Signal Detection in M-PSK Modulation Under Channel's Additive Noise and False Data Injection

Mina Faghih, Zahra Sadat Jamadi, Milad Heshmati and Mehran Attar
COMP 6321: Machine Learning Project Report
Concordia University

***Abstract*—In this project, we address the improvement of the classification accuracy for an M-PSK modulation system in the presence of Additive White Gaussian (AWG) noise and false data injection in the telecommunication channel. M-PSK modulation system refers to a telecommunication system that transmits a sequence of bits from a transmitter to a receiver through a wireless channel. The main goal in these systems is to classify the transmitted bits into M groups (symbols). The transmitted signals (sequences of bits) are affected by the AWG noise in the channel, which can deteriorate the performance of the classification. Moreover, transmitted signals may be exposed to the injection of false samples to the normal samples by cyber attackers. These attackers try to deceive the bit's classifier and affect its performance. In this work, we address two problems. First, we consider AWG noise in the communication channel, which affects the transmitted signal and degrades the classifier's performance. We propose an encoding scheme based on the concept of principal component analysis (PCA) to reduce the effect of noise, which improves classification. Second, we suppose a Cyber attacker in the communication channel injects false samples into the normal transmitted samples. We propose an anomaly detection algorithm using multivariate Gaussian distribution and F-score to detect and filter false samples.**

I. Introduction

M-PSK systems refer to telecommunication systems that include a transmitter and a receiver. The transmitter block transfer sequence of bits to the receiver block using a channel. On the receiver side, the bits predictor block classifies the received bits into M groups. Figure 1 shows the M-PSK system block diagram. In the real-world, M-PSK systems are facing two challenges for the classifications of bits. The first challenge is the AWG noise in the communication channel, and the second is the injection of false samples by a Cyber-attacker. AWG noise and false data injection can significantly affect the performance of the classification. We have addressed two main problems. First, we propose an encoding scheme that is designed using the concept of the PCA method to reduce the effect of noise. This proposed encoding scheme can highly improve the classification of received bits. Second, an anomaly detection algorithm is designed based on a multivariate Gaussian distribution and F-score to detect anomalous samples. The anomaly detector tries to prevent the attackers from deceiving the bit predictor.

To prove the effect of the proposed encoding scheme (i.e., noise reduction using PCA) on the improvement of the classification problem, we have applied the proposed method on Support Vector Machine (SVM) classifier, logistic regression classifier, and multi-layer perceptron (MLP) classifier, and results have been compared with the standard case (i.e., classification without using PCA). The report is organized as follows. In section II, the data generation method is introduced. In section III, the used methods are introduced briefly. Section IV presents and analyzes the achieved results for the two main raised problems. Moreover, this section shows the improvements in the classification problem using the proposed methods. Finally, section V concludes the report.

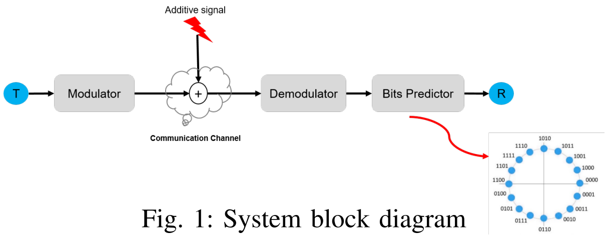

Fig. 1: System block diagram

II. Data Generation Method

In this section, we present our approach for generating a dataset that models the behavior of an M-PSK system. On the transmitter side, a large number of bits (e.g., 10^6 bits) are generated by random function to simulate the data bits. Due to the nature of real telecommunication systems, this kind of simulation is close to practical systems [1]. In the modulator, the bits are grouped with $log(M)$ bits length and name this groups as symbols i.e., 010 is a symbol consist of 3 bits. Then each symbol is mapped to a complex space with specific phase and amplitude by grey mapping. The modulated symbols are transmitted in the wireless channel where the AWG noise is added to the symbols resulting in displacement and distortion in symbols. This noise will be produced using a Gaussian density function. According to channel protocols, the symbols are transmitted by analog signals such as sin as career. This signal has various features such as Power, Signal to Noise Ratio (SNR), Deviation Magnitude and angle. Based on the protocol, the receiver knows the data can be reproduced by phase detection of the symbol signals.

The used data set is consist of 8 features. These features are as follows, two first features are Imaginary and Real parts of the symbols (or x and y on the Cartesian coordinate system), the transmitter phase, the receiver phase, symbol signal power, symbol signal SNR, deviation magnitude and deviation angle. All these features are estimated using well known calculations and distributions used in telecommunication simulations. More details and equations are not added in this report due to lack of space. Since the system uses a 8-PSK modulation schema, each symbol is named from 0,...,7 as the category labels (e.g., symbol $= 001$ should be classified as class 1) [1], [2]. To produce anomalous samples, we supposed that the attacker can change the value of features and in this way, deceive the bit's predictor.

III. Methodology

In this section, the methods for solving the addressed problems are introduced. As mentioned before, the AWG noise in the communication channel affects the transmitted signal and reduces the accuracy of the bit's classifier. We have proposed a method based on PCA to filter the noise and improve the accuracy of the classifier in the classification of transmitted bits. The proposed architecture for noise reduction using PCA has shown in Figure 2. In Figure 2, the encoding block applies PCA on the transmitted data and then finds and removes the components which contain more noise than information, and finally, reconstructs the data and transfers it to the bits predictor block. It is worth mentioning that the encoder block is trained offline and then used online. In what follows, the PCA approach for noise reduction is addressed.

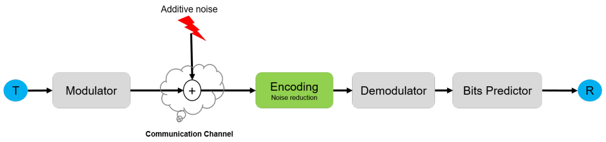

Fig. 2: Proposed architecture for noise reduction

A. Principle Component Analysis for noise reduction

PCA is used to transform datasets from a high dimensional space to a lower dimension space. Moreover, by using PCA, we can find the directions of the maximum variation in a dataset. In other words, PCA can find the directions that contain more information. In general, in a scaled dataset that contains Gaussian noise, the variation of the main information is higher than the variation of noise [3]. As a result, we can find the directions that contain more noise than the main information by using PCA. These directions are called principal components. Let $X \in \mathbb{R}^{m\times n}$ be a dataset where m is the number of samples and n is the number of features, then by solving the following optimization problem, we can find the principal components of X,

$$\begin{aligned} \max_{\phi} \quad & \phi S \phi^T \\ \text{s.t.} \quad & \phi\phi^T = 1 \end{aligned} \tag{1}$$

where $S = X^T X$, is the covariance matrix of X, and $\phi \in \mathbb{R}^{d\times n}$ is an orthonormal matrix. By using a diagonal Lagrange multiplier matrix $\Lambda \in \mathbb{R}^{d\times d}$, the Lagrangian of (1) is

$$L(\phi, \Lambda) = \phi S \phi^T - \Lambda(\phi\phi^T - 1) \tag{2}$$

By imposing $\frac{\partial L}{\partial \phi} = 0$, we have

$$S\phi = \Lambda\phi \tag{3}$$

where Λ contains the eigenvalues λ_j of S, and each column of ϕ i.e., ϕ_{*j}, is an eigenvector of S associated to λ_j. If we re-order λ_j in Λ and ϕ_{*j} in S such that $\lambda_1 \geq \ldots \geq \lambda_p$, then ϕ_{*1} is known as the direction of maximum variation. Similarly, $\phi_{*2}, \ldots, \phi_{*p}$ are, in a descending order, the next directions of the maximum variation. In the literature, ϕ_{*j} are known as the Principal Components (PC) of X [4]. The value of λ_j, represent the amount of information in the direction of the associated eigenvector, ϕ_{*j}. In other words, directions that contain lower information have a smaller eigenvalue. Consequently, if a dataset contains noise, this noise will appear in the components (directions) with smaller eigenvalues.

The proposed encoding scheme contains an encoder and decoder which each of them is linear transformations. These transformation can be found using the columns of the ϕ matrix in (1). By solving the optimization problem in (1), we can find ϕ (components or eigenvectors) and λ_j (associated eigenvalues). Then we choose the eigenvectors which have the largest eigenvalues. The chosen eigenvectors will provide the desired linear transformations for noise reduction. Let $\mathcal{T}$ be the linear transformation of the encoder, then $\mathcal{T}^T$ will be the linear transformation of the decoder and can be found as follows:

$$\mathcal{T} = \begin{bmatrix} \phi_{*1} & \phi_{*2} & \ldots & \phi_{*d} \end{bmatrix} \tag{4}$$

Now, by using the designed encoding scheme, a new sample is going through the encoder and decoder, and in this way we can reduce the effect of the noise from samples. We have used the PCA function from the sklearn.decomposition package to find principal components. This function has two attributes named *components* and *explained variance* which gives the eigenvectors (directions) and eigenvalues (amount of information in the corresponding direction), respectively.

B. Training Optimum Classifiers and Setups

To evaluate the effect of noise reduction using PCA, we checked the results on SVM, MLP, and Logistic regression classifiers before and after applying PCA with identical setups. Moreover, to design optimal classifiers, i.e., classifiers with proper hyper-parameters values, stratified K-fold cross-validation, and grid search technique has been used during the training process. The following setups have been considered for each of the classifiers:

1) SVM: we used 5-folds with shuffling for the cross validation, and SVM has been designed based on RBF kernel. Moreover, the following ranges has been considered for the hyper-parameters search:

$$\alpha = \left[10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3, 10^4, 10^5\right]$$

$$\mathcal{C} = \left[10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3, 10^4, 10^5\right]$$

and other parameters of the SVM have been set to their default values.

2) MLP: we used 5-folds with shuffling for the cross validation, and three hidden layers have been considered in the architecture of the MLP with the following value (number of neurons) ranges in the grid search:

$$L = \left[(\underbrace{20}_{L_1}, \underbrace{15}_{L_2}, \underbrace{10}_{L_3}), (\underbrace{20}_{L_1}, \underbrace{15}_{L_2}, \underbrace{5}_{L_3}), (\underbrace{20}_{L_1}, \underbrace{10}_{L_2}, \underbrace{5}_{L_3})\right]$$

Moreover, we have used adaptive learning rate with 150 iterations, and the range of α is considered as $\alpha = \left[10^{-3} \quad 10^{-2} \quad 10^{-1} \quad 1 \quad 10 \quad 10^{2} \quad 10^{3} \quad 10^{4} \quad 10^{5}\right]$, in the grid search. Other parameters of the MLP have been set to their default values.

3) Logistic Regression: we used 5-folds with shuffling for the cross validation, and the range of $\mathcal{C}$ is considered as $\mathcal{C} = \left[1, 2, 5, 10\right]$, and other parameters of the Logistic Regression have been set to their default values.

C. Designing Anomaly Detector Using Multivariate Gaussian

The proposed architecture for anomaly detection has shown in Figure 3. The anomaly detector model is designed offline based on a multivariate Gaussian distribution using a normal dataset to give the probability of being normal for a sample x. The multivariate Gaussian distribution is as following:

$$P(x;\mu,\Sigma) = \frac{1}{(|\Sigma|^{\frac{1}{2}})(2\pi)^{\frac{q}{2}}} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)} \tag{5}$$

where $\mu \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n\times n}$ are the mean and covariance of the distribution, respectively. To find μ and Σ, the normal samples are used, and by calculating the mean and covariance of the normal dataset, we can achieve μ and Σ. By finding the parameters of the Gaussian distribution, then we can use it online to calculate the probability of being normal for new samples, x_i. Let $P(x_i)$ be the probability of x_i to be normal; then, by using algorithm 1 we can detect anomalous samples. In algorithm 1, ϵ is a threshold found using the F-score.

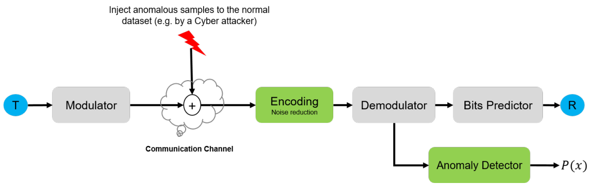

Fig. 3: Proposed architecture for anomaly detection

Algorithm 1: Anomaly Detector

```
Input: $x_i$, $\epsilon$ Output: $\{0 = normal, 1 = anomaly\}$
  Compute the probability of $x_i$'s new samples to be normal:
  $P(x_i;\mu,\Sigma) = \frac{1}{(|\Sigma|^{\frac{1}{2}})(2\pi)^{\frac{q}{2}}} e^{-\frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)}$
if $P(x_i;\mu,\Sigma) \leq \epsilon$ then
    $x_i$ is anomaly;
else
    $x_i$ is normal;
end
```

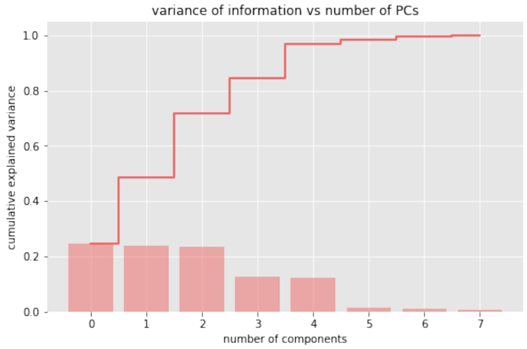

Fig. 4: cumulative explained variance graph

IV. Results

In this section, the achieved results are presented. First, we analyze the effect of noise reduction by using the proposed encoding scheme on the accuracy of the classifiers, and then the results of the anomaly detection problem using multivariate Gaussian distribution are presented.

A. Evaluation of the proposed encoding scheme on the accuracy of the classification

Based on (4), first, we apply PCA to the dataset to find the components that contain more noise than the useful information and in this way, we can find the desired transformation $\mathcal{T}$, (encoder and decoder). Moreover, to analyze the principal components, we used the cumulative explained variance graph, which has been shown in Figure 4. In this figure, the explained variance is equal to the magnitude of the eigenvalues for each of the components. Therefore, components that contain more information have larger eigenvalues, and the cumulative variance is the summation of each component's variance. Figure 4, shows that the last 3 principal components do not contain considerable information compared to other components and the first 5 components contain about 96% of the total information. In other words, the last 3 components are somehow modelling the noise. Consequently, we should expect that by removing the last 3 components and reconstructing the X (dataset) based on the first 5 components, the Gaussian noise is filtered from the X, and the classification accuracy will be increased. Table 1 shows the accuracy of the trained classifiers for different conditions on X. Based on Table 1, the accuracy of all classifiers after applying PCA has increased. Moreover, for the SVM classifier, 2% improvement has been achieved on the test set. Also, with this regard that the last components contain more noise than the information, by adding the last two components (components number 6 and 7), the accuracy of the SVM classifier and logistic regression classifier has slightly decreased. This affair shows that noise has been added to the dataset by adding these two components and caused this reduction.

Remark 1. It is worth noting that we have provided various visualizations of the dataset before and after applying PCA in the Jupyter Notebook file of the project. Moreover, the confusion matrices for all classifiers before and after applying PCA have been shown in that file.

TABLE I: Accuracy of the classifiers on the test set

Condition on X	SVM	MLP	Logistic Regression
without PCA	96.75%	98.25%	98.25%
PCA with first 5 components	98.75%	98.75%	98.75%
PCA with first 6 components	98.50%	98.75%	98.50%
PCA with first 7 components	98.50%	99.00%	98.50%

B. Evaluation of the proposed anomaly detection algorithm

In this section, the results of the proposed anomaly detector are presented. As mentioned before, we should find the parameters of the Gaussian distribution using the normal dataset. Moreover, based on Algorithm 1, we should also find the threshold ϵ. To find the ϵ, first, we use different values based on a try and error method. Figure 5, shows the performance of the anomaly detector using different values of ϵ. For $\epsilon = 10^{-4}$, the anomaly detector has detected all of the anomalous samples, however, many normal samples have been considered as anomalous samples which is not a reasonable prediction. However, if we set $\epsilon = 10^{-7}$, we can see that the anomaly detector has detected all of the normal samples and anomalous samples correctly. It is necessary to mention that the proper value for the ϵ should be found in the presence of normal and anomalous samples. Moreover, a perfect anomaly detector should detect all of the anomalous samples and also should not detect an anomalous sample as a normal sample. In other words, an anomaly detector should have proper values for the *recall* and *precision* rates simultaneously. For this purpose, F-score has been used as an evaluator as follows:

$$F = 2\frac{PR}{P+R} = \frac{TP}{TP + 0.5(FP + FN)} \tag{6}$$

To find the ϵ systematically, we have proposed an algorithm named *select epsilon* which can be found in the codes of the project. This algorithm can find a proper ϵ based on F-score. By using this algorithm, we achieved F-score = 0.91 for $\epsilon = 1.19 \times 10^{-8}$. The results of the anomaly detector has shown in 6. Based on Figure 6, the designed anomaly detector has detected all of the five anomalous samples. Moreover, two normal samples have been detected as anomalous samples, which is an acceptable result. In *select epsilon* algorithm, the accuracy of the detector is dependent to the choice of the step size, and by changing the step size, we may achieve better results.

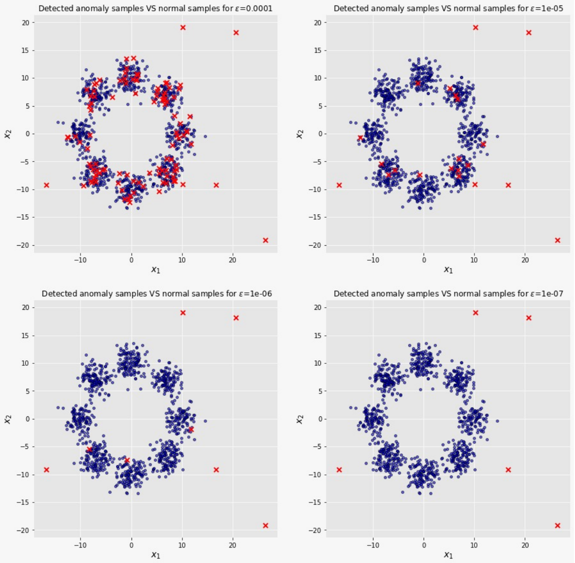

Fig. 5: Detecting anomalous samples-ϵ achieved by try and error

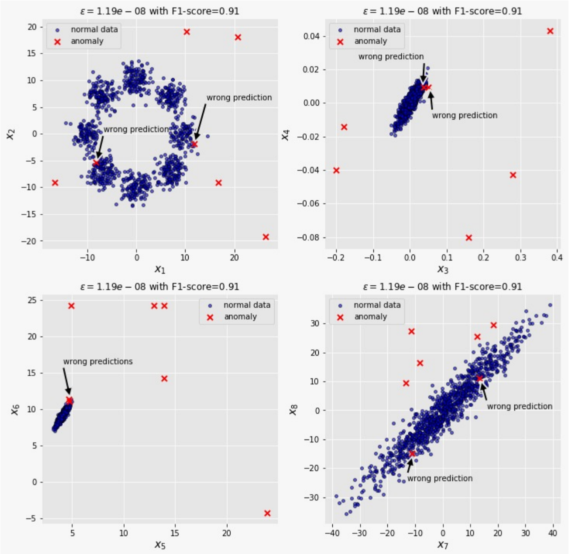

Fig. 6: Detecting anomalous samples-ϵ achieved using F-score

V. Conclusion

In this project, the improvement of the classification accuracy for an M-PSK modulation system in the presence of Gaussian noise in the communication channel has been addressed. An encoding scheme based on the concept of PCA has been proposed to reduce the effect of noise in the dataset. It is shown that, the principal components that contain more noise than the main information can be found and removed from the dataset, which results in an increase in the accuracy of the prediction. Moreover, an anomaly detection algorithm based on multivariate Gaussian distribution and F-score has been proposed to detect anomalous samples injected by a cyber attacker in the communication channel.

References

[1] C.-Y. Huan and A. Polydoros, “Likelihood methods for mpsk modulation classification,” *IEEE Transactions on Communications*, vol. 43, no. 2/3/4, pp. 1493–1504, 1995.

[2] S. Huang, C. Lin, K. Zhou, Y. Yao, H. Lu, and F. Zhu, “Identifying physical-layer attacks for iot security: An automatic modulation classification approach using multi-module fusion neural network,” *Physical Communication*, vol. 43, p. 101180, 2020.

[3] C. Bishop, “Bishop-pattern recognition and machine learning-springer 2006,” *Antimicrob. Agents Chemother*, pp. 03728–14, 2014.

[4] S. Wold, K. Esbensen, and P. Geladi, “Principal component analysis,” *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.