Blind Signal PSK/QAM Recognition Using Clustering Analysis of Constellation Signature in Flat Fading Channel

Gaurav Jajoo, Yogesh Kumar, and Sandeep Kumar Yadav

***Abstract*— A novel method based on constellation structure is proposed to identify PSK and QAM modulation of different orders, in the slow and flat fading channel. The proposed method does not require training for threshold optimization and considers carrier frequency, symbol rate, and phase offset unknown. The symbol rate is estimated using the spectrum of the instantaneous phase of the complex baseband signal. Carrier frequency offset (CFO) is estimated and corrected from the downconverted signal and downsampled to the estimated symbol rate for extraction of constellation points. The phase offset is determined based on the symmetrical structure of constellation. The features extracted using k-medoids are used for classification of the final modulation scheme. Results show that the proposed algorithm outperforms some existing classifiers and offers lower computational complexity compared to algorithms based on subtractive clustering.**

***Index Terms*— Blind Signal, symbol rate estimation, phase offset, rayleigh fading, modulation classification.**

I. Introduction

BLIND signal modulation recognition identifies the modulation format in unknown RF signal and has various applications in the field of the military such as electronic warfare, jamming, finding the hostile signal, etc. It is an integral part of technologies like cognitive radios and software defined radios. In today's era of fixed licensed spectrum, automatic modulation classification (AMC) seems to be an optimal solution for spectrum under-utilization.

Different classification algorithms are developed so far given in [1] and are broadly categorized into Likelihood-based (LB) and Feature-based (FB) methods. LB methods are multiple hypothesis testing problem, in which likelihood function for all modulation schemes considered is evaluated and compared for final modulation classification. LB methods give optimal classification accuracy but require perfect channel knowledge else its accuracy reduces. Different likelihood ratio tests are proposed for computing likelihood function by maximizing over the unknown channel parameters [2]. Likelihood function becomes complex for higher order modulations; therefore, less complicated FB methods are developed.

In FB methods, different features like cumulants, moments, order statistics [3], spectral-based features, graph-based features [4], etc. are extracted from the detected signal and compared with threshold or followed by pattern recognizer for the final classification. Different machine learning techniques like ANN, KNN, stacked encoders, and GP-KNN [5] are used as classifiers, which are complex and require an offline training. Hence goodness-of-fit test based classifiers [6], [7] are used for reliable performance with lower complexity. Unsupervised learning techniques like clustering and regression on constellation signature are also used for modulation classification of blind signal [8].

Many papers in the literature assume symbol duration, phase offset, or noise variance to be known and consider constellation points for modulation classification directly [4]. In this paper, the AMC system is considered as a black box and different parameters viz. carrier frequency offset (CFO), symbol rate, and phase offset of the RF signal are estimated. The carrier frequency of the RF signal is estimated using the center of 90% power of the signal in the desired band and signal is downconverted. CFO from the downconverted signal is evaluated and corrected. CFO corrected signal is downsampled to the estimated symbol rate for extraction of constellation signature. After reconstructing the constellation, different number of medoids are estimated blindly using k-medoids clustering. The final modulation scheme is identified using amplitude and phase of estimated medoids.

II. Signal Model

At the receiving end, downconverted and sampled signal $y[n]$ can be expressed as

$$y[n] = \alpha e^{j2\pi f_0 n t_s} x[n] + w[n] \tag{1}$$

In (1), $x[n]$ is the transmitted baseband signal with root raised cosine (RRC) pulse shaping of random roll-off factor, t_s is the sampling duration, α is the complex-valued channel fading gain used to represent flat fading experienced by the signal. The algorithm is developed for narrowband signal considering signal bandwidth less than coherence bandwidth. Therefore, each frequency component experiences the same value of the block-fading coefficient. Such systems can be represented by a single tap channel filter (α). f_0 is the carrier frequency offset, which occurs due to a mismatch in RF signal frequency and local oscillator frequency. $w(n)$ is the Gaussian noise with zero mean. CFO in $y[n]$ is estimated using eighth order non-linearity and corrected using [8]. CFO corrected signal is resampled to the integer multiple of estimated symbol rate for match filtering and further downsampled to symbol rate for constellation extraction. AWGN and slow and flat fading channels are considered. Received symbol $r[n]$ are expressed as

$$r[n] = \alpha s[n] + w'[n]$$

Manuscript received May 9, 2019; revised June 17, 2019; accepted July 9, 2019. Date of publication July 16, 2019; date of current version October 9, 2019. The associate editor coordinating the review of this letter and approving it for publication was F. Wang. *(Corresponding author: Gaurav Jajoo.)*

The authors are with the Department of Electrical Engineering, Indian Institute of Technology Jodhpur, Jodhpur 342037, India (e-mail: jajoo.1@iitj.ac.in; kumar.33@iitj.ac.in; sy@iitj.ac.in).

Digital Object Identifier 10.1109/LCOMM.2019.2929127

$s[n]$ is extracted symbol from one of the considered modulation scheme. SNR for symbols with unit power can be defined as

$$SNR = 10log_{10}\frac{\alpha^2}{\sigma^2_{w'}}$$

III. Symbol Rate Estimation and Constellation Reconstruction

A new method for symbol rate (SR) estimation has been proposed, which is independent of phase offset. As PSK and QAM signals have a phase change at every symbol duration, the spectrum of the instantaneous phase after passing through average filter shows a peak at the symbol rate. Phase of a complex baseband signal with rectangular pulse shaping has discrete phase levels. Let $\phi[k]$ be the phase represented by:

$$\phi[k] = \sum_{n=1}^{N} A_n p(k - nT_s) \quad (2)$$

In (2), A_n is discrete phase levels which change at every symbol duration T_s. For a pulse shaping filter other than rectangular such as raised cosine or RRC, $p(.)$ take the values accordingly. Fig. 1 is obtained for QPSK with a sampling frequency of 160KHz, symbol rate of 10 KHz, and RRC pulse shaping. In Fig. 1(a), a, b, d, e are constellation points and c, c', f are different instances of transition between symbol points. Fig. 1(b) shows the instantaneous phase ranging from $[-180^0 180^0)$, which has abrupt changes when transition occurs from $c \leftrightarrow c'$. Other abrupt changes come, when transition follows a path $d \leftrightarrow f \leftrightarrow a$ where angle changes from $(-135^0 \leftrightarrow 45^0)$. A moving average filter is used to smoothen these abrupt changes, and $\phi'[k]$ is obtained as shown in Fig. 1(c) and represented by

$$\phi'[k] = \sum_{n=1}^{N} A_n p'(k - nT_s) \quad (3)$$

Squaring (3) and suppressing the DC term, $q[k]$ is obtained shown in Fig. 1(d).

$$q[k] = \phi'^2[k] - E(\phi'^2[k])$$

$Q(\omega)$ is calculated by taking Fourier transform (represented by $\mathscr{F}$) of $q[k]$.

$$\begin{aligned}
Q(\omega) &= \mathscr{F}\sum_n[(A_n^2)p'^2(k-nT_s)] - \mathscr{F}[E(\phi'^2[k])] \\
Q(\omega) &= \mathscr{F}\sum_n[(A_n^2)p'^2(k-nT_s)] - K \\
Q(\omega) &\overset{(a)}{=} \sum_n A_n^2 \mathscr{F}[p'^2(k) \circledast \delta(k-nT_s)] - K \\
&\overset{(b)}{=} \sum_n A_n^2 \mathscr{F}[p'^2(k)]\mathscr{F}[\delta(k-nT_s)] - K \\
&\overset{(c)}{=} \sum_n A_n^2 [P'(\omega) \circledast P'(\omega)]\frac{2\pi}{T_s}\delta(\omega - n\frac{2\pi}{T_s}) - K \\
Q(\omega) &= \frac{2\pi}{T_s}[P'(\omega) \circledast P'(\omega)] \sum_{\forall n, n\neq 0} A_n^2 \delta(\omega - n\frac{2\pi}{T_s}) \quad (4)
\end{aligned}$$

Here $\circledast$ represents convolution. (a) follows from the fact that $h[n] \circledast \delta[n-k] = h[n-k]$. (b) and (c) follows from convolution and multiplication property of Fourier transforms.

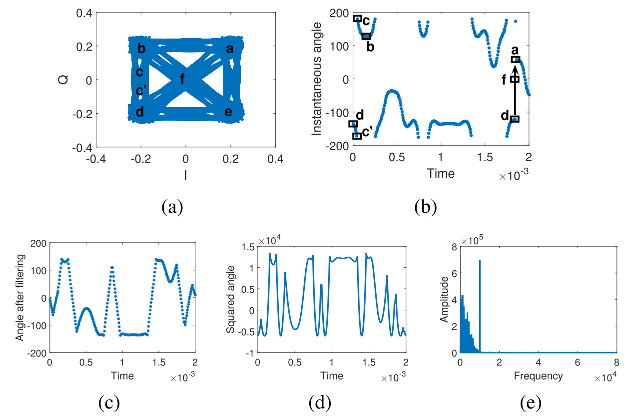


Fig. 1. (a) Root raised cosine pulse shaped QPSK constellation with transitions. (b) Instantaneous angle of 1(a). (c) Phase after passing 1(b) through moving average filter of length 9. (d) Graph obtained after squaring and shifting 1(c) to 0 mean. (e) Spectrum of 1(d) with peak at symbol rate of 10 KHz.

Here K is equivalent to $Q(\omega)$ for $n = 0$. $Q(\omega)$ is shown in Fig. 1(e).

For validation of peak in $Q(\omega)$, length of the moving average filter is varied and corresponding to each length, the sharpness of peak (G_l) is calculated using (5).

$$G_l = \frac{\sum_{k=i-100}^{i+100}(max(Q_l[f]) - Q_l(k))}{200} \quad (5)$$

Here i is index value, corresponding to the maximum of $Q_l(f)$ for a filter length l. The frequency corresponding to a maximum value of G_l for any l is estimated as SR. Signal $y[n]$ given in (1) is used for constellation extraction using estimated SR. Extracted constellation is normalized to zero mean and unit power.

IV. Proposed Algorithm

After extracting the constellation points, the proposed algorithm estimates the domain of modulation, i.e., PSK or QAM followed by phase offset estimation and correction. After correcting phase offset, modulation order is estimated.

A. Modulation Domain Identification

Nine medoids are calculated from constellation points to differentiate PSK and QAM. The amplitudes of all medoids are sorted in ascending order. Standard deviation (SD) of the contiguous amplitude difference (represented by a) is calculated and compared with the threshold ($a_{th} = 0.26$) for classification of the domain, i.e., QAM and PSK.

B. Phase Offset Estimation

If the domain classified is QAM, i.e., for $a > a_{th}$, phase offset is calculated using [9], but this method fails for PSK. Hence, a new method for phase offset estimation of PSK modulation is proposed. Due to phase offset, the constellation rotates with an angle θ_0 for single signal realization in slow fading. Each CFO corrected constellation point $d[n]$, has its own phase with an extra phase of θ_0 ($\angle\alpha = \theta_0$).

$$d[n] = |\alpha|e^{j\theta_0}s[n] \quad (6)$$

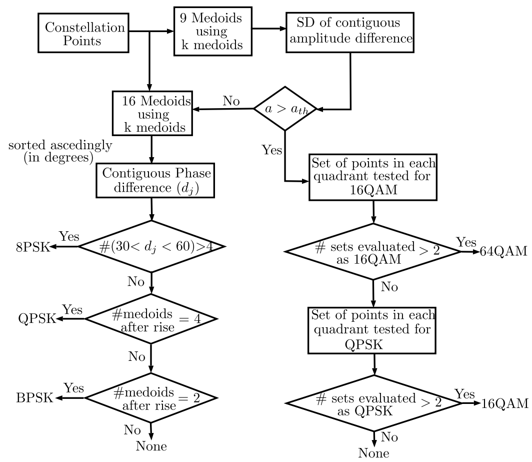


Fig. 2. Block diagram of modulation classification algorithm.

Sum of all arguments of $d[n]$ is defined by ψ.

$$\psi = \sum_{n=1}^{N} arg(d[n])$$
$$\psi = (\theta_1 + \theta_0) + (\theta_2 + \theta_0) \ldots + (\theta_N + \theta_0)$$
$$\psi = \sum_{i=1}^{N} \theta_i + N\theta_0$$

As constellation has symmetric structure, $\frac{1}{N}\sum \theta_i \approx 0$ for large number of equiprobable symbols. Hence, approximate value of θ_0 is given by

$$\theta_0 \approx \frac{\psi}{N} \tag{7}$$

C. Modulation Scheme Estimation

Phase offset corrected constellation points are given to the proposed algorithm shown in Fig. 2. If PSK modulation is observed, 16 medoids are estimated using k medoids clustering. Fig. 3(a) and Fig. 3(b) shows the phase offset corrected BPSK and QPSK constellation, respectively. Fig. 3(c) and Fig. 3(d) shows 16 estimated medoids for Fig. 3(a) and Fig. 3(b), respectively. Phases of all 16 medoids are sorted in ascending order. The contiguous difference of sorted phases is calculated and again sorted in ascending order. If the number of differences between 30^0 and 60^0 is greater than 4, the decision is given in favor of 8-PSK. Otherwise, number of points after a rise provides an estimate of modulation order 4 or 2, i.e., QPSK or BPSK, as shown in Fig. 3(e) and Fig. 3(f), respectively.

If the domain falls under QAM, constellation points of all four quadrants are considered separately for estimation of order. Set of points in each quadrant are shifted to zero mean and tested for 16-QAM. If more than two sets give a decision as 16-QAM, classification result is given in favor of 64-QAM else it is tested for 16-QAM. For classification of 16-QAM, points in each quadrant are shifted to zero mean, and eight medoids are estimated. These shifted medoids are tested for QPSK as points in each quadrant of the 16-QAM constellation has a structure similar to QPSK. Fig. 4 shows the $four$ sets denoted by $Q1$, $Q2$, $Q3$, $Q4$ for each quadrant and testing

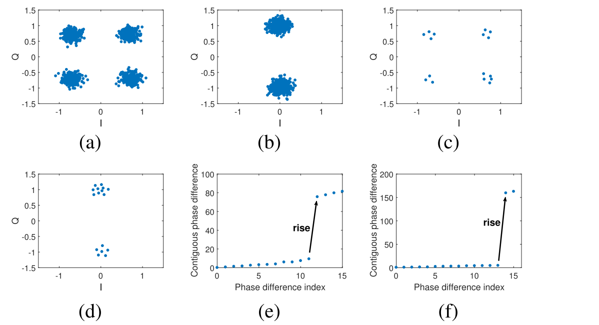


Fig. 3. Constellation at 15dB of (a) QPSK and (b) BPSK. Blind 16 medoids estimated (c) from 3(a) and (d) from 3(b). Contiguous phase differences for (e) medoids estimated in 3(c) and (f) medoids estimated in 3(d).

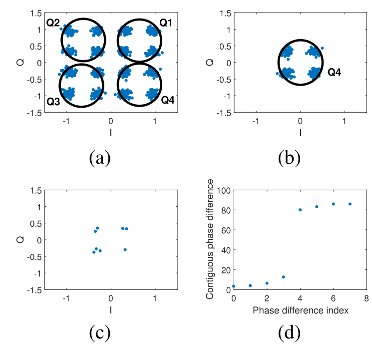


Fig. 4. (a) 16-QAM constellation at 20dB. (b) Fourth quadrant points with zero mean. (c) 8 estimated medoids from 4(b). (d) Contiguous phase difference of 4(c).

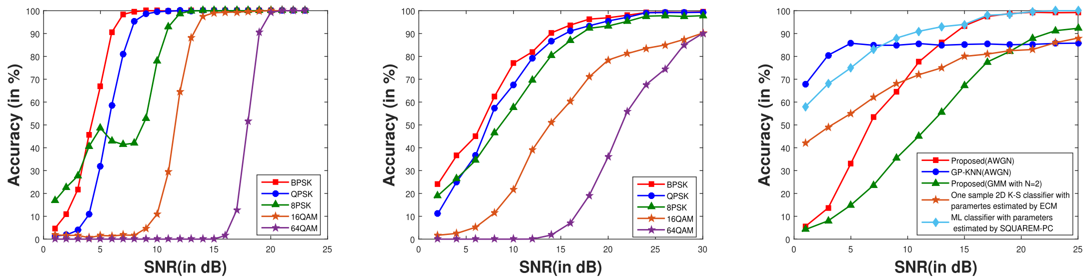
of $Q4$ for QPSK. If 3 or more quadrant points are identified as QPSK, the classification result is given in favor of 16-QAM.

V. Results and Discussion

In this section, the probability of correct classification P_{cc} of the proposed AMC algorithm is assessed through Monte Carlo simulations in LabVIEW. Results obtained are compared with different classifiers used in literature. Modulation candidate set comprises of BPSK, QPSK, 8-PSK, 16-QAM, and 64-QAM. In each trial, P_{cc} is calculated for 1000 recovered random symbols from modulation over 1000 signal realizations for every SNR. All signals are subjected to AWGN and Rayleigh fading with slow phase offset varying uniformly between 0 and 2π.

Fig. 5(a) shows the accuracy for different modulation schemes with SNR under AWGN channel. BPSK, QPSK, and 8-PSK have been identified accurately above 8dB, 9dB, and 12dB SNR respectively while 16-QAM is identified reliably (>90%) above 14dB SNR. 8-PSK has a dip in accuracy due to estimated medoids with the same phase and different amplitude at lower SNR. At higher SNR, estimated medoids are on a circle with values of contiguous phase difference less than 30^0 due to which accuracy decreases slightly. 64-QAM is classified with an accuracy of 90.7% at 21dB in AWGN channel, and in the fading environment, 89.1% accuracy is achieved at 30dB. Fig. 5(b) shows the result of classification accuracy for the fading environment.

Fig. 5(c) compares the proposed algorithm with GP-KNN in AWGN channel over four classes, i.e., BPSK, QPSK, 8-PSK,



(a) (b) (c)

Fig. 5. Accuracy of BPSK, QPSK, 8-PSK, 16-QAM and 64-QAM in (a) AWGN channel and (b) Fading channel. (c) Comparison of proposed algorithm with GP-KNN, K-S, and ML classifiers in Gaussian Mixture Model.

TABLE I

COMPARISON BETWEEN DIFFERENT CLASSIFIERS GIVEN IN [11] AND PROPOSED WORK AT 20dB SNR

	Naive	ANN	Logistic reg.	GBRT	Proposed
BPSK	99.6	20	98.2	96.4	96.9
QPSK	61.4	56.4	54.8	33.4	95.5
8-PSK	16.2	22.6	34.4	57.2	93.3
16-QAM	60.8	43.4	61.6	56.4	78.4

and 16-QAM for 512 symbols. For SNR higher than 12dB, the proposed method achieves greater accuracy compared to GP-KNN. Further, the proposed method is also compared with the algorithm given in [10] for the same set of GMM model. It also works better at higher SNR. In [11], spectral and statistical features are used along with different classifiers like ANN, SVM, k-NN, logistic regression, and gradient boosted regression tree (GBRT). Table I compares the results given in [11] with the proposed classifier for BPSK, QPSK, 8-PSK, and 16-QAM in Rayleigh fading channel without diversity. From the table, it can be observed that the accuracy of the proposed method is higher for all modulations and classifiers except for BPSK with Naive and Logistic regression classifier.

In [5], 16-QAM is classified using fourth and sixth order cumulants with an accuracy of 98.7% at 15 dB with GP-Tree classifier while our method achieves 99.28% without any training. In [12], the polynomial classifier with higher order cumulants is used which needs to optimize the weights at each hierarchy of the binary classification. For AWGN channel with unknown SNR, Conventional and Hierarchical Polynomial classifiers achieve the accuracy of 84% and 91% at 20dB SNR respectively while proposed method provides a classification accuracy of 99.82% which shows that the proposed method performs better than these classifiers.

VI. FURTHER DISCUSSION

The main idea of the paper is to identify the modulation from the received RF signal using extracted constellation and k-medoids clustering. All the modulations are classified by a simple feature based on the consecutive phase of estimated medoids. 64-QAM is classified by treating points in each quadrant as 16-QAM, and 16-QAM modulation is classified by processing points in each quadrant as QPSK. Hence classifier used for QPSK can be directly used for 64-QAM and 16-QAM classification. The complexity of the algorithm with N data points and k clusters is $O(k(N-k)^2)$ per iteration. This complexity does not change for higher order modulations as in likelihood-based methods and is less as compared to subtractive clustering methods [4].

VII. CONCLUSION

In this paper, a new method to classify five modulation schemes has been investigated. It is tested under AWGN and slow and flat fading channel. A new approach to estimate the symbol rate and phase offset has also been proposed. Phase offset is corrected for further mapping of symbols to bits. k-medoids clustering is used which is less sensitive to outliers; hence BPSK, QPSK, 8-PSK, and 16-QAM are recognized reliably above 6, 7.5, 11, and 14 dB SNR respectively. The proposed method does not require training to set thresholds. PSK recognition accuracy is not sensitive to phase offset. The method considers the AMC system as a black box which takes detected RF signal and processes it to give modulation scheme.

REFERENCES

[1] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Survey of automatic modulation classification techniques: Classical approaches and new trends," *IET Commun.*, vol. 1, no. 2, pp. 137–156, Apr. 2007.

[2] J. Zhang, D. Cabric, F. Wang, and Z. Zhong, "Cooperative modulation classification for multipath fading channels via expectation-maximization," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6698–6711, Oct. 2017.

[3] M. Abu-Romoh, A. Aboutaleb, and Z. Rezki, "Automatic modulation classification using moments and likelihood maximization," *IEEE Commun. Lett.*, vol. 22, no. 5, pp. 938–941, May 2018.

[4] X. Yan, G. Zhang, and H.-C. Wu, "A novel automatic modulation classifier using graph-based constellation analysis for-ary QAM," *IEEE Commun. Lett.*, vol. 23, no. 2, pp. 298–301, Feb. 2019.

[5] M. W. Aslam, Z. Zhu, and A. K. Nandi, "Automatic modulation classification using combination of genetic programming and KNN," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2742–2750, Aug. 2012.

[6] F. Wang and X. Wang, "Fast and robust modulation classification via Kolmogorov-Smirnov test," *IEEE Trans. Commun.*, vol. 58, no. 8, pp. 2324–2332, Aug. 2010.

[7] F. Wang, O. A. Dobre, C. Chan, and J. Zhang, "Fold-based Kolmogorov—Smirnov modulation classifier," *IEEE Signal Process. Lett.*, vol. 23, no. 7, pp. 1003–1007, Jul. 2016.

[8] G. Jajoo, Y. Kumar, S. K. Yadav, B. Adhikari, and A. Kumar, "Blind signal modulation recognition through clustering analysis of constellation signature," *Expert Syst. Appl.*, vol. 90, pp. 13–22, Dec. 2017.

[9] K. V. Cartwright, "Blind phase recovery in general QAM communication systems using alternative higher order statistics," *IEEE Signal Process. Lett.*, vol. 6, no. 12, pp. 327–329, Dec. 1999.

[10] W. Chen, Z. Xie, L. Ma, J. Liu, and X. Liang, "A faster maximum-likelihood modulation classification in flat fading non-Gaussian channels," *IEEE Commun. Lett.*, vol. 23, no. 3, pp. 454–457, Mar. 2019.

[11] M. A. Hazar, N. Odabasioglu, T. Ensari, Y. Kavurucu, and O. F. Sayan, "Performance analysis and improvement of machine learning algorithms for automatic modulation recognition over Rayleigh fading channels," *Neural Comput. Appl.*, vol. 29, no. 9, pp. 351–360, May 2018.

[12] A. Abdelmutalab, K. Assaleh, and M. El-Tarhuni, "Automatic modulation classification based on high order cumulants and hierarchical polynomial classifiers," *Phys. Commun.*, vol. 21, pp. 10–18, Dec. 2016.