Genetic Algorithm Optimized Distribution Sampling Test for M-QAM Modulation Classification

Zhechen Zhu\textsuperscript{a}, Muhammad Waqar Aslam\textsuperscript{b}, Asoke K. Nandi\textsuperscript{a,c,*}

\textsuperscript{a}Department of Electronic and Computer Engineering, School of Engineering and Design, Brunel University, Uxbridge, Middlesex, UB8 3PH, UK
\textsuperscript{b}Signal Processing and Communications Group, Department of Electrical Engineering and Electronics, The University of Liverpool, Brownlow Hill, Liverpool L69 3GJ, UK
\textsuperscript{c}Department of Mathematical Information Technology, University of Jyväskylä, Jyväskylä, Finland

Abstract

With the classification performance and computational complexity in mind, we propose a new optimized distribution sampling test (ODST) classifier for automatic classification of M-QAM signals. In ODST, signal cumulative distributions are sampled at pre-established locations. The actual sampling process is transformed into simple counting task for reduced computational complexity. The optimization of sampling locations is based on theoretical signal models derived under various channel conditions. Genetic Algorithm (GA) is employed to optimize distance metrics using sampled distribution parameters for distribution test between signals. The final decision is made based on distances between tested signal and candidate modulations. By using multiple sampling locations on signal cumulative distributions, the classifier’s robustness is enhanced for possible signal statistical variance or signal model mismatching. AWGN channel, phase offset, and frequency offset are considered to evaluate the performance of the proposed algorithm. Experimental results show that the proposed method has advantages in both classification accuracy and computational complexity over most existing classifiers.

Keywords: automatic modulation classification, cognitive radio, distribution test, genetic algorithm, AWGN, flat fading channel.

*Corresponding author email: asoke.nandi@brunel.ac.uk
Email addresses: zhechen.zhu@brunel.ac.uk (Zhechen Zhu), m.w.aslam@liverpool.ac.uk (Muhammad Waqar Aslam)

Preprint submitted to Signal Processing May 29, 2013
1. Introduction

Automatic Modulation Classification (AMC) has been an established research topic for many years. The initial application of AMC was mostly in military electronic warfare, surveillance and threat analysis [1]. The main purpose of AMC is to classify automatically the modulation type of the intercepted signal so that it can be correctly demodulated. Many papers, e.g. [2, 3, 4, 5, 6], have been published suggesting different solutions for this problem. Recently, as intelligent radio communication systems emerge in modern civilian communication applications, AMC, which is an important component in the adaptive modulation module, has attracted much attention from Cognitive Radio (CR) and Software Defined Radio (SDR) developers, e.g. [7, 8, 9].

The fundamental task of AMC remains the same, though new challenges arise in the current CR and SDR development environments. One obvious difficulty comes from the different modulation types being used. In recent years, the use of signal modulations has migrated towards Quadrature Amplitude Modulations (QAM) due to their ability of efficient high capacity data transmission. The popularity of QAM modulations can be easily verified with their presence in many modern radio communication standards. In IEEE 802.11a [10], BPSK, 4-QAM, 16-QAM and 64-QAM are employed as modulations for many wireless communications applications. In Digital Video Broadcasting Terrestrial (DVB-T) [11], 4-QAM, 16-QAM and 64-QAM are also universal selection for digital TV broadcasting.

In this paper, 4-QAM, 16-QAM and 64-QAM have been chosen for the development of the proposed classifier. Nevertheless, modifications can be easily made to accommodate other QAM or wider selection of modulations. The classification of QAM modulations has its unique challenges as most of the signal features are very similar between different M-ary QAM (M-QAM) modulations. Another challenge for AMC is the demand for accurate classification performance under different channel conditions. In addition to channel effects, the demand for short processing time is also an interest for different applications which require real-time reconfiguration of the communication system. In short, the goal is to develop a simple AMC classifier which gives accurate and robust classification performance.
Most existing AMC classifiers can be grouped into two categories: likelihood based (ML) classifiers and feature based classifiers. A likelihood based classifier could give the upper bound of the classification accuracy given the condition of accurate channel estimation. In [12], Wei and Mendel presented the ML method that provides the optimum performance with the correct channel estimation. More ML based classifiers, e.g. [13, 14, 15], have been developed recently to suit different modulations and channel conditions. However, the computational complexity is a major concern. It has led many sub-optimal approaches to be developed in order to have reduced complexity. Wong and Nandi used Minimum Distance (MD) classifier [16] to reduce the complexity. In a different way, Xu, Su and Zhou approached the complexity reduction problem by storing the pre-calculated values in quantized databases for avoiding complex operations [17]. These aforementioned methods have all successfully reduced the complexity to different degrees with certain amount of performance degradation.

For the purpose of further reducing the computational complexity, algorithms based on distribution tests have been developed and presented in some recent publications. F. Wang and X. Wang [18] used Kolmogorov-Smirnov Test (K-S test) [19] to formulate a solution by comparing the testing signal Cumulative Distribution Functions (CDFs) with the reference modulation CDFs. This method successfully achieved an improved performance especially when limited signal length is available. It was pointed out in [20] that the K-S test approach requires the complete construction of signal CDFs which is relatively complex and has the potential to be simplified. In the same paper, an optimized approach was presented which reduces the complexity of K-S classifier by analyzing the CDFs between two modulations at a single given location. When multiple modulations are considered, multiple locations, each responsible for the classification of two modulations, have been used. The classification accuracy is comparable to the K-S classifier and the complexity of the algorithm is reduced significantly. However, it is clear that the embedded information in CDFs is underutilized and the robustness of this approach can be improved. To overcome these limitations, we have developed the optimized distribution sampling test (ODST) classifier which conducts simplified distribution tests at multiple optimized sampling locations to achieve the balance
between simplicity and performance.

A feature based classifier generally consists of few steps including feature extraction, feature selection, and classification. A classic example of feature based AMC method can be found in [21] where multiple features are used with a decision tree classifier and Artificial Neural Network (ANN) classifier. In recent years, higher-order statistics [5] has proven to be well suited for M-QAM signals classification. Notably, the adoption of machine learning techniques show great potential in further enhancing the classification accuracy. Support Vector Machine [22], Artificial Neural Network [23], and Genetic Programming [24] have all been experimented with to select and combine existing features to exploit the full potential in them. In [23], Wong and Nandi proposed to automate the feature selection process with Genetic Algorithm (GA) and successfully reduced the feature dimensions without degrading the classification performance. As distribution parameters estimated at sampling locations could be considered as features, GA has been employed for the optimization of distance metrics for the proposed ODST classifier.

This paper is arranged in the following order. The signal models in different channel conditions are presented in Section 2. It is followed by a detailed description of the classification strategy as well as some extensive analysis in Section 3. Experiment setups are explained in Section 4, with detailed results and analysis given in Section 5. The conclusion is drawn at the end.

2. Signal Models

The general frame of the sampled signal after matched filtering is assumed to be

\[ r(n) = \alpha e^{j(2\pi f_o n T + \theta_o)} \sum_{l=-\infty}^{\infty} s(l) h(nT - lT + \epsilon T) + g(n) \]

where \( \alpha \) is the attenuation factor, \( f_o \) is the frequency offset, \( \theta_o \) is the phase offset and \( g(n) \) being the additive noise. The signal transmitted is assumed to have unit energy. \( s(l) \) are constellations points on an \( I - Q \) plane. The samples are assumed to be normally
distributed among the symbol collections, $h(\cdot)$ is the residual channel effect caused by timing errors $\epsilon_T$ with symbol timing of $T$. In this paper, we collect the signal information from the two constellation dimensions separately. As the observed signal is seen as complex valued, different dimensions are extracted by taking the real and imaginary part of the signal value. Though information is collected and treated separately, both are used in the final decision making for larger statistics. Here we define $r_X(n)$ to be the real part of $r(n)$ extracted using filter $\Re\{\cdot\}$ and $r_Y(n)$ to be the imaginary part of $r(n)$ extracted through $\Im\{\cdot\}$.

$$r_X(n) = \Re\{r(n)\}$$

$$r_Y(n) = \Im\{r(n)\}$$

The modulations in discussion are 4-QAM, 16-QAM and 64-QAM. The general assumption is that the symbol timing is perfectly recovered. Other channel effects including additive noise, phase offset, and frequency offset will be discussed in later sections.

2.1. Additive White Gaussian Noise Channel

The establishment of signal models is necessary for the ODST, since the optimization of sampling locations is based on them. In addition, the quality of the signal modeling greatly affects the quality of the optimized sampling locations. In this section, we consider only Additive White Gaussian Noise (AWGN) channel with unknown channel gain. Other effects are neglected and will be discussed in later sections. The general signal model in AWGN channel can be expressed as

$$r(n) = \alpha s(n) + g(n)$$

where $\alpha$ gives the attenuation factor and $g(n)$ is the additive white Gaussian noise. Here we assume that Signal-to-Noise-Ratio (SNR) is estimated in pre-processing and the transmitted signal to have unit energy. Though most SNR estimation algorithms rely on a given modulation type to calculate the SNR, several blind SNR estimation algorithms have been proposed to operate with unknown signal modulations in [25], [26] and [27]. In this paper, we assume that SNR is known, as assumed in most AMC literature, so that the comparisons
with existing methods are fair. The SNR can be written in the following form where the attenuation factor has been counted as part of the transmitted signal.

\[ SNR = 10 \log_{10} \frac{\alpha^2}{\sigma^2} \]  

(5)

SNR is assumed to be known during the classification. However, the channel gain is presumed to be unknown. Therefore, to have a consistent base model for further analysis, all observed signals have to be normalized. The normalization is done separately on each signal dimension with estimated signal means \((\mu_X \text{ and } \mu_Y)\) and standard deviations \((\sigma_X \text{ and } \sigma_Y)\).

\[ r'_X(n) = \frac{r_X(n) - \mu_X}{\sigma_X} \]  

(6)

\[ r'_Y(n) = \frac{r_Y(n) - \mu_Y}{\sigma_Y} \]  

(7)

To obtain the empirical distribution of the modulated signal, two key parameters are needed: the signal centroids, \(s_X(n) \text{ and } s_Y(n)\), and the noise variance \(\sigma\) around each centroid. The transmitted original signal centroids are drawn from different modulations. For square M-QAM modulations, the resulting real and imaginary centroids share the same pool

\[ s_X(n), s_Y(n) \in \{A_1, \cdots, A_I\} \]  

(8)

where \(I = \sqrt{M}\) is the total number of symbols centroids on one dimension of the signal constellation. Knowing that the transmitted signal have unit energy, the expression of \(A_i\) can be written as

\[ A_i = \frac{2i - I - 1}{\sqrt{\frac{2}{\pi} \sum_{i=1}^{I} (2i - I - 1)^2}} \]  

(9)

The transmitted signal centroids are first altered in the AWGN channel where they are scaled by the unknown attenuation factor \(\alpha\). The second alteration happens when the signals are normalized and divided by the standard deviation of all samples \(\sigma_r\). The resulting signal centroids can then be found as

\[ \hat{s}_X(n), \hat{s}_Y(n) \in \{\hat{A}_1, \cdots, \hat{A}_I\} = \{\frac{\alpha}{\sigma_r}A_1, \cdots, \frac{\alpha}{\sigma_r}A_I\} \]  

(10)
As the noise variance is not affected by the attenuation effect according to Equation (4), it
is only subjected to the normalization process.

\[ \hat{\sigma} = \frac{\sigma}{\sigma_r} \]  

(11)

Though \( \sigma_r \) can be easily estimated, the attenuation factor is unknown and has to be derived
together with the noise variance using the definition of SNR. The process begins with finding
the relationship between the signal variance \( \sigma_r \), transmitted signal centroids and transmitted
noise variances.

\[ \sigma_r^2 = \sum_{n=1}^{N} (X(n) - \mu_X)^2 \]  

(12)

With the assumption of signal mean to be centred around origin and noise to come from
symmetrical Gaussian distribution, for M-QAM modulation, the terms in the above equation
can be replaced by transmitted signal centroids and noise variance. The detailed derivation
can be found in Appendix A.

\[ \sigma_r^2 = \sum_{i=1}^{I} \alpha_i^2 A_i^2 + \sigma^2 \]  

(13)

By combining equation (5), (10), (11) and (13). The estimated noise variance can be written
as

\[ \hat{\sigma} = \frac{1}{\sqrt{1 + 10^{SNR/10}}} \]  

(14)

and the estimated signal centroids

\[ \hat{A}_i = \sqrt{\frac{10^{SNR/10}}{1 + 10^{SNR/10}}} A_i \]  

(15)

Having estimated the two parameters, the combined Probability Density Functions (PDFs)
could be written as

\[ f_i(x) = \frac{1}{\hat{\sigma}\sqrt{2\pi}} e^{-\frac{(x-\hat{A}_i)^2}{2\hat{\sigma}^2}} \]  

(16)

The estimated signal model is then obtained by easily combining the individual PDFs from
each symbol centroids.

\[ \hat{f}(x) = \frac{1}{I} \sum_{i=1}^{I} (f_i(x)) \]  

(17)
and the CDF is expressed as
\[ \hat{F}(x) = \int_{-\infty}^{x} \hat{f}(t)dt \] (18)
with the signal model reduced to depend on only two parameters: SNR and modulation type. This proves that the normalized signals have the same underlying signal model whatever is the channel gain. As long as the SNR stays the same, there is no other factors need to be concerned with and such a model can be used for the optimization of sampling locations.

2.2. Fading Channel with Phase and Frequency Offset

In flat fading channel, phase and frequency offsets are added along with some attenuation and additive noise. The received signal after matched filtering and sampling is given by
\[ r(n) = \alpha e^{j(2\pi f_0 n + \theta_0)} s(n) + g(n) \] (19)
where the residual intersymbol interference is omitted and treated as noise. We first consider the phase offset. It is assumed here that fading is slow, thus the phase offset is consistent for all signal samples. Instead of constructing a signal model with phase offset in mind, it is easier to recover the received data into the transmitted form. As the rotation of the constellation mapping would cause a significant amount of mismatching with the established reference signal model, the Extended Maximum Likelihood (EML) estimator in [28] is used for pre-processing the signal to recover the phase offset. The phase estimation starts with the calculation of fourth-order complex statistics.

\[ \hat{\xi} = \frac{1}{N} \sum_{n=1}^{N} \rho_n^4 e^{j4\phi_n} = \frac{1}{N} \sum_{n=1}^{N} (\Re\{r(n)\} + j\Im\{r(n)\})^4 \] (20)

\( \rho_n \) and \( \phi_n \) come from the polar expression of the nth signal sample \( r(n) = \rho_n e^{j\phi_n} \) among the total number of \( N \) signal samples. The source kurtosis sum \( \hat{\gamma} \) is also needed in the phase estimation.

\[ \hat{\gamma} = \frac{1}{N} \sum_{n=1}^{N} \rho_n^4 - 8 = \frac{1}{N} \sum_{n=1}^{N} (\Re\{r^2(n)\} + \Im\{r^2(n)\})^2 - 8 \] (21)

The phase offset \( \hat{\theta}_{EML} \) is then estimated using the fourth-order complex statistics and source kurtosis sum calculated previously.

\[ \hat{\theta}_{EML} = \frac{1}{4} \text{angle}(\hat{\xi} \cdot \text{sign}(\hat{\gamma})) \] (22)
Once the phase offset is estimated, it can be easily recovered by conducting the following procedure

\[ \hat{r}(n) = r(n)/e^{j\hat{\theta}_{EML}} \]  

(23)

and the PDF could be treated in the same way as in AWGN channel

\[ f_i(x) = \frac{1}{\hat{\sigma}\sqrt{2\pi}} e^{-\frac{(x-A_i)^2}{2\hat{\sigma}^2}} \]  

(24)

Frequency offset is added to the signal separately from the phase offset. Any frequency offset is treated as noise in the investigation.

3. Classification Methodology

The classification procedure starts with the selection of the optimum sampling locations. Once the optimum sampling locations are established, distribution parameters can be collected at different locations and be used for the decision making. The exact procedure in each step will be discussed in the following subsections. It is worth mentioning that due to the small number of modulations in consideration, the multi-class classification problem is handled by dividing it into two 2-class classification steps. The actual decision procedure is demonstrated in Figure 1. As the proposed method exploits the different CDFs between different M-QAM signal modulations and it is the nature of M-QAM signals to exhibit different distribution on their real and imaginary components, the extension of the proposed method for other M-QAM modulations can be easily implemented following the sampling location optimization principle explained in Section 3.1 and the decision value calculation explained in Section 3.2. However, with different modulation candidates, the performance may vary depending on the specific M-QAM modulation being considered. Lower level M-QAM modulations are normally easier to classify. Modulations with similar constellation shape and similar number of symbols are more difficult to distinguish.

3.1. Sampling Location Optimization

In Kolmogorov-Smirnov Test (K-S test), similarity of two distributions is tested by finding the maximum distance between the two distributions. However, it is limited by the fact
that outliers and other irregularities in the test signal distribution can cause the maximum
distance to occur at a location which does not exhibit the best characteristic difference be-
tween them. The effect becomes more significant when the signal length is reduced or the
amount of noise added is increased. Ultimately, the classification accuracy from different
tests could vary dramatically. To overcome this limitation, the ODST uses multiple sam-
pling locations estimated with theoretical analysis to achieve a more robust performance.

As the later distribution test will be based on CDFs from different signal modulations,
the main purpose of the sampling location optimization is to find locations where the two
CDFs from different modulations exhibit the biggest difference. In this paper, we propose
to use the local optima on the CDFs’ differences as sampling locations.

There are two parameters to consider when searching for sampling locations: number of
sampling points and their locations. Though more information from the distribution could
always help to improve the understanding of the signal, some contribute significantly more
than the others. A simple example would be two adjacent sampling locations which are very
close to each other. Though using both of them would better translate the nature of the sig-
nal as compared to using only one of them, any minor advantage using both is often difficult
to justify the added complexity. Figure 2 gives some examples of signal constellations and
differences between these cumulative distributions. With the proposed location optimization
scheme, it can be seen in Figure 2D that there are eight local optima that can be used for
distribution sampling test. These locations are evenly spread over the signal range and each
of them presents distinct differences between two modulations. Both of these characteristics
are desirable qualities when looking for sampling locations.

We define $\mathcal{L}$ as a collection of sampling locations with $l_k$ being the individual points.

$$\mathcal{L} = \{l_k, \text{ for } k = 1, \ldots, K\}$$

Through extended observations of various type of signals and their distributions, we define
the optimum sampling locations to occur when the difference between CDFs from two classes
is locally optimum, for modulation A-QAM and B-QAM, this can be easily transformed into
the calculation of first derivative of their CDFs’ \((F_A \text{ and } F_B)\) difference

\[
\frac{d}{dx} (F_A(L) - F_B(L)) = 0
\] (26)

The derivative of CDFs’s difference can also be replaced by probability distribution for both modulations \((f_A \text{ and } f_B)\)

\[
f_A(L) - f_B(L) = \sum_{i=1}^{I_A} (f_{Ai}(L)) - \sum_{i=1}^{I_B} (f_{Bi}(L)) = 0
\] (27)

where the PDF for each signal centroids \((f_{Ai} \text{ and } f_{Bi})\) are defined previously in (16). \(I_A\) and \(I_B\) correspond to the total number of centroids for each modulation on one signal dimension.

After the optimization of the sampling locations, the theoretical CDF values at sampling locations for different modulations are collected for classification task as the reference data. The reference data for A-QAM while considering the classification between A-QAM and B-QAM is given as

\[
T_{A,B} = [T_{A,B,1}, ..., T_{A,B,k}]
\] (28)

where

\[
T_{A,B,k} = F_A(l_k)
\] (29)

\(F_A\) is the CDF of modulation A-QAM. These values will be stored for later distribution tests.

Once the sampling locations are established, the distribution sampling could be converted to simple counting tasks. The counted distribution parameter \(t_k\) can be written as

\[
t_k = \frac{1}{2N} [\sum_{n=1}^{N} \mathbb{I}(r_X(n) < l_k) + \sum_{n=1}^{N} \mathbb{I}(r_Y(n) < l_k)]
\] (30)

where \(\mathbb{I}(\cdot)\) is an conditional function which returns 1 if the input is true and 0 if input is false. The counting tasks at different locations are also illustrated in Figure 2A.

3.2. Genetic Algorithm for Decision Metric Optimization and Decision Making

Genetic Algorithm (GA) is an evolutionary algorithm inspired by Darwin’s evolutionary theory. As an optimization tool, GA evolves the solutions through generations of selection
and reproduction of individual solutions. The selection process is based on a fitness criteria defined by the problem at hand. Fitter individuals are selected for producing offsprings in a new generation using different genetic reproduction operators. The solution representation is another important factor in GA design. In ODST, weights are given to each distribution parameters to form a custom distance metric. The collections of these weights are considered as genome strings to be trained in GA. Before different types of genetic coding could be discussed, the standard distance metric needs to be established first.

3.2.1. Standard Distance Metric

The counting results are put into the classification context by finding the difference $\Delta t$ between the counted value and the theoretical value from candidate modulations.

$$\Delta t_{A,B,k} = |t_k - T_{A,B,k}|$$  \hspace{1cm} (31)

where $\Delta t_{A,B,k}$ give the difference between testing signal distribution parameter and reference value $T_{A,B,k}$ from candidate $A$. Likewise the difference between testing signal and candidate $B$ can be found as

$$\Delta t_{B,A,k} = |t_k - T_{B,A,k}|$$  \hspace{1cm} (32)

In the standard uniformly weighted distance metric, the decision is made using all sampled results with the same weight. The decision values for different 2-class classification situations are defined as

$$D_{A,B} = \sum_{k=1}^{K} \Delta t_{A,B,k} - \sum_{k=1}^{K} \Delta t_{B,A,k}$$  \hspace{1cm} (33)

where $D_{A,B}$ compares the distance between testing signal and candidate $A$ and the distance between testing signal and candidate $B$. If $D_{A,B} \geq 0$, it means the tested signal is close to candidate $A$ and thus have a higher probability of being classified as candidate $A$. However as there are more than two candidate modulations involved. The final decision can be made according to a set of decision values.

$$\hat{M} = \begin{cases} 4, & D_{4,16} \leq 0 \text{ & } D_{4,64} < 0 \\ 16, & D_{4,16} > 0 \text{ & } D_{16,64} \leq 0 \\ 64, & D_{4,64} \geq 0 \text{ & } D_{16,64} > 0 \end{cases}$$  \hspace{1cm} (34)
The resulting $\hat{M}$ gives the estimated $M$ value for the tested M-QAM signals. $D_{4,16}$, $D_{4,64}$ and $D_{16,64}$ are decision values gathered from the previous stage.

3.2.2. GA Optimized Weighted Distance Metrics

Given the distance definition of equation (31), it is worth questioning the actual contribution of each sampling locations. Though the optimization process attempts to find the best locations with maximum amount of separation while conveying the full characteristic of the CDFs, it is still possible for the local optima to be inefficient. For example, two local optimums can be very close to each other and express the same signal trait. As can be seen in Figure 3, the four locations in the middle get closer when the SNR is less than 9 dB and become effectively same locations at around 7 dB. Then all four locations are no longer selected as optimum sampling locations. Based on the behaviour of these four optimized locations, it is easy to doubt their contribution to the classification task for SNRs between 7 dB to 9 dB. It is also verified in the simulation results that these locations are normally abandoned or given a lower weight.

To justify the use of specific sampling locations, GA has been used to find the best selection of these locations to enhance the decision making procedure. Here the distance metric is redefined with the addition of weights on each sampled distribution parameters.

$$D_{A,B} = \sum_{k=1}^{K} W_{A,B,k} \Delta t_{A,B,k} - \sum_{k=1}^{K} W_{A,B,k} \Delta t_{B,A,k}$$ (35)

There are two types of constraints considered while training the weights. The first limits the weights to binary values (GA-Bin), when $W_{A,B,k} = 0$ the distribution test result at location $k$ will not be included and when $W_{A,B,k} = 1$ the result would be considered. As the training phase evolves for a long time, the trained weights can be an indication of the best selection of sampling locations. The second type was experimented with the linear combinational weights limited to values between 0 and 1 (GA-Lin), so that the trained results could provide a more versatile combination of the decision values. Both cases share the same fitness evaluation approach. The fitness value is obtained through a small classification task using a small set of testing signals. The accuracy of the small classification is directly used the
fitness value. Therefore, fitter individuals always have higher fitness values. Other GA parameters can be found in Table 1.

4. Simulation Setup

4.1. Signal Generation

All experiments are simulated in computer based environment and signals were first created as symbols, randomly drawn from specific modulation mapping in a uniform manner. White Gaussian noise is generated using custom function following the definition of SNR in equation (5). If phase or frequency offset is to be considered, the native MATLAB function is used to implement the channel effects. Additive white Gaussian noise is also included under these channel conditions. Before classification, sampling locations and theoretical reference distribution test values for SNR range from 0 dB to 25 dB with 1 dB step are collected and stored. In the given SNR range, it is discovered that twelve sampling locations are found in each SNR scenario between 0 dB and 7 dB, and sixteen sampling locations are found when SNR is between 8 dB and 25 dB. As two reference CDF values are needed at each sampling location to complete the decision value calculation, there are a total number of 768 reference values prepared for each given signal length.

During GA optimization, the fitness function is defined in the same way as in classification problems where the classification accuracy is used as the fitness value. To reduce the complexity of the training stage, only 1000 realizations from each modulation with a signal length of \( N = 512 \) samples are used in the fitness evaluation process. The training is repeated five times for each SNR value ranging from 0 dB to 10 dB. All signal data are generated randomly at every fitness evaluation, which avoids weights being over-trained for a specific set of signal data. In addition, with the two elites always being passed on to the next generation, the possibly best solutions are always protected to some degree. The training was repeated for five runs under each signal condition. The collections of weights which give the best performance were selected for performance assessment with larger statistics at
4.2. Benchmarking Classifiers

When testing the performance of the proposed solution, Maximum Likelihood (ML) classifier, the K-S test, cumulant based Genetic Programming (GP) and K-Nearest Neighbour (KNN) classifiers were used for benchmarking purpose.

4.2.1. Maximum Likelihood Classifier

To establish an upper bound of the classification performance, ML classifier is used here to help understand the performance of the proposed method. The classification decision comes from the hypothesis $H_M$ with an underlying constellation from M-QAM, whose likelihood function $l(H_M|r_N)$ is maximized. Given that $r_N$ is a set of signal samples received.

$$\hat{H}_M = \arg \max_{H_M} l(H_M|r_N)$$

In the experiments, the likelihood function is used as suggested in [12]. $M$ denotes the total number of possible constellation centroid from modulation M-QAM and $A_i$ being the actual values of these centroids.

$$l(H_M|r_N) = \sum_{n=1}^{N} \left\{ \frac{1}{M} \sum_{i=1}^{M} \exp \left( -\frac{1}{2} \|r_n - A_i\|^2 \right) \right\}$$

ML is only used in AWGN channel to illustrate an upper bound of the classification performance. Due to the fact that ML needs very accurate channel modelling and is very vulnerable against other channel conditions, it is excluded in the performance comparison in flat fading channel.

4.2.2. K-S Test

The K-S test method has been developed in [18]. It is also used for benchmarking to evaluate the performance as well as the computational complexity. The K-S test starts with decision values defined as

$$D_M = \max_{1<n<N} |F(r(n)) - F_M(r(n))|$$
where $F(.)$ is the CDF of the signal being classified and $F_M(.)$ is the theoretical CDF of the candidate signal modulation. Once the decision values from every candidate modulation are collected, the classification decision will be based on the following criteria:

$$\hat{M} = \arg \min_{M=4,16,64} D_M$$

(39)

4.2.3. Cumulant Based GP-KNN

The popularity of cumulant based machine learning AMC classifiers has already been mentioned in the introduction. Here we have used GP-KNN [24] as an example to demonstrate the performance difference between GP-KNN and ODST. Genetic programming is an evolutionary machine learning algorithm inspired by natural evolution [29]. It is used in GP-KNN to evolve a combination of cumulant features to be used in KNN classifier. The cumulants included are $C_{40}, C_{41}, C_{42}, C_{60}, C_{61}, C_{62}$ and $C_{63}$ where each can be calculated by using the following equation.

$$C_{ij} = cum(r(n),..,r(n),r^*(n),..,r^*(n))$$

(40)

4.3. Performance Testing

For the performance test in AWGN channel, two sets of experiments were conducted. The first set of experiments focused on the classification accuracy under different noise levels. Here, the signal length is fixed at 512 samples with the SNR ranging from 0 dB to 25 dB. Then classifications of 100,000 signal realizations from each modulation were tested using ML, K-S test, GP-KNN and the proposed ODST classifier. The successful classification percentage was calculated based on the number of successful classifications and the total number of signal realizations. The results are presented in Figure 4. In the second set of experiments, we tried to understand how the signal length influences the classification performance. In this case, similar settings were used, except for SNR being fixed at 10 dB and sample size to vary from 100 to 1000. The results are presented in Figure 6.

In flat fading channel, the signal length was fixed at $N = 512$ samples and the SNR at 10 dB. Again 100,000 signal realizations from each modulation were tested under separate
conditions of phase and frequency offset. In the experiment for phase offset, the range of offset is controlled within 10 degrees. This is purely for testing the performance of classifiers when handling conditions with inaccurately estimated phase offset. Also, the combination of the proposed method, EML phase estimation and recovery is tested to evaluate its performance. Results are presented in Figure 7. When considering frequency offset, the amount of offset is limited in the range of $1 \times 10^{-4}$ and $2 \times 10^{-4}$.

5. Results and Discussion

In this section, results collected from simulation tests are presented with detailed analysis. The computational complexity is also discussed.

5.1. Classification Performance

The classification performance under different amount of additive noise has always been the prime criteria for an AMC solution. In Figure 4, four different types of AMC classifiers are included. It is clear that ML provides the most accurate classification throughout the SNR range. Excluding the ML classifier, the results show that the proposed ODST classifier has a clear advantage in mid to high SNRs. At 10 dB, the proposed method achieves almost the same accuracy of 98.9% as the ML classifier and the 100% classification is achieved at 11 dB. At the same SNR settings, K-S test provides a successful classification of 95.3% and the perfect classification performance is achieved at 12 dB.

For cumulant based GP-KNN classifier, it can be seen that its performance is limited by the signal length that is available for analysis. In the mid and lower range of SNRs, the proposed ODST classifier maintains the advantage over K-S test. The biggest difference is exhibited at 9 dB where ODST offers an accuracy of 93.9% and K-S test offers 88.6%. However, the accuracy advantage is gradually reduced along with the decreasing SNR until the performance become equivalent below 3 dB. On the other hand, this cumulant based GP-KNN classifier shows a robust performance in low SNRs, offering better classification performance from 3 dB to 8 dB against ODST and from 3 dB to 9 dB than K-S test. The performance at SNR below 3 dB is generally very similar among all classifiers with only ML
classifier having a more than 5% higher accuracy. Complementary results from ODST for different modulations are listed in Table 2. Performance means and standard deviations are collected from 100 sets of tests, each includes 30,000 signal realizations (three modulations times 10,000 signal realizations from each modulation).

In addition to the benchmarking classifiers, several existing classifiers from other literature have been listed in Table 3 for performance comparison with ODST. Results for ODST come from experiments conducted under the same specific condition as each existing classifiers. It is clear that the proposed classifier outperforms the K-S classifier [18], the reduced complexity version of K-S classifier (rcKS) [20], phase based ML classifier [15], as well as cumulant based classifiers [5], [30]. The Minimum Distance (MD) [16] classifier, which is a low-complexity version of the ML classifier, presents similar level of performance at or above 14 dB as compared to the proposed ODST classifier. However, with the SNR at or lower than 10 dB, its classification accuracy is significantly degraded. The comparison between MD classifier and ODST classifier at SNR of 10 dB clearly demonstrates the performance advantage of the proposed method.

Having analyzed the performance of ODST against other existing AMC classifier, let us have a look at the effect of GA optimized weighted decision making on the classification performance. The same experimental setup is used only with SNR limited between 0 dB and 10 dB to investigate the effect of GA optimization on low SNR performance. According to the classification performance in Figure 5, both GA optimized classifiers follow the performance degradation pattern of the original ODST with an increase in classification accuracy of 1% to 3% sustained over the SNR range. The biggest performance improvement is shown between SNR of 7 dB to 10 dB. At 8 dB, GA optimized ODST with analogue weight achieves a classification accuracy of 90.5% providing the largest performance improvement of 4% as compared to the 86.5% classification accuracy of the original ODST classifier. The reason for such improvement can be explained with the analysis of sampling location quality in Section 3. In Figure 3, it is clear that some of the sampling locations start to merge and disappear between 7 dB and 10 dB. The performance improvement provided by GA optimized weights verified these sampling locations need to be given lower weights to
achieve better classification performance. Between the binary weights and analogue weights, analogue weights provide better performance at 8 dB, 9 dB and 10 dB while being almost equal to the binary weights from 0 dB to 7 dB. Overall, both types of optimized weights help to improve the classification by a fair amount.

The robustness against a limited signal length is another important quality for a good AMC classification. In the experiments, same four classifiers are tested and compared in Figure 6. Again, ML excels in all signal length from $N = 100$ to $N = 1000$. Excluding ML classifier, ODST is the best among the remaining classifiers. The largest performance difference of ODST against ML is about 5% at $N = 100$. As the signal length increases the difference starts to reduce and at $N = 600$ ODST achieves performance similar to ML classifier. When compared with K-S test, OSST shows a superior robustness especially when the signal length is in the range from $N = 150$ to $N = 500$. The biggest advantage of ODST is observed at $N = 250$, where K-S test returns a classification accuracy of 93.0%, which is 1.7% below ODST’s 94.7%. Unfortunately, culumant based GP-KNN classifier suffers severely with the reduced signal length. However, as its performance is improving consistently with the increasing signal length, it is clear that, with large enough signal length, GP-KNN classifier is still able to achieve equal level of performance.

In the flat fading channel with unknown phase offset, we have included the original ODST classifier, the original K-S test and ODST classifier with EML phase estimation and recovery. The results are presented in Figure 7. All signals are simulated with a signal length of $N = 512$ and SNR of 10 dB. With no phase error, the classification accuracy difference between the original ODST and K-S test coincide the results in pure AWGN channel. The original ODST starts with an advantage of 3.4%. As more phase offset is introduced, both classifiers’ performance starts to degrade. Nevertheless, ODST sees less degradation before the phase offset reaches $\theta_o = 6^\circ$. Once again, this illustrates the robustness of ODST when compared with K-S test. The degradation of ODST performance accelerates after 6 dB. At $\theta_o = 8.3^\circ$, K-S test surpass ODST to have a better performance with more phase offset. It is an understandable phenomenon, as the ODST relies on an accurate signal model more than the K-S test, when the signal model mismatching exceeds a certain level, the
distribution tests at different locations become barely capable of providing positive contribution towards an accurate classification. Nevertheless, when ODST is teamed up with an accurate phase offset estimation and recovery scheme, this should not be a concern since the mismatching could be limited within a reasonable amount. It is demonstrated with the results from ODST-EML. Regardless of the amount of phase offset experimented with, the classifier delivers a consistent classification accuracy of 98.8%. Under similar conditions, ML classifier and GP-KNN classifier have both exhibited a strong robustness seeing less than 10% degradation in classification accuracy.

As can be seen in Figure 8, both ODST and K-S test perform poorly when frequency offset is considered. With a frequency offset of $1 \times 10^{-4}$ to $2 \times 10^{-4}$, classification accuracy from both classifier drops significantly. For ODST, its classification accuracy is reduced to 95.5% with a frequency offset of $1 \times 10^{-4}$. As the amount of frequency offset increases to $2 \times 10^{-4}$, the classification performance decreases almost linearly to 77%. The K-S test sees similar performance degradation. However, it starts with lower classification accuracy of 92% with frequency offset at $1 \times 10^{-4}$ and reduces to 77% with frequency offset of $2 \times 10^{-4}$. The ODST classifier provides about 3.5% better classification accuracy between $1 \times 10^{-4}$ to $1.3 \times 10^{-4}$. The performance advantage is gradually reduced beyond $1.3 \times 10^{-4}$. One of the causes of this reduced performance comes from the modulations being used, especially 16-QAM and 64-QAM. With their dense signal constellations, there is little room for any frequency offset. The other reason is to do with the nature of distribution test based classifiers, which rely on a solid signal distribution with little frequency shifting. Even though ODST performs better than K-S test, it is difficult to claim its robustness under channels with frequency offsets. Although the frequency offset condition is optimistic, some effective blind frequency offset estimation and compensation approaches for QAM modulated signals have been developed (e.g.,[31]) which would help to achieve the required level of frequency offset.
5.2. Complexity Analysis

The numbers of different operations required by different classifiers are listed in Table 4. It is obvious that the implementation of ML classifier requires exponential and logarithm operation while others do not. The MD classifier significantly reduced the complexity of ML classifier since no exponential or logarithm operation is needed. However, a considerable amount of multiplication and addition are still needed which is similar to the process of cumulant calculation. When comparing K-S test and ODST, given the signal length used and number of different modulation candidates, it is clear that the number of additions used is similar while the memory usage is much lower for ODST. If longer signal length is to be analyzed or more modulations are included, the complexity advantage of ODST will be more evident. Although there is considerable amount of complex computation involved in the training of weights in GA optimized ODST, it is worth clarifying that it is done offline beforehand and will not be repeated for every classification task. Thus only the sampling and decision making should be considered when evaluating the complexity of ODST. With compromised classification performance robustness, the reduced complexity version of K-S classifier, which compares the CDFs at signal signal point, requires fewer additions as well as less memory.

6. Conclusion

A novel distribution based multi-sampling test is proposed for the purpose of classifying M-QAM modulations in a robust manner with little requirement on computational power. It has been demonstrated that the proposed ODST classifier outperforms similar distribution test based algorithms such as K-S test in most channel conditions including AWGN channel, unknown phase offset and unknown frequency offset. In addition, the performance is further enhanced with the adoption of GA and EML estimator. It is also demonstrated in this paper that the proposed ODST classifier has a much lower computation complexity as compared to K-S test, ML and cumulant based classifiers. Comparison with other lower-complexity classifiers methods, including the Minimum Distance classifier and the reduced
complexity version of K-S classifier, has shown that given comparable level of computation resource, the ODST classifier achieve better classification accuracies at different noise level in the AWGN channel. Further development of ODST can be accomplished to suit the requirements of more complex channel conditions better. The requirement of knowledge of channel parameters, such as SNR, can be relaxed to create a more independent classification system.

Acknowledgments

Zhechen Zhu would like to thank the School of Engineering and Design, Brunel University for their financial support. Muhammad Waqar Aslam would like to acknowledge the financial support of the University of Azad Jammu and Kashmir, Pakistan. Asoke K. Nandi would like to thank TEKES for their award of the Finland Distinguished Professorship.

Appendix

The variance of received signal is given as

$$\sigma_r^2 = \frac{\sum_{n=1}^{N} (r_X(n) - \mu_x)^2}{N}$$ (41)

where $N$ is the signal length, $\mu_X$ is the signal mean and $r_X(n)$ being the received signal in AWGN channel.

$$r(n) = \alpha s(n) + g(n)$$ (42)

Because the signals are assumed to have zero mean, the term $\mu_X$ can be removed. Combine Equation (41) and (42), we can get

$$\sigma_r^2 = \frac{\sum_{n=1}^{N} (r_X(n))^2}{N} = \frac{\sum_{n=1}^{N} (\alpha s_X(n) + g_X(n))^2}{N}$$ (43)

$s'(n)$ can be replaced by $A_i$ which come from an equiprobable collection of modulation dependent signal centroids.

$$\sigma_r^2 = \sum_{i=1}^{I} \frac{\sum_{n=1}^{N/I} (\alpha A_i + g_X(n))^2}{N}$$ (44)
\[
\sigma_r^2 = \sum_{i=1}^{I} \sum_{n=1}^{N/I} \frac{\alpha_i^2 A_i^2 + 2\alpha_i A_i g_X(n) + g_X^2(n)}{N} \tag{45}
\]

As we assume the noise is symmetric around each centroids due to the underlying Gaussian distribution, we can get

\[
\sigma_r^2 = \sum_{i=1}^{I} \sum_{n=1}^{N/I} \frac{\alpha_i^2 A_i^2 + g_X^2(n)}{N} \tag{46}
\]

The term with noise energy can be replaced with noise variance.

\[
\sigma_r^2 = \sum_{i=1}^{I} \frac{\alpha_i^2 A_i^2}{I} + \sigma^2 \tag{47}
\]

References


Figure 1: Modulation Classification Decision Tree. Decision making conditions and decision values $D_{416}$, $D_{464}$ and $D_{1664}$ are explained in Section 3.2.
Figure 2: (A) 500 signal samples from 16-QAM at 15 dB on the I-Q plane, (B) 500 signal samples from 64-QAM at 15 dB on the I-Q plane, (C) The CDFs from 16-QAM and 64-QAM, and (D) The difference between the two CDFs. The dashed lines indicate the shared optimized sampling locations.
Figure 3: Optimized sampling locations at different SNRs. The crosses are optimal sampling locations found at different SNRs for the classification between 16-QAM and 64-QAM.
Figure 4: Classification Accuracy vs. SNR with 300,000 signal realizations from 4-QAM, 16-QAM and 64-QAM each consists of 512 samples.
Figure 5: Classification Accuracy vs. SNR with 300,000 signal realizations from 4-QAM, 16-QAM and 64-QAM each consists of signal of length $N = 512$. 

Classification Accuracy (%) vs. SNR (dB)
Figure 6: Classification Accuracy vs. Signal Length with 300,000 signal realizations from 4-QAM, 16-QAM and 64-QAM all under SNR of 10 dB.
Figure 7: Classification Accuracy vs. Phase Offset with 300,000 signal realizations from 4-QAM, 16-QAM and 64-QAM each consists of signal length $N = 512$ under SNR of 10 dB.
Figure 8: Classification Accuracy vs. Frequency Offset with Test of 300,000 signal realizations from 4-QAM, 16-QAM and 64-QAM each consists signal length $N = 512$ under SNR of 10 dB.
Table 1: Settings for the Genetic Algorithm

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Case 1</th>
<th>Case 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constraint</td>
<td>Binary</td>
<td>$0 \leq W \leq 1$</td>
</tr>
<tr>
<td>Generation</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Population</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Elite Count</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Crossover Fraction</td>
<td>60%</td>
<td>80%</td>
</tr>
<tr>
<td>Mutation Type</td>
<td>Uniform</td>
<td>Uniform</td>
</tr>
<tr>
<td>Mutation Rate</td>
<td>60%</td>
<td>40%</td>
</tr>
</tbody>
</table>
Table 2: Classification accuracy in percentage for different modulations with $N = 512$. Each cell contains results from 100 sets of tests conducted with 10,000 signal realizations in each test.

<table>
<thead>
<tr>
<th>Modulations</th>
<th>5 dB</th>
<th>10 dB</th>
<th>15 dB</th>
<th>20 dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-QAM</td>
<td>100.0 ±0.0</td>
<td>100.0 ±0.0</td>
<td>100.0 ±0.0</td>
<td>100.0 ±0.0</td>
</tr>
<tr>
<td>16-QAM</td>
<td>68.2 ±0.4</td>
<td>98.5 ±0.1</td>
<td>100.0 ±0.0</td>
<td>100.0 ±0.0</td>
</tr>
<tr>
<td>64-QAM</td>
<td>65.9 ±0.5</td>
<td>98.1 ±0.1</td>
<td>100.0 ±0.0</td>
<td>100.0 ±0.0</td>
</tr>
</tbody>
</table>
Table 3: Performance comparison with existing classifiers in other publications.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Modulations</th>
<th>Channel</th>
<th>Setting</th>
<th>Accuracy</th>
<th>ODST</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-S 2-D</td>
<td>4-QAM, 16-QAM, 64-QAM</td>
<td>AWGN</td>
<td>$N = 100; 10 \text{ dB}$</td>
<td>78.0%</td>
<td>85.2%</td>
</tr>
<tr>
<td>K-S magnitude</td>
<td>4-QAM, 16-QAM, 64-QAM</td>
<td>AWGN</td>
<td>$N = 100; 14 \text{ dB}$</td>
<td>87.0%</td>
<td>99.6%</td>
</tr>
<tr>
<td>rcKS</td>
<td>4-QAM, 16-QAM, 64-QAM</td>
<td>AWGN</td>
<td>$N = 50; 10 \text{ dB}$</td>
<td>72.5%</td>
<td>77.3%</td>
</tr>
<tr>
<td>Phase Based ML</td>
<td>4-QAM, 16-QAM</td>
<td>AWGN</td>
<td>$N = 1000; 0 \text{ dB}$</td>
<td>73.5%</td>
<td>82.3%</td>
</tr>
<tr>
<td>MD</td>
<td>4-QAM, 16-QAM</td>
<td>AWGN</td>
<td>$N = 100; 10 \text{ dB}$</td>
<td>50.0%</td>
<td>86.5%</td>
</tr>
<tr>
<td>MD</td>
<td>4-QAM, 16-QAM</td>
<td>AWGN</td>
<td>$N = 100; 14 \text{ dB}$</td>
<td>91.5%</td>
<td>97.3%</td>
</tr>
<tr>
<td>Cumulants</td>
<td>16-QAM, 64-QAM</td>
<td>Noise Free</td>
<td>$N = 10,512$</td>
<td>90.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Cumulants Bayes</td>
<td>4-QAM, 16-QAM, 64-QAM</td>
<td>AWGN</td>
<td>$N = 512; 10 \text{ dB}$</td>
<td>87.3%</td>
<td>98.5%</td>
</tr>
</tbody>
</table>
Table 4: Operations needed for different classifiers. $N$ is the signal sample length. $M$ is the number of candidate modulations, $I_m$ represent the total number of centroids for modulation $m$, and $K$ is the number of sampling locations.

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>Multiplier</th>
<th>Addition</th>
<th>Exponential</th>
<th>Logarithm</th>
<th>Memory</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML</td>
<td>5$NM \cdot \sum_{m=1}^{M} I_m$</td>
<td>$6NM \cdot \sum_{m=1}^{M} I_m$</td>
<td>$NM \cdot \sum_{m=1}^{M} I_m$</td>
<td>$NM$</td>
<td>$M$</td>
</tr>
<tr>
<td>MD</td>
<td>2$NM \cdot \sum_{m=1}^{M} I_m$</td>
<td>$NM \cdot (\sum_{m=1}^{M} 3I_m + 1)$</td>
<td>0</td>
<td>0</td>
<td>$M$</td>
</tr>
<tr>
<td>Cumulants</td>
<td>6$N$</td>
<td>6$N$</td>
<td>0</td>
<td>0</td>
<td>$M$</td>
</tr>
<tr>
<td>K-S test</td>
<td>0</td>
<td>2$N(2M + \log 2N)$</td>
<td>0</td>
<td>0</td>
<td>$MN$</td>
</tr>
<tr>
<td>rcKS/rcK</td>
<td>0</td>
<td>$4N \cdot \binom{M}{2}$</td>
<td>0</td>
<td>0</td>
<td>$4M \cdot \binom{M}{2}$</td>
</tr>
<tr>
<td>ODST</td>
<td>0</td>
<td>$4N \cdot \binom{M}{2} \cdot K$</td>
<td>0</td>
<td>0</td>
<td>$4M \cdot \binom{M}{2} \cdot K$</td>
</tr>
</tbody>
</table>