



Research paper

Improving the Classification of MPSK and MQAM Modulations by Using Optimized Nonlinear Preprocess in Flat Fading Channels

I. Kadoun, H. Khaleghi Bizaki*

Electrical and Computer Engineering Department, Malek Ashtar University of Technology, Tehran, Iran.

Article Info

Article History:

Received 11 April 2022
Reviewed 15 May 2022
Revised 15 July 2022
Accepted 30 August 2022

Keywords:

Automatic modulation classification
Linear discriminant analysis
Higher-Order cumulants
Mahalanobis distance

*Corresponding Author's Email
Address: bizaki@yahoo.com

Abstract

Background and Objectives: Intelligent receivers, automatically detect the digital modulation type of the received signals for demodulation purposes where is well known as Automatic Modulation Classification (AMC) module. The performance of AMC algorithms depends on the channel conditions where for example, in fading channel its performance gets worse than the AWGN channel.

Methods: We propose a new algorithm for improving the AMC classification accuracy in flat fading channels. The proposed algorithm consists of an optimizable nonlinear preprocess followed by Linear Discriminant Analysis (LDA) technique. Two Lemmas have been found for extracting the optimization rule. And an optimization algorithm has been built based on the previous Lemmas.

Results: The simulation results show that the proposed algorithm improves the classification accuracy between 8-Phase Shift Keying (8PSK) and 16PSK (as an example of M-array PSK (MPSK) inter-class) for Signal-to-noise ratio (SNR) values greater than 13 dB, and between 16-quadrature amplitude shift modulation (16QAM) and 64QAM (as an example of M-array QAM (MQAM) inter-class) for SNR values greater than 4 dB. On the other hand, the classification accuracy of MPSK and MQAM is improved using the proposed algorithm compared with reference papers. Its improvement is up to 10.79% compared with the [1] and up to 38.552% compared with [2].

Conclusion: By using the proposed optimization algorithm, the AMC classification accuracy has been improved. Other classification problems can use this algorithm. And other nonlinear preprocess functions or optimization algorithms may be found in future work.

This work is distributed under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)



Introduction

With the significant development of modern communication technology, the AMC of the received signal is becoming more critical. Two primary AMC techniques are known: Likelihood-Based (LB) and Feature-Based (FB). LB techniques suffer from high computational complexity and need to estimate the unknown parameters [3], [4]. On the other hand, FB techniques have less complexity, don't need any parameter estimation [5], [6], and can work under different conditions like multipath fading channels [7].

Various studies were done to find good discriminative features for modulation classification like instantaneous time-domain features, Fourier and wavelet transform, higher-order moments, and cumulants [5]-[11]. Comparisons between the performances of these features were made in [12], [13]. According to their results, Higher-Order Cumulants (HOCs) are the best features under different conditions.

Different studies and simulations were done for AMC in fading channels using different HOCs. For example, in

[1], the author shows that the performance accuracy of MPSK and MQAM classification by using HOMs and HOCs is 84.37%. While in [2], the author shows that the performance accuracy of binary phase-shift keying (BPSK), Quadrature Phase-Shift Keying (QPSK), 8-phase-shift keying (8-PSK), 16-PSK, 16-quadrature amplitude modulation (16-QAM), 32-QAM, and 64-QAM classification by using cyclic cumulants is 89.8%, for SNR value of 15 dB.

The most critical inter-class modulation types are MPSK and MQAM [1], [2]. Most of the well-known intra-class modulation types are shown in Table 1.

Table 1: Chosen types of MPSK and MQAM digital modulations

Inter-class modulation	Intra-class modulation
MPSK	BPSK, QPSK, 8PSK, 16PSK
MQAM	8QAM, 16QAM, 32QAM, 64QAM

Our study improves the AMC performance by enhancing the discrimination between some intra-class digital modulation types in Table 1 (like 8PSK and 16PSK, and like 16QAM and 64QAM) for lower SNR values. This improvement is made by optimizing an added nonlinear preprocess function. Two primary optimizable nonlinear functions have been developed: regularized distance-based and nonlinear transformation. The simulation results show that these optimized functions could improve the discrimination between 8PSK and 16PSK for SNR values greater than 13 dB and between 16QAM and 64QAM for SNR values greater than 4 dB. On the other hand, the classification accuracy of MPSK and MQAM has been improved using the proposed algorithm compared with reference papers [1], [2]. The maximum improvement of our proposed algorithm compared with the reference paper [1] is 10.79%, and the maximum improvement of our proposed algorithm compared with the reference paper [2] is 38.552%.

System Model

Consider the received signal in flat fading channel as:

$$r_i(n) = \alpha w_i(n) + v(n) \tag{1}$$

where α is the complex channel fading coefficient which is considered $\alpha \in CN(0,1)$, $w_i(n)$ is the transmitted symbol which is considered an independent and identically distributed (i.i.d) process, and $v(n)$ is the additive white Gaussian noise (AWGN) and is considered $v(n) \in CN(0, \sigma_n^2)$.

The general mathematical form of the HOC is defined as [14], [15]:

$$C_{p,q} = Cum \left[\underbrace{r_1, \dots, r_{p-q}}_{p-q \text{ terms}}, \underbrace{r_{p-q+1}^*, \dots, r_p^*}_{q \text{ terms}} \right] \tag{2}$$

where * denotes the complex conjugate, p is the order of the cumulant, q is the complex conjugate order of the cumulant, and *cum* function is defined as [14]:

$$Cum[r_1, \dots, r_n] = \sum_{\forall v} (-1)^{q-1} (q-1)! E \left[\prod_{j \in V_1} r_j \right] \dots E \left[\prod_{j \in V_q} r_j \right] \tag{3}$$

and the summation is being performed on all partitions $V = (V_1, V_2, \dots, V_q)$ for the set of indexes $(1, 2, \dots, n)$.

To cancel the effect of the power level of the received signal, the first type of normalization must be done [15], [16]:

$$C'_{pq} = \frac{C_{pq}}{(C_{21})^{p/2}} \tag{4}$$

The magnitude of the eighth, sixth, and fourth-order cumulants is greater than that of the second-order cumulants. As a result, we have different values for the other HOC orders. The second normalization can reduce the values range as [15], [17]:

$$\tilde{C}_{pq} = (C'_{pq})^{2/p} \tag{5}$$

According to our simulation results for the selected digital modulation types in Table 1, \tilde{C}_{40} , \tilde{C}_{61} , and \tilde{C}_{80} (equations (6), (7), and (8)) have the most discrimination ability, so they have been chosen in our study [15], [18], [19]:

$$C_{40} = M_{40} - 3M_{20}^2 \tag{6}$$

$$C_{61} = M_{61} - 5M_{21}M_{40} - 10M_{20}M_{41} + 30M_{20}^2M_{21} \tag{7}$$

$$C_{80} = M_{80} - 35M_{40}^2 - 28M_{60}M_{20} + 420M_{20}^2M_{40} - 630M_{20}^4 \tag{8}$$

where [15], [18], [19]:

$$M_{pq} = E \left[r(k)^{p-q} r^*(k)^q \right] \tag{9}$$

is the moment of received signal $r(k)$.

Mathematical Preliminary

A. Linear Discriminant Analysis (LDA)

This technique finds the optimum linear projection vector that maximizes the discrimination between digital modulation types [20]. We define the input features of the two classes for dataset i as:

$$\mathbf{x}_i = \left[\left(\tilde{C}_{40}(r_i) \right)_i \quad \left(\tilde{C}_{61}(r_i) \right)_i \quad \left(\tilde{C}_{80}(r_i) \right)_i \right]^T$$

$$, r_i \in \mathcal{C}_x, i = 1..n_x \quad (10)$$

$$\mathbf{y}_i = \left[\left(\tilde{C}_{40}(r_i) \right)_i \quad \left(\tilde{C}_{61}(r_i) \right)_i \quad \left(\tilde{C}_{80}(r_i) \right)_i \right]^T$$

$$, r_i \in \mathcal{C}_y, i = 1, \dots, n_y \quad (11)$$

These input features can be written as:

$$\mathbf{x}_i := \left[\tilde{C}_{40,x_i} \quad \tilde{C}_{61,x_i} \quad \tilde{C}_{80,x_i} \right]^T \quad (12)$$

$$\mathbf{y}_i := \left[\tilde{C}_{40,y_i} \quad \tilde{C}_{61,y_i} \quad \tilde{C}_{80,y_i} \right]^T \quad (13)$$

The mean vectors of the input features can be calculated as [20]:

$$\boldsymbol{\mu}_x = \left[E(\tilde{C}_{40,x_i}) \quad E(\tilde{C}_{61,x_i}) \quad E(\tilde{C}_{80,x_i}) \right]^T$$

$$:= \left[\mu_{x,1} \quad \mu_{x,2} \quad \mu_{x,3} \right]^T \quad (14)$$

$$\boldsymbol{\mu}_y = \left[E(\tilde{C}_{40,y_i}) \quad E(\tilde{C}_{61,y_i}) \quad E(\tilde{C}_{80,y_i}) \right]^T$$

$$:= \left[\mu_{y,1} \quad \mu_{y,2} \quad \mu_{y,3} \right]^T \quad (15)$$

By defining the projection vector \mathbf{u} , the output features x'_i, y'_i of the first and second classes respectively can be calculated as [20]:

$$x'_i = \mathbf{x}_i^T \mathbf{u}, \quad y'_i = \mathbf{y}_i^T \mathbf{u} \quad (16)$$

Fisher criterion function represents the discrimination measurement between the output features of two classes as [20], [21]:

$$J(\mathbf{u}) := \frac{(\boldsymbol{\mu}'_x(\mathbf{u}) - \boldsymbol{\mu}'_y(\mathbf{u}))^2}{\sigma_x'^2(\mathbf{u}) + \sigma_y'^2(\mathbf{u})} = \frac{\mathbf{u}^T \mathbf{S}_B \mathbf{u}}{\mathbf{u}^T \mathbf{S}_W \mathbf{u}} \in \mathbb{R} \quad (17)$$

where \mathbf{u} is the projection vector, $\boldsymbol{\mu}'_x(\mathbf{u}), \boldsymbol{\mu}'_y(\mathbf{u})$ are the means of the output features for the first and the second classes, respectively, $\sigma_x'^2(\mathbf{u}), \sigma_y'^2(\mathbf{u})$ are the variances of the output features for the first and second classes, respectively. \mathbf{S}_B And \mathbf{S}_W are defined as:

$$\mathbf{S}_B = (\boldsymbol{\mu}_y - \boldsymbol{\mu}_x)(\boldsymbol{\mu}_y - \boldsymbol{\mu}_x)^T \in \mathbb{R}^{d \times d} \quad (18)$$

$$\mathbf{S}_W := \frac{1}{n_x - 1} \sum_{i=1}^{n_x} (\mathbf{x}_i - \boldsymbol{\mu}_x)(\mathbf{x}_i - \boldsymbol{\mu}_x)^T$$

$$+ \frac{1}{n_y - 1} \sum_{i=1}^{n_y} (\mathbf{y}_i - \boldsymbol{\mu}_y)(\mathbf{y}_i - \boldsymbol{\mu}_y)^T \in \mathbb{R}^{d \times d} \quad (19)$$

where n_x, n_y are the numbers of samples for the two classes, respectively. The optimum projection vector \mathbf{u} can be calculated by solving the maximization of the

Fisher criterion function problem of (17) for \mathbf{u} . One of the solutions is using the Lagrange method as [21]:

$$L = \mathbf{u}^T \mathbf{S}_B \mathbf{u} - \lambda (\mathbf{u}^T \mathbf{S}_W \mathbf{u} - 1) \quad (20)$$

where λ is the Lagrange multiplier. Equating the derivative of L to zero gives [21]:

$$\frac{\partial L}{\partial \mathbf{u}} = 2\mathbf{S}_B \mathbf{u} - 2\lambda \mathbf{S}_W \mathbf{u} = 0 \Rightarrow \mathbf{S}_B \mathbf{u} = \lambda \mathbf{S}_W \mathbf{u} \quad (21)$$

which is a generalized eigenvalue problem. One possible solution to the above-generalized eigenvalue problem can be found as [21]:

$$\mathbf{u} = \text{eig}(\mathbf{S}_W^{-1} \mathbf{S}_B) \quad (22)$$

where $\text{eig}(\cdot)$ denotes the eigenvector of the matrix with the largest eigenvalue.

In the following Sections, the LDA algorithm is called the classical LDA.

B. Discrimination measurement

One of the well-known statistical distance measurements between two random variables is Mahalanobis Distance (MD). Suppose \mathbf{v}_x and \mathbf{v}_y are random variables. The MD distance between them can be calculated as [22]:

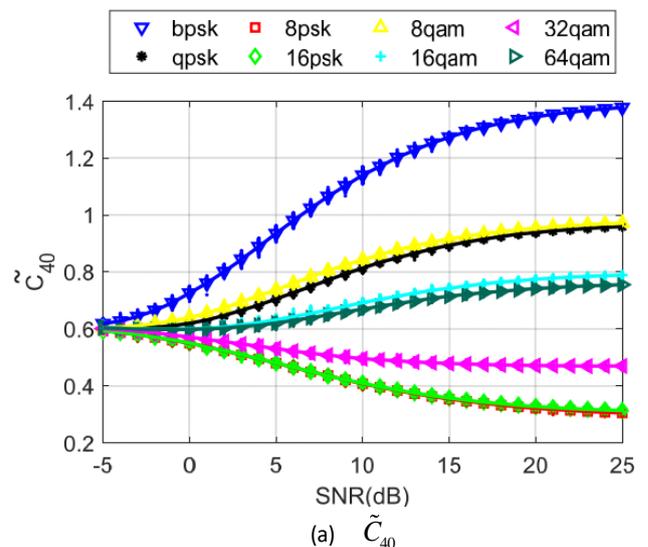
$$d(\mathbf{v}_x, \mathbf{v}_y) = \sqrt{(\boldsymbol{\mu}_x - \boldsymbol{\mu}_y)^T (\mathbf{S}_x + \mathbf{S}_y)^{-1} (\boldsymbol{\mu}_x - \boldsymbol{\mu}_y)} \in \mathbb{R} \quad (23)$$

where $\boldsymbol{\mu}_x, \boldsymbol{\mu}_y$ are mean vectors and $\mathbf{S}_x, \mathbf{S}_y$ are the covariance matrices of the random variables $\mathbf{v}_x, \mathbf{v}_y$, respectively.

This study uses the MD as a discrimination measurement between two random variables.

Conventional Classical LDA-based AMC Problem

The values of the selected HOCs in Section 2, i.e. $\tilde{C}_{40}, \tilde{C}_{61}, \tilde{C}_{80}$, are shown in Fig. 1, for the selected digital modulations in Table 1, and SNR rang [-5:25] dB.



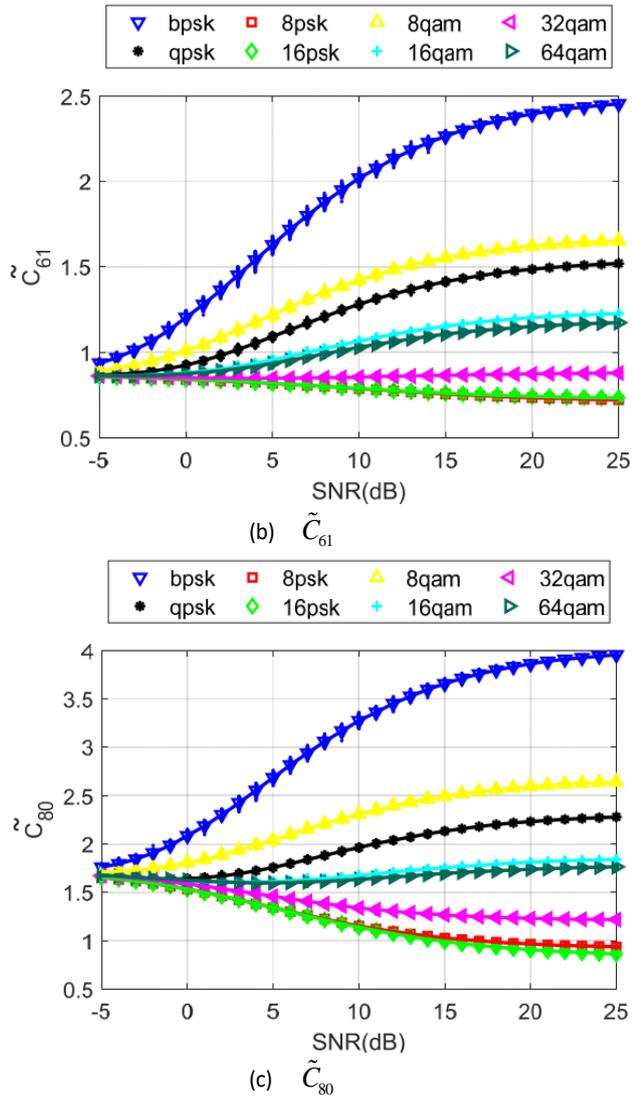


Fig. 1: Values of \tilde{C}_{40} , \tilde{C}_{61} , and \tilde{C}_{80} cumulants respectively.

From Fig. 1, some modulations can be classified easily (like BPSK, QPSK, 8QAM, and 32QAM). In contrast, the others are close to each other (like 8PSK and 16PSK, and like 16QAM and 64QAM). This situation would be worse for lower SNR values.

We define two different problems:

- 8PSK and 16PSK classification as problem p1.
- 16QAM and 64QAM classification as problem p2.

Our work aims to find a new algorithm that separately improves the classification accuracy for the two problems, p1 and p2.

Start with classical LDA to solve the mentioned problems p1 and p2. Calculation of the classification accuracy (ACC) for the problems p1 and p2 using the selected HOCs in Section 2 and the classical LDA algorithm have been done as shown in Fig. 2. As shown in Fig. 2, classical LDA doesn't improve the performance accuracy of 8PSK and 16PSK classification (problem p1) and 16QAM and 64QAM classification (problem p2). As shown in the

next section, we propose modifying the classical LDA algorithm by adding an optimizable nonlinear preprocess.

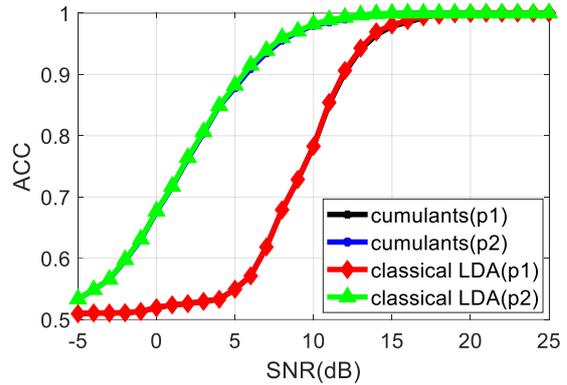


Fig. 2: Classification accuracy using the selected HOCs and the classical LDA for the problems p1 and p2.

Proposed Preprocess LDA Algorithm

The proposed preprocess LDA algorithm consists of an optimizable nonlinear preprocess, followed by the LDA algorithm.

A. The Proposed Mathematical Problem

The selected HOCs can be rewritten as: $C_{x_i,1} := \tilde{C}_{40,x_i}$, $C_{x_i,2} := \tilde{C}_{61,x_i}$, $C_{x_i,3} := \tilde{C}_{80,x_i}$ for the first class C_x , the input features vector becomes $[C_{x_i,1} \ C_{x_i,2} \ C_{x_i,3}]^T$, and $C_{y_i,1} := \tilde{C}_{40,y_i}$, $C_{y_i,2} := \tilde{C}_{61,y_i}$, $C_{y_i,3} := \tilde{C}_{80,y_i}$ for the second class C_y , the input features vector becomes $[C_{y_i,1} \ C_{y_i,2} \ C_{y_i,3}]^T$.

As shown in Section 4, the classical LDA algorithm needs adjustment to improve the discrimination between 8PSK and 16PSK, and between 16QAM and 64QAM modulations for low SNR values. An example of this adjustment is the addition of nonlinear function as follows: $f_1(C_{x_i,1} \text{ or } C_{y_i,1})$, $f_2(C_{x_i,2} \text{ or } C_{y_i,2})$, and $f_3(C_{x_i,3} \text{ or } C_{y_i,3})$ of the selected HOCs in Section 2 as shown in Fig. 3.

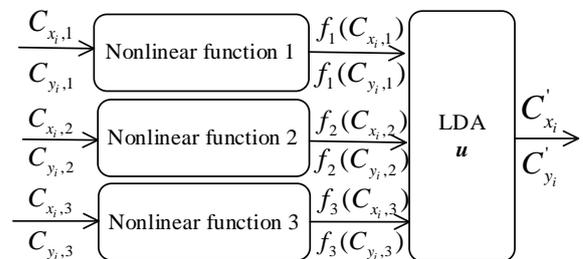


Fig. 3: Block diagram of the proposed nonlinear preprocess LDA algorithm.

where C'_{x_i} , C'_{y_i} are the output features of the first and second classes, respectively. These output features can be calculated as:

$$C'_{x_i} = u_1 f_1(C_{x_i,1}) + u_2 f_2(C_{x_i,2}) + u_3 f_3(C_{x_i,3}) = \mathbf{u}^T \mathbf{f}_{x_i} \quad (24)$$

$$C'_{y_i} = u_1 f_1(C_{y_i,1}) + u_2 f_2(C_{y_i,2}) + u_3 f_3(C_{y_i,3}) = \mathbf{u}^T \mathbf{f}_{y_i} \quad (25)$$

where $\mathbf{u} = [u_1, u_2, u_3]^T$ is the projection vector,

$$\mathbf{f}_{x_i} = [f_1(C_{x_i,1}) \quad f_2(C_{x_i,2}) \quad f_3(C_{x_i,3})]^T := [f_{x_i,1} \quad f_{x_i,2} \quad f_{x_i,3}]^T$$

is the vector of the values of the nonlinear functions for the first class, and:

$$\mathbf{f}_{y_i} = [f_1(C_{y_i,1}) \quad f_2(C_{y_i,2}) \quad f_3(C_{y_i,3})]^T := [f_{y_i,1} \quad f_{y_i,2} \quad f_{y_i,3}]^T$$

is the vector of the values of the nonlinear functions for the second class.

The terms $\mu'_x(\mathbf{u}), \mu'_y(\mathbf{u})$ of the Fisher criterion function (17) are the means of the output features. They are calculated using (24) and (25) as:

$$\mu'_x = \mathbf{u}^T E(\mathbf{f}_x) \quad (26)$$

$$\mu'_y = \mathbf{u}^T E(\mathbf{f}_y) \quad (27)$$

The terms $\sigma_1'^2(\mathbf{u}), \sigma_2'^2(\mathbf{u})$ of the Fisher criterion function (17) are the variances of the output features. They are calculated as:

$$(\sigma'_x)^2 = \mathbf{u}^T \text{cov}(\mathbf{f}_x) \mathbf{u} \quad (28)$$

$$(\sigma'_y)^2 = \mathbf{u}^T \text{cov}(\mathbf{f}_y) \mathbf{u} \quad (29)$$

By using (26), (27), (28), and (29), the Fisher criterion (17) can be written as:

$$J(\mathbf{u}) = \frac{(\mu'_y(\mathbf{u}) - \mu'_x(\mathbf{u}))^2}{\sigma_1'^2(\mathbf{u}) + \sigma_2'^2(\mathbf{u})} = \frac{\mathbf{u}^T [E(\mathbf{f}_y) - E(\mathbf{f}_x)] [E(\mathbf{f}_y) - E(\mathbf{f}_x)]^T \mathbf{u}}{\mathbf{u}^T [\text{cov}(\mathbf{f}_x) + \text{cov}(\mathbf{f}_y)] \mathbf{u}} := \frac{\mathbf{u}^T \mathbf{S}_B \mathbf{u}}{\mathbf{u}^T \mathbf{S}_W \mathbf{u}} \quad (30)$$

where $\mathbf{S}_B := [E(\mathbf{f}_y) - E(\mathbf{f}_x)] [E(\mathbf{f}_y) - E(\mathbf{f}_x)]^T$ and $\mathbf{S}_W := [\text{cov}(\mathbf{f}_x) + \text{cov}(\mathbf{f}_y)]$.

The nonlinear preprocess function allows us to control $\mathbf{S}_B, \mathbf{S}_W$, which affects the discrimination performance.

The task is to find the rules that maximize the discrimination between two classes (Fisher criterion (30)) using the nonlinear preprocesses.

B. Necessary Lemmas

Lemma 1. For $J, \mathbf{S}_B, \mathbf{S}_W$ which are defined in (30) and (17), we find that:

$$\max(J) = \text{trace}(\mathbf{S}_W^{-1} \mathbf{S}_B) \quad (31)$$

Proof: Here, we mention some mathematical analyzes and results:

- The maximum value of the Fisher criterion function (17) is equal to the maximum value of eigenvalues of the matrix $\mathbf{S}_W^{-1} \mathbf{S}_B$ [23]:

$$\max(J) = \lambda_{\max}(\mathbf{S}_W^{-1} \mathbf{S}_B) \quad (32)$$

- The summation of eigenvalues of a matrix is equal to the trace of the matrix [24]:

$$\text{trace}(\mathbf{S}_W^{-1} \mathbf{S}_B) = \sum \lambda_i \quad (33)$$

- The production of eigenvalues of a matrix is equal to the determinant of the matrix [24]:

$$\det(\mathbf{S}_W^{-1} \mathbf{S}_B) = \prod \lambda_i \quad (34)$$

- By noticing \mathbf{S}_B calculation in (18), we find that $\det(\mathbf{S}_B)$ is equal to zero.

- $\det(\mathbf{S}_W^{-1} \mathbf{S}_B) = \det(\mathbf{S}_W^{-1}) \det(\mathbf{S}_B) = 0$, which means (34) is no longer helpful for calculating λ_i .

- According to [21], the rank of $\mathbf{S}_W^{-1} \mathbf{S}_B$ can be calculated as:

$$\text{rank}(\mathbf{S}_W^{-1} \mathbf{S}_B) = \min(T, n-1, L-1) \quad (35)$$

where n is the size of the dataset in each class, L is the number of classes, and T is the number of features. In our case, $L=2, T=3$, and $n \gg L, T$. We find that the rank ($\mathbf{S}_W^{-1} \mathbf{S}_B$) value is equal to 1. Which means we have *one nonzero eigenvalue*. By using (33) we find that:

$$\text{trace}(\mathbf{S}_W^{-1} \mathbf{S}_B) = \lambda \neq 0 \quad (36)$$

- Finally, by using (32) and (36), we find that $\max(J) = \text{trace}(\mathbf{S}_W^{-1} \mathbf{S}_B)$.

Lemma 2: Maximization of the Fisher criterion (30) is equivalent to maximization of the Mahalanobis distance between the values of the nonlinear functions for each feature, i.e Fig. 3.

Proof: According to Lemma 1, maximization of the Fisher criterion function means maximization of $\text{trace}(\mathbf{S}_W^{-1} \mathbf{S}_B)$. So, we have to study the effect of \mathbf{S}_B and \mathbf{S}_W elements on the Fisher criterion function. To simplify it, we study two-class cases where $\mathbf{S}_B, \mathbf{S}_W \in \mathbb{R}^{2 \times 2}$:

$$\mathbf{S}_B = [\boldsymbol{\mu}_y - \boldsymbol{\mu}_x] [\boldsymbol{\mu}_y - \boldsymbol{\mu}_x]^T = \begin{bmatrix} \Delta_1^2 & \Delta_1 \Delta_2 \\ \Delta_1 \Delta_2 & \Delta_2^2 \end{bmatrix} \in \mathbb{R}^{2 \times 2} \quad (37)$$

where $\Delta_1 = \mu_{y,1} - \mu_{x,1}$ is the difference between the means of the two classes for the first feature and $\Delta_2 = \mu_{y,2} - \mu_{x,2}$ is the difference between the means of the two classes for the second feature.

and:

$$\begin{aligned} \mathbf{S}_W &= \text{cov}(\mathbf{x}) + \text{cov}(\mathbf{y}) = \\ & \begin{bmatrix} (\sigma_{x,1})^2 & \rho_x \sigma_{x,1} \sigma_{x,2} \\ \rho_x \sigma_{x,1} \sigma_{x,2} & (\sigma_{x,2})^2 \end{bmatrix} + \begin{bmatrix} (\sigma_{y,1})^2 & \rho_y \sigma_{y,1} \sigma_{y,2} \\ \rho_y \sigma_{y,1} \sigma_{y,2} & (\sigma_{y,2})^2 \end{bmatrix} \quad (38) \\ &= \begin{bmatrix} g_1^2 + h_1^2 & \rho_x g_1 g_2 + \rho_y h_1 h_2 \\ \rho_x g_1 g_2 + \rho_y h_1 h_2 & g_2^2 + h_2^2 \end{bmatrix} \end{aligned}$$

where $g_1 = \sigma_{x,1}$ and $g_2 = \sigma_{x,2}$ are the variances of the first and second features for the first class, $h_1 = \sigma_{y,1}$ and $h_2 = \sigma_{y,2}$ are the variances of the first and second features for the second class. ρ_x and ρ_y are the correlations between the features of the first class and second class, respectively. By using (37) and (38), the trace of $\mathbf{S}_W^{-1} \mathbf{S}_B$ can be calculated as:

$$\begin{aligned} tr &= \text{trace}(\mathbf{S}_W^{-1} \mathbf{S}_B) = \\ & \frac{\Delta_1^2 (g_2^2 + h_2^2) + \Delta_2^2 (g_1^2 + h_1^2) - 2\Delta_1 \Delta_2 (\rho_x g_1 g_2 + \rho_y h_1 h_2)}{(g_1^2 + h_1^2)(g_2^2 + h_2^2) - (\rho_x g_1 g_2 + \rho_y h_1 h_2)^2} \\ &= \frac{\Delta_1^2 (g_2^2 + h_2^2) + \Delta_2^2 (g_1^2 + h_1^2) - 2\Delta_1 \Delta_2 (\rho_x g_1 g_2 + \rho_y h_1 h_2)}{(g_1^2 + h_1^2)(g_2^2 + h_2^2) \left[1 - \frac{(\rho_x g_1 g_2 + \rho_y h_1 h_2)^2}{(g_1^2 + h_1^2)(g_2^2 + h_2^2)} \right]} \\ &= \lambda \neq 0 \quad (39) \end{aligned}$$

Discussion: Starting with \mathbf{S}_B , since Δ_1 and Δ_2 are the differences between the means of the first and second features for the two classes, respectively, we find from (39) that:

$$\lim_{\Delta_1 \rightarrow \mp\infty} (tr) = +\infty, \quad \lim_{\Delta_2 \rightarrow \mp\infty} (tr) = +\infty \quad (39)$$

which means, by increasing the absolute values of the variables Δ_1 and Δ_2 , the Fisher criterion function value (17) will increase and vice versa.

To study the effect of \mathbf{S}_W elements on the Fisher criterion function, by noticing that g_1, h_1 are the variances of the first feature for the two classes, we find from (39) that:

$$\begin{aligned} \text{Since: } & g_1 \geq 0, h_1 \geq 0 \\ \text{When: } & g_1 + h_1 \rightarrow 0 \Leftrightarrow g_1 \rightarrow 0 \text{ and } h_1 \rightarrow 0 \\ \text{Thus: } & \lim_{\substack{g_1 \rightarrow 0 \\ h_1 \rightarrow 0}} (tr) = +\infty \end{aligned} \quad (40)$$

In the same way for g_2, h_2 , we find from (39) that:

$$\begin{aligned} \text{Since: } & g_2 \geq 0, h_2 \geq 0 \\ \text{When: } & g_2 + h_2 \rightarrow 0 \Leftrightarrow g_2 \rightarrow 0 \text{ and } h_2 \rightarrow 0 \\ \text{Thus: } & \lim_{\substack{g_2 \rightarrow 0 \\ h_2 \rightarrow 0}} (tr) = +\infty \end{aligned} \quad (41)$$

which means, by decreasing the values of the variables $g_1 + h_1$ and $g_2 + h_2$, the Fisher criterion function value (17) will increase and vice versa.

From (40), (41), and (42), we find that to maximize the Fisher criterion function, we have to maximize Δ_1, Δ_2 of the matrix \mathbf{S}_B , and minimize $g_1 + h_1, g_2 + h_2$ of the matrix \mathbf{S}_W . The same thing must have been done for \mathbf{S}_B and \mathbf{S}_W in (30).

By defining the means and variances of functions values of features as follows: $\mu_{f,x,i} := E(f_{x,i}), \mu_{f,y,i} := E(f_{y,i}), (\sigma_{f,x,i})^2 := \text{var}(f_{x,i}), (\sigma_{f,y,i})^2 := \text{var}(f_{y,i})$. Maximizing the elements of the matrix \mathbf{S}_B means finding the optimum nonlinear transformation which satisfies:

$$\begin{aligned} f_i &= \underset{f_i}{\text{argmax}} (E(f_{y,i}) - E(f_{x,i}))^2 = \\ & \underset{f_i}{\text{argmax}} (\mu_{f,y,i} - \mu_{f,x,i})^2, \quad i = 1, 2, 3 \end{aligned} \quad (42)$$

Minimizing the elements of the matrix \mathbf{S}_W means finding the optimum nonlinear transformation which satisfies:

$$\begin{aligned} f_i &= \underset{f_i}{\text{argmin}} (\text{var}(f_{x,i}) + \text{var}(f_{y,i})) \\ &= \underset{f_i}{\text{argmin}} ((\sigma_{f,x,i})^2 + (\sigma_{f,y,i})^2), \quad i = 1, 2, 3 \end{aligned} \quad (43)$$

By combining (43) and (44) we find:

$$f_i = \underset{f_i}{\text{optimum}} \left\{ \begin{array}{l} \text{argmax} (\mu_{f,y,i} - \mu_{f,x,i})^2 \\ \text{argmin} ((\sigma_{f,x,i})^2 + (\sigma_{f,y,i})^2) \end{array} \right\} \quad (44)$$

By noticing (23), for each feature $i=1,2,3$, we find that maximizing the MD is equivalent to the condition (45) as:

$$f_i = \underset{f_i}{\text{argmax}} (MD(f_{x,i}, f_{y,i})), \quad i = 1, 2, 3 \quad (45)$$

which is the same as the condition denoted in lemma 2.

C. The Solution to The Proposed Mathematical Problem

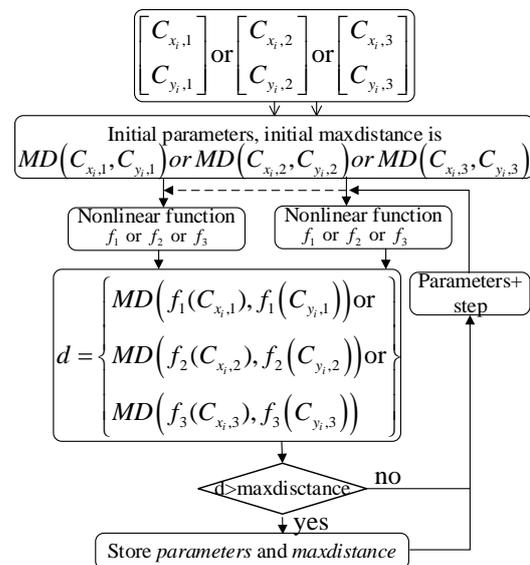


Fig. 4: Optimization of nonlinear preprocess.

Lemma 2 means, to maximize the Fisher criterion, we have to choose the optimum parameters that satisfy (46) for each feature. We propose a simple search algorithm to find these optimum parameters for each selected HOC in Section 2. The proposed algorithm is depicted in Fig. 4.

The proposed optimal nonlinear preprocess LDA algorithm of Fig. 4 consists of two stages where each stage consists of two steps (see Fig. 5):

a- Training stage:

Step 1:

In this step, we calculate the parameters of the nonlinear preprocess (47), (49), and (50):

- Calculate the parameters of the nonlinear preprocess (f_1) for feature1 as $C_{x_i,1}, C_{y_i,1}$.
- Calculate the parameters of the nonlinear preprocess (f_2) for feature2 as $C_{x_i,2}, C_{y_i,2}$.
- Calculate the parameters of the nonlinear preprocess (f_3) for feature3 as $C_{x_i,3}, C_{y_i,3}$.

Step 2:

- Calculate the linear projection vector \mathbf{u} by solving the Eigenvalue problem (22) for S_B and S_W where is presented in (30).

b- Testing stage:

Step 1:

In this step, we apply the nonlinear preprocess (47), (49), and (50) using the calculated parameters in the previous stage as:

- Apply the nonlinear preprocess (f_1) for the feature1, i.e. $C_{x \text{ or } y,1}$, using their calculated parameters.
- Apply the nonlinear preprocess (f_2) for the feature2, i.e. $C_{x \text{ or } y,2}$, using their calculated parameters.
- Apply the optimized nonlinear preprocess (f_3) for the feature3, i.e. $C_{x \text{ or } y,3}$, using their calculated parameters.

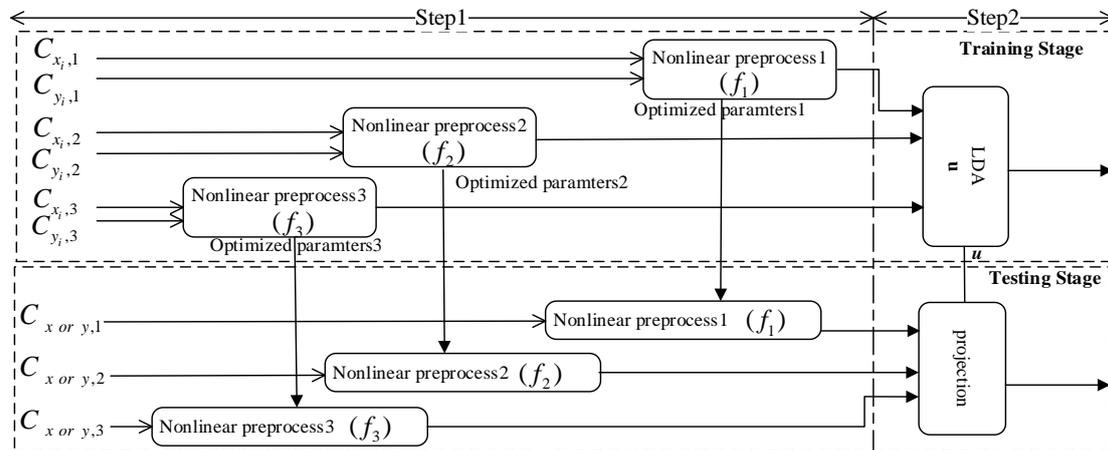


Fig. 5: Optimal nonlinear preprocess LDA-based algorithm.

Step 2:

- Apply the linear projection as (16) using the calculated projection vector in the previous stage.

Two general optimal nonlinear preprocesses have been studied here: regularized distance-based and optimized nonlinear transformation preprocesses.

D. *Regularized Distance-Based Preprocess*

To improve the discrimination between the classes, the distance between these features and the total mean is added to them as:

$$f_j(C_{x_i,j}) = C_{x_i,j} + (\Delta_{x_i,j} + \xi_j)^{-1} (C_{x_i,j} - \mu_j) \quad (46)$$

$$; i = 1..n, j = 1..d$$

$$f_j(C_{y_i,j}) = C_{y_i,j} + (\Delta_{y_i,j} + \xi_j)^{-1} (C_{y_i,j} - \mu_j)$$

$$; i = 1..n, j = 1..d$$

where

$$\Delta_{x_i,j} = (C_{x_i,j} - \mu_j)^2, \Delta_{y_i,j} = (C_{y_i,j} - \mu_j)^2 \quad (47)$$

is the distance between the feature j ($j=1, 2, \text{ or } 3$) and the

total mean of the two classes $\mu_j = \frac{\mu_{x,j} + \mu_{y,j}}{2}$ of the feature j , $C_{x_i,j}$ is the input feature j of the first class, $C_{y_i,j}$ is the input feature j of the second class, $f_j(C_{x_i,j})$ is the regularized distance-based feature value of the first digital modulation type, $f_j(C_{y_i,j})$ is the regularized distance-based feature value of the second class, and ξ_j is the regularizer of the feature j . This regularizer aims to optimize this nonlinear transformation according to (46). We call it the proposed-dist LDA algorithm.

E. *optimized nonlinear transformation*

Another way to find an optimal nonlinear preprocess that satisfies (46), is to add some parameters (here we add two parameters like L_1, L_2) to some known nonlinear transformations. Two nonlinear transformations are used, Box-Cox [25] and tangent hyperbolic (tanh) transformations a [26]:

Box-Cox transformation is defined as [25]:

$$f_j(C_{x_i,j}, L_1, L_2) = \begin{cases} \frac{(C_{x_i,j} + L_2)^{L_1} - 1}{L_1} & \text{if } L_1 \neq 0 \\ \log(C_{x_i,j} + L_2) & \text{if } L_1 = 0 \end{cases} \quad (48)$$

$$f_j(C_{y_i,j}, L_1, L_2) = \begin{cases} \frac{(C_{y_i,j} + L_2)^{L_1} - 1}{L_1} & \text{if } L_1 \neq 0 \\ \log(C_{y_i,j} + L_2) & \text{if } L_1 = 0 \end{cases}$$

We call it the proposed-Box LDA algorithm.

Tangent hyperbolic (tanh) transformation can be defined as [26]:

$$f_j(C_{x_i,j}, L_1, L_2) = \tanh(L_1(C_{x_i,j} + L_2)) \quad (49)$$

$$f_j(C_{y_i,j}, L_1, L_2) = \tanh(L_1(C_{y_i,j} + L_2))$$

We call it the proposed-Tanh LDA algorithm.

The value of parameter L_2 for each feature is close to the total mean of feature $\mu_j = \frac{\mu_{x,j} + \mu_{y,j}}{2}$. This parameter can be determined quickly using the search algorithm in Fig. 4. While the value of L_1 depends on the feature values and the transformation function, it can be determined by using the search algorithm in Fig. 4. Still, it takes more time than itself for L_2 determination.

Time and Space Complexities

Here, we analyze the time and space complexities of the LDA and the proposed algorithms for the training and test stages. Then we compare them.

A. The Classical LDA Algorithm's Time and Space Complexities

To calculate time and space complexities, we suppose that the number of samples is $n_j = n$ for all classes c ($c=2$ in our case), and d is the number of features ($d=3$ in our case). Starting with training complexity, we find [20], [27]-[29]:

Table 2: Time and space complexities for training the classical LDA algorithm

Operation	Time complexity	Space complexity
$\{\mathbf{x}_i\}_{i=1}^n, \{\mathbf{y}_i\}_{i=1}^n$	0	$ncd=6n$
$\boldsymbol{\mu}_x, \boldsymbol{\mu}_y$	$cd(n+1)=6(n+1)$	$cd=6$
$\boldsymbol{\mu}$	$ncd+d=6n+3$	$d=3$
\mathbf{S}_B	$cd^2+cd=24$	$2d^2=18$
\mathbf{S}_W	$ncd^2+ncd=24n$	$2d^2=18$
\mathbf{S}_W^{-1}	$O(d^3)=27$ [29]	$O(d^2)=9$ [28]
$\mathbf{S}_W^{-1}\mathbf{S}_B$	$O(d^3)=27$ [29]	$d^2=9$
$\mathbf{u} = \text{eig}(\mathbf{S}_W^{-1}\mathbf{S}_B)$	$O(d^3)=27$ [20]	$O(d^2)=9$ [30]
Final complexity	$36n+114$	$6n+72$
Our case $n \gg 100$	$36n$	$6n$

This result is similar to the result in [27] and for testing complexity, we find:

Table 3: Time and Space complexities for testing the classical LDA algorithm

Operation	Time complexity	Space complexity
$\{\mathbf{x}_i\}_{i=1}^n, \{\mathbf{y}_i\}_{i=1}^n$	0	$6n$
LDA projection	$12n$	$2n$
Final complexity	$12n$	$8n$

B. The Proposed Algorithm's Time and Space Complexities

Similar to the previous Section, we must calculate the training and testing complexities. According to the proposed nonlinear functions, the maximum number of optimizable variables is two. Suppose a and b are the number of loops for the first and second variables. Starting with training complexity, we find:

Table 4: Time and Space complexities for training the proposed algorithm

Operation	Time complexity	Space complexity
Apply nonlinear preprocess	$2n$	$2n$
$\boldsymbol{\mu}_x, \boldsymbol{\mu}_y$	$6(n+1)$	6
\mathbf{S}_W	$18n$	6
Calculate MD using $\boldsymbol{\mu}_x, \boldsymbol{\mu}_y, \mathbf{S}_W$	12	3
Total complexity of one-time preprocess	$10n+6$	$2n+15$
Repeat for the first variable a times	$a(26n+18)$	$2n+6$
Repeat for the second variable b times	$ab(26n+18)$	$2n+6$
Apply LDA	$36n$	6n
Final complexity	$ab(26n+18)+36n$	$8n+6$
Our case $n \gg 100$	$26abn+36n$	8n

and for testing complexity, we find:

Table 5: Time and Space complexities for testing the proposed algorithm

Operation	Time complexity	Space complexity
$\{\mathbf{x}_i\}_{i=1}^n, \{\mathbf{y}_i\}_{i=1}^n$	0	$2n$
Apply nonlinear preprocess	$2n$	$2n$
LDA projection	$12n$	$8n$
Final complexity	$14n$	$12n$

C. A Comparison Between the Complexities of the Classical LDA and the Proposed Algorithm

Calculating the ratio of the proposed algorithm's complexity over the classical LDA algorithm's complexity is done to compare their complexities, as shown in Table 6.

Table 6: The ratio of the complexity of the proposed algorithm over the complexity of the classical LDA algorithm

stage	Ratio of time complexity	Ration of space complexity
training	$\approx ab+1$	1.25
testing	1.17	1.5

As shown in Table 6, the time complexity of training the proposed algorithm is higher than the time complexity of the classical LDA algorithm due to the optimization process. Otherwise, they are almost similar.

Simulation Results

A. Simulation of the Proposed Algorithm

Three steps for complete simulation:

- I. Optimize the proposed-dist LDA, proposed-Box, and proposed-Tanh algorithms for each selected HOC in Section 2, as shown in Fig. 4.
- II. Calculate of the linear projection vector u as shown in Fig. 5.
- III. Calculate the Number of Misclassified Datasets (NoMD) for the two mentioned problems: problem p1 in Section 4, the classification between 8PSK and 16PSK, and problem p2, which is the classification between 16QAM and 64QAM.

B. The Proposed-Dist LDA Algorithm

Fig. 6 shows the simulation results of the normalized NoMD values of the classical LDA (16) and the proposed-dist LDA algorithms for the problems p1 and p2, and SNR values [-5: 20] dB.

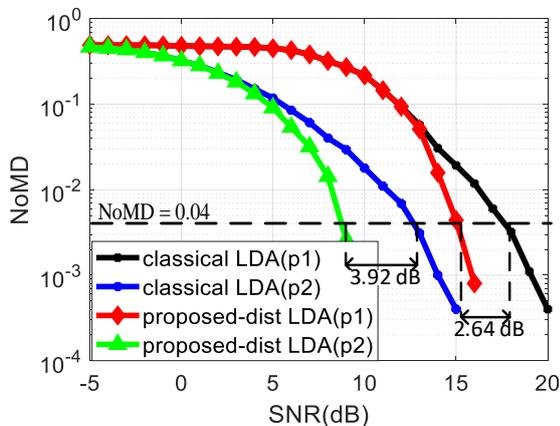


Fig. 6: Normalized NoMD values of the classical LDA and the proposed-dist LDA algorithms for the problems p1 and p2.

As shown in Fig. 6, the proposed-dist LDA algorithm could improve the discrimination between 8PSK and 16PSK for SNR values greater than 13 dB and between 16QAM and 64QAM for SNR values greater than 4 dB and for the normalized NoMD value of 0.04 (as an example), the improvement by using the proposed-dist LDA algorithm compared to the classical LDA algorithm is 2.64 dB for the problem p1 and 3.92 dB for the problem p2.

C. The Proposed-Box LDA Algorithm

Fig. 7 shows the simulation results of the normalized NoMD values of the classical LDA and the proposed-Box LDA algorithms for problems p1 and p2, and SNR values [-5: 20] dB.

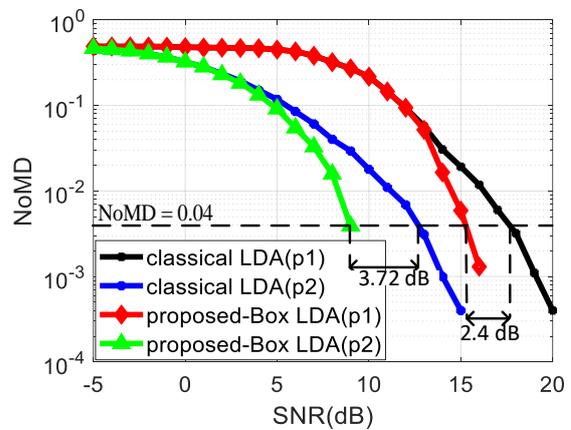


Fig. 7: Normalized NoMD values of the classical LDA and the proposed-Box LDA algorithms for the problems p1 and p2.

As shown in Fig. 7, the proposed-Box LDA algorithm could improve the discrimination between 8PSK and 16PSK for SNR values greater than 13 dB and between 16QAM and 64QAM for SNR values greater than 4 dB and for the normalized NoMD value of 0.04 (as an example), the improvement by using the proposed-Box LDA algorithm compared to the classical LDA algorithm is 2.4 dB for the problem p1 and 3.72 for the problem p2.

D. The Proposed-Tanh LDA Algorithm

Fig. 8 shows the simulation results of the normalized NoMD values of the classical LDA and the proposed-Tanh LDA algorithms for problems p1 and p2, and SNR values [-5: 20] dB.

As shown in Fig. 8, the proposed-Tanh LDA algorithm could improve the discrimination between 8PSK and 16PSK for SNR values greater than 13 dB and between 16QAM and 64QAM for SNR values greater than 4 dB and for the normalized NoMD value of 0.04 (as an example), the improvement by using the proposed-Tanh LDA algorithm compared to the classical LDA algorithm is 2.48 dB for the problem p1 and 4 dB for the problem p2.

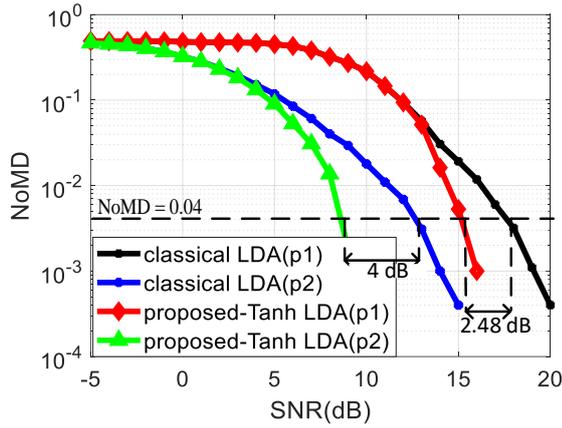


Fig. 8: Normalized NoMD values of the classical LDA and the proposed-Tanh LDA algorithms for the problems p1 and p2.

E. Classification Accuracy Improvement Compared with Reference Papers [1], [2]

MPSK and MQAM have been classified in reference papers [1], [2].

In [1], the author calculated the classification accuracy of MPSK and MQAM over a flat fading channel for SNR values of 10 dB and 0 dB. The classification accuracy of MPSK and MQAM is calculated using our optimized nonlinear LDA algorithm, i.e., the regularized distance-based LDA algorithm. The improvement of our proposed nonlinear LDA algorithm is calculated by subtracting the classification accuracy of the reference paper [1] from the classification accuracy of our proposed algorithm, as shown in Table 7.

Table 7: Comparison between the performance of the reference paper [1] and our proposed algorithm

SNR (dB)	0 dB	10 dB
Classification accuracy in Reference paper [1]	76.8%	84.37%
Classification accuracy of our proposed algorithm	78.25%	95.16%
The improvement of our proposed algorithm	1.45%	10.79%

As shown in Table 7, the classification accuracy of our proposed algorithm is improved compared with the reference paper [1]. The maximum improvement of our proposed algorithm compared with the reference paper [1] is 10.79%.

In [2], the author calculated the classification accuracy of MPSK and MQAM over a flat fading channel for SNR range [0: 20] dB. The improvement of our proposed nonlinear LDA algorithm is calculated by subtracting the classification accuracy of the reference paper [2] from the classification accuracy of our proposed algorithm, as shown in Fig. 9.

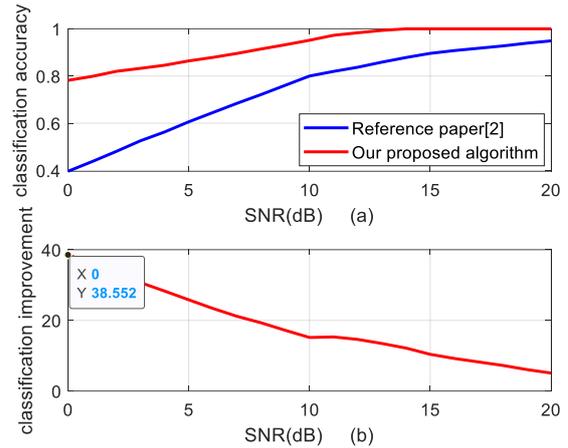


Fig. 9: Comparison between the performance of the reference paper [2] and our proposed algorithm.

As shown in Fig. 9, the classification accuracy of our proposed algorithm is improved compared with the reference paper [2]. The maximum improvement of our proposed algorithm compared with the reference paper [2] is 38.552%.

Conclusion

To improve the classification, an optimized nonlinear preprocess LDA algorithm has been developed. Three optimized functions have been used. These functions have similar performances.

According to Figs. 6, 7, and 8, the proposed preprocess LDA algorithms improve the classification between 8PSK and 16PSK for SNR values greater than 13 dB and between 16QAM and 64QAM for SNR values greater than 4 dB. The proposed-dist LDA algorithm has the best performance for classification between 8PSK and 16PSK. In contrast, the proposed-Tanh LDA algorithm has the best performance for classification between 16QAM and 64QAM. On the other hand, according to Table 7 and Fig. 9, the classification accuracy of our proposed algorithm is improved compared with the reference papers [1], [2]. The maximum improvement of our proposed algorithm compared with the reference paper [1] is 10.79%, and the maximum improvement of our proposed algorithm compared with the reference paper [2] is 38.552%.

By using the proposed optimization algorithm, the AMC classification accuracy has been improved. Other classification problems can use this algorithm. And other nonlinear preprocess functions or optimization algorithms may be found in future work.

Author Contributions

This paper is the result of I. Kadoun’s Ph.d. thesis supervised by H. Khaleghi Bizaki. I. Kadoun and H. Khaleghi Bizaki proposed the main idea of the innovation of the paper. I. Kadoun performed the simulations, carried out the data analysis, interpreted the results and

wrote the manuscript. I. Kadoun and H. Khaleghi Bizaki corrected the proofing the article.

Acknowledgment

The authors would like to thank the editor and anonymous reviewers.

Conflict of Interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Abbreviations

<i>AMC</i>	Automatic Modulation Classification
<i>SNR</i>	Signal-to-Noise Ratio
<i>LDA</i>	Linear Discriminant Analysis
<i>HOCs</i>	Higher-Order Cumulants
<i>MD</i>	Mahalanobis Distance
<i>MPSK</i>	M-array Phase Shift Keying
<i>MQAM</i>	M-array Quadrature Amplitude shift Modulation
<i>LB</i>	Likelihood-Based
<i>FB</i>	Feature-Based
<i>i.i.d</i>	Independent and identically distributed
<i>ACC</i>	Classification accuracy

References

- [1] S. A. Ghauri, I. M. Qureshi, A. Aziz, T. A. Cheema, "Classification of digital modulated signals using linear discriminant analysis on faded channel," *World App. Sci. J.*, 29(10): 1220-1227, 2014.
- [2] O. A. Dobre, A. Abdi, Y. Bar-Ness, W. J. W. P. C. Su, "Cyclostationarity-based modulation classification of linear digital modulations in flat fading channels," *Wirel. Pers. Commun.*, 54(4): 699-717, 2010.
- [3] O. A. Dobre, A. Abdi, Y. Bar-Ness, W. Su, "Survey of automatic modulation classification techniques: classical approaches and new trends," *IET commun.*, 1(2): 137-156, 2007.
- [4] O. A. Dobre, F. Hameed, "Likelihood-based algorithms for linear digital modulation classification in fading channels," in *Proc. 2006 Canadian Conference on Electrical and Computer Engineering: 1347-1350*, 2006.
- [5] A. Hazza, M. Shoaib, S. A. Alshebeili, A. Fahad, "An overview of feature-based methods for digital modulation classification," in *Proc. 2013 1st International Conference on Communications, Signal Processing, and their Applications (ICCSIPA): 1-6*, 2013.
- [6] X.R. Jiang, H. Chen, Y.D. Zhao, W. Q. J. I. J. o. E. Wang, "Automatic modulation recognition based on mixed-type features," *Int. J. Electron.*, 108(1): 105-114, 2021.
- [7] D.-C. Chang and P.-K. J. I. C. Shih, "Cumulants-based modulation classification technique in multipath fading channels," *IET Commun.*, 9(6): 828-835, 2015.
- [8] P. S. Thakur, S. Madan, M. Madan, "Trends in automatic modulation classification for advanced data communication networks," *Int. J. Adv. Res. Comput. Eng. Technol. (IJARCET)*, 4(2): 496-507, 2015.
- [9] Y. Wei, S. Fang, X. Wang, "Automatic modulation classification of digital communication signals using SVM based on hybrid features, cyclostationary, and information entropy," *Entropy*, 21(8): 745, 2019.
- [10] Y. Kumar, M. Sheoran, G. Jajoo, S. K. Yadav, "Automatic modulation classification based on constellation density using deep learning," *IEEE Commun. Lett.*, 24(6): 1275-1278, 2020.
- [11] X. Zhang, J. Sun, X. Zhang, "Automatic modulation classification based on novel feature extraction algorithms," *IEEE Access*, 8: 16362-16371, 2020.
- [12] D. H. Al-Nuaimi, I. A. Hashim, I. S. Zainal Abidin, L. B. Salman, N. A. Mat Isa, "Performance of feature-based techniques for automatic digital modulation recognition and classification—A review," *Electronics*, 8(12): 1407, 2019.
- [13] S. Sobolewski, W. L. Adams, and R. Sankar, "Universal nonhierarchical automatic modulation recognition techniques for distinguishing bandpass modulated waveforms based on signal statistics, cumulant, cyclostationary, multifractal and Fourier-wavelet transforms features," in *2014 IEEE Military Communications Conference: 748-753*, 2014.
- [14] A. Smith, M. Evans, J. Downey, "Modulation classification of satellite communication signals using cumulants and neural networks," in *Proc. 2017 Cognitive Communications for Aerospace Applications Workshop (CCAA): 1-8*, 2017.
- [15] I. Kadoun, H. K. Bizaki, "Advanced features generation algorithm for MPSK and MQAM classification in flat fading channel," *Radioengineering*, 31(1): 127, 2022.
- [16] V. D. Orlic, M. L. Dukic, "Automatic modulation classification algorithm using higher-order cumulants under real-world channel conditions," *IEEE Commun. Lett.*, 13(12): 917-919, 2009.
- [17] A. E. Abdelmutalab, "Learning-based automatic modulation classification," 2015.
- [18] S. A. Ghauri, I. M. Qureshi, A. N. Malik, T. A. Cheema, "Higher order cummulants based digital modulation recognition scheme," *Res. J. Appl. Sci. Eng. Tech.*, 6(20): 3910-3915, 2013.
- [19] A. Abdelmutalab, K. Assaleh, M. J. P. C. El-Tarhuni, "Automatic modulation classification based on high order cumulants and hierarchical polynomial classifiers," *Phys. Commun.*, 21: 10-18, 2016.
- [20] A. Tharwat, T. Gaber, A. Ibrahim, A. E. Hassanien, "Linear discriminant analysis: A detailed tutorial," *AI Commun.*, 30(2): 169-190, 2017.

- [21] B. Ghogh, F. Karray, M. Crowley, "Fisher and kernel Fisher discriminant analysis: Tutorial," arXiv Prepr. arXiv1906.09436, 2019.
- [22] A. Lahav, R. Talmon, Y. J. I. Kluger, I. A. J. o. t. IMA, "Mahalanobis distance informed by clustering," *Inf. Inference A.J. IMA.*, 8(2): 377-406, 2019.
- [23] C. D. Meyer, *Matrix analysis and applied linear algebra*, 71. Siam, 2000.
- [24] G. Strang, "Introduction to linear algebra Fifth," ed: Wellesley-Cambridge Press, 2016.
- [25] A. C. Atkinson, M. Riani, A. J. S. S. Corbellini, "The Box-Cox transformation: Review and extensions," *Stat. Sci.*, 36(2): 239-255, 2021.
- [26] A. H. Namin, K. Leboeuf, R. Muscedere, H. Wu, M. Ahmadi, "Efficient hardware implementation of the hyperbolic tangent sigmoid function," in *Proc. 2009 IEEE International Symposium on Circuits and Systems*: 2117-2120, 2009.
- [27] D. Cai, X. He, J. Han, "SRDA: An efficient algorithm for large-scale discriminant analysis," *IEEE Trans. Knowl. Data Eng.*, 20(1): 1-12, 2007.
- [28] A. Rupp, J. Pelzl, C. Paar, M. Mertens, A. Bogdanov, "A parallel hardware architecture for fast Gaussian elimination over GF (2)," in *Proc. 2006 14th Annual IEEE Symposium on Field-Programmable Custom Computing Machines*: 237-248, 2006.
- [29] L. Zhao, W. Lin, Y. Wang, X. Li, "Recursive local summation of rx detection for hyperspectral image using sliding windows," *Remote Sens.*, 10(1): 103, 2018.
- [30] X. Li, S. Wang, Y. Cai, "Tutorial: Complexity analysis of Singular Value Decomposition and its variants," arXiv paper. arXiv1906.12085, 2019.

Biographies



Iyad Kadoun was born in Damascus, Syria, in 1979 and received his B.S. degree in communication engineering from HIAST in 2002, and his M.Sc. degree in communication engineering from Malek Ashtar University in 2012. His research interests include digital communications.

- Email: idiyad@yahoo.com
- ORCID: [0000-0003-1999-6066](https://orcid.org/0000-0003-1999-6066)
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Homepage: NA



Hossein Khaleghi Bizaki received the Ph.D. degree in electrical engineering and communication systems from Iran University of Science and Technology, Tehran, Iran, in 2008. He is an author or coauthor of more than 100 publications. His research interests include information theory, coding theory, wireless communication, multiple-input-multiple-output systems, space-time processing, and other topics on communication system

and signal processing.

- Email: bizaki@yahoo.com
- ORCID: [0000-0001-9458-8287](https://orcid.org/0000-0001-9458-8287)
- Web of Science Researcher ID: NA
- Scopus Author ID: 23012350800
- Homepage: NA

How to cite this paper:

I. Kadoun, H. Khaleghi Bizaki, "Improving the classification of MPSK and MQAM modulations by using optimized nonlinear preprocess in flat fading channels," *J. Electr. Comput. Eng. Innovations*, 11(1): 141-152, 2023.

DOI: [10.22061/JECEI.2022.8743.550](https://doi.org/10.22061/JECEI.2022.8743.550)

URL: https://jecei.sru.ac.ir/article_1765.html

