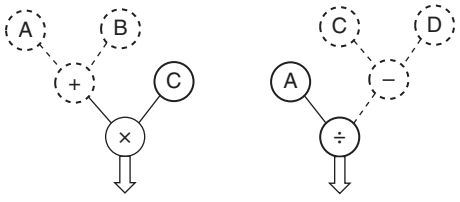



**Figure 6.6** Parent trees selected for crossover operation.

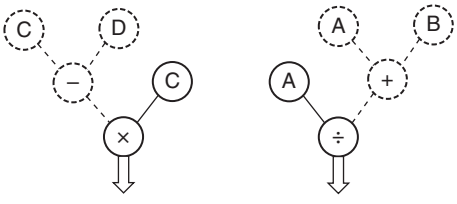



**Figure 6.7** Child trees produced by crossover from trees in Figure 6.6.

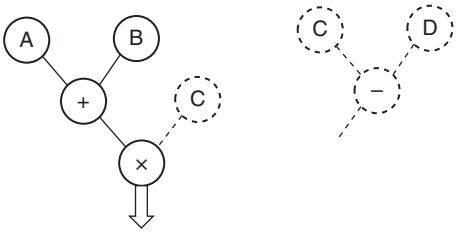



**Figure 6.8** Parent tree selected for mutation and a randomly generated branch.

The first step of the mutation selects a random branch from the parent tree. In the second step, a new branch is randomly generated. Finally, the mutation is completed by attaching the randomly generated new branch to the same position where the old branch was removed from. The resulting tree is the child tree of a mutation operation.

### 6.7.3 Fitness Evaluation

Fitness evaluation is the most important design component because it is directly related to the evaluation of how well an individual in the evolution solves the given problem. If a flawed fitness criterion is used, regardless of how efficient the GP is, the end solution will deviate from the goal of the entire system.

For modulation classification, as we dedicated GP as a feature selector and generator, the goal is to generate a combination of selected features which provides fast and accurate modulation classification. Because of the nature of the task, there have been two different approaches to define the fitness function. The first approach was to evaluate the quality of the new feature by measuring the inter-class tightness and intra-class separation given some training signals. To achieve such evaluation, Aslam *et al.* proposed to use Fisher's criterion as the fitness function for GP (Aslam, Zhu and Nandi, 2011). Assuming there are $L$ number of signal realizations from two different modulations A and B, a new feature acquired through GP can be calculated for each signal realization. Therefore, we have two sets of feature values, $\mathbb{F}_A(1),\mathbb{F}_A(2)\ldots\mathbb{F}_A(L)$ and $\mathbb{F}_B(1),\mathbb{F}_B(2)\ldots\mathbb{F}_B(L)$. To calculate the fitness of this new feature, the following fitness function [equation (6.15) is employed based on Fisher's criterion (Fisher, 1936),

$$\mathcal{F}(\mathbb{F}) = \frac{|\mu_A - \mu_B|}{\sqrt{\sigma_A^2 + \sigma_B^2}} \tag{6.15}$$

where $\mu_A$ and $\mu_B$ are the means of the two sets of the feature values, and $\sigma_A^2$ and $\sigma_B^2$ are the corresponding variances, as defined in equations (6.16) and (6.17).

$$\mu_A = \frac{1}{L}\sum_{l=1}^{L}\mathbb{F}_A(l) \quad \text{and} \quad \mu_B = \frac{1}{L}\sum_{l=1}^{L}\mathbb{F}_B(l) \tag{6.16}$$

$$\sigma_A^2 = \frac{1}{L}\sum_{l=1}^{L}[\mathbb{F}_A(l) - \mu_A]^2 \quad \text{and} \quad \sigma_B^2 = \frac{1}{L}\sum_{l=1}^{L}[\mathbb{F}_B(l) - \mu_B]^2 \tag{6.17}$$

It is obvious in equation (6.15) that the numerator measures the separation of the features from different modulation signals and the denominator measure the tightness of the features from the same modulation signals. Therefore, the fitness function matches the desired property of an effective feature for modulation classification.

However, there are two drawbacks to Fisher's criterion for fitness evaluation. First, the criterion was developed with the assumption of the statistic being normally distributed. In the case of GP-generated features, it is very difficult to establish the distribution of a new feature because the features can be a very complicated combination of many existing features. That is without mentioning the need of normality for the new feature distribution which varies dramatically because of the random nature of GP. The genetic operators constantly maintain the diversity in the populations, resulting in new features of diverse distributions. Secondly, in practice, there are cases where the trained new features may converge to have a minimum amount of difference in their mean difference while having very small variance. In other cases the new features can have very big mean differences with the variance also being infinitely big.

Meanwhile, there is a second approach which does not share the flaws of the Fisher's criterion-based fitness evaluation. As the ultimate goal for the new feature is to enhance classification performance, Zhu *et al.* used a small set of training signals in the GP evaluation and incorporated a computationally efficient classifier in the fitness evaluation (Zhu, Aslam and Nandi, 2010). The fitness, in this case, was evaluated by directly classifying the training signals with a classifier from which the average classification accuracy was used as the fitness value.

## 6.8 Conclusion

In this chapter, several machine learning algorithms are presented which accomplish modulation classification and their feature-enhancement approaches are described. KNN and SVM are two supervised learning algorithms which enable classification of modulations by using multiple input features. The KNN classifier is easier to implement for multi-class problems, while the SVM classifier provides a separation hyperplane that is optimized using training data to easily achieve two-class classification. ANN provides the possibility to combine different features into a smaller figure set. GA using binary coding is considered as a feature selector. GP has both the ability to select features as well as the possibility of creating flexible feature combinations.

## References

Akay, M.F. (2009) Support vector machines combined with feature selection for breast cancer diagnosis. *Expert Systems with Applications*, **36** (2), 3240–3247.

Aslam, M.W., Zhu, Z. and Nandi, A.K. (2011) Robust QAM Classification Using Genetic Programming and Fisher Criterion. European Signal Processing Conference (EUSIPCO), 28 August 2011, Eurasip, pp. 995–999.

Aslam, M.W., Zhu, Z. and Nandi, A.K. (2012) Automatic modulation classification using combination of genetic programming and KNN. *IEEE Transactions on Wireless Communications*, **11** (8), 2742–2750.

Azzouz, E.E. and Nandi, A.K. (1995) Automatic identification of digital modulation types. *Signal Processing*, **47** (1), 55–69.

Espejo, P.G., Ventura, S. and Herrera, F. (2010) A survey on the application of genetic programming to classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, **40** (2), 121–144.

Fisher, R.A. (1936) The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, **7** (2), 179–188.

Guo, H. and Nandi, A.K. (2006) Breast cancer diagnosis using genetic programming generated feature. *Pattern Recognition*, **39** (5), 980–987.

Guo, H., Zhang, Q. and Nandi, A.K. (2007) Feature Generation Using Genetic Programming Based on Fisher Criterion. European Signal Processing Conference, 3 September 2007, Eurasip, pp. 1867–1871.

Hosmer, D.W. and Lemeshow, S. (2000) *Applied Logistic Regression*, John Wiley & Sons, Inc., New York.

Koza, J. R. (1992) *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, The MIT Press, Cambridge.

Mustafa, H. and Doroslovacki, M. (2004) Digital Modulation Recognition Using Support Vector Machine Classifier. Asilomar Conference on Signals, Systems and Computers, November 7–10, 2004, IEEE, pp. 2238–2242.

Nandi, A.K. and Azzouz, E.E. (1997) Modulation recognition using artificial neural networks. *Signal Processing*, **56** (2), 165–175.

Polat, K. and Güneş, S. (2007) Breast cancer diagnosis using least square support vector machine. *Digital Signal Processing*, **17** (4), 694–701.

Theodoridis, S. (2008) *Pattern Recognition*, 4th edn, Academic Press, Boston.

Wong, M.L.D. and Nandi, A.K. (2004) Automatic digital modulation recognition using artificial neural network and genetic algorithm. *Signal Processing*, **84** (2), 351–365.

Zhu, Z., Aslam, M.W. and Nandi, A.K. (2010) Augmented Genetic Programming for Automatic Digital Modulation Classification. IEEE International Workshop on Machine Learning for Signal Processing, 29 August 2010, IEEE, pp. 391–396.

Zhu, Z., Aslam, M.W. and Nandi, A.K. (2011) Support Vector Machine Assisted Genetic Programming for MQAM Classification. International Symposium on Signals, Circuits and Systems, 30 June 2011, IEEE, pp. 1–6.

Zhu, Z., Nandi, A.K. and Aslam, M.W. (2013) Robustness Enhancement of Distribution Based Binary Discriminative Features for Modulation Classification. IEEE International Workshop on Machine Learning for Signal Processing, 22 September 2013, IEEE, pp. 1–6.

# 7

# Blind Modulation Classification

## 7.1 Introduction

From Chapters 3–6, we listed modulation classifiers of different types. These classifiers mostly require the prior knowledge of channel state information (Wei and Mendel, 2000; Wang and Wang, 2010). There are few classifiers which have the certain ability to treat one or two channel parameters as unknown (Panagiotou, Anastasopoulos and Polydoros, 2000). Many classifiers may appear to be able to recognize the modulation type without the need for CSI (Azzouz and Nandi, 1996; Spooner, 1996; Swami and Sadler, 2000). In fact, the classification accuracy is often far inferior if the CSI is not utilized for the preparation of reference values or decision thresholds. Dobre *et al.* reviewed some of the semi-blind classifiers and suggested the necessity of a blind modulation classifier (Dobre, Abdi and Bar-Ness, 2005).

The classification of modulation types in a channel with unknown CSI is normally divided into two steps. In the first step, channel estimation is performed. The estimation can either acquire all of the needed channel parameters or partial CSI. When the entire CSI is estimated, any classifier that we have mentioned in the previous chapters can be employed to complete the second step. If the CSI is partially estimated, a classifier which requires the prior knowledge of all channel parameters will not be able to complete the classification. Instead, a semi-blind classification that can complement the partial channel estimation is required to complete the second step of the blind modulation classification (BMC).

In this chapter, we present a few recently published blind modulation classification approaches which operate with unknown CSI. The first relies on maximum likelihood estimation realized in an unsupervised way through expectation maximization.

*Automatic Modulation Classification: Principles, Algorithms and Applications*, First Edition. Zhechen Zhu and Asoke K. Nandi.

The modulation classification is completed using a likelihood-based classifier which utilizes the estimated CSI. The second is a combination of minimum distance centroid estimation and likelihood-based classifier using non-parametric likelihood function.

## 7.2 Expectation Maximization with Likelihood-based Classifier

As stated in Chapter 3, to achieve optimal classification performance, the likelihood-based classifiers such as the ML classifier require perfect knowledge of the CSI. In normal systems where modulations are known to the receiver, the estimation is relatively easy, especially with pilot samples. In the system with AMC, the modulation is unknown to the receiver and thus cannot be used in the estimator. While an estimator for signals with unknown modulation is available (Gao and Tepedelenlioglu, 2005), the estimation accuracy is not high enough to guarantee high classification accuracy. In addition, such estimators are normally limited to modulations of the same type but different orders. When there are more than two types of modulation in the candidate pool such estimators are not applicable for the CSI estimation for likelihood-based classifiers. It is worth noting that the ALRT, GLRT, and HLRT classifiers could be constructed to achieve maximum likelihood estimation of a channel parameter. The estimation is achieved through an exhaustive manner which requires the evaluation of likelihood for each value of channel parameter in a predefined range, which is highly expensive computationally. The matter is made worse when multiple channel parameters need to be estimated jointly. For this reason, iterative processes such as expectation maximization (EM) (Geoffrey and Peel, 2000) have become a more realistic option for channel estimation (Moon, 1996; Tzikas, Likas and Galatsanos, 2008).

### 7.2.1 Expectation Maximization Estimator

For modulation classification, Chavali and Da Silva first proposed to use EM for CSI estimation (Chavali and Da Silva, 2011) and considered the case of non-Gaussian noises (Chavali and Da Silva, 2013). Here, we present the algorithm in a more recognizable AWGN channel. In EM, we assume there are a set of channel parameters $\Theta = \{\gamma_m, \mu_m, \sigma_m\}$. A Gaussian mixture model is often considered as an integral part of an EM estimator for modulation classification. A GMM is a linear mixture of Gaussian distributed components. The PDF of a GMM model is given by equation (7.1),

$$f_{GMM}(x) = \sum_{m=1}^{M} \gamma_m \mathcal{N}(\mu_m, \sigma_m) \tag{7.1}$$

where $\gamma_m$ is the mixture proportion of the $m$th component, and $\mu_m$ and $\sigma_m$ are the corresponding mean and variance.

The initial step of EM is to provide initial values for the channel parameters. There are two most popular ways to do this. The first method generates random values for the channel parameters $\Theta = \{\gamma_m, \mu_m, \sigma_m\}$. It is easy to implement and adds a minimum amount of extra computation. The disadvantage of this random initialization method is that the time for convergence may be longer and the converged final estimate may be a local optimum instead of a global optimal estimate. The other, more advanced, approach to initialization is to use a fast algorithm to provide a rough estimation for some of the parameters. Soltanmohammadi and Naraghi-Pour used K-means clustering for the initialization step (Soltanmohammadi and Naraghi-Pour, 2013). By clustering all the samples into $M$ number of partitions, the proportion of the Gaussian component can be calculated using the size of the partition, that is, the number of samples assigned to the partition. The mean of the component can be calculated using the mean of all the signal samples in each partition. The variance of each component can be calculated in the same way using the partitioned signal samples.

The expectation step, also known as the E step, evaluates the likelihood of each sampling belongs to each of the Gaussian components. A membership is often assigned using the resulting likelihood values. The membership could be a hard membership which assigns the signal sample to an individual Gaussian component exclusively. Meanwhile, the sample can be given soft memberships for all the Gaussian components, where components with higher likelihood are assigned higher membership values. The membership is often called a latent variable $z$. For the signal samples $r[1], r[2] \ldots r[N]$ the likelihood of the $n$th sampling belonging to the $m$th component of the GMM model is calculated as shown in equation (7.2).

$$\mathcal{L}(r[n],m) = \frac{1}{2\pi\sigma_m^2} e^{-\frac{(r[n]-\mu_m)^2}{2\sigma_m^2}} \tag{7.2}$$

And the corresponding soft membership is calculated as shown in equation (7.3)

$$z(n,m) = \frac{L(r[n],m)}{\sum_{m=1}^{M} L(r[n],m)} \tag{7.3}$$

with the hard membership assigned as shown in equations (7.4).

$$z(n,m) = \begin{cases} 1, \ \mathcal{L}(r[n],m) = \max\limits_{1 \le k \le M} (\mathcal{L}(r[n],k)) \\ 0, \ \mathcal{L}(r[n],m) \ne \max\limits_{1 \le k \le M} (\mathcal{L}(r[n],k)) \end{cases} \tag{7.4}$$

The membership of each signal sample having been evaluated, the maximization step aims to maximize the evaluated log-likelihood function of the current iteration, equation (7.5).

$$\log \mathcal{L}(r,k) = -N \log M - \sum_{m=1}^{M} \sum_{n=1}^{N} z(r[n],m)\gamma_m \left[ \log\left(2\pi\sigma_m^2\right) + \frac{(r[n]-\mu_m)^2}{2\sigma_m^2} \right] \tag{7.5}$$

For modulation classification, there is underlying structure of the component means which can be expressed as a combination of channel gain and transmitted symbols, as given in equation (7.6),

$$\mu_m = hs_m \tag{7.6}$$

where $h$ is the channel coefficient and $s_m$ is the $m$th symbol in the modulation alphabet set. In the meantime, the noise variances of each Gaussian component are often considered identical [equation (7.7)].

$$\sigma_m = \sigma \tag{7.7}$$

The maximization step is achieved by the close form update function of the current likelihood evaluation. Combining equations (7.5)–(7.7), the derivative with respect to channel coefficient and noise variance can be calculated as shown in equations (7.8) and (7.9).

$$\frac{\partial \log \mathcal{L}(r,k)}{\partial h} = \sum_{m=1}^{M} \sum_{n=1}^{N} z(r[n],m)\gamma_m \frac{-2r[n]s_m + 2hs_m^2}{\sigma} \tag{7.8}$$

$$\frac{\partial \log \mathcal{L}(r,k)}{\partial \sigma^2} = \sum_{m=1}^{M} \sum_{n=1}^{N} z(r[n],m)\gamma_m \left( -\frac{1}{\sigma^2} + \frac{(r[n]-hs_m)^2}{\sigma^4} \right) \tag{7.9}$$

When equations (7.8) and (7.9) are set to zero, the update function for the channel gain and the variance are given by equations (7.10) and (7.11),

$$h_{i+1} = \frac{\sum_{m=1}^{M} \sum_{n=1}^{N} z(r[n],m)\gamma_m r[n]s_m}{\sum_{m=1}^{M} \sum_{n=1}^{N} z(r[n],m)\gamma_m s_m^2} \tag{7.10}$$

$$\sigma_{i+1} = \frac{\sum_{m=1}^{M} \sum_{n=1}^{N} z(r[n],m)\gamma_m (r[n]-h_i s_m)}{\sum_{m=1}^{M} \sum_{n=1}^{N} z(r[n],m)\gamma_m} \tag{7.11}$$

where $h_{i+1}$ and $\sigma_{i+1}$ are the updated estimation of the parameters for iteration $i+1$. In equation (7.11), the channel coefficient estimated in the previous iteration is used,

which makes this the update function of the expectation/condition maximization (ECM) algorithm. It is often adopted to deal with coupling parameters such as the channel coefficient and the noise variance in this case. ECM shares the convergence property of EM (Meng and Rubin, 1993) and can be constructed to converge at a similar rate as the EM algorithm (Sexton, 2000).

The iterative process is terminated under two conditions. The first condition terminates the process when the estimation reaches convergence. The condition is represented numerically with the difference between the expected likelihoods of the current iteration and the previous iteration along with a pre-defined threshold. In the second condition, termination is triggered when the pre-defined number of iterations has been reached.

### 7.2.2 Maximum Likelihood Classifier

For modulation classification, given the modulation candidate pool $\mathfrak{M} = \{\mathcal{M}(1), \mathcal{M}(2), \ldots, \mathcal{M}(I)\}$, the EM estimation is performed for each modulation hypothesis $\mathcal{H}_{\mathcal{M}(i)}(i = 1, 2, \ldots I)$. As demonstrated in Figure 7.1, the channel parameter set $\hat{\Theta}_{\mathcal{M}(i)} = \left\{\hat{h}_{\mathcal{M}(i)}, \hat{\sigma}_{\mathcal{M}(i)}\right\}$ is estimated for each modulation hypothesis, with $\hat{h}_{\mathcal{M}(i)}$ being the EM estimated complex channel gain and $\hat{\sigma}^2_{\mathcal{M}_i}$ being the estimated signal variance for modulation candidate $\mathcal{M}(i)$. Subsequently the likelihood for each modulation hypothesis is evaluated using the corresponding channel estimation. Here we use the likelihood function from equation (3.3); however, we replace the known channel parameters with EM estimates, to arrive at equation (7.12).

$$\mathcal{L}(r|\mathcal{H}_{\mathcal{M}}) = \mathcal{L}\left(r|\hat{h}_{\mathcal{M}}, \hat{\sigma}^2_{\mathcal{M}}\right) = \prod_{n=1}^{N}\sum_{m=1}^{M}\frac{1}{M}\frac{1}{2\pi\hat{\sigma}^2_{\mathcal{M}}}e^{-\frac{\left|r[n]-\hat{h}_{\mathcal{M}}sm\right|^2}{2\hat{\sigma}^2_{\mathcal{M}}}} \tag{7.12}$$

It is worth noting that the complex channel gain consists of carrier phase offset estimation. When it is applied to the transmitted symbol $s_m$ the effect of constant phase offset

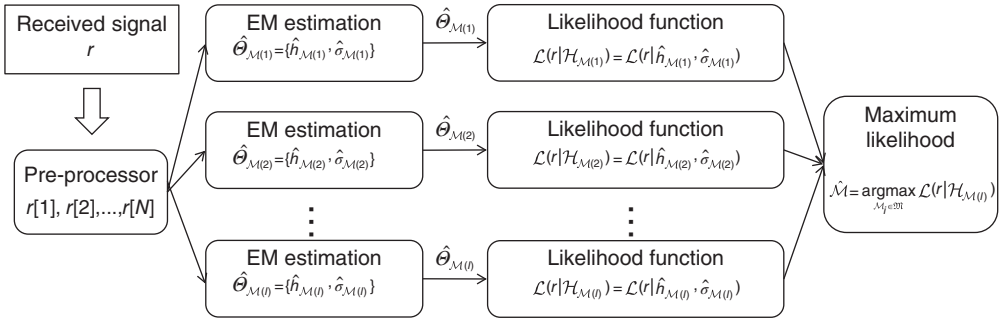



**Figure 7.1** EM estimation and ML classifier.

is effectively compensated. The resulting likelihood values from all modulation hypotheses are then compared in order to conclude the classification decision. Using the maximum likelihood criteria equation (3.13), the modulation hypothesis with the highest likelihood is assigned as the classification decision.

### 7.2.3 Minimum Likelihood Distance Classifier

While the combination of EM estimation and ML classifier seems a perfect match, the combination has some fundamental flaws. The flaw is, in fact, shared with any method that combines a maximum likelihood estimator and a maximum likelihood classifier. As the channel estimation is performed under each modulation hypothesis, the channel estimation maximizes the likelihood evaluation of the modulation hypothesis regardless of whether the hypothesis is true or false. Compared with an ML classifier with given channel knowledge, the likelihood evaluation for the true hypothesis may be accurately achieved. However, because of the ML estimation the likelihood evaluation for false hypothesis is certainly increased, which leads to a reduced separation between the likelihood of the true hypothesis and the false hypotheses. In extreme cases it is even possible for the false modulation hypothesis to provide a higher likelihood value as compared with the true modulation hypothesis.

The above phenomenon is discussed by Soltanmohammadi and Naraghi-Pour, who also suggested a minimum distance likelihood classifier to overcome the issue (Soltanmohammadi and Naraghi-Pour, 2013). Instead of comparing the likelihood value from each modulation hypothesis directly, the distance between the observed likelihood is compared with the empirical likelihood of each modulation with the given channel estimation. The empirical likelihood values can be either computed beforehand and stored in memory or calculated using the channel estimation during the modulation classification processing. For the observed signal, there exist $I$ likelihood values evaluated from different hypotheses denoted as $\{\mathcal{L}(r\,|\,\mathcal{H}_{\mathcal{M}(1)}, \boldsymbol{\Theta}_{\mathcal{M}(1)}), \mathcal{L}(r\,|\,\mathcal{H}_{\mathcal{M}(2)}, \boldsymbol{\Theta}_{\mathcal{M}(2)}),\ldots, \mathcal{L}(r\,|\,\mathcal{H}_{\mathcal{M}(I)}, \boldsymbol{\Theta}_{\mathcal{M}(I)})\}$. Meanwhile, for each hypothesis $\mathcal{H}_{\mathcal{M}(i)}$, there exists a set of empirical likelihood values denoted as $\{\widetilde{\mathcal{L}}(\mathcal{H}_{\mathcal{M}(i)}|\mathcal{H}_{\mathcal{M}(1)}, \boldsymbol{\Theta}_{\mathcal{M}(1)}), \widetilde{\mathcal{L}}(\mathcal{H}_{\mathcal{M}(i)}|\mathcal{H}_{\mathcal{M}(2)}, \boldsymbol{\Theta}_{\mathcal{M}(2)}),\ldots, \widetilde{\mathcal{L}}(\mathcal{H}_{\mathcal{M}(i)}|\mathcal{H}_{\mathcal{M}(I)}, \boldsymbol{\Theta}_{\mathcal{M}(I)})\}$. The distance between the observed likelihood value set and empirical likelihood set from $H_{M(i)}$ is calculated as shown in equation (7.13).

$$\mathcal{D}(r, \mathcal{H}_{\mathcal{M}(i)}) = \sum_{k=1}^{I} \left| \widetilde{\mathcal{L}}(\mathcal{H}_{\mathcal{M}(i)}|\mathcal{H}_{\mathcal{M}(k)}, \boldsymbol{\Theta}_{\mathcal{M}(k)}) - \mathcal{L}(r|\mathcal{H}_{\mathcal{M}(k)}, \boldsymbol{\Theta}_{\mathcal{M}(k)}) \right| \tag{7.13}$$

The subsequent classification decision is then acquired by finding the modulation hypothesis that provides the minimum likelihood distance, equation (7.14).

$$\hat{\mathcal{M}} = \underset{\mathcal{M}(i)}{\arg\min}\, \mathcal{D}(r, \mathcal{H}_{\mathcal{M}(i)}) \tag{7.14}$$

## 7.3 Minimum Distance Centroid Estimation and Non-parametric Likelihood Classifier

While EM estimation provides accurate joint estimation of the channel parameters, the estimation is still based on a known noise model. If unknown noise type with mismatching model is observed, the estimation accuracy is not guaranteed. In addition, being more efficient than an exhaustive estimator, the EM estimator is still relatively complex where multiple instances of likelihood evaluation are required to obtain the estimation. Seeking for alternative blind modulation classification solutions, Zhu and Nandi proposed a scheme to combine joint estimation of channel gain and phase offset with non-parametric likelihood function for BMC (Zhu and Nandi, 2014).

### 7.3.1 Minimum Distance Centroid Estimation

The minimum distance centroid estimator is an iterative process for the joint estimation of channel gain and phase offset. A signal-to-centroids distance metric is proposed to evaluate the mismatch between the estimation being updated and the observed signal. The method is based on the iterative update of a signal centroid collection composed of the original transmitted alphabet, the channel gain and phase shift. We assume the estimated centroids $\mathcal{A}_{\mathcal{M}}$ to possess the original rigid structure after transmission and pre-processing. The mean of centroids $\mu(\mathcal{A}_{\mathcal{M}})$ should remain at 0, the magnitude of two different centroid elements $\mathcal{A}^p_{\mathcal{M}}$ and $\mathcal{A}^q_{\mathcal{M}}$ should follow the original proportions $\|\mathcal{A}^p_{\mathcal{M}}\|/\|\mathcal{A}^q_{\mathcal{M}}\| = \|\boldsymbol{s}^p_{\mathcal{M}}\|/\|\boldsymbol{s}^q_{\mathcal{M}}\|$, and the phase difference between centroids should remain the same, $\phi(\mathcal{A}^p_{\mathcal{M}}) - \phi(\mathcal{A}^q_{\mathcal{M}}) = \phi(\boldsymbol{s}^p_{\mathcal{M}}) - \phi(\boldsymbol{s}^q_{\mathcal{M}})$. For BPSK, QPSK and 8-PSK the centroid collection is given by equations (7.15)–(7.17), respectively,

$$\mathcal{A}_{BPSK} = [-A \quad A] \tag{7.15}$$

$$\mathcal{A}_{QPSK} = \mathcal{A}_{4-QAM} = \begin{bmatrix} jA & A \\ -A & -jA \end{bmatrix} \tag{7.16}$$

$$\mathcal{A}_{8-PSK} = \begin{bmatrix} & -jA & jA^* & \\ -A^* & & & A \\ A & & & A^* \\ & -jA^* & jA & \end{bmatrix} \tag{7.17}$$

where $A$ is defined as the centroid parameter, given by equation (7.18),

$$A = \alpha e^{i\theta_0} s^0_{\mathcal{M}} \tag{7.18}$$

where $s^0_M$ is defined as the transmitted symbol which is nearest to the signal mean.

The signal-to-centroid distance is defined as the sum of the Euclidean distance between each signal sample and its nearest centroid, as defined in equation (7.19).

$$\mathcal{D}_{\mathcal{M}}(r,\mathcal{A}_{\mathcal{M}}) = \sum_{n=1}^{N} \min_{m\in[1,..,M]} \left(\left\|r(n) - \mathcal{A}_{\mathcal{M}}^{m}\right\|\right) \tag{7.19}$$

It is worth noting that the above distance is an equivalent to the evaluation function of minimum distance classifier presented in Chapter 3 (Wong and Nandi, 2008).

With the complex representation of the centroid parameter, $A = x + jy$, the signal-to-centroid distance can be expressed with $x$ and $y$. To find the values of $x$ and $y$ which minimize the signal-to-centroid distance, we use an iterative sub-gradient method based on the sub-gradient calculation shown in equation (7.20).

$$\nabla\mathcal{D}(x,y) = \frac{D(x+\Delta x,y)}{\Delta x} + j\frac{D(x,y+\Delta y)}{\Delta y} \tag{7.20}$$

Meanwhile, the update function for $A_n = x_n + jy_n$ is expressed as given in equation (7.21),

$$x_{n+1} + jy_{n+1} = x_n + jy_n - a\nabla\mathcal{D}(x_n,y_n) \tag{7.21}$$

where $a$ is the update step size. A smaller update step size slows down convergence but provides a more accurate estimation. A bigger update step size helps the estimator converge faster but with the risk of lower estimation accuracy.

The process is repeated for a number of iterations unless the exit condition is triggered by convergence with a given threshold, equation (7.22),

$$\nabla\mathcal{D}(x,y) < \eta_{\mathcal{D}} \tag{7.22}$$

where $\eta_{\mathcal{D}}$ is a pre-defined threshold.

Given the estimated centroid parameter $\hat{A}$, the effective estimation of the channel gain and phase offset can be calculated as shown in equations (7.23) and (7.24), respectively.

$$\hat{\alpha} = \frac{|A|}{|s_m^0|} \tag{7.23}$$

$$\theta_o = \arg(A) - \arg\left(s_m^0\right) \tag{7.24}$$

It is worth noting that the effective estimation of the phase offset is not accuracy. The estimation itself could be offered by a factor of $m\Delta\theta$, where $\Delta\theta$ is the phase difference between adjacent modulation symbols. However, the offset should not affect the modulation classification performance since the likelihood evaluation takes the average of the likelihood from all modulation symbols.

### 7.3.2 Non-parametric Likelihood Function

As demonstrated, the MD estimator is able to find the centroid after transmission; however, the estimation of noise variance is absent. For this reason, the state-of-the-art likelihood classifier is not applicable in this case. To overcome this issue, Zhu and Nandi proposed a non-parametric likelihood function for the evaluation of likelihood without noise variance (Zhu and Nandi, 2014). The NPLF has been suggested in Chapter 3 as an LB approach with reduced complexity. In addition, the NPLF does not impose a hypothesized noise model. The NPLF is defined as given in equation (7.25),

$$L_{NPLF}(r|H_{\mathcal{M}}) = \sum_{n=1}^{N}\sum_{m=1}^{M}\mathbb{I}\left\{\left\|r(n)-\hat{\mathcal{A}}_{\mathcal{M}}^{(m)}\right\| < \mathcal{R}_{\mathcal{M}}\right\} \tag{7.25}$$

where $\mathbb{I}\{\cdot\}$ is an indicator function which returns 1 if the input is true and 0 if input is false, and the test radius $\mathcal{R}_{\mathcal{M}}$ is given by equation (7.26),

$$\mathcal{R}_{\mathcal{M}} = \mathcal{R}_0/\sqrt{M} \tag{7.26}$$

with $\mathcal{R}_0$ being the reference radius. The non-parametric likelihood function is effectively the empirical estimation of the cumulative probability of the given signal in a set of defined local regions. The expectation of the likelihood can be expressed in the following manner [equation (7.27)],

$$E[L_{NPLF}(r|H_{\mathcal{M}})] = \int_{S_{\mathcal{M}}} f(x,y)dS \tag{7.27}$$

where $S_{\mathcal{M}}$ is a limit associated with both estimated centroids $\hat{\mathcal{A}}_{\mathcal{M}}$ and the test radius $\mathcal{R}_{\mathcal{M}}$, and $f(x, y)$ is the probability density function of the testing signal.

It is easy to see that, with the given testing radius, the area of $S_{\mathcal{M}} = M\cdot\pi\mathcal{R}_0^2/M = \pi\mathcal{R}_0^2$ is designed to give each hypothesis an equal area for the cumulative probability calculation. The decision is based on the assumption that matching models should provide maximum cumulative probability in defined regions of the same total area [equation (7.28)].

$$\hat{\mathcal{M}} = \underset{\mathcal{M}\in\mathfrak{M}}{\arg\max}\int_{S_{\mathcal{M}}} f(x,y)dS \tag{7.28}$$

Without examining the centroid estimation for false hypothesis modulations, we evaluate the maximum non-parametric likelihood of different hypotheses in the scenario where each set of estimated centroids has the maximum number of overlaps with the true signal centroids. Such a scenario has been previously examined for the GLRT classifier with unknown channel gain and carrier phase, which results in equal likelihood for nested modulations at high SNR (Panagiotou, Anastasopoulos and Polydoros, 2000).

In the following analysis we will use signals in slow-fading channel with constant phase offset and AWGN noise as an example. The signal PDF is given by equation (7.29).

$$f(x,y)=\frac{1}{M}\sum_{m=1}^{M}\frac{1}{2\pi\sigma^2}e^{-\frac{\left(x-\Re\left(\alpha e^{j\theta}s_{\mathcal{M}}^{m}\right)\right)^2+\left(y-\Im\left(\alpha e^{j\theta}s_{\mathcal{M}}^{m}\right)\right)^2}{2\sigma^2}} \tag{7.29}$$

Approximating the signal distribution at each transmitted signal symbol to a Rayleigh distribution, we arrive at equation (7.30),

$$f(\mathcal{R})=\frac{\mathcal{R}}{\sigma^2}e^{-\mathcal{R}^2/2\sigma^2} \tag{7.30}$$

when the likelihood function estimation becomes a function of the testing radius [equation (7.31)],

$$E[L_{NPLF}(r|H_{\mathcal{M}})]=\mathbb{N}_{\mathcal{M}}\int_{0}^{\mathcal{R}_{\mathcal{M}}}\frac{x}{\sigma^2}e^{-x^2/2\sigma^2}dx \tag{7.31}$$

where $\mathbb{N}_{\mathcal{M}}$ is the maximum number of matching centroids for the hypothesis $\mathcal{M}$. For example, given a piece of QPSK signal, the value of $\mathbb{N}_{\mathcal{M}}$ for different hypotheses would be: $\mathbb{N}_{BPSK}=2$, $\mathbb{N}_{QPSK}=4$ and $\mathbb{N}_{8-PSK}=4$. To simplify the analysis, we generalize the analysis to three general scenarios: hypothesis of lower order $\mathcal{M}^{-}$, hypothesis of matching model and order $\mathcal{M}^{0}$, and hypothesis of higher order $\mathcal{M}^{+}$. In order to satisfy the conditions $E\left[L_{NPLF}\left(r\middle|H_{\mathcal{M}^0}\right)\right]>E[L_{NPLF}(r|H_{\mathcal{M}^-})]$, and $E\left[L_{NPLF}\left(r\middle|H_{\mathcal{M}^0}\right)\right]>E[L_{NPLF}(r|H_{\mathcal{M}^+})]$, the reference radius $\mathcal{R}_0$ should satisfy the condition given in equation (7.32),

$$\mathcal{R}_0=\alpha_{\mathcal{R}}\max_i(\alpha_i) \tag{7.32}$$

where $\alpha_{\mathcal{R}}$ is the radius factor and $\alpha_i$ is the channel gain estimated for modulation $\mathcal{M}(i)$. The determination of the radius factor depends on the range of SNR the classifier

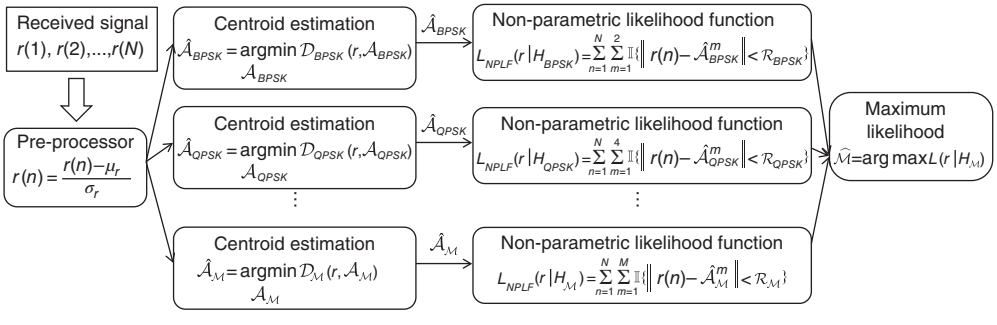



**Figure 7.2** Centroid estimation and NPLF classifier.

operates in, and a detailed analysis is given in Zhu and Nandi (2014). An illustration of the classifier is given in Figure 7.2.

## 7.4 Conclusion

In this chapter we present a few blind modulation classification approaches. The first approach combines EM estimation with likelihood-based classifiers. The ML classifier provides an easy implementation of AMC following the EM channel estimation. The minimum likelihood distance classifier is listed as an alternative approach to the ML classifier, which is able to overcome the performance issues associated with the combination of ML-based estimator and ML-based classifier. The minimum distance centroid estimation provides a quick way to estimate channel gain. It is well complemented by the non-parametric likelihood classifier where knowledge of the noise variance is not required during the classification. The classifier also has the advantage that performance degradation in non-Gaussian noise channel is relatively small.

## References

Azzouz, E.E. and Nandi, A.K. (1996) *Automatic Modulation Recognition of Communication Signals*, Kluwer, Boston.

Chavali, V.G. and Da Silva, C.R.C.M. (2011) Maximum-likelihood classification of digital amplitude-phase modulated signals in flat fading non-Gaussian channels. *IEEE Transactions on Communications*, **59** (8), 2051–2056.

Chavali, V.G. and Da Silva, C.R.C.M. (2013) Classification of digital amplitude-phase modulated signals in time-correlated non-Gaussian channels. *IEEE Transactions on Communications*, **61** (6), 2408–2419.

Dobre, O.A., Abdi, A. and Bar-Ness, Y. (2005) Blind Modulation Classification: A Concept Whose Time Has Come. IEEE Sarnoff Symposium on Advances in Wired and Wireless Communication, 18 May 2005, IEEE, pp. 223–228.

Gao, P. and Tepedelenlioglu, C. (2005) SNR estimation for nonconstant modulus constellations. *IEEE Transactions on Signal Processing*, **53** (3), 865–870.

Geoffrey, M. and Peel, D. (2000) *Finite Mixture Models*, John Wiley & Sons, Inc., New York.

Meng, X.-L. and Rubin, D.B. (1993) Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika*, **80** (2), 267–278.

Moon, T.K. (1996) The expectation-maximization algorithm. *IEEE Signal Processing Magazine*, **13** (6), 47–60.

Panagiotou, P., Anastasopoulos, A. and Polydoros, A. (2000) Likelihood Ratio Tests for Modulation Classification. Military Communications Conference, 22 October 2000, IEEE, pp. 670–674.

Sexton, J. (2000) ECM algorithms that converge at the rate of EM. *Biometrika*, **87** (3), 651–662.

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

Spooner, C.M. (1996) Classification of Co-channel Communication Signals Using Cyclic Cumulants. Conference Record of the Twenty-Ninth Asilomar Conference on Signals, Systems and Computers, 30 October 1995, IEEE, pp. 531–536.

Swami, A. and Sadler, B.M. (2000) Hierarchical digital modulation classification using cumulants. *IEEE Transactions on Communications*, **48** (3), 416–429.

Tzikas, D.G., Likas, A.C. and Galatsanos, N.P. (2008) The variational approximation for bayesian inference. *IEEE Signal Processing Magazine*, **25** (6), 131–146.

Wang, F. and Wang, X. (2010) Fast and robust modulation classification via Kolmogorov–Smirnov test. *IEEE Transactions on Communications*, **58** (8), 2324–2332.

Wei, W. and Mendel, J.M. (2000) Maximum-likelihood classification for digital amplitude-phase modulations. *IEEE Transactions on Communications*, **48** (2), 189–193.

Wong, M.L.D. and Nandi, A.K. (2008) Semi-blind algorithms for automatic classification of digital modulation schemes. *Digital Signal Processing*, **18** (2), 209–227.

Zhu, Z. and Nandi, A.K. (2014) Blind Digital Modulation Classification using Minimum Distance Centroid Estimator and Non-parametric Likelihood Function. *IEEE Transactions on Wireless Communications*, **13** (8), 4483–4494.

# 8

# Comparison of Modulation Classifiers

## 8.1 Introduction

In Chapters 3–7 we listed an array of modulation classifiers. While their mechanisms are distinctly different and intriguing, we are more interested in their actual modulation-classification performance. Modulation classification may be applied in many different scenarios; the traits of a good modulation classifier are shared in most cases.

First, a modulation classifier should be able to classify as many modulation types as possible. Such a trait makes a modulation classifier easily applicable in different applications without needing any modification to accommodate extra modulations. Second, a modulation classifier should provide high classification accuracy. The high classification accuracy is relative to the different noise levels. Third, the modulation classifier should be robust in many different channel conditions. The robustness can be provided by either the built-in channel estimation and correction mechanism or the natural resilience of the modulation classifier against channel conditions. Fourth, the modulation classifier should be computationally efficient. In many applications, there is a strict limitation of computation power which may be unsuitable for over-complicated modulation classifiers. Meanwhile, some applications may require fast decision making, which requires the classification to be evaluated swiftly. Only a modulation classifier with high computational efficiency could meet this requirement. After all, a simple and fast modulation classifier algorithm is always appreciated.

In this chapter, we have tried to benchmark some of the aforementioned modulation classifiers in a simulated testing environment. The goal is to acquire the evaluation of the different traits for different classifiers and to provide a guideline for classifier selection when a specific application arises.

*Automatic Modulation Classification: Principles, Algorithms and Applications*, First Edition. Zhechen Zhu and Asoke K. Nandi.

## 8.2 System Requirements and Applicable Modulations

In this section we examine the system requirements of AMC classifiers and the modulations they are able to classify. The system requirement consists of pilot sample and channel parameters the classifiers needed in order to reach a classification decision. Among them, channel gain, noise variance, carrier phase offset and carrier frequency offset have been picked as the key channel parameters that may be needed by different classifiers.

As the system requirements and applicable modulations have been mentioned in each chapter already, they are only listed again in Tables 8.1–8.5 to provide easy comparison. If the classifier requires a pilot sample for training purposes, it will be labelled "Yes" for "Pilot Samples", otherwise "No" if not. For the channel parameters, there are three possible scenarios: "Known", "Unknown" and "Compensated". When the classifier requires the channel parameter to be known, it means the parameter must be estimated beforehand to complete the classification. When the channel parameter is labelled unknown, it means that the classifier can complete classification without the knowledge of the specific channel parameter. If the channel parameter is labelled compensated, it means that the classifier has the inherent ability to estimate the parameter and use it to improve the classification accuracy. For the applicable modulations, we consider a pool of the most common digital modulations including ASK, PSK, FSK, PAM and QAM. In addition, if the classifier is able to classify the same modulation with different orders, they are instead labelled as M-ASK, M-PSK, M-FSK, M-PAM or M-QAM.

## 8.3 Classification Accuracy with Additive Noise

In this section we examine the classification accuracy of different classifiers against additive noise. In this scenario we assume that we have perfect channel knowledge and that any channel effect has been compensated prior to modulation classification. The signal model used is derived from equation (1.13), such that the only additive noise is considered to be given by equation (8.1),

$$r[n] = \alpha s[n] + \omega[n] \tag{8.1}$$

where the noise model for $\omega[\cdot]$ used is the additive white Gaussian noise. Nine of the most popular modulations are selected to form a candidate pool. They are 2-PAM, 4-PAM, 8-PAM, BPSK, QPSK, 8-PSK, 4-QAM, 16-QAM and 64-QAM. The symbol mappings for all modulations are demonstrated in Figure 8.1.

The alphabet set from the symbol mapping for each modulation is normalized to zero mean and unit power. When simulating the transmitted signal symbols, each symbol was assigned from the alphabet set with equal probability. For each modulation, $L = 1000$ realizations of transmitted signals are first generated. In each signal realization, $N = 1024$ signal samples are sampled at the correct symbol timing without timing mismatch and timing error.

**Table 8.1** System requirements and applicable modulation for LB classifiers

| | Pilot samples | Channel gain | Noise variance | Phase offset | Frequency offset | Modulations |
|---|---|---|---|---|---|---|
| Chapter 3 | | | | | | |
| ML | No | Known | Known | Known | Unknown | M-ASK, M-PSK, M-FSK, M-PAM, M-QAM |
| ML-P | No | Known | Known | Known | Unknown | M-PSK, M-QAM |
| ML-M | No | Known | Known | Unknown | Unknown | M-ASK, M-PAM, M-QAM |
| ALRT | No | Compensated | Compensated | Compensated | Unknown | M-ASK, M-PSK, M-FSK, M-PAM, M-QAM |
| GLRT | No | Compensated | Compensated | Compensated | Unknown | M-ASK, M-PSK, M-FSK, M-PAM, M-QAM |
| HLRT | No | Compensated | Compensated | Compensated | Unknown | M-ASK, M-PSK, M-FSK, M-PAM, M-QAM |

**Table 8.2** System requirements and applicable modulation for distribution test-based classifiers

| | Pilot samples | Channel gain | Noise variance | Phase offset | Frequency offset | Modulations |
|---|---|---|---|---|---|---|
| Chapter 4 | | | | | | |
| KS-one | No | Known | Known | Unknown | Unknown | M-ASK, M-PSK, M-FSK, M-PAM, M-QAM |
| KS-two | Yes | Known | Known | Unknown | Unknown | M-ASK, M-PSK, M-FSK, M-PAM, M-QAM |
| AD | No | Known | Known | Unknown | Unknown | M-ASK, M-PSK, M-FSK, M-PAM, M-QAM |
| CvM | No | Known | Known | Unknown | Unknown | M-ASK, M-PSK, M-FSK, M-PAM, M-QAM |
| ODST | No | Known | Known | Unknown | Unknown | M-ASK, M-PSK, M-FSK, M-PAM, M-QAM |

**Table 8.3** System requirements and applicable modulation for FB classifiers

| | Pilot samples | Channel gain | Noise variance | Phase offset | Frequency offset | Modulations |
|---|---|---|---|---|---|---|
| Chapter 5 | | | | | | |
| Spectral-based features | No | Unknown | Unknown | Unknown | Unknown | ASK, PSK, FSK, PAM, QAM |
| Wavelet-based features | No | Unknown | Unknown | Unknown | Unknown | M-ASK, M-PSK, M-FSK, M-PAM, M-QAM |
| Cyclic cumulants | No | Unknown | Unknown | Unknown | Unknown | M-ASK, M-PSK, M-FSK, M-PAM, M-QAM |
| Moments | No | Unknown | Unknown | Unknown | Unknown | M-ASK, M-PSK, M-FSK, M-PAM, M-QAM |
| Cumulants | No | Unknown | Unknown | Unknown | Unknown | M-ASK, M-PSK, M-FSK, M-PAM, M-QAM |

**Table 8.4** System requirements and applicable modulation for machine learning classifiers

| | Pilot samples | Channel gain | Noise variance | Phase offset | Frequency offset | Modulations |
|---|---|---|---|---|---|---|
| Chapter 6 | | | | | | |
| KNN | Yes | Unknown | Unknown | Unknown | Unknown | M-ASK, M-PSK, M-FSK, M-PAM, M-QAM |
| SVM | Yes | Unknown | Unknown | Unknown | Unknown | M-ASK, M-PSK, M-FSK, M-PAM, M-QAM |
| ANN | Yes | Unknown | Unknown | Unknown | Unknown | M-ASK, M-PSK, M-FSK, M-PAM, M-QAM |
| GA-KNN | Yes | Unknown | Unknown | Unknown | Unknown | M-ASK, M-PSK, M-FSK, M-PAM, M-QAM |
| GP-KNN | Yes | Unknown | Unknown | Unknown | Unknown | M-ASK, M-PSK, M-FSK, M-PAM, M-QAM |

**Table 8.5** System requirements and applicable modulation for blind modulation classifiers

| | Pilot samples | Channel gain | Noise variance | Phase offset | Frequency offset | Modulations |
|---|---|---|---|---|---|---|
| Chapter 7 | | | | | | |
| EM-ML | No | Compensated | Compensated | Compensated | Unknown | M-ASK, M-PSK, M-FSK, M-PAM, M-QAM |
| MD-NPLF | No | Compensated | Unknown | Compensated | Unknown | M-ASK, M-PSK, M-FSK, M-PAM, M-QAM |

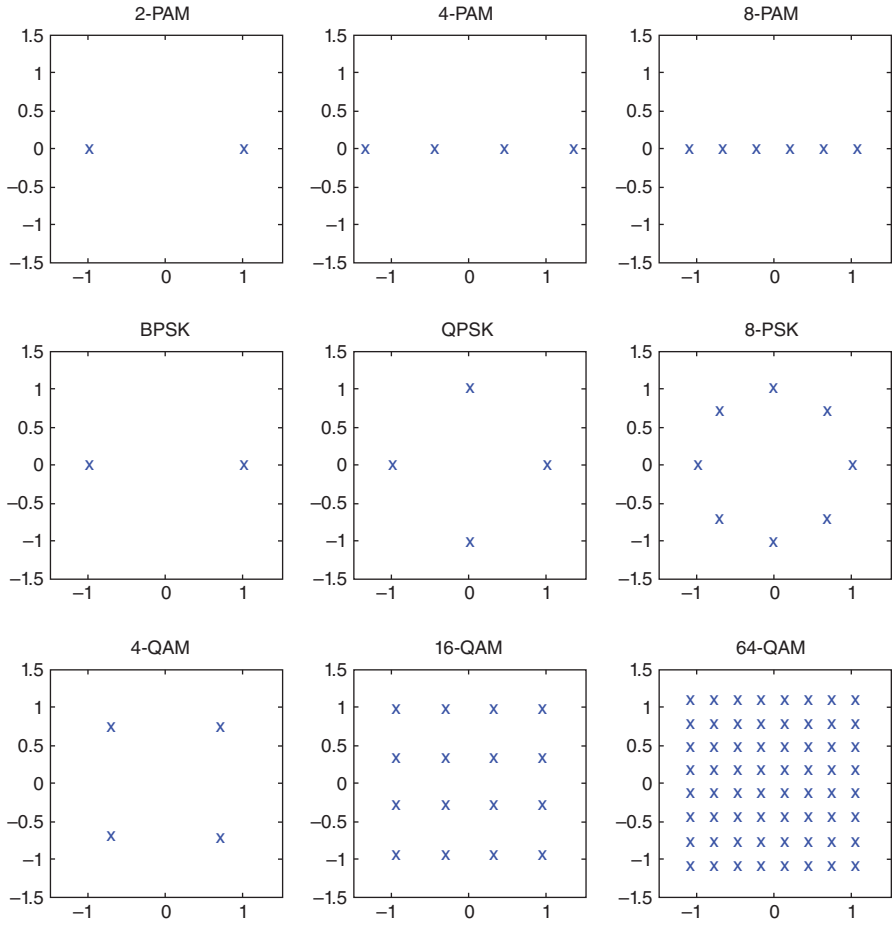



**Figure 8.1** Symbol mapping for different modulations on I-Q plane.

### *8.3.1 Benchmarking Classifiers*

Among all the classifiers, we have selected a small set of classifiers to represent each category of classifier.

For likelihood-based classifiers, the maximum-likelihood classifier (Wei and Mendel, 2000) is used to investigate the performance characteristic of likelihood-based classifiers. In addition, when given perfect channel knowledge, the ML classifier is also used to establish the upper bound of classification accuracy. The likelihood function used is given by equation (3.3). The channel gain and noise variance are assumed to be known to the classifier.

For distribution test-based classifiers, the one-sample KS test classifier (Wang and Wang, 2010) is adopted to represent the distribution test-based classifiers. The channel gain and noise variance are assumed to be known to the classifier.

For feature-based classifiers, we have selected both moments and cumulants (Swami and Sadler, 2000) due to their high classification accuracy for digital modulations of different orders. The classification for both sets of features is accomplished using a KNN classifier. For each signal modulation, 30 signal realizations are generated to construct the reference feature space. No channel knowledge is used for the classification. The list of moments used includes $\mu_{20}$, $\mu_{21}$, $\mu_{40}$, $\mu_{41}$, $\mu_{43}$, $\mu_{60}$, $\mu_{61}$, $\mu_{62}$, and $\mu_{63}$. The list of cumulants used includes $C_{40}$, $C_{41}$, $C_{43}$, $C_{60}$, $C_{61}$, $C_{62}$, and $C_{63}$.

For machine learning classifiers, the combination of GP and KNN is employed to utilize both the moments and cumulants feature sets. In the training state, the same training signals for the KNN classifier is used for the evaluation of feature combinations. The GP process is performed before the signal classification in the training stage in order to achieve the optimized feature selection and combination of the available features. No channel knowledge is used in the classifier.

For blind modulation classifiers, the EM-ML classifier is implemented with an EM stage that provides the joint estimation of complex channel coefficient and noise variance for each modulation hypothesis. The resulting estimates are used for the evaluation of likelihood in each hypothesis for the ML classifier.

### 8.3.2 Performance Comparison in AWGN Channel

In the first set of experiments we examine the robustness of different classifiers when mismatch in noise level is presented. The noise level is measured by the signal-to-noise ratio, which is defined as given in equation (8.2),

$$SNR = 10\log_{10}\frac{P(s)}{P(\omega)}\,\text{dB} \tag{8.2}$$

where $P(s)$ is the power of the power of the transmitted signal, and $P(\omega)$ is the power of the AWGN noise.

From −20 to 20 dB with a step of 1 dB, the AWGN noises are generated for each signal realization according the signal power. For the remainder of this subsection, classification accuracy for all the modulations is detailed for each benchmarking classifier. An overview of the performance comparison between all the classifiers is presented at the end.

In Figure 8.2, the classification accuracy of the ML classifier is listed for each testing signal modulation. The classification accuracy $P_{cc}$ is computed with the aid of equation (8.3),

$$P_{cc} = \frac{L_S}{L}\,(\%) \tag{8.3}$$

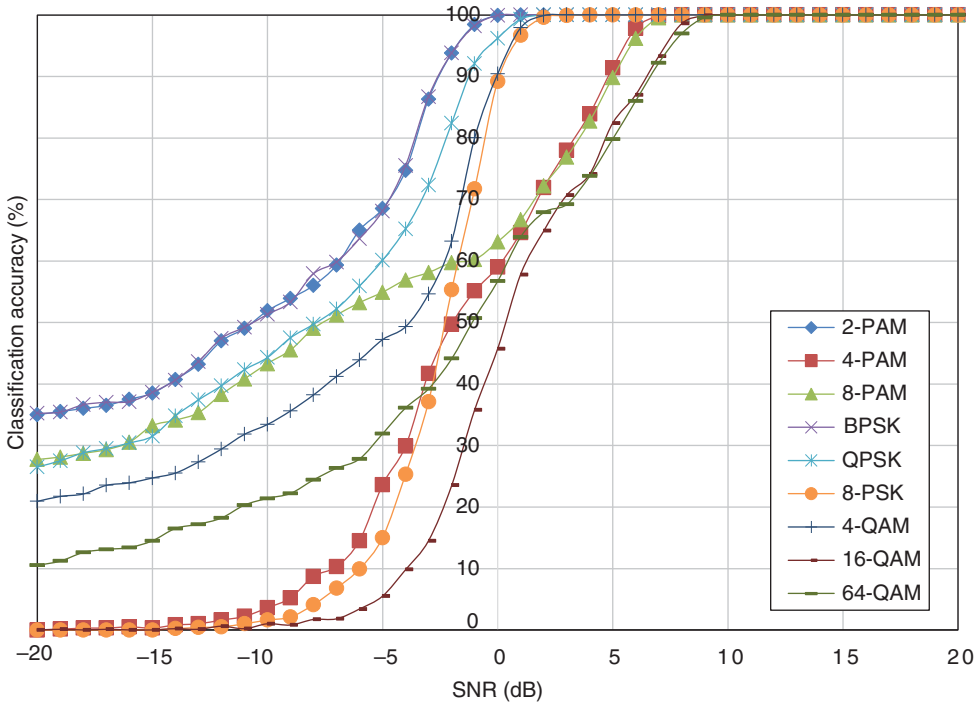



**Figure 8.2** Classification accuracy of the ML classifier in AWGN channel. (*See insert for color representation of the figure.*)

where $L_s$ is the number of signal realizations that have been successfully classified. It is clear that with lower SNR (higher noise level) there is significant performance degradation for all modulations. Among all the modulations, 2-PAM, BPSK and QPSK have relatively higher classification accuracy at all noise levels. The 8-PSK and 4-QAM are also easier to classify with perfect classification (classification accuracy of 100%) achieved below 5 dB. Meanwhile, 4-PAM, 8-PAM, 16-QAM and 64-QAM require high signal power to achieve perfect classification. It is clear that higher-order modulations are more difficult to classify. Under the same noise level, the modulation symbols of a high-order digital modulation are more densely populated and make the symbol states much less distinctive. It is worth noting that the order of the modulation on a single signal dimension has closer correlation to the classification difficulty. Both having four symbol states, 4-PAM has four states on one dimension while QPSK and 4-QAM have fewer states on each dimension. Therefore, the classification of 4-PAM is more difficult than those for QPSK and 4-QAM. The perfect classification of all modulations by the ML classifier is achieved when the SNR level is higher than 9 dB.

In Figure 8.3, the classification accuracy of the one-sample KS test classifier is listed for each testing signal modulation. Similar behaviour is observed for the one-sample

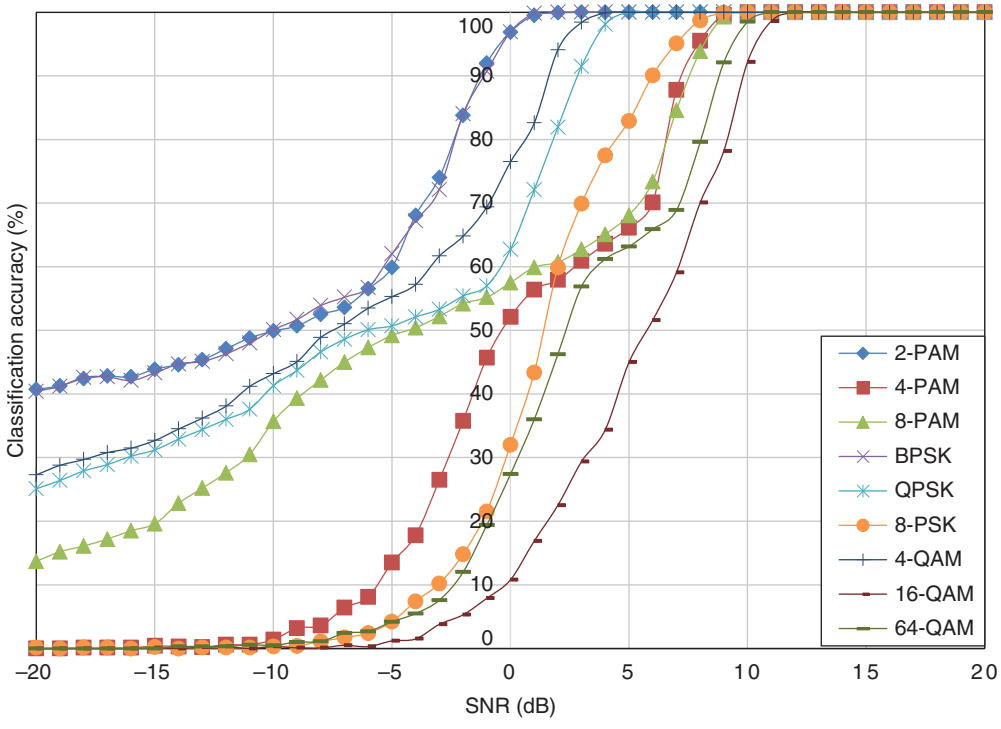



**Figure 8.3** Classification accuracy of the KS test classifier in AWGN channel.

KS test classifier where high-order modulations have lower classification accuracy under the same noise level. Compared with the ML classifier, the performance of the one-sample KS test classifier is inferior but with a small margin. The biggest difference is observed for 8-PSK modulation, where an SNR of 9 dB is required to achieve perfect classification while only 2 dB is need for the ML classifier. The same observation applies to QPSK modulation. The reason for the classification difference for higher-order PSK modulation could be that the one-sample KS test classifier processes the decomposed signal I and Q segments. Therefore, the separation of modulation symbols in their phase is not exploited enough. The perfect classification of all modulations by the one-sample KS test classifier is achieved when the SNR level is higher than 12 dB.

In Figure 8.4, the classification accuracy of the KNN classifier using moments is listed for each testing signal modulation. That for the cumulant-based classifier is shown in Figure 8.5. One significant performance difference between the high-order statistics-based classifiers is that their classification accuracy for certain high-order modulations has an upper bound which cannot be exceeded regardless of the noise level. The reason for the upper bound will be explained in the next section when we investigate the effort of signal length on classification accuracy for different classifiers.

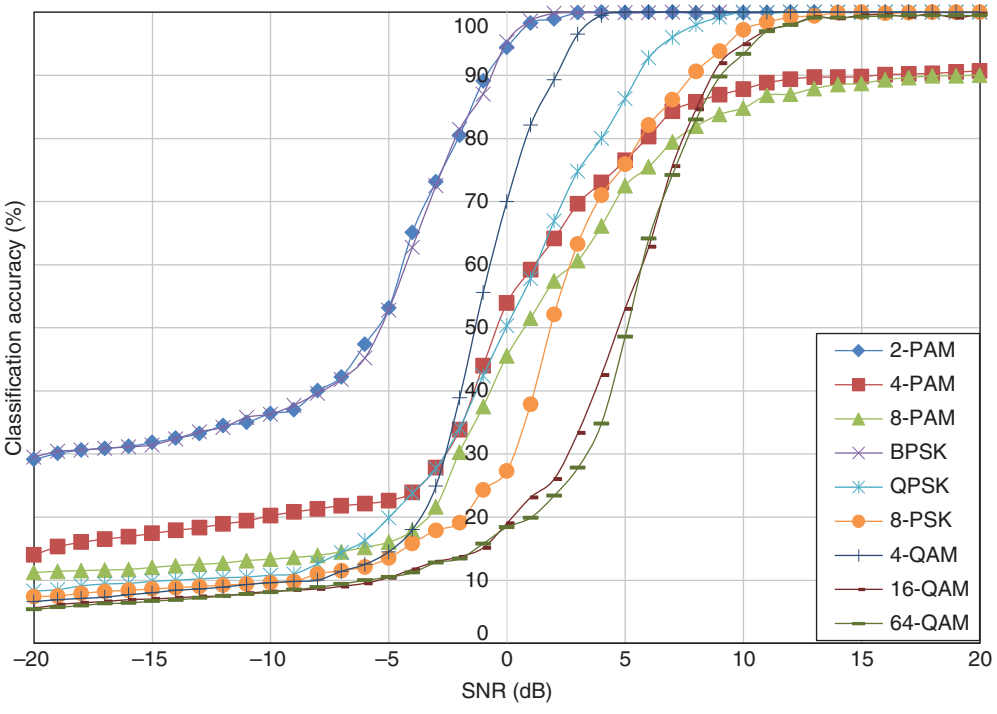



**Figure 8.4** Classification accuracy of the moment-based KNN classifier in AWGN channel.

Apart from that, the moment-based and cumulant-based KNN classifiers share the same degraded performance for high-order PSK modulation for the same reason as the one-sample KS test classifier. In fact the effect is even more obvious for the high-order statistics feature-based classifiers. The performance profiles of the moment-based and cumulant-based classifiers are very similar, although the cumulant-based KNN classifiers have slightly better accuracy in general.

In Figure 8.6, the classification accuracy of the GP-KNN classifier using moment and cumulant features is listed for each testing signal modulation. Compared with the KNN classifier using the same features but without feature selection and feature combination, the performance profile for the GP-KNN classifier shows that the classification accuracy for some of the higher-order modulations is similar. The reason for the change of performance characteristic is the result of the KNN classification fitness evaluation in the GP process. The fitness evaluation calculates the average classification accuracy of a group of training signals from all modulations. Effectively, it encourages the improvement of average classification accuracy. Therefore, the resulting performance pattern is likely to be restructured.

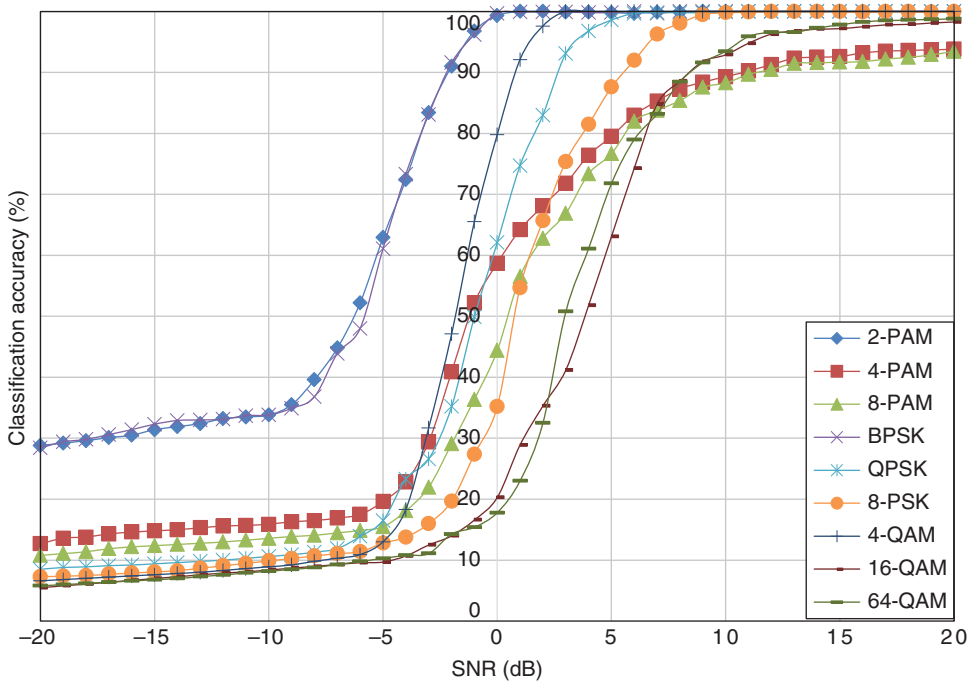



**Figure 8.5** Classification accuracy of the cumulant-based KNN classifier in AWGN channel.

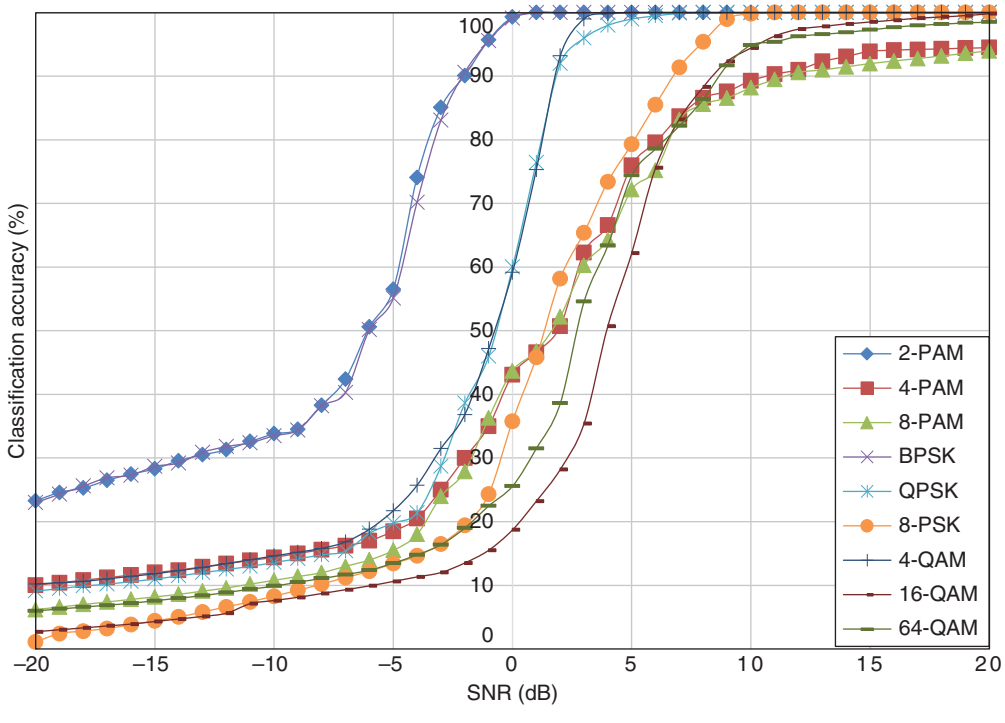



**Figure 8.6** Classification accuracy of the GP-KNN classifier in AWGN channel.

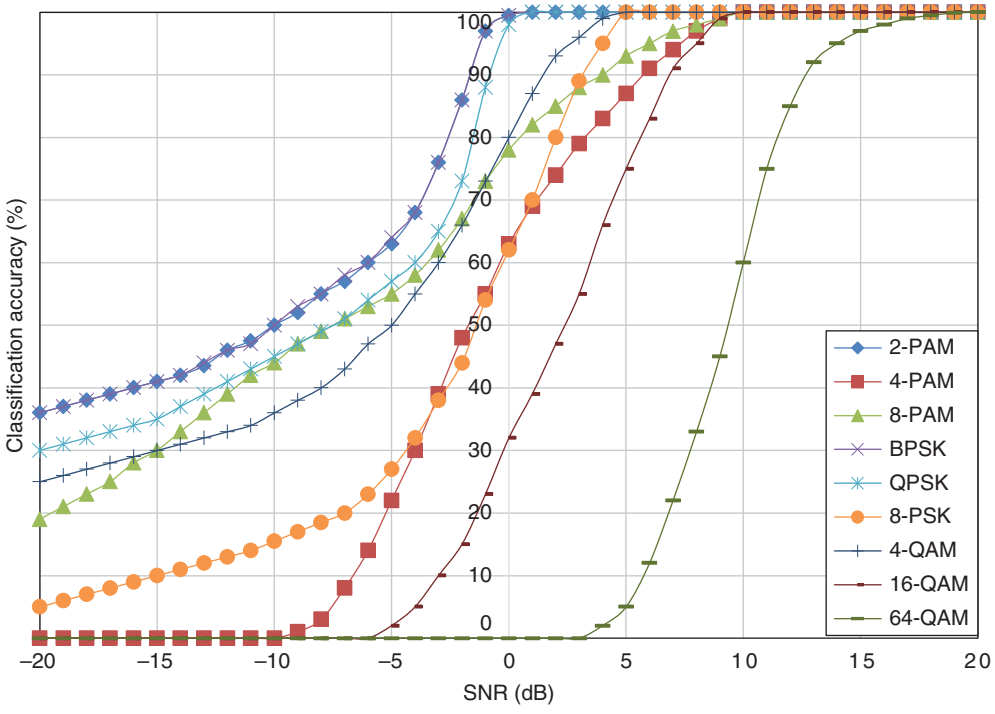



**Figure 8.7** Classification accuracy of the EM-ML classifier in AWGN channel.

In Figure 8.7, the classification accuracy of the EM-ML classifier for different modulations in AWGN channel with varying noise levels is shown. The EM-ML classifier shares some of the performance characteristics of the ML classifier. The optimal classification accuracy in the given noise level range is not limited by the number of samples available for analysis, unlike the moment- and cumulant-based classifiers. However, the high-order modulations have significantly lower classification accuracy as compared with the ML classifier. The difference between the EM-ML classifier and the ML classifier is that, for the ML classifier, a single set of channel parameters is used for the evaluation of likelihood for different hypotheses. Meanwhile, the EM-ML classifier estimates a different set of channel parameters for each modulation hypothesis by using the criteria of maximum likelihood. For false modulation hypotheses, the effect of the EM-ML combination is that the mismatch between the observed signal and the false hypothesis is minimized by the EM estimator. Therefore, classification accuracies for some modulations are reduced.

Taking the average of the classification accuracies of all modulations for each classifier, Figure 8.8 provides an overview of the performance comparison between

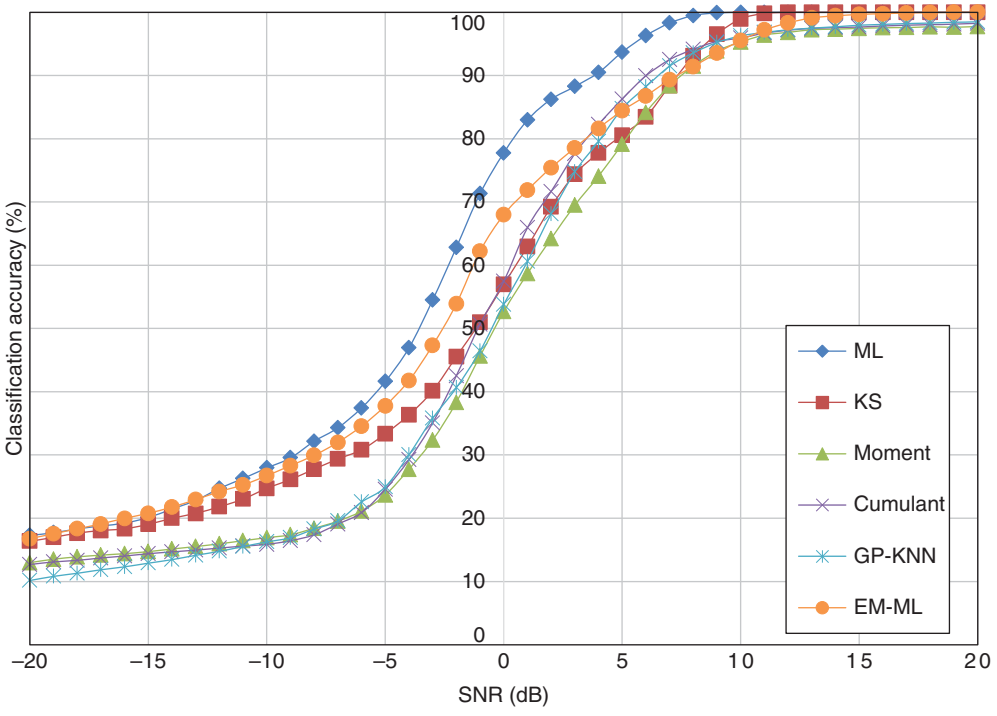



**Figure 8.8** Average classification accuracy of all classifiers in AWGN channel. (*See insert for color representation of the figure.*)

different classifiers in the AWGN channel. It can be seen that the ML classifier has superior classification accuracy at all noise levels. Despite using the same ML classification decision-making method, the EM-ML classifier has significantly lower classification accuracy. However, it is still superior to most of the other classifiers. The KS test classifier is another classifier that is highly accurate in most noise levels. While being superior to the KS test classifier in some noise levels, the moment- and cumulant-based classifiers suffer at high SNR due their limited ability to process signals of small sample size. The GP-KNN classifier shows a small improvement over the basic KNN classifier using either moment or cumulant features.

## 8.4 Classification Accuracy with Limited Signal Length

One aspect of the robustness of a classifier is how it performs with a limited number of signal samples available for analysis. Performance of a classifier is affected by the number of samples. An example would be the distribution test-based classifier.

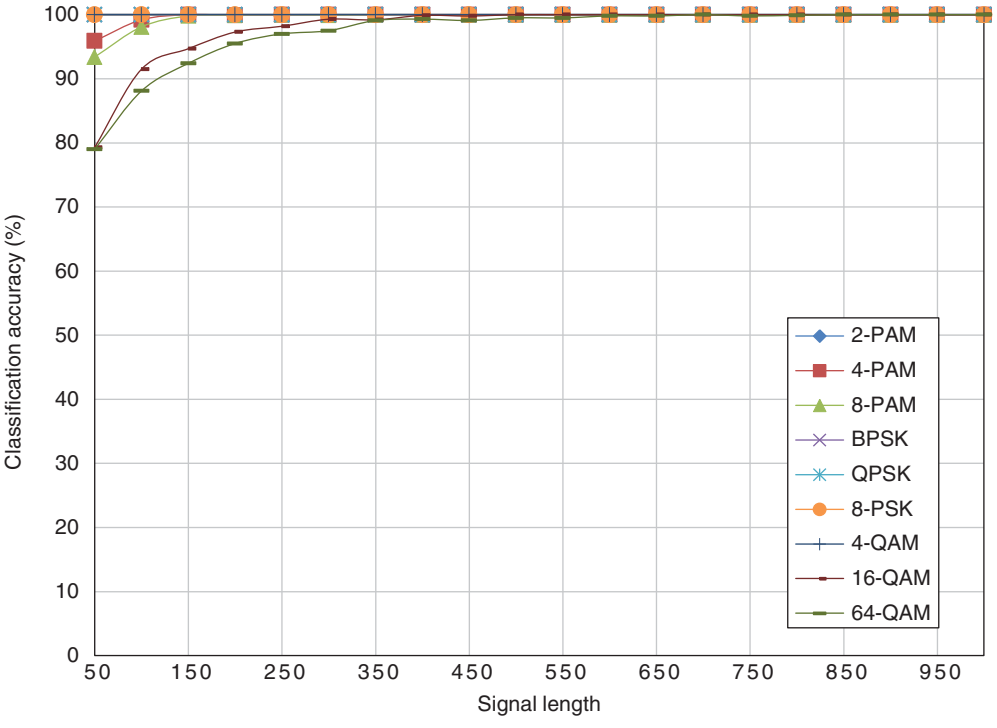



**Figure 8.9** Classification accuracy of the ML classifier with different signal length.

To construct the empirical distribution, there must be a high enough number of signal samples to calculate the distribution. The more samples there are the more accurately would the empirical distribution resemble the true underlying distribution. When a limited number of signal samples is available, outliners in the signal distribution could create distortion to the modelling of signal distribution. Therefore the classification performance can be affected. In this set of experiments we repeat the experiments in Section 8.3. However, the noise level is fixed at SNR = 10 dB while signals of varying lengths are tested. The signal lengths tested vary from 50 to 1000 with a step of 50.

In Figure 8.9, the classification accuracy of the ML classifier is listed for each testing signal modulation given different signal lengths. Most modulations can be classified with accuracies of over 90% with as few as 50 samples available for analysis. The only exception is 16-QAM and 64-QAM modulations. To guarantee perfect classification of the two modulations over 350 samples are needed. Meanwhile, for the other modulation, 150 samples are enough to achieve the same goal. The reason why 16-QAM and 64-QAM modulation are more difficult to classify given a smaller sample size is that

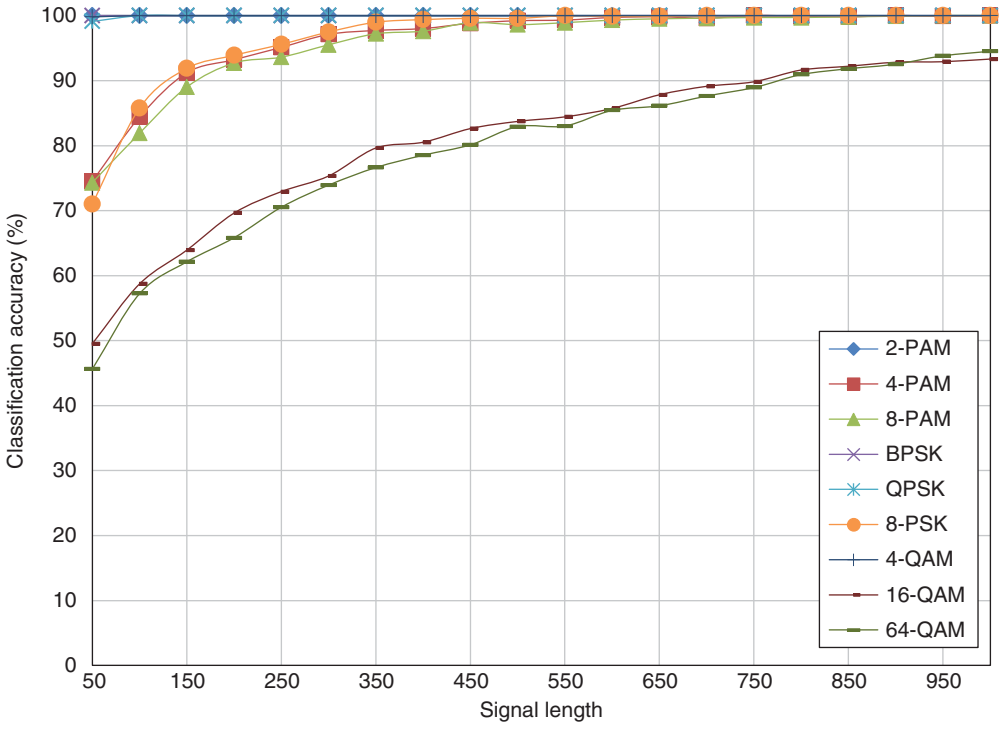



**Figure 8.10** Classification accuracy of the KS test classifier with different signal length.

they both have higher number symbol states. Given 50 samples, it is not even enough to cover all the symbol states in the 64-QAM modulation. To ensure high classification accuracy, there need to be enough signal samples from each symbol state. Thus this property of a receiver can be fully represented. When the same number of signal samples is distributed to different symbol states, it is obvious that the higher-order modulations will have few samples for each symbol states.

In Figure 8.10, the classification accuracy of the KS test classifier is listed for each testing signal modulation given different signal lengths. It is clear that the effect of increased modulation order has a more significant impact on the classification accuracy. Lower-order modulations such 2-PAM, BPSK and QPSK are easily classified with as few as 50 samples. Also, 4-PAM, 8-PAM and 8-PSK have similar performances where a perfect classification is achieved with more than 450 signal samples. Again, the higher-order modulations, that is, 16-QAM and 64-QAM, require a significant number of samples to achieve high classification accuracy. With as few as 1000 samples, neither of the modulations can be perfectly classified.

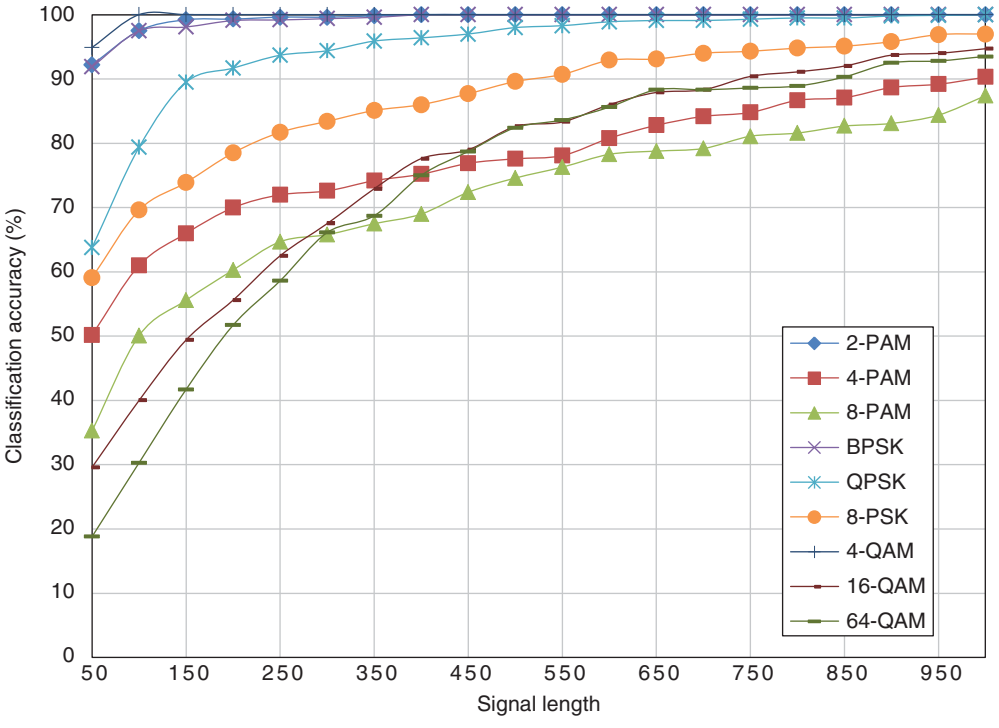



**Figure 8.11** Classification accuracy of the moment-based classifier with different signal length.

In Figure 8.11, the classification accuracy of the moment-based KNN classifier is listed for each testing signal modulation given different signal lengths. The results for the cumulant-based classifier are given in Figure 8.12. In both cases there are clear-performance differences among all the modulations. The lower-order modulations, such as 2-PAM, BPSK and 4-QAM modulations, reach perfect classification when more than 250 samples are available for analysis. Apart from these, QPSK is the only other modulation which can be classified with 100% accuracy when fewer than 1000 samples are given. Among the rest of the modulations, it is interesting that high-order PAM modulations have higher classification accuracy, when fewer than 400 samples are available for analysis. However, the performance improvement with increased sample size is inferior to that found for high-order QAM modulations.

In Figure 8.13, the classification accuracy of the moment- and cumulant-based GP-KNN classifier is listed for each testing signal modulation given different signal lengths. Again, lower-order modulations show better classification accuracy. For 2-PAM, BPSK and 4-QAM, perfect classification is achieved given 150 samples for

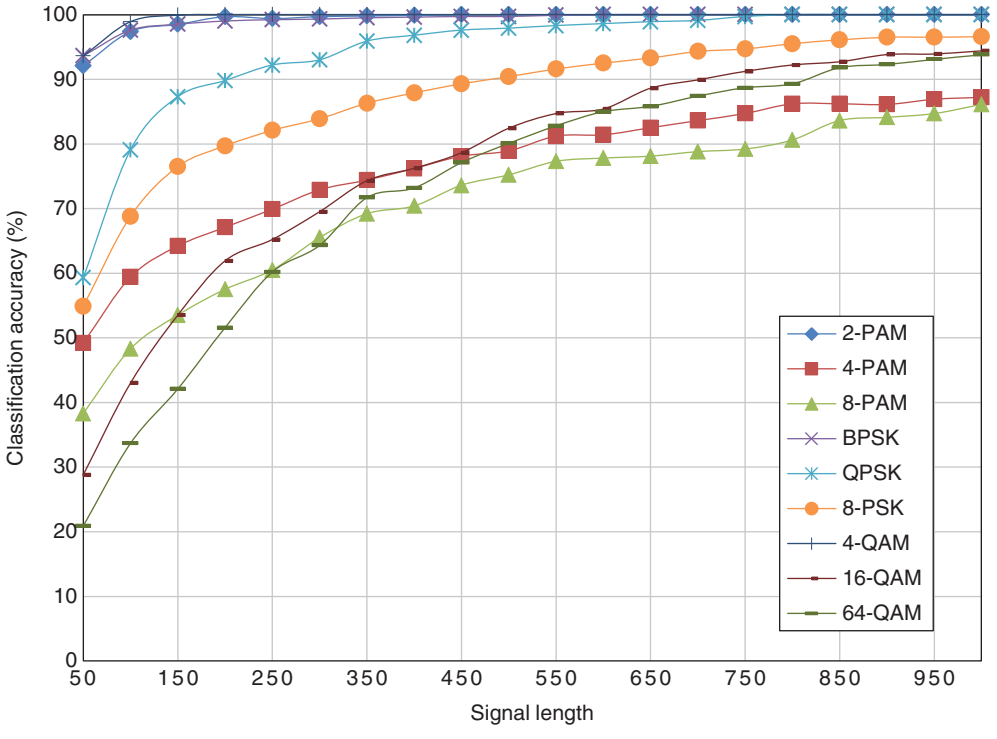



**Figure 8.12** Classification accuracy of the cumulant-based classifier with different signal length. (*See insert for color representation of the figure.*)

analysis. The same condition is met when 300 and 550 samples are available for QPSK and 8-PSK respectively. Different from the moment-based and cumulant-based classifiers, the GP-KNN classifier shows a much more similar performance profile for both higher-order PAM and QAM modulations. Yet none of these modulations can be classified perfectly with fewer than 1000 signal samples.

In Figure 8.14, the results from the EM-ML classifier are given. Like the ML classifier, the classifier is robust given a reduced number of samples for most modulations. However, the 64-QAM is more difficult to classify and shows a noticeable degradation with fewer signal samples for analysis.

Given the average classification accuracy of different classifiers in the fading channel with frequency offset in Figure 8.15, the ML classifier is obviously superior to all the other classifiers. The moment- and cumulant-based classifiers are much limited by the sample size. The limitation is improved when feature selection and combination is performed using GP. The KS test classifier and the EM-ML classifier have similar performance patterns. The EM-ML classifier has higher classification accuracy with

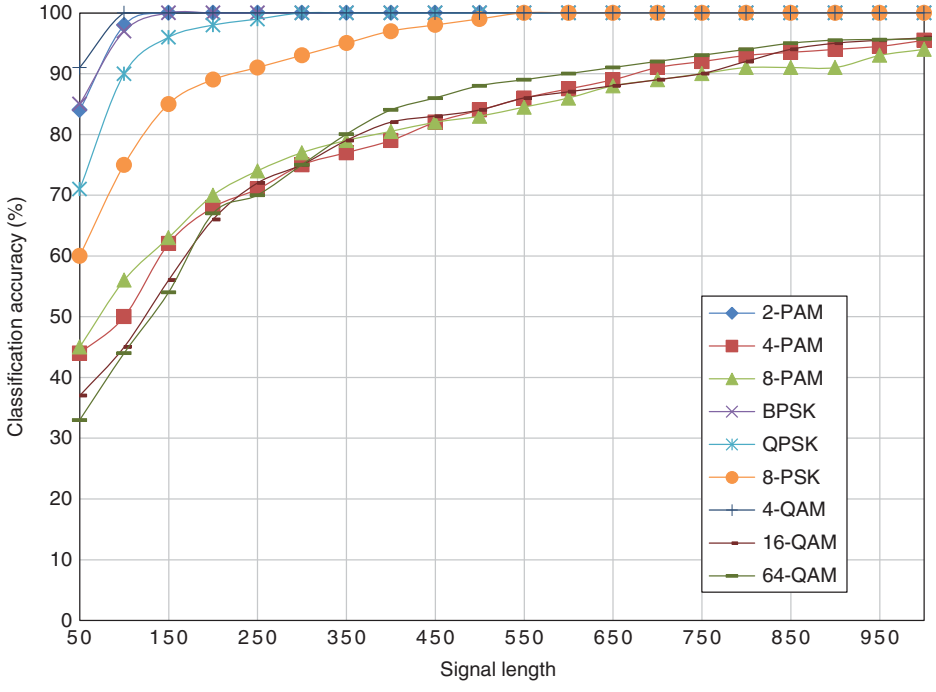



**Figure 8.13** Classification accuracy of the GP-KNN classifier with different signal length.

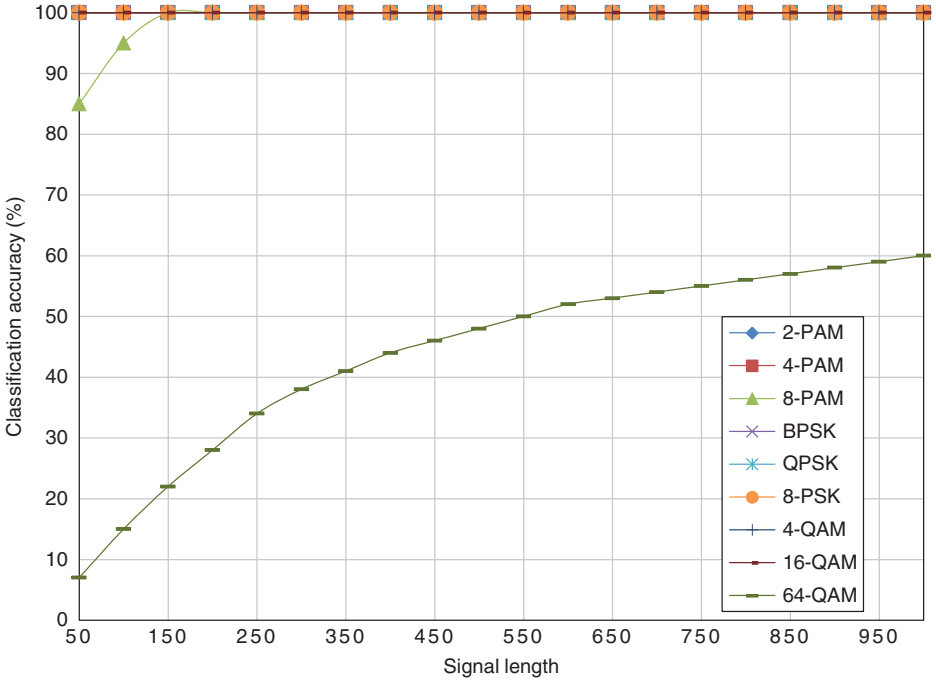



**Figure 8.14** Classification accuracy of the EM-ML classifier with different signal length.

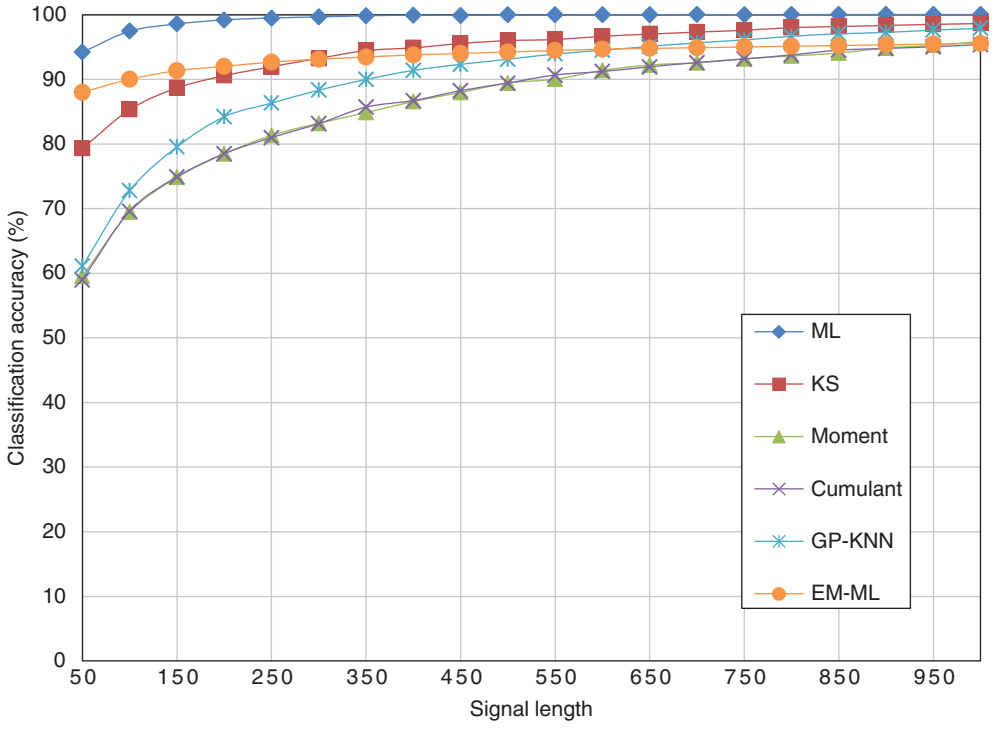



**Figure 8.15** Average classification accuracy of all classifiers with different signal length. (*See insert for color representation of the figure.*)

a small number of signal samples. On the other end, when more signal samples are available for analysis, the KS test classifier shows better performance.

## 8.5 Classification Robustness against Phase Offset

While additive noises are the most commonly considered channel conditions, fading effects are inevitable for systems in wireless channels. To model the fading channel, we use the signal model defined by equation (1.13). In this section we consider only the carrier phase offset. Thus the signal model is given by equation (8.4),

$$r[n] = \alpha e^{j\theta_o} s[n] + \omega[n] \tag{8.4}$$

where $\theta_o$ is the phase offset. In this set of experiments, we fix both the noise level and signal length at SNR = 10 dB and 1024, respectively. Phase offsets, ranging

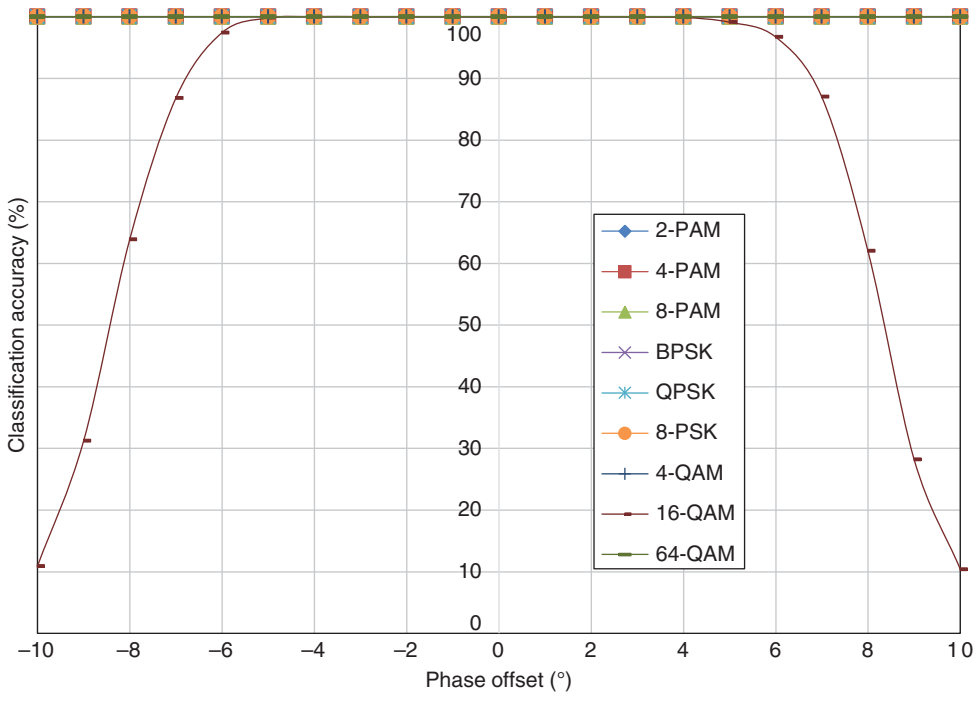



**Figure 8.16** Classification accuracy of the ML classifier with phase offset.

from –10° to 10° with a step of 1°, are considered. For each phase offset, 1000 pieces of signal are generated using equation (8.4).

In Figure 8.16, the classification accuracy of the ML classifier is listed for each testing signal modulation against different levels of phase offset. It can be seen that all modulations have 100% classification accuracy when no phase offset is added. Analytically, it is expected that the performance would degrade when an increasing amount of phase offset is introduced. However, this phenomenon is not observed for the majority of the modulations. It could be concluded that this is a result of the robustness of the ML classifier when moderate amounts of the phase offset are introduced. The only modulation affected by the channel condition significantly is the 16-QAM modulation. Performance degradation is experienced when more than 4° of phase offset is simulated. A dramatic decrease in classification is seen when over 7° of phase offset is considered. In a channel with a phase offset of 10°, the corresponding classification accuracy is reduced to 10%.

In Figure 8.17, the classification accuracy of the KS test classifier is listed for each testing signal modulation against different levels of phase offset. For lower-order modulations, including 2-PAM, 4-PAM, 8-PAM, BPSK, QPSK, 8-PSK and 4-QAM, the KS test classifier is robust enough to provide 100% classification accuracy in the

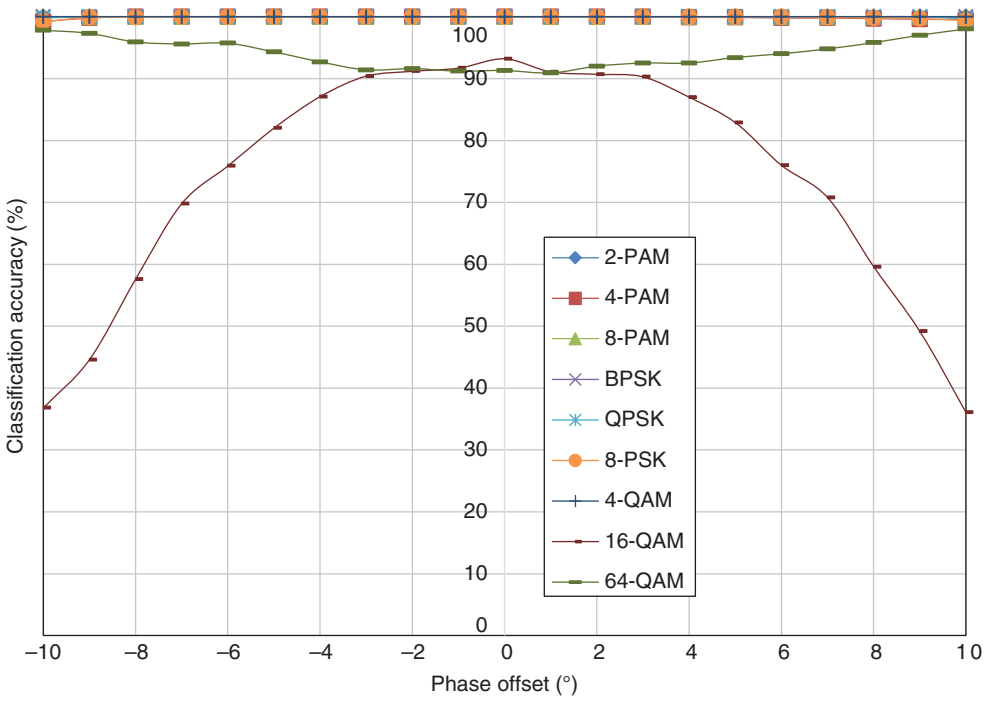



**Figure 8.17** Classification accuracy of the KS test classifier with phase offset.

given setup when a phase offset of less than 10° is considered. Similar to the ML classifier, the 16-QAM modulation is significantly affected by the phase offset. Interestingly, the 64-QAM classification accuracy increases with more phase offset. As the KS test measured the mismatch between the empirical distribution and the reference distribution, phase offset increases the mismatch when the accurate modulation is used for reference. In the meantime, the mismatch between the distorted empirical distribution and a reference distribution from a false modulation hypothesis is also enlarged. Since the mismatch against the false hypothesis increases at a faster rate with increasing level of phase offset, it could be understood that the classification accuracy can increase with more phase offset. However, in these experiments, such a phenomenon is only observed for 64-QAM modulation.

In Figure 8.18, the classification accuracy of the moment-based KNN classifier is listed for each testing signal modulation against different levels of phase offset. The results for the cumulant-based classifier are displayed in Figure 8.19. Between the moment and cumulant features, it is obvious that the cumulant features have higher robustness in the fading channel with phase offset. The biggest difference can be seen for 64-QAM modulation; using cumulants the classifier is able to achieve a constant level of classification accuracy within −10° to 10° of phase offset. Meanwhile, the

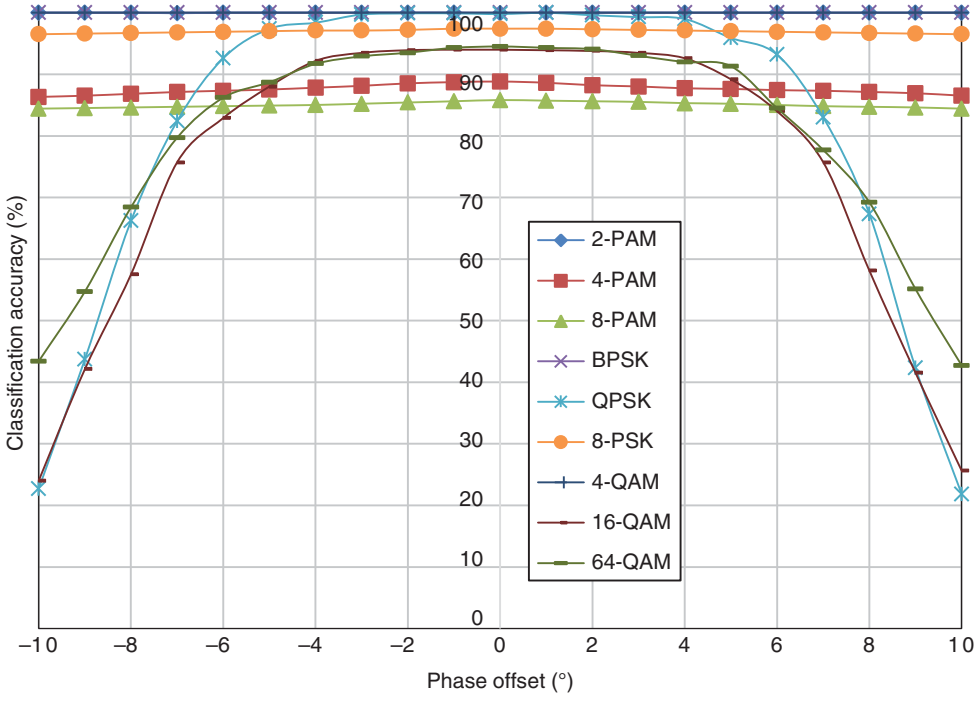



**Figure 8.18** Classification accuracy of the moment-based KNN classifier with phase offset. (*See insert for color representation of the figure.*)

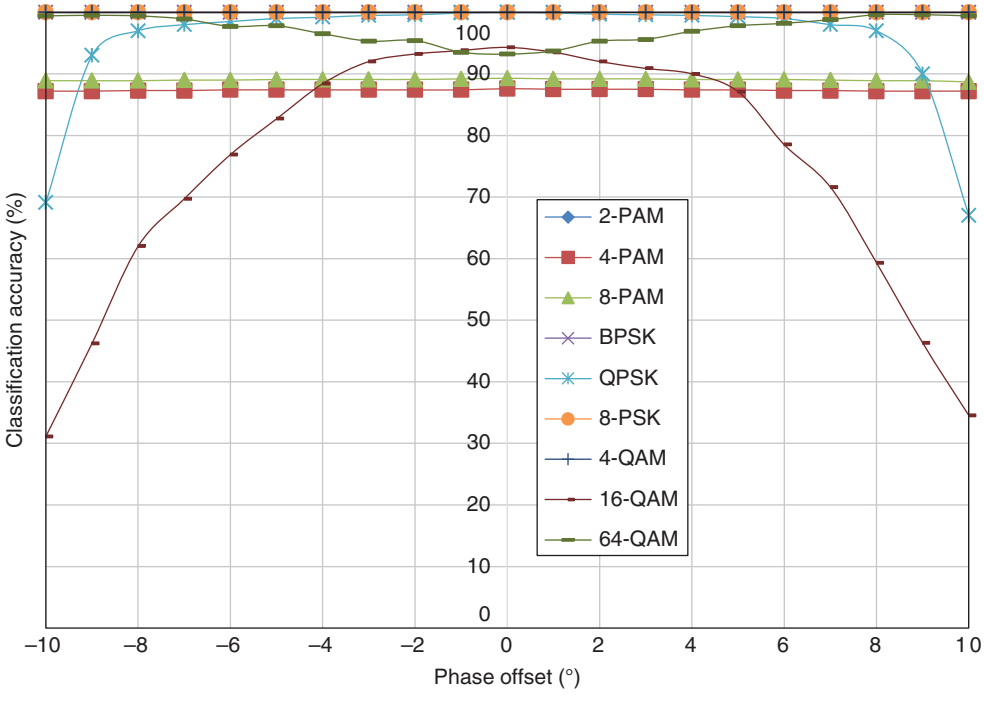



**Figure 8.19** Classification accuracy of the cumulant-based KNN classifier with phase offset.

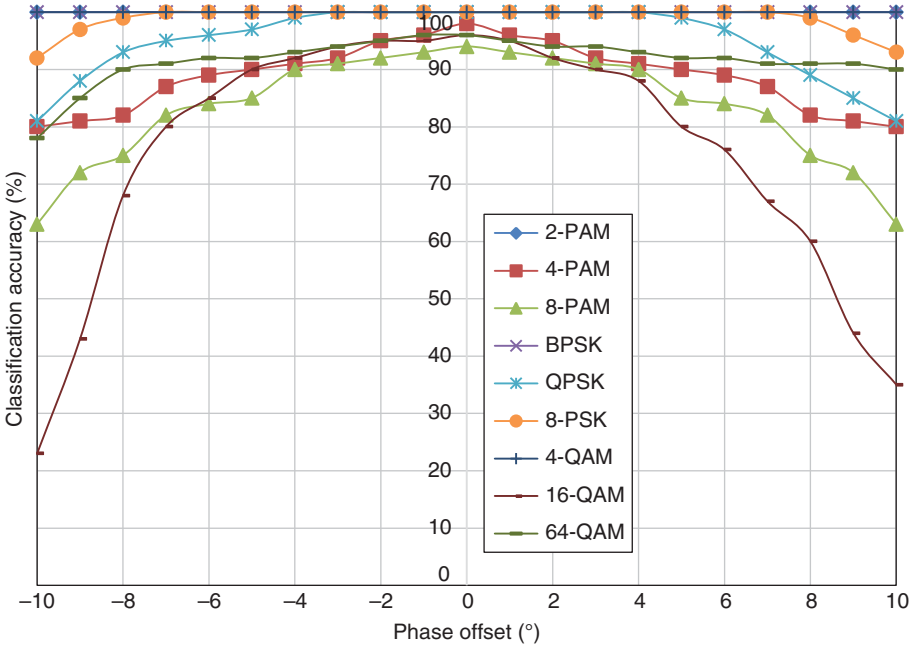



**Figure 8.20** Classification accuracy of the GP-KNN classifier with phase offset.

classification accuracy for 64-QAM sees dramatic degradation when more than 5° of phase offset is introduced. In addition, the robustness of cumulant-based KNN classifier when classifying QPSK is also higher, where severe degradation is only observed when phase offset exceeds 8°.

In Figure 8.20, the classification accuracy of the moment- and cumulant-based GP-KNN classifier is listed for each testing signal modulation against different levels of phase offset. Only 2-PAM, 4-QAM and BPSK enjoy consistent classification accuracy. All other modulations experience different degrees of performance degradation when phase offset is introduced. The reason why feature selection and combination does not improve the performance is that the training of the feature selection and feature combination is conducted in a channel without any phase offset. The resulting feature set is inevitably over-trained for the channel without phase offset. When phase offset is introduced, the mismatch between the feature value of a fading signal and the reference signal space is increased.

In Figure 8.21, the classification accuracy of the EM-ML classifier is listed for each testing signal modulation against different levels of phase offset. The EM-ML classifier shows excellent performance in the fading channel with phase offset. For all modulations, there is no obvious performance degradation. Since the estimation of channel coefficient consists of the estimation of carrier phase offset, the phase offset is

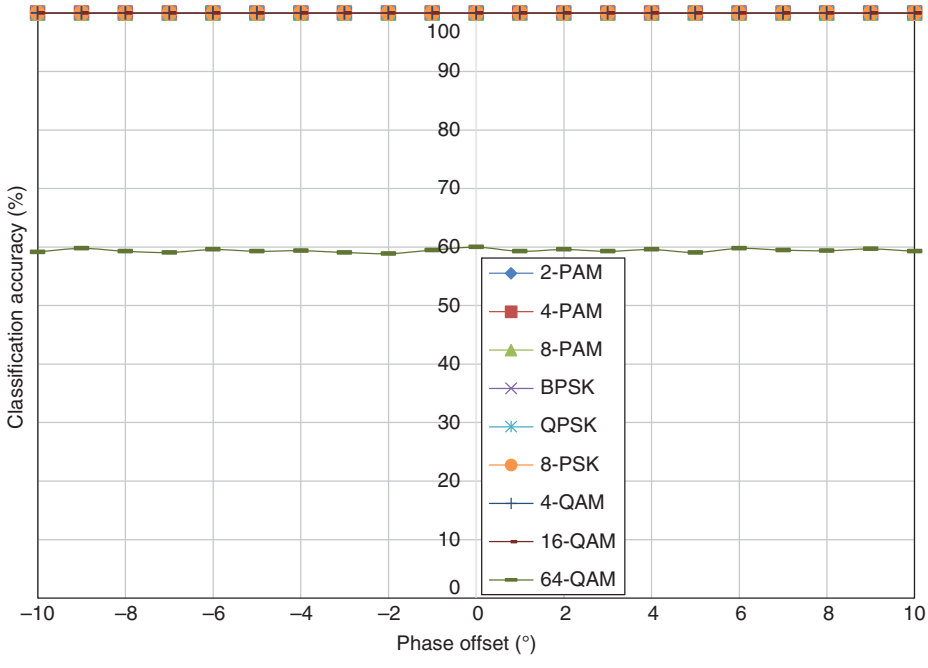



**Figure 8.21** Classification accuracy of the EM-ML classifier with phase offset.

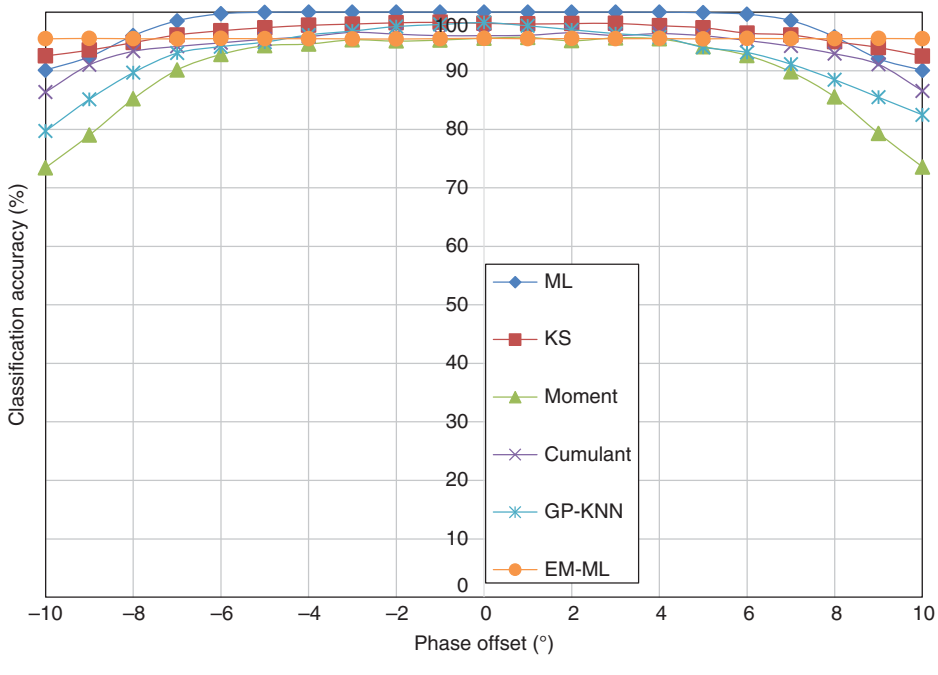



**Figure 8.22** Average classification accuracy of all classifiers with phase offset. (*See insert for color representation of the figure.*)

effectively compensated in the ML classification stage. Thus, the EM-ML classifier has very robust performance in the fading channel with some phase offset.

To conclude this section, we give the average classification accuracy over all modulations for each classifier in the case of a fading channel with carrier phase offset in Figure 8.22. Between $-8^\circ$ and $8^\circ$ of phase offset, the ML classifier is still the best classifier in terms of its classification accuracy. However, beyond $8^\circ$, the EM-ML classifier has higher classification accuracy thanks to its inherent ability to compensate the phase offset estimated in the EM stage. The weakest classifier in the fading channel is the moment-based KNN classifier. Its performance degradation with an increasing amount of phase offset is most severe among all the classifiers.

## 8.6 Classification Robustness against Frequency Offset

Another practical channel effect to be considered in common fading channels is carrier frequency offset. To model the received signal in a fading channel with frequency offset, we derive the following equation (8.5) from equation (1.13). Here,

$$r[n] = \alpha e^{j2\pi \cdot n \cdot f_o/f} s[n] + \omega[n] \tag{8.5}$$

$f_o$ is the frequency offset. It is worth noting that, in the simulation, a relative frequency $f_o/f$, which is calculated from the ratio between the actual frequency offset and the symbol sampling frequency, is used to indicate different levels of frequency offset. The experiment setup in this section is the same as in the previous section, the only difference being that phase offset is neglected and frequency offset is considered. The range of frequency offset used in the simulation is from $1 \times 10^{-5}$ to $2 \times 10^{-4}$.

In Figure 8.23, the classification accuracy of the ML classifier for different modulations with varying levels of frequency offset is given. Different from the other channel conditions we have investigated, the effect of frequency offset is more impulsive in terms of the classification accuracies for different modulations. Most modulations see a dramatic decrease in classification accuracy when a small amount of frequency is added. Surprisingly, high-order modulations, namely 8-PSK and 64-QAM, have much superior classification accuracy when considering frequency offset. This significantly biased performance profile is caused by the unique effect of frequency offset on lower-order modulations. Unlike the phase shift in fading channel, frequency offset introduces a phase shift which is time variant and incremental over time. Therefore, the resulting effect is no longer a shift of the entire signal constellation. Instead, it produces a rotational dispersion of the signal samples. For modulations of lower order, the effective result of frequency offset is the increase in symbols numbers when perceived by the receiver. Therefore, the classification is biased for higher-order modulations. It is not unique to 8-PSK or 64-QAM. If other higher-order modulations

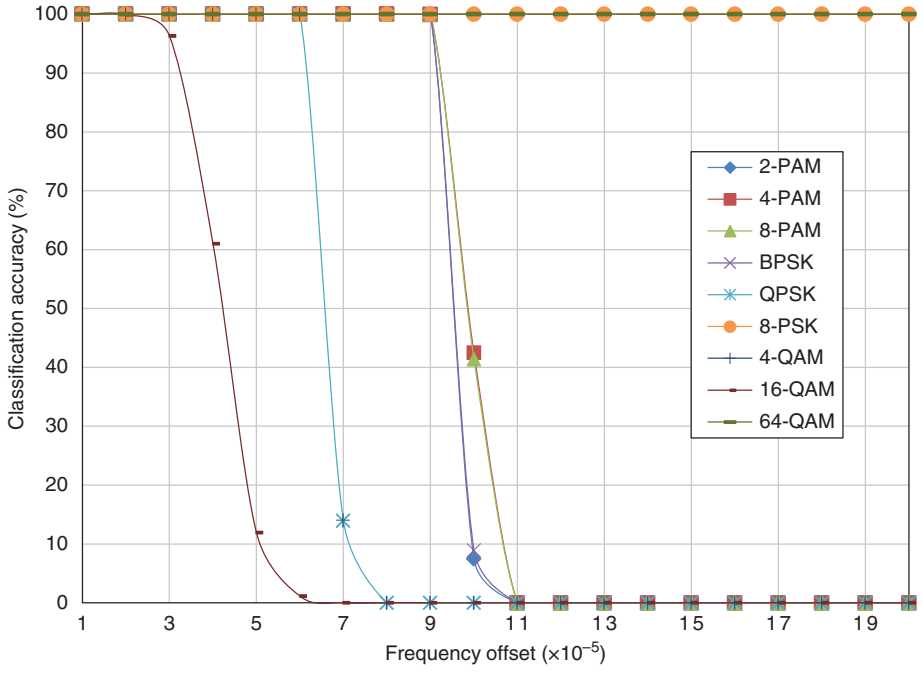



**Figure 8.23** Classification accuracy of the ML classifier with frequency offset.

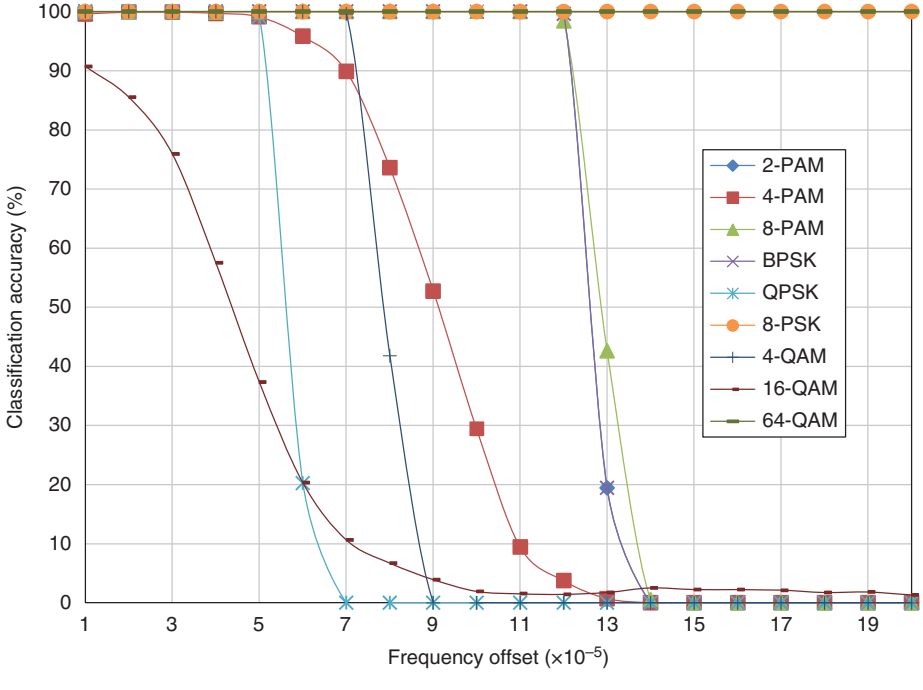



**Figure 8.24** Classification accuracy of the KS test classifier with frequency offset.

are included in the modulation pool, it is likely that they will be correctly classified more often than are 8-PSK or 64-QAM.

In Figure 8.24, the classification accuracy of the KS test classifier for different modulations with varying level of frequency offset is shown. The observation of the results coincides with what has been discussed previously for the ML classifier. In fact, this could be said for the remainder of the benchmarking classifiers. The results for the moment-based KNN classifier, the cumulant-based KNN classifier, the moment- and cumulant-based GP-KNN classier and the EM-ML classifier are shown in Figures 8.25–8.28, respectively.

Figure 8.29 shows the average classification accuracy of each classifier in the simulated fading channel with carrier frequency offset. Owing to the severe distortions caused by frequency offsets, most classifiers have a rather distorted performance profile. However, the trends of the performance changes are shared among most classifiers. The only exception is the EM-ML classifier, while being less accurate when a small amount of frequency offset is added, it is noticeably more robust when a higher level of frequency offset is considered. When most classifiers start to degrade with

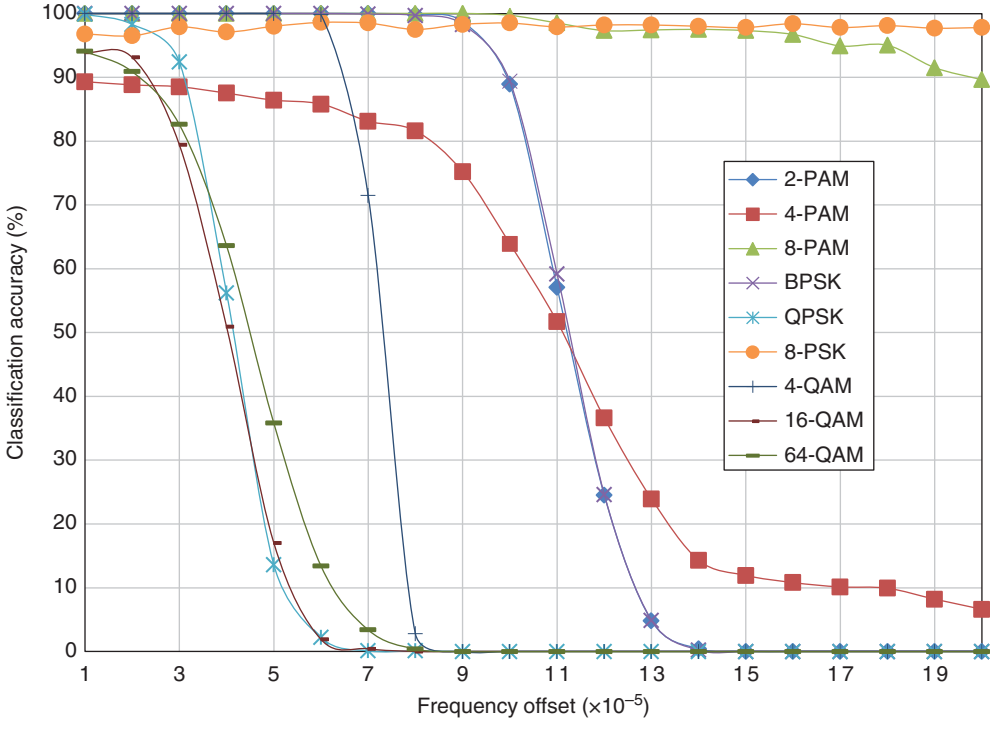



**Figure 8.25** Classification accuracy of the moment-based KNN classifier with frequency offset.

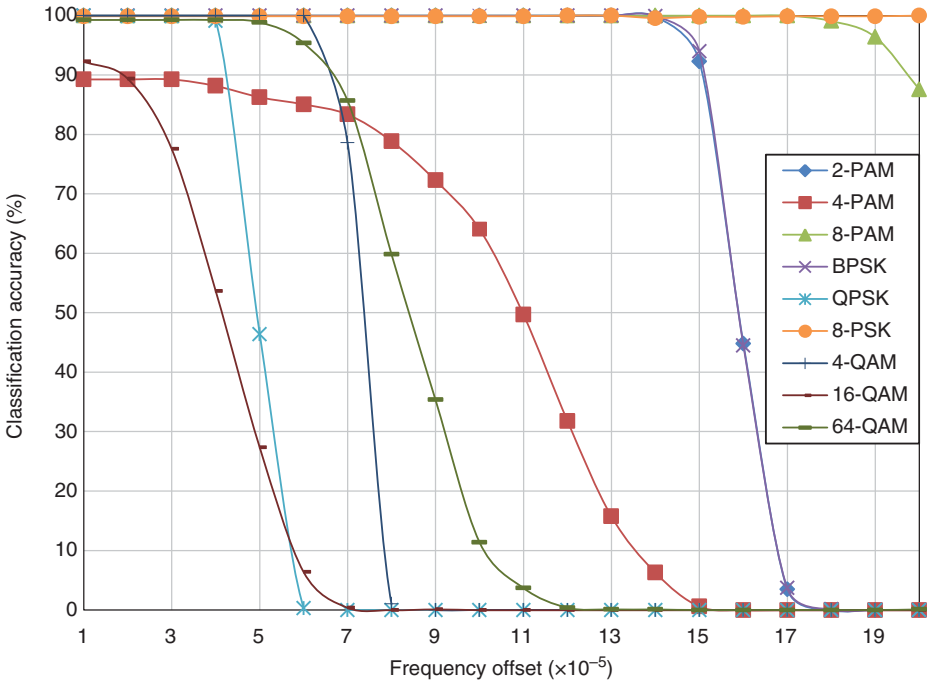



**Figure 8.26** Classification accuracy of the cumulant-based KNN classifier with frequency offset.

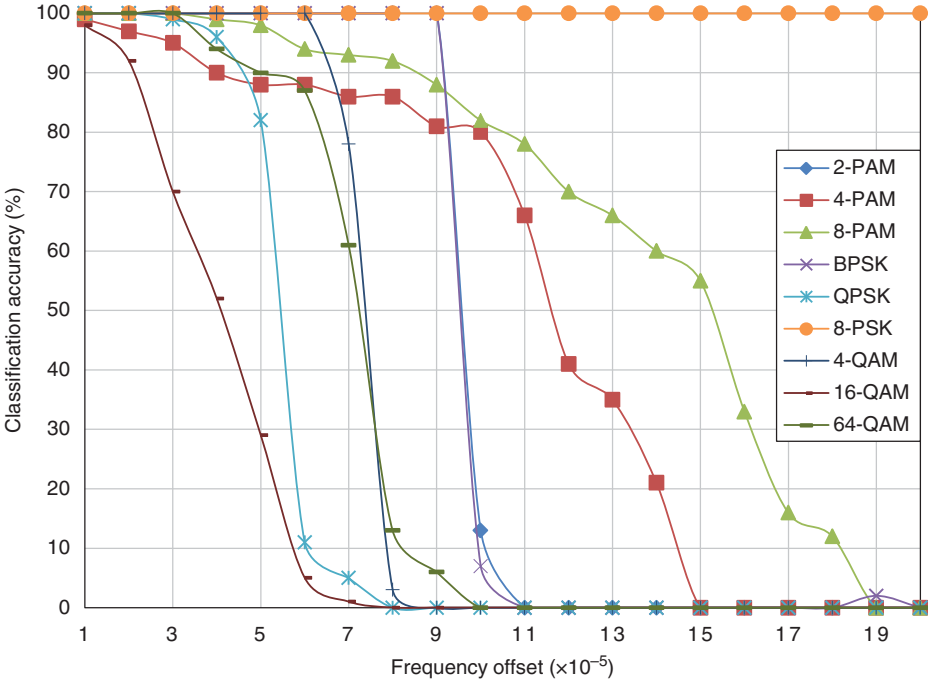



**Figure 8.27** Classification accuracy of the GP-KNN classifier with frequency offset. (*See insert for color representation of the figure.*)

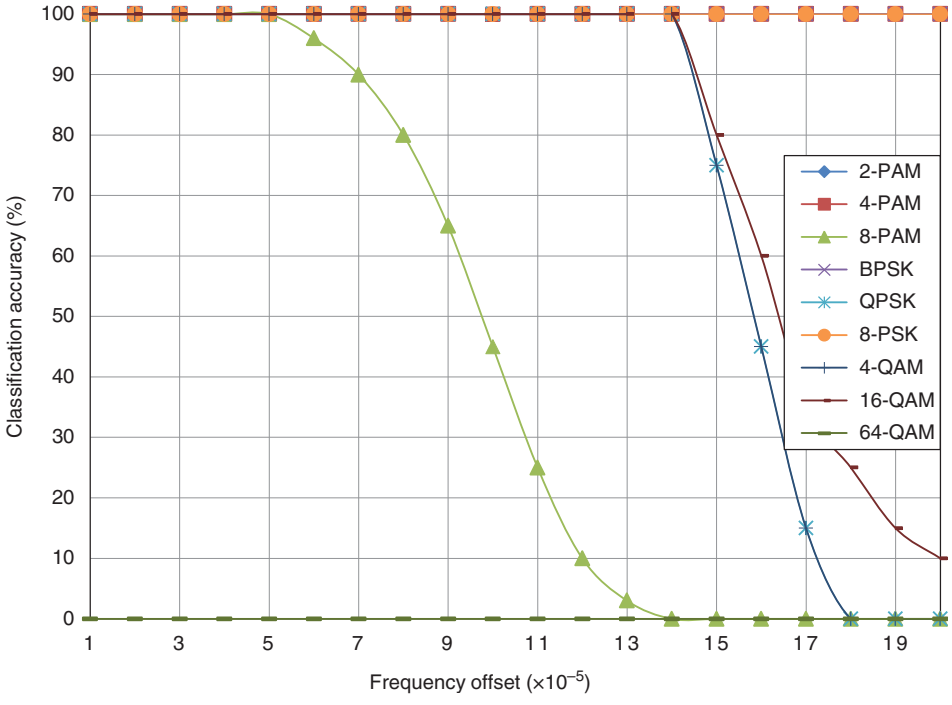



**Figure 8.28** Classification accuracy of the EM-ML classifier with frequency offset.

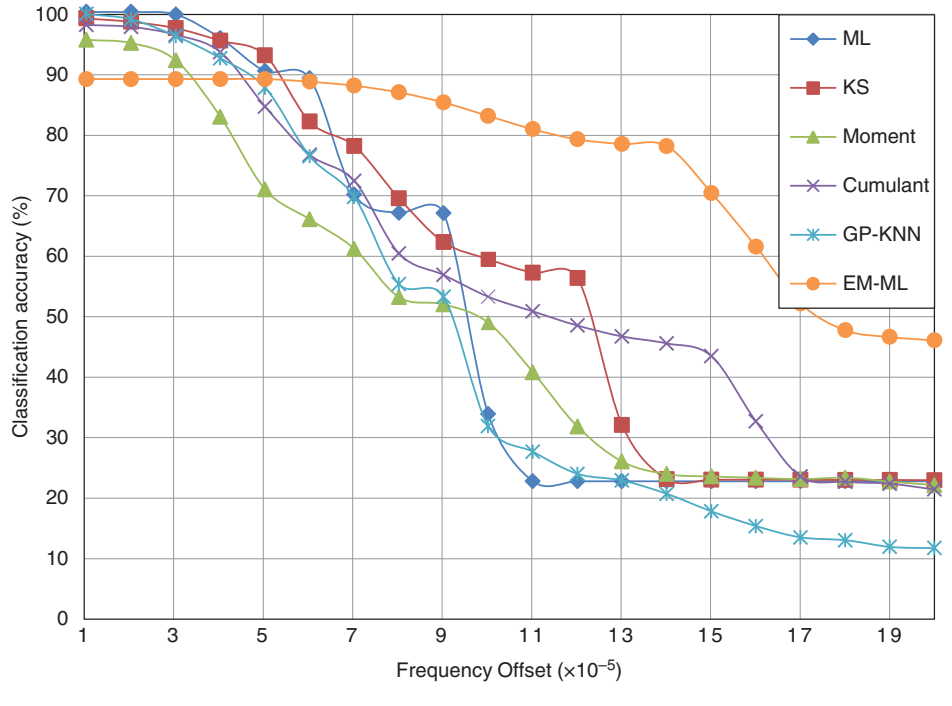



**Figure 8.29** Average classification accuracy of each classifier with frequency offset. (*See insert for color representation of the figure.*)

a frequency offset of $3 \times 10^{-5}$, the EM-ML is able to sustain a consistent level of classification accuracy until more than $1.4 \times 10^{-4}$ frequency offset is considered.

## 8.7 Computational Complexity

The computational complexity of a modulation classifier affects the overall performance of a system in three different ways. First, a complex classifier requires a more powerful processing unit in order to obtain an AMC decision. It imposes a string of limitations on the hardware design. Power supply, cooling system, device form factor, cost, compatibility of other units will all be affected. Secondly, a complex classifier requires more time to process each signal frame. For some applications, classification accuracy is the priority and the processing time could be compromised. However, for some time-critical applications, classifiers with low computational complexity are favoured. Thirdly, a complex classifier requires more power to complete an AMC task. This is especially so for mobile communication units where batteries are used to power the processing units. Clearly, a classifier with lower computational complexity with good accuracy is much desired.

From Chapters 3–7 we presented the implementation of many modulation classifiers. It should give an impression of the complexity of each classifier based on these materials. To provide a clear overview of how these classifiers compare with each other in terms of computational complexity we produced Table 8.6, detailing the number of mathematical operations needed to complete the classification of one piece of signal. To enable the calculation of the operation numbers, we make the following assumptions: (i) the modulation candidate pool has $I$ number of modulation candidates, (ii) the number of symbols of the $i$th candidate modulation is given by $M_i$, (iii) for each signal realization there are $N$ number of signal samples available for analysis, (iv) the number of reference samples for KNN and GP-KNN classifier is given by

**Table 8.6** Number of different operations needed for each classifier

| | Addition | Multiplication | Exponentiation | Logarithm |
|---|---|---|---|---|
| ML | $6NI\sum_{i=1}^{I} M_i$ | $5NI\sum_{i=1}^{I} M_i$ | $NI\sum_{i=1}^{I} M_i$ | $NI$ |
| KS test | $2N(\log 2N + 2I)$ | 0 | 0 | 0 |
| Moments | $10LN + 10N$ | $37LN + 37N$ | 0 | 0 |
| Cumulants | $10LN + 10N + 20L$ | $37LN + 37N + 44L$ | 0 | 0 |
| GP-KNN | $10LN + 10N + 20L + LG + G$ | $37LN + 37N + 44L + LG + G$ | $LG + G$ | $LG + G$ |
| EM-ML | $6KNI\sum_{i=1}^{I} M_i$ | $5KNI\sum_{i=1}^{I} M_i$ | $KNI\sum_{i=1}^{I} M_i$ | $KNI$ |

$L$, (v) the GP-evolved feature combination contains $G$ numbers of each operator under a general assumption, and (vi) the EM estimation stage is repeated for $K$ iterations for the EM-ML classifier.

For the ML classifier, the evaluation of the likelihood for each modulation hypothesis is the most costly process in terms of computation. All the operations are involved in the likelihood. Moreover, they are repeated for each modulation hypothesis, each modulation state and each signal sample. For the EM-ML classifier, the process is then again repeated for a certain number of iterations. Consequently, the computational complexity of the EM-ML classifier is the highest. For the moment- and cumulant-based features, the calculations involve only multiplication and addition. Therefore they are relatively simple. When employing GP for feature selection and combination, the initial investment in the training stage is rather expensive. However, it should be pointed out that once the features are selected and combined, the training step does not need to be repeated for each testing signal realization. In the testing stage, the GP-KNN classifier has a slightly higher computational complexity than does the basic KNN classifier using the same features. The one-sample KS test classifier is known to have lower computational complexity. The construction of the empirical distribution requires only multiplication. However, the requirement of reference theoretical CDF values requires additional calculation. If the reference values are prepared, extra space are needed from the system memory to store all these values.

## 8.8 Conclusion

In this chapter, we analyzed all the classifiers from Chapters 3–7 in terms of their versatility in different system configurations. It is clear that the requirement for channel parameters varies among most classifiers. The best classifier that is able to classify most digital modulations while not needing much prior knowledge of the communication system is the EM-ML classifier. The more important factor of classification accuracy is investigated in Sections 8.3–8.6. Channel conditions including AWGN noise, phase offset, and frequency offset, as well as the limitation of reduced signal sample size, are simulated to test the performance of the ML classifier, the KS test classifier, the moment- and cumulant-based classifiers, the GP-KNN classifier and the EM-ML classifier. Given a matching model, the ML classifier significantly outperforms the rest of the classifiers, when AWGN noise is considered. It also shows higher robustness given a limited number of samples for analysis. However, in complex channels with phase offset and frequency offset, all classifiers experience significant degradation in their classification performance, with the exception of the EM-ML classifier. Thanks to its estimation stage, the phase offset can be completely compensated by the estimation channel coefficient. While not being estimated and compensated, frequency offset also has less impact for the EM-ML classifier. While being versatile and robust, the disadvantage of the ML and EM-ML classifiers is

exposed in Section 8.7, where the computation complexity is evaluated for each classifier. The likelihood-based classifiers all require a high number of exponentiation and logarithms, while these operations are not need by the other classifiers. The classifiers with the least amount of computation requirement are the distribution test-based classifiers, which require no exponentiation or logarithm operations and need a lower number of additions or multiplications.

## References

Swami, A. and Sadler, B.M. (2000) Hierarchical digital modulation classification using cumulants. *IEEE Transactions on Communications*, **48** (3), 416–429.

Wang, F. and Wang, X. (2010) Fast and robust modulation classification via Kolmogorov–Smirnov test. *IEEE Transactions on Communications*, **58** (8), 2324–2332.

Wei, W. and Mendel, J.M. (2000) Maximum-likelihood classification for digital amplitude-phase modulations. *IEEE Transactions on Communications*, **48** (2), 189–193.

# 9

# Modulation Classification for Civilian Applications

## 9.1 Introduction

In previous chapters we have taken a theoretical view of modulation classification. The listed classifiers have mostly been developed based on general assumptions. The performance comparison in Chapter 8 gives a general impression of how each classifier performs in different scenarios.

In this chapter we visit some of the civilian communication systems with special AMC requirements that have not been covered previously. First, we investigate systems where higher-order modulations are deployed. Secondly, the application of modulation classifiers in link adaptation is revised with specific classifiers designed to exploit some of the properties of the system. Lastly, the case of blind modulation classification in MIMO systems is discussed.

## 9.2 Modulation Classification for High-order Modulations

In real world communication systems there are many instances where high-order modulations are employed for high data rate transmission. These systems mostly rely on wired transmission links to achieve ultrahigh spectrum efficiency. Among them, 64-QAM and 256-QAM are used for digital terrestrial television and its high-definition version. Broadband over power line (BPL) uses 1024-QAM and 4096-QAM modulations. To design a classifier for high-order modulation over wired channel, we need to understand the following characteristics of these systems.

*Automatic Modulation Classification: Principles, Algorithms and Applications*, First Edition. Zhechen Zhu and Asoke K. Nandi.

First, the wired channel is relatively stable compared with wireless channels. Therefore, the classifier does not need to adapt to the change of channel conditions. In other words, the channel estimation can be implemented at the initialization of the system. The estimated channel parameters can be used in the remainder of the transmission. Secondly, the high-order modulations are normally more difficult to classify, as concluded in Chapter 8. Therefore, to guarantee successful classification, a classifier of high classification accuracy is needed. Among all the modulation classifiers, the ML classifier, distribution test classifier and the higher-order cumulant features are reported to have superior performance for high-order modulations. Thirdly, because of the large number of symbol states, in a single piece of signal frame, it cannot be guaranteed that the symbol assignment will be equally probable. In certain cases some of the symbol states will not be used in the transmission. This condition has a significant impact on the distribution test-based classifiers. For a KS test classifier, when some of the modulation symbol is not observed in the received signal, the empirical distribution function will exhibit big differences at the missing symbols, since in the theoretical CDF these symbol positions normally possess higher probability values. For higher-order cumulant-based classifiers, it is obvious that the signal length is of great importance. From Figure 8.17 we can see that among three of the QAM modulations considered in the simulation, the higher-order QAM modulations make a higher impact when limited signal length is available for analysis. When the modulation is increased to 1024-QAM or 4096-QAM, it is clear that the higher-order cumulants will not be able to provide the demanded classification accuracy. On the other hand, the ML classifier has suggested superior performance for both high-order modulation as well as limited signal length. Therefore, the ML classifier is the ideal option for the classification of high-order modulations in wired communication systems.

With the above analysis, we construct the AMC solution in a wired communication system using high-order modulations. The overall process is illustrated in Figure 9.1. The heaver dashed line indicates the transmission of pilot samples to the channel estimator in system initialization. Thanks to the pilot samples, many state-of-the-art channel estimators can be used to acquire the CSI rather easily. The thinner dashed line indicates the exchange of CSI and modulation candidate pool from the channel estimator to the ML classifier. The CSI and information of the modulation candidate pool are then saved in the ML classifier to assist modulation classification. They are not updated until the system is reconfigured. During normal transmission, the signals will be transmitted through the wired channel indicated by the heavy solid line. The ML classifier will analyze each signal and evaluate the corresponding likelihood for each modulation candidate in the pool using equation (3.3). When the likelihood values for each modulation hypothesis are obtained, the classification decision can be easily reached using the maximum likelihood criteria given in equation (3.13). The demodulator is then informed of the modulation type of the received signal from the ML classifier. If the classification is correction, the transmission is successfully completed.

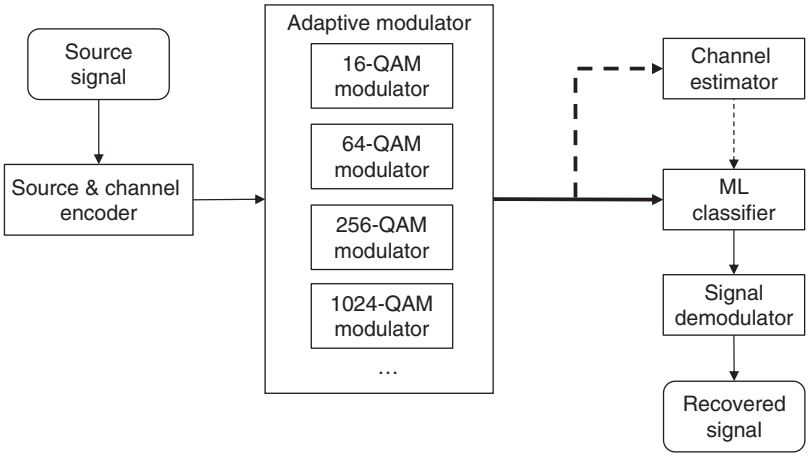



**Figure 9.1** AMC in wired system with high-order modulation.

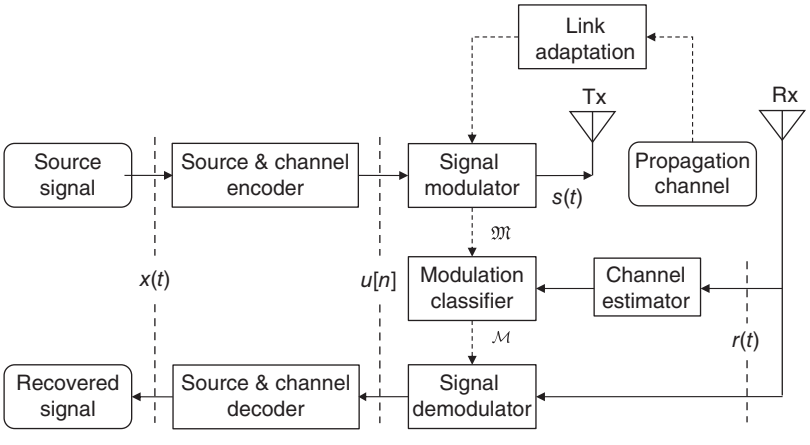



**Figure 9.2** AMC in wired system with link adaptation.

## 9.3 Modulation Classification for Link-adaptation Systems

Link adaptation, also known as adaptive modulation and coding, adaptively selects the modulation method depending on the channel conditions. The mechanism of the LA system with AMC has been briefly discussed in Chapter 1. The system configuration is illustrated in Figure 9.2.

Since the selection of the modulation scheme depends on the channel conditions, the completed posteriori probability of the signal being received with a modulation type

$\mathcal{M}$ can be expressed as a combination of the prior probability of symbol $s_m$ being transmitted and the likelihood of the received signals belonging to the transmitted symbol, as shown in equation (9.1).

$$q(r,s|\Theta) = \int_s p(s_m, r|\Theta) \log p(s_m|\Theta)) ds \tag{9.1}$$

The prior probability can be automatically achieved before the next batch of signals is transmitted. The incorporation of the prior probability in a modulation classifier is most natural for a likelihood-based classifier. Depending on the implementation, the link adaptation could depend on the estimated channel or the acknowledgement and negative acknowledgement protocol. However, this is outside the scope of this book. We assume that the prior probability with regard to the channel information is known to the classifier. Instead of the maximum likelihood criterion one can adopt the maximum a posteriori (MAP) criterion to accommodate the added information of modulation selection prior probability. The application of MAP in modulation classification was first suggested by Haring, Chen and Czylwik (2010). The log MAP likelihood could be calculated using equation (9.2).

$$\log \mathcal{L}_{MAP}(R|S_{\mathcal{M}}, \Theta) = \sum_{n=1}^{N} \log \left( \sum_{m=1}^{M} \frac{1}{M} \frac{1}{2\pi\sigma^2} e^{-\frac{|r[n]-hs_m|^2}{2\sigma^2}} \right) + \log p(S_{\mathcal{M}}|\Theta) \tag{9.2}$$

The classification decision can be given by finding the modulation candidate with the highest a posteriori likelihood, as determined by equation (9.3).

$$\hat{\mathcal{M}} = \underset{\mathcal{M} \in \mathfrak{M}}{\arg\max} \left( \log \mathcal{L}_{MAP}(R|S_{\mathcal{M}}, \Theta_{\mathcal{M}}) \right) \tag{9.3}$$

## 9.4 Modulation Classification for MIMO Systems

Multiple-input and multiple-output (MIMO) systems have been the key technology in many recent communications innovations. A MIMO system employs spatially differing transmission and receiving antenna arrays. MIMO systems enable the transmission of multiple signal streams through spatially different signal paths, which is known as spatial multiplexing (SM). The diversity of the signal paths also provides the possibility of improved link reliability. When space-time coding (STC) is used, multiple versions of the signal symbol are transmitted simultaneously. With accurate channel knowledge, the received signals can be recovered more accurately.

MIMO systems differ from most modulation classifiers due to the multiple paths between the transmission antenna arrays and the receiving antenna arrays. Since each

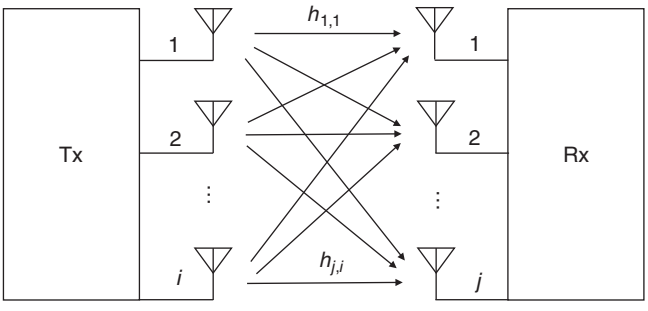



**Figure 9.3** MIMO channel.

receiver provides a mixture of single symbols from all transmitters, the modulation classification approach in single-input and single-out system cannot be used any more. The existing modulation classifiers must be modified to meet the new requirements as shown in Figure 9.3.

The MIMO systems are composed of $N_i$ transmitting antennas and $N_r$ receiving antennas. A Rayleigh fading channel with time-invariant path gains is considered. The resulting channel matrix $H$ is given by an $N_r \times N_i$ complex matrix with the element $h_{j,i}$ representing the path gain between $i$th transmitting antenna and $j$th receiving antenna. Assuming perfect synchronization, the $n$th received MIMO-SM signal sample vector $r_n = [r_n(1), r_n(2), \ldots, r_n(N_r)]^T$ in a total observation of $N$ samples is expressed as given by equation (9.4),

$$r_n = Hs_n + \omega_n \tag{9.4}$$

where $s_n = [s_n(1), s_n(2), \ldots, s_n(N_t)]^T$ is the $n$th transmitted signal symbol vector and $\omega_n = [\omega_n(1), \omega_n(2), \ldots, \omega_n(N_r)]^T$ is the additive noise observed at the $n$th signal sample. The transmitted symbol vector is assumed to be independent and identically distributed with each symbol assigned from the modulation alphabet with equal probability. The additive noise is assumed to be white Gaussian with zero mean and variance $\sigma^2$, which gives $\omega_n \in \mathcal{N}\left(0, \sigma^2 I_{N_r}\right)$, where $I_{Nr}$ is the identity matrix of size $N_r \times N_r$.

From an ML classifier, Choqueuse *et al.* developed the MIMO version of the likelihood-based classifier (Choqueuse *et al.*, 2009), with an updated version of the likelihood function, given in equation (9.5),

$$\mathcal{L}(R|\Theta) = \prod_{n=1}^{N} \frac{1}{M} \sum_{m=1}^{M} \frac{1}{\left(\pi\sigma^2\right)^{N_r}} \exp\left(-\frac{\|r_n - Hs_m\|_F^2}{2\sigma^2}\right) \tag{9.5}$$

where $\|\cdot\|_F^2$ is the Frobenius norm. The possible transmitted symbol set $\mathbf{S} = [S_1, S_2 \ldots S_M]$ gathers all the combinations of transmitted symbols from $N_i$ number of antennas. Given a modulation with $L$ number of states, there exist $M = L^{N_t}$ number of transmitted

symbol vectors and a transmitted symbol set of size $N_t \times L^{N_t}$. The channel matrix and noise variance are assumed to be known. The corresponding log-likelihood function can be derived as equation (9.6).

$$\log\mathcal{L}(R|\Theta) = \sum_{n=1}^{N} \log\left\{ \frac{1}{M} \sum_{m=1}^{M} \frac{1}{(\pi\sigma^2)^{N_r}} \exp\left( -\frac{\|r_n - Hs_m\|_F^2}{2\sigma^2} \right) \right\} \tag{9.6}$$

$$\mathcal{L}(R|\Theta) = \prod_{n=1}^{N} \frac{1}{M} \sum_{m=1}^{M} \frac{1}{(\pi\sigma^2)^{N_r}} \exp\left( -\frac{\|r_n - Hs_m\|_F^2}{2\sigma^2} \right)$$

The subsequent classification decision can be made by finding the modulation candidate which provides the highest likelihood, as calculated using equation (9.7).

$$\widehat{\mathcal{M}} = \underset{\mathcal{M}\in\mathfrak{M}}{\arg\max} \left( \log\mathcal{L}(R|S_{\mathcal{M}}, \Theta_{\mathcal{M}}) \right) \tag{9.7}$$

The method provides high classification accuracy, but with a rather high computational complexity. Alternative classifiers with reduced complexity have been proposed in several publications (Kanterakis and Su, 2013; Mühlhaus *et al.*, 2013). Kanterakis and Su suggested treating independent component analysis (ICA) recovered signal components as independent processes. The recovered source signal components $s_n^s = \left[ s_n^s(1), s_n^s(2), \ldots, s_n^s(N_t) \right]$ are acquired by using equation (9.8),

$$s_n^s = \hat{H}^{-1} r_n = \hat{H}^{-1} H s_n + \hat{H}^{-1} \omega_n \tag{9.8}$$

where $\hat{H}^{-1}$ is the inverse of the estimated channel matrix. For each independent transmitted signal stream, the updated likelihood function is given by equation (9.9).

$$\mathcal{L}(s^s(i)|\Theta) = \prod_{n=1}^{N} \frac{1}{M} \sum_{m=1}^{M} \frac{1}{\pi\sigma^2} \exp\left( -\frac{\left| s_n^s(i) - w(i) H s_n \right|}{2\sigma^2} \right) \tag{9.9}$$

In addition to the likelihood-based classifier, Choqueuse *et al.* also suggested the use of independent component analysis for the estimation of the channel matrix when it is not readily known. Cardoso and Souloumiac's JADE algorithm was suggested (Cardoso and Souloumiac, 1993). The resulting estimation of channel matrix has a phase shift and needs phase correction. The ICA estimation provides channel matrix estimation but the estimation for noise variance is still lacking. In addition, there are limits to the MIMO arrangement that can be considered. When the number of transmitting antennas exceeds the number of receiving antennas, the ICA estimation

cannot be implemented. To overcome these issues related to an ICA-aided modulation classifier for the MIMO system, we suggest an expectation maximization estimator for more practical implementation of the likelihood-based classifier for MIMO systems.

To evaluate likelihood for the ML classifier, the complex channel matrix $H$ and noise variance $\sigma^2$ must be estimated beforehand. Since the modulation is unknown to the receiver, many data-aided approaches using pilot symbols are not suitable. Expectation maximization has been employed for joint channel estimation through an iterative implementation of maximum likelihood estimation (Wautelet *et al.*, 2007; Das and Rao, 2012). In MIMO systems, we consider the received signal $R = [r_1, r_2 \ldots r_N]$ as the observed data. Meanwhile, the membership $Z$ of the observed samples is considered as the latent variable. $Z$ is an $M \times N$ matrix with the $(m,n)$th element being the membership of the $n$th signal sample $\mathbf{r}_n$ , given the transmitted symbol vector $S_m$. The possible transmitted symbol set $\mathbf{S} = [S_1, S_2 \ldots S_M]$ gathers all the combinations of transmitted symbols from $N_t$ number of antennas. Given a modulation with $L$ number of states, there exist $M = L^{N_t}$ number of transmitted symbol vectors and a transmitted symbol set of size $N_t \times L^{N_t}$. With $\Theta = \{H, \sigma^2\}$ representing the channel parameters, the complete likelihood is given by equation (9.10),

$$q(R,S|\Theta) = \int_S p(S|R,\Theta)\log(p(R|S,\Theta)p(S|\Theta))dS \tag{9.10}$$

where $p(R \mid S, \Theta)$ is the probability of the received signal being observed given transmitted symbols vector $S$ and channel parameter $\Theta$. Since the additive noise is assumed to have a complex Gaussian distribution, $p(R \mid S, \Theta)$ can be calculated as given in equation (9.11).

$$p(R|S,\Theta) = \prod_{n=1}^{N} \frac{1}{(\pi\sigma^2)^{N_r}} \exp\left(-\frac{\|r_n - Hs_n\|_F^2}{2\sigma^2}\right) \tag{9.11}$$

Meanwhile, $p(S \mid R, \Theta)$ represents the probability of $S$ being transmitted given the observed signal $R$ and the channel parameter $\Theta$. In Wautelet *et al.* (2007), this probability is acquired by an a posteriori probability computation which is not presented. In this book, we replace the a posteriori probability with a soft membership $z_{nm}$ representing the likelihood of the $n$th transmitted symbol vector being $S_m$, with $\sum_{m=1}^{M} z_{mn} = 1$. Since assignment of the transmitted symbol is independent of the channel parameter, $p(S \mid \Theta)$ is a constant, $1/M$, when equal probability is assumed. The estimation of $\Theta$ is achieved by iterative steps of expectation evaluation and maximization.

The evaluation step (E-step) provides the expected log-likelihood under the current estimate of $\Theta^t$ at the $t$th iteration. The expectation is then subsequently maximized for

the updated estimation of $\Theta$. From equation (9.11) the expected value of the complete log-likelihood is derived as shown in equation (9.12),

$$\mathcal{Q}(R,S|\Theta^t) = \log\prod_{n=1}^{N}\prod_{m=1}^{M} p\left(r_n,S_m|H_t,\sigma_t^2\right)^{z_{mn}} = -\sum_{n=1}^{N}\sum_{m=1}^{M} z_{mn}\left[N_r\log\left(\pi\sigma_t^2\right) + \frac{\|r_n - H_t S_m\|_F^2}{\sigma^2}\right] \tag{9.12}$$

where $p(r_n,S_m|H_t,\sigma_t)$ is the probability of the $n$th received signal vector being observed given the current estimation of channel matrix $H_t$ and noise variance $\sigma_t^2$. The soft membership $z_{nm}$ is evaluated equation (9.13).

$$z_{nm} = \frac{p(r_n|S_m,\Theta^t)}{\sum_{m=1}^{M} p(r_n|S_m,\Theta^t)} = \frac{\exp\left(-\frac{\|r_n - HS_m\|_F^2}{\sigma^2}\right)}{\sum_{m=1}^{M}\exp\left(-\frac{\|r_n - HS_m\|_F^2}{\sigma^2}\right)} \tag{9.13}$$

The update of the parameter estimation is achieved through the maximization of the current expected log-likelihood (M-step). To derive the close form update function for the channel matrix and noise variance, we first find the derivatives of $Q(R,S\,|\,\Theta^t)$ with respect to $H$ and $\sigma^2$ separately. Given that equation (9.14) holds, then

$$\|r_n - HS_m\|_F^2 = \sum_{j=1}^{N_r}\left|r_n(j) - \sum_{i=1}^{N_t} h_{j,i}S_m(i)\right|^2 \tag{9.14}$$

the derivative of $Q(R,S|\Theta^t)$ with respect to the individual element $h_{j,i}$ of the channel matrix is given by equation (9.15).

$$\frac{\partial\mathcal{Q}(R,S|\Theta^t)}{\partial h_{j,i}} = -\sum_{n=1}^{N}\sum_{m=1}^{M} z_{mn}\frac{\sum_{i=1}^{N_t} h_{j,i}{}^*|S_m(i)|^2 - r_n(j)^*S_m(i)}{\sigma^2} \tag{9.15}$$

In the same way, the derivative of $Q(R,S|\Theta^t)$ with respect to the noise variance $\sigma^2$ is found as shown in equation (9.16).

$$\frac{\partial\mathcal{Q}(R,S|\Theta^t)}{\partial\sigma^2} = -\sum_{n=1}^{N}\sum_{m=1}^{M} z_{mn}\left(-\frac{N_r}{\sigma^2} + \frac{\|r_n - HS_m\|_F^2}{\sigma^4}\right) \tag{9.16}$$

When the derivatives are set to zero, the update functions of $h_{j,i}$ and $\sigma^2$ can be derived from equations (9.15) and (9.16). However, it is obvious that different channel parameters are coupled. To simplify the maximization process, the coupled channel parameters are estimated in turn. The path gain $h_{j,i}$ is estimated with the rest of the channel matrix known and is represented with the latest estimate for each path gain. The path gains are updated in ascending order with respect to $j$ and $i$. The resulting update function for $h_{j,i}$ is given by equation (9.17),

$$h_{j,i}^{t+1} = \frac{\sum_{n=1}^{N}\sum_{m=1}^{M} z_{mn}\left[r_n(j)S_m(i)^* - S_m(i)^*\sum_{k=1,k\neq i}^{N_t} h_{k,i}S_m(k)\right]}{\sum_{n=1}^{N}\sum_{m=1}^{M} z_{mn}|S_m(i)|^2} \tag{9.17}$$

where $h_{k,i}$ is the latest estimate of path gain. At the $t$th iteration, $h_{k,i} = h_{k,i}^{t}$ if it has not been updated, or $h_{k,i} = h_{k,i}^{t+1}$ if it has been updated. After the channel matrix is completely updated, $H_{i+1}$ is used to acquire the noise variance estimation, given in equation (9.18).

$$\sigma_{i+1}^2 = \frac{\sum_{n=1}^{N}\sum_{m=1}^{M} z_{mn}\sum_{j=1}^{N_r}\left|r_n(j) - \sum_{i=1}^{N_t} h_{j,i}^{t+1}S_m(i)\right|^2}{N_r\sum_{n=1}^{N}\sum_{m=1}^{M} z_{mn}} \tag{9.18}$$

The EM algorithm with such a maximization process is known as expectation condition maximization. ECM shares the convergence property of EM (Meng and Rubin, 1993) and can be constructed to converge at a similar rate as the EM algorithm (Sexton, 2000). The ECM joint estimation of channel parameters has previously been successfully applied in BMC for SISO systems (Chavali and Da Silva, 2011, 2013; Soltanmohammadi and Naraghi-Pour, 2013).

The final estimation of channel matrix $H$ and noise variance $\sigma^2$ is achieved when the iterative process is terminated by one of two conditions. The first condition terminates the process when the estimation reaches convergence. The condition is represented numerically with the difference between the expected likelihoods of the current iteration and the previous iteration along with a predefined threshold. In the second condition, termination is triggered when the predefined number of iterations has been reached. The estimate at termination is used to replace the known channel matrix and noise variance in equation (9.6) and classification decision to be made using equation (9.7). An illustration of the system is given in Figure 9.4.

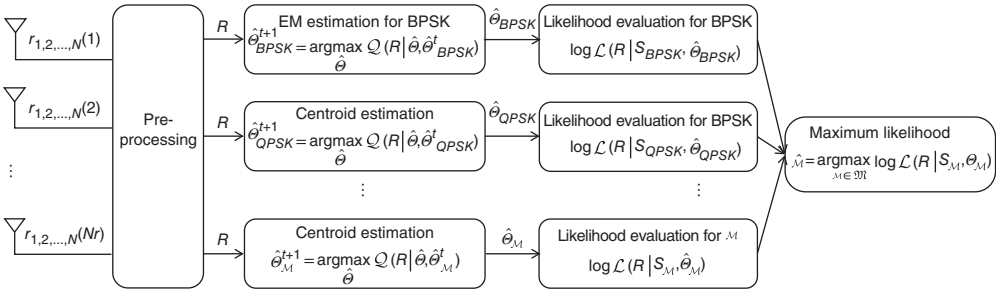



**Figure 9.4** AMC in MIMO system using EM-ML classifier.

## 9.5 Conclusion

In this chapter, three civilian communication systems, which pose their unique requirement on modulation classification, are considered for the practical design of some modulation classifiers. The high-order modulation classification challenge is addressed for wired communication systems with associated channel effect. The knowledge of system configuration in a link adaptation system is exploited for improved modulation classification performance through a maximum a posteriori classifier. The update issue of modulation classification for MIMO systems is solved by an EM-ML classifier. Updated likelihood evaluation and channel estimation are both presented for MIMO systems.

## References

Cardoso, J.F. and Souloumiac, A. (1993) Blind beamforming for non-gaussian signals. *IEE Proceedings – Radar and Signal Processing*, **140** (6), 362–370.

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

Chavali, V.G. and Da Silva, C.R.C.M. (2013) Classification of digital amplitude-phase modulated signals in time-correlated non-gaussian channels. *IEEE Transactions on Communications*, **61** (6), 2408–2419.

Choqueuse, V., Azou, S., Yao, K. *et al*. (2009) Blind modulation recognition for MIMO systems. *Military Technical Academy Review*, **XIX** (2), 183–196.

Das, A. and Rao, B.D. (2012) SNR and noise variance estimation for MIMO systems. *IEEE Transactions on Signal Processing*, **60** (8), 3929–3941.

Haring, L., Chen, Y. and Czylwik, A. (2010) Automatic modulation classification methods for wireless OFDM systems in TDD mode. *IEEE Transactions on Communications*, **58** (9), 2480–2485.

Kanterakis, E. and Su, W. (2013) *Modulation Classification in MIMO Systems*. Military Communications Conference, San Diego, CA, USA, 18 November 2013, pp. 35–39.

Meng, X.-L. and Rubin, D.B. (1993) Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika*, **80** (2), 267–278.

Mühlhaus, M.S., Oner, M., Dobre, O.A. and Jondral, F.K. (2013) A low complexity modulation classification algorithm for MIMO systems. *IEEE Communications Letters*, **17** (10), 1881–1884.

Sexton, J. (2000) ECM algorithms that converge at the rate of EM. *Biometrika*, **87** (3), 651–662.

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

Wautelet, X., Herzet, C., Dejonghe, A. *et al*. (2007) Comparison of EM-based algorithms for MIMO channel estimation. *IEEE Transactions on Communications*, **55** (1), 216–226.

# 10

# Modulation Classifier Design for Military Applications

## 10.1 Introduction

The primary purpose of modulation classification in military scenarios is to ascertain information about an adversary by intercepting radiated energy. The detection and subsequent identification of the signal modulation have three possible uses. First, the knowledge of the modulation scheme can be directly used for the identification of the unit transmitting the signal. Secondly, if decryption and translation is added to the rear end of the modulation classifier and the demodulator, the transmitted message can be recovered if the modulation scheme can be classified. Thirdly, when jammers are deployed to disrupt transmission between adversary units, modulation classification is required to match the modulation of jamming signals and adversary signals.

Jamming is an important part of the electronic attack strategy (Poisel, 2008). The purpose is to override the adversary's communication by introducing a man-made signal in the same communication channel. As the adversary receivers are matched with the transmitter, only the signal of the matching property will be picked up by the adversary receiver. In practice, a jamming signal must be transmitted using the same modulation technique and have the same carrier frequency. In addition, to override the adversary communication, the jamming signal must have higher power to hide the intended transmission. For a modulation classifier, the presence of a jamming signal has no negative effect on the classification performance. In contrast, given that

*Automatic Modulation Classification: Principles, Algorithms and Applications*, First Edition. Zhechen Zhu and Asoke K. Nandi.

the jamming signal has higher power, regardless of the signal being analyzed, the SNR is increased compared with the scenario without the presence of a jamming signal. As demonstrated in Chapter 8, this often improves the classification accuracy. Since there is no new challenge posed by a jamming signal, it is not necessary to develop any specific modulation classification strategy for jammed signals. Most classifiers should suffice in this case.

In the remainder of this chapter we first establish a modulation classifier that narrows down the possible modulation types to be considered in the candidate pool. Next, we focus on classification of the signals transmitted using two different low-probability-of-detection techniques.

## 10.2 Modulation Classifier with Unknown Modulation Pool

Surveillance and threat analysis in electronic warfare monitors the radio transmission channels for communication between an adversary's units. Once communication is detected at a certain frequency band, modulation classification is one of the initial processes required for the retrieval of the transmitted message. Different from civilian communication systems, in military scenarios there is no known modulation candidate pool to assist the classifier. In fact, modulation can be considered as a layer of decryption which prevents the message from being recovered by the adversary unless successful modulation classification is performed. In other words, the adversary is inclined to use any modulation scheme to increase the difficulty of modulation classification.

Historically, modulation classifiers are mostly defined for a finite set of modulation schemes. While it does not provide a direct solution to our problem, it is possible to study the properties of these modulation classifiers and to exploit the true capability of these classifiers. It is unlikely that one classifier will be able to classify all modulation types. However, it is possible to use a few classifiers to perform multi-stage classification. The idea is to classify the signal into general modulation types. The subsequent stage will be accomplished by a different classifier which is more suitable for separation within the specific modulation category.

In this case, we consider seven major types of modulation; namely AM, FM, PAM, ASK, FSK, PSK, and QAM. For digital modulations, that is PAM, ASK, FSK, PSK, and QAM, the modulations of different order are not considered in this section. The classifier aims to narrow down the modulation scheme to an individual type of modulation. If the aforementioned digital modulations are detected, the information of the modulation type can be further processed by another classifier which specialized in the classification of modulations of different order and the same type.

When a piece of unknown single is intercepted, the first step of the proposed scheme is to measure if the phase of the signal has a high variance. Signals such as AM and ASK have constant phase. Therefore, the variance of their signal phase should be very

low. On the other hand, other modulations all have relatively high variance in their signal phase. To achieve the classification, we adopt the standard deviation of the absolute value of the instantaneous signal phase as the feature in the first stage to differentiate AM and ASK from FM, PAM, FSK, PSK, and QAM. The feature $\sigma_{dp}$ is covered in Chapter 5 and the threshold $t(\sigma_{dp})$ can be selected heuristically and optimized according to the channel conditions.

In the first group, to differentiate AM and ASK, the probability density function of their amplitude envelope can be exploited. For AM, being an analogue modulation, the possible state of the amplitude envelope has no prior restriction. Therefore, its PDF is much flatter. In contrast, ASK modulation, depending on the modulation order, have several states in its amplitude envelope representing transmitted symbols in digital communication. Regardless of the noise model, the underlying PDF of ASK signal should have spikes at the mapped amplitude values. The differences between their PDF can be presented numerically with kurtosis and more specifically the fourth-order moment of the signal. This feature has also been listed in Chapter 5. If the fourth-order cumulant $\mu_{42}^{a}$ of the intercepted signal is higher than the pre-defined threshold $t\left(\mu_{42}^{a}\right)$, the signal is classified as AM. Accordingly, if the feature value is below the threshold, the signal is classified as ASK modulation.

In the second group, we can also investigate the amplitude envelope for further classification. It is obvious that FM or FSK modulations have no information in their signal amplitude. Given that the signal amplitude is constant, then, as discussed in Chapter 5, the feature $\gamma_{max}$ should be zero for FM and FSK. The other modulations in the second group, however, have information in either their signal amplitude or their signal phase. Consequently, the feature value $\gamma_{max}$ for these modulations would be non-zero. A pre-defined decision threshold $t(\gamma_{max})$ would enable a simple classification of these two subgroups.

The solutions for the separation of FM and FSK modulations are similar to that for the separation of AM and ASK modulations. The only difference is that instead of studying their amplitude envelope, the signal feature being evaluated is the signal frequency. While both modulations use signal frequency to convey information to be transmitted, being an analogue modulation the possible value of signal frequency is not limited to a few states in FM modulation. The resulting flat PDF of its signal frequency also suggests a lower kurtosis and fourth-order moment of the signal frequency. On the other hand the same feature, $\mu_{42}^{f}$, as covered in Chapter 5, should have higher values among all the FSK modulations. It is not difficult to see that the identification of each of the modulation types can be achieved with a pre-defined threshold $t\left(\mu_{42}^{f}\right)$.

In the second subgroup of PAM, PSK and QAM, the discrimination is not as easily achievable as for the other cases. The properties of different signal features have no distinctive difference between them. Some of the aforementioned features may display different values among these modulations. However, when added noise is considered,

these features do not provide an optimal solution. In this case, we employ the signal vector analyzer for extracting the in-phase and quadrature segments of the received signals. Given the reconstructed signal constellation using their in-phase and quadrature segments, it is easy to see some characteristic features of their symbol mappings. For PAM modulations, there are clear states of magnitude in their symbols mapping. However, there is only one phase for these symbols. For PSK modulations, there are different states in the symbol phases while the magnitude is constant for all symbols. For QAM modulations, both the phase and magnitude have multiple states. The classification of these three types of modulation draws inspiration from the two fourth-order moment features $\mu_{42}^{a}$ and $\mu_{42}^{f}$. In this case, we propose fourth-order moments of the reconstructed signal phase $\mu_{42}^{p}$ and magnitude $\mu_{42}^{m}$. The expressions for the two new features are given by equations (10.1) and (10.2), respectively,

$$\mu_{42}^{p} = \frac{E\{\arg(r[n])\}}{\{E\{\arg^2(r[n])\}\}^2} \tag{10.1}$$

and

$$\mu_{42}^{m} = \frac{E\{|r[n]|)\}}{\left\{E\left\{|r[n]|^2\right\}\right\}^2} \tag{10.2}$$

where $r[n]$ is the complex representation of the $n$th signal sample with $\arg(r[n])$ and $|r[n]|$ being its phase and magnitude. Based on the analysis of each modulation type, PAM should have a low $\mu_{42}^{p}$ value and a high $\mu_{42}^{m}$ value. PSK should have a high $\mu_{42}^{p}$

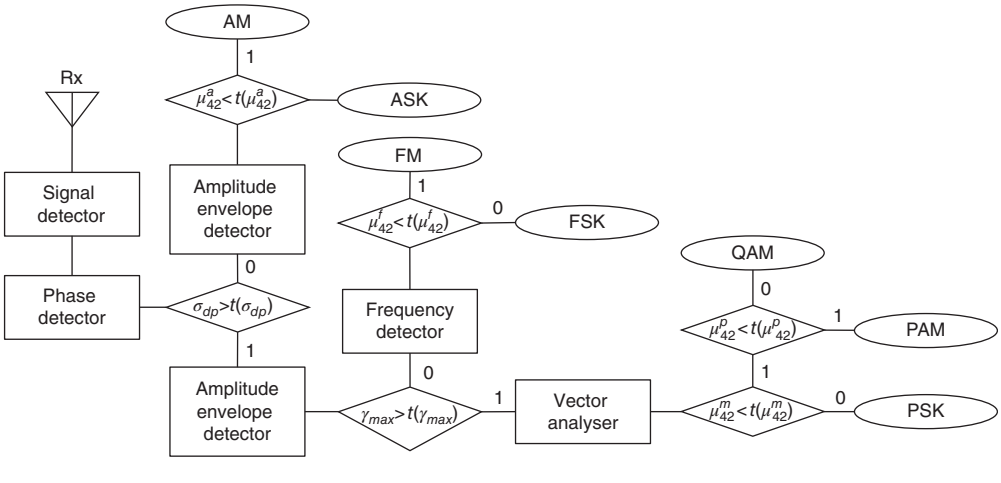



**Figure 10.1** Classification of unknown modulation pool.

value and a low $\mu_{42}^{m}$ value. QAM should have both a high $\mu_{42}^{p}$ value and a high $\mu_{42}^{m}$ value. Given some pre-optimized threshold, the classification decision can be found using the conditions shown in Figure 10.1.

## 10.3 Modulation Classifier against Low Probability of Detection

In the field of EP, it is often required to reduce the chance of the transmitted signal being detected by the adversary. There are two common techniques to achieve that. First, the detection probability can be reduced if the signal power is below the noise power. This is called direct sequence spread spectrum (DSSS). Secondly, the detection can be made difficult if the signal is transmitted with one or more alternating frequencies. This is called frequency-hopping spread spectrum (FHSS). Together with other techniques they form the strategy of low probability of detection (LPD) in military communication systems.

### *10.3.1 Classification of DSSS Signals*

It is easy to see that LPD poses specific challenges to the task of AMC. First, if the DSSS is employed by the transmitter, the SNR is likely to be below 0 dB. Therefore, only the classifiers with robust performance against high noise level can meet the requirement of AMC for DSSS. Therefore, we revisit Figure 8.8 in Chapter 8 to choose the candidates for this scenario. It is clear that when accurate channel state information is available the likelihood-based classifiers have superior classification accuracy and robustness. The drawback of high computational complexity may not be much of a concern given that the detection hardware employed normally does not have much limitation on computational power.

In this context, we limit the discussion to digital modulations, namely PAM, PSK, and QAM. If a wider range of modulations is considered the identification scheme proposed in Section 10.2 can be employed as a prior step of the following process. In the first step, we used the features $\mu_{42}^{p}$ and $\mu_{42}^{m}$ to achieve the classification of modulation type. It is necessary that there be a reduction of computational complexity for the following likelihood classifiers. In addition, performance enhancement may be attained since a more specific approach could be applied to each modulation type. One classification of modulation type is achieved through the process given in Section 10.2; the remaining challenge is the classification of modulations of the same type but different orders.

For PAM modulations, as the signal distribution is unique among different orders thanks to different symbol mappings in their magnitude, the likelihood analysis can be limited to the magnitude of the signal. To make the matter more convenient, it is of interest to convert the signal to a zero-phase one, so that all the symbols have only in-phase segments. According to Section 2.2.2, the distribution of the signal on their

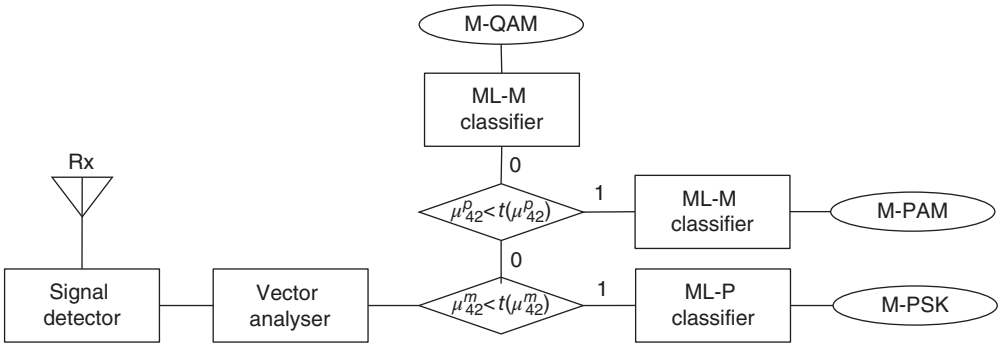



**Figure 10.2** Classification of DSSS signal.

in-phase segments is Gaussian when the added noise has a Gaussian distribution. Therefore we could easily adopt the likelihood function given in equation (3.5) and construct a magnitude-based likelihood classifier (ML-M).

For PSK modulations, it is clear that the phase of the signal complex representation has clear states. By employing phase-based likelihood functions, there are two benefits. First, the channel estimation error can be negated, as the signal magnitude plays a trivial role in the evaluation. Secondly, as much of the distinction between different PSK modulations is displayed in their signal phase, the phase-based likelihood function [equation (3.9)] should provide optimal classification performance on only the dimension of signal phase as part of a phase-based maximum likelihood classifier (ML-P).

For QAM modulations, there are more variations in the symbol mapping in terms of signal in-phase and quadrature segments. It is obvious that the bi-variant likelihood function equation (3.3) for the signal's complex representation is most suitable to exploitation of the different characteristics of different QAM modulation symbols mapping (Figure 10.2).

### *10.3.2 Classification of FHSS Signals*

In the case of FHSS, the carrier frequency of the transmitted signal is changing all the time. The first difficulty is the detection of the signal. While it is not within the scope of this book, the difficulty in signal detection has profound implications in modulation classification. As the frequency detection may be delayed or inaccurate, it is obvious that the performance of the modulation classifier may be affected. In Chapter 8 we benchmarked some state-of-the-art classifiers in the case of frequency offset. There is a common behaviour of significant performance degradation with even the slightest amount of frequency offset. Knowing that the frequency offset is prone in the channel estimator, the selection of modulation classifier is extremely difficult.

Fortunately, there are classifiers with natural robustness against frequency offset. While not being featured as the best option in most cases, these classifiers could be

suitable candidates for the task of classifying FHSS signals. In Chapter 3 the magnitude-based ML classifier is mentioned and its potential robustness is noted. Owing to the fact that signal magnitude is not affected by rotatory distortion of the signal, the performance of a magnitude-based classifier is an obvious candidate for FHSS classification. However, it is also noted that the magnitude-based ML classifier cannot classify PSK modulations since all PSK modulations have the same signal magnitude.

To enable the classification of PSK modulations while keeping the frequency offset in mind, the phase difference classifier is a possible option to combat frequency offset in PSK modulation classification. The phase difference classifiers extract the phase difference between adjacent signal symbols. As the rotatory shift of the signal symbols is progressive, the mismatch between the measured phase difference and the real phase difference can be minimized. Wang and Wang (2010) suggested the phase difference-based classifier using the distribution of phase difference derived by Pawula, Rice and Roberts (1982) for a KS classifier (see Chapter 4).

The classification begins with the fourth-order moment of the signal magnitude $\mu_{42}^m$. Though under heavy rotary dispersion, the signal magnitudes of the M-PSK signal are not affected, which maintains the low value of $\mu_{42}^m$. Clearly, the M-PAM and M-QAM modulations have higher $\mu_{42}^m$ values. To further classify PSK or different orders, the phase difference (PD) classifier can be employed to accomplish the task. To differentiate PAM and QAM modulations, the PD classifier is also a good option. For PAM modulations, despite the frequency offset, the adjacent signal samples maintain a relatively consistent phase difference. In contrast, QAM modulations have a more complex phase difference profile and are easily distinguishable from PAM modulations. The classification of M-QAM and M-PAM can both employ the magnitude-based ML classifier. An illustration of the AMC implementation for classifying FHSS signals is given in Figure 10.3.

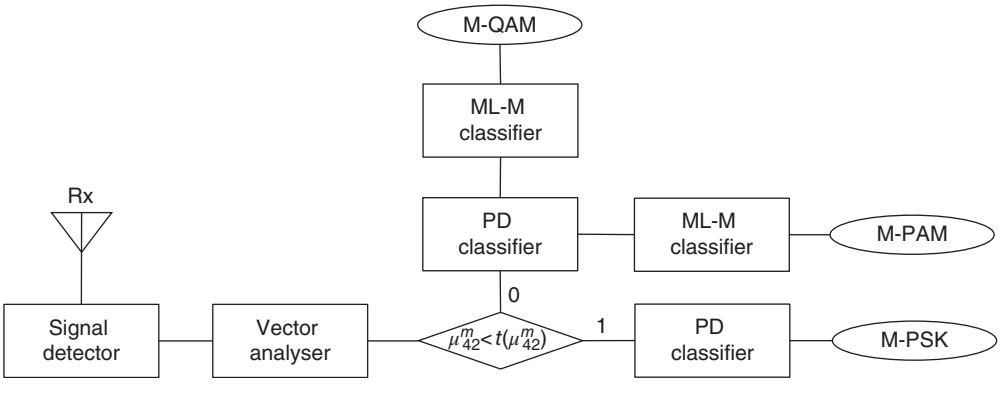



**Figure 10.3** Classification of FHSS signal.

## 10.4 Conclusion

In this chapter we considered several modulation classification tasks that are unique to the military scenarios. The detection of modulation type provides threat analysis and surveillance, as well as the ability to handle a wide range of common modulation types. The task of digital modulation classification is put in the challenge of low-probability-of-detection signals. Two different cases of LPD signal are considered, namely DSSS and FHSS. For DSSS, different types of likelihood-based classifiers are selected for their high performance when the SNR is lower than 0 dB. In the case of FHSS, given the inevitable frequency offset, magnitude-based and phase difference-based classifiers are suggested for a robust classification of the FHSS signal, due to their inherent robustness against frequency offset.

## References

Pawula, R.F., Rice, S.O., and Roberts, J.H. (1982) Distribution of the phase angle between two vectors perturbed by gaussian noise. *IEEE Transactions on Communications*, **30** (8), 1828–1841.

Poisel, A.R. (2008) *Introduction to Communication Electronic Warfare Systems*, Artech House, Norwood, MA.

Wang, F., and Wang, X. (2010) Fast and robust modulation classification via Kolmogorov-Smirnov test. *IEEE Transactions on Communications*, **58** (8), 2324–2332.

# Index



*Automatic Modulation Classification: Principles, Algorithms and Applications*, First Edition. Zhechen Zhu and Asoke K. Nandi.

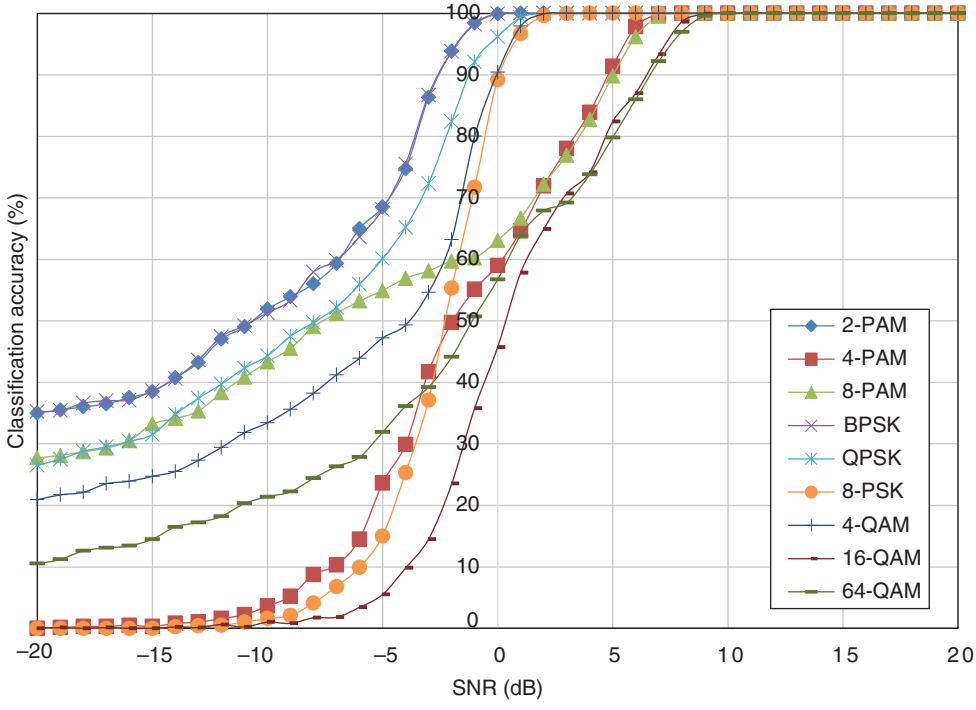



**Figure 8.2** Classification accuracy of the ML classifier in AWGN channel.

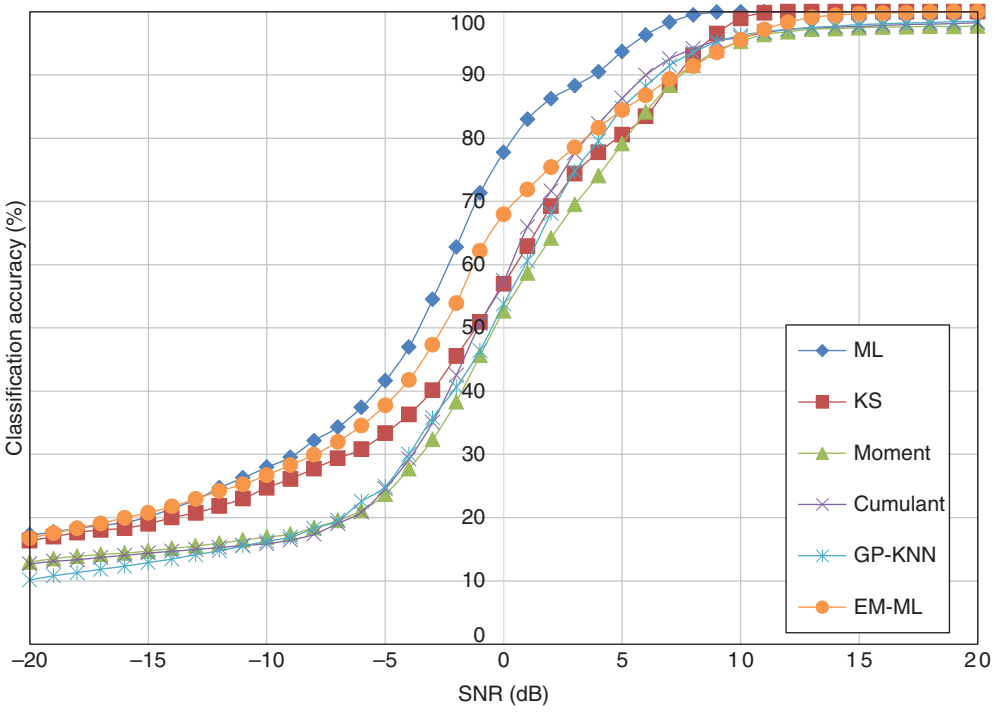



**Figure 8.8** Average classification accuracy of all classifiers in AWGN channel.

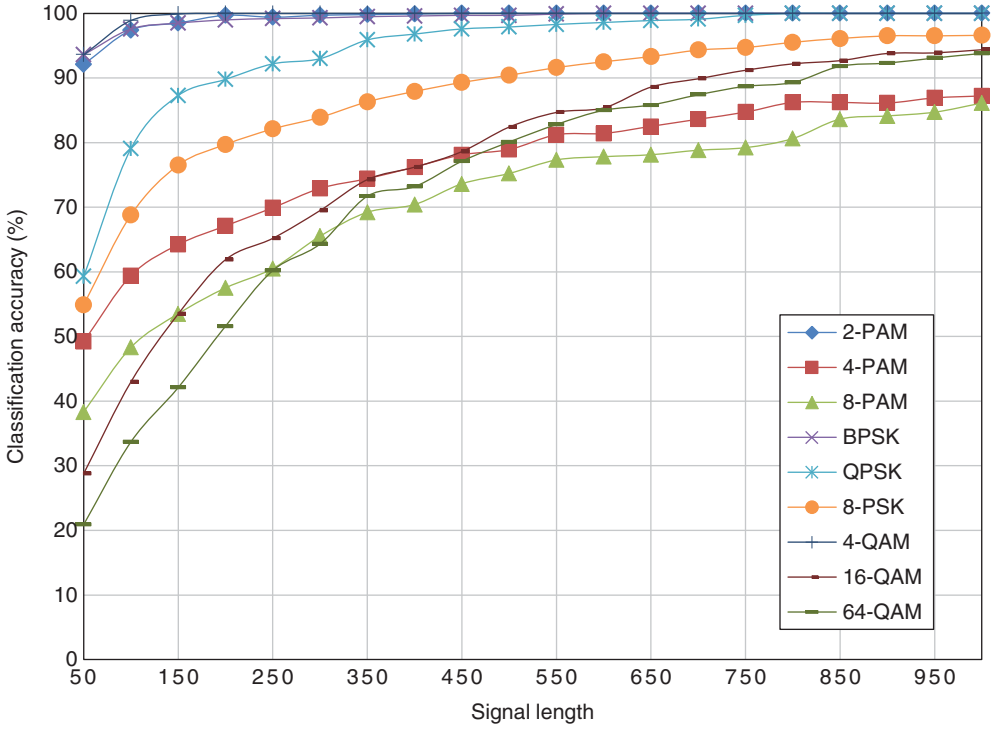



**Figure 8.12** Classification accuracy of the cumulant-based classifier with different signal length.

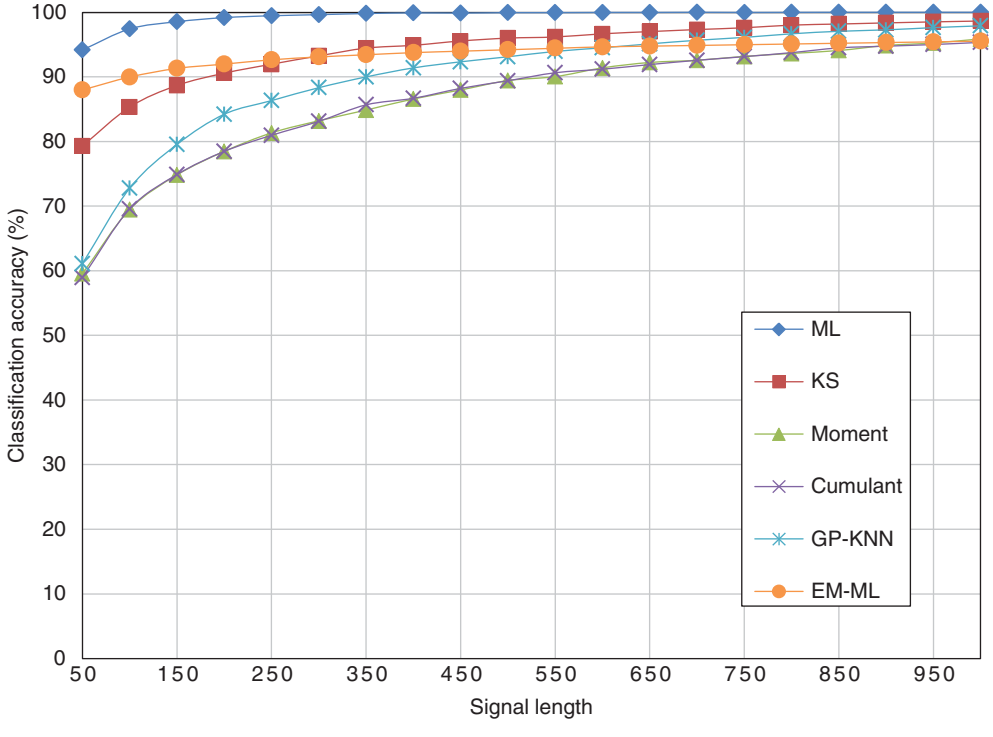



**Figure 8.15** Average classification accuracy of all classifiers with different signal length.

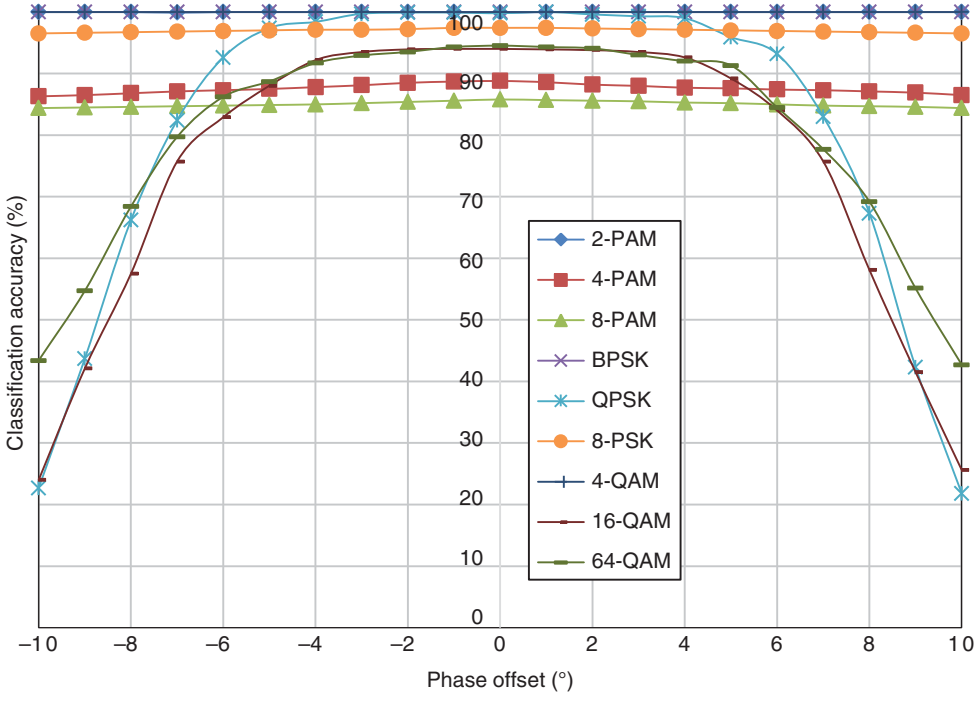



**Figure 8.18** Classification accuracy of the moment-based KNN classifier with phase offset.

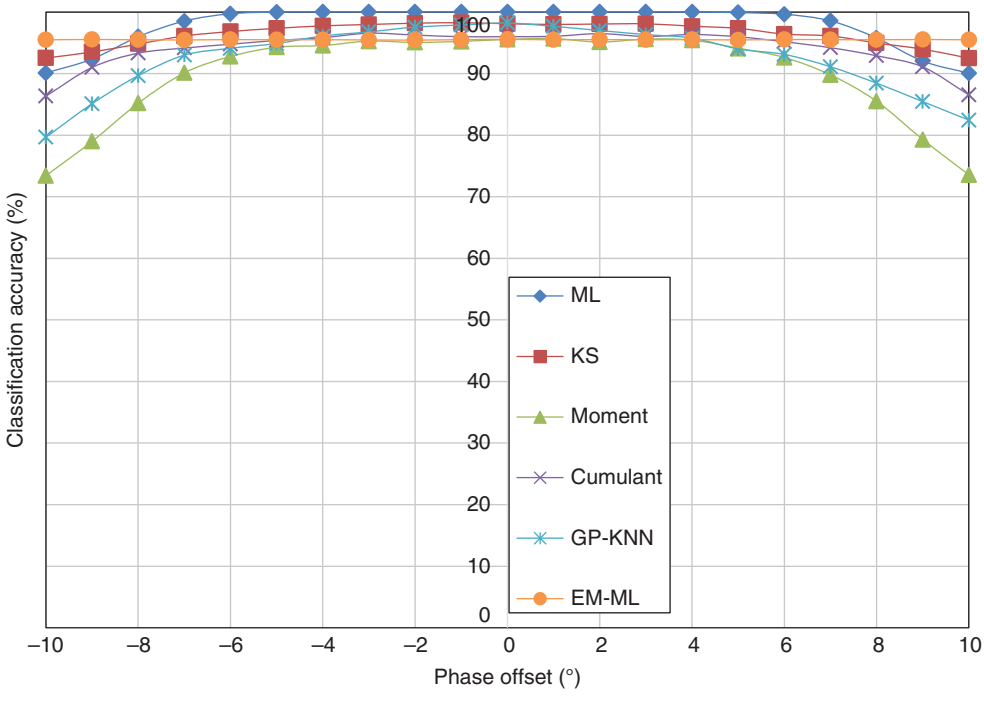



**Figure 8.22** Average classification accuracy of all classifiers with phase offset.

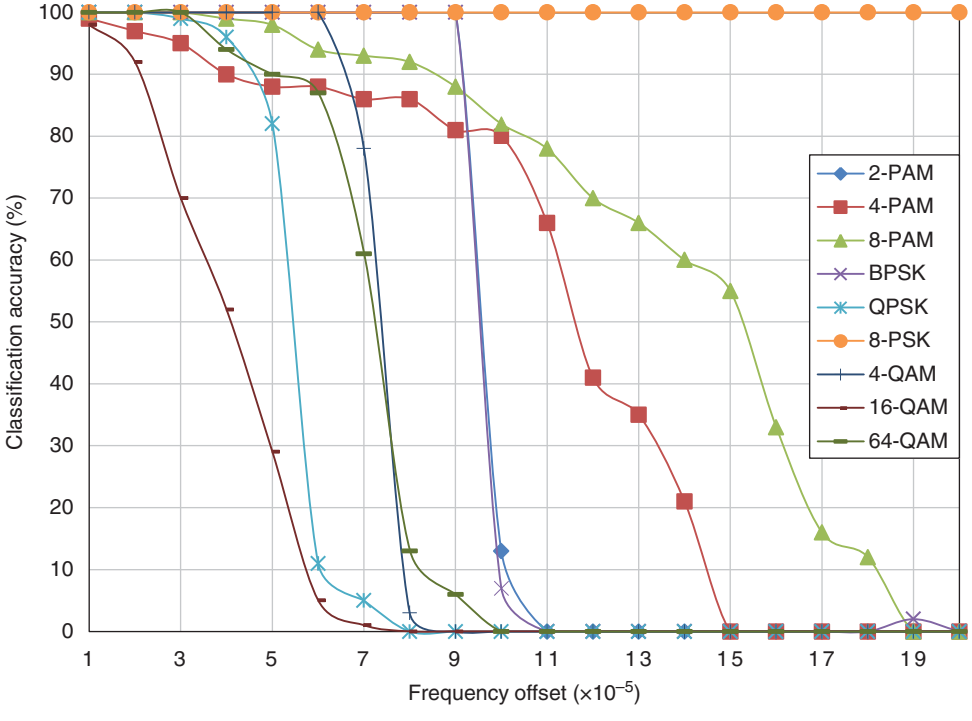



**Figure 8.27** Classification accuracy of the GP-KNN classifier with frequency offset.

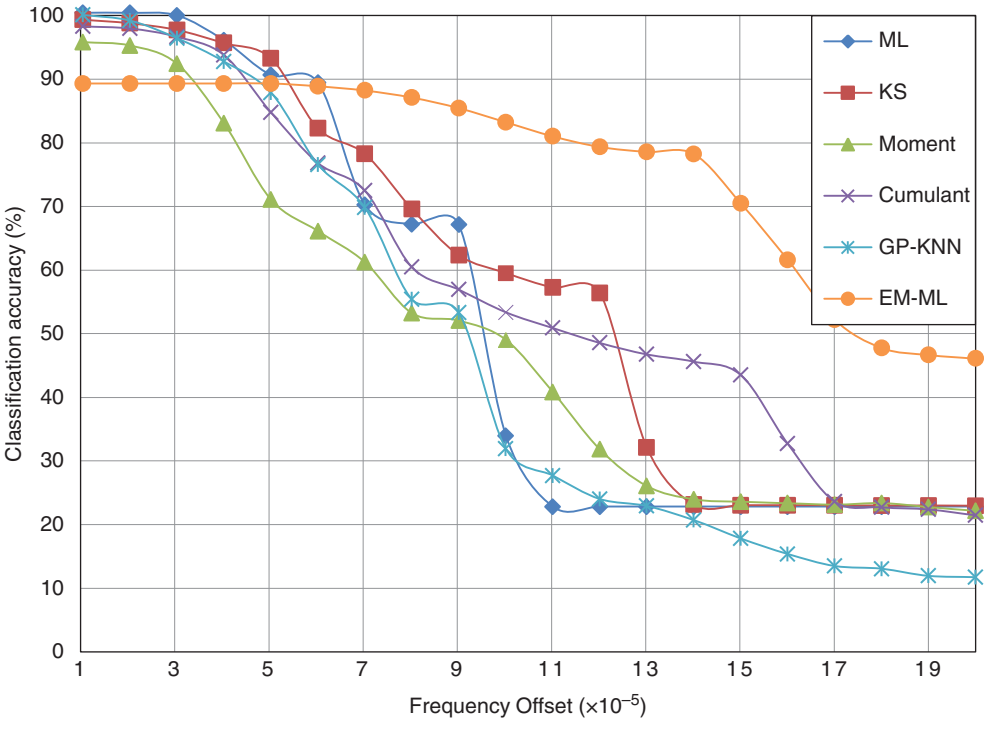



**Figure 8.29** Average classification accuracy of each classifier with frequency offset.

# WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley’s ebook EULA.