Modulation Classification Using a Goodness of Fit Test

Simin Hao¹, Na Wang¹, Ruiqi Sun², Meng Liu¹, Mengjiao Dong¹ and Hongjun Wang¹,*

¹Shandong University, Jinan, Shandong, China.
²University of Electronic Science and Technology of China, Chengdu, Sichuan, China.
*E-mail: hjw@sdu.edu.cn

Abstract. In this paper, we propose an innovative approach using Anderson-Darling (A-D) test for signal modulation classification with less signal samples in AWGN channels. The A-D test is a non-parametric approach to measure the goodness of fit test. It is based on Cramer-Von Mises (CVM) test, using the mean square integral of the difference between the empirical cumulative distribution functions (ECDFs) from received signals and cumulative distribution functions (CDFs) of the signal under different candidate modulation format. In order to avoid the default of CVM, such as it is less sensitive at heads and tails of the distributions, more weights are given to heads and tails in A-D test. Massive simulation results show that compared with the Kolmogorov-Smirnov (K-S) classifiers and the traditional high-order cumulant-based classifiers, the A-D classifiers show better classification performance at different SNRs with less signal samples for M-QAM and M-PSK modulations in AWGN channels.

1. Introduction

The automatic signal modulation classification is the inter-mEDIATE step between signal detection and demodulation, and plays an important role in both civil and military communication fields. In the field of civil communication, the automatic signal modulation classification is applied to solve the problems of interference signal identification and spectrum detection, so as to realize effective radio management and prevent the interference of illegal users. In the military domain, modulation identification is the premise for the non-cooperative party to complete demodulation and obtain the enemy's communication content. Only when the non-cooperative party obtains the modulation mode and demodulates it according to the corresponding mode, can it get the correct transmission information or adopt the corresponding interference and suppression measures. The initial modulation recognition is to analyze the waveform and spectrum of the signal with the help of relevant instruments, then, judges the modulation mode of the signal according to the experience of the staff. This kind of artificial identification method is usually slow with low accuracy and requires higher professional quality. With the development of communication technology, the complexity and diversity of signal modulation methods have further increased the difficulty of modulation identification. Simple waveform features cannot provide sufficient identification basis. Therefore, the study of automatic modulation classification technology is of great practical significance.

Generally, there are two methods of modulation classification, i.e., the likelihood-based (LB) method and the feature-based (FB) method [1]. The LB method takes advantage of the likelihood function from the received signal. According to comparing the likelihood ratio with a threshold to make the decision. It includes the alternatives of average likelihood ratio test (ALRT) [2, 3], the
generalized likelihood ratio test (GLRT) [4], and hybrid likelihood ratio test (HLRT) [4]. These methods may provide excellent classification accuracy but usually require more prior knowledge and have high computational complexity. Furthermore, since LB methods become ineffective under channel fading and phase or frequency offset, these methods usually lack of robustness. The FB method has a lower computational complexity and is much easier to implement than the former one. It uses several features based on spectral features [5], wavelet features [6], high-order statistic features [7], and cyclic features [8]. Then, make the decision according to their observed values. Although it is simple to implement, the accuracy of this method depends much on the signal pre-processing, and the choice of the decision maker, i.e., artificial neural networks (ANNs).

In 2010, Wang and Wang [9] proposed K-S test for higher-order signal classification. According to compare the difference between the ECDFs from received signals and CDFs of the signal under different candidate modulation format to classify modulation. The decision statistic results in the minimum value of the maximum distance between the two CDFs mentioned above. Compared with LB and FB methods, it has less computational complexity. It performs better at various channels than the LB method and it has high accuracy than the FB method.

In this paper, we propose an A-D test for modulation classification. Different from the K-S test which finds the minimum value of the maximum distance between the two CDFs, the A-D test adopts the integral method to find the minimum difference value between two distributions. This method can avoid individual extreme value effects and give more weight to heads and tails. In this paper, we use an A-D method to classify modulation for both QAM and PSK signals in AWGN channels using the amplitude features in real and imaginary components for QAM signals and phase feature for PSK signals. By testing, the A-D test performs better than K-S test and cumulant method in AWGN channels.

2. System Description
In this paper, we consider the following signal model
\[ y_n = x_n + w_n, n = 1, 2, 3...N \]  
where \( x_n \) is the transmitted modulation symbol, \( w_n \) is the signal noise symbol, and \( y_n \) is the received signal symbol. Here, we assume all signals are transmitted in AWGN channels and the noise symbol follows the complex Gaussian distribution, \( w_n \sim N_c(0, \sigma^2) \). The transmitted symbols \( x_n \) are equally derived from unknown modulation constellation \( M_k \), which belongs to a set of possible modulation formats \( \{ M_1, M_2, M_3...M_k \} \). Considering that the power of input signal is normalized to one, the signal-to-noise ratio (SNR) in dB can be written as \( 10 \log_{10} \frac{1}{\sigma^2} \), the modulation classification turns to finding which of the unknown constellation \( M_k \) belongs to base on the received signal.

2.1. Cumulant-based Modulation Classification
In the past, researchers often use cumulant-based modulation classification to identify high-order modulation signals. Swami and Sadler [10] suggested to use the forth-order cumulant functions of the complex-signal as features to classify M-QAM, M-PAM as well as M-PSK modulations, and got good results. The forth cumulants is defined by,
\[ \hat{C}_{20} = \frac{1}{N} \sum_{n=1}^{N} r^3[n] \]  
\[ \hat{C}_{21} = \frac{1}{N} \sum_{n=1}^{N} r[n]^2 \]  
\[ \hat{C}_{40} = \frac{1}{N} \sum_{n=1}^{N} r^4[n] - 3 \hat{C}_{20} \]
The null hypothesis is

\[ H_0 : F_N(z) = F_f(z) \]  

and \( F_N(z) \) is defined as

\[ F_N(z) = \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}(z_n \leq z) \]  

where \( \mathbb{1}(\cdot) \) is defined as a decision formula, which equals to one if \( z_n \) satisfies the inequality requirement, equals to zero if not. Under the null hypothesis, \( F_N(z) \) will be close to \( F_f(z) \) if the number of samples in sample sets is large enough. According to how to measure the distance between \( F_N(z) \) and \( F_f(z) \), there are three commonly used goodness of fit detection methods, including K-S Test [9], CVM test [11], and A-D test.

### 2.2.1. K-S Test

In the K-S test, the test statistic is defined by

\[ D = \sup_{-\infty < z < \infty} |F_N(z) - F_f(z)| \]  

where sup means the supremum of the set of distances.

### 2.2.2. CVM Test

The CVM test is another goodness of fit test method based on the mean square integral of the difference between the \( F_N(z) \) and \( F_f(z) \). It can avoid individual extreme value effects in K-S test, and it is defined by

\[ D = N \int_{-\infty}^{\infty} (F_N(z) - F_f(z))^2 dF_f(z) \]  

### 2.2.3. A-D Test

A-D test is based on CVM test, solving the problem that the difference between the head and the tail are too small. It is defined by

\[ D = N \int_{-\infty}^{\infty} (F_N(z) - F_f(z))^2 \phi(F_f(z))dF_f(z) \]  

where \( \phi(\cdot) \) is a weight formula defined by

\[ \phi(F_f(z)) = \frac{1}{F_f(z)(1-F_f(z))} \]  

These three tests all quantify the distance between \( F_f(z) \) and \( F_N(z) \). The null hypothesis is rejected if \( D \) is greater than a threshold.
In this paper, we use the A-D test for modulation classification and Figure 1 shows a brief flow chart of modulation classification for QAM as well as PSK signals.

![Flow Chart for QAM and PSK Modulations](image)

**Figure 1.** The flow chart for QAM and PSK modulations.

3. **Modulation Classification based on A-D Test in AWGN Channels**

Considering the signal model (1) and noise parameters mentioned in section 2, in order to classify the modulation mode from received signals \(\{y_n\}\), we use its magnitude feature or phase feature for classification. We define \(z_n = \{z_1, z_2, z_3, ..., z_N\}\) as magnitude or phase information derived from \(y_n = \{y_1, y_2, y_3, ..., y_N\}\). For each possible modulation candidate, we can get the standard CDFs \(F(z)\) from derivation and get ECDFs \(F_N(z)\) from received signals. The specific calculation processes are as follow.

3.1. **Classification of M-QAM Signals:**

For M-QAM signals, e.g. 4-QAM, 16-QAM, 64-QAM, the set of signal points of normalized constellations for above modulations are given by

\[
M_{4\text{-QAM}} = \left\{ \frac{1}{\sqrt{2}} (a + bj) | a, b = -1, 1 \right\}
\]

\[
M_{16\text{-QAM}} = \left\{ \frac{1}{\sqrt{10}} (a + bj) | a, b = -3, -1, 1, 3 \right\}
\]

\[
M_{64\text{-QAM}} = \left\{ \frac{1}{\sqrt{42}} (a + bj) | a, b = -7, -3, -1, 1, 3, 7 \right\}
\]

For M-QAM signals, the modulation features are contained in the magnitude. Since the signal constellation points are located symmetrically at a unit square in the complex domain and the white Gaussian noise is chaotic, we use real and imaginary components as the decision statistics. Firstly, from a sequence of \(N\) received signal samples, we get a sequence of \(2N\) samples as

\[
z_{2n-1} = \text{Re}\{y_n\}, \quad z_{2n} = \text{Im}\{y_n\} \quad n = 1, 2, ..., N
\]

where \(\text{Re}\{\cdot\}\) and \(\text{Im}\{\cdot\}\) represent the real and imaginary parts respectively. Then, we have \(z_n\) follow the normal distribution \(N(0, \frac{\sigma^2}{2})\). Secondly, we can obtain the CDFs under modulation \(M_k\) by

\[
F_k(z) = 1 - \frac{1}{\sqrt{M_k}} \sum_{x \in \text{Re}\{M_k\}} Q(\sqrt{2}(z-x) / \sigma), z \in R, k = 1, 2, 3, ..., K
\]

where \(Q(\cdot)\) denotes the Gaussian-Q function, and \(\text{Re}\{M_k\}\) is a set of real components of the received signals in \(M_k\)-order modulation. As the constellation points are located symmetrically, the imaginary components are the same as the real one, so we only use one of them as the decision statistics to get CDFs under certain modulation \(M_k\). Figure 2 shows the CDFs’ distribution diagrams under modulation \(\{M_4, M_{16}, M_{64}\}\).
Thirdly, the ECDF from the data samples is computed as
\[ F_N(z) = \frac{1}{N} \sum_{n=1}^{N} \mathbb{I}(z_n \leq z) \]  
where \( \mathbb{I}(\cdot) \) is defined as a decision formula, which equals to one if \( z_n \) satisfies the inequality requirement, equals to zero if not. Here, we use 2N samples (in step one) to draw ECDFs, and each set of the real and imaginary components occupies 50% proportion. Fourthly, A-D method is used for modulation classification. The A-D statistic is calculated by
\[ D_k = N \int_{-\infty}^{\infty} (F_N(z) - F_k(z))^2 \phi(F_k(z)) dF_k(z) \]  
where \( \phi(\cdot) \) is the weight formula and defined by
\[ \phi(F_k(z)) = \frac{1}{F_k(z)(1 - F_k(z))} \]
the decision on the modulation is given by the minimum A-D statistic which is defined by
\[ k = \arg \min_{1 \leq k \leq K} D_k \]  

3.2. Classification of M-PSK Signals:
For M-PSK signals, e.g. 4-PSK, 8-PSK, 16-PSK, the constellation points are located symmetrically at a unit circle in the complex domain. Then, the above complex signal vectors can be expressed by
\[ M_{M-PSK} = \{ e^{j \frac{2\pi}{M}(m-1)} | m = 1,2,3,...M \} \]  

Since the modulation features of PSK signals are contained in the phase. So, the phase of the received signals is used as the decision statistic
\[ z_n = \arctan \frac{\text{Im}(y_n)}{\text{Re}(y_n)} \quad n = 1,2,3,...N. \]  

With the model in (1), when transmitting the phase \( \theta \), the probability density function (PDF) of \( \phi = \angle(e^{j\theta} + w) \) is defined by [9]
\[
f(\theta | \phi) = \frac{1}{\pi \sigma} \int_{0}^{\infty} \lambda e^{-\frac{\lambda^2}{\sigma^2} + 2\lambda \cos(\phi - \theta)} d\lambda.
\]

So, the CDFs under \( M_k \) modulation is defined by
\[
F^M_f(z) = \frac{1}{M} \sum_{m=1}^{M} \int_{-z}^{z} f(\theta | \phi_m) = \frac{2\pi}{M} (m-1) d\phi \quad z = [0, 2\pi)
\]

Figure 3. shows the CDF’s distribution diagrams under modulation \( \{M_4, M_8, M_{16}\} \).

Finally, formulas (18)-(21) are used to get the decision.

4. Simulation and Results Analysis

In this part, we give the simulation results of A-D test, along with a comparison with other classifiers like cumulant-based and K-S based classifiers at different SNRs. To illustrate the importance of the method for signal classification, we extract the same signal features from three methods. For M-QAM modulations, we consider \( \{4-\text{QAM}, 16-\text{QAM}, 64-\text{QAM}\} \), and for M-PSK modulations, we consider \( \{4-\text{PSK}, 8-\text{PSK}, 16-\text{PSK}\} \). For each modulation type and SNR, we produce 30000 signals for testing and calculate the accuracy rates. The classification performance is shown in Figure 4 and Figure 5 respectively. For each testing signal, we use \( N=100 \) samples and all signals are transmitting in AWGN channels mentioned in (1) with noise follow the complex Gaussian distribution, \( w_n \sim N_c(0, \sigma^2) \). Figure 4. shows QAM modulation classification performance in AWGN channels with different classification methods. As illustrated in the graph, the A-D classifiers outperform other methods at all SNRs. Especially, when the SNR is greater than 10 dB, the A-D classifiers perform better and better than cumulant-based classifiers. In addition, when the SNR is greater than 13 dB, the A-D classification accuracy is around 1. However, compared with K-S classifiers, there is a slight improvement in classification accuracy as A-D classifiers use all signal symbol points to find the minimum value of fitting. For PSK modulation classification performance in Figure 5, the A-D classifiers have improved significantly. When the SNR is greater than 11 dB, the classification accuracy of A-D classifiers is around 1. However, the cumulant-based is just around 0.75 and the K-S based is around 0.9.
Next, we make a brief analysis of misjudgements for QAM and PSK modulations. Table 1 shows the misclassification probabilities when SNR=10 dB. As we can see in the table, most misjudgements occur between 8-PSK and 16-PSK, or 16-QAM and 64-QAM. Because the CDFs of the high-order modulations mentioned above are too similar as shown in Figure 6. So, it is difficult to classify from each other.

Table 1. Misclassification probabilities for SNR=10 dB

|       | 4-QAM | 16-QAM | 64-QAM |
|-------|-------|--------|--------|
| 4-QAM | 1.0000| 0.0000 | 0.0000 |
| 16-QAM| 0.0000| 0.7098 | 0.2933 |
| 64-QAM| 0.0000| 0.2902 | 0.7067 |
|       | 4-PSK | 8-PSK  | 16-PSK |
| 4-PSK | 1.0000| 0.0000 | 0.0000 |
| 8-PSK | 0.0000| 0.8875 | 0.1142 |
| 16-PSK| 0.0000| 0.1125 | 0.8858 |
Figure 6. The CDFs for QAM and PSK modulations when SNR=10 dB.

5. Conclusions
In this paper, we have proposed an innovative A-D test for M-QAM and M-PSK modulation classification. For QAM signals, we have used the real and imaginary components of amplitude to get CDFs (the real and imaginary components occupy 50% respectively). For PSK signals, we have used phase characters for classification. Simulation results have shown that with small number of samples, the A-D classifiers have good performance for high-order modulation classification and outperform the cumulant-based and the K-S based classifiers.

Reference
[1] Dobre O, A Abdi, Y Bar-Ness, and W Su J 2017 Survey of automatic modulation classification techniques: classical approaches and new trends IET Commun. vol 1, pp 137-156.
[2] Polydoros, A and Kim K J 1990 On the detection and classification of quadrature digital modulations in broad-band noise IEEE Trans. Commun. vol 38, pp 1199-1211.
[3] Hong L and Ho K C P 2000 BPSK and QPSK modulation classification with unknown signal level Proc. MILCOM’00 pp. 976-980.
[4] Panagiotou, P Anastasopoulos A and Ploydoros A P 2000 Likelihood ratio tests for modulation classification Proc. MILCOM’00 pp. 670-674.
[5] Jide P and Rex O J 2011 Automatic recognition of analog modulated signals using artificial neural networks Computer Technology and Application pp 29-35.
[6] Konstantinos M, Stavroula V and Philip C P 2007 Interclass and interclass modulation recognition using the wavelet transform Proc. IEEE PIMRC’07, pp. 1-5.
[7] Samir S and Shue-Zen H J 1992 Signal classification using statistical moments IEEE Trans. Commun. vol 40, no 5.
[8] Chad M P 1996 Classification of co-channel communication signals using cyclic cumulants Proc. IEEE ASILOMAR’96, Pacific Grove, CA, USA, 1996, pp. 531-536.
[9] F Wang and X Wang J 2010 Fast and robust modulation classification via Kolmogorov-Smirnov test IEEE Trans. Commun. vol 58, no 8.
[10] Ananthram Swami and Brian M. Sadler J 2000 Hierarchical digital modulation classification using cumulants IEEE Trans. Commun. vol 48, pp 416-429.
[11] Honda C, Oka I and Ata P 2012 Signal detection and modulation classification using a goodness of fit test Proc. ISIT. vol 23, pp 180-183.