# MULTICOLLINEARITY DIAGNOSTICS ON CYCLOSTATIONARY FEATURES

**Anthony Tai, Ph.D.**
Code WXVM
Naval Surface Warfare Center - Crane Division
Crane, IN 47522-5001
anthony.s.tai.civ@us.navy.mil

## ABSTRACT

Cyclostationary features such as high-order moments and cumulants are often applied in detecting and classifying digitally modulated signals. However, these predictor variables may not be mutually uncorrelated, raising the concern that potential multicollinearity may lead to redundant information and significantly affect quality prediction. This report explores machine learning and statistical methods for detecting and mitigating multicollinearity in cyclostationary features. Our study examines features corresponding to synthetic electromagnetic waveforms of nine modulation types over nine levels of signal-to-noise. The empirical results indicate that the Variance Inflation Factor (VIF) method and the Eigendecomposition of the predictor correlation matrix are effective in detecting multicollinearity. Furthermore, employing feature transformation using the Principal Component Analysis (PCA) technique allows us to identify principal components that are less influential in the performance of prediction models.

## 1 INTRODUCTION

Signal classification based on waveform modulation recognition plays an important role in electromagnetic spectrum sensing, interference surveillance, and cognitive radio and radar applications. In target detection and tracking (Butterfield et al., 2016; Bell et al., 2015), a cognitive radar system collects sensor data to produce relevant and actionable information such as pulse descriptor words (PDWs).

A typical PDW may include the captured waveform's carrier frequency, modulation type, signal strength, pulse width, and pulse repetition interval (PRI). In addition to conventional PDW data, cyclostationary features extracted from raw in-phase and quadrature-phase waveforms have also been shown to be useful in modulation recognition tasks (Spooner et al., 2017; O'Shea et al., 2018).

However, strong dependencies between underlying features inherently can mask their individual effects on prediction accuracy, making it difficult to discern critical features from non-essential ones. Moreover, redundant information resulting from these dependencies can lead to degradation in model performance. Finally, data processing over large high-dimensional datasets is computationally expensive. It is therefore imperative to seek and apply effective variable selection methods in signal classification based on cyclostationary features.

In this report, we first describe some properties of cyclostationary signals. This is followed by a discussion on the detection of multicollinearity on cyclostationary features and combating multicollinearity with dimensionality reduction. Lastly, we present some empirical results to illustrate the effectiveness of the proposed diagnostic technique.

## 2 CYCLOSTATIONARY SIGNALS

Cyclostationary signals are signals with periodic statistical parameters such as their mean values, variances, and moments (Spooner, Chad. "Understanding and Using the Statistics of Communication Signals." Cyclostationary Signal Processing, 2022. https://cyclostationary.blog/

`2015/09/28/welcome-to-the-csp-blog/`). As an example, we can consider the autocorrelation function of a digitally modulated signal $x$: $R_x(\tau) = E[x(t+\tau)x(t)]$. If the signal $x$ is wide-sense stationary, that is, $x(t) = x(t+T_0)\ \forall t \in \mathbb{R}$, then its autocorrelation function is periodic with period of cyclostationarity $T_0$: $R_x(t+T_0, \tau) = R_x(t,\tau)\ \forall t \in \mathbb{R}$ (Napolitano, 2014). Since the autocorrelation function is periodic, it can be expressed in Fourier series expansion:

$$R_x(t,\tau) = \sum_{n=-\infty}^{+\infty} R_x^{n/T_0}(\tau) e^{j2\pi(\frac{n}{T_0})t}, \text{ cycle frequencies: } \frac{n}{T_0}, n \in \mathbb{Z}.$$

Correspondingly, the Cyclic Autocorrelation Function (CAF) is defined as

$$R_x^{n/T_0}(\tau) = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} R_x(t,\tau) e^{-j2\pi(\frac{n}{T_0})t}\, dt,$$

and the Spectral Correlation Function (SCF) is given by

$$S_x^{n/T_0}(f) = \int_R R_x^{n/T_0}(\tau) e^{-j2\pi f\tau}\, d\tau.$$

On the other hand, higher order statistics including the temporal and spectral moments and cumulants are described in details in (Spooner and Gardner, 1994). A plot of cyclostationary moments and cumulants associated with various modulation types is given in Figure 1.

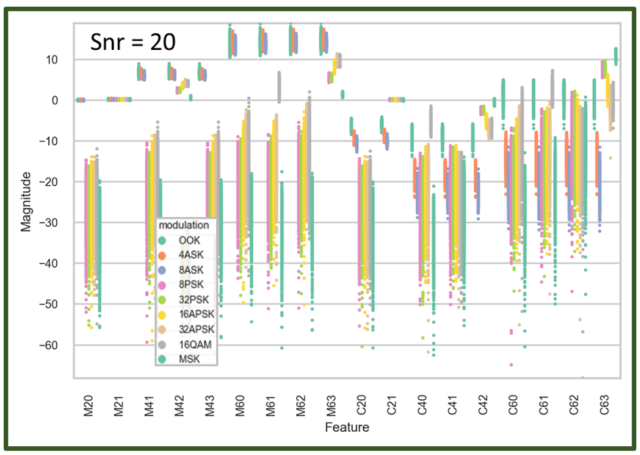


Figure 1: Cyclostationary features associated with various modulation types. Mnm denotes moment of order n with m conjugations, Cnm denotes cumulant of order n with m conjugations.

## 3 Conducting Multicollinearity Diagnostics and Mitigation

Multicollinearity is a situation where linear dependencies occur between multiple variables within a prediction model. Under this condition predictor variables tend to behave similarly. As a result, the effects of individual predictors on the model output become entangled and difficult to control for model improvement.

In this section, we first employ variance inflation factors and eigendecomposition of the correlation matrix of the predictor variables to uncover multicollinearity hidden among cyclostationary features. Then, for mitigation we conduct feature transformation using Principal Component Analysis (PCA) to map the features onto an orthogonal space where the principal components are linearly uncorrelated. This dimensionality reduction method further allows us to prioritize and select features to improve the efficiency in classifying digitally modulated signals.

### 3.1 Detecting Multicollinearity among Cyclostationary Features

In a linear regression problem, the correlation matrix can be computed to determine pairwise correlations between variables. The matrix however may not capture collinearity among multiple variables.

Instead, the quantity variance inflation factor (VIF) (Myers and Myers, 1990) measures the linear relationship of a particular prediction variable with the remaining variables. This makes VIF an effective approach to detect multicollinearity among cyclostationary features.

Consider the matrix $X$ where each column represents the centered and scaled values of a specific cyclostationary feature, its correlation matrix is calculated as $X'X$, where $X'$ denotes the transpose of $X$, with the VIFs as the diagonal elements of $(X'X)^{-1}$. In practice, a valued of VIF greater than 10 would point to potential multicollinearity concern.

Examining the eigenvalues and eigenvectors of the correlation matrix also helps to reveal the existence of multicollinearity. For any given centered and scaled correlation matrix $X'X$, there always exists an orthogonal matrix $V = [v_1, v_2, \cdots, v_m]$ so that $X'X = V\Lambda V'$ where $\Lambda$ is a diagonal matrix $diag(\lambda_1, \lambda_2, \cdots, \lambda_m)$. The sets of $v_i$ and $\lambda_i$ are the eigenvectors and eigenvalues, respectively.

To detect multicollinearity, one could measure the condition number of the correlation matrix $\phi = \frac{\lambda_{max}}{\lambda_{min}}$ where $\lambda_{max} = \max_j \lambda_j$ and $\lambda_{min} = \min_j \lambda_j$. Large values of $\phi$ would indicate high level of multicollinearity. Similarly, individual ratios $\phi_i = \frac{\lambda_{max}}{\lambda_j}$ provide information on the impact of multicollinearity due to individual features.

### 3.2 Combating Multicollinearity with PCA

Principal component analysis (PCA) is often applied in dimensionality reduction and feature extraction (Bishop, 2006; Goodfellow et al., 2016). Through projection onto a linear orthogonal space, PCA produces a set of orthonormal features called principal components each of which a linear combination of the original variables. The principal axes are given by the eigenvectors of $X'X$. The projection of data on these axes are the mutually uncorrelated principal components calculated as $XV$. Correspondingly each eigenvalue $\lambda_i$ gives the variance of its associated eigenvector $v_j$.

To reduce the dimensionality of feature data, we sort the eigenvalues in descending order and select the top values that provide the required total explainable variance. The set of principal components corresponding to these eigenvalues is the desired new feature set.

## 4 Experiment

Synthetic waveforms are generated to demonstrate the detection and mitigation of multicollinearity existing within the set of cyclostationary features. Specifically we evaluate the process over nine modulation types on nine moments and nine cumulants, over nine levels of signal-to-noise.

We first analyze the characteristics of the cyclostationary features based on the waveform's modulation type. Then we examine the resulting correlation matrix and VIF to visually determine the presence of multicollinearity. Finally, we investigate the outcome of the proposed mitigation method for combating multicollinearity.

### 4.1 Experimental Setup

- 9 Modulation types:
  - OOK, MSK, 4ASK, 8ASK, 8PSK, 32PSK, 16APSK, 32APSK, 16QAM
- 18 Cyclostationary features:
  - Moments: M20, M21, M41, M42, M43, M60, M61, M62, M63
  - Cumulants: C20, C21, C40, C41, C42, C60, C61, C62, C63
- 9 Supervised Signal-to-Noise ratio (SNR):
  - (0, ±5, ±10, ±15, ±20)

### 4.2 Cyclostationary Cumulants

The plots that show moments associated with various modulation and feature types are provided in Figure 2

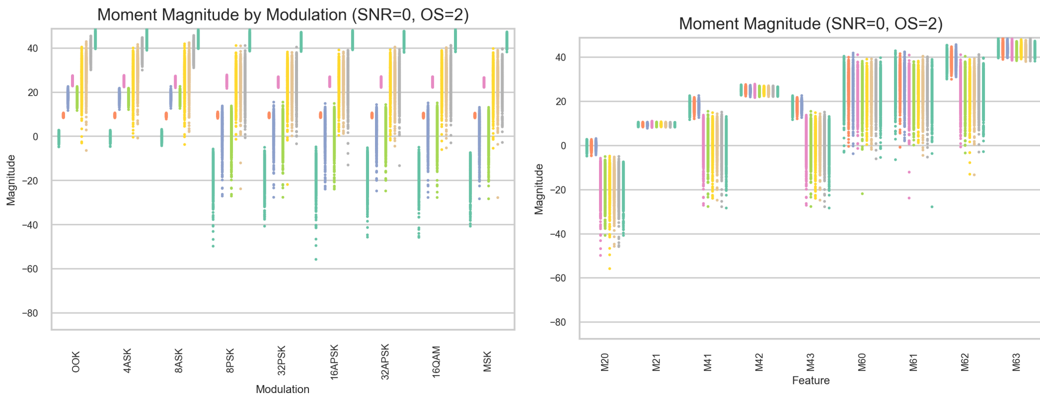


Figure 2: Cyclostationary moments. Left panel: Features associated with various modulation types. Right panel: Features associated with various feature types. Mnm denotes moment of order n with m conjugations, Cnm denotes cumulant of order n with m conjugations.

### 4.3 Cyclostationary Cumulants

The plots that show cumulants associated with various modulation and feature types are provided in Figure 5

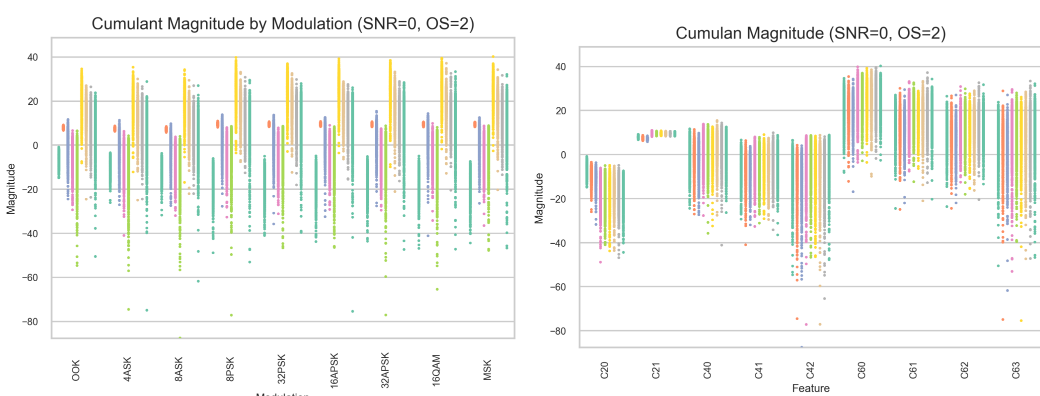


Figure 3: Cyclostationary cumulantss. Left panel: Features associated with various modulation types. Right panel: Features associated with various feature types. Mnm denotes moment of order n with m conjugations, Cnm denotes cumulant of order n with m conjugations.

### 4.4 Feature Paired Plots

A plot that depicts the pairwise correlation between selected features is shown in Figure 4

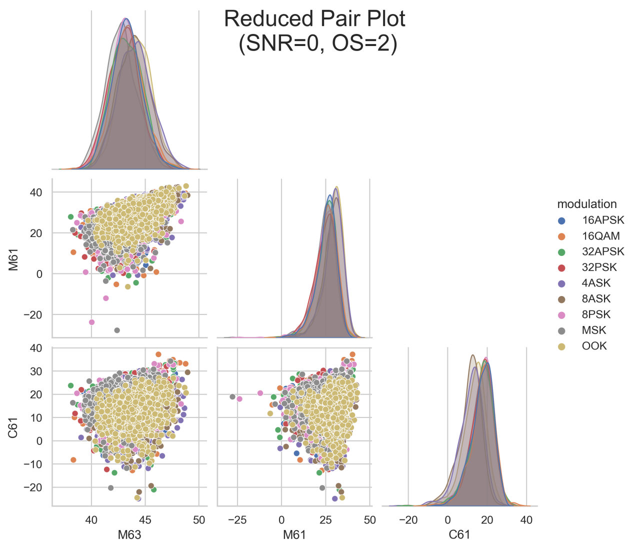


Figure 4: Pairwise correlation between selected moments and cumulant.

### 4.5 Correlation Matrix and Variance Inflation Factor

The proposed detection method reveals collinearity between the cyclostationary features. The figures below provide examples of correlation matrix and variance inflation factor.

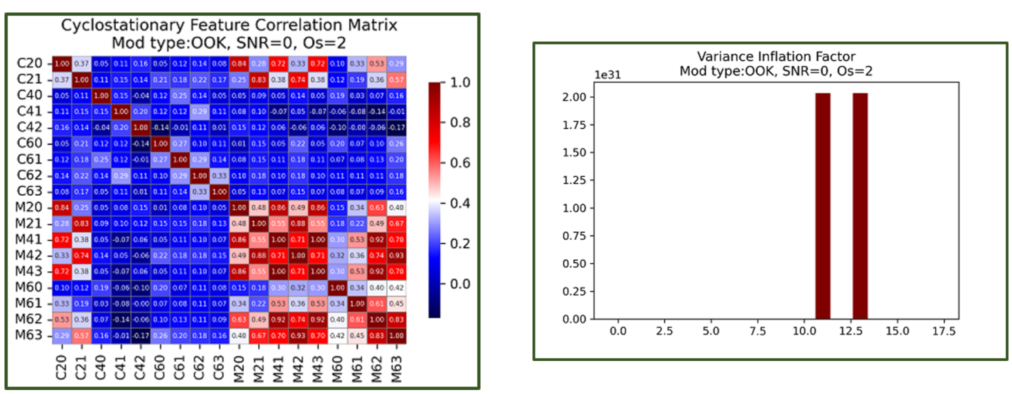


Figure 5: Revealing multicollinearity. Left panel: Correlation matrix between a set of moments and cumulant for the OOK modulation type. Right panel: Variance inflation factor values computed for each cyclostationary feature for the OOK modulation type.

### 4.6 Post-PCA: Explained Variance and Minimum Number of PCs

Figure 6 shows the total and individual variances explained by the principal components with a dotted line indicating the minimum number of principal components needed to meet the required total variance. Higher percentage of explained variance contained in the principal components leads to higher prediction accuracy.

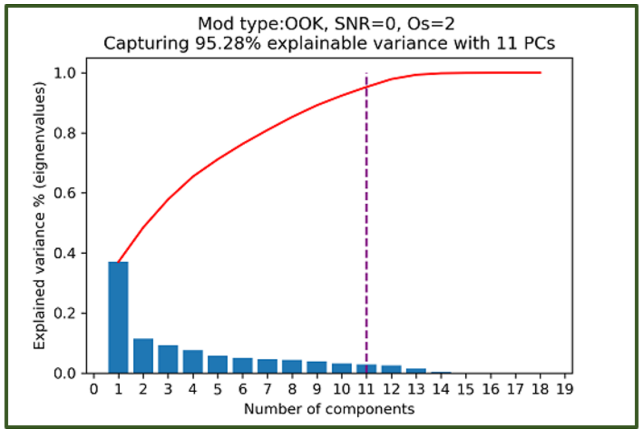


Figure 6: Total and individual variances explained by the principal components. The dotted line shows the minimum number of principal components needed to meet the required total variance.

### 4.7 Minimum Number of PCs – Modulation Type

A plot that depicts the pairwise correlation between selected features is shown in Figure 7

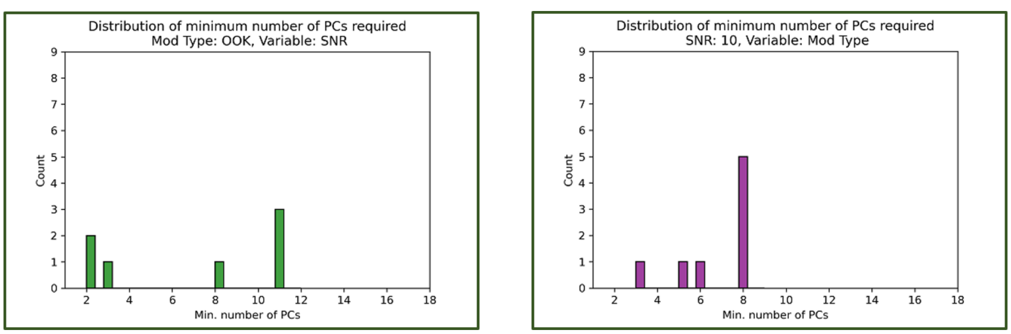


Figure 7: Minimum Number of required principal components. Left panel: based on Modulation Type. Right panel: based on the signal-to-noise ratio.

## 5 Comparative Analysis

Communication waveforms with different modulation types under various noisy conditions naturally exhibit distinct characteristics. The cyclostationary features extracted from these waveforms inherently behave differently.

The following plots compare and contrast results for OOK and MSK modulations.

### 5.1 Correlation Matrices - OOK vs. MSK

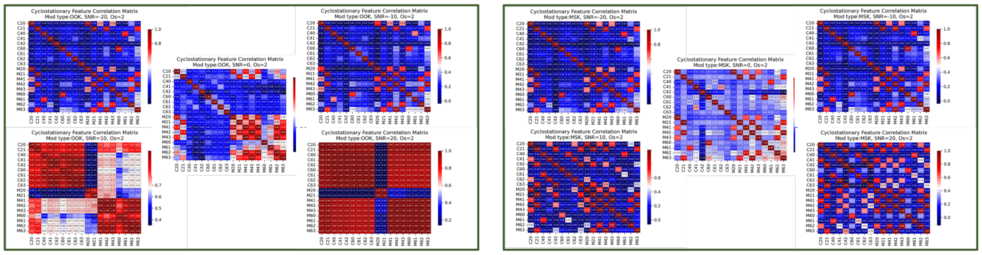

Figure 8: Correlation matrices computed for OOK modulation on the left panel and MSK on the right panel.

### 5.2 Variance Inflation Factor - OOK vs. MSK

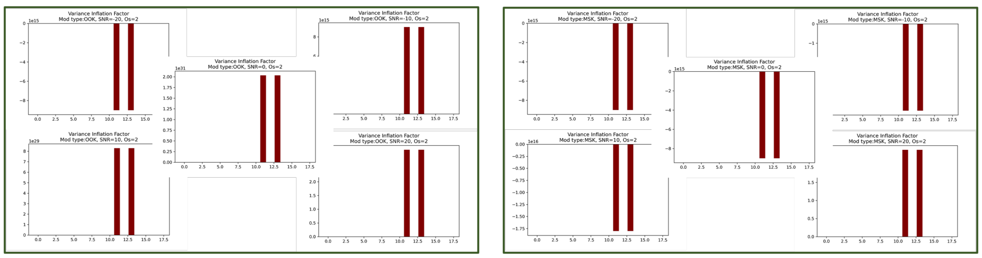

Figure 9: Variance inflation factors computed for OOK modulation on the left panel and MSK on the right panel.

### 5.3 Explained Variance - OOK vs. MSK

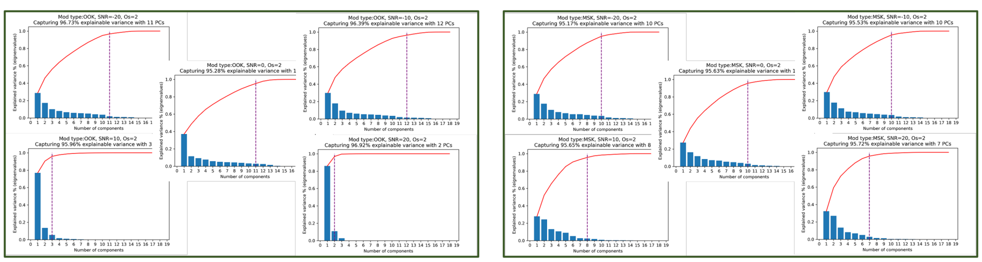

Figure 10: Explained variances computed for OOK modulation on the left panel and MSK on the right panel.

### 5.4 Minimum Number of PCs – Mod Type vs. SNR

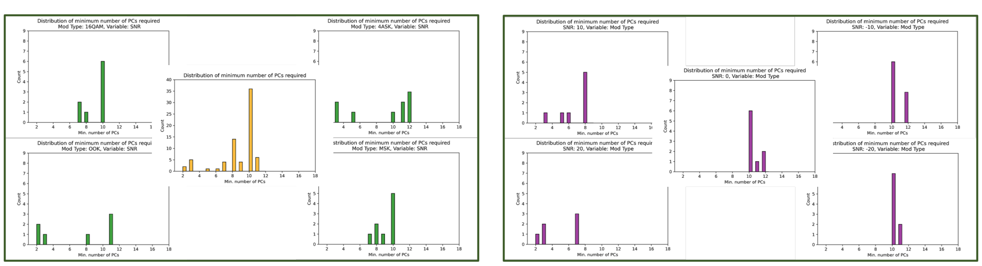

Figure 11: Minimum Number of PCs. Left panel: based on modulation type. Right panel: base on signal-to-noise ratio.

# 6 Discussion and Future Work

In this work, we analyzed cyclostationary moments and cumulants extracted from synthetic communication waveforms. We evaluated correlation matrices and applied the Variance Inflation Factor (VIF) and the Eigendecomposition methods to detect multicollinearity existing among these statistical features. We also employed linear transformation on the feature set using the Principal Component Analysis (PCA) technique to compute new orthonormal features and eliminate data redundancy. This dimensionality reduction method furthermore enabled the selection of the most influential features for downstream classification tasks.

Our empirical results showed that multicollinearity does exist among cyclostationary features and the proposed detection methods were effective in detecting the occurrence and identifying the offenders. In addition, we demonstrated applying PCA for principal component selection to capture desired amount of explained variance.

Future work includes an in-depth research on cyclostationary feature selection over a wide range of waveform modulation types and communication platforms. We would also like to evaluate the impact of feature selection on the performance of various classifiers.

## Acknowledgments

The author gratefully acknowledges the financial support provided by the 2022 Naval Innovative Science & Engineering (NISE) program at the Naval Surface Warfare Center - Crane Division.

## References

Kristine L. Bell, Christopher J. Baker, Graeme E. Smith, Joel T. Johnson, and Muralidhar Rangaswamy. Cognitive Radar Framework for Target Detection and Tracking. *IEEE Journal of Selected Topics in Signal Processing*, 9(8):1427–1439, December 2015. ISSN 1941-0484. doi: 10.1109/JSTSP.2015.2465304. Conference Name: IEEE Journal of Selected Topics in Signal Processing.

Christopher M. Bishop. *Pattern recognition and machine learning*. springer, 2006.

Aaron S. Butterfield, Adam E. Mitchell, Graeme E. Smith, Kristine L. Bell, and Muralidhar Rangaswamy. Metrics for quantifying cognitive radar performance. In *2016 CIE International Conference on Radar (RADAR)*, pages 1–4, October 2016. doi: 10.1109/RADAR.2016.8059298.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

Raymond H Myers and Raymond H Myers. *Classical and modern regression with applications*, volume 2. Duxbury press Belmont, CA, 1990.

Antonio Napolitano. Cyclostationary Signal Processing and its Generalizations, 2014. URL `https://drive.google.com/file/d/1AWk__SK8_hFFD4ERvWIGyXKj6qNd6SNx/view?usp=sharing&usp=embed_facebook`.

Timothy James O'Shea, Tamoghna Roy, and T. Charles Clancy. Over-the-Air Deep Learning Based Radio Signal Classification. *IEEE Journal of Selected Topics in Signal Processing*, 12(1):168–179, February 2018. ISSN 1941-0484. doi: 10.1109/JSTSP.2018.2797022. Conference Name: IEEE Journal of Selected Topics in Signal Processing.

Chad M. Spooner, Apurva N. Mody, Jack Chuang, and Josh Petersen. Modulation recognition using second- and higher-order cyclostationarity. In *2017 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, pages 1–3, March 2017. doi: 10.1109/DySPAN.2017.7920744.

C.M. Spooner and W.A. Gardner. The cumulant theory of cyclostationary time-series. II. Development and applications. *IEEE Transactions on Signal Processing*, 42(12):3409–3429, December 1994. ISSN 1941-0476. doi: 10.1109/78.340776. Conference Name: IEEE Transactions on Signal Processing.