Cyclostationarity Feature-Based Detection and Classification

Amy C. Malady

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
in
Electrical Engineering

A. A. (Louis) Beex, Chair
Tamal Bose
Kathleen Meehan

April 22nd, 2011
Blacksburg, Virginia

Keywords: Automatic Modulation Classification, Detection, Cyclostationarity, Continuous Phase Modulation, Robust Estimation

Cyclostationarity Feature-Based Detection and Classification

Amy C. Malady

ABSTRACT

Cyclostationarity feature-based (C-FB) detection and classification is a large field of research that has promising applications to intelligent receiver design. Cyclostationarity FB classification and detection algorithms have been applied to a breadth of wireless communication signals – analog and digital alike. This thesis reports on an investigation of existing methods of extracting cyclostationarity features and then presents a novel robust solution that reduces SNR requirements, removes the pre-processing task of estimating occupied signal bandwidth, and can achieve classification rates comparable to those achieved by the traditional method while based on only 1/10 of the observation time. Additionally, this thesis documents the development of a novel low order consideration of the cyclostationarity present in Continuous Phase Modulation (CPM) signals, which is more practical than using higher order cyclostationarity.

Results are presented – through MATLAB simulation – that demonstrate the improvements enjoyed by FB classifiers and detectors when using robust methods of estimating cyclostationarity. Additionally, a MATLAB simulation of a CPM C-FB detector confirms that low order C-FB detection of CPM signals is possible. Finally, suggestions for further research and contribution are made at the conclusion of the thesis.

This research was supported by a Bradley Fellowship from Virginia Tech's Bradley Department of Electrical and Computer Engineering and made possible by an endowment from the Harry Lynde Bradley Foundation.

Acknowledgements

First and foremost, my sincerest gratitude and thanks go to my academic advisor Dr. Beex for his continuous support and encouragement throughout the entire research process. It has been a tremendous privilege to work under the guidance of such a talented professor. I am very thankful for all the time he has invested in seeing me succeed and could not have done this without him. I would also like to thank Dr. Meehan and Dr. Bose for their helpful contributions; it is always insightful to see my work from a new prospective. In addition to being wonderful committee members, they are both exceptional professors and have enriched my educational experience here at Virginia Tech. I would also like to thank Dr. Pendleton; her advice has helped me find clarity in times of stress.

I am very thankful for the friends and family who have been there for me to share in the joys and frustrations of graduate school. I don't think I could tell them enough how much it has meant to me that they celebrate my successes. I'd like to thank my lab mates Areg, Arjun, and Roshin for their valuable encouragement and help. I'd especially like to thank my sisters: Christine for paving the road before me, Katherine for being a source of inspiration to me every day, and Julia, who has brought me such happiness.

I'd like to thank Ben Hilburn for his love and support. He has always believed in me, and at times it has helped me believe in myself.

I am eternally indebted to my dad – thank you for reading to me every night when I was young, taking me to the science museums on the weekends, and always being eager to share in my interests. I credit my curiosity and love of math and science to him. He has been the most important influence on my character, and I have looked up to him my whole life. He is my hero.

Contents

List of Abbreviations

1REC	Denotes a full response CPM signal
AM	Amplitude Modulation
BPSK	Binary Phase Shift Keying
CF	Cycle Frequency
CMAD	Complex Median Absolute Deviation from the Mean
CPM	Continuous Phase Modulation
CTCF	Cyclic Temporal Cumulant Function
CTCFE	Cyclic Temporal Cumulant Function Estimate
CTMF	Cyclic Temporal Moment Function
CTMFE	Cyclic Temporal Moment Function Estimate
DSB	Double Sideband
FB	Feature-Based
FM	Frequency Modulation
FSK	Frequency Shift Keying
MAD	Median Absolute Deviation from the Mean

M-PSK	M-ary Phase Shift Keying
M-QAM	M-ary Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
RC	Root Raised Cosine
SNR	Signal to Noise Ratio
SSB	Single Sideband

List of Figures

List of Tables

1 Introduction

1.1 Research Motivation and Objective

Recently, there has been much focus on automatic modulation classification and detection [1-18]. As the requirements for any classifier are high, there have been many proposed solutions. The universal solution to this problem, however, has yet to surface; contributions in the area of automatic modulation classification and detection remain an important area of focus.

Cyclostationarity is a promising feature for signal classification and detection as extracting cyclostationarity features can be done with minimal pre-processing tasks – tasks that often require a priori knowledge of channel characteristics and signal parameters that are unavailable during detection and classification stages [19]. The objective of this work, in particular, is to improve the practicality of using cyclostationarity feature-based detection and classification by alleviating SNR and observation time requirements. Additionally, this work develops a low-order cyclostationarity feature-based detector for CPM signals in an effort to expand the existing realm of application so that it includes the practical possibility of CPM classification based on cyclostationarity.

1.2 Literature Review

Signal detection and signal classification are inherent steps necessary for automatic signal demodulation. Automatic signal detection and classification have inherent civilian and military applications. As technology aims for smarter radios, automatic modulation classification (AMC) has become a significant task for a receiver. Many hypothesize that smart radios capable of dynamic spectrum access could be the solution to the crowded spectrum problem [23-26]; AMC would be an important function of any intelligent, dynamic receiver. In the automatic modulation classification problem the receiver attempts to classify a third-party signal - therefore, there is limited a priori information about the signal available to the receiver. In addition to the typically hostile channel environment, the receiver must also estimate parameters normally only known during friendly communication.

In general, there are two classes of AMC: Likelihood-Based (LB) and Feature-Based (FB). Likelihood-based algorithms form a likelihood ratio that is then compared to a well-determined threshold. Likelihood-based algorithms are optimal in that they minimize the probability of misclassification [1-3]. This method of finding an optimal solution, however, is computationally complex. As a result, it is common to use a suboptimal feature-based algorithm. This work uses the suboptimal feature-based approach to both classification and detection.

Since most modulation schemes are based on systematically controlling the value and variance of the instantaneous signal properties, it is logical to try to detect and monitor these features as a means of signal classification. For example, digital PSK and analog FM signals exhibit constant instantaneous complex amplitude. Conversely, digital ASK and analog AM signals carry information in their fluctuating instantaneous complex amplitude [27]. Similar observations can be made about the zero-crossing rate and instantaneous phase for digital and analog modulation schemes.

There are various popular signal features considered for FB AMC, such as, but not limited to, instantaneous amplitude/phase/zero crossing rate [4-8], high order statistics [9-13], the wavelet transform of the signal [14-16], and cyclostationarity [17-21]. A thorough literature review of current research focuses and developments in AMC is available [1].

A broad division of modulation types would be separation into analog or digital. Within each of these broad categories, there exist numerous specific modulation schemes. While the literature tends to treat analog and digital modulations separately, the ultimate goal of a universal automatic modulation classifier would require an algorithm that simultaneously addresses both schemes. The goal of this work is to develop a reliable classifier that can jointly consider analog and digital signals.

All of the previously published classifiers assume some knowledge of the channel and/or signal parameters. Solutions that minimize assumptions about the channel and the underlying signal are of particular interest, since the end application would either need a method of estimating the assumed parameters or an algorithm that is robust to channel influence and varying signal parameters – such as carrier frequency, phase offset, bandwidth, and timing information.

Cyclostationarity is an attractive feature for classification and detection as estimating cyclostationary features can be successfully done without requiring many of the pre-processing tasks that rely on unavailable channel information [17].

Dobre has previously proposed a FB classifier that uses a decision tree algorithm that exploits cyclostationarity by testing for higher order cyclostationarity at each node [17]. The classifier handles AM, DBS/SSB, BPSK, MPSK, and M-QAM. A highlight of this cyclostationary feature-based algorithm is its robustness to phase, frequency, and timing offset estimation errors.

One advantage of decision tree algorithms is easy expandability. Usually higher-order modulations can be considered by adding higher order statistics as features [9] and new modulation schemes (digital vs analog) can be easily added to a decision tree by adapting another algorithm as an alternative to an early decision block [4]. Redundant tests are avoided by quickly eliminating large sets of possible modulation schemes [4]. The computational effort to extract some features is greater than others, so by streamlining the execution of the algorithm, cumbersome feature extraction can be avoided in some cases.

The joint consideration of analog and digital signals, the relatively low SNR requirements, the limited pre-processing requirements, and easy expandability of Dobre's classifier motivated the choice to use this existing decision tree algorithm as the foundation for the direction of the research pursued in this thesis [17]. For the potential applications of interest, the long observation time and pre-processing assumption on the signal bandwidth are the two most restrictive requirements of the Dobre classifier [17].

Promising results in using robust estimation techniques to improve the performance of cyclostationary feature extraction [28] motivated the decision to investigate the impact of incorporating robust estimation techniques into the Dobre classifier. The ultimate goal of incorporating robust estimation techniques is to reduce the requirements on SNR and/or observation time as – all else being equal – both expand the realm of utility.

The literature on CPM cyclostationarity suggests that only higher-order features are useful for detection and classification [29]. Higher-order feature extraction typically has higher SNR requirements, longer observation time requirements, and increased computational complexity. CPM signals have good power and bandwidth efficiency, and they also exhibit a constant modulus. Because of these properties, CPM is a natural choice when constant modulus is a design criterion, for example, as in peak-power limited systems. CPM is a common modulation scheme in satellite transmission systems [30-32]. Another goal of this research, therefore, is to develop the groundwork towards using low order cyclostationarity features for the eventual goal of joint classification of CPM signals and the existing signals considered in the work by Dobre.

2 Background Material

2.1 Introduction

In this chapter, the technical background material is presented which is important to understanding later results. In particular, important mathematical expressions for the lag product, Cyclic Temporal Moment Function (CTMF), and Cyclic Temporal Cumulant Function (CTCF) as considered in this work will be presented [33]. A statistical test for cyclostationarity is also presented, as this will be used throughout the work [34]. Finally, higher-order statistics of CPM signals are presented as a motivator for developing a detection algorithm that can exploit lower order statistics.

2.2 Cyclostationarity Definitions

The *n*-th order *q*-conjugate lag product of $x(t)$ – a generally complex signal with unit variance – is defined in (2.2.1)

$$L_x(t,\boldsymbol{\tau})_{n,q} = \left(\prod_{j=1}^{n} x^{(*)}(t+\tau_j) \right); \boldsymbol{\tau} = \begin{bmatrix} \tau_1 & \cdots & \tau_n \end{bmatrix} \tag{2.2.1}$$

The symbol (*) denotes optional conjugation, such that there are *q* conjugations in total.

The spectral representation of the lag product is the cyclic temporal moment function $R_x^{\alpha}(\boldsymbol{\tau})_{n,q}$ (CTMF).

$$R_x^{\alpha}(\boldsymbol{\tau})_{n,q} = \lim_{Z\to\infty} \frac{1}{2Z+1} \sum_{t=-Z}^{Z} L_x(t,\boldsymbol{\tau})_{n,q} e^{-j2\pi\alpha t} \tag{2.2.2}$$

A signal is said to exhibit *n*-th order *q*-conjugate cyclostationarity if there exists some pair $\begin{bmatrix} \alpha & \boldsymbol{\tau} \end{bmatrix}$, $\alpha \neq 0$, for which the CTMF produces discrete spectral components [33]. The α corresponding to a non-zero $R_x^{\alpha}(\boldsymbol{\tau})_{n,q}$ is said to be an *n*-th order *q*-conjugate cycle frequency (CF). The CTMF contains impure sine wave contributions, resulting from lower order lag products.

The cyclic temporal cumulant function (CTCF) $C_x^{\beta}(\boldsymbol{\tau})_{n,q}$ considers only pure sine wave contributions, and is conceptually equivalent to the CTMF minus lower order contributions. The CTCF contains contributions from spectral components that result strictly from cyclostationarity introduced by *n*-order *q*-conjugate lag products [33].

$$C_x^{\beta}(\boldsymbol{\tau})_{n,q} = \sum_{\mathbf{P}} \left[(-1)^{p_m - 1}(p_m - 1)! \sum_{\boldsymbol{\alpha}_{p_m}^T 1 = \beta} \prod_{j=1}^{p_m} R_x^{\alpha_j}(\boldsymbol{\tau}_{\mathbf{v}_j})_{\hat{n}_j, \hat{q}_j} \right] \qquad (2.2.3)$$
$$\boldsymbol{\alpha}_{p_m}^T = [\alpha_1; ...; \alpha_{p_m}]$$

$\mathbf{P}$ denotes the collection of unique subsets $\{P_1, P_2, ...P_m\} \in \mathbf{P}$. The subset P_k has p_k elements, or cells, $\{\mathrm{v}_1, ..., \mathrm{v}_{p_k}\}$ such that $\bigcup_{j=1}^{p_k} \mathrm{v}_j = P_k$. The subset P_k is a partition of the index set $I = \{1, 2, ..n\}$ which implies the relationships in (2.2.4).

$$\begin{aligned} &1. \bigcup P_k = I \\ &2. \bigcap \mathrm{v}_j = \varnothing \end{aligned} \qquad (2.2.4)$$

As an example, consider $I = \{1, 2, 3\}$. There are 5 unique subsets such that m = 5 and $\{P_1, P_2, ...P_5\} \in \mathbf{P}$. The subsets, denoted by $(\{\mathrm{v}_1\}, ..., \{\mathrm{v}_{p_k}\})_{P_k}$, are organized in (2.2.5) according to their value of p_k. Let P_1 always be associated with p_1 = 1 and P_m always be associated with $p_m = n$.

$$\begin{aligned} p_1 \quad &= 1 : (\mathbf{v}_1 = \{1,2,3\})_{P_1} \\ p_2, p_3, p_4 &= 2 : \begin{cases} (\mathbf{v}_1 = \{1,2\}, \mathbf{v}_2 = \{3\})_{P_2} \\ (\mathbf{v}_1 = \{1,3\}, \mathbf{v}_2 = \{2\})_{P_3} \\ (\mathbf{v}_1 = \{2,3\}, \mathbf{v}_2 = \{1\})_{P_4} \end{cases} \\ p_5 \quad &= 3 : (\mathbf{v}_1 = \{1\}, \mathbf{v}_2 = \{2\}, \mathbf{v}_3 = \{3\})_{P_5} \end{aligned} \qquad (2.2.5)$$

The cells $\mathbf{v}_j$ index the length n vector $\boldsymbol{\tau} = [\tau_1, ..., \tau_n]$ such that $\boldsymbol{\tau}_{\mathbf{v}_j} = \left[\tau_{\mathbf{v}_j(1)}, ..., \tau_{\mathbf{v}_j(length(\mathbf{v}_j))}\right]$. The order $\hat{n}_j$ of $R_x^{\alpha_j}(\boldsymbol{\tau}_{\mathbf{v}_j})_{\hat{n}_j, \hat{q}_j}$ satisfies $\hat{n}_j = length(v_j) \forall P_k$ such that $\sum_j \hat{n}_j = n$. The order of conjugation $\widehat{q}_j$ satisfies $\sum_j \widehat{q}_j = q$. As an example, consider the delay vector $\boldsymbol{\tau} = [\varepsilon_1\ \varepsilon_2\ \varepsilon_3]$ and suppose we are interested in the third-order zero-conjugate CTCF. Instead of grouping the index, as in (2.2.5), in (2.2.6) the delays are grouped into the appropriate delay vectors $\boldsymbol{\tau}_{\mathbf{v}_j}$.

$$\begin{aligned} p_1 \quad &= 1:\left(\boldsymbol{\tau}_{\mathbf{v}_1} = [\varepsilon_1\ \varepsilon_2\ \varepsilon_3]\right)_{P_1}; \\ p_2, p_3, p_4 &= 2:\begin{cases} \left(\boldsymbol{\tau}_{\mathbf{v}_1} = [\varepsilon_1\ \varepsilon_2]; \boldsymbol{\tau}_{\mathbf{v}_2} = [\varepsilon_3]\right)_{P_2} \\ \left(\boldsymbol{\tau}_{\mathbf{v}_1} = [\varepsilon_1\ \varepsilon_3]; \boldsymbol{\tau}_{\mathbf{v}_2} = [\varepsilon_2]\right)_{P_3} \\ \left(\boldsymbol{\tau}_{\mathbf{v}_1} = [\varepsilon_2\ \varepsilon_3]; \boldsymbol{\tau}_{\mathbf{v}_2} = [\varepsilon_1]\right)_{P_4} \end{cases} \\ p_5 \quad &= 3:\left(\boldsymbol{\tau}_{\mathbf{v}_1} = [\varepsilon_1], \boldsymbol{\tau}_{\mathbf{v}_2} = [\varepsilon_2], \boldsymbol{\tau}_{\mathbf{v}_3} = [\varepsilon_3]\right)_{P_5} \end{aligned} \tag{2.2.6}$$

In (2.2.3), the summation over $\boldsymbol{\alpha}_{p_k}^T 1 = \beta$ indicates all length p_k vectors $\boldsymbol{\alpha}_{p_k}^T = [\alpha_1, \ldots, \alpha_{p_k}]^T$ that satisfy $\boldsymbol{\alpha}_{p_k}^T 1 = \beta$ need to be considered if each member α_j has a nonzero $R_x^{\alpha_j}(\boldsymbol{\tau}_{\mathbf{v}_j})_{\widehat{n}_j, \widehat{q}_j}$ associated with it. There is always only one possible length $p_k = 1$ to be considered which is $\boldsymbol{\alpha}_1^T = [\beta]$. If the order of interest, *n*, is the lowest order of cyclostationarity, then $C_x^{\beta}(\boldsymbol{\tau})_{n,q} = R_x^{\alpha=\beta}(\boldsymbol{\tau})_{n,q}$ since all terms of the CTMF when $\widehat{n}_j < n$ will evaluate to zero, $R_x^{\alpha_j}(\boldsymbol{\tau}_{\mathbf{v}_j})_{\widehat{n}_j < n, \widehat{q}_j} = 0$. The β corresponding to a non-zero $C_x^{\beta}(\boldsymbol{\tau})_{n,q}$ is said to be a pure *n*-th order *q*-conjugate CF.

The previous definitions all rely on continuous, infinite time duration signals. In practice, we have discrete, finite time duration signals. The classical CTMF and CTCF estimators are defined below in (2.2.7) and (2.2.8), respectively [33].

$$\widehat{R}_x^{\alpha}(\boldsymbol{\tau})_{n,q} = \frac{1}{T}\sum_{t=0}^{T} L_{x_n}(t, \boldsymbol{\tau})_{n,q}\, e^{-j2\pi\alpha t} \tag{2.2.7}$$

$$\begin{aligned} C_x^{\beta}(\boldsymbol{\tau})_{n,q} &= \sum_{\mathbf{P}} \left[(-1)^{p_m-1}(p_m-1)! \sum_{\boldsymbol{\alpha}_{p_m}^T 1=\beta} \prod_{j=1}^{p_m} \hat{R}_x^{\alpha_j}(\boldsymbol{\tau}_{v_j})_{\hat{n}_j,\hat{q}_j} \right] \\ \boldsymbol{\alpha}_{p_m}^T &= [\alpha_1;...;\alpha_{p_m}] \end{aligned} \qquad (2.2.8)$$

For a uniformly spaced set of alphas, the operation in (2.2.7) can be evaluated with an FFT. An acceptable resolution (number of bins) can be found by zero padding the lag product.

The ideal CTMF is only non-zero when $\alpha = CF$; the CTMF estimate (CTMFE), however, is nonzero at $\alpha \neq CF$. Although non-zero values in the CTMFE do not necessarily indicate a cycle frequency, we still must consider those contributions when estimating the CTCF (CTCFE) from the CTMFE. Consequently, there are theoretically infinitely many considerations of length p_k vectors $\boldsymbol{\alpha}_{p_k}^T = [\alpha_1,...,\alpha_{p_k}]^T$ that satisfy $\boldsymbol{\alpha}_{p_k}^T 1 = \beta$ for each partition $P_k; k = 2,..,m$, since the corresponding values of $R_x^{\alpha_j}(\boldsymbol{\tau}_{v_j})_{\hat{n}_j<n,\hat{q}_j} \neq 0$ even when α_j is not an order n_j cycle frequency.

Practically, the number of considerations for calculating the CTCFE depends on the number of frequencies of interest when calculating the CTMFE. For example, if the CTMFE is found using an FFT, there are a limited number of frequency bins in the resolution of the CTMFE to consider when calculating the CTCFE. If prior knowledge of the location of lower order cycle frequencies is known, a reasonable approximation of the CTCFE can be found by only considering those values of $R_x^{\alpha_j}(\boldsymbol{\tau}_{v_j})_{\hat{n}_j<n,\hat{q}_j} \neq 0$ that are known to reflect the presence of an actual cycle frequency in the calculation of the CTCFE.

As an example of the impact of finite data set estimation, consider $\left|\widehat{R}_x^{\alpha'}(\tau)_{2,1}\right|$, the magnitude of the CTMFE, estimated from 100,000 samples of an RC pulse shaped QPSK signal vs. when it is estimated from only 1,000 samples, as illustrated in Fig. 2.2.1a – b, respectively.

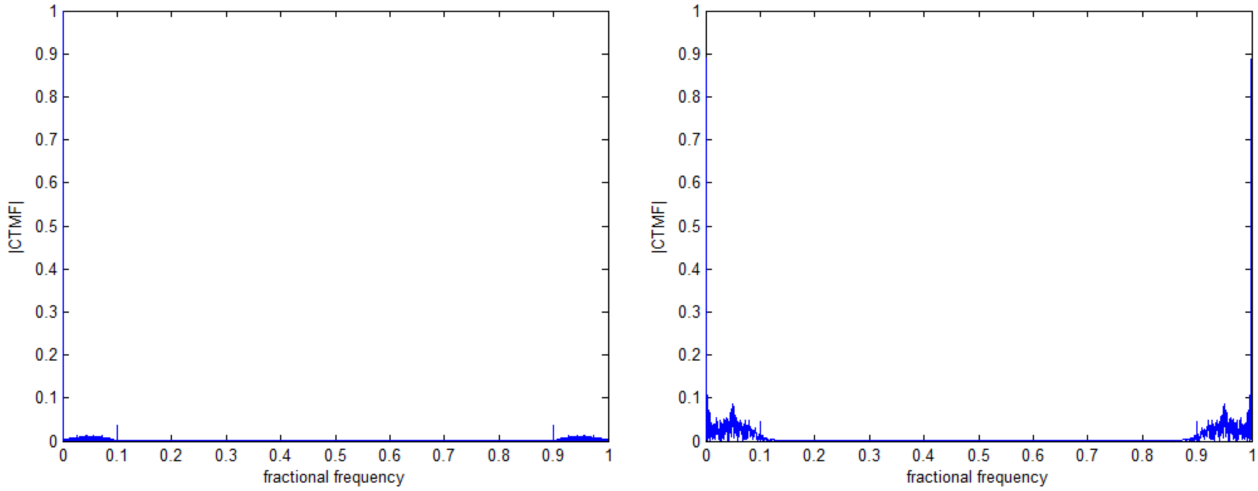


Fig. 2.2.1 a. (left) The second-order first-conjugate CTMFE of an RC pulse shaped QPSK signal based on 100,000 samples; b. (right) the second-order first-conjugate CTMFE of the same RC pulse shaped QPSK signal based on 1000 samples.

A QPSK signal has a cycle-frequency corresponding to the symbol rate [33]. In this example, the cycle frequency is at the fractional frequency 0.1. When the CTMFE is estimated from a large data set, such as 100,000 samples, the cycle frequency in $\left|\widehat{R}_x^{\alpha'}(\tau)_{2,1}\right|$ is prominent. When the CTMFE is estimated from a shorter data set, such as 1,000 samples, the cycle frequency in $\left|\widehat{R}_x^{\alpha'}(\tau)_{2,1}\right|$ is not as prominent. Even when the CTMFE is estimated from a large data set, errors in estimation are significant at lower frequencies, specifically in the fractional frequency range 0 – CF; for example, the cycle frequency is 0.1 in the illustration above.

In addition to finite data length estimation errors, additive noise can introduce further estimation errors. This work specifically considers complex additive white Gaussian noise. As an example of the impact of AWGN, consider $\left|\hat{R}_x^{\alpha'}(\tau)_{2,1}\right|$, the magnitude of the CTMFE, when estimated from the same length 100,000 sample finite data set with additive noise, as shown in Fig. 2.2.2b. The noise-free finite data estimate of $\left|\hat{R}_x^{\alpha'}(\tau)_{2,1}\right|$ is shown again, in Fig. 2.2.2a, for direct comparison. The SNR is 5 dB.

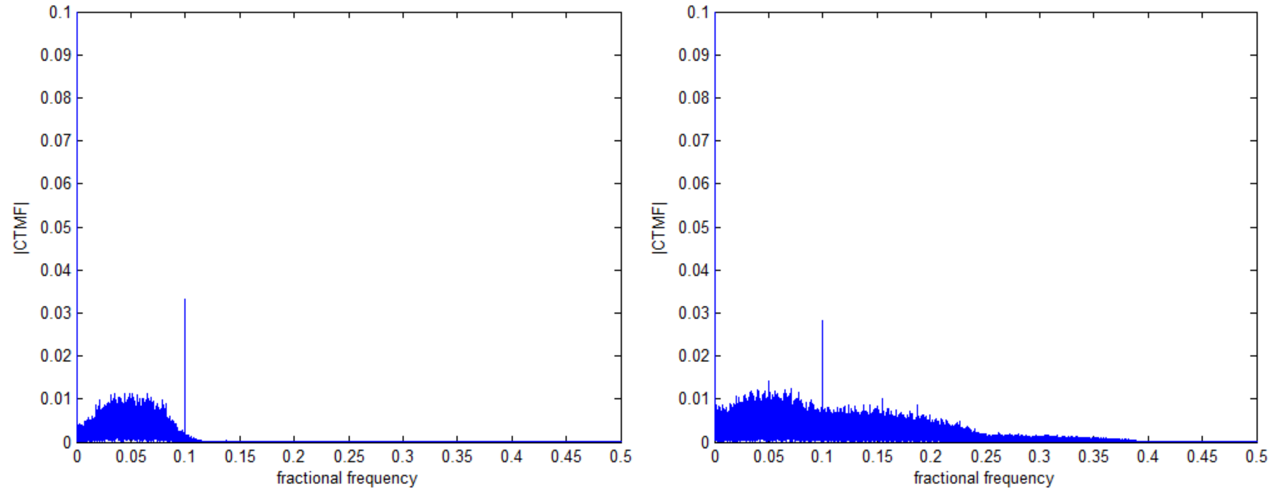


Fig. 2.2.2 a (left) Estimate of the CTMF from a finite, noise-free data set; b.(right) estimate of the CTMF from a finite, noisy data set.

Because of finite data length errors, the general statement that nonzero values of $\left|\hat{R}_x^{\alpha'}(\tau)_{2,1}\right|$ correspond to cycle frequencies does not hold in practice. Rather than only considering the magnitude of the CTMFE as an indicator of the presence of cycle frequencies, a statistical test can be employed [34].

2.3 Statistical Test for Cyclostationarity

A statistical test was proposed [34] for detecting a cycle frequency in the noisy CTMFE. The presence of a cycle frequency is an indication of cyclostationarity. At a particular frequency, a test statistic is calculated and its value compared to a threshold to determine statistical significance. Justification for the test is provided elsewhere [35].

The test procedure to determine the presence or absence of a cycle frequency is as follows. Calculate $\hat{R}_x^{\alpha'}\left(t,\boldsymbol{\tau}\right)_{n,q}$, according to (2.2.7), where α' is a candidate CF. From $\hat{R}_x^{\alpha'}\left(t,\boldsymbol{\tau}\right)_{n,q}$, create the following row vector $\hat{\mathbf{c}}$.

$$\hat{\mathbf{c}} = \left[\mathrm{Re}\left(\hat{R}_x^{\alpha'}\left(\boldsymbol{\tau}\right)_{n,q} \right) \quad \mathrm{Im}\left(\hat{R}_x^{\alpha'}\left(\boldsymbol{\tau}\right)_{n,q} \right) \right] \tag{2.3.1}$$

Next, estimate the unconjugated $\hat{S}_{n,q}^{(T)}$ and conjugated $\hat{S}_{2f}^{(T*)}$ cyclic spectrum.

$$\hat{S}_{n,q}^{(T)} = \frac{1}{TL} \sum_{s=-\frac{(L-1)}{2}}^{s=-\frac{(L-1)}{2}} W(s) \hat{R}_x^{\alpha'-\frac{2\pi s}{T}} \left(\boldsymbol{\tau}\right)_{n,q} \hat{R}_x^{\alpha'+\frac{2\pi s}{T}} \left(\boldsymbol{\tau}\right)_{n,q} \tag{2.3.2}$$

$$\hat{S}_{n,q}^{(T*)} = \frac{1}{TL} \sum_{s=-\frac{(L-1)}{2}}^{s=-\frac{(L-1)}{2}} W(s) \hat{R}_x^{\alpha'+\frac{2\pi s}{T}} \left(\boldsymbol{\tau}\right)_{n,q} \left(\hat{R}_x^{\alpha'+\frac{2\pi s}{T}} \left(\boldsymbol{\tau}\right)_{n,q} \right)^{*} \tag{2.3.3}$$

where $W(s)$ is a window of length L, used for smoothing in the "spectral" domain. A 61-point Kaiser window is used here.

Using (2.3.2) and (2.3.3) assemble covariance matrix estimate $\widehat{\Xi}_{2c}$.

$$\widehat{\Xi}_{n,q} = 0.5\begin{bmatrix} \mathrm{Re}\left(\hat{S}_{n,q}^{(T)} + \hat{S}_{n,q}^{(T*)}\right) & \mathrm{Im}\left(\hat{S}_{n,q}^{(T)} - \hat{S}_{n,q}^{(T*)}\right) \\ \mathrm{Im}\left(\hat{S}_{n,q}^{(T)} + \hat{S}_{n,q}^{(T*)}\right) & \mathrm{Re}\left(\hat{S}_{n,q}^{(T*)} - \hat{S}_{n,q}^{(T)}\right) \end{bmatrix} \tag{2.3.4}$$

The test statistic $\Upsilon_{n,q}$ is then found as follows.

$$\Upsilon_{n,q} = T\hat{\mathbf{c}}\widehat{\Xi}_{n,q}^{-1}\hat{\mathbf{c}}^{T} \tag{2.3.5}$$

At $\alpha \neq CF$, i.e. at non-cycle frequencies, the asymptotic distribution of the test statistic follows a chi-squared distribution with two degrees of freedom. At a cycle frequency, the asymptotic distribution of the test statistic follows a multivariate normal distribution with mean and variance dependent on the distribution of the data set [34].

The threshold for statistical significance is, therefore, designed to provide an acceptable trade-off between a low probability of false alarm and a high probability of detection. The probability of false alarm for a given threshold can be taken from a chi-squared distribution table.

Figure 2.3.1 illustrates one possible realization of the test statistic at each fractional frequency in the resolution of the CTMFE $\widehat{R}_x^{\alpha'}(\tau)_{2,1}$, when estimated from the noise-free finite length 100,000 sample data set, and when the CTMFE $\widehat{R}_x^{\alpha'}(\tau)_{2,1}$ is calculated from a noisy finite length 100,000 sample data set. Again, the signal of interest is an RC pulse shaped QPSK signal, and for the noisy case SNR is 5 dB.

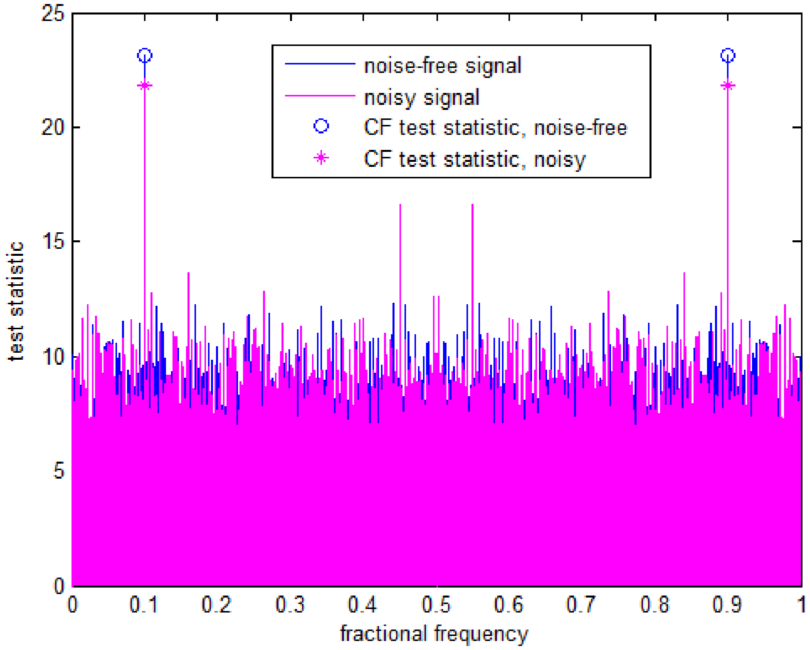


Fig. 2.3.1 Comparison of the behavior of the test statistic when calculated from the CTMFE in the presence and absence of additive noise.

The cycle frequencies occur at the fractional frequencies 0.1 and 0.9 (and 0, which is not shown here), so the test statistics at those two frequencies are expected to be statistically significant. Figure 2.3.1 highlights that the asymptotic distribution of the test statistic at non-cycle frequencies is mostly unaltered by the underlying distribution of the data. While the value of the test statistic at a cycle frequency is slightly less in the presence of noise, the non-cycle frequency test statistics provide little insight as to whether the CTMFE $\hat{R}_x^{\alpha'}(\tau)_{2,1}$ was calculated from a noisy or noise-free signal.

In the classification problem, the statistical test for *n*-order, *q*-conjugate cyclostationarity will be applied to various communication signals that may or may not contain the specific cyclostationarity of interest, but could possibly contain other orders of cyclostationarity. It is important and useful that the behavior of the statistical test at non-cycle frequencies is not – or not much – affected by varying the distribution of the underlying data set of interest.

2.4 CPM Signal Definitions and High Order Cyclostationarity

Much work has been done on using cyclostationarity for classification and detection of common wireless communication signals [17] – [21]. One wireless communication signal that has not been as extensively included in proposed cyclostationarity feature-based solutions to the detection and classification problem is the Continuous Phase Modulation or CPM signal.

CPM is popular for satellite and mobile communications because of its excellent power and bandwidth efficiency [30]-[32]. CPM signals are also constant modulus. While these qualities of CPM are good for communications, they affect the cyclostationarity of the signal in ways that are not ideal for using cyclostationarity in feature-based classifiers and detectors.

The complex baseband CPM signal is expressed below in (2.4.1) [36]

$$s(t) = \exp\left(j\left(2\pi \sum_{k=-\infty}^{\infty} \rho_k h_{k \bmod K} q(t - kT_{sym}) \right)\right) \tag{2.4.1}$$

where T_{sym} is the symbol period, ρ_k is a sequence of equiprobable symbols drawn from the set $\pm 1, 3..(M-1)$, $h_{k \bmod K}$ is the cyclically varying modulation index, and $q(t - kT_{sym})$ is the phase pulse response of the system during the *k*-th symbol period [29]. The modulation indices $\mathbf{h}$ are rational numbers of the form $\mathbf{p}/q$, where $\mathbf{p}$ is a vector of K integers that have no factors in common with q and that all satisfy $p_k < q$. The phase pulse response is typically expressed as an integral of the frequency pulse $g(\upsilon)$.

$$q(t) = \int_0^t g(\upsilon) d\upsilon \tag{2.4.2}$$

The frequency pulse is nonzero only when $0 \le t \le LT_{sym}$ and is normalized so that $g(LT_{sym}) = 0.5$. The pulse duration L refers to the number of symbol periods that is affected by a single symbol; such a signal is referred to as L-REC CPM or partial response CPM when $L > 1$ and full response when $L = 1$. In this work, we restrict the discussion to 1REC CPM. The 1REC CPM frequency baseband pulse is defined in (2.4.3).

$$g(t) = \begin{cases} \dfrac{1}{2T_{sym}} & 0 \le t < T_{sym} \\ 0 & otherwise \end{cases} \tag{2.4.3}$$

Previous work has shown that the lowest order of cyclostationarity, when $\boldsymbol{\tau} = \mathbf{0}$, in a complex CPM signal depends on the choice of modulation indices [29]. For example, single-h CPM signals will exhibit n-order cyclostationarity, with some possible conjugation configuration, when $h\mathbf{1}^{(*)T}\mathbf{1}$ is an integer, where $\mathbf{1}$ is a vector of length n and (*) denotes possible conjugation [29]. Using this property, when $h = \{1/3\}$ the lowest order of cyclostationarity is $n = 3$, whereas when $h = \{12/16\}$, the lowest order of cyclostationarity is $n = 4$. For multi-h CPM signals, the lowest order of cyclostationarity, when $\boldsymbol{\tau} = \mathbf{0}$, is the lowest value of n such that the product $n\mathbf{h}$ returns a vector of whole numbers. Table 2.4.1 lists the lowest order of cyclostationarity found in common CPM configurations [29].

h	Lowest order of cyclostationarity
{4/16 5/16}	16
{7/16 10/16}	16
{12/16 13/16}	16
{1/4}	4
{1/2}	2
{2/9}	9

Table 2.4. 1 Relationship between modulation index and lowest order of cyclostationarity when $\tau = 0$.

To understand the correspondence between modulation indices and lowest order of cyclostationarity, consider the following alternative mathematical representation for 1REC CPM,

$$s(t) = \exp\left(j \left(2\pi\rho_k h_{k \bmod K} \left[\frac{t - kT_{sym}}{2T_{sym}} \right] + \pi \sum_{i=-\infty}^{k-1} \rho_i h_{i \bmod K} \right) \right) \tag{2.4.4}$$

As an example, we now consider $h = \{2/9\}$. From Table 2.4.1, the lowest order of cyclostationarity is n = 9, which implies that the spectrum of the ninth-order lag product will contain discrete spectral lines. The ninth-order lag product is given in (2.4.5).

$$L_s(t, \mathbf{0})_{9,0} = \left[\exp\left(j \left(2\pi\rho_k \frac{2}{9} \left[\frac{t - kT_{sym}}{2T_{sym}} \right] + \pi \sum_{i=-\infty}^{k-1} \rho_i \frac{2}{9} \right) \right) \right]^9 \tag{2.4.5}$$

After algebraic manipulation, recognizing that $\left[\exp(\bullet)\right]^9 = \exp(9\bullet)$, canceling terms, and using the property $\exp(A + B) = \exp(A)\exp(B)$, (2.4.5) becomes (2.4.6).

$$L_s(t, \mathbf{0})_{9,0} = \exp\left(j2\pi\rho_k \left[\frac{t - kT_{sym}}{T_{sym}} \right] \right) \exp\left(j2\pi \sum_{i=-\infty}^{k-1} \rho_i \right) \tag{2.4.6}$$

Since the symbols ρ_i are always expressed as integers, the term $\exp\left(j2\pi\sum_{i=-\infty}^{k-1}\rho_i\right)$ evaluates to 1. As this is the case, the ninth-order lag product of this CPM signal is accurately described as a 2^M-FSK signal with the important property that the frequency deviation is an integer multiple of T_{sym}^{-1}, specifically, $f_d = \rho_i \frac{1}{T_{sym}}$. Another interpretation for this signal is that it is a CPM signal with integer modulation index $h = 2$. Previous work has already shown that even integer h CPM signals exhibit first order cyclostationarity [29].

Also note that for any $n < 9$ in this example, the lag product can be thought of as a new CPM signal with $h = \{2n/9\}$, see (2.4.7).

$$L_s(t,\mathbf{0})_{n,0} = \exp\left(j\left(2\pi\rho_k\frac{2n}{9}\left[\frac{t-kT_{sym}}{2T_{sym}}\right]+\frac{\pi}{T_{sym}}\sum_{i=-\infty}^{k-1}\rho_i\frac{2n}{9}\right)\right) \tag{2.4.7}$$

Although the choice of $h = \{2n/9\}$ may violate selection criteria for a suitable h, the lag product will still be of the same form as for a non-integer h 1REC CPM signal.

Returning to $L_s(t,\mathbf{0})_{9,0}$, suppose that instead of $h = \{2/9\}$, the modulation index is given as $h = \{1/9\}$. The new expression for $L_s(t,\mathbf{0})_{9,0}$ is

$$L_s(t,\mathbf{0})_{9,0} = \exp\left(j\left(2\pi\rho_k\left[\frac{t-kT_{sym}}{2T_{sym}}\right]+\pi\sum_{i=-\infty}^{k-1}\rho_i\right)\right) \tag{2.4.8}$$

In the work showing the properties of cyclostationarity for even integer h CPM, it has been shown that odd integer h CPM also contains first order cyclostationarity at cycle frequencies $\rho_i\frac{f_{sym}}{2f_{sam}}$ [29], as shown in Fig. 2.4.1.

Figures 2.4.1a and 2.4.1b below illustrate the ninth order CMTFE for the cases $h=\{1/9\}$ and $h=\{2/9\}$. The sampling frequency is $f_{sam}=48000$ Hz, and the symbol frequency is $f_{sym}=6000$.

Applying this same analysis to other modulation index choices yields similar conclusions.

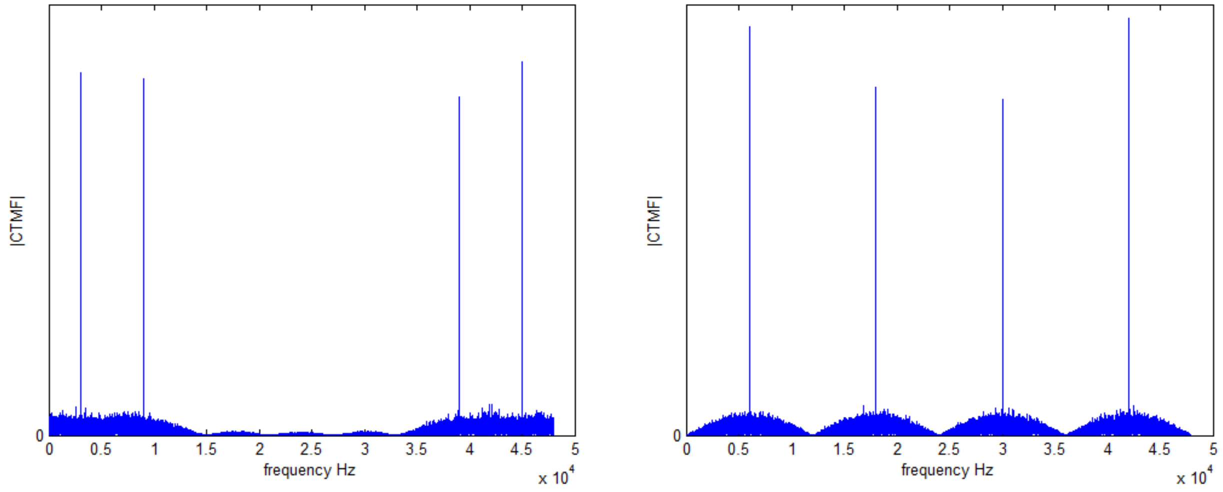


Fig. 2.4.1 a (left) $\hat{R}_s^{\alpha'}(\mathbf{0})_{9,0}$ when $h=\{1/9\}$ and b (left) when $h=\{2/9\}$.

The peaks in the CTMFE are prominent when estimated from a finite data set with no additive noise, which suggests that cyclostationarity may be a useful tool for classifying or detecting CPM signals. In the presence of additive Gaussian noise, however, the peaks become less prominent, which will be investigated further in Chapter 5. Additive noise is a compounding problem when higher orders of the lag product must be considered [33]. Additionally, without knowledge of the modulation index, the use of cyclostationarity features for the detection of CPM requires a search over all possible values of *n*-order cyclostationarity and conjugation configurations. The computational complexity of using cyclostationarity features increases as the number of possible values of $\mathbf{h}$ increases; the computational complexity of finding the lag product increases linearly as n increases. The choice of $\mathbf{h}$ completely determines the lowest order of cyclostationarity, so it is possible that a detector or classifier would need to consider CPM configurations where the lowest order of cyclostationarity is greater than 9, for example.

Considering the lack of a uniform choice of n and q such that $\widehat{R}_s^{\alpha'}(\mathbf{0})_{n,q}$ contains discrete spectral components – together with the compounding error caused by additive Gaussian noise for increasingly larger values of n – motivates the search for, and development of, a feature exhibiting uniform low-order of cyclostationarity such that $\widehat{R}_s^{\alpha'}(\boldsymbol{\tau} \neq \mathbf{0})_{n,q}$ contains discrete spectral components for all possible CPM configurations.

3 Background: Robust Estimation

3.1 Introduction

The purpose of this chapter is to provide background information on maximum likelihood estimators and robust M-estimators. In this chapter, important properties of robust estimators at the distributions of particular interest to the joint analog and digital classifier previously published [17] will be presented.

3.2 Robustness Concepts

In signal processing, only a finite number of samples can be taken to represent an infinite data set. For example, noise in a wireless channel is commonly modeled as zero-mean complex Gaussian, which is to say that given an infinite observation time of a noise channel, the distribution of the noise follows the complex Gaussian [27]. In practice, though, the noise is observed for only a finite – and sometimes relatively short – amount of time. The finite set of N-

observations drawn from an underlying distribution $f(z)$ is referred to as the sample set $z_N = z_1, z_2, \ldots, z_N$ [37].

In some estimation problems, the underlying PDF of a sample set is known, but specific parameters $\vec{\Theta}$ of the PDF are unknown [38]. For example, the parameters of location θ and scale σ are of particular interest in this work. The location parameter of a PDF is a useful description of where a distribution resides in space – where space can be a one dimensional line, a two dimensional plane, or some other higher dimensional surface. The scale of a PDF is a useful description of the spread of a distribution, around its location, in space [38]. The following discussion will be restricted to continuous one dimensional distributions[1].

Common measures of location are the mean and the median. For symmetric distributions, these descriptions of location are mathematically equivalent. The standard measure of the scale of a distribution is the variance. The mean θ_{mean}, median θ_{median}, and variance σ^2 are defined in (3.2.1) – (3.2.3), respectively [38].

$$\theta_{mean} = \int_{-\infty}^{\infty} z f_z(z) dz \tag{3.2.1}$$

$$\int_{-\infty}^{\theta_{median}} f_z(z) dz = \frac{1}{2} \tag{3.2.2}$$

$$\sigma^2 = \int_{-\infty}^{\infty} (z - \theta_{mean})^2 f_z(z) dz \tag{3.2.3}$$

---

[1] The distributions of interest in this work will be two-dimensional (complex), but the distributions of the real and imaginary components will be assumed independently distributed such that each estimate of the two dimensional (real + imaginary) location estimates can be made independently. Estimates of scale will also be scalar estimates, without further concern of how the scalar estimates translate to two dimensional space.

The likelihood functions *L*, as expressed in (3.2.4), indicate how likely it is that an observed sample set came from a certain PDF with the parameters $\hat{\Theta}$ [38].

$$L(z_m;\hat{\Theta}) = \eta \prod_i f_z(z_i;\hat{\Theta}) \tag{3.2.4}$$

The term η is included as a weighting parameter if multiple PDFs are being considered and external information provides insight into how likely it is a certain PDF occurs. For convenience, consider η = 1, which is just to say that all underlying PDF assumptions $f_z(z)$ are equally likely. When using likelihood functions, there is an assumption that the true values of the parameters Θ exist. The goal of the likelihood function is to find the estimate $\hat{\Theta}$ that is most likely to be, or closest to, the true Θ. For example, the Gaussian distribution is completely defined by its mean and variance [33], so if $\hat{\Theta} = \begin{bmatrix} \theta_{mean} & \sigma^2 \end{bmatrix}$, this implies that $\hat{\Theta} = \Theta$.

The maximum likelihood estimate $\hat{\Theta}_{ML}$ of the parameters Θ is the value of $\hat{\Theta}$ that maximizes the likelihood function [38].

$$\hat{\Theta}_{ML} = \arg\max \left\{ L(\vec{z}_m;\hat{\Theta}) \right\} \tag{3.2.5}$$

The log-likelihood estimate is sometimes easier to work with, and is found by taking the (natural) log of the likelihood function; this monotone function preserves the maximization and also changes multiplication operations into additions. The maximization problem can be turned into a minimization problem by multiplying (3.2.5) by negative one.

$$\hat{\Theta}_{ML} = \arg\min\left\{-\sum_m \log\left(f_z(z_i;\hat{\Theta})\right)\right\} \tag{3.2.6}$$

A necessary condition for the solution to (3.2.6) is that the first derivative of the log-likelihood function equals zero, see (3.2.7), at the maximum likelihood estimate $\hat{\Theta}_{ML}$. Because the log-likelihood function may not be strictly convex, (3.2.7) is not a sufficient condition for optimality [38].

$$\frac{d}{d\hat{\Theta}_{ML}}\left(\sum_m -\log\left(f_z(z_i;\hat{\Theta}_{ML})\right)\right) = 0 \tag{3.2.7}$$

Since summation and derivation are linear operations, the order of summation and derivation can be interchanged.

$$\sum_m \frac{d}{d\hat{\Theta}_{ML}}\left(-\log\left(f_z(z_i;\hat{\Theta}_{ML})\right)\right) = 0 \tag{3.2.8}$$

A final step is to apply the chain rule to the derivative in (3.2.8) to eliminate the log operation, resulting in

$$\sum_m \frac{-f_z'(z_i;\hat{\Theta}_{ML})}{f_z(z_i;\hat{\Theta}_{ML})} = 0 \tag{3.2.9}$$

where the notation $f_z'(z_i;\hat{\Theta}_{ML})$ indicates the first derivative with respect to $\hat{\Theta}_{ML}$.

The maximum likelihood parameter vector $\hat{\Theta}_{ML}$ is often found by using iterative searches [38], which is not ideal in time sensitive analysis. An important assumption when using ML-estimators is knowledge of the underlying PDF. In problems like the classification problem, it is normally the case that the underlying PDF could be one of several possibilities [38]. In those cases, the iterative search for $\hat{\Theta}_{ML}$ must be done over multiple PDFs $f_z(z)$. The final classification decision is based on the most likely PDF, from a pool of potential PDFs, whose individual likelihood was found by doing an iterative search. The log-likelihood functions themselves can also be computationally expensive to calculate.

At the Gaussian, where $f_z(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(z-\mu)^2}{2\sigma^2} \right)$, the maximum likelihood estimators $\hat{\Theta}_{ML} = \left[\hat{\theta}_{mean} \quad \hat{\theta}_{\mathrm{var}} \right]$, when the parameters over which (3.2.9) is being optimized are $\Theta = \left[\mu \quad \sigma^2 \right]$, are given by the equations in (3.2.10) [37].

$$\hat{\Theta}_{ML} = \left[\sum_i \left(z_i - \hat{\theta}_{mean} \right) = 0 \quad \sum_i \left(\left(z_i - \hat{\theta}_{mean} \right)^2 - \hat{\theta}_{\mathrm{var}} \right) = 0 \right] \tag{3.2.10}$$

Solving for the estimates themselves, the maximum likelihood estimates of location and scale are the familiar sample mean $\hat{\theta}_{mean} = \frac{1}{N} \sum_i z_i$ and sample variance $\hat{\theta}_{\mathrm{var}} = \frac{1}{N} \sum_i \left(z_i - \hat{\theta}_{mean} \right)^2$. This is to say that, if the underlying PDF is Gaussian, given a sample set, the most likely description of the underlying PDF is $f_z(z) = N\left(\hat{\theta}_{mean}, \hat{\theta}_{\mathrm{var}} \right)$.

In signal processing, an alternative to maximum likelihood classification is feature-based classification [1]. Feature-based classification exploits certain features of known signals of interest, such as cyclostationarity, to make classification decisions. While feature-based classifiers are not attempting to estimate maximum likelihood parameters $\hat{\Theta}_{ML}$ to find the likelihood, or probability, that a sample set was drawn from a certain PDF, feature-based classifiers still use estimates of parameters to calculate features of interest, as is the case for the CTMFE.

Sample sets, regardless of the distribution they were drawn from, have their own parameters of location and scale, referred to as sample estimates. Since the Gaussian distribution is such an important distribution and is often the assumed underlying PDF, the classic estimates of location and scale are the sample mean and sample variance. These estimates are usually used without regard to the underlying distribution. At the Laplacian distribution, for example, the maximum likelihood estimator of location is actually the sample median $\hat{\theta}_{med}$, defined in (3.2.11). Therefore, the sample median is a better estimate of the true mean than the sample mean, at the Laplacian [38].

$$\sum_{i} sign(z_i - z_{med}) = 0\,,\ \hat{\theta}_{med} = z_{med} \qquad (3.2.11)$$

Another consequence of using a sample estimate that may not be the ML-estimate of that parameter, given the PDF of the observed signal, is that the sample estimate is used with the expectation that it is a reflection of the true value; for example, the expectation that the sample mean reflects the true mean. Since there are some distributions for which, for example, the mean does not exist (Cauchy [40]), it is not meaningful to use the sample mean to infer information about the underlying distribution. The efficiency of an estimator is a measure of how accurately the estimator estimates the true value [38]. At the Cauchy, the sample mean has zero accuracy in estimating the true median, and therefore, the sample mean has zero efficiency at the Cauchy.

Another property of interest for estimators is Fisher consistency. An estimator is Fisher consistent if the estimate approaches the true value as the length of the sample set approaches infinity. Furthermore, efficient estimators can be made consistent [38]. A final property of an estimator is its bias. An estimator is biased if the mean of the estimates does not equal the true value.

Robustness concepts arise from the question of what happens to the validity, usefulness, or general behavior of a sample estimate when the underlying assumptions are violated. Huber developed a class of M-estimators as a generalization of the ML-estimators to answer that question [41]. M-estimators also try to estimate parameters of location and scale, but can trade off optimality, in the maximum likelihood sense, for robustness to violated assumptions.

The general form of the M-estimators of location is given below [37],

$$\sum_{n=0}^{N-1} \psi(z_i - \hat{\theta}) = 0 \tag{3.2.12}$$

where $\psi(x)$ is known as the score function. When an M-estimator of location is the ML-estimator of location, at a given PDF, the score function satisfies the relationship given in (3.2.13) as a direct result of (3.2.9)

$$\psi(z_i - \hat{\theta}) = \frac{-f_z\,'(z_i - \hat{\theta})}{f_z(z_i - \hat{\theta})} \tag{3.2.13}$$

For example, based on the definition in (3.2.13), the ML score function of the Gaussian, is $\psi(z_i - \hat{\theta}) = z_i - \hat{\theta}$. This agrees with the previously stated ML estimator derived from the likelihood function in (3.2.9).

Violated assumptions can result from incorrectly assuming an underlying PDF, or from collecting bad samples, referred to as outliers. Outliers can result from noisy samples, as in cases where a sample is collected with bad equipment or in the presence of strange phenomena, or from unlikely samples. In this work, the primary source of outliers is unlikely samples. Consider, for example, a finite sample set drawn from a Gaussian random process. It is possible, though highly unlikely, that a sample in the set could be 5 (or more) standard deviations away from the mean. This unlikely sample is not a "good" indication of the underlying PDF, even though it was drawn from the actual distribution. The influence function $q(x)$ is a measure of the influence of outliers on an estimate [38].

$$q(x) = \frac{\psi(x)}{x} \tag{3.2.14}$$

Robust score functions have influence functions that taper as they move from the origin. Tapering of the influence function reflects the robustness of the score function. The score function and influence function of the sample mean and the sample median are illustrated in Fig. (3.2.1) below. The sample mean gives equal influence to all samples when finding the estimate of location. The sample median, however, highly favors the samples that are close to the center of the underlying distribution. Samples that are farther from the center of the distribution have decreasing influence over the estimate of location.

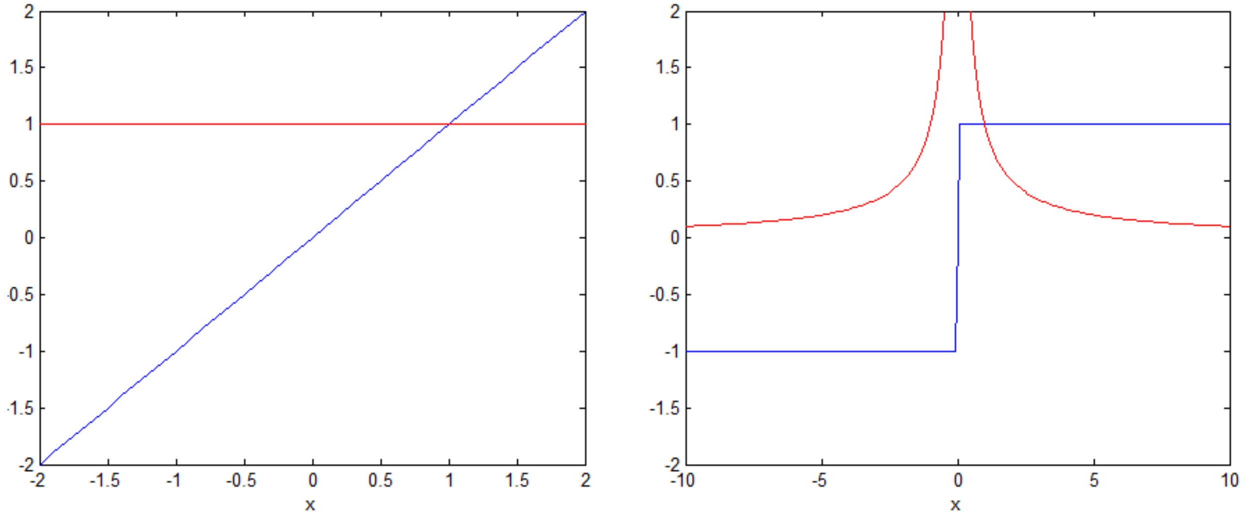


Fig. 3.2.1 Score function (blue) and influence function (red) of sample mean (left) and sample median (right).

As illustrated in Fig. 3.2.1, when calculating the sample mean, extreme outliers have as much influence on the final estimate as well-behaved samples. As sample sets decrease in length, the unbounded score function exacerbates the difficulty of reliable estimation in the presence of outliers.

A bounded, and therefore more robust, score function that forms a tradeoff between the unbounded score function of the sample mean and the highly robust sample median score function is Huber's Score Function [28], defined as in (3.2.15).

$$\psi(x)_{Huber} = \begin{cases} x & for\,|x| \le a \\ a\,\mathrm{sgn}(x) & for\,|x| > a \end{cases} \tag{3.2.15}$$

To simplify notation, we let $x = z_i - \hat{\theta}$. When a is relatively large, Huber's function behaves like the sample mean, but when a is small, Huber's function behaves like the sample median, a more robust estimate of location. The choice of a is a reflection on the likely range of x.

A standard choice for a is 1 when the sample set is normalized by its sample variance [28]. The merit of the Huber estimator is unique to each underlying PDF, but for the work done here the Huber estimator is a good tradeoff between optimality and robustness as performance improvements were achieved using the Huber estimator. The Huber estimator certainly, though, is not the only choice. However, it is the only robust M-estimator considered in this work.

An M-estimator of scale is Huber's "Proposal 2," defined in (3.2.16) [42].

$$\sum_{n=0}^{N-1} \psi\left(\frac{z_n}{\hat{\theta}_{Huber}}\right) \psi\left(\frac{z_n^*}{\hat{\theta}_{Huber}}\right) = Nc \qquad (3.2.16)$$

When the score function is $\psi(x) = x$, and the data set is real valued, the M-estimator of scale in (3.2.16) yields the familiar sample variance estimate, $Nc = \sum_{i=0}^{N-1} \left(\frac{z_i - \hat{\theta}_{mean}}{\hat{\sigma}^2} \right)^2$. The constant c is a term included for Fisher consistency, and equals 1 at the Gaussian distribution, such that (3.2.16) reduces to $\hat{\sigma}^2 = \frac{1}{N} \sum_{i=0}^{N-1} \left(z_i - \hat{\theta}_{mean} \right)^2$, which agrees with the ML estimator of scale, at the Gaussian, as derived in (3.2.10). The constant c is unique for different score functions, and for Huber's function, will vary with a.

One method of solving (3.2.16) is to use an iterative method. As the signals of interest in this work are complex, if a one-step estimate is preferred, for computational reasons, the robust estimate of the Complex Median Absolute Deviation from the median θ_{CMAD} is a suitable one-step estimate θ_{CMAD} of location for complex data.

$$\hat{\theta}_{CMAD} \triangleq \gamma median\,|\mathbf{z}| \tag{3.2.17}$$

The notation $median|\mathbf{z}|$ indicates the sample median of the absolute value of the N samples in $\mathbf{z}$. Recall that the score function for the sample median is $\psi(x) = sign(x)$. The factor γ is included for Fisher consistency.

If the time and resources are available to use an iterative estimate to solve Huber's "Proposal 2", the following estimate can be iterated until it converges. Take the initial estimate $\hat{\theta}_{Huber,k}$ of scale to be the $\hat{\theta}_{CMAD}$.

$$\hat{\theta}^2_{Huber,k+1} = \frac{\hat{\theta}^2_{Huber,k}}{Nc} \sum_{n=0}^{N-1} \psi\left(\frac{z_n}{\hat{\theta}_{Huber,k}}\right) \psi\left(\frac{z_n^*}{\hat{\theta}_{Huber,k}}\right) \tag{3.2.18}$$

3.3 Fisher Consistency of Estimators at Three Distributions of Interest

The efficiency and consistency of an estimator are directly related to the underlying distribution. The signals of interest to the classifier in Chapter 4 are well modeled as one of the following three random processes, Z_1, Z_2, or Z_3, defined below.

$$Z_1 = S + jS + X + jY \tag{3.3.1}$$

$$Z_2 = 2S + X + jY \tag{3.3.2}$$

$$Z_3 = D + X + jY \tag{3.3.3}$$

The random processes $X \sim Y \sim N\left(0, 0.5\sigma_n^2\right)$, where the notation $N\left(0, 0.5\sigma_n^2\right)$ describes a real zero-mean Gaussian random process with variance $0.5\sigma_n^2$. The random process

$S \sim N\left(0, \frac{\sigma_s^2}{2}\right)$. The (potentially complex) discrete random process $D \sim \bar{U}(\mathrm{s})$ is a discrete uniform random process with equiprobable symbols drawn from the vector **s**, such that the variance of D is $\sigma_s^2 = \mathrm{var(s)}$. For $i = 1,2,3$ the random processes Z_i satisfy $\sigma_s^2 + \sigma_n^2 = 1$. For visualization purposes, it may be useful to identify the processes with the communication signals they describe. Z_1 fits the definition of a SSB signal in additive white Gaussian noise, as modeled in [17], Z_2 describes a DSB signal in additive white Gaussian noise, as modeled in [17], and Z_3 describes M-QAM and M-PSK signals in additive Gaussian noise, where the distinction between constellations is described by the vector $\mathbf{s}$ [27]. In Z_1, Z_2, or Z_3, $X + jY$ models the additive white Gaussian noise; the other processes model the communication signal.

The estimators of interest in this work are the sample mean, the θ_{CMAD}, and the scale estimate in (3.2.18), the score function is Huber's function with *a = 1*. For convenience, the definition of the estimate $\hat{\theta}_{CMAD}$ is provided again, (3.3.4).

$$\hat{\theta}_{CMAD} = \gamma \; median|\mathbf{z}| \qquad (3.3.4)$$

An important property of Gaussian random processes is that the mean of a sum of independent Gaussian processes is the sum of the means of the component processes [43]. Additionally, the variance of the sum of independent Gaussian processes is the sum of the variances of the component processes [43].

Z_1 and Z_2 are sums of independent Gaussian random processes, and as such the ML-estimator of location, for sample sets drawn from either process, is the sample mean. The true mean of Z_1 and Z_2 is 0. While not as immediately obvious, Z_3 is also a sum of complex independent Gaussian processes. The number of processes corresponds to the length of the vector **s** and the mean of each individual process corresponds, one to one, to the complex elements of **s**. The mean of Z_3, therefore, is the mean of the complex elements of **s**. Typical digital constellations are symmetric and have zero mean [27]. Under the latter assumption, the mean of Z_3 is zero, even though the mean of each of the Gaussian processes being summed to arrive at Z_3 is not.

Because the distributions of interest are all sums of Gaussians, and the mean of the sums can be found as the sum of the means, the sample mean is an efficient and consistent estimate of location for Z_1, Z_2, and Z_3.

The next estimator to consider is the θ_{CMAD}. Recall the relationship $\sigma_s^2 + \sigma_n^2 = 1$. The Fisher consistency of the θ_{CMAD} at each Z_i will be evaluated for the minimum and maximum values of σ_s^2.

At Z_1, the distribution of the signal $f_{z1}(z)$ is equivalent for $\sigma_s^2 = 0$ and $\sigma_s^2 = 1$ (so $\sigma^2 = 1/2$). The distribution of the real and imaginary components are independent and identical and are given as

$$f_{z1}(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-z^2}{2\sigma^2} \right) \tag{3.3.5}$$

Since the θ_{CMAD} considers the median of the absolute value of **z**, it is important to note that $f_{|z1|}(z)$, the distribution of |**z**|, follows the Rayleigh distribution with variance equal to $\sigma^2 = 1/2$ [43].

$$f_{|z1|}(z) = \frac{z}{\sigma^2}\exp\left(\frac{-z^2}{2\sigma^2}\right) \tag{3.3.6}$$

When the score function $\psi(x) = x$, the estimate θ_{CMAD} at the Gaussian, is the median of the Rayleigh distribution (3.3.7) [43].

$$\int_0^{\theta_{med}} \frac{z}{\sigma^2}\exp\left(\frac{-z^2}{2\sigma^2}\right)dz = \frac{1}{2} \Rightarrow \theta_{med} = \sqrt{\sigma^2 \ln(2)} \tag{3.3.7}$$

Therefore, for the θ_{CMAD} to be a Fisher consistent estimate of the variance, at Z_1, a correction factor of $\gamma = \frac{1}{\sqrt{\ln(2)}}$ is needed in (3.2.17). As the relative ratio σ_s^2 / σ_n^2 fluctuates, the distribution $f_{z1}(z)$ will vary, but will always be well modeled as a circular symmetric complex Gaussian. For this reason, the same correction factor is used independent of σ_s^2 / σ_n^2. Also, if the score function is changed to a Huber function, as long as less than ½ of the sample set is clipped, the θ_{CMAD} will stay consistent with the given correction factor. If over ½ of the sample set is clipped, the θ_{CMAD} will always be equal to the correction factor - for the case of *a* = 1 in (3.2.15).

At Z_2, $f_{z2}(z)$ is equivalent to $f_{z1}(z)$ when $\sigma_s^2 = 0$. When $\sigma_s^2 = 1$, the distribution of Z_2 is given as

$$f_{z2}(z) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2} \right) \tag{3.3.8}$$

Because (3.3.8) is real valued, θ_{CMAD} reduces to the θ_{MAD} – an estimate of the median absolute deviation from the mean. It is well known that at the Gaussian, the θ_{MAD} requires a correction factor of 1.4826 [39]. As the ratio σ_s^2 / σ_n^2 decreases, the distribution of Z_2 moves away from the description in (3.3.8) and closer to that in (3.3.5). Therefore, the necessary correction factor ranges from $\gamma = 1.2011$ (for when the distribution is like that in (3.3.5)) to $\gamma = 1.4826$ (for when the distribution is like that in (3.3.8)). This varying of the necessary correction factor holds for the Huber score function as long as less than half of the sample set is clipped. If more than half of the data set is clipped, the θ_{CMAD} will always equal the correction factor - for the case of $a = 1$ in (3.2.15).

Again, at Z_3, $f_{z3}(z)$ is equivalent to $f_{z1}(z)$ when $\sigma_s^2 = 0$. When $\sigma_s^2 = 1$, the distribution of Z_3 is given as

$$f_{z3}(z) = \frac{1}{m} \sum_{i=1}^{m} \delta(z - s_i) \tag{3.3.9}$$

The elements s_m are generally complex and belong to the length *m*-vector **s**. When **s** describes a PSK constellation, $\hat{\theta}_{CMAD} = \gamma$ since the absolute value of all members of **s** is one. When **s** describes a QAM constellation, $\hat{\theta}_{CMAD} = \gamma \, median|s|$, which in general depends on the signal constellation. In this work, the two constellations of interest are BPSK and QPSK, and $\hat{\theta}_{CMAD} = \gamma$ for both. The value of $\hat{\theta}_{CMAD}$ will fluctuate, therefore, over the range $\gamma\sqrt{\ln(2)} < \hat{\theta}_{CMAD} < \gamma$ in response to the ratio of σ_s^2 / σ_n^2, regardless of the percentage of the sample set that is clipped by the score function.

In the blind classification problem, it is not possible to adjust the correction factor based on knowledge of the signal of interest (as the purpose of the classification is to determine which modulation scheme describes the signal); the correction factor has to be constant. If the correction factor is chosen such that $\gamma = \frac{1}{\sqrt{\ln(2)}}$, and assuming the Huber function clips less than half the data, θ_{CMAD} ranges as described in Table 3.3.1.

SSB Model	$Z_1 = S + jS + X + jY$	$\hat{\theta}_{CMAD} = 1$
DSB Model	$Z_2 = 2S + X + jY$	$0.81 < \hat{\theta}_{CMAD} < 1$
Digital Signal Model	$Z_3 = D + X + jY$	$1 < \hat{\theta}_{CMAD} < 1.2011$

Table 3.3.1 Range of asymptotic estimate of $\hat{\theta}_{CMAD}$ as a function of the underlying signal of interest.

This behavior of $\hat{\theta}_{CMAD}$ will be further explored in Chapter 4.

Because of the application of Huber's "Proposal 2" in this work, it is not of interest to ensure that the robust estimator of scale (3.2.18) is consistent. This will be proven in the next section.

3.4 Robust Estimator of Second-Order First-Conjugate Cyclostationarity and a Corresponding Robust Statistical Test

A robust method of calculating the second-order first-conjugate CTMFE of a signal was proposed by Biedka and Mili [28]. In the latter, the authors exploit the known presence of second-order first-conjugate cyclostationarity in QPSK signals to obtain timing estimates. While the primary intention of the paper was to propose an estimation method that is robust to additive contaminated Gaussian noise, the authors noted that they also saw performance improvements from the robust estimator in additive white Gaussian noise for SNR greater than roughly 0 dB. The conclusion drawn by the authors was that the overall distribution of the data (signal plus noise) is similar to contaminated Gaussian noise, as shown in the previous section, when the signal power and additive Gaussian noise power are roughly equal.

The robust estimate of location in (3.2.18) can be extended to a robust second order-first conjugate CTMFE, with $\tau = \mathbf{0}$, by multiplying the argument of the summation in (3.2.19) by an exponential at a candidate cycle frequency of interest and applying the definition to the data set **x**.

$$\widetilde{R}_x^{\alpha}(\mathbf{0})_{2,1} = \frac{\hat{\theta}^2_{Huber,1}}{Nc} \sum_{n=0}^{N-1} \psi\left(\frac{x_n}{\hat{\theta}_{Huber,1}} \right) \psi\left(\frac{x_n^{\,*}}{\hat{\theta}_{Huber,1}} \right) e^{-j2\pi\alpha n} \qquad (3.4.1)$$

In (3.4.1), $\psi(x) = \psi(x)_{Huber}$ with a = 1. The robust CTMFE $\widetilde{R}_x^{\alpha}(\mathbf{0})_{2,1}$ differs from the classic CTMFE $\hat{R}_x^{\alpha}(\mathbf{0})_{2,1}$ for a = 1 in (3.4.2) only in that the data set is clipped according to the Huber function, and normalized by a robust estimate of scale. For the sake of continuity, the definition of the Huber function is repeated in (3.4.2).

$$\psi(x)_{Huber} = \begin{cases} x & for\,|x| \le a \\ a\,\mathrm{sgn}(x) & for\,|x| > a \end{cases} \tag{3.4.2}$$

Calculating the robust CTMFE is a two step algorithm. First, calculate the robust scale estimate $\hat{\theta}_{Huber,k}$, according to Huber's "Proposal 2," of the data set with the score function $\psi(x) = x$. Normalize the data set by that robust estimate of scale $\hat{\theta}_{Huber,k}$. This normalized data set is then clipped according to the Huber score function before being used to calculate the CTMFE. It is important to normalize the data so that the same value of a in (3.4.2) that specifies the bounds of the Huber function can be used on different data sets with relatively the same effect. If the data is not normalized, the value of a needs to be scaled to reflect non-unit variance in the data set. The two step algorithm in Table 3.4.1 summarizes the steps in calculating the robust CTMFE.

Step 1: Normalize the data set by a complex estimate of scale.	$\hat{\theta}_{Huber,0} = \hat{\theta}_{CMAD}$ $\hat{\theta}^2_{Huber,1} = \frac{\hat{\theta}^2_{Huber,0}}{Nc} \sum_{n=0}^{N-1} \frac{x_n x_n^*}{\hat{\theta}^2_{Huber,0}}$ (3.4.3) $\tilde{x}_n = \frac{x_n}{\hat{\theta}_{Huber,1}}$
Step 2: Apply Huber's score function to $\tilde{x}_n$ to calculate the CTMFE.	$\widetilde{R}_x^{\alpha}(\mathbf{0})_{2,1} = \frac{\hat{\theta}^2_{Huber,1}}{Nc} \sum_{n=0}^{N-1} \psi_{Huber}\left(\tilde{x}_n\right) \psi_{Huber}\left(\tilde{x}_n^*\right) e^{-j2\pi\alpha n}$

Table 3.4.1 Summary of two step algorithm to calculate the robust CTMFE.

Note that in step 1, the score function is $\psi(x) = x$. In step two, the score function is Huber's Score Function $\psi_{Huber}(x)$ and $a = 1$. The term c is included for Fisher consistency, but the consistency of the estimate $\widetilde{R}_x^{\alpha}(\mathbf{0})_{2,1}$ turns out to be of no consequence in this work, as will be shown shortly.

If the CTCFE is not equivalent to the CTMFE, the robust CTCFE $\widetilde{C}_x^{\beta}(\mathbf{0})_{2,1}$ can be found as in the classic case, but with $\hat{R}_x^{\alpha}(\mathbf{0})_{2,1}$ replaced by $\widetilde{R}_x^{\alpha}(\mathbf{0})_{2,1}$, so that

$$\widetilde{C}_x^{\beta}(\mathbf{0})_{2,1} = \sum_{P_2} \left[(-1)^{p-1}(p-1)! \sum_{\boldsymbol{\alpha}^T 1=\beta} \prod_{j=1}^{p} \widetilde{R}_x^{\alpha_j}(\mathbf{0}_{v_j})_{n_j,q_j} \right] \tag{3.4.4}$$

For the distributions of interest in this work, if the data set is zero mean, the second-order first-conjugate CTCFE is equivalent to the CTMFE. The sample mean is an efficient and consistent estimator of the true mean, as was shown in Section 3.3. Therefore, it is well worth the reduction in computational burden to find the CTMFE based on the zero mean sample set. The CTCFE, in that case, is equivalent to the CTMFE.

In the work done here, the iterative solution for the robust estimate of scale is a useful feature when the observation time is decreased to well below 1 second. For a long observation time, for the distributions of interest, there is no noticeable performance gain in using the iterative solution to (3.4.3) vis-à-vis using the sample variance. The one minor design difference is that $\hat{\theta}^2_{Huber,1}$ can be biased such that the estimator of scale is no longer Fisher consistent. Manipulating the estimate of scale increases or reduces the percentage of data being clipped according to the Huber score function in step 2 of the algorithm when calculating the CTMFE.

The effects of biasing the estimate of location, as given in (3.4.3), will be investigated thoroughly later in this section.

To illustrate how the robust estimator improves the CTMFE, consider the following example. The second-order first-conjugate lag product of a BPSK signal has a cycle frequency at the inverse of its symbol period T_{sym}^{-1} Hz, which implies that a BPSK signal exhibits second-order first-conjugate cyclostationarity. Therefore, the second-order first-conjugate CTMF contains a discrete spectral component at T_{sym}^{-1} Hz. Given a sufficient observation time and SNR, the CTMFE will thus contain a discrete spectral component at the appropriate cycle frequency. Figure 3.4.1 illustrates the classic and robust second–order, first-conjugate CTMFEs of a BPSK signal based on 1 second of observation in an AWGN environment with SNR = 0 dB. For this particular signal and sampling frequency, the cycle frequency is $0.125 f_s$.

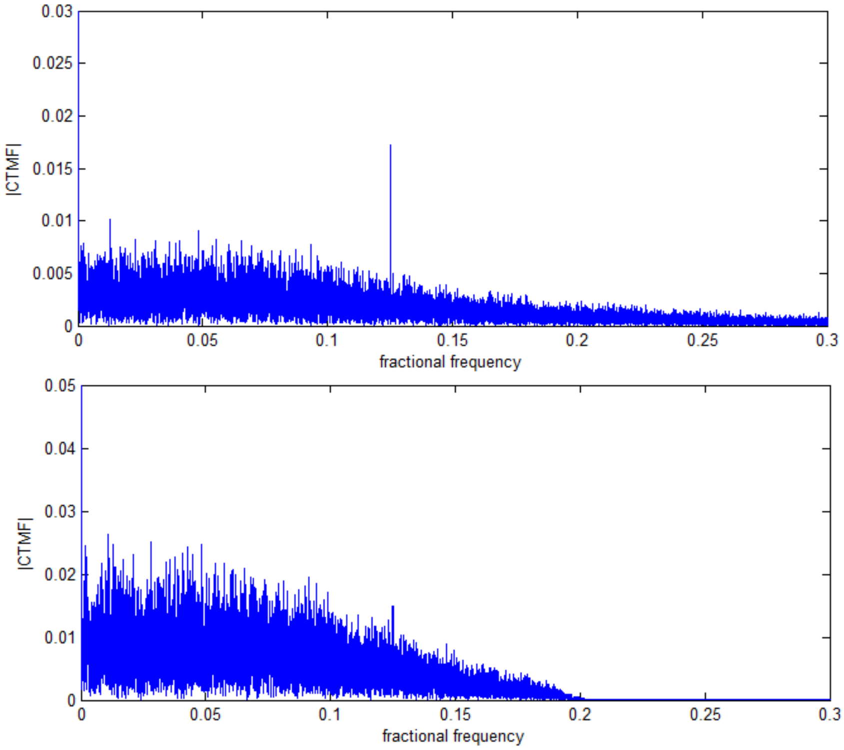


Fig. 3.4.1 Robust (top) vs. classic (bottom) CTMFE of a BPSK signal in a 0 dB additive Gaussian noise environment.

Note that in the classic CTMFE, the cycle frequency peak is buried by the noise floor, while – for the same signal and based on the same data samples – the noise floor is suppressed in the robust CTMFE, and the peak at the cycle frequency is very distinct.

The robust statistical test is the same as for the classic test, with the exception that the robust estimate $\widetilde{R}_x^\alpha(\mathbf{0})_{2,1}$, is used in place of the classic estimate $\hat{R}_x^\alpha(\mathbf{0})_{2,1}$, defined in (2.2.7), throughout the test. The notation '~' will commonly replace the '^' to indicate robust estimates in place of classic estimates, respectively.

Recall that the robust CTMFE $\widetilde{R}_x^\alpha(\mathbf{0})_{2,1}$ is scaled by $\frac{\hat{\theta}^2_{Huber,1}}{Nc}$ for Fisher consistency. As stated before, though, the choice of c_2 does not impact the value of the final test statistic. Consider the following reorganization of $\widetilde{R}_x^\alpha(\mathbf{0})_{2,1}$.

$$\begin{aligned}\tilde{F}_x^\alpha(\mathbf{0})_{2,1} &= \sum_{n=0}^{N-1}\psi\left(\frac{x_n}{\hat{\theta}_{Huber,1}}\right)\psi\left(\frac{x_n^*}{\hat{\theta}_{Huber,1}}\right)e^{-j2\pi\alpha n}\\ \widetilde{R}_x^\alpha(\mathbf{0})_{2,1} &= \frac{\hat{\theta}^2_{Huber,1}}{Nc}\tilde{F}_x^\alpha(\mathbf{0})_{2,1}\end{aligned} \tag{3.4.5}$$

The robust cumulant estimate $\tilde{\mathbf{c}}$ is expressed as follows

$$\begin{aligned}\tilde{\mathbf{c}} &= \frac{\hat{\theta}^2_{Huber,1}}{Nc}[\mathrm{Re}(\tilde{F}_x^\alpha(\mathbf{0})_{2,1}) \quad \mathrm{Im}(\tilde{F}_x^\alpha(\mathbf{0})_{2,1})]\\ &= \frac{\hat{\theta}^2_{Huber,1}}{Nc}\tilde{\mathbf{d}}\end{aligned} \tag{3.4.6}$$

The robust estimates of the conjugated $\tilde{S}_{2,1}^{(T^*)}$ and unconjugated $\tilde{S}_{2,1}^{(T)}$ cyclic spectrum are given below

$$\begin{aligned}\tilde{S}_{2,1}^{(T*)} &= \frac{1}{TL}\left(\frac{\hat{\theta}^2_{Huber,1}}{Nc}\right)^2 \sum_{s=-\frac{(L-1)}{2}}^{s=-\frac{(L-1)}{2}} W(s)\tilde{F}_x^{\alpha'+\frac{2\pi s}{T}}(\mathbf{0})_{2,1}\left(\tilde{F}_x^{\alpha'+\frac{2\pi s}{T}}(\mathbf{0})_{2,1}\right)^* \\ &= \left(\frac{\hat{\theta}^2_{Huber,1}}{Nc}\right)^2 \tilde{P}_{2,1}^{(T*)}\end{aligned} \qquad (3.4.7)$$

$$\begin{aligned}\tilde{S}_{2,1}^{(T)} &= \frac{1}{TL}\left(\frac{\hat{\theta}_{Huber,1}{}^2}{Nc}\right)^2 \sum_{s=-\frac{(L-1)}{2}}^{s=-\frac{(L-1)}{2}} W(s)\tilde{F}_x^{\alpha'-\frac{2\pi s}{T}}(\mathbf{0})_{2,1}\tilde{F}_x^{\alpha'+\frac{2\pi s}{T}}(\mathbf{0})_{2,1} \\ &= \left(\frac{\hat{\theta}_{Huber,1}{}^2}{Nc}\right)^2 \tilde{P}_{2,1}^{(T)}\end{aligned} \qquad (3.4.8)$$

Using (3.4.7) – (3.4.8), the robust covariance estimator $\tilde{\Xi}_{2,1}$ is then

$$\begin{aligned}\tilde{\Xi}_{2,1} &= 0.5\left(\frac{\hat{\theta}^2_{Huber,1}}{Nc}\right)^2 \begin{bmatrix} \mathrm{Re}\left(\tilde{P}_{2,1}^{(T)}+\tilde{P}_{2,1}^{(T*)}\right) & \mathrm{Im}\left(\tilde{P}_{2,1}^{(T)}-\tilde{P}_{2,1}^{(T*)}\right) \\ \mathrm{Im}\left(\tilde{P}_{2,1}^{(T)}+\tilde{P}_{2,1}^{(T*)}\right) & \mathrm{Re}\left(\tilde{P}_{2,1}^{(T*)}-\tilde{P}_{2,1}^{(T)}\right) \end{bmatrix} \\ &= \left(\frac{\hat{\theta}^2_{Huber,1}}{Nc}\right)^2 \tilde{\Theta}_{2,1}\end{aligned} \qquad (3.4.9)$$

Using (3.4.7) – (3.4.9), the robust test statistic $\tilde{\Upsilon}_{2,1}$ is expressed as

$$\tilde{\Upsilon}_{2,1} = T\frac{\hat{\theta}^2_{Huber,1}}{Nc}\tilde{\mathbf{d}}\left(\left(\frac{\hat{\theta}^2_{Huber,1}}{Nc}\right)^2 \tilde{\Theta}_{2,1}\right)^{-1} \frac{\hat{\theta}^2_{Huber,1}}{Nc}\tilde{\mathbf{d}}^T \qquad (3.4.10)$$

Using basic rules of algebra and the property that $\left[kA\right]^{-1} = k^{-1}A^{-1}$, where k is a constant and A is a square matrix, the final test statistic reduces to

$$\tilde{\Upsilon}_{2,1} = T\tilde{\mathbf{d}}\tilde{\Theta}_{2,1}^{-1}\tilde{\mathbf{d}}^T \qquad (3.4.11)$$

The test statistic in (3.4.11) is then compared against a threshold Γ of statistical significance to make a binary decision on the presence or absence of cyclostationarity, just as described in Chapter 2 for the classic case.

From (3.4.11) we observe that the normalization constant $\frac{\hat{\theta}^2_{Huber,1}}{Nc}$ in (3.4.1) has no impact on the final test statistic. Recall from Section 2.3 that the distribution of the test statistic at non-cycle frequencies is chi-squared with two degrees of freedom [34]. Figure 3.4.2 verifies that using robust estimates in the calculation of the test statistic does not alter the distribution of the test statistic at non-cycle frequencies. These results are based on calculating the test statistic at 100,000 non-cycle frequencies using both the robust and classic estimates, and the underlying distribution is SSB, (though the underlying distribution does not impact the distribution at non-cycle frequencies [34]). The robust test statistic at non-cycle frequencies maintains the same distribution as the classic test statistic. This is useful and in a sense indicates that the robust statistical test is consistent with respect to the classic statistical test.

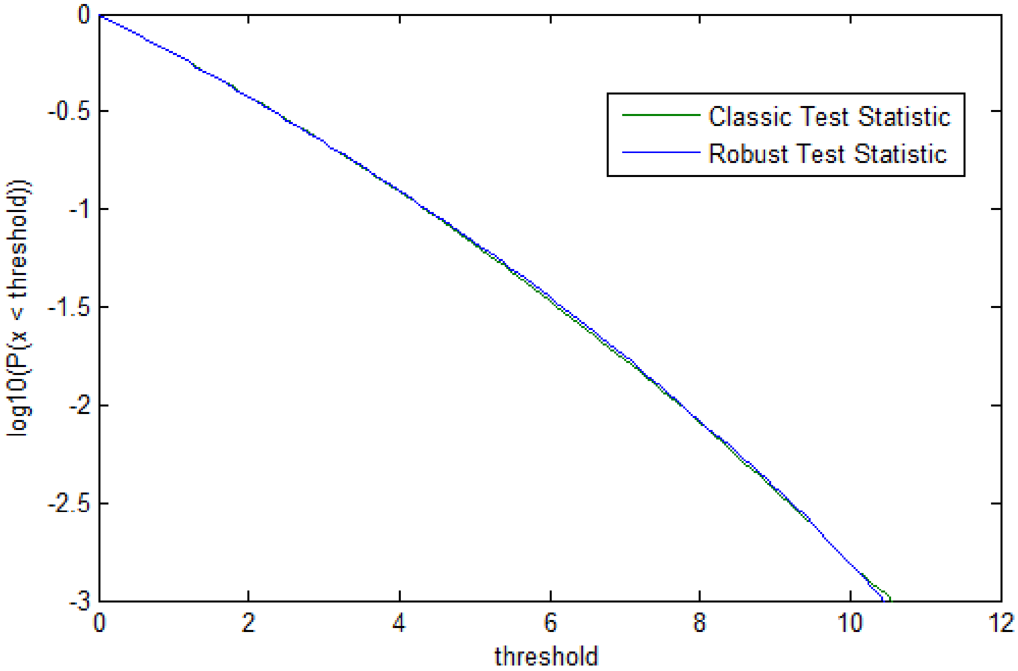


Fig. 3.4.2 Comparison of the distribution of the classic test statistic and the robust test statistic calculated from the same data set at 100,000 non-cycle frequencies.

One last robust measure, which artificially controls the test statistic in a useful way because of the underlying assumptions on the distribution of our signals of interest, is to multiply the robust test statistic in (3.4.11) by $\frac{\hat{\theta}^2_{CMAD}}{c}$, when $\psi(\tilde{x}_n)$ indicates the Huber function with $a=1$ applied at the robustly normalized data set $\tilde{x}_n$.

$$\widehat{\Upsilon}_{2,1} = T\frac{\hat{\theta}^2_{CMAD}}{c}\tilde{\mathbf{d}}\tilde{\Theta}_{2,1}{}^{-1}\tilde{\mathbf{d}}^T \qquad (3.4.12)$$

where c is the same c as that is used in step 1 in Table 3.4.1.

For any distribution, the $\hat{\theta}_{CMAD}$, when the score function is Huber's score function, is upper bounded by γa, where γ is the constant chosen in (3.2.17) and a is the bound on the Huber function, and equal to 1 here, based on suggestions by Mili [28].

The signals of interest to the statistical test that do not contain cyclostationarity are SSB and DSB analog signals. The signals which do exhibit cyclostationarity are digital signals. From Table 3.3.1, the value of the $\hat{\theta}_{CMAD}$ (when $c = 1$) is less than or equal to 1 for the analog signals, but greater than 1 for the digital signals when $\gamma = \frac{1}{\sqrt{\ln(2)}}$, as long as less than ½ of the data set is clipped. Since it is the digital signals that produce test statistics that are larger than a threshold of statistical significance and the analog signals that produce test statistics that are lower than a threshold of significance, the bounds on the $\hat{\theta}_{CMAD}$ are useful scaling values in (3.4.12) for this strategic choice of γ.

The previously derived behavior of the $\hat{\theta}_{CMAD}$ is only valid for $c = 1$. When $c = 1$, the estimate of scale is Fisher consistent and when the sample set is normalized according to the robust estimate of scale, the assumed relationship $\sigma_s^2 + \sigma_n^2 = 1$ holds. If the robust estimate of scale is made inconsistent, i.e. $c \neq 1$, then after scaling by $\hat{\theta}_{Huber,k}$, in general $\sigma_s^2 + \sigma_n^2 = \frac{1}{c}$. Additionally, as c grows larger, the Huber function clips an increasing amount of data. If c grows large enough, the $\hat{\theta}_{CMAD}$ will become equal to 1.2011, for any of the distributions. Because of the nature of the underlying distributions though, $\hat{\theta}_{CMAD}$ will not saturate until c grows much larger than 1.2011. Therefore, the inequality $\frac{\hat{\theta}_{CMAD}^2}{c} < 1$ holds for analog signals even as c grows larger.

Conversely, as c grows smaller, the $\hat{\theta}_{CMAD}$ of the analog signals grows smaller more quickly, since $\tilde{x}_n$ is also scaled by c in step 1 of Table 3.4.1. Figure 3.4.3 verifies the aforementioned behavior of the non-cycle frequency test statistic when it is multiplied by the constant $\frac{\hat{\theta}^2_{CMAD}}{c}$ at the SSB and DSB distributions, for an SNR of 2 dB.

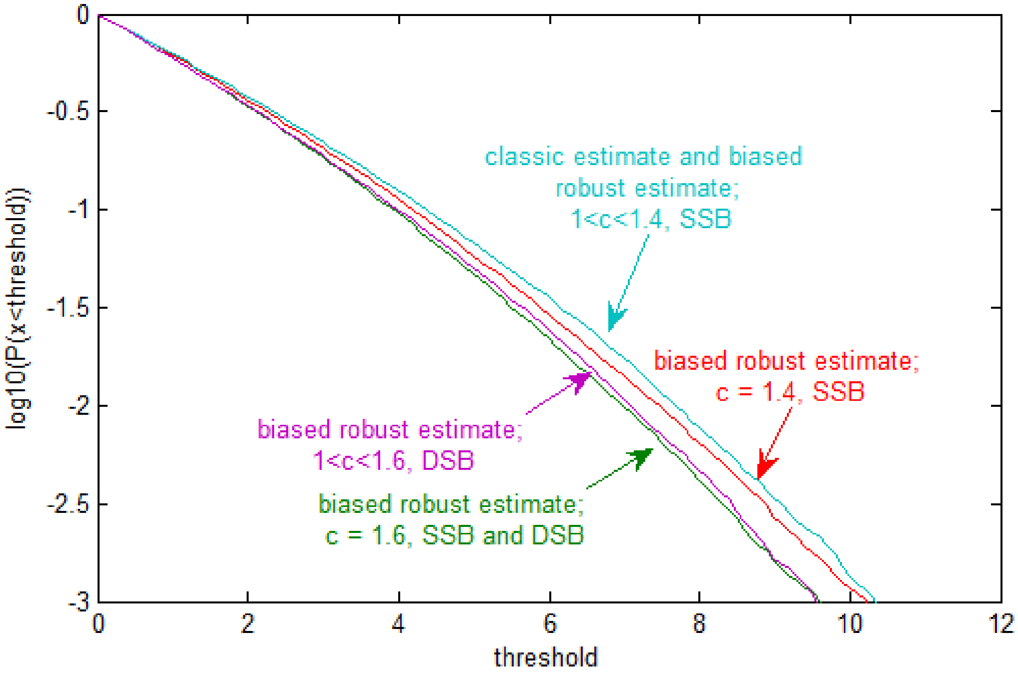


Fig. 3.4.3 Experimental verification of the bounds on the scaling factor. For the range 1<c<1.6, the value of the biased test statistic is never scaled above the classic test statistic. In the DSB case, the biased test statistic is always less than the classic test statistic. In the SSB case, the biased robust test statistic is consistent with the classic test statistic.

Given the behavior of the θ_{CMAD} for digital signals, the inequality $1 \leq \frac{\hat{\theta}_{CMAD}^{2}}{c}$ is valid over the range of $1 \leq c < 1.2011$, where the upper bound on equality is dependent on the SNR. As SNR increases, the upper bound approaches 1.2011.

The value of c is chosen based on the end goal for the desired behavior of the test statistic; such design considerations for the choice of c will be covered in Chapter 4. This artificial control of the test statistic is useful only because of underlying assumptions on the signals of interest and may not be applicable under other assumptions. Exploiting the inconsistency of the $\hat{\theta}_{CMAD}$ is almost a separate test in and of itself for distinguishing digital and analog signals, under these assumptions. The choice to artificially drive the test statistic should be made based on system considerations. In the absence of knowledge of incoming signals, the test statistic should be considered as defined in (3.4.11). The test statistic itself is robust, in a sense, to any knowledge of the underlying distributions of the data in that a Fisher consistent estimate of the $\hat{\theta}_{CMAD}$ is not a requirement. Chapter 4 presents results of the statistical test for cases when an assumption on the distribution of the analog signals can be made, as well as for cases where it cannot be made.

4 Robust Feature-based Cyclostationarity Classifier

4.1 Introduction

Cyclostationarity is a powerful tool for feature-based signal classification [1]. Some published results suggest simultaneous classification of analog and digital signals in complex additive Gaussian noise environments with signal to noise ratios as low as 5 dB [17]. In addition to low SNR requirements, cyclostationarity feature-based classifiers have low pre-processing task requirements.

One algorithm in particular [17], proposed by Dobre, performs blind classification without the pre-processing tasks of timing recovery, carrier phase and frequency offset estimation, and signal and noise power estimation by exploiting the presence or absence of low order cyclostationarity as indicated by the presence or absence of a cycle frequency in the n-order q-conjugate CTMF. The performance of this algorithm does, however, depend on accurate signal bandwidth estimation, as one assumption is that out-of-band noise is blocked.

In this chapter the mathematical concepts laid out in the previous chapter are used to develop a robust cyclostationarity feature-based classifier. The new procedure uses the robust CTMFE in both steps of the previously published algorithm: the candidate cycle frequency is selected based on the behavior of the robust CTMFE, and the statistical test uses robust estimates in place of classic estimates of the cyclic spectrum. The robust estimator in conjunction with the robust statistical test improves cyclostationarity based signal classification by reducing SNR requirements, reducing observation time requirements, and relaxing noise bandwidth assumptions. A specific example is provided that compares the performance of the robust classifier against previously published results, given the same environment. In this example, the robust estimator enjoys a 4 dB improvement in SNR requirements for detecting cyclostationarity, while retaining the same or better tolerable false alarm rates [17]. A portion of the content of this chapter was published in GLOBECOM 2010 [44].

4.2 Comparison of Previously Published Classifier Design and Robust Classifier Design

The previously published classifier is a four node binary decision tree that exploits different orders of cyclostationarity at each node to distinguish various analog and digital signals, as illustrated in Fig. 4.2.1. The classifier assumes that the input is one of the following signals: AM, DSB, SSB, BPSK, QPSK, or some higher order M-PSK/M-QAM signal. Note that the noise-only case is not in the scope of this classifier, so it is assumed that some previous detection algorithm has established the presence of a signal.

Of particular interest here is the performance at nodes three and four. Nodes three and four test second-order first-conjugate cyclostationarity to distinguish DSB from BPSK, and SSB from higher order M-PSK and M-QAM, respectively. Nodes three and four have substantially higher SNR requirements than the previous nodes in the classifier for achieving $P_{cc}^{(l)}$, probability of correct classification, at node *l*, of 100%. The probability of correct classification at nodes three and four, $P_{cc}^{(3)}$ and $P_{cc}^{(4)}$ respectively, is virtually identical, and 100% correct classification requires an SNR of 5 dB, as shown in Fig. 4.2.2.

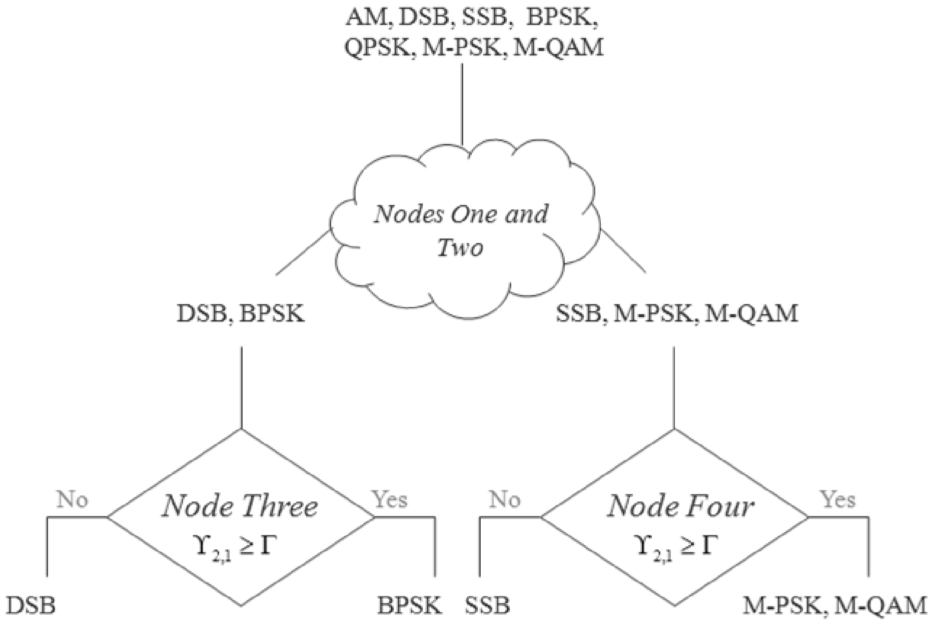


Fig. 4.2.1 Nodes Three and Four from Dobre's work [17]. Both nodes perform the same test, but use knowledge of previous decisions to determine the potential inputs and outputs.

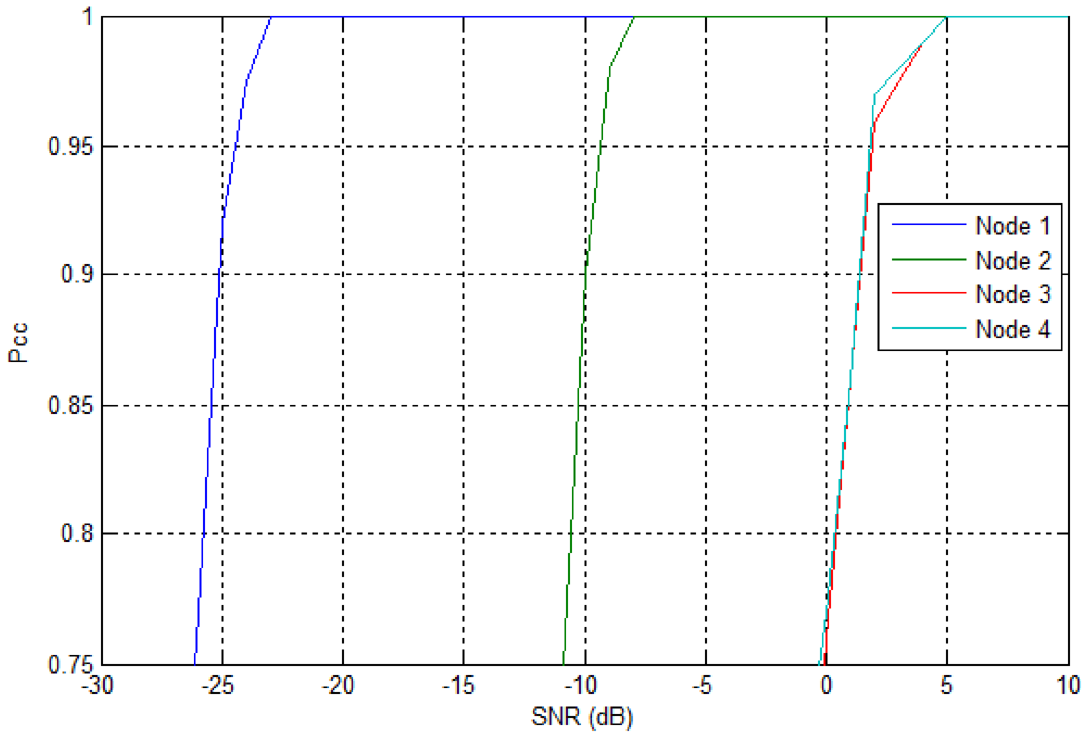


Fig. 4.2.2 Published SNR requirements of each node in the four node binary decision tree [17].

As nodes three and four are the limiting cases, any improvement in SNR requirement at these two nodes causes an overall reduction in SNR requirements for the classifier. The goal of using a robust estimator of the second-order first-conjugate CTMFE is reducing the SNR requirements at nodes three and four.

As suggested by Fig. 4.2.1, each node performs a statistical test to draw conclusions about the presence or absence of cyclostationarity. Specifically, each node executes a two step algorithm to determine the presence or absence of a cycle frequency – an indicator of cyclostationarity. The first step of the test is to calculate the *n*-order *q*-conjugate cyclic temporal moment function and search for a peak. If there is no peak, the decision at the node is that cyclostationarity is not present. If a peak is found, the second step is to calculate a test statistic

$\Upsilon_{n,q}$ at the candidate cycle frequency $\tilde{\alpha}$ that corresponds to the peak in the CTMF, as described in Section 2.3. If the test statistic exceeds Γ, the threshold of statistical significance, the decision at the node is that *n*-order *q*-conjugate cyclostationarity is present. The decision of the presence or absence of cyclostationarity then dictates the flow through the decision tree.

Nodes three and four consider the case of $n = 2$ and $q = 1$. The test statistic $\Upsilon_{2,1}$ is found as in (2.3.5) and the threshold for statistical significance is taken from the chi-squared distribution with two degrees of freedom. Dobre uses a threshold of 13.814, which corresponds to a false alarm rate of 10^{-3} [43].

It is noteworthy that the peak detection algorithm was not explicitly stated in Dobre's work [17]. A search for local maximums by comparing the magnitude of the CTMFE against the average of its neighborhood (the size of the neighborhood depends on the resolution of the CTMF) yields comparable results. If any frequency has a corresponding magnitude of the CTMFE that is substantially larger than the local average, substantially in this case being defined as in excess of 4.25 times the average, it is declared a candidate cycle frequency. The choice of 4.25 was found heuristically based on the best observed tradeoff between detecting peaks and ignoring non-CFs. It is possible to have multiple candidates for a given sample set.

In the robust CTMFE the global maximum is prominent at the cycle frequency, as observed in Fig. 3.4.1, and the noise floor is relatively flat. In the classic CTMFE, the noise floor has shaping that reflects the receive filter shape. The need to search for a local maximum arises from this noise floor shaping. Based on this observation, we note that the search for a peak in the case of the robust CTMFE can be the much simpler search for the global maximum.

Figure 4.2.3 reflects the modifications necessary to form a robust node. The most significant differences between the classic test and the robust test are the identification of candidate cycle frequencies and the use of the robust CTMF.

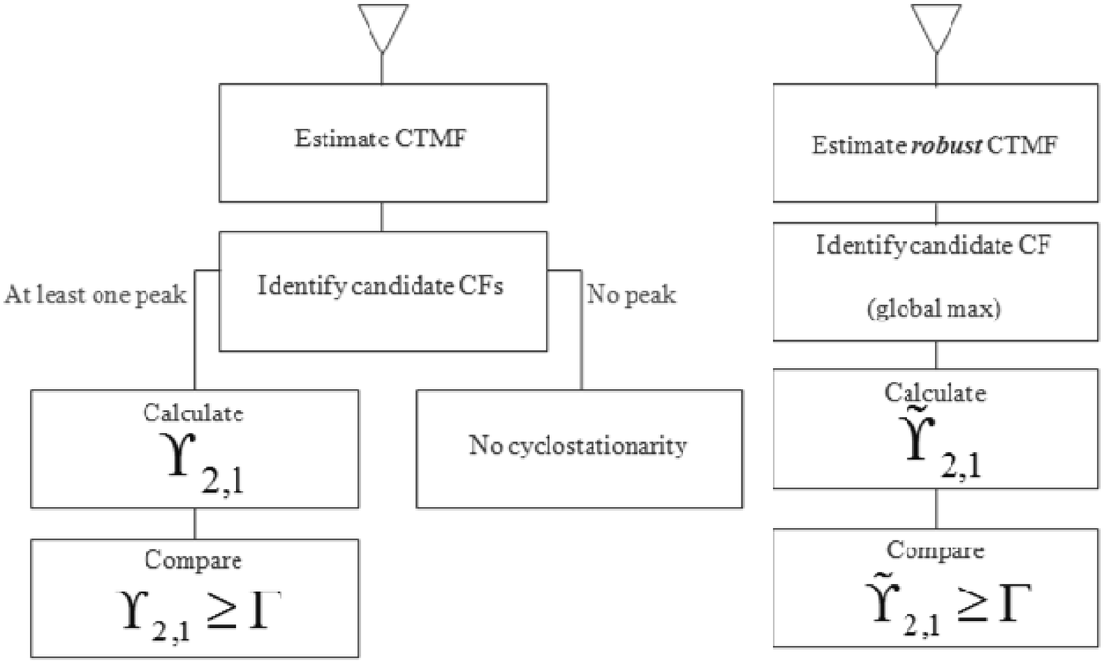


Fig. 4.2.3 Comparison of classic node algorithm vs. robust algorithm.

As stated, the distribution of the test statistic at non-cycle frequencies follows a chi-squared distribution with two degrees of freedom. Using this knowledge of the distribution of the non-cycle frequency test statistic, a fixed threshold can be chosen such that the false alarm rate is fixed. It is important to pay some attention to the definition of Probability of False Alarm (PFA) though, before selecting that threshold.

One interpretation of the distribution of the test statistic at non-cycle frequencies is that given a single sample set, if a test statistic were calculated at every frequency over a continuous spectral range, the distribution of those test statistics would be chi-squared with two degrees of freedom. An equally correct interpretation is that the test statistic associated with a particular non-cycle frequency over an infinite number of sample sets, drawn from the same underlying process, would vary over the ensemble according to the chi-squared distribution with 2 degrees of freedom. For example, a threshold of 13.814 would correspond to a false alarm rate of 10^{-3} [43].

If the data rate of the digital signals of interest were known a priori, such that the statistical test could always be applied at the same candidate cycle frequency, the false alarm rate at that particular frequency would directly translate to the trial false alarm rate. As this information is not assumed to be a priori knowledge the latter definition of false alarms is perhaps not useful. For example, if the resolution of the CTMFE is larger than 1000 candidate cycle frequencies, it is likely that at least one non-cycle frequency will generate a test statistic above 13.814. To increase the odds of identifying a candidate cycle frequency, much larger resolutions of the CTMFE are used, which further increases the number of frequencies to likely generate test statistics above the threshold of significance. As it only takes one non-cycle frequency producing a test statistic above the threshold of statistical significance to trigger a false alarm, if the test were applied at every frequency resolution, it would be highly likely that a CTMFE with over 1000 frequency bins in its resolution would trigger a trial level false alarm.

It is perhaps more useful to consider the false alarm rate on a trial by trial basis. To do this, it is important to know whether there is any relationship between the magnitude of the CTMFE at non-cycle frequencies and the magnitude of the test statistic.

Figure 4.2.4 shows the distribution of the test statistic at a non-cycle frequency corresponding to the largest magnitude member of the CTMFE over 100,000 trials. The CTMFE had 48000 non-cycle frequencies in its resolution. This distribution is not chi-squared, which is evident by the 10^{-3} false alarm rate requiring a threshold greater than 16 instead of the expected 13.814. This indicates that there is some level of correlation between the magnitude of the CTMFE and the magnitude of the test statistic. If the threshold of statistical significance is set to 13.814, the false alarm rate will be over 10^{-1}, which is 100 times the theoretical value predicted by chi-squared tables.

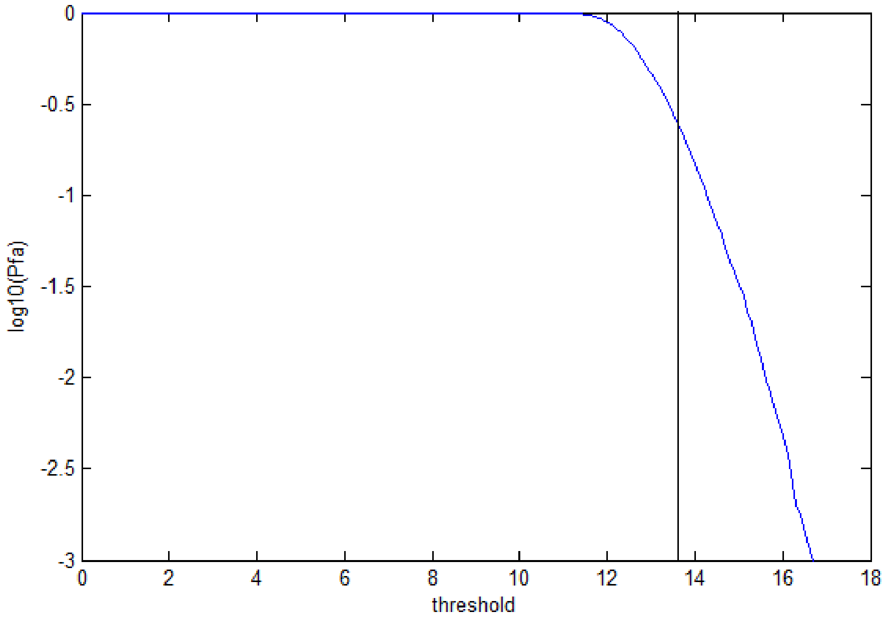


Fig. 4.2.4 The distribution of the test statistic associated with the largest magnitude member of the CTMFE found from 100,000 independent trials. The underlying process does not have second-order cyclostationarity.

Given the relationship between the magnitude of the CTMFE and the magnitude of the test statistic, it is no longer possible to guarantee *trial* false alarm rates that correspond to the chi-squared distribution with two degrees of freedom. The step of searching for a significantly large member of the CTMFE in Dobre's algorithm is an important step to drive down the number of observed false alarms. Even though the CTMFE at a cycle frequency is the global maximum, if the robust test uses the global maximum as the candidate cycle frequency, it can expect to have false alarm rates comparable to the classic false alarm rates. To match the same false alarm rates, the robust test must use a purposefully designed control mechanism.

If no assumptions can be made about the distribution of the signal that does not contain second-order cyclostationarity, the only option to control false alarm rates in the robust estimator is to increase the requirements on declaring a potential candidate cycle frequency. Rather than doing only a global maximum search, an extra requirement can be made on the relative magnitude of the global maximum, as compared to its neighbors. This approach is similar to that in the classic method.

If assumptions can be made on the distribution of the signal that does not contain second-order cyclostationarity, a strategic scaling of the test statistic can be done to drive it below the chosen threshold, as described in Chapter 3. Both techniques of controlling the false alarm rate will be considered in the following sections.

4.3 Results – Classic vs. Robust Estimator Performance

To compare the classification performance of the robust vs. classical estimators, we consider BSPK, QPSK, DSB, and SSB modulations, since these are the signals of interest in the previously published classifier [17]. The (baseband) signals $r(t)$ at the input of the receiving filter are of the form

$$r(t) = \exp\left(j\left(2\pi\Delta f_c t + \varphi\right)\right) s(t) + n(t) \qquad (4.3.1)$$

where Δf_c is the carrier frequency offset and φ is the phase offset. The signal of interest is $s(t)$ and $n(t)$ is uncorrelated complex zero mean Gaussian noise. The entire received signal is then filtered such that the noise is no longer white.

There are four signals of interest. The analog signals, DSB $s_{DSB}(t)$ and SSB $s_{SSB}(t)$ are defined in (4.3.2) and (4.3.3).

$$s_{DSB}(t) = Am(t) \tag{4.3.2}$$

$$s_{SSB}(t) = A\left(m(t) - j\breve{m}(t)\right) \tag{4.3.3}$$

where A is the amplitude of the signal. The message $m(kT)$ was generated by low pass filtering Gaussian random variables, and $\breve{m}(kT)$ denotes the Hilbert transform of $m(kT)$.

The digital signals, BPSK $s_{BPSK}(t)$ and QPSK $s_{QPSK}(t)$ are defined in (4.3.4) and (4.3.5).

$$s_{BPSK}(t) = A\sum_n \cos\left(\pi(1-s_n)\right) p_{TX}\left(t - nT_{sym} - \varepsilon T_{sym}\right) \tag{4.3.4}$$

$$s_{QPSK}(t) = A\sum_n \left(\cos\left(\frac{\pi s_n}{2} + \frac{\pi}{4}\right) + j\sin\left(\frac{\pi s_n}{2} + \frac{\pi}{4}\right)\right) p_{TX}\left(t - nT_{sym} - \varepsilon T_{sym}\right) \tag{4.3.5}$$

where $p_{TX}(t)$ is the pulse shape, ε is the timing error, T_{sym} is the symbol time, and $F_{sym} = \dfrac{1}{T_{sym}}$. The n-th data symbol s_n is randomly selected from the set $\{0,1\}$ for BPSK signals and from $\{0,1,2,3\}$ for QPSK signals, where all possible symbol values are equiprobable. The pulse shape used in this work is an RC pulse, as in (4.3.6), with shaping parameter $\alpha = 0.25$.

$$p_{TX}(t) = \frac{\sin\left(\pi t F_{sym}\right)}{\pi t F_{sym}} \frac{\cos\left(\pi \alpha t F_{sym}\right)}{\left(1-\left(2\pi \alpha t F_{sym}\right)^2\right)} \tag{4.3.6}$$

To facilitate a fair comparison between the robust classifier results and those published in [17], the signal parameters are the same as in [17], $\Delta f_c = 480$ Hz, signal bandwidth is 3 kHz, $F_s = 48$ kHz, and $\varepsilon = 0.8$. The observation interval is one second, which corresponds to 48000 samples.

The SNR is defined at the output of the receiving filter, and the total received signal power is normalized such that the variance of (4.3.7) at the output of the receiving filter is one. The receive filter is a windowed sinc pulse FIR filter. The windowing function is a 242 length Kaiser Window with parameter $\beta = 5.21$. The final signal at the output of the receiver is

$$\hat{r}(t) = \left(\exp\left(j\left(2\pi \Delta f_c t + \varphi\right)\right) s(t) + n(t)\right) * w(t) \tag{4.3.7}$$

where $w(t)$ is the receive filter. Although not specified in the previous work, using a 3 dB noise bandwidth of 0.09 F_S yields similar performance in these trials as what was reported in Dobre's work satisfies the assumption of out-of-band noise cancellation [17].

A thorough investigation of the behavior of the unbiased robust test statistic at non-cycle frequencies was provided in Section 3.4. Additionally, a discussion of the biased robust test statistic was provided for distributions of interest in this work. It is now of interest to verify that the robust estimator yields better classification rates than the classic estimator when using either a biased or unbiased test statistic, for ease of reference, again defined in (4.3.8) and (4.3.9), respectively for our digital and analog signals of interest.

$$\tilde{\Upsilon}_{2,1} = T\tilde{\mathbf{d}}\tilde{\Theta}_{2,1}{}^{-1}\tilde{\mathbf{d}}^{T} \tag{4.3.8}$$

$$\hat{\Upsilon}_{2,1} = T\frac{\hat{\theta}^{2}_{CMAD}}{c}\tilde{\mathbf{d}}\tilde{\Theta}_{2,1}{}^{-1}\tilde{\mathbf{d}}^{T} \tag{4.3.9}$$

Recall that when $c = 1$ the test statistic found in (4.3.9) is still Fisher consistent for a SSB signal at any SNR. For any SNR, the scaling term $\frac{\hat{\theta}^{2}_{CMAD}}{c}$ evaluates to less than 1 for a DSB signal. The consistency of the scaling term depends on the SNR of the DSB signal. At high SNR, the term $\frac{\hat{\theta}^{2}_{CMAD}}{c}$ will evaluate to less than 1 for any value of $1 < c < 1.6$, as was shown in Fig. 3.4.3. For either analog signal, though, neither the choice of c nor the SNR causes the biased test statistic to increase the value of the test statistic (beyond maybe 0.01 or 0.02%, which arises from errors in estimating the $\hat{\theta}^{2}_{CMAD}$ from finite sample sets).

Figure 4.3.1 illustrates the impact of the choice of c on the probability of a trial false alarm, when the analog signals follow the assumptions made in the previously published classifier and the test statistic is calculated at the global maximum. The data represents the results of 10,000 independent trials. For reference, the distribution using the unbiased robust test statistic and the classic test statistic are overlaid – each of which are also found based on the candidate cycle frequency corresponding to the global maximum of the CTMFE.

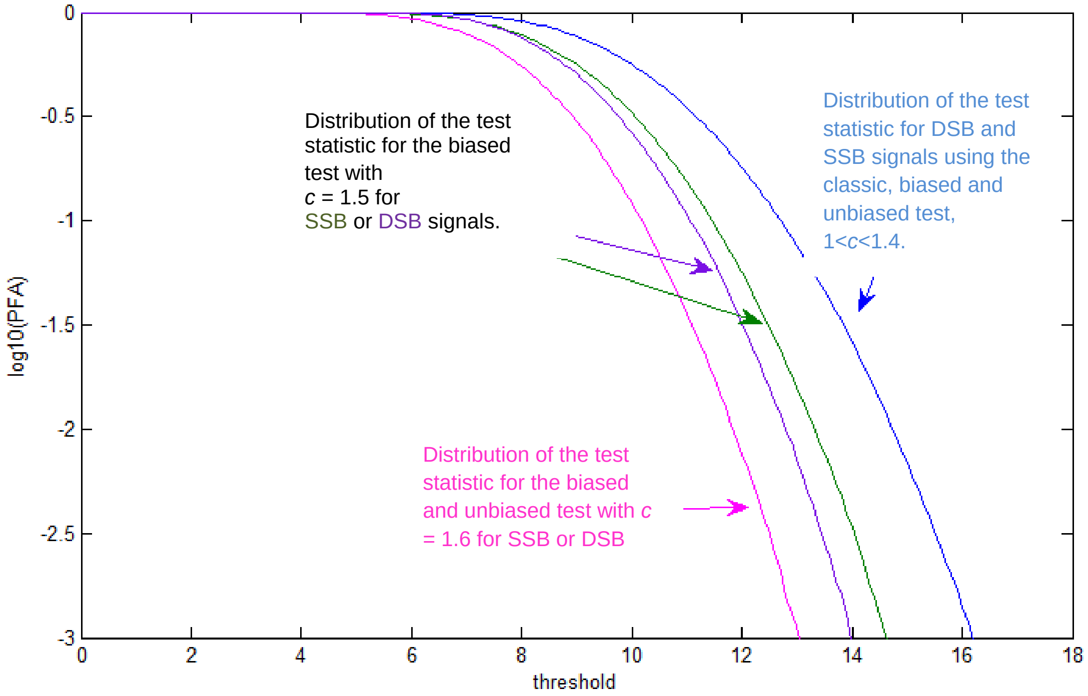


Fig. 4.3.1 Illustration of the impact of *c* on the trial level false alarm rate. Choosing $c = 1.6$ biases the test statistic enough that the threshold of 13.814 satisfies a trial false alarm rate of 10^{-3}.

From Fig. 4.3.1, it is obvious that choosing $c = 1$, the biased test performs the same as the unbiased and classic test. To achieve a false alarm rate of 10^{-3} in the Fisher consistent cases and the classic case, the threshold should be set at $\Gamma \cong 16.2$, which was found heuristically. In later analysis, it will be noted when results arise from checking known cycle frequencies vs. when results arise from running the entire algorithm, as laid out in Section 4.2.3. Figure 4.3.1 can be used to select thresholds when $c = 1$ and the selection criterion is the global maximum of the CTMFE.

The behavior of $\frac{\hat{\theta}^2_{CMAD}}{c}$ for the digital signals of interest was derived in Chapter 3. To summarize, at low SNR, the choice of *c* will at best not alter the magnitude of the test statistic (for $c = 1$), but as c moves away from 1, it can actually decrease the magnitude of the test statistic, as was the case for the analog signals. At higher SNR, though, the choice of c can actually increase the magnitude of the test statistic. A choice of $c = 1$ will either leave the non-cycle frequency test statistics of SSB and DSB unchanged, or in some SNR cases for DSB *decrease* the test statistic, while simultaneously either leaving the cycle frequency test statistics unchanged (in low SNR) or increasing the value of the cycle frequency test statistics. For this reason, $c = 1$ is a good choice for controlling the behavior of the scaling term $\frac{\hat{\theta}^2_{CMAD}}{c}$. Figure 4.3.2 illustrates how $c = 1$ is a good choice when considering the behavior of the test statistic at cycle frequencies. The test statistics in the histograms in Fig. 4.3.2 were found at known cycle frequencies of a BPSK signal in an SNR of 2 dB.

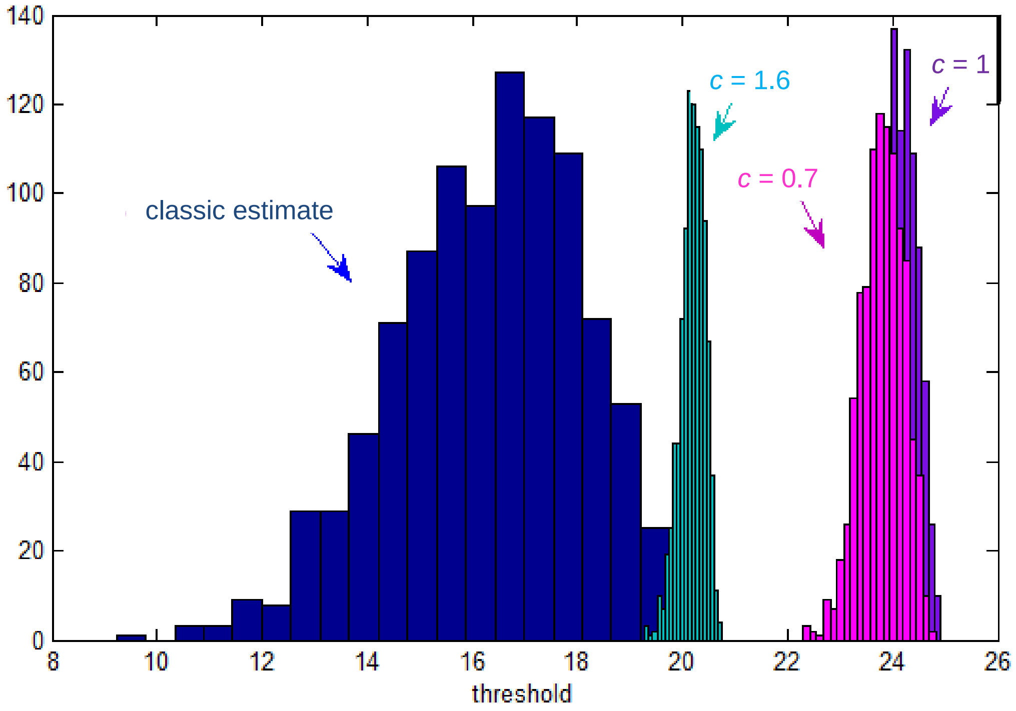


Fig. 4.3.2 Histograms illustrating the impact of the choice of c on the biased test statistic.

The biased test statistic calculated from digital signals has a slightly larger mean when $c = 1$, than for other choices of c. For this same choice of $c = 1$, the false alarm rate remains unchanged, as compared to the unbiased test statistic as seen in Fig. 4.3.1. Therefore, we have gained in probability of detection without sacrificing false alarm rates.

The rest of the comparisons between the classic estimator and the robust estimator will be for the case of $c = 1$, as this choice can increase the test statistic in the digital case without doing so in the analog case.

Figure 4.3.3 illustrates probability of detection vs. SNR for the biased and unbiased robust test statistic and the classic test statistic, given perfect knowledge of the digital signals' cycle frequency and a tolerable false alarm rate of 10^{-3}. The results, therefore, do not reflect the ability of the classifier to select the correct candidate cycle frequency based on peaks in the CTMFE. Also, the chi-squared distribution with two degrees of freedom correctly reflects the natural trial false alarm rate since the statistical test is always applied at the same non-cycle frequency in the analog case; the threshold for statistical significance, therefore can be taken from the chi-squared charts and was chosen to be 13.814. The observed false alarm rates were well below the theoretical rates, thus satisfying the tolerable rates.

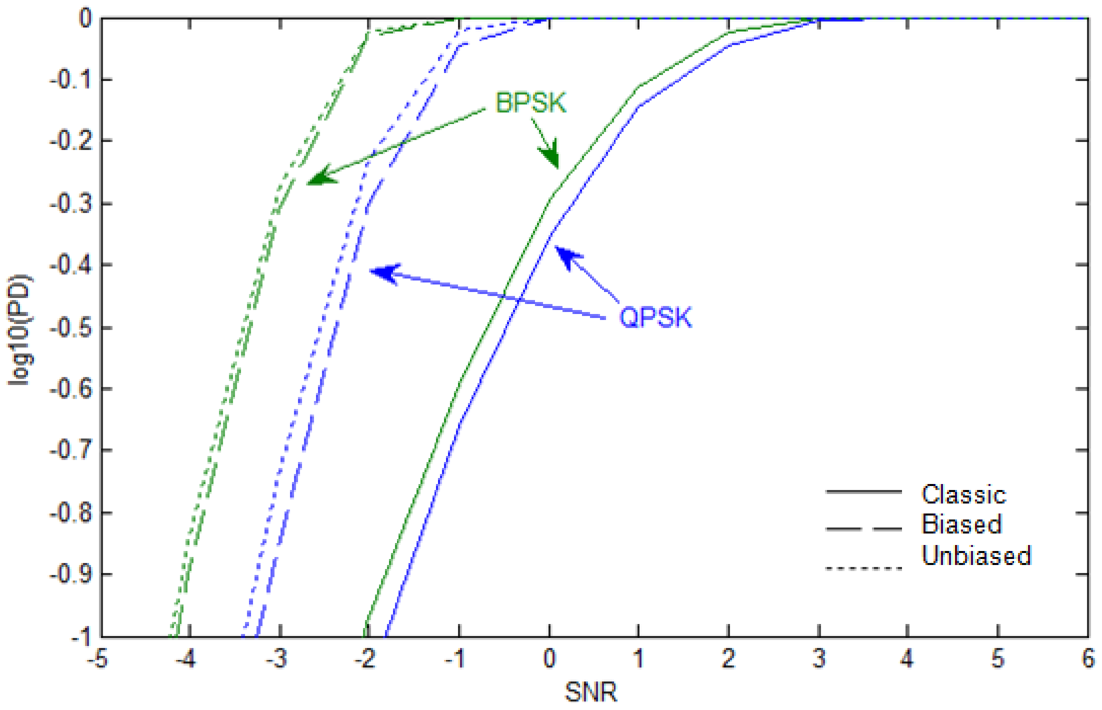


Fig. 4.3.3 Impact of the choice of statistical test on PD, the probability of detecting second-order cyclostationarity in digital signals given perfect knowledge of the cycle frequency ($\Gamma = 13.814$).

Notice that even when underlying assumptions on the data cannot be made, such that the test statistic cannot be biased in a meaningful way, the robust test still outperforms the classic test. In fact, the major benefits of biasing the test statistic are best realized at higher SNRs, when classification is already easier. Figure 4.3.4, below, illustrates the dependence on SNR for noticeable improvements in detection between the biased an unbiased robust test statistic.

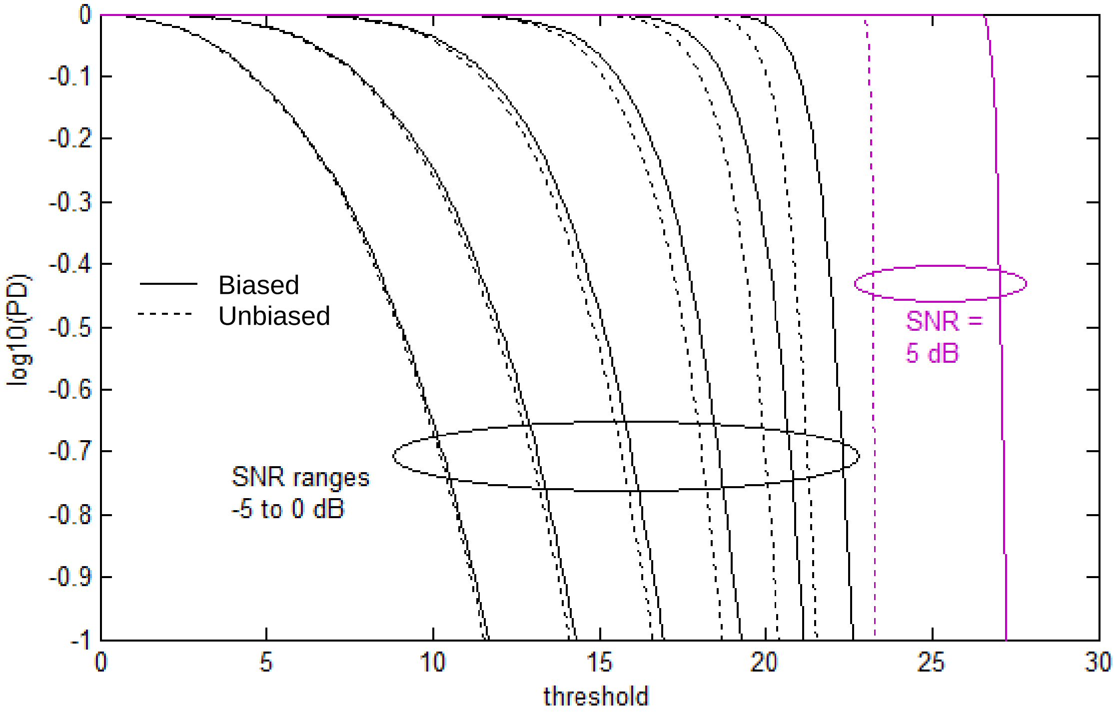


Fig. 4.3.4 Demonstration of the impact of the calculation of the robust test statistic, biased or unbiased, on the distribution of the robust test statistic. The biased test statistic, solid line, achieves better probability of detection, as compared to the unbiased test statistic, as SNR increases.

From observation of Fig. 4.3.4, and as was suggested in Chapter 3, the improvements from biasing the test statistic vanish at low SNRs. Since the value of the bias of $\frac{\hat{\theta}^2_{CMAD}}{c}$ increases as the underlying distribution moves further from a Gaussian and closer to a mixed Gaussian, its value increases above one in relation to the SNR. This detection performance dependence on SNR is similar to the false alarm reduction requirements dependence on SNR, in the DSB case.

A priori knowledge of cycle frequencies is not always available. Correct identification of candidate cycle frequencies, therefore, cannot be assumed. The ability to identify the correct cycle frequency as the candidate cycle frequency impacts the overall probability of detection for the classifier. Relating the magnitude of the CTMFE to the candidate cycle frequency also increases the probability of trial false alarm, since the magnitude of the CTMFE is loosely related to the magnitude of the test statistic – even at non-cycle frequencies.

Figure 4.3.5 illustrates the improvements – that result from using the robust CTMFE to select candidate cycle frequencies – in the ability of the classifier to identify the correct cycle frequency. In Fig. 4.3.5, only the probability of identifying the correct cycle frequency is being reported at each SNR. This data does not speak to the results of the statistical test. The candidate cycle frequency is chosen as described by Dobre's method, i.e. in terms of locally significant maximums [17]. The noise bandwidth is 0.25 F_S. Note that the selection process for both the biased and unbiased robust statistical test is the same, so the results reported below apply to both methods.

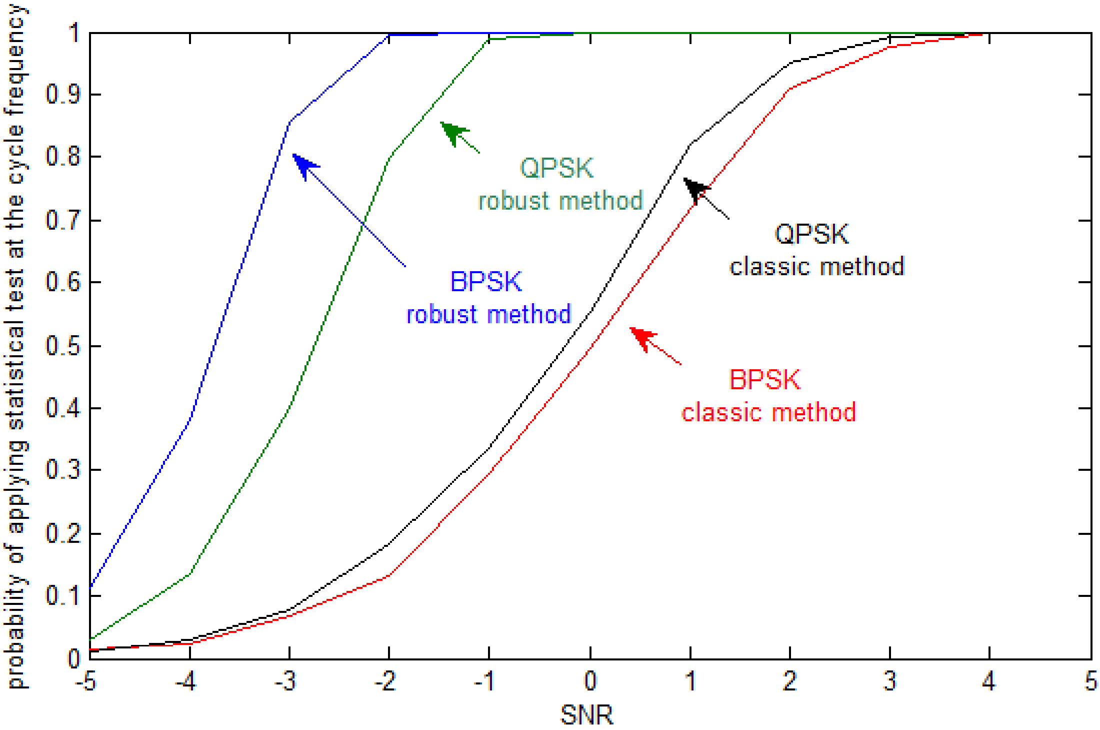


Fig. 4.3.5 Demonstration of the improvements in identifying the correct cycle frequencies when using the robust CTMFE, as opposed to the classic estimate.

In the case of the robust estimator, the local maximum criterion yields better candidate cycle frequency detection in low SNR, as compared to the global maximum, but the test statistic is not yet statistically significant at these low SNRs. Through experimental trials, it was found that when the test statistic is statistically significant, the CTMFE at the cycle frequency is also the global maximum. For this reason, the robust classifier can use the global maximum to simplify selection without noticeable degradation in the probability of correctly classifying the digital signals. For the classic estimator, the test statistic is significant before the magnitude of the CTMFE at the cycle frequency becomes the global maximum. The selection criteria, therefore, cannot be simplified for the classic estimator without sacrificing correct classification of digital signals.

The selection criteria impacts the probability of correctly classifying digital signals *and* analog signals in that it can artificially control false alarm rates by restricting the probability of applying the statistical test at non-cycle frequencies. It was found in the case of both the robust and classic estimators that the best overall performance is achieved when local maximum searches are done, as this reduces false alarms more frequently than causing missed detection of cycle frequencies. The tradeoff for avoiding applying the statistical test at non-cycle frequencies while not inflicting too strict of a criterion so that cycle-frequencies are not also passed over, is an important design consideration. The best results in this work correspond to requiring that the magnitude of the CTMFE be at least 4.25 times as large as that of its neighbors before applying the statistical test. While this negatively impacts the ability to flag cycle frequencies at low SNR, the test statistics associated with cycle frequencies that are not tested, due to the candidate cycle frequency selection criteria, are not typically statistically significant. Also, reducing the selection criteria would require increasing the threshold of statistical significance to maintain the same false alarm rate. In this sense, enabling the classifier to detect smaller peaks at cycle frequencies (that are not likely to be statistically significant) in turn raises the probability of trial level false alarms.

Consider the limiting case where all frequencies in the resolution of the CTMFE are tested, then the trial false alarm rate for a given threshold is dependent on the number of frequencies being tested. In cases where only 100 frequencies are being tested, the threshold could be much lower than in cases where 10000 frequencies are being tested. Without a priori knowledge of the actual cycle frequency, though, limiting the resolution of the CTMFE reduces the ability of the classifier to identify cycle frequencies.

The results in Fig. 4.3.2 – Fig. 4.3.5 were found for a receiver BW of 0.25 F_s, which violates the assumption made by Dobre [17] of out of band noise blocking. The robust estimator, though, also enjoys robustness to changes in receiver bandwidth. The results below, in Fig. 4.3.6, illustrate the change in the performance of the classic estimator when even more out of band noise is blocked, specifically, when the receiver bandwidth is 0.09 F_s. The performance of the robust estimator remains unchanged. When the receiver bandwidth is 0.09 F_s, the results we found for the classic method match the results found in the previous work [17].

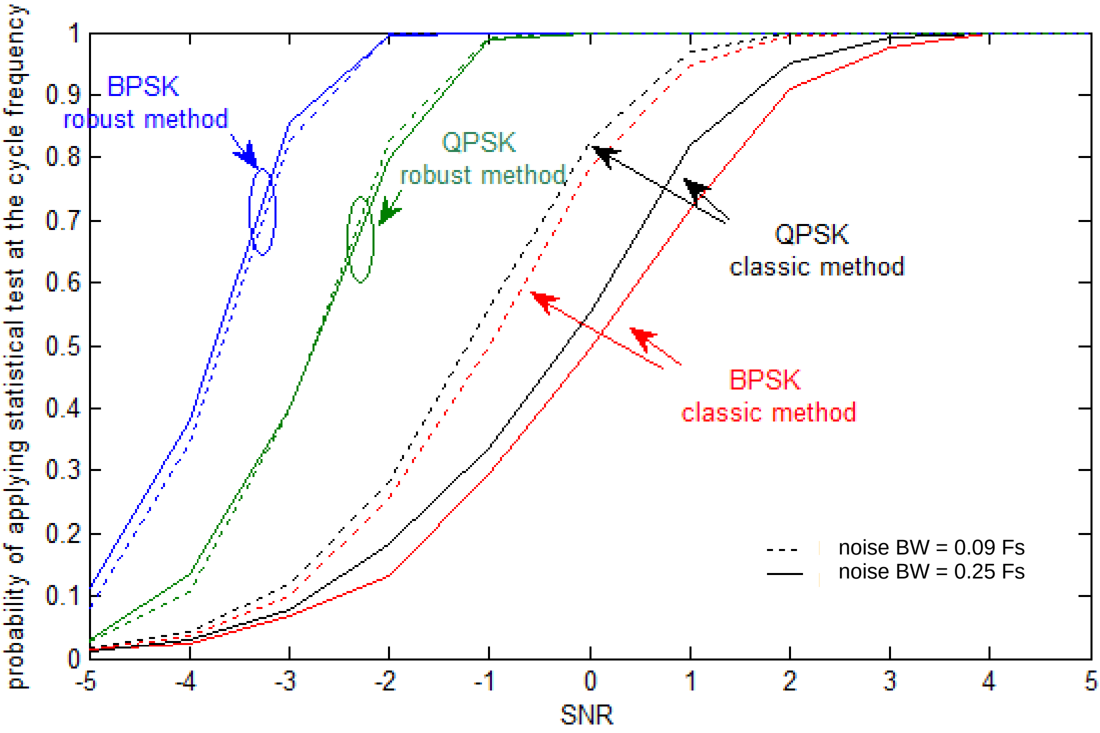


Fig. 4.3.6 Demonstration of the impact of the noise bandwidth on the behavior of the CTMFE estimate.

Even with the performance improvement from blocking out of band noise, though, the classic estimator does not perform as well as the robust estimator. The effect of the receiver bandwidth on classifier performance is not limited to its ability to identify candidate cycle frequencies. The bandwidth also directly impacts the distribution of the test statistic at a cycle frequency. Figure 4.3.7 illustrates the impact of the receiver BW on the distribution of the test statistic when the classifier is given perfect knowledge of the cycle frequency. The signal of interest is a BPSK with an SNR of 3 dB at the output of the receiver. The results reflect the value of the test statistic calculated at the cycle frequency over 100,000 trials.

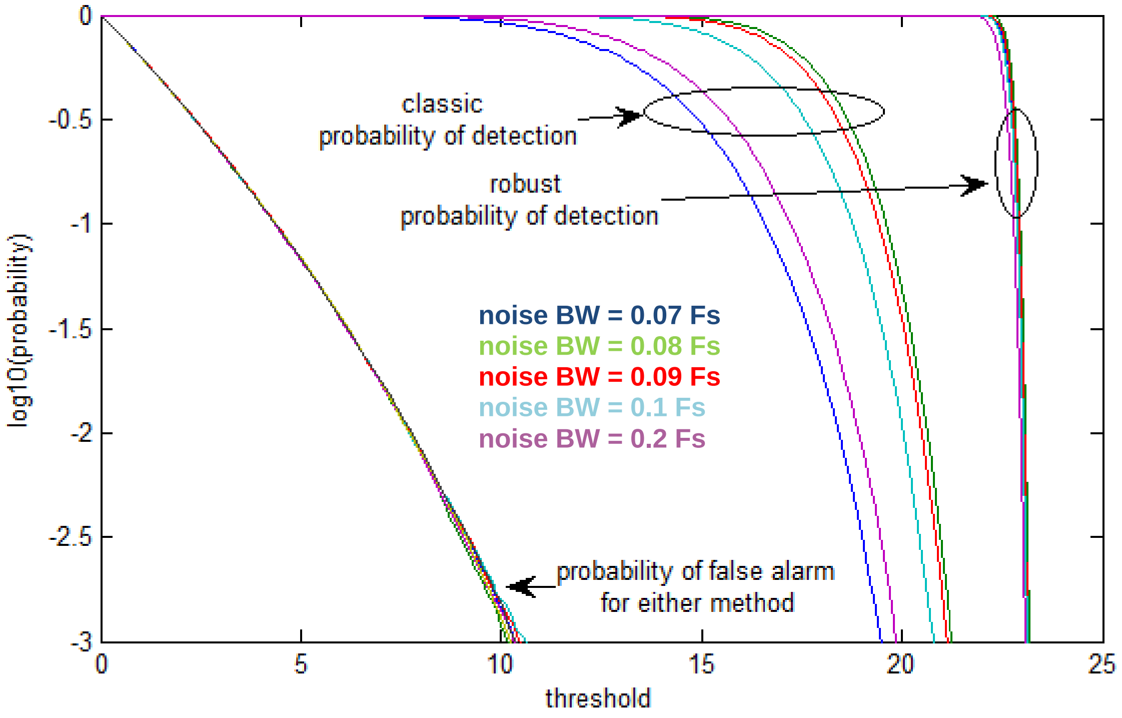


Fig. 4.3.7 Demonstration of the impact of the noise bandwidth on the probability that the cycle frequency test statistic exceeds a certain threshold.

The bandwidth occupied by the signal is approximately 0.08 F_s, and the classic estimator achieves the best performance, of the listed BW cases, when the receiver bandwidth closely matches the signal bandwidth. When the bandwidth of the receiver is even slightly narrower than the signal, the classic estimator suffers a significant loss in performance. Notice that the robust estimator maintains the same performance for the entire range of bandwidths investigated here.

As a final performance comparison, in Fig. 4.3.8 the performance of the classic classifier and of the unbiased and biased robust classifiers is compared.

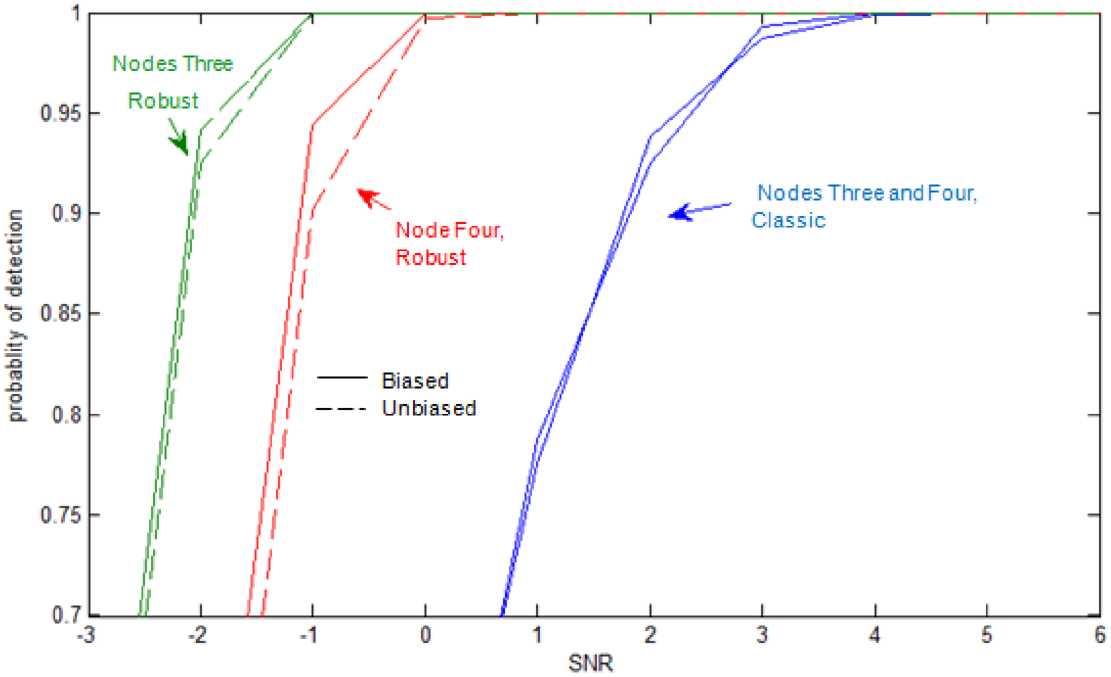


Fig. 4.3.8 Performance comparison of the robust node and classic node. This performance plot is a reflection of both the statistical significance of test statistics as well as the ability of the classifier to identify the correct candidate cycle frequency.

The plots in Fig. 4.3.8 reflect the performance of the entire algorithm, as illustrated in the classifier node algorithms, Fig. 4.2.3. The receiver bandwidth is 0.09 F_s. The trial false alarm rate no longer naturally follows a chi-squared distribution with 2 degrees of freedom, since the candidate cycle frequency selection is related to the magnitude of the CTMF. Both the classic and robust tests use the local maximum selection criterion to suppress false alarms. The probability of detection is a simultaneous reflection of the performance of the selection criterion as well as the statistical significance of the test statistic. $\Gamma = 13.814$ corresponds to a trial level false alarm rate of 10^{-3} because of the strict candidate cycle frequency selection criterion.

Alternative results are shown in Fig. 4.3.9, when the global maximum is used and the estimator is biased with c = 1.6. The non-Fisher consistent results with a simpler detection strategy result in a loss of about 1 dB in performance. The classic results are plotted for reference.

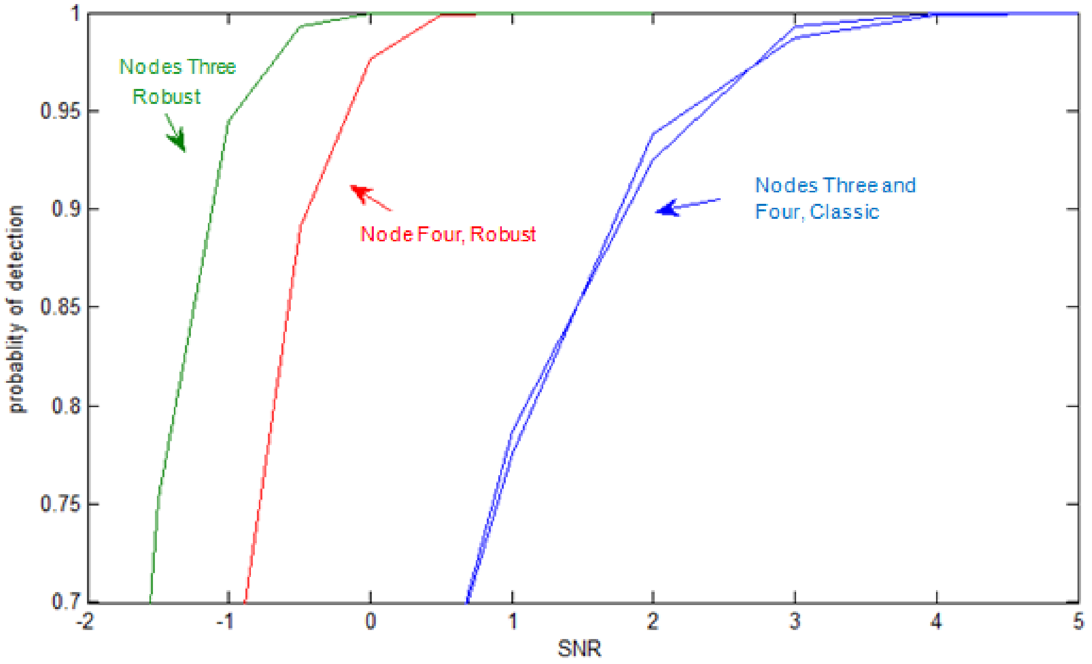


Fig. 4.3.9 Performance comparison of the robust node and classic node when the robust test uses a biased test statistic with $c = 1.6$ and the global maximum selection criteria.

The performance of the robust classifier can be traded off in exchange for a simpler peak detection algorithm. Additionally, in the simplified algorithm, there is always only one candidate cycle frequency chosen. In theory, the search for local maximums can return multiple candidate cycle frequencies, which increases the computational complexity and time requirements. In time sensitive/limited resource applications, it could be advantageous to trade of one dB in performance for speed.

Finally, a further motivator to consider robust estimates of cyclostationarity is its potential to reduce the long observation times typically required in classic estimators of cyclostationarity. Initial testing shows that the robust estimator has the potential to perform accurate classification with observation times much shorter than one second, while the performance of the classic estimator is much more sensitive to shortening of the observation window. Figure 4.3.10 reflects the outcome of 15000 trials where the classic and robust estimators were presented the same 4800 samples (1/10 of the observation time compared to any of the previous results above) and given a priori knowledge of the location of the cycle frequency.

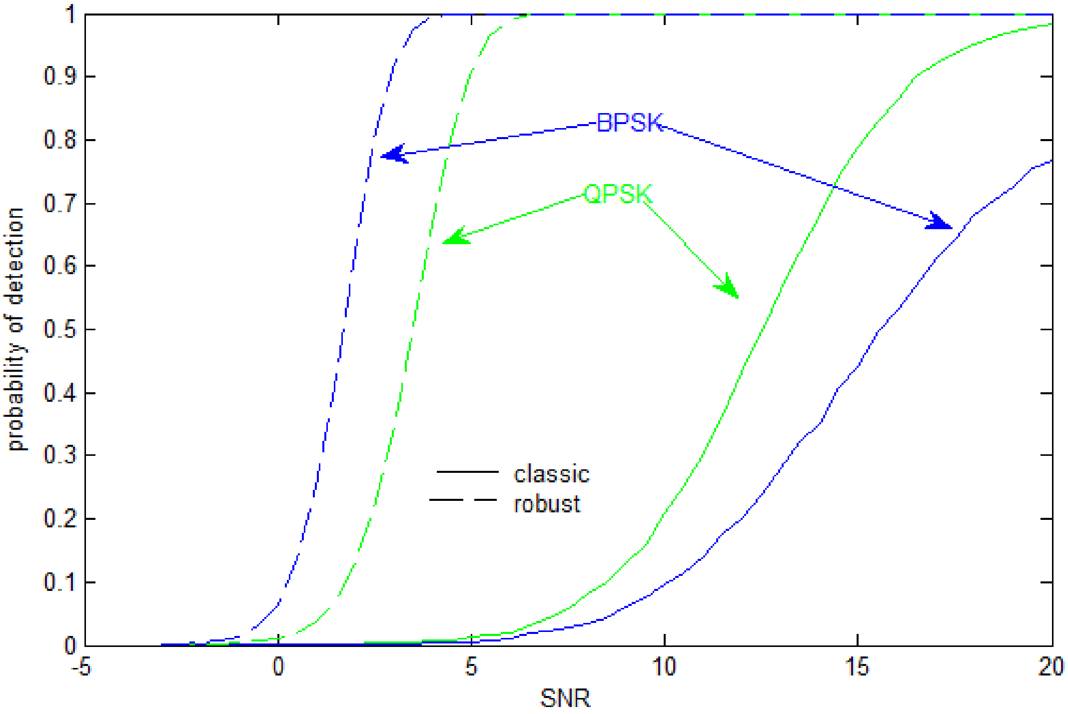


Fig. 4.3.10 Performance comparison of the robust node and classic node when the observation time is 0.1 seconds (versus the 1 second observation window in [17]).

The biased robust estimator achieves 100% detection at an SNR of 5 dB. The unbiased robust estimator only requires 6 dB. Also, the BPSK SNR requirements are 1-2 dB less than the QPSK SNR requirements, depending on the biasing of the test statistic. The classic estimator, on the other hand, does not achieve 100% correct detection at an SNR of even 15 dB in either case.

Initial results motivate the continued consideration of the robust classifier for shorter observation windows. While only the second-order first-conjugate case of cyclostationarity was studied here, as it was the only feature requiring SNR above 0 dB for detection, the significant SNR improvements gained by using the robust estimator during short observation times may be possible at lower order estimates of cyclostationarity as well, where classic estimator SNR requirements are likely to surpass 0 dB.

4.4 Conclusion

The proposed robust classifier of analog and digital signals based on robust estimates of second-order first-conjugate cyclostationarity achieves an SNR improvement of 4-5 dB, depending on some design considerations when compared to the classic classifier which does not exploit any robust statistics.

In addition to reducing SNR requirements, the robust classifier is also robust to variations in noise bandwidth. In channels where the bandwidth of the signal can fluctuate and is not readily estimated, it is advantageous that a classifier maintain performance as the relative ratio of signal bandwidth to noise bandwidth fluctuates. Initial testing also shows promising results for using robust estimators of cyclostationarity to mitigate the growing SNR requirements as observation time decreases, as may be important for cognitive radio applications for example.

Cyclostationarity Feature-Based CPM Detection

5.1 Introduction

From Chapter 2, the definition of a CPM signal is repeated in (5.1.1).

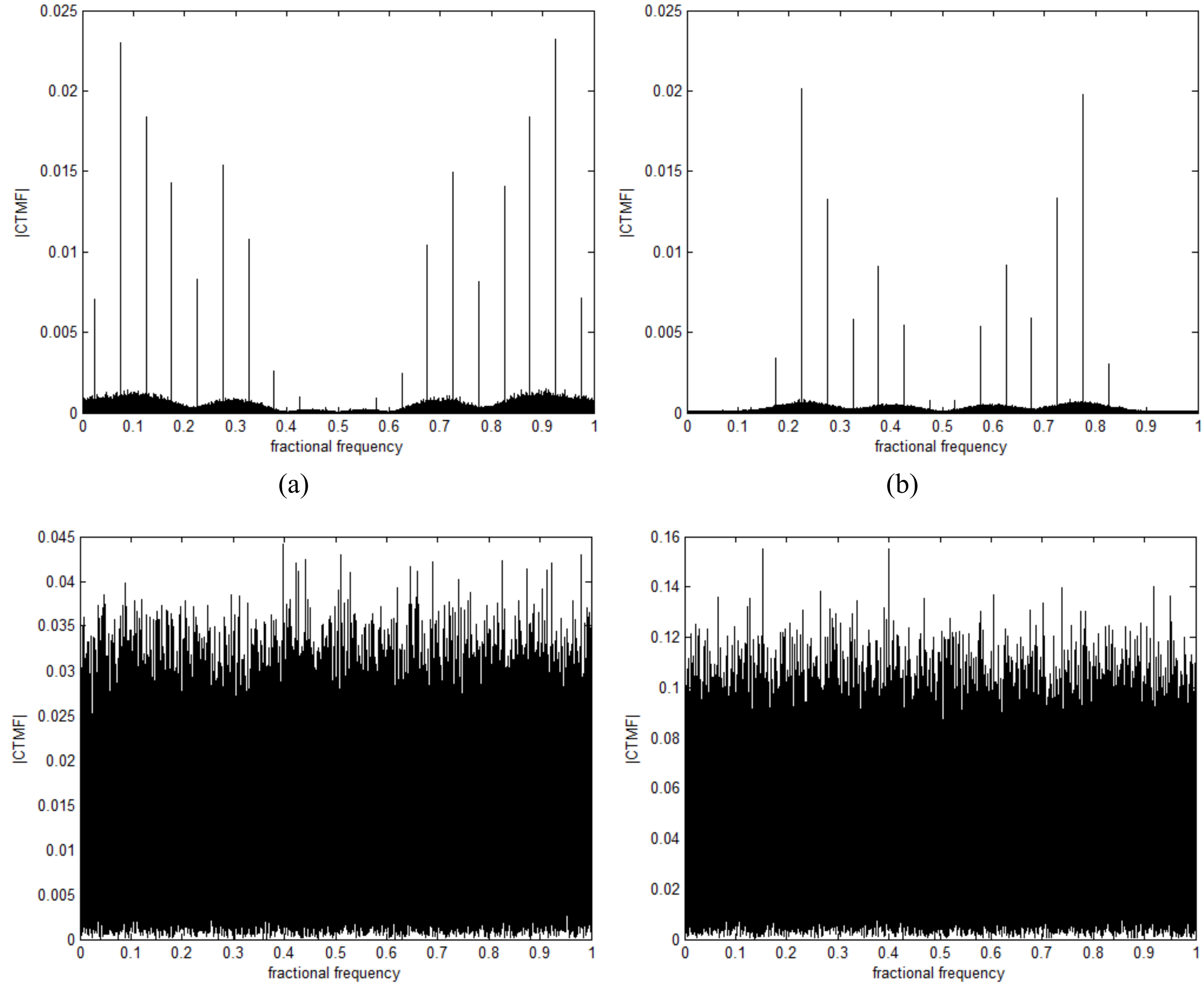
$$s(t) = \exp\left(j\left(2\pi \sum_{k=-\infty}^{\infty} \rho_k h_{\langle k \rangle_H} q(t - kT_{sym}) \right)\right) \tag{5.1.1}$$

In this work we consider 1-REC – full response – CPM signals with a phase response $q(t) = \int_0^t g(\tau)d\tau$. The baseband frequency pulse $g(t)$ is defined as in (5.1.2)

$$g(t) = \begin{cases} \frac{1}{2T_{sym}} & 0 \le t < T_{sym} \\ 0 & otherwise \end{cases} \tag{5.1.2}$$

As stated in Chapter 2, it is well known that CPM signals contain high-order cyclostationarity when the delay vector of the lag product is all zeros. The order n of cyclostationarity of a CPM signal is directly related to the modulation index. The locations of the cycle-frequencies are directly related to the modulation index and the symbol alphabet [29].

Estimating high-order cyclostationarity from finite noisy data sets is a difficult and sometimes impossible task. Figs. 5.2.1a-d illustrate the impact of complex AWGN on the estimated n-th order CTMFE of two multi-h 4-ary CPM signals. The noise-free signals, $s_1(t)$ and $s_2(t)$, have length H = 2 modulation indices $h_1 = \{1/9 \quad 2/9\}$ and $h_2 = \{4/16 \quad 5/16\}$ respectively. Accordingly, the lowest orders of cyclostationarity are n = 9 and n = 16, as explained in Chapter 2. The noisy signals $r_1(t) = s_1(t) + n(t)$ and $r_2(t) = s_2(t) + n(t)$, are associated with the signals $s_1(t)$ and $s_2(t)$ observed in complex additive white Gaussian noise.



(a) (b)

(c) (d)

Fig. 5.2.1 a. 9th order CTMFE of $s_1(t)$ from clean finite sample set. (b) 16th order CTMFE of $s_2(t)$ from clean finite sample set. (c) 9th order CTMFE of $r_1(t)$ from noisy finite sample set, SNR = 5 dB. (d) 16th order CTMFE of $r_2(t)$ from noisy finite sample set, SNR = 10 dB.

In both noisy cases, the SNR – of 5 dB and 10 dB, respectively – is relatively high from a practical standpoint, but in both of these cases the cycle frequency spectral lines are not apparent. The work in this chapter presents a novel consideration of second order CPM cyclostationarity resulting from non-zero delay vectors.

5.2 Second-Order Cyclostationarity of CPM Signals with non-zero Delay Vectors

The piecewise second-order first-conjugate lag product of the continuous, infinite time CPM signal *s*(*t*) with delay vector $\boldsymbol{\tau} = \begin{bmatrix} 0 & \Delta t \end{bmatrix}$ where $\Delta t < T$ and T is the symbol time, as a function of the delay Δt, is given in (5.2.1).

$$L_s(t,\boldsymbol{\tau})_{2.1} = \begin{cases} \sum_k e^{j\left(2\pi\rho_k h_{\langle k\rangle_H}\left[\frac{t-kT}{2T}\right]+\pi\sum_{i=1}^{k-1}\rho_i h_{\langle i\rangle_H}\right)} e^{-j\left(2\pi\rho_{k-1} h_{\langle k-1\rangle_H}\left[\frac{t-kT+(T-\Delta t)}{2T}\right]+\pi\sum_{i=1}^{k-2}\rho_i h_{\langle i\rangle_H}\right)} & 0 \le t < \Delta t \\ \sum_k e^{j\left(2\pi\rho_k h_{\langle k\rangle_H}\left[\frac{t-kT}{2T}\right]+\pi\sum_{i=1}^{k-1}\phi_i\right)} e^{-j\left(2\pi\rho_k h_{\langle k\rangle_H}\left[\frac{t-kT+\Delta t}{2T}\right]+\pi\sum_{i=1}^{k-1}\rho_i h_{\langle i\rangle_H}\right)} & \Delta t \le t < T \end{cases} \tag{5.2.1}$$

The second-order first-conjugate CTMF of the complex CPM signal, as a function of the delay (see the full derivation in Appendix), is given in (5.2.2) below. Let $\theta_{m,n,q} = \left(\rho_n h_{\langle q\rangle_H} - \rho_m h_{\langle q-1\rangle_H}\right)$, $\phi_{m,q} = \rho_m h_{\langle m\rangle_H}$, and $C_1 = \frac{1}{TH2^{M+1}}, C_2 = \frac{1}{TH2^M}$.

$$\begin{aligned} R_s^\alpha(\boldsymbol{\tau})_{2,1} = {} & C_1 \sum_{m=0}^{2^M-1}\sum_{n=0}^{2^M-1}\sum_{q=1}^{H} e^{j\pi\left(\frac{\Delta t\phi_{n,q-1}}{T}\right)} \sum_l \delta\left(\alpha - \frac{l}{HT}\right)\left(\frac{e^{\frac{j2\pi\Delta t}{T}\left(\frac{\theta_{m,n,q}}{2}-\frac{l}{H}\right)}-1}{\frac{j2\pi}{T}\left(\frac{\theta_{m,n,q}}{2}-\frac{l}{H}\right)e^{j\frac{2\pi ql}{H}}}\right) \\ & + C_2 \sum_{m=0}^{2^M-1}\sum_{q=1}^{H} e^{-j\pi\frac{\Delta t\phi_{m,q}}{T}} \sum_l \delta\left(\alpha - \frac{l}{HT}\right)\left(\frac{e^{-j\frac{2\pi l}{H}} - e^{-j\frac{2\pi l\Delta t}{HT}}}{-\frac{j2\pi l}{HT}e^{j\frac{2\pi ql}{H}}}\right) \end{aligned} \tag{5.2.2}$$

The theoretical expression for the discrete infinite time CTMF is given in (5.2.3) below.

In the discrete case, in addition to satisfying $\Delta t < T$, only delay vectors of the form $\boldsymbol{\tau} = \begin{bmatrix} 0 & \Delta t \end{bmatrix}$, $\Delta t = \lambda T_{sam}$, where Δt is an integer λ multiple of the sample time T_{sam}, are considered.

$$\begin{aligned} R_s^{\alpha}(\boldsymbol{\tau})_{2,1} = {} & C_1 \sum_{m=0}^{2^M-1} \sum_{n=0}^{2^M-1} \sum_{q=1}^{H} \frac{\mathrm{e}^{j\pi\left(\frac{\Delta t \phi_{m,q-1}}{T_{sym}}\right)}}{\mathrm{e}^{\frac{j2\pi q}{H}}} \sum_{l} \delta\left(\alpha - \frac{l}{HT_{sym}}\right) \sum_{v=0}^{(\lambda-1)} \mathrm{e}^{\frac{j2\pi v}{T_{sym}}\left(\frac{\theta_{m,n,q}}{2} - \frac{l}{H}\right)} \\ & + C_2 \sum_{m=0}^{2^M-1} \sum_{q=1}^{H} \frac{\mathrm{e}^{-j\pi\left(\frac{\Delta t \phi_{m,q}}{T_{sym}}\right)}}{\mathrm{e}^{\frac{j2\pi q}{H}}} \sum_{l} \delta\left(\alpha - \frac{l}{HT_{sym}}\right) \sum_{v=\lambda}^{L-1} \mathrm{e}^{\frac{-j2\pi v}{HT_{sym}}} \end{aligned} \tag{5.2.3}$$

To verify the accuracy of (5.2.3), the theoretical CTMF, as defined in (5.2.3), is compared to the CTMFE, found as described in Chapter 2 in (2.2.7).

Figs. 5.2.2 through 5.2.5 demonstrate the accuracy of (5.2.3) for various CPM configurations. The discrepancies between the theoretical and estimated values can be attributed to the finite data sequence and the unavoidable violation of the assumption of completely equiprobable symbols and symbol transitions. As the length of the modulation index increases, or as *M* gets larger, these assumptions are more noticeably violated. In an effort to mitigate the effect of finite sample lengths, for each CPM signal, the CTMFs from 1000 different realizations of the signal are averaged. For visualization purposes, connections are drawn between cyclic moments – nonzero peaks in the CTMFE – at the same frequency but found from different delays.

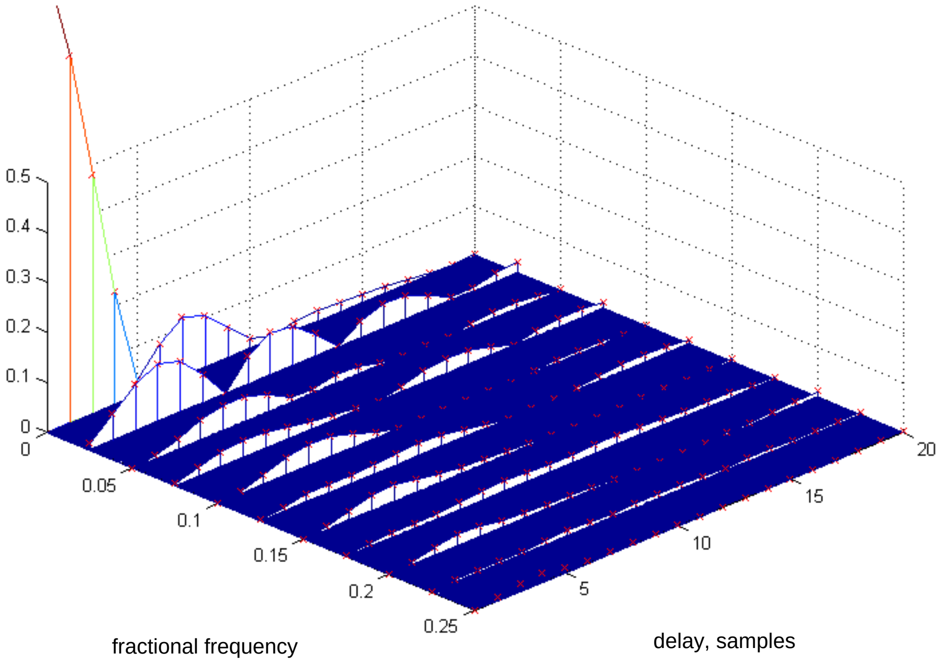


Fig. 5.2.2 Second-order first-conjugate CTMFE for a CPM signal with $M = 3$, $h = [7\ 10]/16$, and $T_{sym} = 20T_{sam}$. Theoretical values are shown as red asterisks; observed values are shown as a 3-D plot.

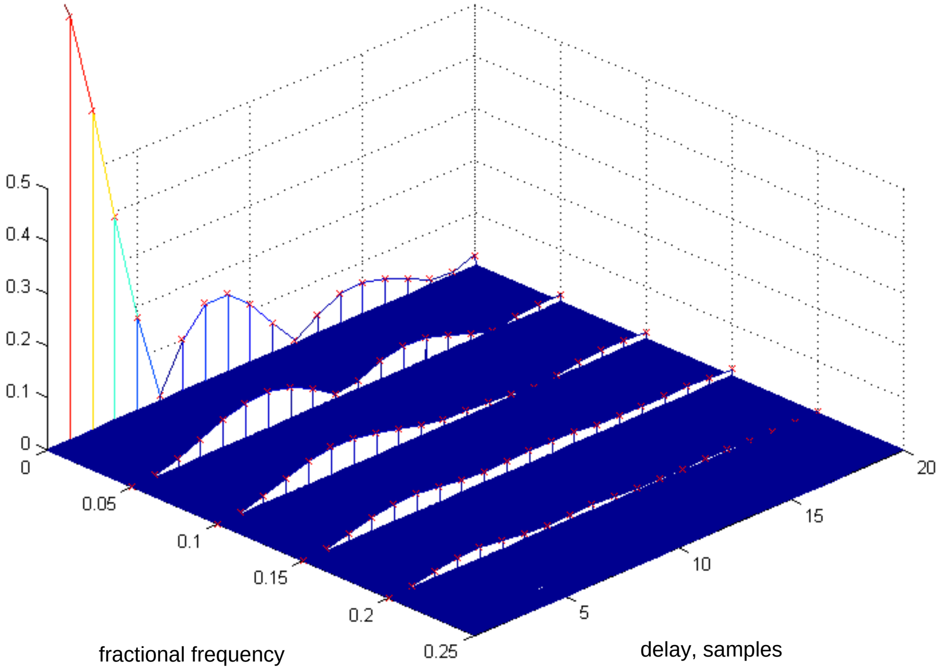


Fig. 5.2.3 Second-order first-conjugate CTMFE for a CPM signal with $M = 4$, $h = 2/9$, and $T_{sym} = 20T_{sam}$.Theoretical values are shown as red asterisks; observed values are shown as a 3-D plot.

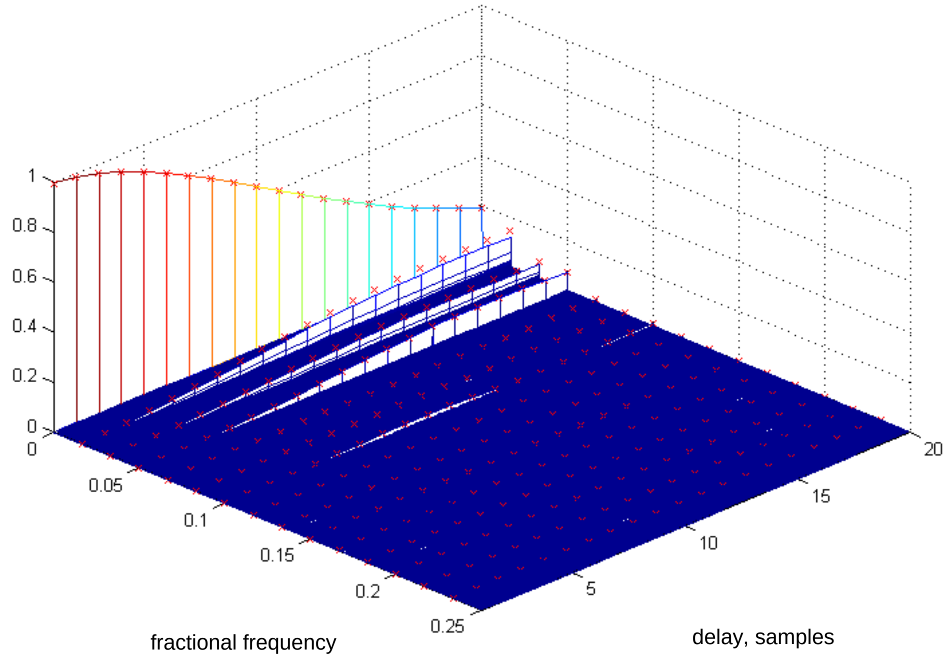


Fig. 5.2.4 Second-order first-conjugate CTMFE for a CPM signal with $M = 2$, $h = [2\ 3\ 4]/11$, and $T_{sym} = 20T_{sam}$. Theoretical values are shown as red asterisks; observed values are shown as a 3-D plot.

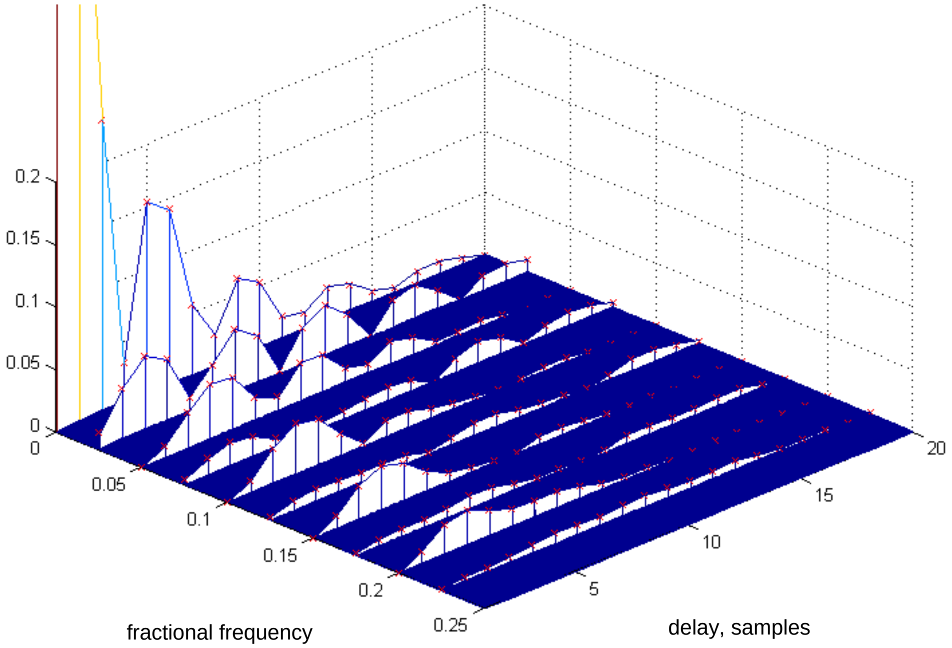


Fig. 5.2.5 Second-order first-conjugate CTMFE for a CPM signal with $M = 4$, $h = [5\ 6]/16$, and $T_{sym} = 20T_{sam}$. Theoretical values are shown as red asterisks; observed values are shown as a 3-D plot.

From Figs. 5.2.2 through 5.2.5, it is apparent that the choice of modulation index, alphabet length, and delay greatly impact the behavior of the CTMF. While some configurations have many choices where the peak at the cycle frequency for a given delay is very prominent, as in Fig. 5.2.5, for other choices of modulation index and alphabet lengths the magnitude of the peaks at cycle frequencies diminishes quickly, as in Fig. 5.2.4.

The CTMFs of the second-order zero-conjugate lag-product of the isolated real $R_{re(s)}^{\alpha}(\tau)_{2,0}$ and isolated imaginary $R_{im(s)}^{\alpha}(\tau)_{2,0}$ components of a CPM signal are equivalent. When the CPM signal of interest is circularly symmetric, the following hold true,

$$\begin{aligned} R_{re(s)}^{\alpha}(\tau)_{2,0} &= R_{im(s)}^{\alpha}(\tau)_{2,0} = \frac{1}{2} R_{s}^{\alpha}(\tau)_{2,1} \\ R_{s}^{\alpha}(\tau)_{2,1} &= R_{re(s)}^{\alpha}(\tau)_{2,0} + R_{im(s)}^{\alpha}(\tau)_{2,0} \end{aligned} \tag{5.2.4}$$

Without loss of generality, an example is provided with a modulation index of $h = [5/16 \quad 6/16]$; the length of the symbol alphabet is 16. Figs. 5.2.6 and 5.2.7 demonstrate the validity of (5.2.4) by comparing the theoretical value $\left|\frac{1}{2} R_{s}^{\alpha}(\tau)_{2,1}\right|$ with the observed $\left|\hat{R}_{re(s)}^{\alpha}(\tau)_{2,0}\right|$ and $\left|\hat{R}_{im(s)}^{\alpha}(\tau)_{2,0}\right|$, respectively. In Fig. 5.2.8 $\left|\hat{R}_{re(s)}^{\alpha}(\tau)_{2,0} + \hat{R}_{im(s)}^{\alpha}(\tau)_{2,0}\right|$ and $R_{s}^{\alpha}(\tau)_{2,1}$ are compared. The difference $\left|\hat{R}_{re(s)}^{\alpha}(\tau)_{2,0} - \hat{R}_{im(s)}^{\alpha}(\tau)_{2,0}\right|$ is shown in Fig. 5.2.9. Any non-zero difference between $\hat{R}_{re(s)}^{\alpha}(\tau)_{2,0}$ and $\hat{R}_{im(s)}^{\alpha}(\tau)_{2,0}$ is attributed to errors resulting from estimation based on a finite data set (this is confirmed by diminishing error magnitudes for estimation from increasingly larger data sets).

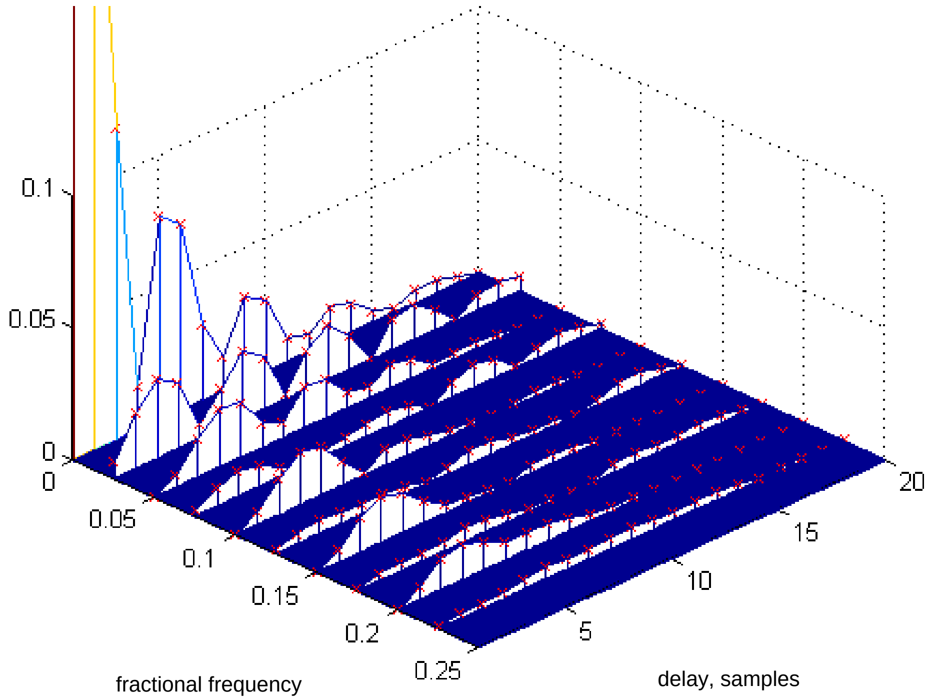


Fig. 5.2.6 Theoretical $\left|\frac{1}{2}R_s^{\alpha}(\tau)_{2,1}\right|$ (red asterisk) and observed/estimated $\left|\hat{R}^{\alpha}_{re(s)}(\tau)_{2,0}\right|$ (3-D).

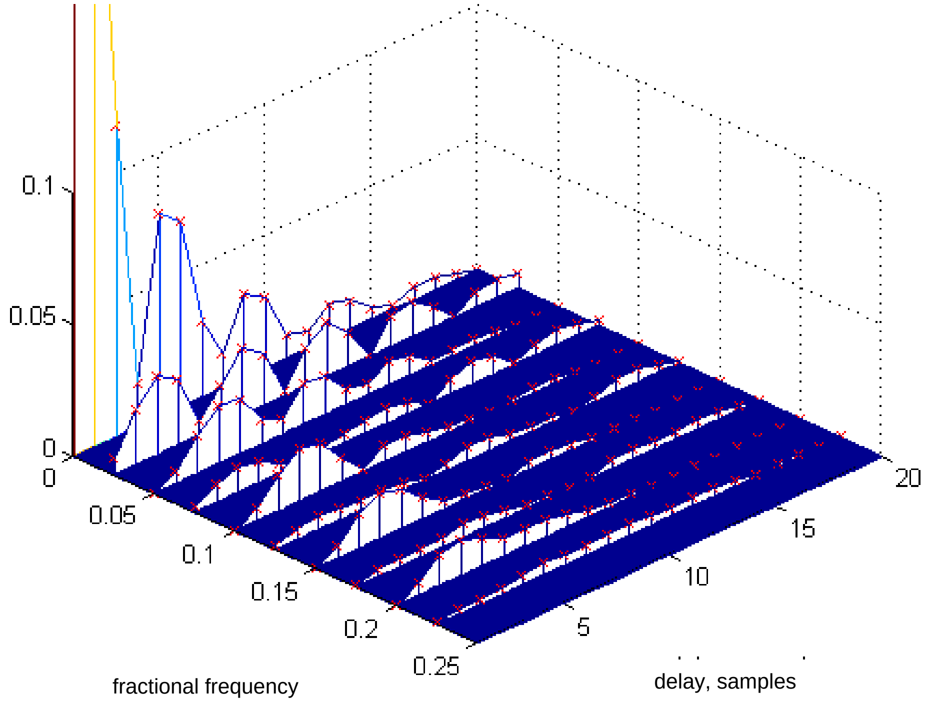


Fig. 5.2.7 Theoretical $\left|\frac{1}{2}R_s^{\alpha}(\tau)_{2,1}\right|$ (red asterisk) and observed/estimated $\left|\hat{R}^{\alpha}_{im(s)}(\tau)_{2,0}\right|$ (3-D).

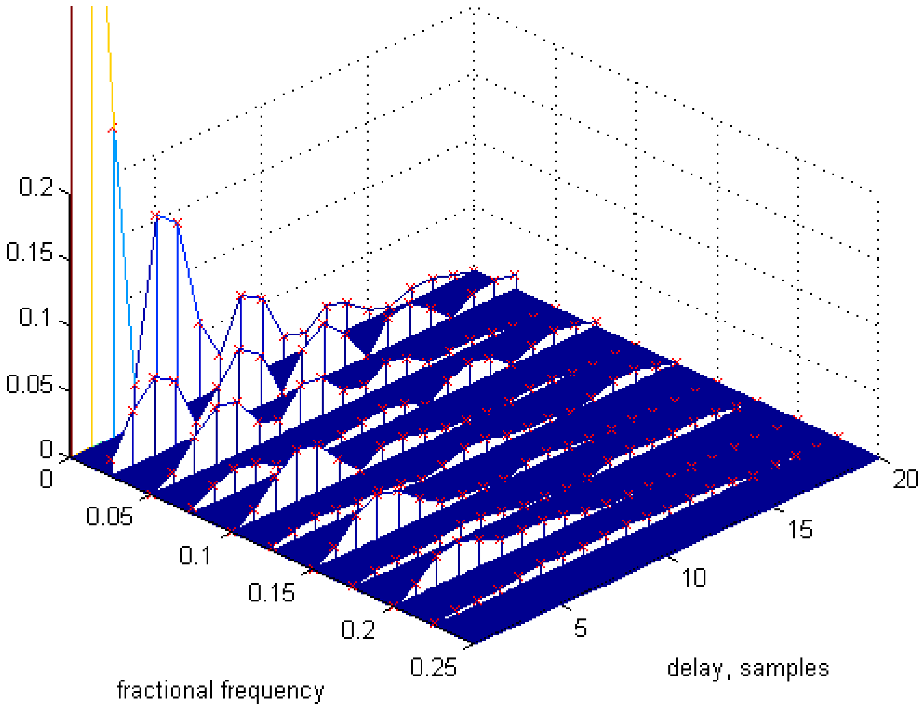


Fig. 5.2.8 Sum $\left|\hat{R}^{\alpha}_{re(s)}(\tau)_{2,0}+\hat{R}^{\alpha}_{im(s)}(\tau)_{2,0}\right|$ (3-D) and $\left|R^{\alpha}_{s}(\tau)_{2,1}\right|$ (red asterisk).

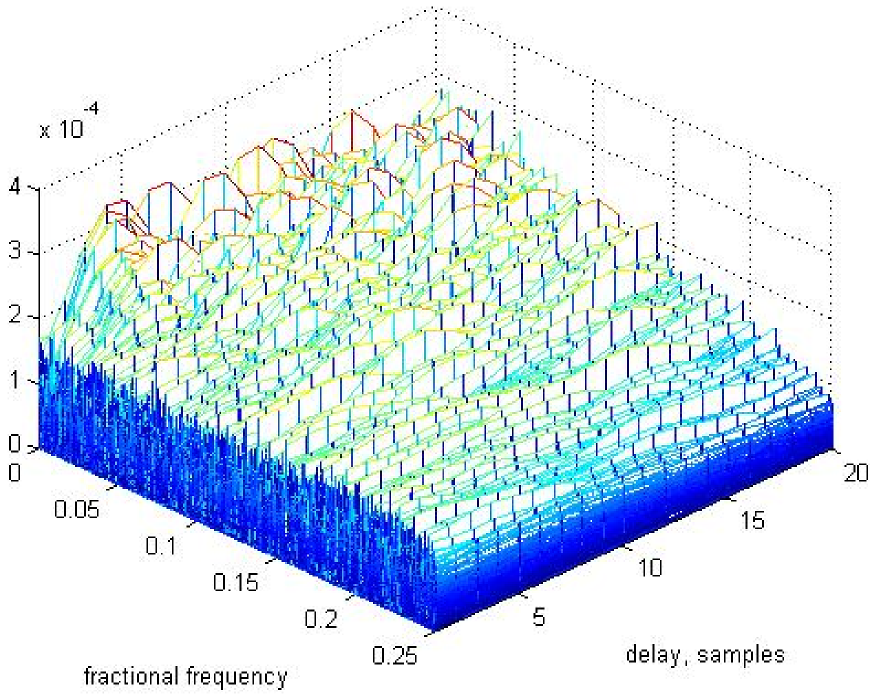


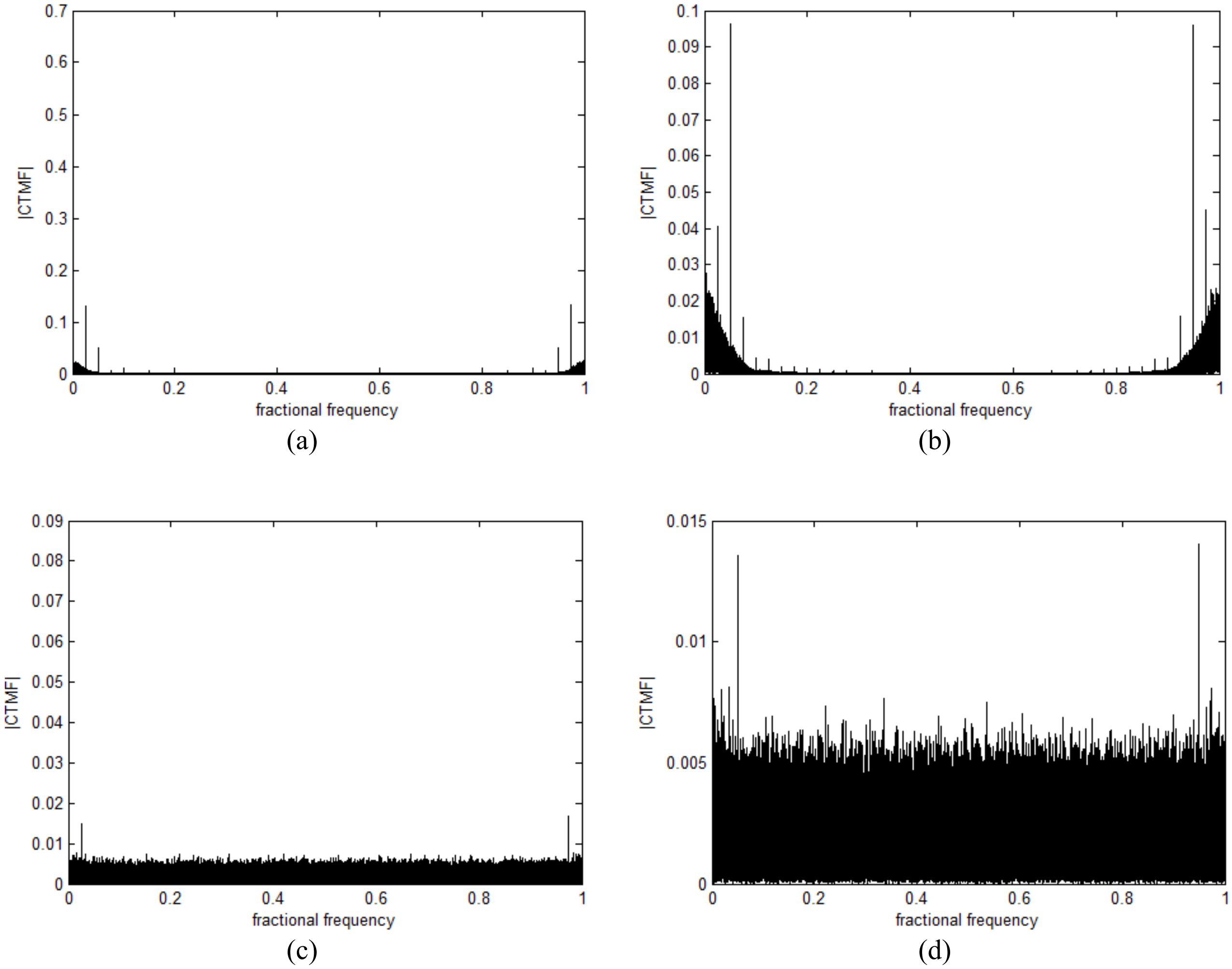
Fig. 5.2.9 Difference $\left|\hat{R}^{\alpha}_{re(s)}(\tau)_{2,0}-\hat{R}^{\alpha}_{im(s)}(\tau)_{2,0}\right|$.

While only one example is provided here, many other cases were studied, all of which validated the relationships in (5.2.4). For a full derivation of the relationships, see Appendix.

Having established the existence of low order cyclostationarity and a corresponding mathematical description, it is of interest to study the impact of complex additive noise on the prominence of the peaks in the CTMF.

5.3 Impact of Complex Additive Gaussian Noise on the second-order CTMFE of a CPM signal

Recall from Section 5.1 that the high-order cyclostationarity of $s_1(t)$ and $s_2(t)$ is not apparent when the SNR is 5 and 10 dB respectively, as was observed in Fig. 5.1.1 a – d. Using the same definition of $s_1(t)$ and $s_2(t)$, but decreasing the SNR of $r_1(t) = s_1(t) + n(t)$ and $r_2(t) = s_2(t) + n(t)$ to -8 dB, the second-order first-conjugate CTMFE of the same complex CPM signals $s_1(t)$ and $s_2(t)$, when the symbol time $T_{sym} = 20T_{sam}$ and delay vector $\boldsymbol{\tau} = \begin{bmatrix} 0 & T_{sym} \end{bmatrix}$, with and without additive noise, is shown in Fig. 5.3.1 a – d.



(a) (b)

(c) (d)

Fig. 5.3.1 (a) 2nd order 1st conjugate CTMFE of $s_1(t)$ (b) 2nd order 1st conjugate CTMFE of (c) 2nd order 1st conjugate CTMFE of $r_1(t)$ (d) 2nd order 1st conjugate CTMFE of $r_2(t)$ all from a one second observation at 200,000 samples / second.

Unlike the peaks in the high order CTMFEs, the peaks in the second order CTMFE remain prominent in $r_1(t)$ and $r_2(t)$ when SNR is as low as -8 dB. While the cyclic moments at higher cycle-frequencies are buried in the noise floor, at least two cycle frequencies remain prominent in each case. This finding agrees with the expectation that lower order cyclostationarity, in general, is easier to detect in complex AWGN than higher order cyclostationarity [33].

This promising discovery suggests that second-order cyclostationarity of CPM signals can be exploited for practical applications. One such application, which is explored in depth in the next section, is the detection problem. A portion of the results presented in this chapter were published in ASILOMAR 2010 [45].

5.4 Application – Blind Detection of CPM in AWGN

In a practical system, there will always be additive noise such that the received signal $r(t)$ is a summation of the desired signal $s(t)$ and the noise $n(t)$. This work considers complex additive Gaussian noise.

A statistical test was proposed for detecting a CF in the noisy CTMFE [34]. At a candidate cycle frequency, a test statistic is calculated and its value is compared to a threshold to determine statistical significance. Justification for the test is provided elsewhere [35]. Some modifications to the test, which originally considers the generally complex signal, are necessary when dealing with the isolated real and imaginary components. In total, two test statistics are calculated, one associated with $\hat{R}^{\alpha}_{re(r)}\left(\tau\right)_{2,0}$ and one associated with $\hat{R}^{\alpha}_{im(r)}\left(\tau\right)_{2,0}$, and found via (2.2.7).

To find the two test statistics, the original test [34] is applied to the second-order zero-conjugate isolated real and imaginary lag products, as illustrated in Fig. 5.4.1. Using the same delay vector τ, the CTMFs of the real component $\hat{R}^{\alpha}_{re(r)}\left(\tau\right)_{2,0}$ and of the imaginary component $\hat{R}^{\alpha}_{im(r)}\left(\tau\right)_{2,0}$ are calculated. Candidate cycle frequencies are chosen from the peaks in the CTMFs.

Then the original statistical test for second-order cyclostationarity is applied to both real $r_{re}(t)$ and imaginary $r_{im}(t)$ at each candidate cycle frequency; this means that peaks identified in $R^{\alpha}_{im(r)}(\tau)_{2,0}$ will trigger a statistical test at the corresponding candidate CFs in both $r_{re}(t)$ and $r_{im}(t)$ and vice versa. The final test statistic, $\Upsilon_{2,0}$, is the sum of $\Upsilon_{2,0\,\mathrm{Im}}$ and $\Upsilon_{2,0\,\mathrm{Re}}$ evaluated for the same candidate CF.

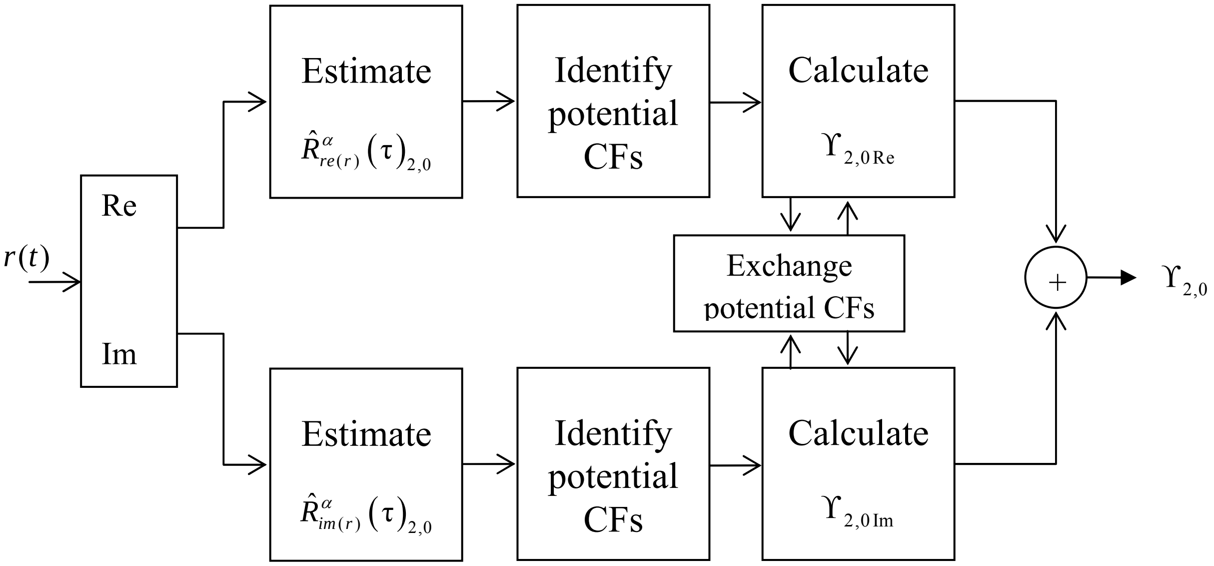


Fig. 5.4.1 Schematic representation of the novel statistical test for identifying cyclostationarity in CPM signals.

It is noteworthy that the candidate cycle frequencies chosen based on $R^{\alpha}_{re(r)}(\tau)_{2,0}$ and $R^{\alpha}_{im(r)}(\tau)_{2,0}$ will not necessarily match. In the event that the signal is a CPM signal, the true CFs will be the same for the isolated real and imaginary components of the signal. Due to errors from noise and a finite data set, however, one or both attempts to identify all the CFs may fail. If the signal under observation is only noise, there will be no correlation between selected candidate cycle-frequencies as the peaks in the CTMFE are random. Attempting to identify peaks in both CTMFEs increases the likelihood of finding a true CF. *Since only the detection problem is being considered, and not the classification problem, the only two possible signals are CPM or noise.*

The asymptotic distribution of the test statistics $\Upsilon_{2,0\mathrm{Im}}$ and $\Upsilon_{2,0\mathrm{Re}}$ at $\alpha \neq CF$ is a chi-squared distribution with two degrees of freedom; therefore, the asymptotic distribution of the final test statistic, $\Upsilon_{2,0}$, is chi-squared with four degrees of freedom. The asymptotic distribution of the test statistics at a cycle-frequency is a multivariate normal distribution with mean and variance dependent on the distribution of the data set [34].

Given the distribution of the test statistic for the noise-only case and the CPM case, a threshold Γ of statistical significance is chosen to satisfy some acceptable probability of false alarm. For example, the threshold $\Gamma = 18.47$ corresponds to a false alarm rate of 10^{-3}. Using this threshold, the test would be as follows: if $\max \Upsilon_{2,0} > \Gamma$, then the signal is declared to be CPM, otherwise it is declared to be noise only. The detection rate would depend on SNR and the properties of the CPM signal [34].

While the asymptotic distribution of the test statistic at non-cycle frequencies is well defined, it varies for finite data sets. In the absence of derived thresholds, chi-squared tables are a good starting place for selecting threshold values. Observations show, however, that in practice – in the absence of a sophisticated peak-selection algorithm for testing potential cycle frequencies – the false alarm rates are somewhat higher than what is indicated by chi-squared tables. In this work, the statistical test is applied to significantly large local maxima; heuristically derived, significantly large is taken to mean at least 4.25 times larger than the local mean.

Early testing showed that considering only the real or imaginary component yielded a lower probability of detection for the same false alarm rate. While those results are not reported here, that finding served as motivation for simultaneously using the test statistic associated with $r_{re}(t)$ and $r_{im}(t)$.

Based on the above detection procedure, probability of false alarm vs. detection curves for CPM signals based on the novel feature of second-order cyclostationarity found in the isolated real $r_{re}(t)$ and imaginary $r_{im}(t)$ components of CPM signals were found heuristically.

In the simulations, the signal is brought down to baseband and sampled at 8 times the symbol rate, $T_{sym} = 6000$ symbols per second, yielding a sampling rate T_{sam} of 48000 samples per second. The signal is received with a frequency offset Δf of $\Delta f = 0.01F_{sym}$, and a timing offset Δt of $\Delta t = 0.8T_{sam}$. The noise environment is complex additive Gaussian, and the receiving filter is large enough to preserve the bandwidth of the signal; the latter means that the receive filter does not truncate a significant portion of the highest frequency content of the signal.

The probability of detection is the probability of $\max \Upsilon_{2,0}$ exceeding the threshold Γ when $\max \Upsilon_{2,0}$ is calculated from a noisy CPM signal. The probability of false alarm is the probability of $\max \Upsilon_{2,0}$ exceeding the threshold Γ when $\max \Upsilon_{2,0}$ is calculated from noise only.

As seen in Section 5.3, the choice of the delay vector greatly impacts the magnitude of the peaks in the CTMFE at cycle frequencies. Fig. 5.4.2 illustrates the impact of delay choice on detection and false alarm rates for a CPM signal with modulation index $h = [7 \quad 10]/16$, alphabet length of 8, and SNR = -7 dB. The results presented in Fig. 5.4.2 are based on the results of 10,000 independent trials where the classifier is presented 5000 CPM signals and 5000 noise-only signals and runs the entire detection algorithm, as illustrated in Fig. 5.4.1.

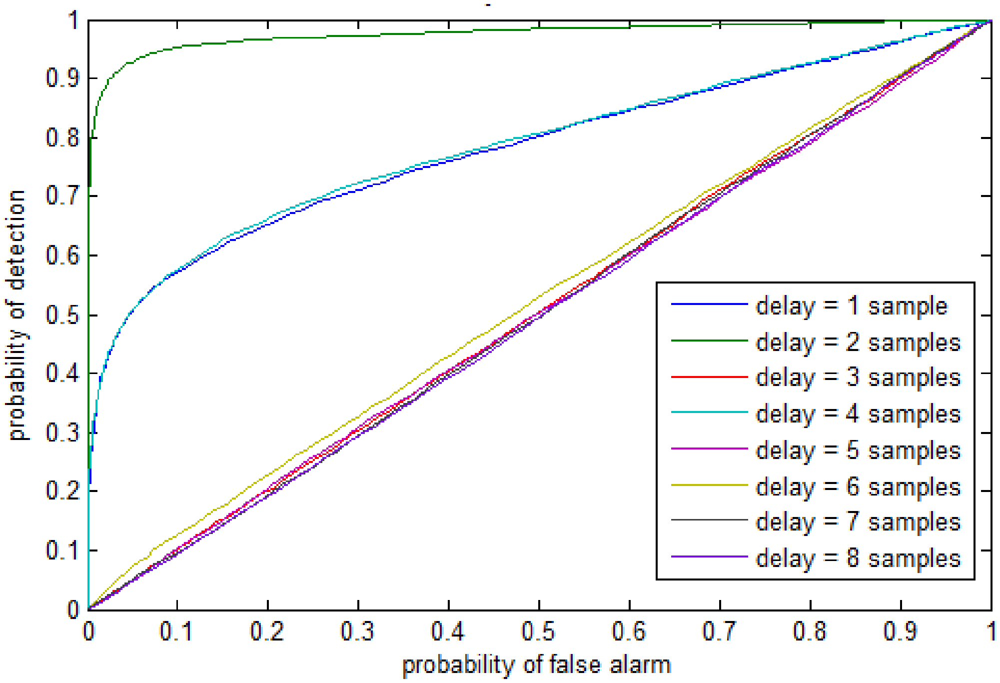


Fig. 5.4.2 Probability of detection vs. false alarm curves for a 1REC 8-ary CPM signal with $h = [7 \quad 10]/16$ over a range of delays, at SNR = -7 dB.

From Fig. 5.4.2 it is apparent that the selection of the delay is not arbitrary; in fact, the delay choice greatly impacts the ability of the detector to correctly distinguish a CPM signal from noise. When considering the zero delay vector, the search for cyclostationarity is over higher orders and possible conjugation configurations. The search for second-order cyclostationarity, when using a non-zero delay vector, is over different possible delay choices and then considering the maximum of the associated test statistics. Searching for higher orders of cyclostationarity is not only more computationally expensive, but, as shown before, has increasingly higher SNR requirements as the lowest order of cyclostationarity present increases. Overall, the search over possible delay choices is much more tractable.

Figure 5.4.3 illustrates the impact of modulation index on the probability of detection. In Fig. 5.4.3, the threshold for detection was set at 25; the corresponding observed false alarm rate was less than 0.1%. The results are based on 10,000 independent trials. These results reflect the best possible detection rate by using the optimal delay vector for each modulation configuration. Rather than performing the extensive search over all delay choices, presumably the detector has perfect knowledge of the best delay choice in each trial. The delay choice does not impact the false alarm rates since the distribution of the test statistic at non-cycle frequencies is independent of delay.

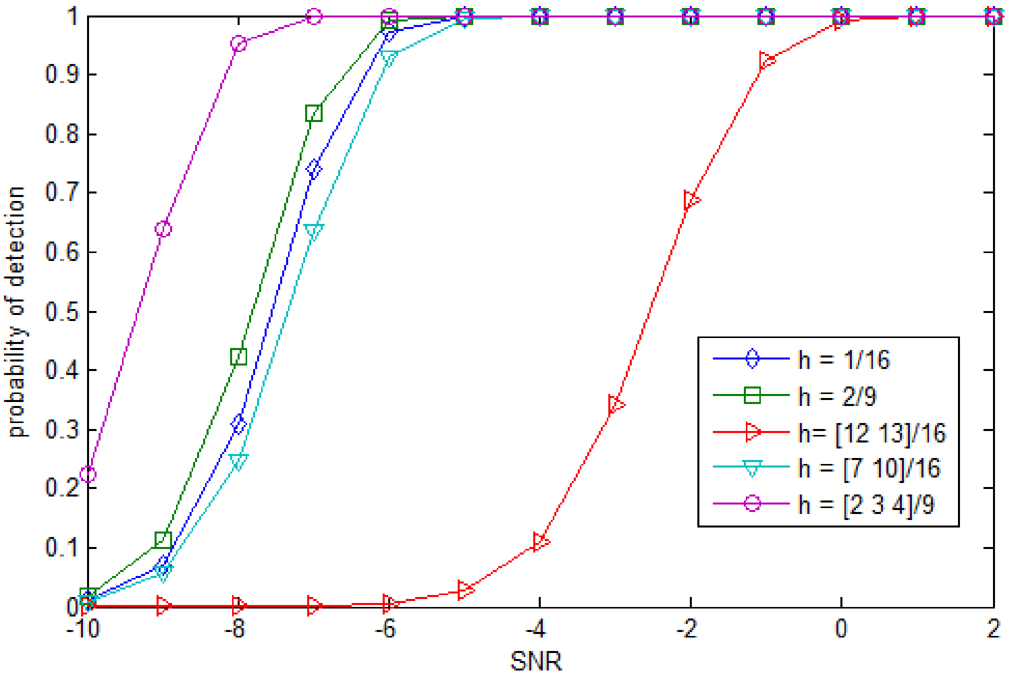


Fig. 5.4.3 Probability of detection curves at varying SNRs for different modulation index choices. For all modulation index choices, $M = 3$.

While not all complex baseband CPM signals contain the same order of cyclostationarity, a common order of cyclostationarity in the isolated real and imaginary components of the CPM signal exists for non-zero delay vectors. This novel consideration of CPM cyclostationarity has allowed the development of a feature to detect CPM signals in complex additive Gaussian noise. For some modulation indices and orders of M, reliable detection is possible at SNRs as low as -6 dB.

It is suspected that similar results could be achieved by considering the second-order first-conjugate CTMFE of the complex CPM signal with a non-zero delay vector. One possible advantage of using both the isolated real and imaginary CTMFs is the potential to use the correlation between cycle frequency selections as a further feature for detection. While this possibility was not explored here, it is certainly a potential area for further research. Additionally, while only the detection problem was considered, since the CTMFs are distinct for different modulation indices and alphabet lengths, it is likely that it is possible to extend this work to the classification problem with reasonable success.

6 Conclusion

6.1 Conclusion

In this work, a robust estimator was developed that improved the overall performance of extracting second-order first-conjugate cyclostationarity features by reducing SNR requirements, and removing the a priori assumption on knowledge of signal bandwidth [44]. Given the same observation time of one second, the robust estimator is able to achieve the same probability of false alarm and correct classification as the original estimator at 4 dB lower SNR.

In addition to reducing SNR requirements, preliminary results show that if the observation time is reduced from the original proposed observation time of one second to one-tenth of a second, the robust classifier can achieve reliable classification at SNR of 5 dB. The original classifier needed a full second of observation time to achieve reliable classification at SNR of 5 dB. In preliminary trials, the original classifier does not achieve reliable classification for the shorter observation time – of 1/10 s – even when SNR is as high as 15 dB.

The robust estimator is also less sensitive to changes in the occupied signal bandwidth. We showed that as long as the receiving filter is at least the bandwidth of the incoming signal, the robust estimator performs roughly the same under varying receiver bandwidths. The original classifier, however, suffers performance losses when the noise bandwidth exceeds the signal bandwidth.

Since low-order cyclostationarity is a promising feature for detection and classification, this work also focused on developing mathematical expressions for second-order cyclic moments of complex CPM signals, as well as for the isolated real and imaginary components. A specific application example of blind asynchronous detection demonstrated the success of using this novel feature for reliable detection [45]. Additionally, the work done here provides a framework for further investigating the possibility of using second order cyclostationarity features for CPM classification.

6.2 Suggestions for Further Research

Using robust estimation techniques improved classification based on second-order first-conjugate cyclostationarity. A further consideration is to extend the work done here to develop robust estimators of other choices of *n*-order *q*-conjugate cyclostationarity. Higher–order statistics typically have requirements for high SNR and long observation times. Based on the improvements seen in this work, robust estimation techniques may offer a promising path towards alleviating the severe requirements on using reliable high–order statistic feature extraction for detection and classification.

Based on the promising results found in this work, another area of further research is to investigate the feasibility of using second-order cyclostationarity features of CPM signals for classification. Robust estimation methods could potentially be included when estimating these features to further improve the reliability of low order cyclostationarity feature-based detection and classification of CPM.

References

[1] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Survey of automatic modulation classification techniques: classical approaches and new trends," Communications, IET, vol. 1, no. 2, pp. 137-156, April 2007. doi: 10.1049/iet-com:20050176.

[2] W. Wei and J. M. Mendel, “Maximum-likelihood classification for digital amplitude-phase modulations,” IEEE Trans. Commun., vol. 48, pp. 189-193, 2000.

[3] J. A. Sills, “Maximum-likelihood modulation classification for PSK/QAM,” in Proc. IEEE MILCOM, 1999, pp. 57-61.

[4] D. Grimaldi, S. Rapuano, and L. De Vito, "An Automatic Digital Modulation Classifier for Measurement on Telecommunication Networks," Instrumentation and Measurement, IEEE Transactions on, vol. 56, no. 5, pp. 1711-1720, Oct. 2007.

[5] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, “The classification of joint analog and digital modulations,” in Proc. IEEE MILCOM, 2005, pp. 1-6.

[6] G. Arulampalam, V. Ramakonar, A. Bouzerdoum, and D. Habibi, "Classification of digital modulation schemes using neural networks," Proceedings of the Fifth International Symposium on Signal Processing and Its Applications (ISSPA '99), vol. 2, pp. 649-652, vol. 2, 1999.

[7] A. K. Nandi and E. E. Azzouz, “Modulation recognition using artificial neural networks,” Signal Processing, pp. 165-175, 1997.

[8] A. Hossen, F. Al-Wadahi, and A. J. Joseph, “Classification of modulation signals using statistical signal characterization and artificial neural networks,” Eng. Appli. Artif. Intell., 2007, doi: 10.1016/j.engappai.2006.08.004.

[9] M. R. Mirarab and M. A. Sobhani, "Robust modulation classification for PSK /QAM/ASK using higher-order cumulants," 6th International Conference on Information, Communications & Signal Processing, 2007, pp.1-4, 10-13 Dec. 2007.

[10] A. K. Nandi and E. E. Azzouz. "Algorithms for Automatic Modulation Recognition of Communication Signals." IEEE Transactions on Communications Vol. 46. No. 4, April 1998.

[11] O. A. Dobre, Y. Bar-Ness, and S. Wei, "Higher-order cyclic cumulants for high order modulation classification," Military Communications Conference (MILCOM 2003), IEEE, vol. 1, pp. 112-117, Vol. 1, 13-16 Oct. 2003.

[12] A. Punchihewa, O. A. Dobre, S. Rajan, and R. Inkol, "Cyclostationarity-based Algorithm for Blind Recognition of OFDM and Single Carrier Linear Digital Modulations," *IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications, (PIMRC 2007)*, pp. 1-5, 3-7 Sept. 2007.

[13] A. Swami and B. M. Sadler, "Hierarchical digital modulation classification using cumulants," IEEE Trans. Commun., vol. 48, pp.416-429, 2000.

[14] L. Hong and K. C. Ho, "Identification of digital modulation types using the wavelet transform," Military Communications Conference (MILCOM 1999) Proceedings. IEEE, vol. 1, pp. 427-431, 1999.

[15] W. Dan, G. Xuemai, and G. Qing, "A new scheme of automatic modulation classification using wavelet and WSVM," 2nd International Conference on Mobile Technology, Applications and Systems, pp. 15-17, Nov. 2005.

[16] K. C. Ho, W. Prokopiw, and Y. T. Chan, "Modulation identification by the wavelet transform," Military Communications Conference (MILCOM '95), Conference Record, IEEE , vol. 2, pp. 886-890, 7 Nov 1995.

[17] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Cyclostationarity-Based Blind Classification of Analog and Digital Modulations," Military Communications Conference, (MILCOM 2006), IEEE , pp. 1-7, 23-25 Oct. 2006.

[18] O. A. Dobre, Q. Zhang, S. Rajan, and R. Inkol, “Second-order cyclostationarity of cyclically prefixed single carrier linear digital modulations with applications to signal recognition,” pp. 1-5. Proc. IEEE GLOBECOM, 2008.

[19] Q. Zhang, O. A. Dobre, S. Rajan, and R. Inkol, "On the second-order cyclostationarity for joint signal detection and classification in cognitive radio systems," Canadian Conference on Electrical and Computer Engineering, 2009. (CCECE '09). pp. 204-208, 3-6 May 2009.

[20] A. Fehske, J. Gaeddert, and J. H. Reed, "A new approach to signal classification using spectral correlation and neural networks," First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN 2005), pp. 144-150, 8-11 Nov. 2005.

[21] C. M. Spooner, “Classification of cochannel communication signal using cyclic cumulants”, Proc. Asilomar, pp. 531-536, 1995.

[22] W. A. Gardner, A. Napolitano, and L. Paura, “Cyclostationarity: Half a century of research, Signal Processing”, Vol. 86, Issue 4, pp. 639-697, April 2006, ISSN 0165-1684, doi: 10.1016/j.sigpro.2005.06.016.

[23] S. Srinivasa and S. A. Jafar, "Cognitive radios for dynamic spectrum access - The Throughput Potential of Cognitive Radio: A Theoretical Perspective," Communications Magazine, IEEE, vol. 45, no. 5, pp. 73-79, May 2007, doi: 10.1109/MCOM.2007.358852.

[24] J. Mitola III, "Cognitive radio for flexible mobile multimedia communications," (MoMuC '99) 1999 IEEE International Workshop on Mobile Multimedia Communications, 1999. pp. 3-10, 1999, doi: 10.1109/MOMUC.1999.819467.

[25] S. Geirhofer, L.Tong, and B. M. Sadler, " Cognitive radios for dynamic spectrum access - Dynamic Spectrum Access in the Time Domain: Modeling and Exploiting White Space," Communications Magazine, IEEE , vol. 45, no. 5, pp. 66-72, May 2007, doi: 10.1109/MCOM.2007.358851.

[26] A. Ghasemi and E. S. Sousa, "Spectrum sensing in cognitive radio networks: requirements, challenges and design trade-offs," Communications Magazine, IEEE, vol. 46, no. 4, pp. 32-39, April 2008, doi: 10.1109/MCOM.2008.4481338.

[27] S. Haykin and M. Moher, Introduction to Analog & Digital Communications, 2nd ed., Wiley, 2007.

[28] T. Biedka, L. Mili, and J. H. Reed, “Robust estimation of cyclic correlation in contaminated Gaussian noise”, Proc. 29th Asilomar Conference on Signals, Systems and Computers, pp. 511-515, Pacific Grove, CA, 1995.

[29] A. Napolitano and C. M. Spooner, “Cyclic Spectral Analysis of Continuous-Phase Modulation Signals”, IEEE Trans. Signal Process, pp. 30-44, 2001.

[30] M. K. Nezami and B. Peterson. “Performance of efficient tactical UHF-SATCOM waveforms: Occupied bandwidth, coding gain, spectral efficiency, bit error rate, and adjacent channel interference”, Proc IEEE MILCOM , pp. 152-159, 2002.

[31] A. Svensson, and C. Sundberg, "Serial MSK-Type Detection of Partial Response Continuous Phase Modulation," IEEE Transactions on Communications, vol. 33, no. 1, pp. 44-52, Jan 1985.

[32] G. de Veciana and A. Zakhor, "Neural net-based continuous phase modulation receivers," IEEE Transactions on Communications, vol. 40, no. 8, pp. 1396-1408, Aug 1992. doi: 10.1109/26.156644.

[33] W. A. Gardner, Cyclostationarity in Communications and Signal Processing. New York: IEEE Press, 1993.

[34] A. V. Dandawade and G. B. Giannakis, "Statistical tests for presence of cyclostationarity," IEEE Trans. Signal Process., vol. 42, pp. 2355-2369, 1994. doi: 10.1109/78.317857.

[35] A. V. Dandawate and G. B. Giannakis, "Nonparametric polyspectral estimators for kth-order (almost) cyclostationary processes," IEEE Transactions on Information Theory, vol. 40, no. 1, pp. 67-84, Jan 1994. doi: 10.1109/18.272456.

[36] T. Aulin and C. Sundberg, "Continuous Phase Modulation--Part I: Full Response Signaling," IEEE Transactions on Communications, vol. 29, no. 3, pp. 196- 209, Mar 1981. doi: 10.1109/TCOM.1981.1095001.

[37] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, Robust Statistics: The Approach Based on Influence Functions. New York: Wiley, 1986.

[38] R. A. Maronna, R. D. Martin, and V. J. Yohai, Robust Statistics: Theory and Methods, New York: Wiley, 2006.

[39] R. G. Staudte, and S. J. Sheather, Robust Estimation and Testing. New York: JohnWiley and Sons, 1990.

[40] T. J. Rothenberg, F. M. Fisher, and C. B. Tilanus, “A Note on Estimation from a Cauchy Sample”, Journal of the American Statistical Association, vol. 59, June 1964.

[42] P. J. Huber. Robust Statistical Procedures. Philadelphia, PA: SIAM Press, 1997.

[43] A. Leon-Garcia, Probability and Random Processes for Electrical Engineering. Don Mills, Ont., Canada: Addison-Wesley, 1989.

[44] A. C. Malady and A. A. (Louis) Beex, "AMC Improvements from Robust Estimation", in Proc. GLOBECOM, pp.1-5, 2010.

[45] A. C. Malady, and A. A. (Louis) Beex, “Detection of CPM Based on Second-Order Cyclostationarity,” in Proc. Asilomar, pp. 501-505, 2010.

Appendix

The second-order first-conjugate lag product of a complex CPM signal, as given in (2.4.4), with delay vector $\tau = \begin{bmatrix} 0 & \Delta t \end{bmatrix}$ when $0 < \Delta t < T$ is given in (A.1). For compactness, the following is defined $\phi_k \triangleq \rho_k h_{\langle k \rangle_H}$.

$$L_s(t,\boldsymbol{\tau})_{2.1} = \begin{cases} \sum_k e^{j\left(2\pi\phi_k\left[\frac{t-kT}{2T}\right]+\pi\sum_{i=1}^{k-1}\phi_i\right)} e^{-j\left(2\pi\phi_{k-1}\left[\frac{t-kT+(T-\Delta t)}{2T}\right]+\pi\sum_{i=1}^{k-2}\phi_i\right)} & 0 \le t < \Delta t \\ \sum_k e^{j\left(2\pi\phi_k\left[\frac{t-kT}{2T}\right]+\pi\sum_{i=1}^{k-1}\phi_i\right)} e^{-j\left(2\pi\phi_k\left[\frac{t-kT+\Delta t}{2T}\right]+\pi\sum_{i=1}^{k-1}\phi_i\right)} & \Delta t \le t < T \end{cases} \tag{A.1}$$

After algebraic manipulation, (A.1) is written more compactly as (A.2).

$$L_s(t,\boldsymbol{\tau})_{2,1} = \begin{cases} \sum_k e^{j\left(2\pi(\phi_k-\phi_{k-1})\left[\frac{t-kT}{2T}\right]+\frac{\pi\Delta t\phi_{k-1}}{T}\right)} & 0 \le t < \Delta t \\ \sum_k e^{-j\pi\left(\frac{\Delta t\phi_k}{T}\right)} & \Delta t \le t < T \end{cases} \tag{A.2}$$

The limited memory lag product is still continuous during a symbol period. The lag product is constant over $\Delta t \le t < T$. The lag product is also continuous at symbol changes. The following rectangular pulses *f*(*t*) and *g*(*t*) are defined

$$\begin{aligned} f(t) &\triangleq \begin{cases} 1 & 0 \le t < \Delta t \\ 0 & otherwise \end{cases} \\ g(t) &\triangleq \begin{cases} 1 & \Delta t \le t < T \\ 0 & otherwise \end{cases} \end{aligned} \tag{A.3}$$

Using the definitions of *f*(*t*) and *g*(*t*) (A.2) can be written more compactly as (A.4).

$$L_s(t,\tau)_{2,1} = \sum_k f(t-kT) e^{j\left(2\pi(\phi_k - \phi_{k-1})\left[\frac{t-kT}{2T}\right] + \frac{\pi \Delta t \phi_{k-1}}{T}\right)} + \sum_k g(t-kT) e^{-j\pi\left(\frac{\Delta t \phi_k}{T}\right)} \tag{A.4}$$

The second-order first-conjugate CTMF of the CPM signal with delay vector $\boldsymbol{\tau} = \begin{bmatrix} 0 & \Delta t \end{bmatrix}$ is found by taking the Fourier transform of the lag product, see (A.5).

$$\begin{aligned} R_s^{\alpha}(\boldsymbol{\tau})_{2,1} &= \int \sum_k f(t-kT) e^{j\pi\left((\phi_k - \phi_{k-1})\left[\frac{t-kT}{T}\right] + \frac{\Delta t \phi_{k-1}}{T}\right)} e^{-j2\pi\alpha t} dt \\ &\quad + \int \sum_k g(t-kT) e^{-j\pi\left(\frac{\Delta t \phi_k}{T}\right)} e^{-j2\pi\alpha t} dt \end{aligned} \tag{A.5}$$

We can remove the dependence of the symbols ρ_k and modulation index, on *k*, by considering that the symbols are equiprobable and from a finite set and that the modulation indices are cyclical. As such, the integration will evaluate the same if instead we consider the average value taken by the lag product during each of the *k* symbol periods. Since the modulation indices are cyclic, the average value is cyclic. As the modulation indices are cyclically varying, let $k = lH + q$, where *l* and *q* are integers. The average is taken over the equally likely 2^M values of ρ_k for each of the *H* possible values of $h_{\langle k \rangle_H}$. Equation (A.5) can be rewritten as follows

$$R_s^{\alpha}(\boldsymbol{\tau})_{2,1} = C_1 \int A e^{-j2\pi\alpha t} dt + C_2 \int B e^{-j2\pi\alpha t} dt \tag{A.6}$$

where, for ease of presentation, we let $C_1 = \frac{1}{2^{M+1}}, C_2 = \frac{1}{2^M}$, $\theta_{m,n,q} = \left(\rho_m h_{\langle q \rangle_H} - \rho_n h_{\langle q-1 \rangle_H}\right)$, and $\phi_{m,q} = \rho_m h_{\langle q \rangle_H}$, so that in addition

$$\begin{aligned} A &= \sum_l \sum_{m=0}^{2^M-1} \sum_{n=0}^{2^M-1} \sum_{q=1}^{H} f(t-qT-lHT) e^{j\pi\left(\theta_{m,n,q}\left[\frac{t-qT-lHT}{T}\right] + \frac{\Delta t \phi_{n,q-1}}{T}\right)} \\ B &= \sum_l \sum_{m=0}^{2^M-1} \sum_{q=1}^{H} g(t-qT-lHT) e^{-j\pi\frac{\Delta t \phi_{m,q}}{T}} \end{aligned} \tag{A.7}$$

The argument of $f(\bullet)$ and $g(\bullet)$ is now $t-qT-lHT$ instead of $t - kT$ to account for the cyclically varying modulation index. Since all the expressions of time are based on the argument $t-qT-lHT$, (A.6) can be rewritten in terms of convolution, which results in (A.8).

$$\begin{aligned} R_s^{\alpha}(\boldsymbol{\tau})_{2,1} &= C_1 \int \sum_{m=0}^{2^M-1} \sum_{n=0}^{2^M-1} \sum_{q=1}^{H} \left(f(t) e^{j\pi\left(\frac{\theta_{m,n,q}}{T}t + \frac{\Delta t\phi_{n,q-1}}{T}\right)} * \sum_{l} \delta\left(t-qT-lHT\right) \right) e^{-j2\pi\alpha t} dt \\ &+ C_2 \int \sum_{m=0}^{2^M-1} \sum_{q=1}^{H} \left(g(t) e^{-j\pi\frac{\Delta t\phi_{m,q}}{T}} * \sum_{l} \delta\left(t-qT-lHT\right) \right) e^{-j2\pi\alpha t} dt \end{aligned} \tag{A.8}$$

where $*$ represents the convolution operator and $\delta(t)$ the Dirac delta function. Next performing the convolution, leads to (A.9).

$$\begin{aligned} R_s^{\alpha}(\boldsymbol{\tau})_{2,1} &= C_1 \int \sum_{m=0}^{2^M-1} \sum_{n=0}^{2^M-1} \sum_{q=1}^{H} \int \left(f(v) e^{j\pi\left(\frac{\theta_{m,n,q}}{T}v + \frac{\Delta t\phi_{n,q-1}}{T}\right)} \sum_{l} \delta\left((t-v)-qT-lHT\right) \right) dv e^{-j2\pi\alpha t} dt \\ &+ C_2 \int \sum_{m=0}^{2^M-1} \sum_{q=1}^{H} \int \left(g(v) e^{-j\pi\frac{\Delta t\phi_{m,q}}{T}} \sum_{l} \delta\left((t-v)-qT-lHT\right) \right) dv e^{-j2\pi\alpha t} dt \end{aligned} \tag{A.9}$$

The order of integration can be swapped because of the convergence properties of the double integral. Define $u=t-v$ and make the substitution in (A.9) to arrive at (A.10).

$$\begin{aligned} R_s^{\alpha}(\boldsymbol{\tau})_{2,1} &= C_1 \int \sum_{m=0}^{2^M-1} \sum_{n=0}^{2^M-1} \sum_{q=1}^{H} \frac{f(v) e^{j\pi\left(\frac{\theta_{m,n,q}}{T}v + \frac{\Delta t\phi_{n,q-1}}{T}\right)}}{e^{j2\pi\alpha v}} \int \sum_{l} \delta\left(u-qT-lHT\right) e^{-j2\pi\alpha u} du dv \\ &+ C_2 \int \sum_{m=0}^{2^M-1} \sum_{q=1}^{H} \frac{g(v) e^{-j\pi\frac{\Delta t\phi_{m,q}}{T}}}{e^{j2\pi\alpha v}} \int \sum_{l} \delta\left(u-qT-lHT\right) e^{-j2\pi\alpha u} du dv \end{aligned} \tag{A.10}$$

Note that the inner integral is the Fourier transform of a time shifted train of impulses or Dirac comb. We can use properties of the Fourier transform (a time-domain Dirac comb transforms into a frequency-domain Dirac comb) to solve the integral in (A.10) and arrive at (A.11).

$$
\begin{aligned}
R_s^{\alpha}(\boldsymbol{\tau})_{2,1} &= C_1 \int \sum_{m=0}^{2^M-1} \sum_{n=0}^{2^M-1} \sum_{q=1}^{H} \frac{f(v)e^{j\pi\left(\frac{\theta_{m,n,q}}{T}v+\frac{\Delta t\phi_{n,q-1}}{T}\right)}}{e^{j2\pi\alpha(v+qT)}} \frac{1}{HT}\sum_{l}\delta\left(\alpha-\frac{l}{HT}\right)dv \\
&+ C_2 \int \sum_{m=0}^{2^M-1} \sum_{q=1}^{H} \frac{g(v)e^{-j\pi\frac{\Delta t\phi_{m,q}}{T}}}{e^{j2\pi\alpha(v+qT)}} \frac{1}{HT}\sum_{l}\delta\left(\alpha-\frac{l}{HT}\right)dv
\end{aligned}
\qquad \text{(A.11)}
$$

For brevity, redefine $\tilde{C}_1 = \frac{1}{HT2^{M+1}}$, and $\tilde{C}_2 = \frac{1}{HT2^{M}}$. Next, move only the expressions that are a function of the variable of integration inside the integrand and swap the order of summation and integration to arrive at (A.12).

$$
\begin{aligned}
R_s^{\alpha}(\boldsymbol{\tau})_{2,1} &= \tilde{C}_1 \sum_{m=0}^{2^M-1} \sum_{n=0}^{2^M-1} \sum_{q=1}^{H} \frac{e^{j\pi\left(\frac{\Delta t\phi_{n,q-1}}{T}\right)}}{e^{j2\pi\alpha qT}} \sum_{l}\delta\left(\alpha-\frac{l}{HT}\right)\int f(v)e^{j2\pi\left(\frac{\theta_{m,n,q}}{2T}-\alpha\right)v}dv \\
&+ \tilde{C}_2 \sum_{m=0}^{2^M-1} \sum_{q=1}^{H} \frac{e^{-j\pi\frac{\Delta t\phi_{m,q}}{T}}}{e^{j2\pi\alpha qT}} \sum_{l}\delta\left(\alpha-\frac{l}{HT}\right)\int g(v)e^{-j2\pi\alpha v}dv
\end{aligned}
\qquad \text{(A.12)}
$$

Since the functions $f(t)$ and $g(t)$ are nonzero over finite ranges, the limits of integration can be changed to reflect the interval over which $f(t)$ and $g(t)$ are nonzero to arrive at (A.13)

$$
\begin{aligned}
R_s^{\alpha}(\boldsymbol{\tau})_{2,1} &= \tilde{C}_1 \sum_{m=0}^{2^M-1} \sum_{n=0}^{2^M-1} \sum_{q=1}^{H} \frac{e^{j\pi\left(\frac{\Delta t\phi_{n,q-1}}{T}\right)}}{e^{j2\pi\alpha qT}} \sum_{l}\delta\left(\alpha-\frac{l}{HT}\right)\int_{0}^{\Delta t} e^{j2\pi\left(\frac{\theta_{m,n,q}}{2T}-\alpha\right)v}dv \\
&+ \tilde{C}_2 \sum_{m=0}^{2^M-1} \sum_{q=1}^{H} \frac{e^{-j\pi\frac{\Delta t\phi_{m,q}}{T}}}{e^{j2\pi\alpha qT}} \sum_{l}\delta\left(\alpha-\frac{l}{HT}\right)\int_{\Delta t}^{T} e^{-j2\pi\alpha v}dv
\end{aligned}
\qquad \text{(A.13)}
$$

Also, because of the dirac delta function, we know that $R_s^{\alpha}(\boldsymbol{\tau})_{2,1}$ is nonzero only when $\alpha = l\left(HT\right)^{-1}$ where l is an integer. We can apply this substitution to arrive at (A.14).

$$\begin{aligned} R_s^{\alpha}(\tau)_{2,1} &= \tilde{C}_1 \sum_{m=0}^{2^M-1} \sum_{n=0}^{2^M-1} \sum_{q=1}^{H} e^{j\pi\left(\frac{\Delta t \phi_{n,q-1}}{T}\right)} \sum_{l} e^{-j\frac{2\pi q l}{H}} \delta\left(\alpha - \frac{l}{HT}\right) \int_0^{\Delta t} e^{j\frac{2\pi}{T}\left(\frac{\theta_{m,n,q}}{2} - \frac{l}{H}\right)v} dv \\ &+ \tilde{C}_2 \sum_{m=0}^{2^M-1} \sum_{q=1}^{H} e^{-j\pi\frac{\Delta t \phi_{m,q}}{T}} \sum_{l} e^{-j\frac{2\pi q l}{H}} \delta\left(\alpha - \frac{l}{HT}\right) \int_{\Delta t}^{T} e^{-j\frac{2\pi l}{HT}v} dv \end{aligned} \tag{A.14}$$

After performing integration, we arrive at the final expression for the second-order first-conjugate CTMF of the CPM signal in (A.15).

$$\begin{aligned} R_s^{\alpha}(\tau)_{2,1} &= \tilde{C}_1 \sum_{m=0}^{2^M-1} \sum_{n=0}^{2^M-1} \sum_{q=1}^{H} e^{j\pi\left(\frac{\Delta t \phi_{n,q-1}}{T}\right)} \sum_{l} \delta\left(\alpha - \frac{l}{HT}\right) \left(\frac{e^{\frac{j2\pi\Delta t}{T}\left(\frac{\theta_{m,n,q}}{2} - \frac{l}{H}\right)} - 1}{\frac{j2\pi}{T}\left(\frac{\theta_{m,n,q}}{2} - \frac{l}{H}\right) e^{j\frac{2\pi q l}{H}}} \right) \\ &+ \tilde{C}_2 \sum_{m=0}^{2^M-1} \sum_{q=1}^{H} e^{-j\pi\frac{\Delta t \phi_{m,q}}{T}} \sum_{l} \delta\left(\alpha - \frac{l}{HT}\right) \left(\frac{e^{-j\frac{2\pi l}{H}} - e^{-j\frac{2\pi l \Delta t}{HT}}}{-\frac{j2\pi l}{HT} e^{j\frac{2\pi q l}{H}}} \right) \end{aligned} \tag{A.15}$$

We have established that the second-order first-conjugate CTMF is discrete in α with possible cycle frequencies at $\alpha = l\left(HT\right)^{-1}$ where l is an integer. The magnitudes of these discrete spectral components are a function of the modulation indices, the symbol alphabet, and the integer l.

The following definitions in (A.16a-c) and (A.17a-c) are useful for deriving the second-order zero-conjugate cyclostationarity of the isolated real and imaginary components of the signal from the second-order first-conjugate CTMF of the complex CPM signal.

$$\begin{aligned} L_s(t,\tau)_{2,1} &= \exp(j \arg_0)\exp(-j \arg_{\Delta t}) \\ &= \exp\left(j\left(\arg_0 - \arg_{\Delta t}\right)\right) \\ &= \cos\left(\arg_0 - \arg_{\Delta t}\right) + j\sin\left(\arg_0 - \arg_{\Delta t}\right) \end{aligned} \tag{A.16a}$$

$$L_{real(s)}(t,\boldsymbol{\tau})_{2,0} = \cos(\mathrm{arg}_0)\cos(\mathrm{arg}_{\Delta t}) \tag{A.16b}$$

$$L_{imag(s)}(t,\boldsymbol{\tau})_{2,0} = \sin(\mathrm{arg}_0)\sin(\mathrm{arg}_{\Delta t}) \tag{A.16c}$$

$$R_s^{\alpha}(\boldsymbol{\tau})_{2,1} = \mathscr{F}\left(L_s(t,\boldsymbol{\tau})_{2,1}\right) \tag{A.17a}$$

$$R_{real(s)}^{\alpha}(\boldsymbol{\tau})_{2,0} = \mathscr{F}\left(L_{real(s)}(t,\boldsymbol{\tau})_{2,0}\right) \tag{A.17b}$$

$$R_{imag(s)}^{\alpha}(\boldsymbol{\tau})_{2,1} = \mathscr{F}\left(L_{imag(s)}(t,\boldsymbol{\tau})_{2,0}\right) \tag{A.17c}$$

The steps shown in (A.18) illustrate how expressions for $R_{real(s)}^{\alpha}(\alpha,\boldsymbol{\tau})_{2,0}$ and $R_{imag(s)}^{\alpha}(\alpha,\boldsymbol{\tau})_{2,0}$ arise from $R_s^{\alpha}(\alpha,\boldsymbol{\tau})_{2,1}$. To go from line three to line 4, apply the sum-difference trigonometric identity which relates $\cos(A-B) = \cos(A)\cos(B) + \sin(A)\sin(B)$. To go from line 4 to line 5, consider that the Fourier transform satisfies $\mathscr{F}\left(A+B+C\right) = \mathscr{F}\left(A\right) + \mathscr{F}\left(B\right) + \mathscr{F}\left(C\right)$, and use the definitions in (A.16b) and (A.16a).

$$\begin{aligned}
R_s^{\alpha}(\boldsymbol{\tau})_{2,1} &= \mathscr{F}\left(L_s(t,\boldsymbol{\tau})_{2,1}\right) \\
&= \mathscr{F}\left(re(L_s(t,\boldsymbol{\tau})_{2,1}) + jim(L_s(t,\boldsymbol{\tau})_{2,1})\right) \\
&= \mathscr{F}\left(\cos(\mathrm{arg}_0 - \mathrm{arg}_{\Delta t}) + j\left(\sin(\mathrm{arg}_0 - \mathrm{arg}_{\Delta t})\right)\right) \\
&= \mathscr{F}\left(\cos(\mathrm{arg}_0)\cos(\mathrm{arg}_{\Delta t}) + \sin(\mathrm{arg}_0)\sin(\mathrm{arg}_{\Delta t}) + j\left(\sin(\mathrm{arg}_0 - \mathrm{arg}_{\Delta t})\right)\right) \\
&= \mathscr{F}\left(L_{real(s)}(t,\boldsymbol{\tau})_{2,0}\right) + \mathscr{F}\left(L_{imag(s)}(t,\boldsymbol{\tau})_{2,0}\right) + \mathscr{F}\left(j\left(\sin(\mathrm{arg}_0 - \mathrm{arg}_{\Delta t})\right)\right)
\end{aligned} \tag{A.18}$$

The expression for $\Im\left(j\left(\sin(\mathrm{arg}_0 - \mathrm{arg}_{\Delta t})\right)\right)$ is equivalent, by Euler's identity, to $\mathscr{F}\left(\dfrac{e^{j(\mathrm{arg}_0 - \mathrm{arg}_{\Delta t})} - e^{-j(\mathrm{arg}_0 - \mathrm{arg}_{\Delta t})}}{2}\right)$. The expression for $\mathscr{F}\left(L_s(t,\boldsymbol{\tau})_{2,1} = e^{j(\mathrm{arg}_0 - \mathrm{arg}_{\Delta t})}\right)$ was already derived, (A.15). Note that the term $e^{-j(\mathrm{arg}_0 - \mathrm{arg}_{\Delta t})}$ is the complex conjugate of $e^{j(\mathrm{arg}_0 - \mathrm{arg}_{\Delta t})}$. The expression for

$L_{s*}(t,\boldsymbol{\tau})_{2,1} = e^{-j(\arg_0 - \arg_{\Delta t})}$ is given in (A.19) and differs from (A.2) only in that the argument is negated, so that

$$L_{s*}(t,\boldsymbol{\tau})_{2,1} = \begin{cases} \sum_k e^{-j\left(2\pi(\phi_k - \phi_{k-1})\left[\frac{t-kT}{2T}\right] + \frac{\pi\Delta t\phi_{k-1}}{T}\right)} & 0 \le t < \Delta t \\ \sum_k e^{j\pi\left(\frac{\Delta t\phi_k}{T}\right)} & \Delta t \le t < T \end{cases} \tag{A.19}$$

Without providing every step of the derivation, but by following the same steps shown in (A.3) – (A.15), the expression for $\mathscr{F}\left(L_{s*}(t,\boldsymbol{\tau})_{2,1}\right)$ is given in (A.20) . The expression can be directly found from (A.13), by applying the Fourier transform identity that if $\mathscr{F}\left(x(t)\right) = X(f)$ then $\mathscr{F}\left(x^*(t)\right) = \left[X(-f)\right]^*$. Under the summation over *l*, only $\alpha = -l(HT)^{-1}$ produces a non-zero result.

$$\begin{aligned} \mathscr{F}\left(L_{s*}(t,\boldsymbol{\tau})_{2,1}\right) &= \left[R_s^{-\alpha}(\boldsymbol{\tau})_{2,1}\right]^* \\ &= \tilde{C}_1 \sum_{m=0}^{2^M-1} \sum_{n=0}^{2^M-1} \sum_{q=1}^{H} e^{-j\pi\left(\frac{\Delta t\phi_{n,q-1}}{T}\right)} \sum_l \delta\left(-\alpha - \frac{l}{HT}\right) \left(\frac{e^{-\frac{j2\pi\Delta t}{T}\left(\frac{\theta_{m,n,q}}{2} - \frac{(-l)}{H}\right)} - 1}{-\frac{j2\pi}{T}\left(\frac{\theta_{m,n,q}}{2} - \frac{(-l)}{H}\right) e^{-j\frac{2\pi q(-l)}{H}}} \right) \\ &\quad + \tilde{C}_2 \sum_{m=0}^{2^M-1} \sum_{q=1}^{H} e^{j\pi\frac{\Delta t\phi_{m,q}}{T}} \sum_l \delta\left(-\alpha - \frac{l}{HT}\right) \left(\frac{e^{j\frac{2\pi(-l)}{H}} - e^{j\frac{2\pi(-l)\Delta t}{HT}}}{\frac{j2\pi(-l)}{HT} e^{-j\frac{2\pi q(-l)}{H}}} \right) \end{aligned} \tag{A.20}$$

Apply a change of variables $\tilde{\alpha} = -\alpha$ and distribute the negative signs to arrive at (A.21).

$$\mathscr{F}\left(L_{s*}(t,\boldsymbol{\tau})_{2,1}\right)=\tilde{C}_1\sum_{m=0}^{2^M-1}\sum_{n=0}^{2^M-1}\sum_{q=1}^{H}e^{-j\pi\left(\frac{\Delta t\phi_{n,q-1}}{T}\right)}\sum_{l}\delta\left(\tilde{\alpha}-\frac{l}{HT}\right)\left(\frac{e^{\frac{j2\pi\Delta t}{T}\left(\frac{-\theta_{m,n,q}}{2}-\frac{l}{H}\right)}-1}{\frac{j2\pi}{T}\left(\frac{-\theta_{m,n,q}}{2}-\frac{l}{H}\right)e^{j\frac{2\pi ql}{H}}}\right)$$

$$+\tilde{C}_2\sum_{m=0}^{2^M-1}\sum_{q=1}^{H}e^{j\pi\frac{\Delta t\phi_{m,q}}{T}}\sum_{l}\delta\left(\tilde{\alpha}-\frac{l}{HT}\right)\left(\frac{e^{-j\frac{2\pi l}{H}}-e^{-j\frac{2\pi l\Delta t}{HT}}}{-\frac{j2\pi l}{HT}e^{j\frac{2\pi ql}{H}}}\right) \quad \text{(A.21)}$$

The expression for the CTMF of the complex conjugate of the lag product differs from the CTMF of the lag product only in that all of the variables that depend on the symbol alphabet, ρ, are negated: $\phi_{m,q}$ and $\theta_{m,n,q}$. Recall that the symbol alphabet is always of the form $\rho=\pm1,3,...,M-1$; therefore, $\rho_m=-\rho_{M-m-1}$ for $m=0,1,...,2^{M-1}$. Note that in the final CTMF, the variables dependent on the symbol alphabet are only considered with respect to their average value taken over $m=0,1,...,2^{M-1}$, and because $\rho_m=-\rho_{M-m-1}$ for $m=0,1,...,2^{M-1}$, the average of ρ_m equals the average of $-\rho_m$. As a result, for the CPM signals of interest, the following in (A.22) holds,

$$\mathscr{F}\left(L_s(t,\boldsymbol{\tau})_{2,1}=e^{j(\arg_0-\arg_{\Delta t})}\right)=\mathscr{F}\left(L_{s*}(t,\boldsymbol{\tau})_{2,1}=e^{-j(\arg_0-\arg_{\Delta t})}\right) \quad \text{(A.22)}$$

Using the relationship in (A.22) and Euler's identity, we conclude the relationship in (A.23) holds for the CPM signals of interest.

$$\mathscr{F}\left(j\left(\sin(\arg_0-\arg_{\Delta t})\right)\right)=0 \quad \text{(A.23)}$$

Finally, we conclude using (A.18) and (A.23) that the relationship in (A.24) holds for CPM signals.

$$R_s^{\alpha}(\alpha,\boldsymbol{\tau})_{2,1}=R_{real(s)}^{\alpha}(\boldsymbol{\tau})_{2,0}+R_{imag(s)}^{\alpha}(\boldsymbol{\tau})_{2,0} \quad \text{(A.24)}$$

Since the CPM signals of interest are constant envelope and circular symmetric, the observed relationship in (A.25), consistently holds.

$$\left|R^{\alpha}_{real(s)}(\tau)_{2,0}\right| = \left|R^{\alpha}_{imag(s)}(\tau)_{2,0}\right| \tag{A.25}$$