**SP**
**RE**

# A Linear Discriminant Analysis based Modulation Recognition Method for Linear Modulations

**Mohamad Reza Ahvaraki[1], Seyed Ali Ghorashi[1], Behrad Mahboobi[2*]**

[1] Cyberspace Research Institute, Shahid Beheshti University, Tehran, Iran.
[2] Mechanic, Electric and Computer Department, Science and Research Branch, Islamic Azad University, Tehran, Iran.

**Abstract**

In this paper, a new method for signal modulation classification based on Linear Discriminant Analysis for multi-class cases is introduced and its results are compared with two other methods of automatic modulation classification. Higher-order complex cumulants are used as feature vectors in the first place. Then, linear Fisher discriminant analysis maps these vectors to another space to separate different classes efficiently. One versus all support vector machine classifier and kernel Fisher discriminant analysis methods with radial basis function kernel is used. The achieved results in comparison with two research papers show that the proposed method classifies the classes in a shorter time with equal or better accuracy.

## 1. INTRODUCTION

Automatic Modulation Classification (AMC) plays an important role in Communication Intelligence (COMINT) [1] [2] [3] and is an intermediate step between detection and de-modulation of a signal. It has applications such as signal confirmation and spectrum monitoring. AMC also is a key technology in cognitive radio and software radio [4]. AMC is a difficult task due to lack of data about the transmission parameters such as signal power, phase offset, timing information, and carrier frequency at the receiver. Also, in real world, the difficulty increases because of presence of factors such as multipath fading, time-varying, and frequency-selective channels. Various methods have been proposed to

determine unknown parameters of a detected signal [5] [6] [7] [8] [9].

In general, modulation classification can be divided into two categories of inter-class and intra-class [10]. Inter-class identification refers to distinguishing between ASK, FSK, PSK, etc., whereas intra-class identification refers to distinguishing within a single class, such as between BPSK and QPSK.

AMC is performed in two general steps: preprocessing and classification. Preprocessing task includes operations for noise reduction, estimation of the carrier frequency, symbol period, signal power and equalization [2]. Depending on the level of sensitivity of the classification algorithm to unknown parameters, preprocessing task with different levels may be applied [2].

There are two main approaches to classify the modulation type of a received signal: maximum likelihood-based approach and feature-based (FB) approach [2]. The former offers an optimal solution in a Bayesian sense. It is based on the likelihood function of the received signal and decision is made by comparing the likelihood ratio against a threshold. The main drawback of likelihood-based approach in general is computational cost [2] [1]. In FB approach, several features of the received signal are extracted and then a decision is made according to the feature values. A disadvantage of FB approach is its non-optimality, however, it is easy to implement, and in a proper implementation, it can give a near optimal performance [11].

It was shown that by using high order cumulants, a wide variety of modulated signals can be classified [12] [13] [14]. Relevant AMC works based on FB approach are briefly reviewed in following.

Many types of features like instantaneous amplitude, phase and frequency, Fourier and wavelet transform, high order moments and cumulants and cyclostationary have been already used for classification [15] [16] [17] [18] [19]. Also, pattern recognition methods including artificial neural networks [20] [21], clustering, decision tree [22] and support vector machines (SVMs) [18] are used to make decisions in related cases.

As machine learning field is growing in recent years, using kernel functions makes linear classifiers cope with non-linear problems, which was called the $3^{rd}$ revolution of pattern analysis methods [23]. In kernel method, the following steps should be followed [24]: first, adjust the algorithm to the form of the inner product of the input vector and then combine the algorithm and kernel function. Therefore, the kernel method falls in two parts: the design of kernel function and design of the algorithm. In [4] some fourth-order complex cumulants and their ratios were chosen in order to build a feature matrix in the first place. Then, the kernel function is used to map the feature matrix impliedly to a high dimensional feature space and finally, linear Fisher discriminant analysis (FDA) is applied to do the classification. Radial basis kernel is chosen as the kernel function. Cross-validating grid search method is used to select the kernel function's parameters. Comparing with SVM, the kernel FDA (KFDA) method is appropriate and much faster because there is no need to solve an optimization problem and train the classifier. There is a good accuracy in low signal to noise ratios (SNRs) because of using higher order cumulants, which are robust against noise.

In [25] cyclostationary features was used for classification. Barathram et al. [25]used peak values of second spectral correlation called cyclic domain profile as features and

compared neural networks (NNs) and hidden Markov model (HMM) for classification of BPSK, QPSK, MSK, and FSK modulations. Cyclostationary feature extraction, in this case, makes classifier robust against stationary noise and separates signals with overlapping power spectral densities. The second order cyclic spectrum is not capable of classifying higher order QAM or higher order PSK modulations.

In this paper, the multi-class FDA method is used in order to classify the MPSK and MQAM modulated signals. Also, a comparison between some AMC methods is done. The FB approach is employed, using higher order cumulants as features and three different classifier results are compared. Once the modulation format is recognized, the next step would be demodulation and data extraction.

The rest of this paper is organized as follows: Background knowledge, including problem statement and cumulants, are given in Section 2. SVM classifier and FDA are explained in Section 2.3. Finally, simulation results and the paper conclusion are given in Sections 3 and 4, respectively.

## 2. BACKGROUND

### 2.1. Problem Statement

In this paper, the problem of the classification of linear modulated signals under existing additive white Gaussian noise (AWGN) is considered. The environment is considered interrelated and synchronized and received signal has carried out carrier frequency and timing synchronization.

Consider the complex baseband model as,

$$x(t) = v(t) + w(t) \tag{1}$$

where w(t) is the complex AWGN and v(t) is a complex baseband linear modulated signal:

$$v(t) = \sum_k c_k h(t - kT - t_0) \tag{2}$$

and T is the symbol period, $t_0$ the timing offset, h(t) is the shaping filter and $\{c_k \in \mathbb{C}\}$ is an iid (independent and identically distributed) symbol sequence. The following six types of linear modulations are considered: MPSK, M = 2,4,8, and M-QAM, M = 16,64,256.

The problem is to decide which one of these six modulations, the noisy received signal belongs to. The performance is computed as the ratio of correct classification versusall classifications.

### 2.2. Feature Extraction

### 2.2.1. Cumulants

A real signal moment generating function is defined as:

$$M(t) = E\left[ e^{tX(t)} \right] \tag{3}$$

and for a complex modulated signal, single moments are calculated as:

$$M_{pq} = E\left[ X(k)^{p-q} X(k)^q \right] \tag{4}$$

Cumulant generating function also is defined as [26]:

$$K(t) = \log\left( E\left[ e^{tX(t)} \right] \right) \tag{5}$$

where the coefficients of power series expansion of the cumulant generating function are cumulants. Cumulants have additivity property. White noise's 4th and higher order cumulates are zero and because of additivity property of cumulants, it would be a well-defined feature for classifying signals in AWGN channels. The joint cumulant of sev-

eral random variables is defined by a similar cumulant generating function:

$$K(t_1, t_2, ..., t_n) = \log\left(E\left[e^{\sum_{j=1}^{n} t_j X_j}\right]\right) \quad (6)$$

A consequence is that:

$$\kappa(x_1, ..., x_n) = \sum_{\pi}(|\pi|-1)!(-1)^{|\pi|-1}\prod_{B\in\pi}E\left(\prod_{i\in B}X_i\right) \quad (7)$$

where $\pi$ runs through the list of all partitions of $\{1,...,n\}$ B runs through all the list of blocks of partition $\pi$ and $|\pi|$ is the number of partitions. Under the assumption of independent, identically distributed data symbols, normalized complex cumulants of order $n$ can be computed by Eq. (7). The fourth and sixth-order cumulants are chosen because of their robustness against additive Gaussian noise and because they are invariant with respect to shift, scale and rotation of MPSK signal constellation [27]. The norm of ideal higher-order cumulants of a signal is expressed in Table. 1 [28].

### 2.2.2 Feature Selection

After forming the feature matrix, the value of each feature should be specified. Applying principal component analysis (PCA) on the feature matrix, the most important and efficient features that make classes more separable are selected. The final feature matrix is:

$$F_x = [f_{x1}, f_{x2}, f_{x3}] = \left[|C_{40}|, |C_{61}|, \left|\frac{C_{40}}{C_{42}}\right|\right] \quad (8)$$

By using the result in Table.1, the ideal values of $F_x$ are:

$$F_x = \begin{cases} [2,16,1] & BPSK \\ [1,4,1] & QPSK \\ [0,0,0] & 8-PSK \\ [0.68, 2.08, 1] & 16-QAM \\ [0.619, 1.797, 1] & 64-QAM \\ [0.604, 1.734, 1] & 256-QAM \end{cases}$$

**Table 1.** *Theoretical Cumulants for signal constellations under the constraint of unit variance.*

| Mod. $C_{m,n}$ | BPSK | QPSK | 8-PSK | 16-QAM | 64-QAM | 256-QAM |
|---|---|---|---|---|---|---|
| $C_{4,0}$ | 2 | 1 | 0 | 0.68 | 0.619 | 0.604 |
| $C_{4,1}$ | 2 | 0 | 0 | 0 | 0 | 0 |
| $C_{4,2}$ | 2 | 1 | 1 | 0.68 | 0.619 | 0.604 |
| $C_{6,0}$ | 16 | 0 | 0 | 0 | 0 | 0 |
| $C_{6,1}$ | 16 | 4 | 0 | 2.08 | 1.797 | 1.734 |
| $C_{6,2}$ | 16 | 0 | 0 | 0 | 0 | 0 |
| $C_{6,3}$ | 16 | 4 | 4 | 2.08 | 1.797 | 1.734 |
| $C_{4,0}/C_{4,2}$ | 1 | 1 | 0 | 1 | 1 | 1 |
| $C_{4,1}/C_{4,2}$ | 1 | 0 | 0 | 0 | 0 | 0 |

## 2.3. Classifiers Based On Kernel Function

### 2.3.1. Support Vector Machine(SVM)

SVM is a two-class classifier and it is based on the idea of "large margin" and "mapping the data to a higher dimensional space" and the kernel functions. The cost function of this optimization problem is defined as $\|w\|^2$, so it should be minimized to find an optimal discriminant hyperplane that maximizes the margin between the two classes and satisfy,

$$\begin{cases} J(w, w_0) = \dfrac{1}{2}\|w\|^2 \\ y_i(w^T x_i + w_0) \geq 1 \qquad i = 1, 2, ..., N \end{cases} \qquad (9)$$

KKT (Karush–Kuhn–Tucker) conditions should also be satisfied to find an optimal solution for the hyperplane:

$$\begin{cases} \dfrac{\partial}{\partial w}\zeta(w, w_0, \lambda) = 0 \\ \dfrac{\partial}{\partial w_0}\zeta(w, w_0, \lambda) = 0 \\ \lambda_i \geq 0, \ i = 1, 2, ..., N \\ \lambda_i\left[ y_i\left(w^T x_i + w_0\right) - 1\right], i = 1, 2, ..., N \end{cases} \qquad (10)$$

In cases that classes are not linearly separable, SVM maps the data into a higher dimensional feature space with a non-linear mapping and find an optimal hyperplane in that space as [29] :

$$\min \varphi(w, \xi) = \frac{1}{2}(w.w) + C\sum_{i=1}^{n}\xi_i \quad i = 1, 2, ..., n$$

$$\begin{cases} y_i\left[w.x_i + b\right] \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases} \qquad (11)$$

According to the Eq. (11), and Lagrange theorem, the quadratic problem can be represented by kernel function:

$$\max Q(\alpha) = -\frac{1}{2}\sum_{j=1}^{m}\sum_{i=1}^{n}\alpha_i\alpha_j y_i y_j k(x_i, x_j) + \sum_{i=1}^{n}\alpha_i$$

$$s.t.\begin{cases} \displaystyle\sum_{i=1}^{n}\alpha_i y_i = 0 \qquad i = 1, 2, ..., n \\ 0 \leq \alpha_i \leq C \end{cases} \qquad (12)$$

where $k(x_i, x_j) = \varphi(x_i)\varphi(x_j)$ is the kernel function that fulfills the Mercer theorem, so the discriminant function can be expressed as:

$$\begin{aligned} f(x) &= \text{sgn}\left\{(w.x) + b\right\} \\ &= \text{sgn}\left\{\sum_{x_i \in SV}\alpha_i y_i k(x_i, x_j) + b\right\} \end{aligned} \qquad (13)$$

where

$$w = \sum_{x_i \in SV}\alpha_i y_i \varphi(x_i) \qquad (14)$$

According to the optimal problem (11), the complexity of SVM is not related to the dimension of features but it is restricted by the number of samples. The reason is that SVM needs to compute the kernel function between every two samples of training data, to generate a kernel matrix of $n \times n$ elements, where $n$ is the number of training samples. So for big value of $n$, sorting the kernel matrix in SVM needs big memory. As a conclusion, matrix operations makes SVM too costly to implement.

### 2.3.2. Fisher Discriminant Analysis for Multi Class Cases (Proposed Method)

In this method, the main goal is finding a vector and multiply it to the feature matrix and map it to a new feature space with more separability [30]. In multi-class cases, the mapping vector is replaced with a transformation matrix. The major task can be summarized as follows: If x is an m-dimensional

vector of samples, transform it into another *l*-dimensional vector y, so that an adopted class separability criterion is optimized. The first step is to find between class scatter matrix $S_w$ and within-class scatter matrix $S_b$ that indicates the separability between classes and samples of each class,

$$S_w = \sum_{i=1}^{M} P_i \Sigma_i \tag{15}$$

$$S_b = \sum_{i=1}^{M} P_i (\mu_i - \mu_0)(\mu_i - \mu_0)^T \tag{16}$$

where

$$\mu_0 = \sum_{i}^{M} P_i \mu_i \tag{17}$$

The mixture scatter matrix would be the sum of Eqs. (15) and (16):

$$S_m = E\left[ (x - \mu_0)(x - \mu_0)^T \right]$$
$$S_m = S_w + S_b \tag{18}$$

A separability measure called FDR (Fisher Discriminant Ratio) is defined as [30]:

$$J = trace\left\{ \frac{S_b}{S_w} \right\} \tag{19}$$

In the next step, *y* vector that maps the feature matrix should be found

$$y = Ax \tag{20}$$

Now, the scatter matrixes (15) and (16) are computed for y according to values for feature data:

$$\begin{cases} S_{yw} = A^T S_{xw} A \\ S_{yb} = A^T S_{xb} A \end{cases} \tag{21}$$

As the goal is maximizing the between-class scatter matrix and minimizing the within-class scatter matrix, the element A should satisfy:

$$\frac{\partial J(A)}{\partial A} = 0 \tag{22}$$

According to Eqs. (20) and (21), we have:

$$\frac{\partial J(A)}{\partial A} =$$
$$-2S_{xw} A \left( A^T S_{xw} A \right)^{-1} \left( A^T S_{xb} A \right) \left( A^T S_{xw} A \right)^{-1} + 2S_{xb} A \left( A^T S_{xw} A \right)^{-1} = 0$$

or

$$\left( S_{xw}^{-1} S_{xb} \right) A = \left( S_{yw}^{-1} S_{yb} \right) A \tag{23}$$

This is an eigenvalue problem and to solve this, the matrixes $S_{yb}$ and $S_{yw}$ are to be diagonalized simultaneously by a linear transformation:

$$\begin{cases} B^T S_{yw} B = I \\ B^T S_{yb} B = D \end{cases} \tag{24}$$

The transformed features would be like:

$$\hat{y} = B^T$$

$$y = B^T A^T x \tag{25}$$

and here is the verification of equality of (y and $\hat{y}$ )

$$J(\hat{y}) = trace\left\{ S_{\hat{y}w}^{-1} S_{\hat{y}b} \right\}$$
$$= trace\left\{ \left( B^T S_{yw} B \right)^{-1} \left( B^T S_{yb} B \right) \right\}$$
$$= trace\left\{ B^{-1} S_{yw}^{-1} S_{yb} B \right\}$$
$$= trace\left\{ S_{yw}^{-1} S_{yb} B B^{-1} \right\} = J(y)$$

The final vector to be multiplied to x is $A^T B^T$ and this matrix with size of $m \times l$ maps the features to another space with higher reparability.

After mapping, each feature would be assigned to the class that has the shortest distance between its mean and the feature sample.

### 2.3.3. Kernel Fisher Discriminant Analysis

The procedures of KFDA is given in [31]: first, map the feature data samples through a non-linear mapping to a higher dimensional feature space. Then, perform the linear Fisher discriminant analysis to realize the nonlinear discriminant analysis relative to input space. According to the Mercer theorem [30], non-linear mapping $\phi(x)$ maps input data into a high dimensional feature space, $x \rightarrow \phi(x) \in H$, where H is a Hilbert space. For each two data samples as the input, the inner product operation is represented as

$$\langle \varphi(x), \varphi(z) \rangle = K(x, z) \tag{26}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product operation in H and K(x,z) is an asymmetric condition function which satisfy,

$$\iint_{C\,C} K(x,z)g(x)g(z)dxdz \geq 0 \tag{27}$$

For any $g(x), x \in C \subset R^l$ such that

$$\int_C g(x)^2 dx \leq +\infty \tag{28}$$

where C is a finite subset of $R^1$. In this case, training data are converted from $x$ to $\varphi(x)$. Then one can perform linear FDA (Section 2.3.2) in the new feature space.

### 2.3.4 The Design of a Robust Multi Class Classifier

The design of a multi-class classifier has two essential stages: first the selection of kernel function and second, multi-class classifier decomposition.

### 2.3.4.1. Selection of the Kernel Function

Selection of the kernel function is important because the value of every parameters in kernel function affect the mapping function and change the complexity of samples distribution in feature space. Some of the kernels used in pattern recognition applications are as follows [30]:

*1) Polynomials*

$$K(x, z) = \left(x^T z + 1\right)^q, \quad q > 0 \tag{29}$$

where *p* and *d* are two constant parameters. In the case of *p*=0 and *d*=1, it is called a linear kernel function. The operation speed of this kernel is fast.

*2) Radial Basis Functions*

$$K(x, z) = \exp\left(-\frac{\|x - z\|^2}{\sigma^2}\right) \tag{30}$$

where $\sigma$ controls the width of the kernel function and needs to be obtained.

*3) Hyperbolic Tangent*

$$K(x, z) = \tanh\left(\beta x^T z + \gamma\right) \tag{31}$$

It is also called neural network kernel function. Due to the limitlessness of radial basis function feature space, the limit sample in this space must be linearly discriminable, so it is most commonly used in classification [4]. There are many algorithms such as grid searching, evolution method, and simulated annealing approach [4] to obtain the kernel function parameters

### 2.3.4.2. Multi-Class Classifier Decomposition

As SVM and linear FDA are binary classifiers, for multiclass classification, one should combine them in one of these ways [30]; 1- One against one 2- One against all 3- Binary Tree 4- Error-correcting output code

## 3. SIMULATION RESULTS

Parameter selection procedure: At first 200 every MPSK (M =2, 4, 5) and M-QAM (M = 16, 64, 256) signals every 2 dB from -20 to 40 in AWGN channel are generated. Cumulant features are extracted from signals and made up feature matrix. The samples of each class are grouped randomly into train and test.Using the multi-class FDA method, RBF kernel with $\sigma = 1$ is chosen as initialization. In the grid search, the logarithm of the initial value is computed, then the best value in $\log \sigma^2 \in \left\{ \log \sigma_0^2, \log \sigma_0^2 - 1, ..., \sigma_0^2 + 1 \right\}$ is chosen. In this case, as written in Table 2, $\sigma = 4.47$ is chosen. Similarly, as the parameter selection for KFDA, SVM kernel function parameters are initialized. Searching after initialization results in C = 10 and $\sigma = 1$ as the RBF kernel function parameters (see Table. 3).

**Table 2.** *Classification accuracy of parameter selection of KFDA classifier for 100 samples as train and test.*

| $\log \sigma$ | **-2** | **-1** | **0** | **1** | **1.3** |
|---|---|---|---|---|---|
| Classification accuracy | 0.75 | 0.71 | 0.82 | 0.84 | 0.84 |

**Table 3.** *Classification accuracy of parameter selection of SVM classifier for 100 samples as train and test.*

| $\log \sigma$ / $\log(C)$ | **-2** | **-1** | **0** | **1** | **1.3** |
|---|---|---|---|---|---|
| -2 | 0.51 | 0.67 | 0.67 | 0.62 | 0.66 |
| -1 | 0.65 | 0.67 | 0.67 | 0.83 | 0.71 |
| 0 | 0.53 | 0.5 | 0.67 | 0.83 | 0.83 |
| 1 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 |
| 1.3 | 0.34 | 0.58 | 0.5 | 0.5 | 0.5 |

Simulations: In this experiment, the classification accuracy of MPSK modulation and MQAM modulations in AWGN channel, under different SNR are obtained.

- *Test (1)*

Phase 1: Simulations are done in different SNR,signal length, and data samples for training and testing. Training data are generated under SNRs from -10 to 20 dB, and for each training, data classification is done for test data under SNRs from -20 to 40 dB. The number of samples and signal length are constant.

Phase 2: for signals in phase 1, the number of samples used in training data varies from 10 to 1000 for a specific SNR and a constant signal length. the results are recorded.

Phase 3: In this phase, the length of samples varies between 1000 and 10000 data points. This results in a specific SNR and a constant signal length is recorded.

Classifications are done with SVM classifier in one against all mode and KFDA and multi-class FDA classifier (the proposed method) at the same time to do the comparison. In this test, all modulations (MPSK and MQAM) are included.

From Fig. 1 it can be concluded that although in high SNR, the performance of all classifiers are similar, in low SNR case, multi-class FDA method is outstanding as well.

It can be seen from Fig. 2 that when classifiers trained at a low SNR (e.g. 0 dB), increasing the signal length of test signals, causes a better accuracy. Also, according to Fig. 2, when classifiers are trained at low SNR, SVM does not perform as well as multi-class FDA and KFDA.
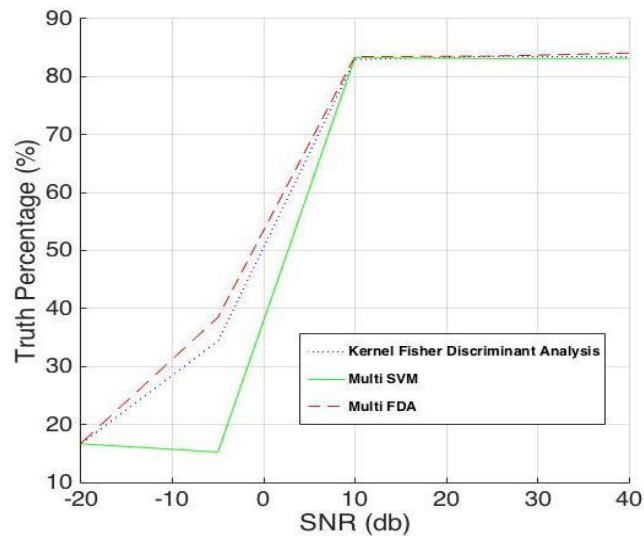
***Fig. 1. Classification accuracy using 1000 training and 1000 testing samples in different SNR for SVM and KFDA and multi-class FDA classifiers.***
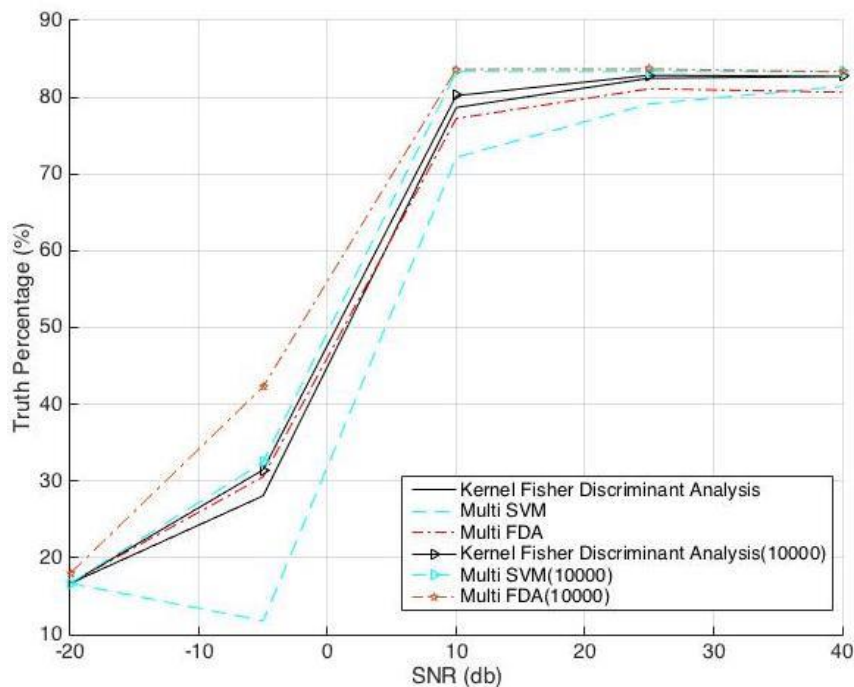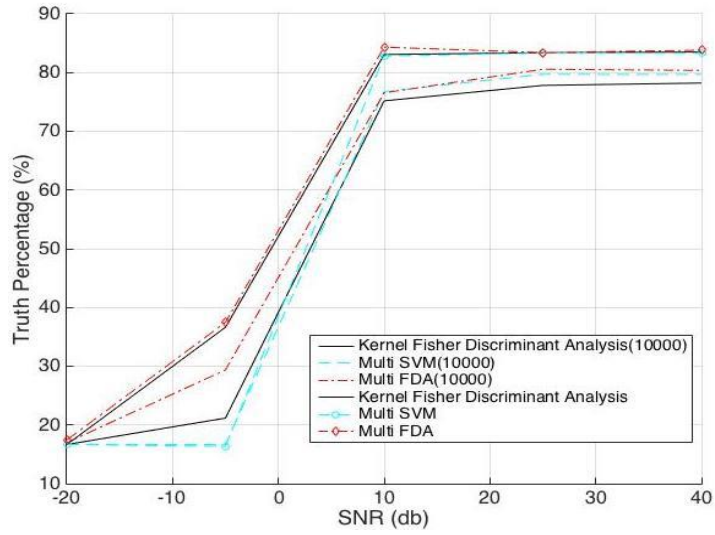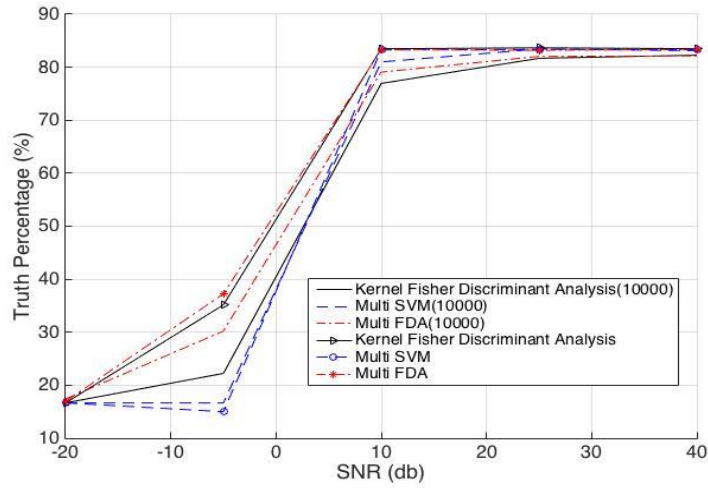


***Fig. 2. Classification accuracy by using 1000 training and 1000 testing samples in different signal lengths (1000 and 10000) for SVM and KFDA and multi-class FDA classifiers.***

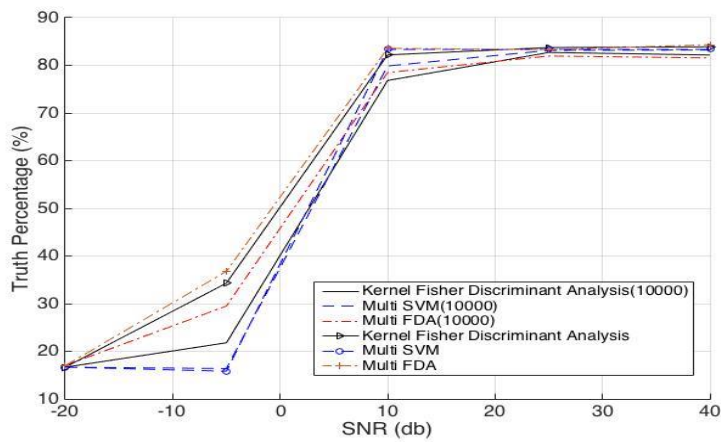Increasing the number of samples from 100 to 1000, as shown in Fig. 3 does not affect the classification accuracy, significantly. On the other hand, it takes much more time to do classification.

(a)



(b)



(c)

**Fig. 3. Classification accuracy by using different data samples for test data (a). 10 samples, (b). 100 samples, (c).1000 samples, for SVM and KFDA and multi-class FDA classifiers.**

**Table 4.** *Comparison of the performance time of different classification methods in the different number of testing samples.*

| Classifier | Number of test samples | Time (s) |
|---|---|---|
| SVM | 10 | 0.865 |
| | 100 | 0.844 |
| | 1000 | 1.18 |
| Multi-class FDA | 10 | 0.045 |
| | 100 | 0.024 |
| | 1000 | 0.116 |
| KFDA | 10 | 0.144 |
| | 100 | 0.155 |
| | 1000 | 0.246 |

**Table 5.** *Comparison of the performance accuracy of different classification methods with 10 testing samples and 10000 and 100000 training samples.*

| Classifier | Accuracy (10000 training samples) | Accuracy (100000 training samples) |
|---|---|---|
| SVM | 80% | 82% |
| Multi-Class FDA | 81% | 83% |
| KFDA | 78% | 82% |

## 4. CONCLUSION

In this paper, the performance of the proposed FDA method for the multi-class case, i.e. the classification of 6 classes, is evaluated and a comparison between SVM, KFDA, and multi-class FDA classifier without using a kernel function, in the case of classification of MPSK (M=2,4,8) and MQAM (m=16,64,256) signals, is given. Multi-class FDA and KFDA perform better than SVM in different SNR and with different training and testing data samples.

In cases that data is normally distributed and all groups are identically distributed and have different covariance matrixes, linear discriminant analysis methods perform better.

However, in SVM with the aim of generalizing the optimally separating hyperplane, the assumption is all groups are separable. In cases that groups are not linearly separable, SVM does not perform as good as LDA methods. SVM is an optimization problem, but LDA has an analytical solution. The op-

timization problem for SVM has a dual and a primal formulation that allows the user to optimize over either the number of data points or the number of variables, depending on which method is the most computationally feasible. It takes much more time for SVM classifier to solve the optimization problem, but in comparative, the LDA classifier computational time is really shorter. In this case, as shown in Table 4, multi-class FDA and KFDA perform faster than SVM and also more accurate than SVM in low SNRs. As the results written in Table 5, the number of training and testing samples is not affecting the classification accuracy for high SNRs, though the proposed method (Multi-Class FDA) is appropriate in terms of achieved accuracy for low SNRs and small sample size as well.

## REFERENCES

[1]    H. Jon and M. K. Simon, Autonomous SoftwareDefined Radio Receivers for Deep Space Applications (JPL Deep-Space Communications and Navigation Series), Hoboken, New Jersey: John Wiley & Sons, Inc., 2006.

[2]    A. Y. B.-N. W. S. O. Dobre, "Survey of automatic modulation classification techniques: classical approaches and new trends,," Communications , IET 1 (2), pp. 137-156, 2007.

[3]    Z. Zhu and A. K. Nandi, Automatic Modulation Classification: Principles, Algorithms and Applications, London: John Wiley & Sons, 2015.

[4]    X. Zhu, Y. Wu and B. Wang, "The MPSK Signals Modulation Classification Based on Kernel Methods," in 2008 8th International Symposium on Antennas, Propagation and EM Theory, Kumming,China, 2008.

[5]    U. Mengali and A. N.D'Andrea, Synchronization Techniques for Digital Receivers, New York, 1997.

[1]    E. El-Mahdy and N. M. Namazi, "Classification of multiple M-ary frequency-shift keying signals over a Rayleigh fading channel,," IEEE Transactions on Communications, vol. 50, no. 6, pp. 967-974, June 2002.

[6]    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.

[7]    Q. Shi, Y. Gong and Y. L. Guan, "Asynchronous Classification of High-Order QAMs," in IEEE Wireless Communications and Networking Conference, Las Vegas, 2008.

[8]    W. C. Headly and C. R. C. M. da Silva, "Asynchronous Classification of Digital Amplitude-Phase Modulated Signals in Flat-Fading Channels," IEEE Transactions on Communications, vol. 59, no. 1, pp. 7-12, 2011.

[9]    K. Ho, W. Prokopiw and Y. Chan, "Modulation identification of digital signals by the wavelet transform," IEE Proceedings-Radar, Sonar and Navigation, pp. 169-176, 2000.

[10]   F. YongMing, C. Liang, J. Zhu and S. Wang, "Extended Modulation Classification Scheme based on Approximate Entropy," in IEEE Symposium on Wireless Technology & Applications (ISWTA), Kuching, 2013.

[11]   W. A. Gardner, "Signal interception: a unifying theoretical framework for feature detection," IEEE Transactions on Communications, vol. 36, no. 8, pp. 897-906, August 1998.

[12] M. Spooner, W. A. Brown and G. K. Yeung, "Automatic radio-frequency environment analysis," in Conference Record of the Thirty-Fourth Asilomar Conference on Signals, Systems and Computers, Pacific Grove, 2002.

[13] M. Spooner, "On the utility of sixth-order cyclic cumulants for RF signal classification," in Conference Record of Thirty-Fifth Asilomar Conference on Signals, Systems and Computers, Pacific Grove, 2002.

[14] C.-S. Park, W. Jang, S.-P. Nah and D. Y. Kim, "Automatic Modulation Recognition using Support Vector Machine in Software Radio Applications," in The 9th International Conference on Advanced Communication Technology, Okamoto, 2007.

[15] F. Xie, C. Li and G. Wan, "An Efficient and Simple Method of MPSK Modulation Classification,," in 2008 4th International Conference on Wireless Communications, Networking and Mobile Computing, 2008.

[16] C.-S. Park, J.-H. Choi, S.-P. Nah, W. Jang and D. Y. Kim, "Automatic Modulation Recognition of Digital Signals using Wavelet Features and SVM," in 2008 10th International Conference on Advanced Communication Technology, Gangwon-Do, 2008.

[17] H. Gang, L. Jiandong and L. Donghua, "Study of modulation recognition based on HOCs and SVM," in 2004 IEEE 59th Vehicular Technology Conference. VTC 2004-Spring, Milan, 2004.

[1] C. Malady and A. A. Beex, "AMC Improvements from Robust Estimation," in 2010 IEEE Global Telecommunications Conference GLOBECOM 2010, Miami, 2010.

[18] Avci, "Selecting of the optimal feature subset and kernel parameters in digital modulation classification by using hybrid genetic algorithm–support vector machines: HGASVM," Expert Systems with Applications, vol. 36, no. 2, pp. 1391-1402, 2007.

[19] M. Pedzisz and A. Mansour, "Automatic modulation recognition of MPSK signals using constellation rotation and its 4th order cumulant," Digital Signal Processing, vol. 15, no. 3, pp. 295-304, 2005.

[20] J. Yuan, Z. Zhao-Yang and Q. Pei-Liang, "Modulation classification of communication signals," in IEEE MILCOM 2004. Military Communications Conference, Monterey, 2004.

[21] J. Shawe-Taylor and N. Cristianini, Kernel Methods for Pattern Analysis, London: Cambridge University Press, 2004.

[22] H. Wang and J. Yu, "Study on the kernel-based methods and its model selection," Journal of Southern Yangtze University (Natural Science Edition), vol. 5, no. 4, pp. 500-504, 2006.

[23] B. Ramkumar, "Automatic modulation classification for cognitive radios using cyclic feature detection," IEEE Circuits and Systems Magazine, vol. 9, no. 2, pp. 27-45, 2009.

[24] P. O. Amblard, M. Gaeta and J. L. Lacoume, "Statistics for complex variables and signals — Part I: Variables," Signal Processing, vol. 53, no. 1, pp. 1-13, August 1996.

[25] Y. Shaoquan and C. Weidong, "Classification of MPSK signals using cumulant invariants," Journal of Electronics (China), vol. 19, no. 1, pp. 99-103, January 2002.

[26] O. A. Dobre, Y. Bar-Ness and W. Su, "Higher-order cyclic cumulants for high order modulation classification," in IEEE Military Communications Conference, 2003. MILCOM, Boston, 2003.

[27] Z. Q. Bian and X. G. Zhang, Pattern Recognition, 2000.

[28] S. Theodoridis and K. Koutroumbas, Pattern Recognition, 4th Edition ed., vol. V, Elsevier Inc., 2008.

[29] Z. W. Y and G. Y. J, "Generalized kernel fisher function optimal discriminate in pattern recognition," Pattern Recognition and Artificial Intelligence, vol. 15, pp. 429-433, April 2002.

[30] Swami and B. M. Sadly, "Hierarchical digital modulation classification using cumulants," IEEE Trans. Commun, vol. 48, no. 3, pp. 416-429, 2000.