Clustering based distribution fitting algorithm for Automatic Modulation Recognition

Kam-Tim Woo  
Department of Electronic and Computer Engineering  
Hong Kong University of Science and Technology

Chi-Wah Kok  
Department of Electronic and Computer Engineering  
Hong Kong University of Science and Technology

Abstract

Automatic Modulation Recognition (AMR) is an expert in modulation type identification. Many existing algorithms attempt to recognize the modulation candidates using phase and magnitude feature extraction. Performance is a major drawback of this feature extraction under noisy environment. In this paper, we proposed a new algorithm using a modified Chi-squared test on clustered received signals as components to its performance function. Simulations show that even under low SNR environment, our proposed algorithm achieved higher recognition rate than other existing algorithms.

1. Introduction

Recently, advances in adaptive modulation techniques have fueled the growth of software radio. This make a challenging task to design multi-mode handset that can dynamically reconfigures itself in global roaming [13]. Traditionally, either training sequence or pilot signal is employed to adapt the modulation setting. This inevitably reduces the channel capacity. Hence, algorithms that do not require any training sequence in training process, became a popular research topic in the past decade.

In the absence of prior knowledge of the received signals, we are interested in getting the global signature of the modulation type - the constellation shape. As each modulation type has its own set of basis functions such as shape and number of constellation signals, collection of these basis functions can form a modulation model library.

Automatic Modulation Recognition (AMR) is powerful in distinguish different modulation model from the received signals, and can be found in many applications in modern communication systems [1] to fulfill our aspects.

In general, the AMR method can be classified into two categories: Pattern recognition and Decision theoretic. Decision theoretic was the main methodology used by researchers like [4, 6, 7, 10] in the past. In recent years, pattern recognition methods are dominant especially with the use of neural networks [8, 15]. There are several different modulation recognition approaches as mentioned in [8] like quasi-log-likelihood to differentiate BPSK and QPSK, log-likelihood function to classify between Offset-QPSK (OQPSK) / BPSK / QPSK, and sequential probability ratio test in the context of hypothesis testing to classify among several QAM signals, etc. Authors in [14] use energy vectors derived from wavelet packet decomposition as feature vectors to distinguish between ASK, PSK, and FSK modulation types.

In this paper, we proposed a new algorithm for recognition of modulation types from a larger number of the constellation candidates. Our algorithm using modified Chi-square test on clustered received signals to illustrate the modulation recognition and its effectiveness. Speedup could be obtained by parallelizable of these separable recognition sub-algorithms.

A system model and assumption of the AMR algorithm is briefly reviewed in Section 2. In Section 3, a feature based recognition technique is employed to distinguish the constellation signals from either QAM or PSK in modulation technique. In addition, the validation of the constellation signals can be implemented by introducing a fitting technique on the distribution of received signals. Towards the end of this paper, simulations show that our algorithm outperform other existing algorithms. In addition, simulations also show that our algorithm works fine even work with a few number of received signals. Successful recognition of these systems give us opportunities to extend this algorithm to other communication systems.

2. System Model

Figure 1 shows the basic building blocks of our proposed algorithm. Modulated signals are transmitted over the channel. The received signals are usually corrupted by the channel impairment, such as AWGN noise and ISI. In recovering unknown signals, equalization should be taken into ac-
count. In our work, we assume the signals at the receiver output are already equalized. Although the phase error of the equalized signals may introduce a rotation of the reconstructed constellation signals, we assume the phase error is eliminated by Phase Locked Loop (PLL).

Without loss of generality, we will focus on the modulation signals in a 2-dimensional space. This can be achieved by extracting the amplitude and phase features of each modulation signal and projecting them onto a 2-dimensional space on an I-Q plane. It can be extended to higher dimensional signal space for other features, such as time or frequency.

A pre-defined library consists of all candidates of our target (4-PSK, 8-PSK, 8-QAM, 16-PSK, 16-QAM and 64-QAM) signal constellations will be used. We further assume that all QAM constellation in the library is rectangular QAM. Note that our algorithm is not limited to recognize the above library candidates, more library candidates can be included if necessary. The above modulation models are simple and most frequently used in digital communication system. We assume that the maximum modulus of each candidate is set to be one. The transmitted symbols are further assumed to be equally probable.

3 Proposed Algorithm

For any received constellation, the best match model can be found by conducting a full search of all candidates of the modulation models in the library. However, a full-search algorithm is expensive because the computational complexity grows rapidly with the increasing number of candidates in the library. In order to speed up the search procedure, an efficient algorithm on limiting the number of searched candidates is suggested.

Figure 2 shows the architecture of our proposed algorithm, which consists of two sub-routines of recognition - the modulation type recognition and the constellation size identification. The modulation type of the received signals (either PSK or QAM) is first to be recognized. With the success in identifying the modulation type, the number of possible candidate models can be reduced. As there is no need to match those modulation candidates with other modulation type, the computational complexity is reduced. The details of modulation type recognition will be discussed in Section 3.1.

Another sub-routine of recognition - the constellation size identification - is used to identify the valid constellation size in the identified modulation type which obtained in Section 3.1. A best match of valid constellation size, $M'$, is obtained by associating with a minimum cost function $J(M')$, which is modified from Chi-square test. We will discuss the details of constellation size identification in Section 3.2. With the recognized modulation type and the number of constellation signals, the modulation model can be uniquely identified.

3.1 Modulation Type Recognition

An effective approach to recognize the modulation type is an important part of this work. If a digitally modulated signal can be uniquely characterized by its constellation, it should also be identifiable by the recovered constellation at the receiver. This is the premise of the constellation type recognition.

Consider a noiseless channel, the received constellations of PSK should all lie on a single circle, while QAM constellations lie on some circles with different radii. We imposed a modulus feature criterion that can be used to discriminate the modulation type from their amplitude information. Suppose $r_1, r_2, \ldots, r_M$ are the modulus of each constellation signal in $M$-QAM, such that

$$r_1 \leq r_2 \leq \cdots \leq r_M = 1. \tag{1}$$

Noted the assumption that the maximum modulus of the modulated signal is set to one which is stated in Section 2. As a result, the average modulus of $M$-QAM is expected to be less than 1, because $\frac{1}{M} \sum_{i=1}^{M} r_i \leq \frac{1}{M} \sum_{i=1}^{M} r_M = 1$. In addition, the modulus of M-PSK should be equal to 1.

Since the received signal assumed to be corrupted by the AWGN noise, the constant modulus condition still held.
Therefore, the decision rule of the modulation type from the observation of the received signal \( X \) is proposed as

\[
E[\|X\|] \leq_{\text{PSK}} \frac{\text{GAM}}{1 - \epsilon} \tag{2}
\]

where \( E \) is the expectation operator, \( \| \cdot \| \) is the 2-norm distance metric, \( X = \{x_1, x_2, \ldots, x_N\} \) is the collection of \( N \) equalized signals, and \( \epsilon \) is small positive constant depends on noise variance. A hierarchical based decision criterion is required if more different modulation types are included in the library [2].

The modulation type of each received constellation is identified according to the rule in Equation (2). After the modulation type is identified, those with the same modulation type in the library become possible candidate models.

### 3.2 Constellation Size Identification

With the prior knowledge about the modulation type, a sub-routine is proposed to identify the constellation size in the received signals. Firstly, the constellation is partitioned into \( M \) non-overlapping clusters, which can be formulated as a \( M \)-clustering problem. At the end of the clustering process, each cluster centroid will be assigned as the reconstructed constellation signal. Secondly, a distribution fitting algorithm is used to measure the goodness of fit of the data in that cluster by comparing to the Gaussian model. The above procedure will be repeated for all possible modulation candidate models in the library of the same modulation type determined in Section 3.1.

In the following sections, we will discuss how data is being partitioned, and how the distribution fitting algorithm works.

#### 3.2.1 Clustering in Constellation

As discussed in Section 2, recognition of an unknown modulation is relevant to a clustering problem. Using clustering algorithms can assist the constellation signal reconstruction by minimizing the sum of squared error (SSE). Either KM [11], KHM [17] or binary hyperplane-based algorithms [16] could be employed to assist the clustering process. The recognition of the number of modulation signals starts with the first possible model with \( M \) modulation signals. The received signals are partitioned into \( M \) non-overlapped regions with initial centroids.

Figure 3(a) is an 8-PSK received constellation signals. With the use of any clustering algorithm, the received constellation signals is then partitioned into possible number of clusters associated with the potential PSK candidates in the library. Figure 3(b) - 3(d) show the different clusters in the 4-PSK, 8-PSK and 16-PSK modulation candidates identification respectively. In Figure 3(b) - 3(d), the boundaries of adjacent clusters and the centroids of clusters are denoted by dash-dot lines and circles respectively. The cluster centroids are regarded as reconstructed constellation signals. The clusters can be viewed as the decision regions which is used to validate the constellation signals by distribution fitting algorithm in next section.

#### 3.2.2 Distribution Fitting Algorithm

Observing from Figure 3(b) - 3(d), the best match of the constellation size (number of constellation signals) could be achieved when all the tested clusters are Gaussian distributed around the corresponding centroids.

The clustering results are validated by the Gaussian fitting algorithm which quantifies how close the data distribution in each cluster is when compared to the Gaussian distribution. Since both inphase and quadrature phase components are assumed to be independent, the two dimensional probability density function \( g(\cdot) \) is given by

\[
g(x) = \frac{1}{2\pi(\sigma_j)^2} \exp(-\frac{\|x - c_j\|^2}{2(\sigma_j)^2}) \tag{3}
\]

where \( x \in C_j \), \( c_j \) and \( (\sigma_j)^2 \) are the centroid and variance of cluster \( C_j \), \( j = 1, 2, \ldots, M \).
It is quite difficult to obtain a continuous distribution function from the received signals as they are all discrete random signals. There are three basic elements to determine the goodness of fit of a two dimensional distribution. Firstly, a two dimensional distribution fitting problem can be reduced into a collection of one dimensional distribution fitting problems. To achieve this, the one dimensional distributed data can be obtained by splitting the two dimensional distributed data into either vertical or horizontal slices. Secondly, a Chi-square test is used to test the goodness of fit of the Gaussian distribution to the sliced signals. Thirdly, the Chi-square test is repeated for another sliced signals in each cluster. These steps are repeated for all candidate models with the same modulation type.

To extend the use of Chi-square test from a one dimensional distribution to a two dimensional distribution fitting test of each cluster, an adaptive \( L \times L \) grid is employed. Each cluster will be masked by the \( L \times L \) grids for the goodness of fit test. The width of the grid, \( \Delta \), is varied according to the data variance in the clusters, and the center of the grid is coincided with the centroid of the cluster.

The goodness of fit test is defined by a modified Chi-square based cost function, \( J(M^i) \), is defined as

\[
J(M^i) = \frac{1}{M^i} \sum_{j=1}^{M^i} \sum_{u=1}^{L} \sum_{v=1}^{L} \frac{[N_j(u,v) - \bar{N}_j(u,v)]^2}{\bar{N}_j(u,v)}
\]

where \( M^i \) is the total number of clusters obtained in each modulation candidate \( i \), \( N_j(u,v) \) and \( \bar{N}_j(u,v) \) is the actual and expected number of signals that fall in the \((u,v)^{th}\) grid in the \(j^{th}\) cluster respectively.

This cost function, \( J(M^i) \), measures the weighted difference between the number of received signals and the expected number of received signals that fall in the same interval. The larger the value of this cost function, the more the mismatch of that distribution.

### 4 Simulation Results

In this section, we investigate the performance of the AMR algorithm. The candidates of modulation models 4-PSK, 8-PSK, 8-QAM, 16-PSK, 16-QAM and 64-QAM are included in our simulation. Without prior knowledge of the constellations feature, we will use AMR to recognize the most likely candidate modulation in our pre-defined library. The modulation type of the received signals is first determined by the constant modulus criterion. Then, the received signals will be partitioned into a certain number of clusters, and the reconstructed constellation signals are assigned according to the cluster centroids. Finally, the received signals in each cluster are mapped onto the \( L \times L \) squared grids for distribution fitting test.

In the distribution fitting procedure, two important parameters, the number of grids \( L \) and the grid size \( \Delta \), are required to estimate the probability density function of random received signals. The wider the base, \( L \times \Delta \), the larger the contribution of those farthest received signals. This leads to a smooth density estimation. However, the narrow base leads to an erratic estimation because of very few or no data points are included. In addition, the density function will smear if the grid size, \( \Delta \), is too large. On the other hand, a too narrow grid will lead to many empty resolution elements and a spiky estimation. In our simulation, the setting of the two parameters are

\[
L = 11 \quad \text{and} \quad \Delta = \frac{1}{2} \max(\sigma_I, \sigma_Q)
\]

where \( \sigma_I \) and \( \sigma_Q \) are the standard deviation of the inphase and quadrature phase components of the received signals that fall in the cluster. A larger value \( \sigma_I \) or \( \sigma_Q \) indicates a wider spread of the received signals.

The performance of AMR is evaluated by the number of successful recognized packages over 1500 independent trails. Each package consisting of \( N \) equally probable modulated signals, is sent over an AWGN channel with \( \text{SNR} \) in the range from 0 dB to 20 dB. The successful recognition rate is given by,

\[
\text{Successful rate} = \frac{\text{number of successful recognized packages}}{\text{number of received packages}},
\]

and

\[
\text{SNR} = 10 \log_{10} \frac{E_s}{N_o} = 10 \log_{10} \frac{1}{N_o} \sum_{j=1}^{N} |c_j|^2
\]

where \( E_s \) is the power of the additive Gaussian noise, \( c_j \) is the constellation signal, and \( E_s \) is symbol energy.

Figure 4 show the performance of successful recognition rates of the AMR algorithm with packages size of 200 (dashed line) and 300 (solid line). The package size is indicated inside the brackets in the legend. Without loss of generality, more simulation points can provide a higher recognition rate as they support a better estimation of probability density function. From the observation of the performance curve in Figure 4, similar successful recognition rate can be achieved by both 200 and 300 received signals especially for PSK constellations. This demonstrates that our proposed AMR algorithm is robust to the variation of the number of received symbols.

Simulation results show that the proposed AMR algorithm can effectively recognize the candidate models when \( \text{SNR} \) is above 13 dB. Even the \( \text{SNR} \) is as low as 10 dB, 80\% of the candidate models except 16 PSK can be successfully recognized. When the \( \text{SNR} \) is further increased to 18 dB or above, all modulation models in the library can be recognized successfully.
4.1 Comparison with The Previous Work

Modulation recognition is a relatively young research topic yet there are not many existing performance measurements. Comparing our performance with results in other publications are not that straightforward. There are two main reasons for this. The first one is there is no established standard digital modulation library. Hence, different researchers applied their algorithms to their own choice. For instance, the authors in [3] proposed an algorithm in identification of digital modulation types (QAM, PSK and FSK) using the wavelet transform. Meanwhile, the authors in [4, 15] proposed to classify different modulation models within the same modulation type. [4] proposed the algorithm which is used to distinguish the models between BPSK and QPSK, while the authors in [15] applied the neural network algorithm on the classification of 16, 64 and 256-QAM. In [2], the authors proposed an algorithm for multi-models classification between M-PSK, M-FSK, M-ASK. However, model candidates are limited to $M = \{2, 4\}$ in their algorithm.

The second reason is about the successful recognition rate of a specified modulation. The successful recognition rate is strongly affected by candidates of modulation models in the library. As a result, even identical recognition rates obtained by two different algorithms, variation still occurs if their candidate models used are different.

In this section, we compare the performance of our proposed algorithm and the algorithm proposed in [2]. To have a fair comparison, we limit the candidates of modulation models to 2-PSK and 4-PSK only. As both candidates have the same modulation type, PSK, we can implement our constellation size recognition with the Gaussian Fitting algorithm.

In the simulation, 300 samples of unknown signals are sent through an AWGN channel. The performance of the algorithm is measured by an average successful recognition rate (ASRR), which is evaluated by averaging out the successful recognition rate of 2-PSK and that of 4-PSK. The successful recognition rate is obtained over 1500 realizations. Figure 5 shows the average successful recognition rate of our proposed algorithm compare with the algorithm used in [2]. Simulation results have demonstrated that our algorithm outperform the algorithm in [2]. The main reason is our distribution fitting algorithm overcome the high sensitive issue of the phase variance feature in [2] in low SNR environment.

Now, let’s extend both algorithms into the modulation recognition with the library candidates of 2-PSK, 4-PSK and 8-PSK. Figure 6 shows ASRR of our proposed algorithm and the algorithm in [2]. Similar to the previous simulation results in Figure 5, our algorithm still outperform the algorithm in [2].

Other recognition algorithms based on higher order statistical moments in [12] are also being compared. In this simulation, we would compare the performance of ASRR
of the algorithm in the modulation recognition between 2-PSK and 4-PSK. Similar to the previous work, 300 samples of received signals are used in the recognition. Figure 7 shows the average successful recognition rate of different algorithms. Higher successful recognition rate can be achieved by using 4th statistical moments than 2nd statistical moments. As shown in Figure 7, the slope of the ASRR is also constant. This demonstrates that the algorithm [12] is less sensitive to noise variance compared to the algorithm in [2]. Under low SNR (from 0 to 3 dB), the statistical moments algorithm has a better performance compared to the algorithm in [2]. However, the algorithm in [2] outperform the algorithm in [12] after the cross-over points. As the noise variance is involved in the cost function of our proposed model fitting algorithm, it provides a robustness of noise sensitivity. This is the reason why our proposed algorithm could outperform the algorithms proposed in [2] and [12].

5 Summary

We have presented an approach for automatic modulation recognition that uses the shape of the rebuilt constellation as the key signature. By considering this key signature feature on the I-Q plane, the proposed AMR algorithm consists of two sub-routines in modulation type recognition and constellation size identification. Simulation results showed a successful recognition rate for as high as 90% at SNR 10 dB of 8-PSK, 8-QAM and 16-QAM. When SNR increases to 13 dB, the successful recognition rate of 16-PSK can reach 90%. The simulations also show our algorithm outperform other algorithms.

References