

Distribution Based Spectrum Sensing in Cognitive Radio

Jørgen Berle Christiansen

Master of Science in Electronics Submission date: July 2010 Supervisor: Tor Audun Ramstad, IET Co-supervisor: Aawatif Hayar, Eurecom

Norwegian University of Science and Technology Department of Electronics and Telecommunications

Problem Description

Cognitive radio is a promising field for efficient spectrum utilization. In order for a cognitive radio to function, it needs information about whether a given band of the spectrum is vacant or occupied. Hence a cornerstone component in cognitive radio technology is efficient and robust spectrum sensing algorithms. To prevent hidden terminal problems and similar kinds of interference, it is crucial that the spectrum sensing algorithms provide reliable and robust detection performance also in very low signal to noise ratio conditions.

The thesis shall undertake investigation of distribution based spectrum detection methods under low signal to noise ratio conditions. The underlying hypothesis being that poor performance can be expected in low Signal to Noise Ratio (SNR) environments from conventional detectors such as the energy- or autocorrelation based

ones. The study will focus on this, with the aim of investigating the problem with current approaches, and exploit these in devising a new strategy to potentially solve the problem.

The spectral sensing problem is a binary detection problem, where the detector must decide whether the spectrum is vacant or if there is a signal present. Thus the received signal will have two different conditional probability density distributions, conditioned on the detection hypothesis. This difference in distributions can be

exploited by methods utilizing model selection, information theoretic distance measures, or approaches incorporating higher order statistics to distinguish the distributions. Model selection approaches have already been explored, so it is recommended that higher order statistics or information theoretic distance measures are the primary areas to search for potential solutions to the problem at hand.

Assignment given: 15. January 2010 Supervisor: Tor Audun Ramstad, IET

Abstract

Blind spectrum sensing in cognitive radio is being addressed in this thesis. Particular emphasis is put on performance in the low signal to noise range. It is shown how methods relying on traditional sample based estimation methods, such as the energy detector and autocorrelation based detectors, suffer at low SNRs. This problem is attempted to be solved by investigating how higher order statistics and information theoretic distance measures can be applied to do spectrum sensing. Results from a thorough literature survey indicate that the information theoretic distance Kullback-Leibler (KL) divergence is promising when trying to devise a novel cognitive radio spectrum sensing scheme. Two novel detection algorithms based on Kullback-Leibler divergence estimation are proposed. However, unfortunately only one of them has a fully proven theoretical foundation. The other has a partial theoretical framework, supported by empirical results.

Detection performance of the two proposed detectors in comparison with two reference detectors is assessed. The two reference detectors are the energy detector, and an autocorrelation based detector. Through simulations, it is shown that the proposed KL divergence based algorithms perform worse than the energy detector for all the considered scenarios, while one of them performs better than the autocorrelation based detector for certain signals. The reason why the detectors perform worse than the energy detector, despite the good properties of the estimators at low signal to noise ratios, is that the KL divergence between signal and noise is small. The low divergence stems from the fact that both signal and noise have very similar probability density distributions.

Detection performance is also assessed by applying the detectors to raw data of a downconverted UMTS signal. It is shown that the noise distribution deviates from the standard assumption (circularly symmetric complex white Gaussian). Due to this deviation, the autocorrelation based reference detector and the two proposed Kullback-Leibler divergence based detectors are challenged. These detectors rely heavily on the aforementioned assumption, and fail to function properly when applied to signals with deviating characteristics.

Acknowledgments

I would first and foremost like to thank my supervisor Tor Audun Ramstad at the Norwegian University of Science and Technology and my cosupervisor Aawatif Hayar at the EURECOM research institute. Both have sacrificed much of their time in order to provide me with directions and helpful advice during the progress of my work. I would also like to thank EURECOM for having me as a visiting researcher while working on this thesis. The environment at EURECOM has been highly stimulating and inspiring. In addition, I would like to thank Bassem Zayen at EURECOM for helpful discussions and for him facilitating the sharing of signal data acquired in a previous EURECOM experiment. The analysis of these signals have added an additional dimension to the results presented in this thesis.

Contents

A	bstra	\mathbf{ct}							i
Acknowledgments							iii		
Acronyms					ix				
C	onter	\mathbf{nts}							ix
Li	st of	Table	5						xi
Li	st of	Figur	es						xiv
1	Intr	oducti	ion						1
	1.1	Resear	rch Approach						2
	1.2	Thesis	Objectives and Structure		•	•	•	•	2
2	Theoretical Preliminaries								5
	2.1	The T	wo Fundamental Theorems of Probability						5
		2.1.1	The Law of Large Numbers						5
		2.1.2	The Central Limit Theorem			•		•	6
	2.2	Statist	tical Properties of Random Processes			•		•	6
		2.2.1	Wide Sense Stationarity			•		•	6
		2.2.2	Circular Symmetry			•		•	6
	2.3	2.3 Statistical Propagation Models				•	•	•	7
		2.3.1	Rayleigh Multipath Fading Model		•	·	·	•	8
		2.3.2	Rician Multipath Fading Model		•	•	·	•	8
	~ .	2.3.3	Shadow Fading		•	·	·	•	9
	2.4	Proba	bility Distribution of a Communication Signal		•	·	·	·	9
	2.5	Kullba	ack Leibler Divergence		•	·	·	·	10
	2.6	Higher	r Order Statistics		•	·	·	·	10
		2.6.1	Cumulants		•	·	·	·	11
		2.6.2	Polyspectra		•	·	·	·	12
	~ -	2.6.3	Estimating Cumulants		•	•	·	·	12
	2.7	Statist	tical Detection Theory		•	·	·	·	12
		2.7.1	Binary Detection		•	·	·	·	12

		2.7.2	Constant False Alarm Rate Detector	14
	2.8	Ortho	gonal Frequency Division Multiplexing	14
		2.8.1	Distribution of an OFDM Signal	14
3	Bac	kgrou	nd	15
	3.1	Cognit	tive Radio	15
	3.2	Spectr	cum Sensing and Two Reference Sensing Algorithms	16
		3.2.1	Challenges in Spectrum Sensing	17
	3.3	Refere	ence Detectors	18
		3.3.1	Energy Detector	19
		3.3.2	Autocorrelation detector	20
		0.0.2		
4	Ana	alysis a	and Survey Results on Higher Order Statistics and In-	
	forr	nation	Theoretic Distance Measures	23
	4.1	Challe	enges When Applying Autocorrelation Estimators in Low SNRs	24
	4.2	Potent	tial Solutions to Overcome the Low SNR Detection Problem	28
		421	Higher Order Statistics	$\frac{-0}{28}$
		422	Information Theoretic Distance Measures	31
		1.2.2		01
5	Kul	lback-	Leibler Divergence Based Spectrum Sensing Detectors	33
-	5.1	A Not	e on Signal and Estimator Dimension	33
	5.2	Two P	Proposed Detectors	34
	0.2	521	Histogram Based Detector	34
		5.2.1	CDF Based Detector	30
	53	J.2.2 Implo	montation and Complexity	10
	0.0	5 2 1	Proudo Codo	40
		520	Complexity Analyzia	41
	F 4	0.0.4 D-44	Complexity Analysis	40
	5.4	Detect	for an OEDM Cimelia AWCN Channel	4.4
		gence	for an OFDM Signal in AWGN Channel	44
	5.5	Kullba	ack-Leibler Divergence Based Detectors Under Noise Uncertainty	47
	5.6	Summ	ary of Proposed Algorithms	49
		5.6.1	KL CDF	49
		5.6.2	KL HIST	50
c	C :	latio		51
0				51
	0.1	Simula		51
		0.1.1	Scenario I: Raised Cosine Puise Snaped QAM Signal in	۳1
		010	AWGN Channel	51
		6.1.2	Scenario 2: OFDM Signal in AWGN Channel	52
		6.1.3	Scenario 3: OFDM Signal in AWGN Channel Including Ri-	
		-	cian Multipath Fading and Shadowing	53
	6.2	Simula	ation Results	53
		6.2.1	Detection, Scenario 1	55
		6.2.2	Detection, Scenario 2	55
		6.2.3	Detection, Scenario 3	56
		6.2.4	Detection Under Noise Uncertainty, Scenario 1	56

	 .3 Discussion of Simulation Results 6.3.1 Discussion of Simulation Results Under No Noise Uncertainty 6.3.2 Discussion of Simulation Results Under Noise Uncertainty 	59 59 61
7	Detection Results for an Authentic Captured Signal .1 Description of the Signal and its Distribution .2 Detection Results .3 Discussion of Detection Results for the Captured UMTS Signal	63 63 68 69
8	Summarizing Discussion 1 Potential of Kullback-Leibler estimation in Spectrum Sensing	71 71
9	Conclusion	75
10	Future Work and potential 0.1 Complete the Theoretical Analysis of the KL HIST Algorithm 0.2 Other Distributions and Multidimensional Estimators	77 78 78 78 78 79 79
Re	erences	80
Α	MatLab [®] Code A.1 KL CDF Algorithm A.2 KL HIST Algorithm	87 87 88
в	Proofs 3.1 CDF Based Kullback-Leibler Divergence Estimator 3.2 Fourier Transform of a Complex Exponential	91 91 93
С	Derivation of Error Variance for Autocorrelation Estimation of $AR(1)$ process	95

Acronyms

ADC	Analog to Digital Converter
AR	Autoregressive
AWGN	Additive White Gaussian Noise
CDF	Cumulative Density Function
CSMA	Carrier Sense Multiple Access
FCC	Federal Communications Commission
FFT	Fast Fourier Transform
FPGA	Field Programmable Gate Array
iid	independent and identically distributed
KL	Kullback-Leibler
NTIA	National Telecommunications and Information Administration
OFDM	Orthogonal Frequency Division Multiplexing
PDF	Probability Density Function
PT	Post og Teletilsynet
SDR	Software Defined Radio
SNR	Signal to Noise Ratio
UMTS	Universal Mobile Telecommunications System
WRAN	Wireless Regional Area Network
WSS	Wide Sense Stationarity

List of Tables

6.1	Table gives noise uncertainty parameters used for the simulations in figure 6.3 in dB and linear scale. It also gives an approximate required block length for the number of samples needed to estimate the noise variance so it is confined within the given uncertainty set with an approximate probability of 0.99	58
7.1	Table gives test statistics for the three detection algorithms applied to the signal block, noise block and filtered noise block respectively. The appropriate detection thresholds n are also given.	68
7.2	Table states whether the binary detection decision is correct or wrong based on the signal frame and the computed test statistic (Corresponding test statistics and detection thresholds are given in table 7.1)	68
	taure (.1)	00

List of Figures

Figure shows conditional distributions of the test statistic Υ under the two hypotheses H_0 and H_1 along with a detection threshold η to illustrate probability of detection $P_{\rm D}$ and probability of false detection $P_{\rm FD}$.	13
Required number of samples to estimate $\hat{r}_{xx}[1]$ with a fixed esti- mation variance for a unit variance AR1 process with parameter $\rho_{AR1} = 0.4$ in white Gaussian noise.	27
Estimation errors and variances for the histogram based Kullback- Leibler divergence estimator. The underlying theoretical distribu- tion has variance $\sigma^2 = 2$. <i>P</i> denotes the pre-loading constant applied	
to avoid divide by zero issues in the estimator	36 39
Theoretical and simulated bias and variance for two different estima- tors for estimating Kullback-Leibler divergence between the envelope of complex Gaussian noise and a theoretical Rayleigh distribution. Both distributions have Rayleigh parameter $\sigma^2 = 2$	41
Complexity analysis of four spectrum sensing algorithms Theoretical detection results for OFDM signal in AWGN channel. For the Receiver Operating Characteristic (ROC) curves, SNR is	44
-11 dB	47 49
Welch estimates of the power spectrum. Channel is AWGN with SNR= 8 dB for both signals	52
Monte Carlo simulation results assessing detection performance of a number of spectral detection algorithms. See the section on simulation scenarios for details regarding implementation.	54
	Figure shows conditional distributions of the test statistic Υ under the two hypotheses H_0 and H_1 along with a detection threshold η to illustrate probability of detection P_D and probability of false detection P_{FD}

6.3 Scenario 1 Monte Carlo simulation assessing energy detector under noise uncertainty. Colors denote upper and lower uncertainty level of the uncertainty interval $\left[\frac{1}{\rho}\sigma^2, \rho\sigma^2\right]$, where ρ is given in the legend. 57

7.1 Figures show the time domain components of the captured signal. . 64

Chapter 1 Introduction

Modern communication relies heavily on wireless technology. As a multitude of wireless technologies have been developed, and continuously increased capacity is being demanded, a severe problem has emerged. The electromagnetic spectrum is a fixed physical quantity, and only a certain part of it is suitable for radio communication. Available electromagnetic spectrum for wireless transmission has become a scarce and highly valuable resource.

The traditional way of governing this resource has been to administer licenses for portions of the spectrum, usually by a national agency such as Post og Teletilsynet (PT). Recent research published by the US Federal Communications Commission (FCC), see for instance [13] or [14], shows that a large part of the spectrum is not being effectively utilized. A typical reason being the fact that system demand among licensed users varies significantly with time and location. Such research has spurred the development of a next generation wireless technology commonly referred to as cognitive radio. A device using cognitive radio technology will intelligently determine whether a certain part of the frequency spectrum is idle, or if it is being utilized. If the cognitive radio can successfully determine with a high degree of certainty that a specific part of the spectrum is being idle, it can then transmit on these frequencies without interfering with the licensed owner of the spectrum, thus achieving better spectral resource efficiency. The requirement of no interference is extremely rigid to avoid disturbing licensed users. It is thus key for the development of cognitive radio to invent fast and highly robust ways of determining whether a frequency band is available or being occupied. This is the area of spectrum sensing for cognitive radio.

Spectrum sensing will be the backbone of any autonomous cognitive radio. In order to avoid interference with primary users, robust spectrum sensing should be available with adequate performance at very low signal to noise ratios. For instance, the upcoming cognitive radio based standard IEEE 802.22, which is a high speed wireless regional network standard, requires the detector to sense a primary user at -116 dBm [50, 57]. Many current spectrum sensing algorithms do not meet this requirement, hence there is a motivation to investigate why existing techniques do not provide satisfactory performance, and to investigate new approaches to provide

a solution to the problem. More detailed background information on cognitive radio and spectrum sensing will be presented in chapter 3.

1.1 Research Approach

This section provides a summary of the approach chosen to attack the topic, and how the research was structured. The problem at hand is very open as it is presented in the problem description. It is stated that current spectrum sensing techniques suffer from challenges in the low signal to noise range, requires the reasons for this to be analyzed and suggests that higher order statistics or information theoretic criteria are possible areas to look for a solution to overcome the problem. It is apparent that the problem at hand is wide and challenging. To meet the outlined demands, it is important that the scope is limited to provide a tangible base for a master's thesis. Hence the first step in the research has been to analyze the problem and decide on the correct approach.

The research was split in the following sections:

- 1. Analysis of the problem at hand to limit the scope.
- 2. Literature survey on background information and current techniques in spectrum sensing.
- 3. Analysis of a selection of the conventional approaches to identify problems in the low signal to noise region.
- 4. Literature survey in the areas suggested in the problem outline related to higher order statistics and information theoretic distance measures to decide on a potential new approach.
- 5. Proposing a novel spectrum sensing scheme and providing insight through a theoretical analysis and simulations.

It becomes obvious from the list above that the research is divided in two main parts. The first part revolving around literature surveys and theoretical analysis, and the second part being founded on computer aided simulation. All simulations have been performed utilizing the software package MatLab[®] R2008a, The Mathworks, Inc. As of this, the software will simply be referred to as MatLab[®].

1.2 Thesis Objectives and Structure

The purpose of the thesis is to present an analysis of the problem along with a proposed solution, while maintaining a limited scope to provide coherency and depth. Emphasis has been put on providing an intuitive and chronological presentation. This is reflected in the structure, which follows the enumerated list presented in the previous section.

The thesis starts by briefly introducing a number of theoretical concepts of importance to the following analysis. It is assumed that the reader is familiar with basic concepts from signal processing and communications, so the theory chapter will be structured more as a review of essential fundamental topics and a brief introduction to peripheral topics that the reader might not be familiar with. A number of references providing further depth are provided.

Following the theory chapter is a presented overview of the problem context and two current solutions that will act as references. The overview section starts by introducing the historical background for the term cognitive radio, and progresses naturally into contemporary research problems and current solutions. The following chapter contains the first real research part, which starts by analyzing the reference approaches introduced in the background chapter. After the analysis, results from an extensive literature survey on higher order statistics and information theoretic distance measures are presented. Alongside the presentation of the survey results, a simultaneous discussion of their relevance is given. A conclusion is made on which results that were important enough to pursue further. Based on the findings from the literature survey, two novel detectors are proposed and analyzed.

In the last part of the thesis, the two proposed novel detectors are compared with two reference detectors in terms of detection performance. Performance is mainly assessed through simulations utilizing synthetic signals, but one chapter is devoted to applying the detectors on an authentic captured UMTS signal in order to provide perspective and strengthen the findings from the simulations.

This thesis has unfortunately become very long. This is because it is meant to be self contained for readers that are not familiar with cognitive radio or spectrum sensing. For readers with extensive knowledge related to the field and who are in a hurry, the following is recommended: Read 4.1 and 5.6, have a look at the figures 6.2 and 6.3 before reading the discussion in the same chapter (A quick look at the sections describing the reference algorithms in the background chapter might be necessary to understand notation in the figures and discussion). Quickly skim through the first sections of chapter 7 before reading the chapter discussion. Then read the conclusion in chapter 9. This material would have been the backbone if a condensed summary paper was to be written.

Chapter 2 Theoretical Preliminaries

This section will introduce a number of theoretical concepts of importance. The treatment of the theory will be brief, and thus only serves as an introduction to establish a common starting point for readers of various backgrounds. It is assumed that the reader has at least a basic understanding of elementary statistics and signal processing. As most of the treatment is introductory, experienced readers can most likely skip this entire section, with the possible exception of the sub-section dealing with higher order statistics, a topic not commonly taught in general signal processing courses. The most general advice is for the reader to consult the table of contents to establish whether review of the theoretical preliminaries is necessary, or whether the chapter can be skipped and rather used as reference when relevant topics are encountered in the following chapters.

2.1 The Two Fundamental Theorems of Probability

It is assumed that the reader is familiar with the fundamental theorems of probability. However, these two laws are frequently referred to in this thesis, so this section will provide a brief review in addition to references [4].

2.1.1 The Law of Large Numbers

The law of large numbers is the first of the two fundamental theorems of probability. The law has two definitions; the weak and the strong law of large numbers. Only the weak law will be reviewed here. It states that the sample average of a sequence of independent and identically distributed random variables converges in probability to the true mean.

$$\frac{1}{n}\sum_{i=1}^{n}X_{i} = \bar{X}_{n} \xrightarrow{p} \mu \quad \text{when} \quad n \to \infty$$
(2.1)

That is

$$\lim_{n \to \infty} \Pr\left(\left|\bar{X}_n - \mu\right| < \epsilon\right) \tag{2.2}$$

Which has the interpretation that the difference between the sample average of the independent and identically distributed (iid) sequence and the true mean can always be made smaller than some arbitrarily small constant for large enough n.

2.1.2 The Central Limit Theorem

The central limit theorem states that the sum of a sequence of iid random variables converges to a normal distribution.

$$\sum_{i=1}^{n} X_{i} \xrightarrow{p} \mathcal{N}\left(n\mu_{X}, n\sigma_{X}^{2}\right) \quad \text{when} \quad n \to \infty$$
(2.3)

where μ_X is the true mean and σ_X^2 is the true variance of X_i . The central limit theorem is an important fundamental concept, and is commonly encountered.

2.2 Statistical Properties of Random Processes

Most of the signals encountered in digital communications are not deterministic. This section will in a very brief fashion introduce Wide Sense Stationarity (WSS) and circular symmetry.

2.2.1 Wide Sense Stationarity

For a strictly stationary stochastic process, all statistical properties are time invariant. A weaker requirement, which is often sufficient for signal analysis, is referred to as wide sense stationarity. For a signal to be wide sense stationary [19] its mean must be constant

$$\mu_y[n] = \mu_y, \tag{2.4}$$

the variance must be finite

$$\sigma_y^2[n] < \infty \tag{2.5}$$

and the autocorrelation of the signal $r_y[m,l]$ must only depend on the difference k=m-l

$$r_{y}[m,l] = r_{y}[m-l]$$
(2.6)

A wide sense stationary signal has a real and even autocorrelation function [19].

2.2.2 Circular Symmetry

An important statistical concept for complex random variables is circular symmetry. Circular symmetry intuitively means that the joint probability density function of the random variable, which will be a two-dimensional function in the complex plane, will remain invariant with respect to any rotation. Hence if X is a circularly symmetric random variable, $Y = \exp(j\theta)X$ has an identical joint probability density function as X. This implies that the mean value of the circularly symmetric random variable X is 0, $\mu_X = 0$ and the the real and imaginary components of the random variable are independent [18].

2.3 Statistical Propagation Models

Wireless communication signals suffer from multipath propagation during transmission. The multipath effect is caused by reflection and scattering of the transmitted electromagnetic waves by objects surrounding the transmitter or receiver. Constructive and destructive interference causes what is mainly referred to as fading of the received signal. This section will give a brief review of the main concepts of multipath propagation along with two important statistical multipath models. The material in this section is mainly adopted from [18] and [37].

A single pulse transmitted over a multipath channel will appear as a pulsetrain at the transmitter. The strength of the pulses varies, and depends on whether the pulse is a direct line of sight component or a reflection. The time between the first and the last pulse is called the delay spread of the channel, and is an important property. If the delay spread is small compared to the inverse bandwidth of the transmitted signal, the time spreading of the signal is small, but if the delay spread is large the time spreading can lead to significant signal distortion.

It is often typical to have moving transmitters, receivers or both in wireless communication systems. This will lead to a time varying channel response.

Recall the model of a transmitted wireless signal

$$s[t] = \operatorname{Re}\left\{u[t]\exp\left(j2\pi f_{c}t\right)\right\}$$

$$(2.7)$$

where u[t] is the equivalent baseband signal of s[t] and f_c is the carrier frequency. Extending the model to include multipath yields

$$r[t] = \operatorname{Re}\left\{\sum_{n=0}^{N(t)} \alpha_n[t] u[t - \tau_n[t]] \exp\left(j \left(2\pi f_c \left(t - \tau_n[t]\right) + \phi_{D_n}\right)\right)\right\}$$
(2.8)

where $\alpha_n[t]$ is the time varying amplitude and $\tau_n[t]$ is the time varying time delay of the multipath components. ϕ_{D_n} is the Doppler phase shift of the channel.

$$\phi_{D_n} = \int_t 2\pi f_{D_n}[t]dt \tag{2.9}$$

where $f_{D_n}[t]$ is the time varying Doppler frequency shift caused by moving transmitters or receivers.

Assuming that the delay spread is small compared to the inverse of the signal bandwidth, the fading can be assumed to have a narrow bandwidth so $u[t - \tau_i] \simeq$

u[t]. Hence (2.8) can be simplified to

$$r[t] = \operatorname{Re}\left\{u[t] \exp\left(j2\pi f_c t\right) \left(\sum_{n=0}^{N(t)} \alpha_n[t] \exp\left(-j\phi_{D_n}\right)\right)\right\}$$
$$= \operatorname{Re}\left\{u[t] \exp\left(j2\pi f_c t\right) \alpha_{N(t)}\right\}$$
(2.10)

It can be seen from (2.10) that the effect of the multipath reduces to the complex scalefactor $\alpha_{N(t)}$ when the delay spread is small compared to the inverse of the signal bandwidth.

2.3.1 Rayleigh Multipath Fading Model

The complex amplitudes $\alpha_n[t]$ and the time delays $\tau_n[t]$, corresponding to the different multipath components can be considered independent. Hence when N(t) is large, one can invoke the central limit theorem on $\alpha_{N(t)}$. If the scattering environment is dense, with no line of sight components, the magnitude of $\alpha_{N(t)}$ has a Rayleigh distribution and the channel is said to be a Rayleigh fading channel.

A Rayleigh distributed random variable Y is defined as $Y = \sqrt{X_1^2 + X_2^2}$, where $X_1 \sim \mathcal{N}(0, \sigma_r^2)$ and $X_2 \sim \mathcal{N}(0, \sigma_r^2)$, with

$$f_Y(y) = \frac{y}{\sigma_r^2} \exp\left(-\frac{y^2}{2\sigma_r^2}\right)$$
(2.11)

Note that X_1 and X_2 must be statistically independent.

2.3.2 Rician Multipath Fading Model

If the scattering environment contains fixed permanent scatterers or signal reflectors, the random variable $\alpha_{N(t)}$ no longer has zero mean, and its distribution becomes Rician.

A Rice distributed random variable Y is defined as $Y = \sqrt{X_1^2 + X_2^2}$, where $X_1 \sim \mathcal{N}(\nu \cos \theta, \sigma_r^2)$ and $X_2 \sim \mathcal{N}(\nu \sin \theta, \sigma_r^2)$, where θ is any real number, with

$$f_Y(y) = \frac{y}{\sigma_r^2} \exp\left(-\frac{\left(y^2 + \nu^2\right)}{2\sigma_r^2}\right) I_0\left(\frac{y\nu}{\sigma_r^2}\right)$$
(2.12)

where I_0 is the modified Bessel function of the first kind, with order 0. Note that X_1 and X_2 must be statistically independent. It is common to define a Rician fading channel through what is called the K_{rice} factor. K_{rice} is defined as

$$K_{\rm rice} = \frac{\nu}{2\sigma_r^2},\tag{2.13}$$

which is the ratio of the power in the line of sight component over the scattered power.

2.3.3 Shadow Fading

The statistical fading models introduced previously describe signal attenuation due to multipath propagation. In addition to multipath, communication signals also suffer from physical objects affecting the wave propagation directly. The fading caused by such objects is referred to as shadow fading, or simply shadowing. Location, size and reflective properties of the blocking objects are generally not known, so the effect of shadowing must, as for the multipath fading, be modeled statistically.

The log-normal shadowing model, is one of the most common statistical shadow fading models, and it is the one that will be applied in simulations in this thesis. In log-normal shadowing, the ratio of received power to transmitted power $\phi = \frac{P_r}{P_t}$ follows a log-normal distribution

$$f_{\Phi}(\phi) = \frac{10}{\phi \ln (10)\sigma_{\phi_{\rm dB}}\sqrt{2\pi}} \exp\left(-\frac{(10\log_{10}\phi - \mu_{\rm dB})^2}{2\sigma_{\phi_{\rm dB}}^2}\right)$$
(2.14)

2.4 Probability Distribution of a Communication Signal

The probability distribution of communication signals is of vital importance to the analysis in this thesis, as the research is aimed at finding distribution based methods to perform spectrum sensing in cognitive radio. It is hard to completely characterize such distributions due to the stochastic nature of many communication signals, however there are some common properties. The main property that will be addressed in this section is the fact that signals have zero mean.

Recall from (2.7) that a digitally modulated signal can be written as

$$s[t] = \operatorname{Re}\left\{u[t]\exp\left(j2\pi f_{c}t\right)\right\}$$

$$(2.15)$$

where u[t] is the equivalent baseband signal (Typically a complex number representing a symbol, pulse shaped with a linear filter). It is not easy to explicitly characterize the probability distribution of this signal in general, but it is easy to say something about its mean. To avoid DC on the output, the constellation used for the signal symbols is chosen such that it has mean 0. Hence s[t] has mean zero. This turns out to be an important property in determining the efficiency of the two novel spectrum sensing algorithms that are to be presented later.

Another important property when considering the distribution of the communication signal is the contribution from noise. The noise is generally assumed to be additive with a zero-mean Gaussian distribution, independent of the signal. For noise impact analysis of communication systems, this additive noise typically consists mainly of thermal noise from the receiver input stage [20]. The mean of the received signal y[t] is

$$\mathbf{E}[y[t]] = \mathbf{E}[s[t] + n[t]] = \mathbf{E}[s[t]] + \mathbf{E}[n[t]] = 0$$
(2.16)

Thus the distribution of the received signal will have a zero mean. Do not confuse this mean with the non-zero mean in the discussion about channel fading. The mean value discussed in relation to the channel fading is the mean value of the random variable $\alpha_n[t]$ which is the complex scale factor associated with the fading. $\alpha_n[t]$ does not necessarily have zero mean in general. This does not contradict the claim that a received communication signal has zero mean since the signal and fading are independent, recall that $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$ when X and Y are independent random variables. This is also intuitively understood by recalling that any band limited communication channel can be modeled as a linear filter [37]. A linear filter only affects amplitude and phase of frequencies present, it can not introduce new spectral content. Hence since a DC signal was not present, it can not be introduced by the fading channel.

This section provided motivation as to why communication signals have zeromean distributions.

2.5 Kullback Leibler Divergence

Kullback-Leibler divergence, or relative entropy, is a measure of the distance between two probability distributions. This distance is however not symmetric in general, so it is not a distance in the Euclidean sense. The Kullback-Leibler divergence between the two continuous probability density functions f(x) and g(x) is defined as

$$D(f||g) = \mathbf{E}\left[\log\frac{f(x)}{g(x)}\right]$$
(2.17)

where the expectation is taken with respect to f. D(f||g) is only finite if the support set of f is contained in the support set of g [10]. Another important property of the Kullback-Leibler divergence is that it is non negative,

$$D(f||g) \ge 0 \tag{2.18}$$

and in general, non-symmetric

$$D(p||q) \neq D(q||p).$$

2.6 Higher Order Statistics

Everyone with experience from signal processing or time series analysis, is familiar with the concepts of autocorrelation and power spectral density, and the relationship between them. Over the past decades, these concepts have been generalized to higher dimensions through cumulants and corresponding polyspectra. The following sections present an introduction to the theory of higher order statistics and polyspectra, before ending the presentation with a quick introduction on cumulant estimation. As these concepts are simply higher order equivalents of the already mentioned autocorrelation and power spectral density, they should be easily accessible to the reader.

2.6.1 Cumulants

Recall that the moment generating function of a random variable is

$$m_X(t) = \mathbf{E}\left[\exp\left(tX\right)\right] \tag{2.19}$$

From the moment generating function, the raw moments $\mathbf{E}[X^n]$ of the random variable X can easily be derived.

$$\mathbf{E}\left[X^{n}\right] = \frac{d^{n}}{dt^{n}}m_{X}(t)|_{t=0}$$
(2.20)

The cumulant generating function is defined as the natural logarithm of the moment generating function.

$$c_X(t) = \ln \mathbf{E} \left[\exp \left(tX \right) \right] \tag{2.21}$$

where the raw cumulants are obtained as the coefficients of the Taylor series expansion of the cumulant generating function equivalent to the method of obtaining moments defined previously in (2.20). The following simple example provides some intuitive understanding of the relationship between cumulants and moments

$$\sigma_X^2 = m_X^{''}(0) - m_X^{'}(0)^2 = c_X^{''}(0)$$
(2.22)

The definition extends naturally to random vectors $\mathbf{X}_k = [X_1, X_2, ..., X_m]^T$:

$$c_{k,X}(\mathbf{t}) = \ln \mathbf{E} \left[\exp \left(\mathbf{t}^T \mathbf{X} \right) \right]$$
(2.23)

where $\mathbf{t} = [t_1, t_2, ..., t_m]^T$.

For zero-mean complex random variables from a discrete stochastic process x[n], the first cumulants are given as

$$c_{2,x}(k) = \mathbf{E} [x[n]x^*[n+k]]$$

$$c_{3,x}(k_1, k_2) = \mathbf{E} [x^*[n]x[n+k_1]x[n+k_2]]$$

$$c_{4,x}(k_1, k_2, k_3) = \mathbf{E} [x[n]x[n+k_1]x^*[n+k_2]x^*[n+k_3]]$$

$$- c_{2,x}(k_1)c_{2,x}(k_2 - k_3)$$

$$- c_{2,x}(k_2)c_{2,x}(k_3 - k_1)$$

$$- c_{2,x}(k_3)c_{2,x}(k_1 - k_2)$$
(2.24)

For a non zero-mean process x[n], $x[n] - \mathbf{E}[x[n]]$ in the cumulant expressions is simply substituted for x[n]. It must be noted that different definitions of complex cumulants exist, depending on the choice of which signals to conjugate. The definitions above are common.

Cumulants provide a measure of higher order correlation, and $c_{2,x}$ is easily recognized as the well known autocorrelation, which measures the amount of linear correlation of the signal. For a stochastic process with a Gaussian distribution, all cumulants of higher order than 2 are zero. Hence cumulants also conveniently provide a measure of the distance from a given stochastic process to a stochastic process with a Gaussian distribution.

2.6.2 Polyspectra

Polyspectra are the extension of the regular power spectrum to higher orders. The familiar power spectrum is defined through the Wiener-Khinchin theorem as the Fourier transform of the autocorrelation function.

$$S_{xx}[f] = \sum_{k=-\infty}^{\infty} r_{xx}[k] \exp\left(-j2\pi kf\right)$$
(2.25)

This definition can be generalized to higher order spectra

$$S_{xx}^{(m)}[\mathbf{f}] = \sum_{k_1 = -\infty}^{\infty} \cdots \sum_{k_m = -\infty}^{\infty} c_{m-1,x} \exp\left(-j2\pi \mathbf{k}^T \mathbf{f}\right)$$
(2.26)

2.6.3 Estimating Cumulants

Cumulants can be estimated from a finite number of samples with sample averages if the following conditions are met [31]: The underlying channel must be exponentially stable, the process stationary and the first 2m cumulants are absolutely summable. If these requirements hold, the sample average of the cumulant will converge in probability to the true cumulant. What is meant by underlying channel is the linear time invariant system that generates the process if it is represented as an innovations process. Recall that every WSS stochastic process has a real, even and periodic power spectrum, which implies that it can be spectrally factored, which again implies that it can be represented as an innovations process. I. e. a process generated by white noise driving a linear time invariant system [19].

A consistent sample average has a well known form, such as this example (Which is a sample average estimate of the third cumulant)

$$\hat{c}_{3,x}(k_1,k_2) = \frac{1}{N-k_2} \sum_{n=1}^{N-k_2} x^*[n]x[n+k_1]x[n+k_2] \quad k_2 > k_1$$
(2.27)

For an extensive treatment of Higher order spectra and related signal processing, see [34].

2.7 Statistical Detection Theory

Statistical detection and estimation theory is an important topic in both research and applied signal processing. Such a topic is too wide to address in this section, but a few concepts important for later analysis sections are presented.

2.7.1 Binary Detection

The most basic detection problem is binary. That is deciding between two different detection hypotheses. A problem typically encountered is detecting the presence

of a signal in noise. For the cognitive radio spectrum sensing problem that is to be presented later, the typical detection hypotheses are

$$\begin{aligned} H_0: & y[n] = w[n] \\ H_1: & y[n] = x[n] + w[n] \end{aligned}$$
 (2.28)

where y[n] denotes the received signal, x[n] a transmitted signal and w[n] noise. In order to perform detection, a test statistic Υ_r is computed from a block of the received samples. The most basic test statistic can for instance be to estimate the power the received signal block. This is the energy detector. The energy detector will be presented later. The test statistic will have a conditional probability distribution $f_{\Upsilon_r|H_i}(y)$ conditioned on the detection hypothesis. Figure 2.1 shows an arbitrary example aimed at illustrating the conditional distributions of a binary detection problem including a detection threshold. The challenge encountered in



Figure 2.1: Figure shows conditional distributions of the test statistic Υ under the two hypotheses H_0 and H_1 along with a detection threshold η to illustrate probability of detection $P_{\rm D}$ and probability of false detection $P_{\rm FD}$.

the detection problem is typically how to determine the detection threshold. From figure 2.1, it is apparent that the choice of threshold is subject to a tradeoff between deciding that the signal is present when it actually is present as often as possible, while at the same time making as few decisions that the signal is present when it is not. These two events are called detection and false detection respectively. More formally

$$P_{\rm D} = \Pr\left(\text{Decide Signal Present}|H_1\right)$$

$$P_{\rm FD} = \Pr\left(\text{Decide Signal Present}|H_0\right)$$
(2.29)

Correspondingly, one can also implicitly define the probability of missed detection $P_{\rm MD} = 1 - P_{\rm D}$, and probability of true false detection $P_{\rm TF} = 1 - P_{\rm FD}$.

A thorough treatment of detection and estimation theory can be found in [54].

2.7.2 Constant False Alarm Rate Detector

Correctly determining an optimal threshold requires knowledge of the conditional distributions of the test statistic. Often, it can be substantially easier to obtain knowledge of $f_{\Upsilon_r|H_0}(y)$ than $f_{\Upsilon_r|H_1}(y)$. This has motivated the term constant false alarm rate detector (CFAR). This detector has a detection threshold determined from $f_{\Upsilon_r|H_0}(y)$ only. It has a controlled probability of false alarm. The threshold is not necessarily optimal, but it is a common technique, since a controlled probability of false detector in practice. Another argument for not further optimizing the threshold is that while the conditional distribution under H_0 for many detectors can be stationary (For instance if w occurs from thermal noise in the receiver), the distribution under H_1 is often time varying, for instance due to varying signal power, varying signal distortion or varying signal attenuation.

2.8 Orthogonal Frequency Division Multiplexing

Orthogonal Frequency Division Multiplexing (OFDM) is a popular modulation scheme, widely applied in various wireless standards. A wideband channel is partitioned into a number of narrowband orthogonal sub channels. The sub channels are typically picked to have much smaller bandwidth than the inverse of the channel delay spread, hence the fading on each sub channel can be assumed to be approximately flat and independent of the other sub channels. This diminishes the need for complex equalization schemes, and is one of the main reasons for the popularity of OFDM [18]. Another main reason for the widespread use of OFDM is that the modulation can be performed in the digital domain through the use of the Fast Fourier Transform (FFT) algorithm. An OFDM signal, using a cyclic prefix, can be modeled as [60]

$$s[n] = \sum_{l=-\infty}^{\infty} \sum_{k=0}^{N-1} d_{l,k} g[n-lP] \exp\left(-j\frac{2\pi k}{N} (n-lP-L)\right)$$
(2.30)

where $d_{l,k}$ is the databit on the k th sub-carrier during the l th OFDM symbol, g[n] is the pulse shaping filter at the transmitter, N is the FFT length, P is the OFDM symbol length and L is the cyclic prefix length. A cyclic prefix is a technique used to prevent inter symbol interference on the sub-channels, for more information see [18].

2.8.1 Distribution of an OFDM Signal

As long as the FFT length is large enough, and the data symbols used to modulate the sub-carriers are independent, the central limit theorem can be invoked, yielding a Gaussian distribution for the OFDM signal. This is a common assumption when analyzing OFDM based communication schemes [6, 60].

Chapter 3 Background

This chapter provides background material to understand the problem at hand and the results to be presented in this thesis. The concept of cognitive radio will be explained along with the principles of spectrum sensing. In addition, the chapter will end by introducing two existing spectrum sensing algorithms. These two algorithms will serve as references when evaluating the novel approaches that resulted from the research.

3.1 Cognitive Radio

Radio communication has evolved around a rigid structure stringently defined by the choice of modulation, encoding and the type of hardware to implement these choices [11]. However, the revolution in processor technology in the 1980's and 1990's gradually allowed more flexibility in the design of radio systems as a larger part of the necessary signal processing could be performed digitally. This flexibility created a developing field, researching highly dynamic and adaptive radio systems. These radio systems were defined in software but were implemented through reconfigurable hardware such as Field Programmable Gate Arrays (FPGAs). This branch of wireless technology was named *Software Defined Radio (SDR)*. Cognitive radio, is an extension to SDR, where one allows the radio system to adapt to its environment through learning or artificial intelligence with the aim of increasing performance. An early paper introducing the concept of cognitive radio is [32].

The original definition of cognitive radio is wide, as it envisions the wireless node as a device with cognitive capabilities utilizing all available environmental parameters. According to [32], examples of parameters the cognitive radio can exploit are knowledge such as time, user location, user preferences, knowledge of its own hardware and limitations, knowledge of the network and knowledge of other users in the network. This initial definition of cognitive radio is conceptual, and deviates somewhat from the common contemporary working definition of cognitive radio.

A sub set of cognitive radio that has received a substantial amount of focus is

the Spectrum Sensing Cognitive Radio. This is a radio that dynamically monitors activity in its available electromagnetic spectrum and adapts its transmission to available spectral resources. The most common scenario being an unlicensed secondary user wishing to utilize idle parts of the spectrum when transmission from the licensed primary user is absent. It has become standard practice to simply use the wide term cognitive radio also when referring to limited sub definitions such as Spectrum Sensing Cognitive Radio. This is for instance reflected in modern redefinitions. A typical example is this definition of cognitive radio from the U.S. National Telecommunications and Information Administration (NTIA) [42]:

Cognitive Radio: A radio or system that senses its operational electromagnetic environment and can dynamically and autonomously adjust its radio operating parameters to modify system operation, such as maximize throughput, mitigate interference, facilitate interoperability, access secondary markets.

This definition is a slight misnomer, since it only refers to a more limited adaptive radio, and not the complete cognitive device, utilizing all available parameters from its environment, as presented by the pioneer Mitola in [32]. However, this redefinition of cognitive radio appears to have been widely adopted, and to stick with this practice, the NTIA definition of cognitive radio will be the working definition in this thesis. However, the reader should still be aware of the fact that the original concept of cognitive radio was coined around a concept where a complete set of environmental parameters, and not only spectral parameters, were considered.

This section introduced the term cognitive radio by explaining its historical origin, and further emphasized how development in recent years has focused largely on Spectrum Sensing Cognitive Radio. With the context established, one can quickly move on to the area of spectrum sensing, which is the target topic for this thesis.

3.2 Spectrum Sensing and Two Reference Sensing Algorithms

This section will introduce spectrum sensing in cognitive radio. The different areas of spectrum sensing will be addressed, especially focusing on blind spectrum sensing, which is the area of concentration chosen for the research. In addition, two common blind spectrum detection algorithm are presented. These two sensing algorithms will provide a reference for the two novel approaches that have been developed by the author during the research presented in this thesis.

In order for a cognitive radio to dynamically utilize available spectrum, it must be able to quickly and robustly determine which parts of the relevant spectrum that are available or not. All further processing and decision making performed by the communicating device is based on the results from the initial sensing. It is obvious that spectrum sensing is extremely important for a cognitive radio device to perform satisfactorily. Hence spectrum sensing is a cornerstone of cognitive radio. A large amount of research effort has been put into the area of spectrum sensing over the past decade. The research can be divided in two main parts:

- 1. Blind Spectrum Sensing
- 2. Non-Blind Spectrum Sensing

As the names imply, blind spectrum sensing algorithms make sensing decisions without any prior knowledge, whereas non-blind approaches utilize some form of a priori knowledge about the underlying signals. Typical known signal features can be modulation type, carrier frequency or pulse shape. Two main categories of nonblind spectrum sensing techniques are based on waveform- [51] or cyclostationarity detection [12]. A main category of blind spectrum sensing is based on energy detection [53]. Another main category is based on signal autocorrelation, where there are both blind [23, 33] and non-blind approaches [6]. In addition, there exists a plethora of smaller separate categories, such as for instance subspace based sensing [59] or model selection based sensing [56, 58]. The following tutorial papers provide a thorough treatment of spectrum sensing approaches in addition to important background information [57, 30, 21]. The papers also include an extensive amount of additional references.

Non-blind spectrum sensing has received a lot of attention in spectrum sensing research. This is not surprising as one can obtain optimal detection results with the right signal knowledge. It is for instance elementary knowledge for anyone involved with communications that the matched filter is the optimal receiver for a known pulse shape in an AWGN channel [37]. However, non-blind spectrum sensing narrows the opportunities and applicability due to the need for a priori information of signal features. This to some extent contradicts the original idea of the cognitive radio as an agile and flexible device able to adapt to its environment, since it through utilizing non-blind spectrum sensing, tailored to specific signals, will be limited in terms of operating spectrum bands. It is crucial to limit the scope of this research, and due to the desire to maintain flexibility for the proposed approaches, it has thus been chosen to only focus on *blind* spectrum sensing techniques.

This section presented the topic spectrum sensing for cognitive radio and explained how spectrum sensing algorithms can be divided in the two groups *blind* and *non-blind*. In addition, the section ended by providing motivation for why only blind spectrum sensing is being investigated in this research.

3.2.1 Challenges in Spectrum Sensing

It was previously argued that blind spectrum sensing techniques are the most universal when designing dynamic cognitive radio systems, since the designer not necessarily has knowledge of primary user's signal features. The lack of a priori signal knowledge obviously is an additional disadvantage for blind spectrum sensing approaches as opposed to non-blind. However, the two additional challenges to be presented are common to both.

Since the cognitive radio autonomously makes decisions to transmit, often in licensed frequency bands, it is essential to prevent the cognitive radio from interfering with other users. This is an important networking and resource allocation challenge, and to solve this challenge, it is essential that accurate spectrum sensing algorithms are utilized. This sort of resource allocation problem is very similar to the one experienced in networks based on Carrier Sense Multiple Access (CSMA) [24]. In these networks, a main problem is what is referred to as the hidden node problem or hidden terminal problem [24]. It refers to the fact that while two nodes A and B in a network can both hear node C, they can be hidden to each other. Assume that A decides to transmit to C, it listens for activity, the channel is clear and it starts to transmit. Then while A is transmitting, B also decides to transmit. B listens, and perceives the bandwidth as available since it can not hear A. B starts transmitting to C as well, and a collision occurs. The hidden node problem is primarily caused by physical distance (I.e. Node A and B are placed far apart on each side of node C) or by channel effects such as fading and shadowing. To prevent the *hidden node problem* and similar interference related problems, the spectrum sensing algorithms must be able to detect the presence/absence of signals at very low signal to noise ratios.

The upcoming IEEE standard 802.22, a Wireless Regional Area Network (WRAN) standard employing cognitive radio technology, is a good example of the stringent requirements imposed on spectrum sensing algorithms. The 802.22 standard exploits white spaces in the spectrum licensed for TV transmission to provide long range wireless broadband Internet. It is not fully developed, but the preliminary standard requires the cognitive radio to sense TV transmissions at -116 dBm with a probability of detection $P_{\rm D} \geq 0.9$ and probability of false detection $P_{\rm FD} \leq 0.1$ [50, 57]. Spectral detection at such a low signal to noise ratio is a very challenging requirement.

The last challenge is computational complexity. A large number of emerging wireless devices where cognitive radio can provide a potential future benefit are hand held. Hand held devices usually operate on batteries and have limited computational resources. Hence it is a challenge to develop fast and robust spectrum sensing algorithms with low computational complexity.

Three main challenges for spectrum sensing have been presented. The lack of a priori knowledge of the signal is limited to blind spectrum sensing, while robust performance in low signal to noise ratios and maintaining a low computational complexity are essential to both. The requirement for reliability and accuracy in the low SNR region is emphasized in the research presented in this thesis.

3.3 Reference Detectors

An informative introduction to cognitive radio and spectrum sensing in general, including historical context and contemporary research problems, has been presented. With this information as base, the thesis will progress and provide more perspective by introducing two specific spectrum sensing algorithms that are widely used. The energy detector is the first. This detector is very simple and has been extensively researched and applied. The other detector relies on signal autocorrelation, which is also a very common approach. The following sections will provide
insight in the principles underlining the two detectors and provide the most important related mathematics. It should be noted that these two reference algorithms have been chosen because they are widely applied and referred to in the literature, and that they not necessarily represent state of the art. The alternative would have been to apply more narrow state of the art solutions. However, state of the art solutions tend to be be very specialized and non-blind. Hence choosing the energy detector and an autocorrelation based detector provides a better fit with the aim of the thesis.

3.3.1 Energy Detector

The energy detector is the simplest spectral detection algorithm. It has an advantage in that it is easy and intuitive to comprehend, and has low computational complexity. For these reasons, the energy detector is therefore a good reference when evaluating new and more elaborate spectral sensing methods. The results presented here for the energy detector are mainly taken from [25]. The test statistic is given as

$$\Upsilon_{\rm ED} = \sum_{n=1}^{N} |y[n]|^2$$
(3.1)

As seen from the test statistic above, the energy detector, as the name implies, simply measures the energy of the received signal y[n]. The given test statistic presented here is defined in the time domain, but by recalling Parseval's theorem, it is obvious that an equivalent test statistic can be defined in the frequency domain.

Under H_0 , the received signal is assumed to be circularly symmetric complex white Gaussian noise. Hence the test statistic has a central χ^2_{2N} distribution, where 2N denotes the number of degrees of freedom. A constant false alarm detector can easily be derived. The threshold for the detector is given as

$$\eta_{\rm ED} = \frac{\sigma_n^2}{2} F_{\chi_{2N}^2}^{-1} \left(1 - P_{\rm FD} \right) \tag{3.2}$$

where $F_{\chi^2_{2N}}^{-1}$ denotes the inverse cumulative probability function of a central χ^2 distribution with 2N degrees of freedom, $P_{\rm FD}$ is the probability of false detection and σ_n^2 the noise variance. By applying the threshold (3.2), the probability of detection can be derived, yielding

$$P_{\rm D} = 1 - F_{\chi^2_{2N}} \left(\frac{\eta_{\rm ED}}{\sigma_n^2 + \sigma_s^2} \right), \tag{3.3}$$

where σ_s^2 denotes the signal variance.

Note that the analysis of the energy detector in parts of the literature invoke the central limit theorem, when deriving the threshold. This is also a correct approach, and the thresholds computed with the two different methods converge for large N. A more in depth general treatment of the energy detector can be found in for instance [53], and more spectral sensing specific treatment in for instance [30], which is tutorial paper on spectral sensing.

Noise Uncertainty

As previously mentioned, it is common to assume independent received samples and invoke the central limit theorem when analyzing the energy detector. The previous section only addressed the received signal under H_0 , but this is also common to assume under H_1 . H_0 and H_1 are the binary detection hypotheses, where one under H_0 assumes the received signal to consist of only noise and under H_1 a transmitted signal plus noise. By making the assumption of independent samples, important insight into the major problem of the energy detector is given. In theory, the energy detector can perform reliable detection for any SNR if a sufficient number of samples N is utilized for the estimation. Practical experiments in for instance [5] have shown this result not to hold. The energy detector turns out to provide deteriorating performance when the SNR is decreased. For sufficiently low SNR, robust detection becomes impossible. These results stem from the fact that the theoretical analysis for the energy detector assumes the noise variance to be known, and the underlying noise to have a perfect stationary Gaussian distribution. This assumption does not hold. In reality, the noise variance will usually not be completely stationary. The assumption about the distribution of the noise is also known to be weak. Impulsive noise, aliasing from imperfect filters, leakage from other spectral bands etc. all add to the existing thermal noise, and in many cases create a distribution for the total noise which deviates from Gaussian [50].

To model this noise uncertainty, [50] proposes to allow the actual noise distribution to be confined within a closed set around a nominal noise distribution. The parameter defining the set is denoted ρ , and is called the noise uncertainty parameter. Assume the nominal noise distribution is Gaussian, with variance σ_n^2 . Then the noise uncertainty set will be given as

$$\sigma_n^2 \in \left[\frac{1}{\rho}\sigma^2, \rho\sigma^2\right] \quad , 1 < \rho \tag{3.4}$$

The noise uncertainty model is used to prove the existence of what is called SNR walls. An SNR wall is defined as a signal to noise ratio, for a given noise uncertainty, where the number of samples N required for robust detection approaches infinity. SNR walls, and the consequences of them have received significant attention in [50, 47, 48, 49]. Reading these papers is highly recommended for in-depth study, but a lot of the content goes beyond the scope of this thesis.

This section gave a very brief introduction to the concept of noise uncertainty. The idea of assuming that the noise distribution can not be known completely, but for analysis purposes can be confined to a set known as the noise uncertainty set, will be used to assess the two novel detection algorithms that are to be presented later i this thesis.

3.3.2 Autocorrelation detector

Assuming independence of the received samples is not necessarily true. For most real world communication signals, there usually exists correlation between samples. Typical causes of correlation are for instance dependence of data caused by channel coding, frequency dependent fading and oversampling analog to digital converters on the receiver end [41]. At this point an important difference from statistics must be stressed; correlation and statistical independence are not equivalent. While statistical independence implies a correlation of 0, the converse is not generally true. The reason being that a correlation coefficient measures linear dependence. Higher order dependencies are ignored. Hence one exception to the rule stands out. If samples are jointly normal, statistical independence is equivalent to zero correlation.

A number of different detection schemes, exploiting the correlation of received signals, have been devised for cognitive radio. The schemes can be divided in two main groups, where only one will be addressed in this thesis. Group one is the signal feature based autocorrelation detection algorithms. These typically require knowledge of autocorrelation features of the underlying signal. A typical example of such an approach is the cyclic prefix autocorrelation based algorithm for detecting OFDM signals. It is introduced in [6]. Since this group of detection algorithms require a priori knowledge about the received signal, they are not blind and are therefore not directly relevant to the work presented in this thesis.

The other group of autocorrelation based detection algorithms is the group that blindly exploits signal correlation. The underlying hypothesis assumes the signal to be circularly symmetric complex white Gaussian noise under H_0 . Hence

$$r_{yy}[k] = 0 \quad \forall \ k \neq 0 \quad \text{under } H_0 \tag{3.5}$$

where $r_{yy}[k]$ denotes the autocorrelation function of the received signal y[n].

The detector used as reference in this thesis is presented and analyzed in [33] and [23]. The algorithm will be referred to as the IM algorithm, after the last names of the two authors. Its test statistic is

$$\Upsilon_{\rm IM} = \sum_{k=1}^{K} w_k \frac{\text{Re}\{\hat{r}_{yy}[k] \exp\left(-j\omega k\right)\}}{\hat{r}_{yy}[0]}$$
(3.6)

where

$$\hat{r}_{yy}[k] = \begin{cases} \frac{1}{N-k} \sum_{n=1}^{N-k} y[n] y^*[n+k] & k \ge 0\\ \hat{r}_{yy}^*[-k] & k < 0 \end{cases}$$
(3.7)

$$w_k = \frac{K+1-|k|}{K+1}$$
(3.8)

K is the number of lags of the autocorrelation function to be utilized in computing the test statistic, $\hat{r}_{yy}[k]$ is the maximum likelihood consistent and unbiased estimate of the autocorrelation function of y[n] and w_k is a weighting function for providing varying weight to different lags. In [33], it is shown that the optimal weights w_k are proportional to the underlying autocorrelation function $r_{yy}[k]$. Since the autocorrelation function is generally unknown, an appropriate guess must be made for the weights w_k . The authors of [33] and [23] suggest a triangular function for w_k , motivated by an assumption that the autocorrelation function for real world signal is likely to be decaying. ω is a scanning frequency introduced to allow the baseband frequency of y[n] to have a non zero center frequency. A non zero center frequency can for instance occur as a residual artifact from non ideal downconversion.

It can be seen from (3.6) and (3.7) that evaluating the test statistic can be done blindly. Hence the algorithm can be applied without a priori signal knowledge. Optimality is not achieved for the blind algorithm, since the weights w_k are non optimal, but the authors argue that this has a non vital influence on detection performance. However, there is one caveat to the algorithm, making it only pseudo blind. The choice of K, which is the number of autocorrelation lags to add has to be chosen during implementation. Recall from the theory chapter that a WSS random process has a real and even autocorrelation function. It is further proven in [23] that if the WSS random process is complex with independent real and complex parts, it has a real and even autocorrelation function, which is greater than zero for all lags $k < |N_c|$. N_c is some non zero cut-off integer. The major problem here is of course that N_c is unknown in general when y[n] is an arbitrary signal. Hence N_c must be guessed during implementation. Choosing N_c too small simply gives reduced information embedded in the test statistic, but choosing N_c too large can be devastating. If N_c is too large, one risks adding negative lags of the autocorrelation function to the test statistic, and thus drastically reducing the detection capability of the algorithm.

The IM algorithm is a constant false alarm rate (CFAR) detection algorithm and the threshold can be computed as

$$\eta_{\rm IM} = Q^{-1} \left(P_{\rm FD} \right) \sqrt{\frac{W^{(K)}}{2 \left(N - Q^{-1} \left(P_{\rm FD} \right)^2 \right)}}; \tag{3.9}$$

$$W^{(K)} = \sum_{k=1}^{K} w_k^2 \tag{3.10}$$

N is the number of samples, $P_{\rm FD}$ is the probability of false detection and Q is the standard cumulative Gaussian function

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} \exp\left(-\frac{u^2}{2}\right) \mathrm{d}u \tag{3.11}$$

Two reference detectors, the energy detector and the IM autocorrelation detector have been presented. An intuitive explanation of the algorithms along with the important mathematical results should provide the reader with a sound perspective of common blind spectrum sensing algorithms. This is important as the following chapter will start by analyzing the problems with these algorithms in the low signal to noise ratio region, before the subsequent chapter will present two novel approaches in an attempt to mitigate these problems.

Chapter 4

Analysis and Survey Results on Higher Order Statistics and Information Theoretic Distance Measures

Previous chapters have introduced the spectrum sensing problem in Cognitive Radio, along with two reference detectors. This chapter begins with an analysis demonstrating some of the major issues with established sensing algorithms in the low signal to noise ratio region. The analysis starts by examining the estimation of the autocorrelation function. This estimation is used directly in the IM autocorrelation based algorithm and implicitly in the energy detector (The energy detector is in reality an autocorrelation based detector with k = 0), and in a number of other algorithms not addressed in this thesis [57, 30, 21]. By making some realistic assumptions on signal structure, it is shown how algorithms relying on conventional sample based estimates of the autocorrelation function suffer at decreasing signal to noise ratios because the estimation variance depends on signal- and noise variance.

Based on analysis of the autocorrelation estimator, it is concluded that potential new sensing techniques should rely on estimators with an estimation variance independent of noise- or signal variance in order to remain efficient for finite number of samples at low signal to noise ratios. This was a key factor when performing a literature survey on higher order statistics and information theoretic distance measures. Areas that were suggested to have potential solutions in the initial problem statement. When assessing the results from the survey, it is judged that information theoretic distance measures appear to be more promising than higher order statistics to overcome the problem of increased estimation variance at low SNR. Findings from the survey indicates that Kullback-Leibler divergence, the most common information theoretic distance measure, can be estimated with a low estimation variance that is independent of underlying signal- or noise variance. This becomes the foundation for further research. Two novel spectrum sensing algorithms based on Kullback-Leibler divergence estimation are devised and analyzed in the following chapter.

4.1 Challenges When Applying Autocorrelation Estimators in Low SNRs

To analyze autocorrelation based spectral sensing algorithms, the mean squared error (MSE) for the estimation of the autocorrelation function of a general signal will be derived. Let x[n] be an arbitrary, possibly complex signal, assumed to be wide sense stationary over the analysis interval. The MSE of the estimator is

$$\mathbf{E}\left[\epsilon_{r}^{2}[k]\right] = \mathbf{E}\left[|r_{yy}[k] - \hat{r}_{yy}[k]|^{2}\right]$$

=
$$\mathbf{E}\left[|r_{yy}(k)|^{2} - r_{yy}^{*}[k]\hat{r}_{yy}[k] - r_{yy}[k]\hat{r}_{yy}^{*}[k] + |\hat{r}_{yy}[k]|^{2}\right]$$
(4.1)

Since $r_{yy}[k]$ is a deterministic function and the estimator $\hat{r}_{yy}[k]$ is unbiased ($\mathbf{E}[\hat{r}_{yy}[k]] = r_{yy}[k]$), the above simplifies to

$$\mathbf{E}\left[\epsilon_r^2[k]\right] = \mathbf{E}\left[\left|\hat{r}_{yy}[k]\right|^2\right] - \left|r_{yy}[k]\right|^2 \tag{4.2}$$

(4.2) can be evaluated by considering the standard unbiased maximum likelihood sample based estimator

$$\hat{r}_{yy}[k] = \frac{1}{N-k} \sum_{l=0}^{N-k-1} y[n-l]y^*[n-l-k]$$
(4.3)

which gives the following

$$\mathbf{E}\left[\epsilon_{r}^{2}[k]\right] = \mathbf{E}\left[\frac{1}{N-k}\sum_{l=0}^{N-k-1}y[n-l]y^{*}[n-l-k]\frac{1}{N-k}\sum_{m=0}^{N-k}y^{*}[n-m]y[n-m-k]\right] - |r_{yy}[k]|^{2} = \frac{1}{(N-k)^{2}}\left(\sum_{l=0}^{N-k-1}\sum_{m=0}^{N-k-1}\mathbf{E}\left[y[n-l]y^{*}[n-l-k]y^{*}[n-m]y[n-m-k]\right]\right) - |r_{yy}[k]|^{2}$$
(4.4)

If one assumes that y[n] is a WSS complex Gaussian stochastic process with zero mean, the above simplifies to

$$\mathbf{E}\left[\epsilon_r^2[k]\right] = \frac{1}{(N-k)^2} \left(\sum_{l=0}^{N-k-1} \sum_{m=0}^{N-k-1} \left(|r_{yy}[k]|^2 + r_{yy}[m-l]r_{yy}[l-m]\right)\right) - |r_{yy}[k]|^2$$
$$= \frac{1}{(N-k)^2} \sum_{l=0}^{N-k-1} \sum_{m=0}^{N-k-1} r_{yy}[m-l]r_{yy}[l-m]$$
(4.5)

The above used the following identity [40], which holds for zero mean complex Gaussian random variables

$$\mathbf{E}[X_1^*X_2^*X_3X_4] = \mathbf{E}[X_1^*X_3] \mathbf{E}[X_2^*X_4] + \mathbf{E}[X_2^*X_3] \mathbf{E}[X_1^*X_4]$$
(4.6)

For a wide sense stationary random process y[n], the autocorrelation function is conjugate symmetric [19], that is

$$r_{yy}[k] = r_{yy}^*[-k] \tag{4.7}$$

Hence

$$\mathbf{E}\left[\epsilon_{r}^{2}[k]\right] = \frac{1}{(N-k)^{2}} \sum_{l=0}^{N-k-1} \sum_{m=0}^{N-k-1} r_{yy}[m-l]r_{yy}[l-m]$$

$$= \frac{1}{(N-k)^{2}} \sum_{l=0}^{N-k-1} \sum_{m=0}^{N-k-1} r_{yy}[m-l]r_{yy}^{*}[m-l]$$

$$= \frac{1}{(N-k)^{2}} \sum_{l=0}^{N-k-1} \sum_{m=0}^{N-k-1} |r_{yy}[m-l]|^{2}$$
(4.8)

The above can be used to derive a useful relationship between the estimation variance and signal to noise ratio. Let the received signal y[n] = x[n] + w[n] consist of the arbitrary complex WSS signal x[n] with a zero-mean Gaussian distribution, plus additive white complex Gaussian noise. It is assumed that signal and noise are circularly symmetric, and mutually independent. These assumptions are realistic for a range of signals, for instance for an OFDM signal [6, 60]. The autocorrelation function of the received signal is $r_{yy}[k] = r_{xx}[k] + \sigma_w^2 \delta[k]$. Substituting into (4.8) yields

$$\mathbf{E}\left[\epsilon_{r}^{2}[k]\right] = \frac{1}{(N-k)^{2}} \sum_{l=0}^{N-k-1} \sum_{m=0}^{N-k-1} |r_{yy}[m-l]|^{2}$$

$$= \frac{1}{(N-k)^{2}} \sum_{l=0}^{N-k-1} \sum_{m=0}^{N-k-1} \left(r_{xx}[m-l] + \sigma_{w}^{2}\delta[m-l]\right) \left(r_{xx}^{*}[m-l] + \sigma_{w}^{2}\delta[m-l]\right)$$

$$= \frac{1}{(N-k)^{2}} \sum_{l=0}^{N-k-1} \sum_{m=0}^{N-k-1} |r_{xx}[m-l]|^{2} + \frac{1}{(N-k)^{2}} \sum_{l=0}^{N-k-1} \left(r_{xx}[0]\sigma_{w}^{2} + \sigma_{w}^{4}\right)$$
(4.9)

By remembering that $r_{yy}[0] = \sigma_y^2$, 4.9 can be reduced to

$$\mathbf{E}\left[\epsilon_{r}^{2}[k]\right] = \frac{1}{(N-k)^{2}} \sum_{l=0}^{N-k-1} \sum_{m=0}^{N-k-1} |r_{xx}[m-l]|^{2} + \frac{2\sigma_{x}^{2}\sigma_{w}^{2} + \sigma_{w}^{4}}{N-k}$$

$$= \frac{\sigma_{x}^{4}}{N-k} + \left(\frac{1}{(N-k)^{2}} \sum_{l=0,m\neq l}^{N-k-1} \sum_{m=0,m\neq l}^{N-k-1} |r_{xx}[m-l]|^{2}\right) + \frac{2\sigma_{x}^{2}\sigma_{w}^{2} + \sigma_{w}^{4}}{N-k}$$

$$\geq \frac{\sigma_{x}^{4}}{N-k} + \frac{2\sigma_{x}^{2}\sigma_{w}^{2} + \sigma_{w}^{4}}{N-k}$$

$$= \frac{\left(\sigma_{x}^{2} + \sigma_{w}^{2}\right)^{2}}{N-k}$$
(4.10)

The inequality in (4.10) reduces to the equality

$$\mathbf{E}\left[\epsilon_r^2[k]\right] = \frac{\sigma_w^4}{N-k},\tag{4.11}$$

if the signal y[n] consists solely of white complex circularly symmetric Gaussian noise. When assessing detection performance for decreasing signal to noise ratios, it is equivalent whether the Signal to Noise Ratio (SNR) is reduced by fixing the signal variance and increasing the noise variance or fixing the noise variance and decreasing the signal variance. Using the former illustrates an important problem with the sample based estimators. Notice how a linear increase of the noise variance when the signal variance is fixed requires a quadratic increase in the number of signal samples used in the estimation in order to keep the lower bound on the estimation variance fixed. There is obviously practical limitations on the number of samples that can be utilized for estimation in the cognitive radio, these limitations can for instance be hardware related or the signal can be assumed wide sense stationary only over limited time intervals. Hence it becomes infeasible to achieve reasonable lower bounds for the estimation variance when the SNR drops sufficiently. Keeping a low estimation variance is crucial in order to perform reliable detection.

Since (4.10) is a general inequality, it is not suitable for making specific illustrations of the number of samples required for a given MSE. However, in [8], which is unpublished previous work by the author, the variance of the estimation error when estimating the autocorrelation of an autoregressive process of order 1 (AR1) with parameter ρ in white Gaussian noise has been derived. For more information on autoregressive processes, see for instance [19] or [38]. The estimation variance for estimating the autocorrelation function of a unit variance AR1 process in white Gaussian noise is given as

$$\mathbf{E}\left[\epsilon_{r}^{2}[k]\right] = \frac{1}{N} \left(\frac{\varrho^{2} - \varrho^{-2} + 4N^{-1} \left(1 - \varrho^{2N}\right) + 2\varrho^{2k} \left(1 - \varrho^{-2}\right)}{\left(1 - \varrho^{2}\right) \left(1 - \varrho^{-2}\right)} \right) + \frac{1}{N} \left(\varrho^{2k} \left(2\sigma_{n}^{2} + 2k - 1\right) + 4\sigma_{n}^{2} + \sigma_{n}^{4} (1 + \delta[k]) \right)$$
(4.12)

Where N is the number of samples used for the estimation, σ_n^2 is the variance of the noise, k is the autocorrelation lag and ρ is the AR parameter. The derivation can be found in appendix C. By assuming N >> 4, (4.12) can be approximated as

$$\mathbf{E}\left[\epsilon_{r}^{2}[k]\right] = \frac{1}{N} \left(\frac{\varrho^{2} - \varrho^{-2} + 2\varrho^{2k} \left(1 - \varrho^{-2}\right)}{\left(1 - \varrho^{2}\right) \left(1 - \varrho^{-2}\right)} \right) + \frac{1}{N} \left(\varrho^{2k} \left(2\sigma_{n}^{2} + 2k - 1\right) + 4\sigma_{n}^{2} + \sigma_{n}^{4} (1 + \delta[k]) \right)$$
(4.13)

An AR process can be used as a good model of a correlated communication signal, see discussion in [8]. It is seen from the above equation, that when the SNR decreases, it becomes infeasible to estimate the autocorrelation with sufficient accuracy. For an illustration see figure 4.1. Figure 4.1 plots the required number



Figure 4.1: Required number of samples to estimate $\hat{r}_{xx}[1]$ with a fixed estimation variance for a unit variance AR1 process with parameter $\rho_{AR1} = 0.4$ in white Gaussian noise.

of samples N versus SNR for estimation of the autocorrelation of an AR1 process in white Gaussian noise using (4.13). Recall that $\mathbf{E}\left[\epsilon_r^2[k]\right] = \text{MSE}$ and since (4.13) applies to a unit variance AR1 process in white Gaussian noise with variance σ_n^2 , SNR= $1/\sigma_n^2$. It is easily seen that the number of required samples N increases drastically when the SNR decreases below approximately -5 dB (Notice the logarithmic scale on the Y axis).

This section showed how conventional sample based estimation of a signal's autocorrelation function suffers from an estimation variance that depends on the underlying signal- and noise variance. It was further shown how the number of samples needed for reliable estimation increased dramatically at decreasing signal to noise ratios, hence rendering reliable estimation infeasible at sufficiently low SNRs. Note however, that in order to make this general analysis, some fundamental assumptions on signal and noise structure had to be made. The analysis assumed both the signal and the noise to be WSS, to have circularly symmetric Gaussian distributions and be statistically independent of each other. Recall from the theory section how it was argued that communication signals have zero-mean, and that the underlying noise is assumed to be circularly symmetric complex white Gaussian noise. It is also fair to assume that the signal and noise are statistically independent and WSS over sufficient time intervals to perform detection. Hence the only questionable part of the argument is assuming the signal to have a circularly symmetric complex Gaussian distribution. This assumption is not true in all cases, but the assumption is realistic for some signals, such as OFDM as long as the number of sub-carriers is large enough [6, 60]. Although the assumptions on signal structure do not always hold, it is fair to assume that there none the less is some universality to the results derived above. For simplicity in the further analysis in this thesis, the results above will be assumed to hold for general signals. Remember however that the results have only been proved strictly for signals that that have a WSS circularly symmetric Gaussian distribution.

4.2 Potential Solutions to Overcome the Low SNR Detection Problem

The previous section analyzed the shortcomings of conventional block based sample average estimates. Such estimators are applied in both the energy detector and the IM autocorrelation detector. It was shown how detection at low signal to noise ratios became infeasible because the estimators had estimation variances relying on signal- and noise variance. A starting point when trying to devise new detection algorithms, to overcome this problem, is to search for estimators where estimation performance does not rely on signal- or noise variance. This has been the main focus when performing a broad literature survey on higher order statistics and information theoretic distance measures. The survey is the starting point before proposing two novel spectrum sensing algorithms.

The two following sections will provide brief summaries of the important findings from the literature survey, and the conclusions drawn by them. Following the results from the survey is an extensive introduction and analysis of the two proposed novel spectrum sensing algorithms.

4.2.1 Higher Order Statistics

Initially, higher order statistics was investigated for the purpose of spectrum detection in cognitive radio. It is well known that including higher order statistics brought improvements to many areas of signal processing [31, 34]. However, the improvements seem to be mostly related to underlying signal structures that do not apply to the cognitive radio detection problem. An area that especially has benefited from higher order statistics is estimation in colored Gaussian noise, a fact that is obvious when remembering from the theory chapter that higher order cumulants of a Gaussian distribution are zero. Hence methods based on higher order statistics will boost the signal to noise ratio for non-Gaussian signals in Gaussian noise [15]. For the cognitive radio problem on the other hand, this thesis uses the standard system model assuming the underlying noise to be circularly symmetric white Gaussian noise. If the communication signals also have distributions that deviate little from Gaussian, higher order statistics do not necessarily provide any benefits, and may even deteriorate the results. This section will provide the main findings from a thorough literature survey aimed at investigating the potential of higher order statistics for blind spectrum sensing in cognitive radio.

[52] presents results for general detection of arbitrary random signals. The algorithms presented are relying on higher order statistics and a hybrid of higher order statistics and conventional correlation based approaches. The results for different signal models span from a moderate performance increase of about 1 dB to a more significant decrease of approximately 5 dB compared to the energy detector. In total, the results presented in [52] are not very promising. Another relevant paper [45], presents a similar analysis, although less complete than [52]. The results are more ambiguous, but in total also appear to be discouraging. The research summarized in the paper shows cumulant spectra of a vestigal sideband modulated TV signal, obtained through simulations. Results in the paper show that information is present in the higher order spectra, but it appears that little information is present when the SNR drops below -10 dB. This is not promising as existing detection algorithm of substantially lower complexity already perform satisfactorily at such SNRs. However, the paper indicates that cyclic frequencies are present in the higher order spectra. Cyclic frequencies can be exploited when performing non-blind spectrum sensing. The benefit of higher order spectra for non-blind spectrum sensing is further documented in [29]. [29] reports increased performance over the energy detector for a third order cyclic cumulant detector. Although the results on higher order cyclic frequencies are positive, they are not relevant for the research presented in this thesis since this research is aimed at blind spectrum sensing.

After a thorough review of all discovered papers with a direct link between higher order statistics and spectrum sensing in cognitive radio, the survey was expanded to comprise other relevant areas involving higher order statistics that might prove themselves useful to the problem at hand. Particularly useful for the preliminary conclusion about the potential of higher order statistics were revealed after studying a number of papers on modulation classification. Higher order statistics have in a number of papers been shown to be helpful in performing modulation classification of digitally modulated communication signals [46, 28, 27, 7]. This itself is an interesting finding, but results presented, especially in [46] and [7] became pivotal. The papers suggest that while higher order statistics are applicable for modulation classification, robustness of the methods deteriorate drastically below 10 dB. This implies what was already mentioned briefly before, that for many modulation types, communication signals often have distributions that are very close to Gaussian. Especially in lower signal to noise ratios. Hence higher order statistics provide little information. Recall that the distribution of the received signal is the convolution between a Gaussian noise distribution and a communication signal that is typically close to Gaussian. Hence after the convolution, the received signal has a distribution very close to Gaussian, and at signal to noise ratios below 10 dB, it becomes increasingly hard to distinguish between the modulation types through conventional estimation. This leads to the most important argument against the potential of higher order statistics for the cognitive radio detection problem in the low signal to noise ratio range. The fact is that all the higher order methods that were reviewed rely on conventional sample average based estimation methods equivalent to the ones prevalent among energy and autocorrelation based detectors. For instance cumulants are estimated with the same sample average based rectangular window as the one applied in the energy detector and the IM autocorrelation algorithm reviewed in the background section. [46] derived the estimator used for estimating the fourth order cumulant. The expression applies to a signal in noise, and is rather involved. However, if the signal is assumed to be zero, the variance expression reduces to

$$MSE_{C4} \ge \frac{4\sigma_w^8}{N},\tag{4.14}$$

which shows a striking structural resemblance to a result seen previously in (4.11).

$$MSE_{C2} \ge \frac{\sigma_w^4}{N-k} \tag{4.15}$$

Note the similarity to (4.14). However, the result for the fourth order cumulant is more severe since the noise variance is raised to the power of four, as opposed to the power of two for the second order cumulant estimated in the energy detector and the autocorrelation based algorithm. This supports the argument that the sample based estimators used to estimate higher order cumulants suffer from the same problems in low SNR as the autocorrelation estimator. It is also obvious that the variance of the fourth order cumulant estimate is higher than the second order cumulant estimate (I.e. the autocorrelation). This is an obvious fact, and it is stated explicitly in for instance [31], how the variance of the cumulant estimate increases with the cumulant order.

This section provided relevant points from a thorough literary survey aimed at investigating the potential of higher order statistics in spectrum sensing at low signal to noise ratios. Overall, after analyzing the findings of the survey related to blind spectrum sensing based on higher order statistics in general, and in cognitive radio in particular, it was judged that no apparent potential has been revealed. The higher order methods reviewed were all structured around the same conventional sample average based estimation as autocorrelation detectors, which means that they suffer from the same problems at low signal to noise ratios. In addition, estimating higher order cumulants require more computation and the estimates have a higher variance than the second order cumulant estimates applied in the energyand autocorrelation detector. However, it should be emphasized that positive results were encountered for non-blind spectrum sensing, but non-blind spectrum sensing is beyond the scope of this research. It must also be stressed that while the conclusion after this survey is that no apparent potential has been *discovered*, it must not be interpreted as a conclusion that possible potential does not exist.

4.2.2 Information Theoretic Distance Measures

The findings when investigating blind spectrum sensing with higher order statistics were discouraging, and only strengthened the idea of finding estimators of low computational complexity that did have an estimation variance independent of signal or noise variance. Conditional distributions of the received signal under the two hypotheses H_0 and H_1 have already been discussed. To continue this approach, it is appealing to investigate whether information theoretic distance measures can be applied to distinguish between them. To limit the scope of the analysis, only the most common information theoretic distance measure, the Kullback-Leibler divergence, has been chosen to be addressed due to its widespread use in information theory and communications [10].

The Kullback-Leibler divergence is a well known information theoretic measure that provides a metric for the distance between two probability distributions [10] (recall that this is not a distance in the Euclidean sense as $D(f||g) \neq D(g||f)$ in general). Previous research indicates that KL divergence estimation can be used for detection purposes [2, 3], which establishes motivation to investigate such methods for the cognitive radio blind spectrum sensing problem at hand. Recall from previous sections that the cognitive radio detection problem investigated in this report assumes an underlying circularly symmetric complex Gaussian noise model. Hence the KL divergence should be estimated from an empirical distribution of the received samples $\hat{f}_Y(y)$ to a theoretical noise distribution $g_N(y)$, $D(\hat{f}||g)$.

For the spectrum sensing detection problem, only one vector of input samples is to be compared to the theoretical distribution under one of the different hypotheses; signal present (H_1) and signal vacant (H_0) . H_0 is used since H_1 has a non-stationary nature due to channel effects. Hence the KL divergence estimation algorithm must take two vectors as inputs, where the first vector represents the received signal samples and the second vector represents parameters of the theoretical conditional distribution under H_0 . Kullback-Leibler divergence can be estimated in a number of ways. The most straightforward is to first estimate the empirical distribution of the received samples at hand using a normalized histogram, and then approximate the KL divergence of the continuous true probability density functions with the KL divergence between the the empirical distribution and a corresponding noise distribution derived from the relevant theoretical noise model (I.e. theoretical conditional distribution under H_0). This approach will be explored and analyzed. In addition to the traditional histogram based method, it is desirable to have one other more advanced estimation method as reference.

A literature survey resulted in the two papers [36] and [39] as good candidates for more elaborate Kullback-Leibler divergence estimators. [39] suggests estimating the characteristic function of a normalized version of the input signal, composing a toeplitz matrix of the characteristic function, compute its eigenvalues and use these eigenvalues to estimate the KL divergence. The estimation procedure is founded in the relationship between the sum of the eigenvalues of an autocorrelation matrix, and the integral of the spectrum given by Szego's theorem (For intuition on this relationship, see for instance [55]). In the algorithm proposed in [39], the characteristic function serves as the autocorrelation sequence, while its Fourier transform (I. e. probability density function) serves as the spectrum. This idea is appealing, and the original authors suggest good results obtained through simulations. However, the method does not appear to be easily extended to handle one empirical and one theoretical distribution. In addition, the paper does not present a satisfactory analysis of robustness and bias so the method is deemed as inapplicable.

[36] presents a completely different approach. The algorithm given suggests estimating the KL divergence between two distributions through estimating their cumulative density functions. The analysis and ideas presented in the paper are thorough and consistent and the author implies that the estimation variance of the algorithm only scales with the number of input samples, but the estimator is not directly applicable to the cognitive radio spectrum sensing problem. The [36] method only estimates the KL divergence between two empirical distributions represented by two vectors of signal samples. This calls for the development of a method that can work with one theoretical and one empirical distribution. Inspired by the method from [36], the author decided to develop a novel approach to perform this estimation that handles one empirical and one theoretical distribution. This method will be presented and analyzed later.

Summaries and conclusions from extensive literature surveys on higher order statistics and information theoretic distance measures, represented by the Kullback-Leibler divergence have been presented. The main conclusion for higher order statistics based methods was that they relied on the same conventional sample average based estimators as the reference detectors, and thus suffered from the same problems. The main conclusion for Kullback-Leibler divergence is that there is potential for estimators only relying on the number of samples and not signalor noise variance, but that there is a need for developing novel estimation methods as a reference to the standard histogram based estimators. With these results in mind, two novel spectrum sensing detection algorithms based on Kullback-Leibler estimation will be presented in the following sections, where one of them relies on a new estimation method inspired by [36], but completely developed by the author during the research presented in this thesis.

Chapter 5

Kullback-Leibler Divergence Based Spectrum Sensing Detectors

The previous chapter discussed how Kullback-Leibler divergence was the most promising area discovered in the literature survey on higher order statistics and information theoretic distance measures. This chapter will present and analyze two proposed detectors. One will be based on empirical probability density function estimation using conventional histograms, the other will be a novel approach developed by the author utilizing the empirical cumulative density function. The first part of the chapter is devoted to introducing the detectors. The second part of the chapter provides a performance analysis. During the performance analysis, the probability of detectors. The chapter ends with a summary of the derived mathematical results for the two proposed detectors.

5.1 A Note on Signal and Estimator Dimension

Communication signals are usually two dimensional due to the use of in-phase and quadrature components. Hence the signals are represented in the complex domain, and a two dimensional probability density function is needed to completely characterize the signal. Some initial attempts on utilizing Kullback-Leibler divergence estimation for multivariate signals following the outlines given in [36] have been performed. However, these algorithms relied on nearest neighbor searching, which for the relatively large sample sets (Usually in the size of thousands) used for the simulations to be presented later in this thesis posed a significant computational challenge. It was thus chosen to focus on univariate estimation for simplicity, although univariate estimation is expected to provide worse performance than multivariate estimation. However, utilizing univariate estimation is to a certain extent supported by the assumption of circular symmetry. If a complex signal has a circularly symmetric distribution, its real and imaginary components are statistically independent.

Possible distributions that can be utilized are for instance signal amplitude or signal magnitude. More elaborate schemes such as utilizing the distribution of the signal's one sample autocorrelation function can also be used. Results from the author's previous work [8], suggests that schemes based on the autocorrelation distribution will have little effect when there is little correlation in the signal. Hence it was chosen in this thesis to focus on the more straightforward schemes. The final choice is to use the distribution of the signal magnitude. Since the signal under H_0 is assumed to be circularly symmetric white complex Gaussian noise, the signal magnitude under H_0 will have a Rayleigh distribution. The choice to use signal magnitude was partly made in order to utilize both dimensions of the signal, without necessarily having to assume circular symmetry, and partly due to the fact that the Kullback-Leibler divergence between two different Rayleigh distributions is twice as large as the Kullback-Leibler divergence between the Gaussian distributions used to generate them. Hence better detection capabilities can be achieved. This will be documented and elaborated through a mathematical derivation later. The following sub-sections will now introduce the univariate Kullback-Leibler divergence based spectrum sensing detectors.

5.2 Two Proposed Detectors

This section will present the two detectors. The first sub-section introduces a detector based on estimating Kullback-Leibler divergence through the use of a histogram. The theoretical analysis for this detector is unfortunately not complete, but is motivated through empirical results. Due to this, the sub-section becomes rather long. The second sub-section introduces a detector based on estimating Kullback-Leibler divergence by the aid of cumulative density functions. This detector has a fully proven theoretical foundation, yielding a concise presentation.

5.2.1 Histogram Based Detector

Recall that the Kullback-Leibler divergence from the probability mass function f(y) to g(y) is [10]

$$D(f||g) = \mathbf{E} \left[\log \frac{f(y_i)}{g(y_i)} \right]$$

= $\sum_i f(y_i) \log \frac{f(y_i)}{g(y_i)}$ (5.1)

The above can used as an approximation to the KL divergence for continuous distributions (2.17). A discrete probability mass function can be obtained through quantizing a continuous probability density function. In practice, such quantization can be approximated through histograms. The choice of histogram bin width Δ

and the number of bins N_{Δ} is of apparent importance to the accuracy of the approximation. Too few bins will give a crude estimate, while too many bins will give very noisy histograms. Another crucial aspect is that to obtain a finite KL divergence, $f(y_i) \neq 0$ and $g(y_i) \neq 0 \forall y_i$. Note that this implies the assumption that the underlying true KL divergence is finite, hence that the support set of the true underlying probability density function f(y) is contained in the support set of g(y). To assure that the KL estimate remains finite, no histogram bin can contain zero elements. To prevent this, a technique named preloading is applied [26]. Preloading simply means that some non-zero constant is added to all bins of the histogram.

After the histogram H_{Δ} has been estimated and preloading applied, the histogram is normalized to a probability mass function

$$\hat{f}(y_i) = H_{\Delta}(y_i) / \left(\sum_i H_{\Delta}(y_i)\right)$$

For the spectrum sensing problem, only f(y) needs to be approximated as explained above. g(y) is given by parameters according to the detection hypothesis and the probability mass function is computed as

$$g(y_i) = \begin{cases} \int_{y_i - \frac{\Delta}{2}}^{y_i + \frac{\Delta}{2}} g_{Y|H_0}(y) \mathrm{d}y & i \neq 1, i \neq N_\Delta \\ \int_{-\infty}^{y_i + \frac{\Delta}{2}} g_{Y|H_0}(y) \mathrm{d}y & i = 1 \\ \int_{y_i - \frac{\Delta}{2}}^{\infty} g_{Y|H_0}(y) \mathrm{d}y & i = N_\Delta \end{cases}$$
(5.2)

where $g_{Y|H_0}(y)$ is the true underlying continuous conditional probability density function for the received signal under H_0 .

[2] and [3] provide a discussion on estimating KL divergence through histograms, and insight into the accuracy of the probability density function approximation when using histograms. In this thesis, an analysis aided by Monte-Carlo simulation has been performed in order to pick good values for the number of bins used in the histogram and the preloading constant. It is argued in |26| that 0.5 is an appropriate choice for the preloading constant, so this value was used as the starting point in the analysis. Intuitively, reducing the preloading constant reduces the bias of the algorithm. A number of simulations were performed utilizing preloading constants 0.5, 0.01 and 0.001 to determine the appropriate choice of parameters. The results of the simulations can be seen in figure 5.1. The figure shows bias and variance when estimating the Kullback Leibler divergence with a histogram between empirical samples drawn from a Rayleigh distribution with parameter $\sigma^2 = 2$ and a theoretical Rayleigh distribution with σ^2 varying to represent different signal to noise ratios (This implicitly assumes that the signal will also have a circularly symmetric Gaussian distribution, which is not necessarily true, but assumed for simplicity). This represents a realistic scenario to the effect of bin width and preloading. When interpreting the results, it is important to keep in mind that the ideal situation is to pick a preloading constant and number of bins to get a good tradeoff between low estimation variance, low bias and consistent bias for



Figure 5.1: Estimation errors and variances for the histogram based Kullback-Leibler divergence estimator. The underlying theoretical distribution has variance $\sigma^2 = 2$. *P* denotes the pre-loading constant applied to avoid divide by zero issues in the estimator.

low signal to noise ratios (I.e. when the parameter of the input signal is close to the parameter of the theoretical underlying distribution.) It can be seen from the figures that the estimation variance remains almost constant (Although appears to be slightly decreasing with decreased preloading constant), hence the important evaluation criterion is to have the lowest possible bias, while at the same time having consistent bias when the Rayleigh parameters are in a close range of the theoretical parameter. The important range requiring consistency has been chosen to be SNR $\in [-\infty, -5]$ dB for the given simulation. This ensures that the algorithm provides the same bias for a range of low signal to noise ratios. When the SNR is high, consistent bias becomes less important because the KL divergence often will be several orders of magnitude above the detection threshold. From the figures, it can be seen that using a preloading constant of 0.01 and 18 bins in the histogram, provides a satisfactory tradeoff for the previously mentioned criteria. Thus these values are chosen permanently for all simulations provided in this report. Note that the choice of parameters is affected by the number of samples used in the estimation. 2048 samples have been utilized in these simulations as this is the block length used for detection in the simulations that are to be presented later in this report. Increasing the number of samples provides simulation results following the same trend as seen in figure 5.1, but yielding a different choice of parameters.

The observant reader has probably also noted that the estimation variance is strictly decreasing with reduced number of histogram bins. This may appear odd at first, since one expects the performance to rely on the histogram accurately approximating the underlying probability density function, but turns out to be correct. Recall that for a constant false alarm rate detector, the distribution under H_0 must be considered (I.e. when the input samples come from the underlying noise distribution only.) Thus the Kullback-Leibler divergence is supposed to be 0, which requires $\hat{f}(y_i)/f(y_i) = 1 \forall i$. Note that this result only requires the estimates $\hat{f}(y_i)$ to be as accurate as possible, and is independent of the number of estimates (I.e. number of bins). Hence it becomes apparent that reducing the number of bins will reduce the estimation variance since fewer bins give a lower variance of $\hat{f}(y_i)$. But as expected, the bias increases dramatically (I.e the estimation accuracy is reduced because the histogram does not give a good approximation to the underlying distribution) when lowering the number of bins when the two distributions are not identical (I.e. in these figures when SNR $\neq -\infty$).

The author has not been able to provide a solid theoretical analysis of the histogram based Kullback-Leibler divergence estimator, but a series of Monte Carlo simulations has provided very good insight in its behavior. The estimator appears to have bias and variance

$$\mu_{\text{KLHIST}} = f_1(N_\Delta, P, N) \tag{5.3}$$

$$\sigma_{\text{KLHIST}}^2 = f_2(N_\Delta, P, N) \tag{5.4}$$

where $f(N_{\Delta}, P, N)$ here denotes a function of the number of bins N_{Δ} , the preloading constant P and the number of samples N. Simulations have revealed a striking structure for these two functions. After extensive analysis of Monte Carlo simulations, the bias and variance seemingly can be described as

$$\mu_{\rm KLHIST} = k\theta \tag{5.5}$$

$$\sigma_{\rm KLHIST}^2 = k\theta^2 \tag{5.6}$$

where k and θ are constants common for both the bias and variance. It appears that k depends on the choice of preloading constant and number of bins in the histogram (And possibly other unknown factors that have not been identified), while $\theta = 1/N$, recall that N is the number of received samples. For the parameters chosen for the simulations in this thesis, $N_{\Delta} = 18$ and P = 0.01, $k \simeq 8.51$. A fit between these heuristic theoretical parameters compared to simulated bias and error variance is extremely accurate. This will be shown after the following section. However as mentioned previously, these parameters are heuristic and the author has unfortunately not been able to prove them nor derive equations to calculate them. But the structure of these parameters will be further explored in the following paragraph.

When looking at (5.5) and (5.6) one recognizes the parameters to be those of a Gamma distribution. This corresponds well with empirical results. The left figure seen in 5.2 shows a Quantile-Quantile plot of $1 \cdot 10^4$ Kullback-Leibler divergence estimates versus the quantiles of a theoretical Gamma distribution with parameters k = 8.51 and $\theta = 1/N = 1/2048$. The Kullback-Leibler divergence estimated is the one from a vector of 2048 samples drawn from a Rayleigh distribution with parameters $\sigma^2 = 2$ to the corresponding theoretical Rayleigh distribution. It is seen that a very good model for the bias, variance and distribution for the histogram based Kullback-Leibler divergence estimator under H_0 has been devised with the aid of simulations. Hence it is straightforward to determine a detection threshold for a constant false alarm detector. Deriving a detection threshold for a constant false alarm detector for the histogram based Kullback-Leibler estimator is straightforward when assuming the Kullback-Leibler divergence estimates under H_0 to have a Gamma distributed error as presented above. The threshold is given by solving

$$P_{\rm FD} = \int_{\eta_{\rm KLHIST}}^{\infty} f_{\hat{\kappa}|H_0}(y) \mathrm{d}y \tag{5.7}$$

where $f_{\hat{\kappa}|H_0}(y)$ is the conditional distribution of the Kullback-Leibler divergence estimate $\hat{\kappa}$ under H_0 . When making the assumption that the Kullback-Leibler divergence estimate $\hat{\kappa}$ has a Gamma distribution with shape parameter k and scale parameter $\theta = 1/N$ under H_0 , (5.7) can be easily evaluated and reduces to

$$\eta_{\text{KLHIST}} = F_{\gamma(k,\theta)}^{-1} (1 - P_{\text{FD}})$$
(5.8)

where $F_{\gamma(k,\theta)}^{-1}$ is the inverse cumulative density function of a Gamma distribution with parameters k and θ . For the simulations performed in this report, the block length is N = 2048 and the shape parameter was evaluated to 8.51 through Monte



(a) Empirical Data (Y) VS. Th. Gamma Distribu-(b) Empirical KL Divergence Estimate Distribution (X) tions

Figure 5.2: Left: Figure shows how well the gamma distribution fits the histogram based KL divergence estimate when the input data comes from the theoretical distribution. Right: Densities of histogram based KL divergence estimates.

Carlo simulations when the preloading constant is P = 0.01 and the number of histogram bins $N_{\Delta} = 18$.

The Gamma distribution hypothesis should have been explored further, which might have been the key to a rigorous theoretical analysis of the histogram based Kullback-Leibler divergence estimator. Unfortunately, due to time constraints, it is not possible to pursue this further. However, a brief discussion presenting a possible idea for research is given in the future potential chapter, section 10.1.

This section presented a spectrum sensing algorithm based on estimating the Kullback-Leibler divergence between the empirical distribution of the envelope of the received signal and a theoretical Rayleigh distribution using a histogram. The author has not been able to derive a complete theoretical analysis of the approach, but through simulations, it has been shown how the distribution of the Kullback-Leibler divergence estimate under H_0 appears to follow a Gamma distribution, and how a theoretical expression for the constant false alarm detection threshold can be derived when assuming this to hold. The following section will present the other sensing algorithm, which is based on estimating a Cumulative Density Function (CDF) rather than a Probability Density Function (PDF).

5.2.2 CDF Based Detector

With the ideas presented in [36] as motivation, a novel approach to estimate KL divergence, applicable to the problem at hand has been fully developed by the author as part of this research. This approach is developed with a full rigorous

theoretical foundation. The estimator requires N to be large. It is given as

$$\hat{D}(f||g) = -\gamma - \frac{1}{N} \sum_{i=1}^{N} \ln(N\Delta G(y_i))$$
(5.9)

where $\gamma \simeq 0.577215$ is the Euler-Mascheroni constant, $\Delta G(y_i) = G(y_i) - G(y_{i-1})$ and G denotes the cumulative density function such that g(y) = G'(y). Recall that g is the conditional distribution under H_0 (I.e. the noise distribution). If the received samples $\mathbf{y} = [y_1 \cdots y_N]$ come from G (I.e. f = g), \hat{D} has an asymptotically normal distribution and the bias and variance of the estimator is

$$\lim_{N \to \infty} \mu_{\hat{D}} = 0 \tag{5.10}$$

$$\sigma_{\hat{D}}^2 = \frac{1}{N} \left(\frac{\pi^2}{6} - 1 \right) \quad \text{for large } N \tag{5.11}$$

A proof can be found in appendix B.1.

With the above results it is straightforward to derive a threshold for a constant false alarm detector. The derivation of the threshold is done by simply following standard procedures for inverting the CDF of a Gaussian distribution with parameters found in (5.10) and (5.11). For intuition behind this consult basic literature on detection theory, [54] in its entirety is recommended. The detection chapter in [37] is also a good introduction, and approaches the problem of detection from a more application specific viewpoint. The threshold is

$$\eta_{\text{KLCDF}} = \sqrt{\frac{\pi^2/6 - 1}{N}} Q^{-1}(P_{\text{FD}});$$
 (5.12)

Figure 5.3 shows simulation results demonstrating the estimation performance of both the CDF and the histogram based Kullback-Leibler divergence estimators under H_0 . The figure shows that while the CDF based algorithm is asymptotically unbiased, it suffers from higher estimation variance than the histogram based one. It can also be seen from figure 5.3 that the heuristic theoretical analysis of the histogram based KL divergence estimator is consistent with the simulation results. To a certain degree it is surprising that the intuitively simple histogram based estimator has several orders of magnitude lower estimation variance than the CDF based estimator. However, remember that the histogram based estimator suffers from bias (Recall that a biased estimator can have variance lower than the Cramer-Rao bound, which represents a lower bound on the mean squared error of an unbiased estimator [10]).

5.3 Implementation and Complexity

The previous section introduced two proposed spectrum sensing detectors based on Kullback-Leibler divergence estimation. This section will discuss the implementation of the actual algorithms by presenting pseudo code. It will also address performance in terms of computational complexity.



Figure 5.3: Theoretical and simulated bias and variance for two different estimators for estimating Kullback-Leibler divergence between the envelope of complex Gaussian noise and a theoretical Rayleigh distribution. Both distributions have Rayleigh parameter $\sigma^2 = 2$.

5.3.1 Pseudo Code

This section will present pseudo code of the proposed spectrum sensing algorithms.

The algorithm based on utilizing the CDF is straightforward to understand by considering (5.9). Pseudo code for an algorithm estimating the Kullback-Leibler divergence based on this method is given as

 $\text{KL-CDF}(S, F_p)$

- 1 // S is array of input samples, F_p is array containing parameters defining a
- 2 // theoretical cumulative density function F
- 3 S = PROCESS(S) // pre-process the input samples.

4 // E.g.
$$S[i] = \sqrt{\operatorname{Re} \{S[i]\}^2 + \operatorname{Im} \{S[i]\}^2}$$

- 5 $S = \operatorname{Sort}(S)$
- 6 REMOVE(S) // Remove identical entries in S
- 7 N = S. length
- 8 $\hat{\kappa} = 0$
- 9 **for** i = 2 **to** N

```
10 \hat{\kappa} - = \ln \left( F(S[i]) - F(S[i-1]) \right) /\!\!/ F is a cumulative density function
11 /\!\!/ defined by the parameters in F_p
```

- 12 $\hat{\kappa} = \hat{\kappa}/N \ln N \gamma \not / \gamma$ is the Euler-Mascheroni constant,
- 13 // where $\gamma \simeq 0.577215664901533$

```
14 return \hat{\kappa}
```

It can be seen from the pseudo code above that the algorithm is basically a direct computation of (5.9). However, two implementation issue should be noted. In line

6, a function named *Remove* is called. The purpose of this function is to remove identical entries in the vector of input samples S. Identical entries must be removed to prevent F(S[i]) - F(S[i-1]) = 0, which will yield $\hat{\kappa} = -\infty$. Removing the identical samples is just a security measure, as for a finite number of samples, the probability of two samples being absolutely identical is small. In the MatLab[®] simulations performed during this research, no samples are removed. When devising this algorithm, it was assumed that the input samples came from a continuous distribution. In retrospect, it is obvious that this was an unfortunate assumption. Implications of this will be discussed later in this report. The second issue is implementation specific. The function *Process* depends on which distribution that is being used for the KL divergence estimation. Possible choices are for example signal amplitude (I.e. No processing function is needed), signal magnitude (The example given in the pseudo code) or the distribution of some correlation estimate (In which case the processing would be to estimate this correlation).

Estimating the Kullback-Leibler divergence utilizing histograms can be done in a number of ways. This pseudo code describes the actual implementation used in this research.

 $\text{KL-HIST}(S, F_p, N_{\Delta}, p)$

 $/\!\!/ S$ is array of input samples, F_p is array containing parameters defining a 1 $\mathbf{2}$ // theoretical cumulative density function F, N_{Δ} is the number of histogram 3 # bins, p is the preloading constant. 4 S = PROCESS(S) // pre-process the input samples.# E.g. $S[i] = \sqrt{\text{Re}\{S[i]\}^2 + \text{Im}\{S[i]\}^2}$ 5 $S_{min} = MIN(S)$ 6 7 $S_{max} = MAX(S)$ 8 $\Delta = (S_{max} - S_{min})/(N_{\Delta} - 1)$ 9 Get histogram bin centers: for i = 1 to N_{Δ} 10 $x[i] = S_{min} + (i-1) \cdot \Delta$ 11 H = HISTOGRAM(S, x) // Get histogram with bins centered at points in x12 $\hat{p}_1 = (S[1 \dots S. length] + = p) / \sum (S[1 \dots S. length] + = p)$ 1314for m = 1 to N_{Δ} 15**if** m == 1 $\hat{p}_2[m] = F\left(x[m] + \Delta/2\right)$ 16elseif $m == N_{\Delta}$ 17 $\hat{p}_2[m] = F(x[m] - \Delta/2)$ 1819else $\hat{p}_2[m] = F(x[m] + \Delta/2) - F(x[m] - \Delta/2)$ 2021 $\hat{\kappa} = 0$ 22for i = 1 to N_{Δ} $\hat{\kappa} + = \hat{p}_1[i] \ln \left(\hat{p}_1[i] / \hat{p}_2[i] \right)$ 23**return** $\hat{\kappa} \not\!\!/$ Estimate of $D(p_1||p_2)$ 24Note that the histogram based algorithm could have been implemented in a number of ways, depending on how to choose the bin width and number of bins.

For further insight, the actual MatLab[®] implementations of the algorithms can be found in appendix A.1 and A.2.

5.3.2 Complexity Analysis

The previous section presented pseudo code for the two proposed algorithms. This section will provide a brief discussion on computional complexity. It will also provide simulation results assessing the performance in terms of execution time for the two proposed algorithms in comparison with the reference algorithms described in the background chapter.

Complexity terminology will be the asymptotic *O*-notation, which is standard when analyzing algorithms. For readers who are not familiar with this notation, it will be briefly introduced. The notation is used to describe an asymptotic upper bound, and is defined as

$$O(g(n)) = \{f(n) : \exists \text{ positive constants} \\ c \text{ and } n_0 \text{ such that } 0 \leq f(n) \leq cg(n) \forall n \geq n_0\}$$

$$(5.13)$$

This definition is taken from [9]. This book is an excellent reference on algorithms and analysis of algorithms.

It is difficult to say anything exact about the computational complexity of the proposed algorithms since this depends on the implementation of the sub functions. However, when considering the pseudo code, some main points can be noted. Complexity of the KL CDF algorithm is dominated by the sorting operation. Typical high performance general sorting algorithms (E.g. Quicksort) run in $O(N \log_2 N)$ [9]. This is the dominating operation, as the other operations of KL CDF run in O(N) time (The running time of *Remove* depends on the implementation, but can in general be done in O(N) time since it only requires iterating the array once).

It is harder to pinpoint a dominating operation for KL HIST, but depending on the implementation of the sub routines, KL HIST appears to be dominated by O(N) terms. Hence one should normally expect KL HIST to perform better than KL CDF when the number of input samples N grows. Also note that the actual Kullback-Leibler computation is O(N) for KL CDF, while it is only $O(N_{\Delta})$ for KL HIST. Recall that typically $N_{\Delta} \ll N$. Since this computation is a significant part of the algorithm, this also indicates faster running time for KL HIST.

To get an impression of the relative performance, the execution times have been recorded for various input sizes. The input signal is circularly symmetric complex Gaussian noise. Execution time has been measured by using the MatLab[®] stop-watch function *tic/toc*. Simulations were performed on a laptop computer with a single core 1.6 GHz CPU. Results from the simulations can be seen in figure 5.4. The results seen in the figure support the discussion above, indicating that the running time of the KL CDF algorithm dominates when the number of input samples increases. It is also seen that both KL based algorithms have execution times that are of one to two orders of magnitude greater than the reference algorithms. This is to be expected as the amount of computation to be performed for the energy



Figure 5.4: Complexity analysis of four spectrum sensing algorithms.

detector and the autocorrelation based detector is very limited (Approximately N complex multiplications and N additions in total for each of the energy detector and the IM autocorrelation detector respectively).

This section provided a discussion on the computational complexity of the proposed algorithms. It was argued that the KL HIST algorithm asymptotically should have a better running time than the KL CDF algorithm. This argument was further strengthened by simulation results. The simulations also showed that the KL algorithms have running times of approximately one to two order of magnitudes above the two reference algorithms.

5.4 Detection Probability for Detector Based on Kullback-Leibler Divergence for an OFDM Signal in AWGN Channel

Evaluating a theoretical probability of detection for an arbitrary signal is challenging. However, due to the approximate Gaussian distribution of an OFDM signal, the calculation becomes tractable. Assuming the OFDM signal to have a Gaussian distribution is a very common assumption often utilized in the literature, see for instance [6] or [60]. The assumption arises from invoking the central limit theorem on the resulting output from the inverse FFT.

For simplicity when calculating the probability of detection, only an Additive White Gaussian Noise (AWGN) channel is assumed. The received signals under the two detection hypotheses are

$$\begin{aligned} H_0: & y[n] = w[n] \\ H_1: & y[n] = x[n] + w[n] \end{aligned}$$
 (5.14)

where x[n] is an OFDM signal and w[n] is circularly symmetric complex white

Gaussian noise. The signal x[n] and noise w[n] are assumed to be statistically independent, and the inverse FFT used in the OFDM modulation is assumed to be large enough for the Gaussian assumption to hold. y[n] will have a zero-mean Gaussian distribution with variance $\sigma_y^2 = \sigma_w^2$ under H_0 and variance $\sigma_y^2 = \sigma_w^2 + \sigma_x^2$ under H_1 . Thus the envelope of the signal under both H_0 and H_1 has a Rayleigh distribution.

The Kullback-Leibler divergence between two Rayleigh distributions with parameters $\theta_1 = \sigma_{y|H_1}^2/2$ and $\theta_2 = \sigma_{y|H_0}^2/2$ is computed as

$$\kappa_R = \int_0^\infty \frac{y}{\theta_1} \exp\left(-\frac{y^2}{2\theta_1}\right) \ln\left(\frac{\frac{y}{\theta_1} \exp\left(-\frac{y^2}{2\theta_1}\right)}{\frac{y}{\theta_2} \exp\left(-\frac{y^2}{2\theta_2}\right)}\right) dy$$

$$= \int_0^\infty \frac{y}{\theta_1} \exp\left(-\frac{y^2}{2\theta_1}\right) \ln\left(\frac{\theta_2}{\theta_1}\right) + \frac{y^3}{2\theta_1} \exp\left(-\frac{y^2}{2\theta_2}\right) \left(\frac{1}{\theta_2} - \frac{1}{\theta_1}\right) dy$$
(5.15)

The integral is easily solved by first applying the substitution $u = \frac{y^2}{2}$ and then performing integration by parts on the second term, yielding

$$\kappa_R = \frac{\theta_1}{\theta_2} - 1 - \ln\left(\frac{\theta_1}{\theta_2}\right) \tag{5.16}$$

This Kullback-Leibler divergence can be compared to the well known divergence between two Gaussian distributions with mean values μ_1 and μ_2 and variance θ_1 and θ_2 given as [10]

$$\kappa_G = \frac{(\mu_1 - \mu_2)^2}{2\theta_2} + \frac{1}{2} \left(\frac{\theta_1}{\theta_2} - 1 - \ln\left(\frac{\theta_1}{\theta_2}\right) \right)$$
(5.17)

By recalling that a communication signal has zero-mean, it is seen that the Kullback Leibler divergence between the magnitude of $X_1^{(1)} + jX_1^{(2)}$ and $X_2^{(1)} + jX_2^{(2)}$ (Assuming statistical independence between the real and imaginary components) is twice than the divergence between X_1 and X_2 . Where X_1 and X_2 are two independent Gaussian random variables. Hence, as discussed previously, there is motivation to apply the magnitude of the signal rather than the amplitude in the Kullback-Leibler divergence based spectrum sensing detectors.

By using (5.16), and the bias, variance and conditional distributions of the two previously proposed estimation algorithms under H_0 given for the CDF based algorithm in (5.10) and (5.11) and the histogram based algorithm in (5.5) and (5.6) respectively, one can compute a theoretical detection threshold and probability of detection for the estimation algorithms. However, recall from the previous analysis that while the bias, variance and distribution expression for the estimate $\hat{\kappa}$ with the CDF based estimator are analytically proven, the bias, variance and distribution expressions for $\hat{\kappa}$ using the histogram based estimator are only motivated by simulations.

By using the definition of signal to noise ratio,

$$SNR = 10 \log \frac{\sigma_s^2}{\sigma_n^2} \tag{5.18}$$

and

$$\theta_1 = \frac{\sigma_x^2 + \sigma_w^2}{2}$$
$$\theta_2 = \frac{\sigma_w^2}{2},$$

one can, after some basic algebraic manipulation, substitute into (5.16) and obtain the following theoretical Kullback-Leibler divergence between the distributions under H_1 and H_0 as a function of the SNR for an OFDM signal in AWGN

$$\kappa_R = 10^{\text{SNR/10}} - \ln\left(10^{\text{SNR/10}} + 1\right)$$
(5.19)

The theoretical detection probability of an OFDM signal in AWGN is computed as ∞

$$P_{\rm D}({\rm SNR}) = \int_{\eta}^{\infty} f_{\hat{\kappa}|\kappa_R}(y) \,\mathrm{d}y \tag{5.20}$$

where κ is the true Kullback-Leibler divergence and $\hat{\kappa}$ is the estimate. κ is a constant, while $\hat{\kappa}$ is a random variable.

By applying the threshold given in (5.12) and the theoretical KL divergence in (5.19) one obtains the following probability of detection for the CDF detector

$$P_{\rm D} = \frac{1}{\sqrt{2\pi}\sigma_{\hat{D}}} \int_{\eta_{\rm KLCDF}}^{\infty} \exp\left(-\frac{(\hat{\kappa} - \kappa_R)^2}{2\sigma_{\hat{D}}^2}\right) d\hat{\kappa}$$

$$= Q\left(Q^{-1}(P_{\rm FD}) - \frac{\kappa_R}{\sigma_{\hat{D}}}\right)$$
(5.21)

where $\sigma_{\hat{D}}$ is given in (5.11) (Recall from the analysis of the CDF estimator that its error has an asymptotic normal distribution with variance $\sigma_{\hat{D}}$ and zero mean under H_0).

A similar approach applied to the histogram detector, utilizing (5.19) and the heuristic approach assuming a Gamma distribution with parameters k and $\theta = 1/N$ as discussed previously, yields

$$P_{\rm D}({\rm SNR}) = \frac{1}{\theta^k \Gamma(k)} \int_{\eta_{\rm KLHIST}-\kappa_R}^{\infty} \hat{\kappa}^{k-1} \exp\left(\frac{-\hat{\kappa}}{\theta}\right) d\hat{\kappa}$$

= 1 - F_{\gamma(k,\theta)} (\eta_{\rm KLHIST}-\kappa_R) (5.22)

where $F_{\gamma(k,\theta)}$ denotes the cumulative density function of a Gamma distribution with shape parameter k and scale parameter θ , κ_R is a function of SNR and is given by (5.19) and the detection threshold η_{KLHIST} is given by (5.8). This calculation assumes that the distribution of the estimation error under H_1 for an OFDM signal in AWGN is the same as under H_0 . This can be seen from the right hand plot in figure 5.2 to obviously not hold in general. However, it is assumed to provide a good indication of theoretical performance as the right handed tail behavior of the distributions, which is what is important for the P_{D} calculation, behaves approximately similarly under H_1 as under H_0 for low signal to noise ratios. For larger signal to noise ratios, making the assumption $f_{\hat{\kappa}|H_0}(y) = f_{\hat{\kappa}|H_1}(y)$, is obviously wrong, but it typically has little effect on the correctness of the calculations as the Kullback-Leibler divergence in this region is very high relative to the estimation variance. With the above expressions for the probability of detection,



Figure 5.5: Theoretical detection results for OFDM signal in AWGN channel. For the Receiver Operating Characteristic (ROC) curves, SNR is -11 dB.

the expected detection performance can be evaluated. Figure 5.5a shows receiver operating characteristics for the detection and figure 5.5b shows $P_{\rm D}$ versus SNR for an OFDM signal in AWGN. It can be seen that despite the independence of signal or noise variance in the expression for the estimator variance, the Kullback-Leibler divergence based detectors still appear to perform worse than the energy detector. A number of simulations will be performed to confirm and further assess these performance findings.

5.5 Kullback-Leibler Divergence Based Detectors Under Noise Uncertainty

During theoretical design of communication systems, one often assumes the underlying noise to be stationary circularly symmetric complex with a white Gaussian distribution. This assumption has been shown to be weak in reality, see for instance [5]. In reality, the noise variance will usually not be completely stationary. The assumption about the distribution of the noise is also known to be weak. Impulsive noise, aliasing from imperfect filters, leakage from other spectral bands etc. all add to the existing thermal noise, and in many cases create a distribution for the total noise which deviates from Gaussian. Recall how section 3.3.1 introduced the concept of noise uncertainty to create an effective model to treat the problem. Section 3.3.1 treated noise uncertainty for the energy detector. This section will provide a similar analysis in order to investigate the effect of noise uncertainty on the Kullback-Leibler divergence based detectors.

There is a major difference in the nature of the energy and the KL based detectors. The energy detector suffers under noise uncertainty because computing the detection threshold requires knowledge of the underlying noise variance. The two proposed KL detectors on the other hand only rely on a priori knowledge in order to compute the threshold. However, the KL divergence estimators require exact knowledge of the underlying theoretical noise distribution. Hence, uncertainty in this knowledge will affect the Kullback-Leibler divergence estimate, and not the detection threshold.

Recall from section 3.3.1 that noise uncertainty is modeled by letting the actual noise variance be confined within a set given by a nominal noise variance and an uncertainty parameter ρ such that $\sigma_n^2 \in \left[\frac{1}{\rho}\sigma^2, \rho\sigma^2\right]$. To get ρ in dB, the following definition is used $\rho_{dB} = 10 \log_{10} \rho$. The effect of uncertainty on the KL divergence based detectors can be understood by considering the equations presented in section 5.4 and including distribution uncertainty. Recall that the Kullback-Leibler divergence between an OFDM signal and circularly symmetric complex white Gaussian noise can be written as

$$\kappa_R = \frac{\theta_1}{\theta_2} - 1 - \ln\left(\frac{\theta_1}{\theta_2}\right) \tag{5.23}$$

where

$$\theta_1 = \left(\sigma_x^2 + \sigma_w^2\right)/2$$

and

$$\theta_2 = \rho \sigma_w^2 / 2.$$

Notice that the uncertainty, represented by the parameter ρ is only present in the expression for θ_2 (I.e. the theoretical parameter used by the detector), since θ_1 represents the true received signal. Denote the linear scale SNR

$$\sigma_x^2 / \sigma_w^2 = \gamma_{\rm lin}$$

By performing substitution and simple algebra, one can easily represent κ_R as a function of γ_{lin} and ρ

$$\kappa_R = \frac{1}{\rho} \left(\gamma_{\rm lin} + 1 \right) - 1 - \ln \left(\gamma_{\rm lin} + 1 \right) + \ln \rho \tag{5.24}$$

Under no noise uncertainty ($\rho = 1$), (5.24) decays monotonically with decreasing SNR. Differentiating (5.24) in terms of γ_{lin} , yields

$$\frac{\partial \kappa_R}{\partial \gamma_{\rm lin}} = \frac{1}{\rho} - \frac{1}{\gamma_{\rm lin} + 1} \tag{5.25}$$

Since by definition $\rho \ge 1$ and $\gamma_{\text{lin}} \ge 0$, (5.24) can when there is no uncertainty, by considering (5.25), be seen to be monotonically decaying for reduced SNR ($\rho = 1$),

while it decreases to a minimum before increasing again when there is uncertainty. This can cause significant trouble for the detector. If the estimate increases sufficiently when the SNR is decreased, it will cross the detection threshold and trigger a false detection. This is expected to be expressed as a breakdown behavior characterized by a probability of detection that is not monotonically decaying with decreased SNR. This can clearly be seen in figure 5.6. The figure illustrates the theoretical KL divergence between an OFDM signal and circularly symmetric complex white Gaussian noise under no noise uncertainty versus the KL divergence under an uncertainty of $\rho = 2$ dB. It is clearly seen how false detection will occur when the SNR drops below approximately 6 dB. Detection performance under



Figure 5.6: Figure shows the effect on KL divergence based detection under 2 dB noise uncertainty. Threshold is computed for the KL CDF algorithm with $P_{\rm FD} = 0.05$ utilizing 2000 samples.

noise uncertainty will be thoroughly investigated through simulations that are to be presented in chapter 6. Note that this analysis applied the upper limit (ρ) of the uncertainty interval to illustrate the problems occurring for the KL divergence based estimators under noise uncertainty. Analysis could also have been performed for the lower limit $(1/\rho)$, or any other arbitrary value in the interval.

5.6 Summary of Proposed Algorithms

This section will briefly summarize the previous important findings for the two proposed algorithms.

5.6.1 KL CDF

The proposed algorithm depending on estimating the Kullback-Leibler divergence through the cumulative density function is given on closed form as

$$\hat{D}(f||g) = -\gamma - \frac{1}{N} \sum_{i=1}^{N} \ln(N\Delta G(y_i))$$
(5.26)

where $\gamma \simeq 0.577215$ is the Euler-Mascheroni constant, N is the number of input samples, $\Delta G(y_i) = G(y_i) - G(y_{i-1})$ and G denotes the cumulative density function such that g(y) = G'(y). Recall that g is the conditional distribution under H_0 (I.e. the noise distribution). The detection threshold for a constant false alarm detector is given as

$$\eta_{\text{KLCDF}} = \sqrt{\frac{\pi^2/6 - 1}{N}} Q^{-1}(P_{\text{FD}}); \qquad (5.27)$$

For an OFDM signal in AWGN channel, the probability of detection is given as

$$P_{\rm D} = Q \left(Q^{-1}(P_{\rm FD}) - \frac{\kappa_R}{\sigma_{\hat{D}}} \right)$$
(5.28)

with $\sigma_{\hat{D}}$ given in (5.11). Q is the standard cumulative Gaussian which is defined in (3.11).

Pseduocode for the actual algorithm was given in section 5.3.1 and the corresponding MatLab[®] implementation can be found in appendix A.1.

5.6.2 KL HIST

The theoretical framework for the histogram based detector has not been fully proven such as the framework for the CDF based detector. However, for the KL HIST detector, the following results have been motivated empirically through simulations. A constant false alarm detection threshold is given as

$$\eta_{\text{KLHIST}} = F_{\gamma(k,\theta)}^{-1} (1 - P_{\text{FD}}) \tag{5.29}$$

where $F_{\gamma(k,\theta)}^{-1}$ is the inverse cumulative density function of a Gamma distribution with parameters k and $\theta = 1/N$. For the simulations performed in this report, the block length N = 2048 and that the shape parameter k was evaluated to 8.51 through Monte Carlo simulations, when the preloading constant P = 0.01 and the number of histogram bins $N_{\Delta} = 18$. Using the same parameters an approximate probability of detection for an OFDM signal in AWGN is given as

$$P_{\rm D}({\rm SNR}) = 1 - F_{\gamma(k,\theta)} \left(\eta_{\rm KLHIST} - \kappa_R \right)$$
(5.30)

where $F_{\gamma(k,\theta)}$ denotes the cumulative density function of a Gamma distribution with shape parameter k and scale parameter θ and κ_R is given by (5.19) and the detection threshold η_{KLHIST} is given by (5.8).

Pseduocode for the actual algorithm is given in section 5.3.1 and the corresponding MatLab[®] implementation can be found in appendix A.2.

Chapter 6 Simulations

Previous chapters have introduced the problem of blind spectrum sensing in cognitive radio, presented an analysis of problems with existing detectors utilizing sample average based estimators and proposed two novel algorithms based on Kullback-Leibler divergence. The proposed detectors were hoped to improve performance, especially in the low signal to noise region. This chapter will provide a number of simulations aimed at assessing the performance of the proposed detectors in comparison with the two reference detectors. The chapter is split in two main sections. The first section will introduce three common simulation scenarios. All scenarios have different properties in order to get a thorough impression of detector performance. Section two presents the results for the simulations. The chapter ends with a thorough discussion of the simulation results.

6.1 Simulation Scenarios

This section will introduce the common simulation scenarios used to test the detection algorithms. Three different scenarios, with different properties have been chosen to evaluate spectral detection performance. The reader is assumed to be familiar with common digital modulation and communication principles. All simulation scenarios follow the Monte Carlo principle, where detection results are obtained as the average of a number of simulations. For each iteration of the Monte Carlo simulation, a test statistic is computed on the basis of the signal samples in one block. A binary decision is made by comparing the test statistic to a predetermined detection threshold. The threshold is computed for the detectors to have a probability of false detection $P_{\rm FD} = 0.05$.

6.1.1 Scenario 1: Raised Cosine Pulse Shaped QAM Signal in AWGN Channel

The first scenario is the simplest. A signal consisting of 512 16-QAM information symbols is generated by mapping random integers from 0 to 15 to a Gray coded

16-QAM constellation. After modulation, the signal containing the complex information symbols is pulse shaped with a raised cosine filter. The filter has roll-off parameter $\beta = 0.5$, and upsamples the signal by a factor of 4. Hence the total length of one signal block is $4 \cdot 512 = 2048$ samples. Complex white Gaussian noise is then added to the signal. A Welch power spectrum estimate of a one block realization of the signal can be seen in figure 6.1.



Figure 6.1: Welch estimates of the power spectrum. Channel is AWGN with SNR= 8 dB for both signals.

6.1.2 Scenario 2: OFDM Signal in AWGN Channel

OFDM is the modulation of choice for the two last simulation scenarios to be used as evaluation tools in this thesis. In OFDM, a wideband channel is divided into a set of narrowband orthogonal subchannels. If the inverse bandwidth of each subchannel is much larger than the delay spread of the channel, the subchannels can be approximated as having independent flat fading. This prevents the need for expensive and complex wideband equalization schemes. OFDM modulation can also be efficiently implemented through digital signal processing due to the FFT algorithm. These two reasons have made OFDM a very popular modulation scheme in recent wireless standards [18].

While the QAM signal in scenario 1 was arbitrarily picked for illustrative purposes, the OFDM signal used for scenario 2 and 3 follows an existing standard. The specific standard is called DVB-T. DVB-T abbreviates *Digital Television Broadcast* - *Terrestrial*, and as the name implies is a standard for wireless digital transmission of TV signals. The standard is administered by the European Telecommunications Standards Institute (ETSI). The official ETSI web page can be found at [35].

The actual implementation of the DVB-T OFDM signal is largely adapted from an existing MatLab[®] implementation from Georgia Institute of Technology, described and fully documented in [1]. The implementation includes upconversion, downconversion, lowpass filtering and sampling of the transmitted OFDM signal. Figure 6.1 shows the Welch power spectrum estimate of one realization of a signal block in AWGN with SNR = 8 dB. For the Monte Carlo simulation, each signal block consists of one symbol, which in the 2k mode implemented here contains 2048 samples. 500 iterations are performed in the simulation.

6.1.3 Scenario 3: OFDM Signal in AWGN Channel Including Rician Multipath Fading and Shadowing

The third simulation scenario utilizes the same DVB-T OFDM signal as scenario 2, but to make the simulations more realistic, the signal is subjected to Rician multipath fading and shadowing following a log normal distribution. It is assumed that the detection performance in AWGN (I.e. Scenario 1 and 2) will provide a good impression of the performance, but it is necessary to extend the simulations to include signal distortion due to multipath- and shadow fading.

The maximum Doppler shift of the channel is 100 Hz, the K-factor for the Rician fading is 10 (Which represents a very strong line of sight component) and the standard deviation for the log normal shadowing is 10 dB.

Since the fading causes the channel to be time variant, it is necessary to apply longer averaging than in scenario 2 to obtain good simulation results. To reflect this, the number of iterations in the Monte Carlo simulation is increased from 500 to 1000.

6.2 Simulation Results

Simulations are important in assessing the performance of the proposed spectrum sensing algorithms. The previous section presented three different simulation scenarios, chosen to investigate spectrum sensing. The three scenarios provide different attributes aimed at yielding a framework for diverse assessment of the detection algorithm. The simulations are split in two main parts. Part one is a regular assessment of detection performance for the three proposed simulation scenarios, assuming that the underlying noise variance is known to all algorithms. The results from these simulations can be seen in the batch figure 6.2. Part two evaluates the algorithms under noise uncertainty in scenario 1. It was judged sufficient to only include scenario 1 for the noise uncertainty trial to limit the scope of the thesis. However, it is fair to assume that behavior in the other simulation scenarios to a large degree can be inferred from the results from scenario 1 under noise uncertainty combined with the results under no noise uncertainty from scenario 2 and 3. The simulation results for scenario 1 under noise uncertainty can be seen in the batch figure 6.3. It can be noted from figure 6.3, that the IM algorithm has been excluded. This stems from the fact that the IM algorithm does not depend on knowing the underlying noise variance, and therefore does not suffer under noise uncertainty.



Figure 6.2: Monte Carlo simulation results assessing detection performance of a number of spectral detection algorithms. See the section on simulation scenarios for details regarding implementation.
6.2.1 Detection, Scenario 1

This subsection will list the simulation results under scenario 1 as seen in figure 6.2a. All listed dB quantities when addressing relative performance are given as approximate values as they have been obtained through manual inspection of the figures.

The IM algorithm performs the best, with IM(3) at the top with IM(2) and IM(1) following approximately 0.5 dB and 1.5 dB behind respectively. Recall that the index of the IM algorithm denotes how many lags of the estimated autocorrelation function that are used to compute the test statistic. Subsequent to the IM algorithm is the energy detector, with approximately 3 dB reduced performance compared to IM(3). Following IM(3) is the histogram based Kullback-Leibler divergence detector, approximately 5 dB behind IM(3). The worst performance is displayed by the CDF based KL divergence algorithm, which shows a performance reduction of approximately 11 dB compared to IM(3). The $P_{\rm D}$ curves for all detectors can be observed to have very similar slopes.

6.2.2 Detection, Scenario 2

This subsection will list the simulation results under scenario 2 as seen in figure 6.2c. The simulation results in this figure for the energy detector and the two KL divergence based detectors are complemented with solid curves denoting theoretical detection results. All listed dB quantities when addressing relative performance are given as approximate values as they have been obtained through manual inspection of the figures.

It is slightly more difficult to comparatively assess the detectors in scenario 2 than scenario 1 in terms of a quantitative dB measure since the different detectors display different slopes for the $P_{\rm D}$ curves. The curves can be divided in three pairs, where each pair displays a slope different from the two other pairs. The three pairs are; energy detector and IM(1), the two Kullback-Leibler based algorithms and IM(2) and IM(3).

The best performance is obtained from the energy detector. Subsequent is the histogram based Kullback-Leibler divergence algorithm, which has a performance in the range from approximately 2 dB to approximately 5 dB below the energy detector. Following the KL histogram algorithm is IM(1), with a steady performance loss of approximately 7 dB compared to energy detection. The worst performance is displayed by IM(2), IM(3) and KL CDF. IM(2) and IM(3) have virtually identical curves, while KL CDF differs from the IM(2) and IM(3) curves with as much as approximately 4 dB. The KL CDF algorithm is the best of the three when the signal to noise ratio is $SNR \ge -7$ dB, with the opposite result for SNR < -7 dB. In total, IM(2), IM(3) and KL CDF can be seen to perform on average about 9 dB worse than the best performance, which is obtained by the energy detector.

6.2.3 Detection, Scenario 3

Relative detection results for scenario 3 are to a large extent aligned with the previously presented results for scenario 2. This is expected as the underlying signal used is the same. The main difference is in absolute performance, which is caused by the addition of multipath- and shadow fading. It is obvious from figure 6.2b how the absolute detection performance deteriorates when the signal is subjected to channel fading. The $P_{\rm D}$ slopes for all detectors start dropping at higher SNR values than for the AWGN case. While the $P_{\rm D}$ curves started dropping off in the range from approximately -10 dB to about 1 dB for the four detectors in the AWGN channel of scenario 2, all curves start dropping off before 5 dB under the fading applied in scenario 3 (5 dB is the upper limit of the evaluated SNR for all three scenarios).

In terms of relative performance, results are aligned with scenario 2. The energy detector performs the best, with the KL HIST algorithm as number two, with a performance in the range of approximately 1 dB to 4 dB below. Following is IM(1), which has a steady performance approximately 5 dB below the energy detector. As for scenario 2, the worst performance is obtained by IM(2), IM(3) and the KL CDF algorithm. However, there is one difference when comparing to scenario 3. In scenario 2, the $P_{\rm D}$ curves of KL CDF and the two IM detectors crossed. This does not happen within the SNR range visible in figure 6.2b. IM(2) and IM(3) perform approximately 9 dB below the energy detector. The bottom performance is obtained by the KL CDF algorithm at approximately 10 dB to 11 dB below the energy detector.

The previous section has presented a summary of detection results for the three simulation scenarios when the underlying noise variance is assumed to be known by the detector. In the following section, simulation results under noise uncertainty will be addressed.

6.2.4 Detection Under Noise Uncertainty, Scenario 1

A number of spectrum sensing algorithms assume knowledge of the underlying noise variance. The noise distribution might be unknown or non-stationary, which will degrade the performance of the sensing. [50] introduced the term *noise uncertainty* set to try to understand and quantify this degradation. Recall that the main points of this analysis are presented in the background chapter, section 3.3.1. This section will assess the performance of the proposed detectors along with the energy detector as reference under noise uncertainty. The IM algorithm does not depend on noise variance knowledge, and is hence not affected by noise uncertainty. Results from the simulations can be seen in the batch figure 6.3. There is one sub figure for each detector, with the energy detector seen in figure 6.3c, the KL CDF algorithm in figure 6.3b and the KL HIST algorithm in 6.3a. Each sub figure has a black solid line as reference, denoting performance of the respective detectors without noise uncertainty. Performance under noise uncertainty is assessed through giving the detector the two worst case variance figures under a given uncertainty set. Recall that under noise uncertainty, the actual noise distribution is confined within



Figure 6.3: Scenario 1 Monte Carlo simulation assessing energy detector under noise uncertainty. Colors denote upper and lower uncertainty level of the uncertainty interval $[\frac{1}{\rho}\sigma^2, \rho\sigma^2]$, where ρ is given in the legend.

a set given by a nominal noise variance and an uncertainty parameter ρ such that $\sigma_n^2 \in \left[\frac{1}{\rho}\sigma^2, \rho\sigma^2\right]$. To get ρ in dB, the following definition is used $\rho_{\rm dB} = 10\log_{10}\rho$. In the simulation, the noise uncertainty definition has been interpreted as σ^2 being the true noise variance, and σ_n^2 as being the estimate of the noise variance that is used by the detector. The lower- and upper worst case noise variance estimates, $\frac{1}{\rho}\sigma^2$ and $\rho\sigma^2$ respectively, are used by the detector in the simulations, denoted in the figures by $\rho_{\rm dB}$ L (Lower) and $\rho_{\rm dB}$ U (Upper).

Three different values of the uncertainty parameter ρ are used in the simulations. They are given in table 6.1. The table also gives an approximate required number

ρ		Required Block Length, $P \simeq 0.99$		
dB	Linear	Ν		
0.1	$\simeq 1.0233$	$1.2 \cdot 10^4$		
0.5	$\simeq 1.1220$	450		
1	$\simeq 1.2589$	100		

Table 6.1: Table gives noise uncertainty parameters used for the simulations in figure 6.3 in dB and linear scale. It also gives an approximate required block length for the number of samples needed to estimate the noise variance so it is confined within the given uncertainty set with an approximate probability of 0.99.

of samples needed to estimate the noise variance such that the estimate is confined within the uncertainty set with an approximate probability of 0.99. The number of required samples N given in table 6.1 has been computed using the following equation

$$P_{1/2} = \frac{1}{\sqrt{2\pi}\sigma_{\epsilon}} \int_{\rho\sigma^{2}}^{\infty} \exp\left(-\frac{1}{2} \frac{\left(x - \sigma^{2}\right)^{2}}{\sigma_{\epsilon}^{2}}\right) dx$$
$$= Q\left(\left(\rho - 1\right)\sqrt{N}\right)$$
(6.1)

Here, σ^2 denotes the true noise variance, N the number of samples needed for the estimate, $P_{1/2}$ the half sided probability for the estimate to be confined within the noise uncertainty set and σ_{ϵ}^2 is the estimator variance. For circularly symmetric complex white Gaussian noise $\sigma_{\epsilon}^2 = \sigma^4/N$ (See section 4.1). Hence the required number of samples to estimate the noise variance under a given noise uncertainty ρ , if one requires the estimate to be confined within the set with probability $1-2P_{1/2}$, is given as

$$N \simeq \left(\frac{Q^{-1}\left(P_{1/2}\right)}{\rho - 1}\right)^2 \tag{6.2}$$

The use of \simeq stems from the fact that the noise uncertainty set is not symmetric $\left(1 - \frac{1}{\rho} \neq \rho\right)$, but symmetry has been assumed for simplicity in the calculation since only a rough estimate of N is needed to illustrate the impact of the noise uncertainty.

Figure 6.3 shows that the effect of noise uncertainty for the spectrum sensing algorithms varies with the uncertainty parameter ρ . For the smallest value of the parameter $\rho = 0.1$ dB, the energy detector has the worst performance, with approximate deviation from the no uncertainty reference ranging from -4 dB to 7 dB. The KL CDF algorithm on the other hand, performs the best with a minor deviation approximately ranging from -0.5 dB to 0.5 dB.

However, the relative performance in terms of sensitivity to noise uncertainty changes somewhat when the value of ρ increases. It can be seen from the figures that while the energy detector has a performance that consistently degrades through widening the area between the upper and lower $P_{\rm D}$ curves for increased ρ , the Kullback-Leibler divergence based detectors suffer from a breakdown when the uncertainty is large enough. The breakdown can be seen as an inconsistency where the probability of detection suddenly starts to increase when the signal to noise ratio decreases below a certain threshold. This behavior occurs for both KL based algorithms, but can be seen from the figure to be more severe for the KL histogram based detector.

The previous sections went through results from a number of simulations aimed at assessing detection performance and detector robustness towards noise uncertainty. The two proposed spectrum sensing algorithms are compared with the two references. Reasons behind these results, along with potential implications, will be thoroughly discussed in the following section.

6.3 Discussion of Simulation Results

A common simulation framework has been provided to evaluate the two proposed spectrum sensing algorithms in comparison with the two reference detectors. The following sections will discuss and explain the findings from these simulations.

6.3.1 Discussion of Simulation Results Under No Noise Uncertainty

In terms of the reference detectors, the behavior is as expected. The two different test signals, the OFDM DVB-T signal and the 16-QAM signal have different spectra. Recall that Welch power spectrum density estimates of the two signals were shown in figure 6.1. The detection performance of the two reference detectors can to a large extent be inferred simply by considering the signal spectra. Both signals have approximately flat spectra, but the OFDM signal displays a sharp roll-off compared to the QAM signal. It can also be seen that due to the oversampling in the raised cosine filter for the QAM signal, it has a normalized bandwidth of approximately 0.2π rad/sample, which is much narrower than the OFDM signal, which with no oversampling has a normalized signal bandwidth of approximately 0.8π rad/sample. Two things can be inferred from this. It is expected that the autocorrelation detector will perform better for the QAM signal, since it due to the lower signal bandwidth has an autocorrelation function with a slower decay than the OFDM signal. It should also be expected that more lags of the autocorrelation

function can be applied for detection. The other thing to infer from the signal spectra is to expect that the energy detector is at the top end performance wise for the OFDM signal, since the OFDM signal has a Gaussian distribution and is nearly white. Recall that the energy detector is the optimal detector for detecting a white Gaussian signal in white Gaussian noise. These expectations are confirmed when considering the simulation results seen in figure 6.2. As expected, the autocorrelation detector gives the best performance for the QAM signal (figure 6.2a), and utilizing more lags improves the detection. However, remember that increased performance for utilizing an increased number of lags only occurs until the first zero crossing of the autocorrelation function. The opposite effect of adding more lags to compute the test statistic occurs for the OFDM signal. Due to the large signal bandwidth (I.e. the lack of oversampling in the receiver), the autocorrelation function of the signal decays rapidly, and utilizing more than the first lag of the autocorrelation function actually deteriorates the detection performance (figure 6.2c and figure 6.2b). Further, the expectations of the energy detector for the OFDM signal were also correct. In figures 6.2c and 6.2b, the energy detector can be seen to provide the best performance of all the detectors.

As for the two proposed Kullback-Leibler detection algorithms, relative performance compared to the energy detector was already established mathematically in the theory section. However, when deriving the probability of detection, it was assumed that the conditional distribution of the Kullback-Leibler divergence estimates were the same under H_1 as H_0 . This assumption is known not to be correct, but it was argued that it should be sufficient to obtain good theoretical results for the probability of detection. This assumption can be seen to hold to a large extent when considering the performance for the OFDM signal in AWGN seen in figure 6.2c and 6.2d. For the CDF based algorithm, the simulated detection performance matches the theoretical results with a high degree of accuracy. This is true for both the $P_{\rm D}$ versus $P_{\rm FD}$ curves and the $P_{\rm D}$ versus SNR curves. The curve from the Monte-Carlo simulation can be seen to display slightly worse performance than the theoretical curve in both figures. This deterioration in performance compared to the theoretical results is expected, because when calculating the probability of detection, it was assumed that the KL estimator is unbiased, while it in reality is only asymptotically unbiased (See figure 5.3). However, the discrepancy is so small that for all practical purposes, the theoretical detection performance can be considered correct.

For the histogram based detector, the theoretical results are not as strong as for the CDF based detector. Recall from section 5, addressing the Kullback-Leibler estimators, that while the CDF based algorithm is presented with a complete and rigorous theoretical framework, the results for the histogram based estimator have not been proved theoretically. However, it can be seen from figure 6.2c how simulation results for the KL HIST algorithm are within approximately 0 to 1 dB of the theoretical detection results. The reason for this discrepancy is most likely resulting from assuming the conditional distribution for the estimation error to be the same under H_1 as H_0 , and hence shows that making this assumption is not sufficient. This is strengthened when comparing deviation with the tail behavior of the estimate distributions seen in figure 5.2 (The large difference in tail behavior for the conditional estimate distribution when the SNR> -8 dB has no effect since the KL divergence here is so large that it does not matter whether the threshold is exact or not). It is not until considering the receiver operating characteristic in figure 6.2d, it becomes obvious that the current theoretical analysis is clearly inadequate. At least when the probability of false detection P_{FD} is greater than $\simeq 0.05$. Again, the most likely cause for this is the fact that the distribution of the estimate under H_1 differs from the distribution under H_0 (As seen in figure 5.2).

When considering the simulation results for scenario 3, an obvious fact is observed. It is clearly seen how introducing channel distortion in terms of multipathand shadow fading clearly deteriorates the detection performance. While the detection performance under AWGN dropped rapidly from 1 to $P_{\rm FD}$ over a range of about 5 dB, the slope of the detection curve falls off considerably slower, extending the SNR range of the drop to at least 25 dB (The rest is outside the range of the figure, so there are no grounds to make conclusions on performance). Recall that the Rice factor for the multipath fading in this scenario is $K_{\rm rice} = 10$, and that this corresponds to a very strong line of sight component compared to the multipath components. Hence the Rician multipath fading is expected not to cause significant performance degradation. The shadow fading on the other hand, has a standard deviation of 10 dB, and can be expected to decrease performance over a wide range of SNRs. This is clearly seen as the case in figure 6.2b.

6.3.2 Discussion of Simulation Results Under Noise Uncertainty

It is already known that performance of the energy detector deteriorates under noise uncertainty. It is important to assess this property, as the underlying noise distribution is rarely known exactly for a detector. The simulation results under noise uncertainty are shown in figure 6.3. Both KL divergence based detectors are suffering less under noise uncertainty than the energy detector when the noise uncertainty is around 0.1 dB. To determine noise variance to such an accuracy requires about $1.2 \cdot 10^4$ samples (See table 6.1). This is not unreasonable, but the problem is that one never knows whether only noise is present. The noise distribution may also be time variant. Larger uncertainties might occur, which is a great weakness of the KL divergence based detectors. This weakness under uncertainty is essential when trying the understand the applicability of Kullback Leibler divergence in spectrum sensing. Under moderate to high noise uncertainties, the robustness of the KL based algorithms breaks down. This occurs under an uncertainty of about 0.5 dB for the CDF algorithm and about 1 dB for the HIST algorithm.

The breakdown behavior was described mathematically in section 5.5 and is a result of the different nature of the Kullback-Leibler divergence estimators compared to for instance the conventional energy detector. The energy detector suffers from noise uncertainty because determining the detection threshold relies on knowing the noise variance. For both the KL divergence based detectors, determining the threshold only relies on parameters that are always known a priori. Hence the threshold is always correct. The problem is that to estimate the divergence from the empirical distribution of the received samples, one must have a reference distribution. This creates the need for having complete knowledge of the underlying noise distribution. If the underlying distribution suffers from uncertainty, the actual estimate will be wrong. Hence one is estimating the wrong divergence, and thus making wrong decisions. Under large uncertainty, even if the received samples are from the true underlying distribution, the estimated Kullback-Leibler divergence, based on using the erroneous underlying theoretical distribution due to uncertainty, will be large at low signal to noise ratios. Hence the detector falsely starts detecting again when the signal to noise ratio drops. This is the breakdown that is observed in figure 6.3.

The reason why the breakdown behavior is more severe for the KL HIST algorithm is the same as why it performs better than the KL CDF algorithm, which is that its estimation variance is significantly lower $(O(1/N) \text{ vs. } O(1/N^2))$. The lower estimation variance gives a more accurate estimation and a lower detection threshold. Thus the detector is more sensitive to the small deviations in the estimated KL divergence due to uncertainty in the underlying theoretical conditional distribution under H_0 .

Noise uncertainty, as addressed in this report, is equivalent to noise variance uncertainty since the underlying noise is always assumed to be circularly symmetric white complex with a Gaussian distribution. Recall that a Gaussian distribution is fully described by its mean and variance, and that circular symmetry implies that the mean is zero and the real and imaginary components are statistically independent. Hence the variance fully describes circularly symmetric white Gaussian noise. However, the KL divergence estimators are obviously also affected by any other type of uncertainty, as one requires perfect knowledge of the underlying theoretical conditional distribution under H_0 to make a correct estimate of the Kullback-Leibler divergence. The use of variance uncertainty here was simply applied for simplicity.

This section discussed the impact of noise uncertainty on the spectrum sensing detectors. It was argued why the KL based algorithms perform better than the energy detector when the noise uncertainty is small, but that they suffer from a breakdown under large uncertainty when the SNR decreases below a certain threshold. This breakdown behavior was expected from the mathematical considerations in section 5.5. The last part of the section stressed the point that even though distribution uncertainty was only expressed through uncertainty in estimating variance in this thesis, any type of uncertainty in the conditional distribution under H_0 will reduce performance of the KL divergence based detectors.

Chapter 7 Detection Results for an Authentic Captured Signal

The previous chapter discussed detection performance of the proposed Kullback-Leibler divergence based spectrum sensing algorithms through simulations using synthetic signals. It is important to assess the new algorithms by using synthetic signals as this provides complete control over the test environment, preventing potential contamination of the signals by for instance interference or measurement equipment. However, only assessing algorithms through computer simulations is dangerous, particularly for two main reasons. The aforementioned interference or measurement distortion might be unavoidable in a real world application and should hence be included in the original analysis. The other reason being that it is important to test underlying assumptions. Any theoretical analysis starts by making assumptions to simplify the problem. For instance, in the spectrum sensing problem, the main assumption is that the underlying signal under H_0 consists of circularly symmetric complex white Gaussian noise. By acquiring real world signals, one can ensure that the original assumptions hold and that the algorithm behaves as expected from computer simulations. This chapter will present detection results for the four detectors addressed in this thesis on two frames of raw data from a Universal Mobile Telecommunications System (UMTS) receiver. One frame has a signal present and one contains noise, hence detection under H_0 and H_1 can be performed. The two frames have been extracted from a longer sequence of captured raw data.

7.1 Description of the Signal and its Distribution

A sequence of raw captured data of a downconverted UMTS signal has been acquired from a previous cognitive radio experiment performed at the EURECOM research institute. The signal has a bandwidth of 60 MHz and is modulated with OFDM. The center frequency before downconversion is 1950 MHz. The raw data is represented as unsigned 16 bit integers. For details on the experiment, including signals and hardware used to capture them, see [17]. It can be discussed whether experiments should have been carried out for this research specifically. However, the trials were aimed at cognitive radio, and the documentation of the signals and experiment is sufficient for the results to be illustrated in this thesis. The substantial amount of planning and time required to perform controlled wireless hardware experiments should also be taken into account.

When assessing the raw data, it can be observed that both the in-phase and the quadrature components have mean values. It was shown in the theory chapter how communication signals have zero mean, it is thus assumed that these mean values are artifacts from the actual capturing circuit. The mean values are consistent for both the noise and signal frames. In order to prepare the signal for further processing, these mean values are subtracted. After this subtraction, the signals are represented with the ordinary 64 bit floating point format which is the MatLab[®] default. Figure 7.1 shows the in-phase and quadrature components of the signal after the mean has been subtracted. There is a spike present around $n = 0.7 \cdot 10^4$ for the imaginary signal component. This is assumed to be a distortion of unknown origin, and is not considered further during the following analysis. The actual signal block can be seen around $n = 1 \cdot 10^4$. A signal block containing 2501 samples has been extracted. A corresponding noise block is acquired by retrieving the 2501 first samples. The signal to noise ratio between these two blocks is 2.9 dB. Note however that there is a degree of uncertainty to this estimate do to the limited number of samples. Spectral analysis of the signal and noise blocks reveals that



Figure 7.1: Figures show the time domain components of the captured signal.

there is powerful spectral content at a few distinct frequencies that appears to be residual carrier frequency artifacts from the down conversion. Welch power spectral density estimates of the signal, and estimates of the autocorrelation function can be seen in figure 7.2. Note that the spectral estimate is not symmetric, this is expected since the underlying signal is complex. From the spectral estimates, especially for the noise block (figure 7.2c), three sinusoidal components are clearly

present, seen as peaks around the normalized frequencies $\omega = 0.01\pi$ rad/sample, $\omega = 1.4\pi$ rad/sample and $\omega = 1.8\pi$ rad/sample. Recall from Fourier analysis that non-symmetric spikes in the frequency domain correspond to complex exponentials in the time domain (See appendix B.2), a fact that enforces the suspicion of the spikes as being artifacts from the modulation. It turns out that these unwanted frequency components create a problem for the basic assumption made in the previous theoretical analysis that the signal under H_0 is complex circularly symmetric white Gaussian noise. Hence an additional noise block is generated by bandpass filtering the original noise block. This bandpass filtering is done in the frequency domain by replacing the spikes by complex Gaussian noise with the same variance as adjacent samples. 6 samples were replaced for the component around $\omega = 0.01\pi$ rad/sample, and 4 samples were replaced for the two other components respectively. Hence a total of 14 samples out of 2501 samples were replaced with circularly symmetric complex white Gaussian noise, thus it is assumed that this manipulation, apart from removing the sinusoidal components, has negligible effect on the nature of the resulting signal. A fact that can also be seen by comparing the spectral estimates before and after filtering (figure 7.2c versus figure 7.2d). Another result of the Fourier domain bandpass filtering is that the inverse Fourier transform to retrieve a time domain signal interpolates the previously quantized signal. From the spectra it is apparent that the noise is not white. A distinct rolloff can be seen from $\omega = 0.5\pi$ rad/sample to $\omega = \pi$ rad/sample. The same roll off can be identified both for the signal block and noise block, and is assumed to be resulting from anti-aliasing or other bandpass filters in the receiver. It was already established that assuming the noise to be white does not hold here. However, to investigate the assumption of the underlying signal under H_0 as being circularly symmetric Gaussian, the distribution of the signal magnitude has been compared to a theoretical Rayleigh distribution. The histogram of the magnitude of the samples in the noise block before filtering can be seen in figure 7.3a. The effect of signal quantization is apparent. Remember that it is the real and imaginary signal components that are quantized, hence the quantization does not appear as symmetric intervals for the magnitude. It is seen that the shape of the histogram resembles a Rayleigh distribution in shape, but due to quantization it is impossible to justify a conclusion (In addition, the distribution is distorted from Rayleigh due to the residual frequency components). However, recall that besides from removing unwanted spectral components, a side effect of the bandpass filtering applied was to interpolate the time domain samples. Figure 7.3b shows an estimate of the probability density function of the bandpass filtered noise block compared to a theoretical Rayleigh distribution. The Rayleigh distribution used a maximum likelihood estimate parameter σ^2 estimated from the samples of the bandpass filtered noise block. Correspondingly, figure 7.3c shows a Quantile-Quantile plot between the quantiles of the envelope of the bandpass filtered noise block samples and an identical number of samples drawn from a theoretical Rayleigh distribution given by the maximum likelihood parameter estimated from the bandpass filtered noise block. It is clearly seen from the two figures that the magnitude samples are well described with a Rayleigh distribution. This strengthens the assumption that the



Figure 7.2: Figures show the Welch method power spectral density estimates of the signal block, noise block and the filtered noise block. In addition, one figure shows the real value of the estimated autocorrelation function of the three blocks respectively (Recall from the theory and background chapters that the imaginary component is zero).

signal under H_0 consists of circularly symmetric complex Gaussian noise. Although the contribution from the sinusoidal content believed to be residual artifacts from the downconversion needs to be disregarded. This section described a captured



(c) Quantile-Quantile plot, Filtered Noise Block

Figure 7.3: Figures show histogram and probability density function estimates and Quantile-Quantile plot for the envelope of the captured noise block samples. The theoretical Rayleigh distribution seen in the right figure has parameter $\sigma^2 = \hat{\sigma}_n^2/2$, where $\hat{\sigma}_n^2$ is the estimated variance of the noise block.

sequence of raw UMTS data. It further presented an investigation of the signal to check whether the assumption of the signal under H_0 as being circularly symmetric complex white Gaussian noise holds. It was concluded that this assumption in its originality does not hold due to artifacts most likely to stem from the receiver. However, when removing sinusoidal components, believed to be residual carrier frequency artifacts, the assumption of the signal being circularly symmetric complex Gaussian noise appears to be justified (Note that *white* has been omitted). The following section will present detection results.

7.2 Detection Results

The previous section described the captured raw data UMTS signal, and the preprocessing performed before applying the spectrum sensing. This section will present the detection results for the two proposed Kullback-Leibler divergence based algorithms and the two reference detectors. All detectors have, whenever applicable, been applied to all three signal blocks. The first block contains the UMTS signal frame plus noise and receiver (assumed) artifacts, the second contains noise plus receiver (assumed) artifacts and the third is the noise frame after bandpass filtering (and implicit interpolation).

Results from the detection are summarized in table 7.2 while the corresponding computed test statistics and detection thresholds can found in table 7.1. The

	Signal Block	Noise Bl.	η	Filtered Noise B.	η Filt.
ED	$6.43 \cdot 10^{4}$	$3.32 \cdot 10^{4}$	$3.43 \cdot 10^4$	$3.20 \cdot 10^4$	$3.31 \cdot 10^{4}$
KL HIST	0.2744	0.0288	0.0055	0.0031	0.0055
KL CDF	NA	NA	0.0264	0.0237	0.0264
IM(1)	0.1870	0.1072	0.0116	0.0946	0.0116

Table 7.1: Table gives test statistics for the three detection algorithms applied to the signal block, noise block and filtered noise block respectively. The appropriate detection thresholds η are also given.

	Signal Block	Noise Block	Filtered Noise Block
ED	Correct Decision	Correct Decision	Correct Decision
KL HIST	Correct Decision	Wrong Decision	Correct Decision
KL CDF	NA	NA	Correct Decision
IM(1)	Correct Decision	Wrong Decision	Wrong Decision

Table 7.2: Table states whether the binary detection decision is correct or wrong based on the signal frame and the computed test statistic (Corresponding test statistics and detection thresholds are given in table 7.1).

tables show that the energy detector chooses correctly for all three blocks, the KL histogram algorithm chooses correctly for the signal block and the filtered noise block while the IM autocorrelation detector only chooses correctly for the signal block. The KL CDF algorithm is in a special position as it due to its structure requires a signal with a continuous amplitude spectrum. It is thus only applicable for the filtered signal block, where it makes a correct detection decision.

7.3 Discussion of Detection Results for the Captured UMTS Signal

Detection results for the captured raw UMTS signal were presented in the previous chapter. This section will address and discuss these findings.

As expected, the energy detector decides correctly for all signal blocks. The energy detector does not make any assumptions on the underlying signal distributions or correlation properties, and hence is bound to be correct in this case where its detection threshold is directly calibrated from the noise blocks. The other reference detector on the other hand, the IM autocorrelation based detector, fails automatically due to the inherent correlation of the signals. In figure 7.2b and figure 7.2c it is clearly seen that the signal is correlated, and not white. Hence the IM detector is not directly applicable. However, this shows that higher order statistics, a topic previously disregarded in favor of KL divergence to limit the scope of this thesis, could be more applicable to real world system. Real world systems depend on realizable filters, which are bound to introduce correlation. Recall that cumulants of Gaussian signals of order higher than two are always zero, regardless of whether the second order cumulant is zero not (I.e. if the Gaussian signal is correlated or white).

The Kullback-Leibler divergence based algorithms also perform as expected. KL HIST decides correctly for the signal block and the filtered noise block, while it makes the wrong decision for the unfiltered noise block. This is expected as its underlying assumption of Gaussian noise does not hold for the unfiltered case due to the residual sinusoidal content. However, this underlines an important point for the KL algorithms. These algorithms can be tailored to have any distribution as the underlying H_0 default distribution. Hence if it was known that a certain receiver had residual modulation artifacts, the detector could have been calibrated to this. It should also be noted how efficient the KL HIST algorithm is, as just a little spectral contribution believed to be an artifact from the downconversion triggers a detection.

A more grave problem is however uncovered for the KL CDF algorithm, and it clearly shows the dangers of disregarding the actual practical applications when designing engineering solutions. A practical communication system has an Analog to Digital Converter (ADC) with a limited resolution at the receiver, which will quantize the received signal. Hence the CDF algorithm becomes directly inapplicable since it requires a signal with a continuous amplitude distribution (The 64 bit floating point default format of MatLab[®] is here for all practical purposes considered as being continuous). Of course an interpolation can be performed digitally if the processor in the receiver supports it, but this is not a viable solution due to the increased complexity.

The above results clearly show that some of the fundamental assumptions made in the initial theoretical analysis do not hold for this particular receiver. The receiver noise is not white, a fact that probably stems from anti aliasing or other low pass filters in the receivers having a slow roll-off. It is also experienced that the received signal appears to have artifacts from the hardware in terms of DC content and carrier residue. Hence these mentioned receiver characteristics cause the received signal under H_0 to deviate from the assumed pure circularly symmetric white complex Gaussian noise. This causes problems for all the discussed detectors, except the energy detector, since these detectors rely heavily on the aforementioned assumption. However, it was also stressed that the KL divergence based algorithms can be tailored to any underlying distribution if it is known a priori. It was also seen that the KL HIST estimator is efficient as just a little spurious additional spectral content in the noise block was enough to trigger a detection.

Chapter 8 Summarizing Discussion

The previous two chapters presented detection results for synthetic signals and captured frames of an authentic UMTS signal. Both chapters ended with thorough individual discussions of the respective results. This chapter will present a summarizing discussion based on the two previous chapters aimed at illustrating whether the two proposed Kullback-Leibler divergence based detectors solved the problem at hand. Recall that the purpose of the research was to investigate spectrum sensing at low SNRs with the intention of proposing new solutions to overcome existing problems.

8.1 Potential of Kullback-Leibler estimation in Spectrum Sensing

When initiating this cognitive radio spectrum sensing research, it was with the purpose of devising a detection strategy that could improve performance over existing schemes in the low signal to noise range. An analysis was done, indicating that a major problem for conventional sample average based estimates, used in both the energy detector and the IM autocorrelation detector, was an estimation variance depending on signal to noise ratio. This dependence made detection at low signal to noise ratios infeasible. To maintain a fixed estimation variance, the estimation block length had to be increased at a higher rate than the decrease in SNR (Quadratic versus linear). It was thus concluded that if possible, a new detection scheme should be based on estimators that were independent of signal and noise variance. Such estimators were devised for estimating the Kullback-Leibler divergence between an empirical distribution and a theoretical distribution. These estimators have very good behavior in terms of estimation variance, and do not depend on signal- or noise variance. Hence these estimators function equally well for arbitrary signal to noise ratios. This has been proven for the CDF based estimator, which has an estimation variance that only depends on the number of samples. The theoretical analysis for the histogram based estimator is unfortunately not complete. However, it has been motivated through simulations that the estimation variance of the histogram based algorithm depends on the number of samples, number of bins used in the histogram and the choice of pre-loading constant.

A problem that has become the fundamental issue as to whether Kullback-Leibler divergence can be used efficiently in spectrum sensing is the nature of the conditional distributions under the detection hypotheses H_0 and H_1 . All results obtained in this thesis point towards the fact that communication signals have distributions that are very similar to the underlying Gaussian noise. A worst case example is an OFDM signal, which is well approximated as having a Gaussian distribution as long as the number of subcarriers is large [6, 60]. Since both signal and noise have zero mean, the only difference between the two conditional distributions is variance. It is obvious that simply doing a maximum likelihood estimate of the variance (I.e. applying the energy detector) should be sufficient.

Consider the following example to get an impression of the challenges with applying Kullback-Leibler divergence for spectrum sensing in low signal to noise ratios. For an OFDM signal in AWGN, the Kullback-Leibler divergence between the signal envelope and the theoretical noise envelope is derived in (5.19). Evaluating (5.19) at an SNR of -22 dB renders a divergence of $\approx 1.98 \cdot 10^{-5}$. This is a small value for these estimators, and the only reason good performance is obtained at reasonably low signal to noise ratios (Down to about -5 dB for the CDF algorithm and -10 dB for the HIST algorithm) is because of the estimation variance of these estimators is low (O(1/N) for the KL CDF estimator and $O(1/N^2)$ for the KL HIST estimator). It is thus obvious that Kullback-Leibler divergence in spectrum sensing is significantly challenged by the similarity of the conditional distributions under H_0 and H_1 .

In addition to the similar nature of the conditional distributions under H_0 and H_1 , another problem with the KL based detectors is that they are sensitive to knowledge about the underlying theoretical distribution. Section 6.2.4 demonstrated this sensitivity through applying noise uncertainty. It was shown mathematically how a breakdown occurred at a sufficiently low SNR when uncertainty was introduced. This is unfortunate, and severely limits the applicability of the detectors for spectrum sensing. In a real world application, it is very hard to have full knowledge of the conditional distribution under H_0 . The effect of uncertainty in the theoretical distribution was further supported when analyzing the captured UMTS signal. The captured signal showed several artifacts inconsistent with the assumption of the distribution under H_0 as being circularly symmetric complex Gaussian. Hence the Kullback-Leibler divergence algorithms do not function as intended as they estimate the wrong divergence. Even if there is no signal present, the detector will estimate a divergence since it has erroneous information on the underlying theoretical distribution under H_0 .

However, it is very attractive that the proposed Kullback-Leibler divergence estimators have estimation variances that do not depend on signal- or noise variance. They are not disadvantaged at low signal to noise ratios, and in addition the estimation variance is low $(O(1/N^2)$ for KL HIST). These properties should yield the estimators applicable for other detection applications than cognitive radio, where the underlying conditional distributions are more distinct. The sensitivity of the detectors was further illustrated in chapter 7, when detection was triggered for the KL HIST algorithm due to spurious additional spectral content in the noise block. To increase potential performance of the Kullback-Leibler divergence based detectors, one needs to apply the estimators to conditional distributions with a larger inherent difference.

Chapter 9 Conclusion

This thesis set forth exploring undiscovered ground in spectrum sensing for cognitive radio. Specifically, the aim of the research has been to investigate whether higher order statistics or information theoretic distance measures could be used to improve spectrum detection performance in the low signal to noise region. Through a thorough research effort, two novel spectrum sensing algorithms based on Kullback-Leibler divergence estimation were proposed and analyzed.

Based on the previous chapters, a number of conclusions can be made. Based on the theoretical analysis, it was shown how detectors relying on conventional sample average based estimation suffer when the signal to noise ratio decreases. In order to maintain a fixed estimation variance, a linear decrease in signal to noise ratio requires a quadratic increase in the number of samples used for the estimation. Hence, accurate estimation becomes infeasible at low signal to noise ratios. Note that the proof is only done for signals with a circularly symmetric Gaussian distribution.

It can be concluded that Kullback-Leibler divergence based detection is hard due to the similarity of signal and noise distributions. However, the KL HIST algorithm is to some extent promising. It has a performance which is lower than the energy detector, and a complexity which is higher, but it suffers less under moderate noise uncertainty. This assumes that it is possible to keep the noise uncertainty at a moderate level. The work on the KL HIST algorithm is unfortunately not complete. A full theoretical analysis is needed in order to make an absolute conclusion on the applicability for spectrum sensing.

The KL CDF algorithm on the other hand can be concluded not to perform satisfactory for the spectrum sensing problem. A main reason is that it requires the received signal to have a continuous amplitude distribution, a feature which is difficult in an actual application due to quantization in the AD converter. Further, the KL CDF algorithm yields the lowest detection performance and the highest computational complexity of all the discussed detectors.

Experiments with real data revealed that the the very common assumption in academia when doing spectrum sensing research, that the signal under H_0 can be modeled as circularly symmetric white Gaussian noise, can be dangerous. The

actual raw data analyzed here showed to some extent considerable deviation from this assumption. This implies that care should be taken when making assumptions on the noise distribution.

Although, it was shown in this thesis that Kullback-Leibler divergence did not appear as a very promising spectrum sensing approach, due to the inherent similarity of the distributions of noise and communication signals, it is important to stress the features of the developed estimators. The two estimators that were proposed in this thesis will have very good detection properties for detection problems were the conditional distributions of the hypotheses are less similar.

Chapter 10

Future Work and potential

This thesis has explored a number of novel aspects of spectrum sensing for cognitive radio. A number of interesting results have been revealed. However, several areas are left only partially explored. This leaves numerous opportunities for further research. In the following section, some of these opportunities will be briefly discussed.

10.1 Complete the Theoretical Analysis of the KL HIST Algorithm

The KL HIST algorithm introduced in this thesis showed partially promising behavior. However, the theoretical analysis of the algorithm was unfortunately not completed. Very interesting results were previously discussed regarding the probable Gamma distribution under H_0 . It is known that a Gamma distribution occurs when summing k independent exponentially distributed random variables, if k is an integer. In this case, k was evaluated to approximately 8.51 when the preloading constant is 0.01 and the number of bins in the histogram is 18. 18 random variables are added together to generate the resulting, possibly Gamma distributed, random variable representing the estimate of the Kullback-Leibler divergence. Hence, the distribution is not generated by summing k independent exponentially distributed random variables (since $k \simeq 8.51$ in this case). However, one idea that could have been explored further is an idea presented in [41]. It is argued how a sum of ncorrelated random variables can be represented as the sum of m independent random variables, where m < n. This idea and others should be explored further and explained. With a solid explanation for the Gamma assumption, it is believed that the rest of the theoretical analysis of the histogram based Kullback-Leibler divergence estimator will follow.

10.2 Other Distributions and Multidimensional Estimators

Only one-dimensional estimators were considered in this thesis. It was argued how preliminary trials with multidimensional estimators involved extensive search operations, causing significant computational complexity. It was further argued that if the assumption of circular symmetry holds for the communication signals, the real and imaginary components are independent, and hence multidimensional estimators should yield little or no improvement compared to one dimensional estimators. However, this choice followed from simple reasoning, and no controlled experiments were performed. It would thus be interesting to extend the research presented in this thesis to multidimensional Kullback-Leibler estimators.

10.3 Other Areas to Apply the Kullback-Leibler Divergence Estimators

Novel estimators for estimating Kullback-Leibler divergence between one empirical distribution and one theoretical distribution were derived during this research. These estimators have very good specifications, comprising especially the low estimation variance which only depends on parameters such as number of samples, which is known a priori. These estimators unfortunately turned out to not be as efficient for the cognitive radio problem since there was not sufficient discrepancy between the conditional probability distributions of the received signal under the two hypotheses H_0 and H_1 . It becomes apparent how it would be interesting to search for other areas than cognitive radio comprising detection problems with larger inherent discrepancy between the conditional distributions.

10.4 Distributed Interference Metric

Previous chapters have discussed how the proposed Kullback-Leibler divergence based detectors compute the KL divergence between the distribution of the received signal's envelope and a reference distribution, typically chosen to be a Rayleigh distribution. Especially in the chapter addressing the captured UMTS signal, it becomes apparent that any additional interference present in the receiver will be expressed through increased divergence. This can be exploited in itself. If one assumes the distribution of the signal at the receiver can be calibrated when there is no input at the receiver antenna. Then the KL divergence will be a metric of the amount of interference at the receiver. This is itself an interesting property that can be exploited in a distributed network. Instead of receivers making binary detection decisions, the computed KL divergences can be transmitted to a fusion center. If the fusion center knows the position of the receiver nodes, it will have a map illustrating interference present at receivers. From this information, it might be able to infer something about the position and nature of the interference. The information can also be used for resource allocation.

10.5 More Trials on Real Signals

Chapter 7 became pivotal in making the final conclusions in this thesis. When applying the spectrum sensing algorithms on an authentic captured signal, a number of weaknesses were revealed. The most severe weakness turned out to be assuming that the signal under H_0 is circularly symmetric complex white Gaussian noise. However, the captured data only contained one sequence from one specific receiver. It is obvious that it would be highly interesting to perform more thorough trials on larger sets of data, comprising different signals captured with different hardware.

Real time operation should also be explored.

10.6 Investigating Fundamental Limits

The results obtained during this research further strengthen the assumption that there might be fundamental physical limits to spectrum sensing in low signal to noise ratios. This assumption is further strengthened by results in papers [22, 61]. These papers address a technique referred to as stochastic resonance. Stochastic resonance has been known for decades in physics and has, especially over the past decade, received significant attention in various fields of engineering [16]. The reason for this attention is that stochastic resonance presents a theory describing how, counter intuitively, adding a certain amount of noise to a system can actually improve the signal to noise ratio. Such a technique might prove itself useful in order to overcome what seems to be physical limitations and boost performance in low signal to noise ratios. Exploring stochastic resonance specifically, or fundamental limits in spectrum sensing in general, would be highly interesting topics for future research.

Bibliography

- G. Acosta. Ofdm simulation using matlab. http://www.ece.gatech.edu/ research/labs/sarl/tutorials/OFDM/Tutorial_web.pdf, 2000.
- [2] M. Afgani, S. Sinanovic, and H. Haas. Anomaly detection using the kullbackleibler divergence metric. pages 1 –5, Oct. 2008.
- [3] M. Afgani, S. Sinanovic, and H. Haas. Information theoretic approach to signal feature detection for cognitive radio. pages 1 –5, 30 2008-dec. 4 2008.
- [4] D. Bertsekas and J. Tsitsiklis. Introduction to Probability, pages 213,269,273– 274. Athena Scientific, 2002.
- [5] D. Cabric, A. Tkachenko, and R. Brodersen. Experimental study of spectrum sensing based on energy detection and network cooperation. *The 2nd Annual International Wireless Internet Conference (WICON)*, Aug. 2006.
- [6] S. Chaudhari, V. Koivunen, and H.V. Poor. Autocorrelation-based decentralized sequential detection of ofdm signals in cognitive radios. *Signal Processing*, *IEEE Transactions on*, 57(7):2690–2700, july 2009.
- [7] Xiaoqian Chen, Hongyuan Wang, and Qiao Cai. Performance analysis and optimization of novel high-order statistic features in modulation classification. In Wireless Communications, Networking and Mobile Computing, 2008. WiCOM '08. 4th International Conference on, pages 1-4, 12-14 2008.
- [8] J. Christiansen. Recursive estimation of first and second order statistics for spectral detection with applications in cognitive radio. Unpublished Work, 2009.
- [9] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. Introduction to Algorithms, pages 41,145. MIT Press, 2001.
- [10] T. Cover and J. Thomas. Elements of Information Theory, pages 14–43, 251. John Wiley and Sons, 2006.
- [11] M. Cummings and T. Cooklev. Tutorial: Software-defined radio technology. In Computer Design, 2007. ICCD 2007. 25th International Conference on, pages 103 –104, 7-10 2007.

- [12] A.V. Dandawate and G.B. Giannakis. Statistical tests for presence of cyclostationarity. Signal Processing, IEEE Transactions on, 42(9):2355-2369, sep 1994.
- [13] FCC. Cognitive radio technologies proceeding. http://www.fcc.gov/oet/ cognitiveradio/, 2003.
- [14] FCC. Spectrum access and the promise of cognitive radio technology. ftp://ftp.fcc.gov/pub/Bureaus/Engineering_Technology/Documents/ cognitive_radio/vanwazercognitive_radio_presentation.ppt, 2003.
- [15] Mei Feng and K.-D. Kammeyer. Suppression of gaussian noise using cumulants: a quantitative analysis. volume 5, pages 3813 –3816 vol.5, apr 1997.
- [16] L. Gammaitoni, P. HÅďnggi, P. Jung, and F. Marchesoni. Stochastic resonance. *Review of Modern Physics*, 70:223âÅŞ87, 1998.
- [17] M. Ghozzi, B. Zayen, and A. Hayar. Experimental study of spectrum sensing based on distribution analysis. June 2009.
- [18] A. Goldsmith. Wireless Communications, pages 64–94,322,374–396. Cambridge University Press, 2005.
- [19] M. Hayes. Statistical Signal Processing, pages 83,104–105,111–115. John Wiley and Sons, 1996.
- [20] S. Haykin. Communication Systems, pages 58–71. John Wiley and Sons, 2001.
- [21] S. Haykin, D.J. Thomson, and J.H. Reed. Spectrum sensing for cognitive radio. *Proceedings of the IEEE*, 97(5):849-877, may 2009.
- [22] Di He, Yingpei Lin, Chen He, and Lingge Jiang. A novel spectrum-sensing technique in cognitive radio based on stochastic resonance. Vehicular Technology, IEEE Transactions on, 59(4):1680-1688, may 2010.
- [23] T. Ikuma and M. Naraghi-Pour. Autocorrelation-based spectrum sensing algorithms for cognitive radios. pages 1 –6, aug. 2008.
- [24] H. Karl and A. Willig. Protocols and Architechtures for Wireless Sensor Networks, page 113. Wiley, 2007.
- [25] V. Koivunen, S. Chaudhari, J. Ryynanen, M. Kosunen, E. Larsson, D. Danev, A. Hayar, and B. Zayen. Spectrum sensing algorithm evaluation. Technical report, EURECOM and NTNU and LiU, 2010.
- [26] R. Krichevsky and V. Trofimov. The performance of universal encoding. Information Theory, IEEE Transactions on, 27(2):199 – 207, mar 1981.
- [27] C.J. Le Martret and D.M. Boiteau. Modulation classification by means of different orders statistical moments. In *MILCOM 97 Proceedings*, volume 3, pages 1387–1391 vol.3, 2-5 1997.

- [28] P. Li, F. Wang, and Z. Wang. Algorithm for modulation recognition based on high-order cumulants and subspace decomposition. In *Signal Processing*, 2006 8th International Conference on, volume 3, 16-20 2006.
- [29] Yingpei Lin, Chen He, Lingge Jiang, and Di He. A spectrum sensing method in cognitive radio based on the third order cyclic cumulant. pages 1 –5, nov. 2009.
- [30] Jun Ma, G.Y. Li, and Biing Hwang Juang. Signal processing in cognitive radio. Proceedings of the IEEE, 97(5):805-823, may 2009.
- [31] J.M. Mendel. Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications. *Proceedings of* the IEEE, 79(3):278 –305, mar 1991.
- [32] III Mitola, J. and Jr. Maguire, G.Q. Cognitive radio: making software radios more personal. *Personal Communications*, *IEEE*, 6(4):13–18, aug 1999.
- [33] M. Naraghi-Pour and T. Ikuma. Autocorrelation-based spectrum sensing for cognitive radios. Vehicular Technology, IEEE Transactions on, 59(2):718-733, feb. 2010.
- [34] C.L. Nikias and J.M. Mendel. Signal processing with higher-order spectra. Signal Processing Magazine, IEEE, 10(3):10 –37, jul 1993.
- [35] Official Web Page. European telecommunications standards institute. http: //www.etsi.org.
- [36] F. Perez-Cruz. Kullback-leibler divergence estimation of continuous distributions. pages 1666 –1670, july 2008.
- [37] J. Proakis and Salehi M. Digital Communications, pages 177–178,262,597,831– 835. McGraw Hill, 2007.
- [38] J. Proakis and D. Manolakis. *Digital Signal Processing*, page 836. Prentice Hall, 2007.
- [39] D. Ramirez, J. Via, I. Santamaria, and P. Crespo. Entropy and kullbackleibler divergence estimation based on szego's theorem. 17th European Signal Processing Conference (EUSIPCO 2009), Aug 2009.
- [40] I. Reed. On a moment theorem for complex gaussian processes. Information Theory, IRE Transactions on, 8(3):194-195, april 1962.
- [41] J.E. Salt and H.H. Nguyen. Performance prediction for energy detection of unknown signals. Vehicular Technology, IEEE Transactions on, 57(6):3900 -3904, nov. 2008.
- [42] S. Seidel. Raytheon, ieee 802 tutorial. http://www.ieee802.org/802_ tutorials/05-July/IEEE%20802%20CR%20Tutorial%207-18-05%20seidel% 20input.pdf, 2005.

- [43] Y. Shao and M. Hahn. Limit theorems for the logarithm of sample spacings. Statistics and Probability Letters, 24:121–132, 1995.
- [44] D. Stefanica. A Primer for the Mathematics of Financial Engineering, page 19. FE Press, 2008.
- [45] Yongliang Sun, Yutao Liu, and Xuezhi Tan. Spectrum sensing for cognitive radio based on higher-order statistics. pages 1–4, oct. 2008.
- [46] A. Swami and B.M. Sadler. Hierarchical digital modulation classification using cumulants. *Communications, IEEE Transactions on*, 48(3):416–429, mar 2000.
- [47] R. Tandra and A. Sahai. Fundamental limits on detection in low snr under noise uncertainty. volume 1, pages 464 – 469 vol.1, June 2005.
- [48] R. Tandra and A. Sahai. Noise calibration, delay coherence and snr walls for signal detection. pages 1 –11, Oct. 2008.
- [49] R. Tandra and A. Sahai. Overcoming snr walls through macroscale features. pages 583 –590, Sept. 2008.
- [50] R. Tandra and A. Sahai. Snr walls for signal detection. Selected Topics in Signal Processing, IEEE Journal of, 2(1):4–17, Feb. 2008.
- [51] H. Tang. Some physical layer issues of wide-band cognitive radio systems. pages 151 –159, Nov. 2005.
- [52] J.K. Tugnait. Detection of random signals by integrated polyspectral analysis. Signal Processing, IEEE Transactions on, 44(8):2102 –2108, aug 1996.
- [53] H. Urkowitz. Energy detection of unknown deterministic signals. Proceedings of the IEEE, 55(4):523 – 531, april 1967.
- [54] H. Van Trees. Detection, Estimation and Modulation Theory, part 1. John Wiley and Sons, 2001.
- [55] M. Vetterli and J. Kovacevic. Wavelets and Subband coding, page 54. Prentice Hall, 1995.
- [56] Haiquan Wang, En-Hui Yang, Zhijin Zhao, and Wei Zhang. Spectrum sensing in cognitive radio using goodness of fit testing. *Trans. Wireless. Comm.*, 8(11):5427–5430, 2009.
- [57] T. Yucek and H. Arslan. A survey of spectrum sensing algorithms for cognitive radio applications. *Communications Surveys Tutorials*, *IEEE*, 11(1):116–130, first 2009.
- [58] B. Zayen, A. Hayar, H. Debbabi, and H. Besbes. Application of smoothed estimators in spectrum sensing technique based on model selection. In Ultra Modern Telecommunications Workshops, 2009. ICUMT '09. International Conference on, pages 1-4, 12-14 2009.

- [59] B. Zayen, A. Hayar, and K. Kansanen. Blind spectrum sensing for cognitive radio based on signal space dimension estimation. In *Communications*, 2009. *ICC '09. IEEE International Conference on*, pages 1–5, 14-18 2009.
- [60] Wei Zeng and Guangguo Bi. Robust detection of ofdm signals for cognitive uwb in low snr with noise uncertainty. pages 1 –5, sept. 2008.
- [61] Kun Zheng, Husheng Li, S.M. Djouadi, and Jun Wang. Spectrum sensing in low snr regime via stochastic resonance. In *Information Sciences and Systems* (CISS), 2010 44th Annual Conference on, pages 1 –5, 17-19 2010.

Appendix A MatLab[®] Code

A.1 KL CDF Algorithm

function lambda = SD_KL_cdf(x_p, SDparam)

%Function estimates the Kullback Leibler divergence between the empirical %distribution of the samples in x_p and the theoretical rayleigh %distribution defined by the parameter nVar. (Where nVar is the variance of %the complex Gaussian noise that has a rayleigh magnitude. I.e. The %rayleigh sigma^2 parameter is nVar/2)

```
nVar = SDparam{4}; %Make sure there are no index conflicts in the current
%sendora implementation
x_p = sort(abs(x_p), 'ascend'); %Sorting samples, to generate empirical cdf
nInit = length(x_p);
*Quick fix to remedy the problem of epsilon=0 which occurs when two
%identical entries exist in x_p
delList = [];
for i=2:length(x_p)
    if x_p(i) == x_p(i-1)
        delList(end+1) = i;
    end
end
x_p(delList) = [];
n = length(x_p);
if length(delList)>round(0.1*nInit)
    msq1 = 'More than 10% of the samples of the original input signal';
    msg2 = ' have been discarded, algorithm might not be applicable to';
    msg3 = ' the current problem. Quantization of the input signal is';
    msg4 = ' a likely cause.';
    error(strcat(msg1,msg2,msg3,msg4));
end
```

delQxi_ray = exp(-(x_p(1:end-2).^2)/(nVar))-exp(-(x_p(2:end-1).^2)/(nVar)); lambda = -1/n*sum(log(n*delQxi_ray))-0.577215664901533;

A.2 KL HIST Algorithm

```
function KLest = SD_KL_hist(s1, SDparam)
```

```
%Function estimates the Kullback Leibler divergence between the empirical
%distribution of the samples in s1 and the theoretical Rayleigh
%distribution defined by the parameter nVar. (Where SDparam{4} is the
%variance of the circularly symmetric complex Gaussian noise that has a
%Rayleigh magnitude. I.e. The Rayleigh parameter is SDparam{4}/2)
nVar = SDparam{4}/2; %nVar is variance parameter of theoretical Rayleigh
%distribution
Nbins = SDparam{6}; %Number of bins to use in histogram.
s1 = abs(s1); %Compute envelope of signal
Lmin = min(s1);
Lmax = max(s1);
x = linspace(Lmin, Lmax, Nbins); %Centers of histogram bins that will be used
%Get histogram:
hS1 = hist(s1, x);
%Get the integral of the theoretical distribution over each bin given by x:
pS2 = getTheoreticalRayleighHist(x, nVar, Nbins);
plEst = (hS1+0.01)/sum(hS1+0.01); %Compute estimate of pmf, using 0.01 as
%preloading constant
KLest = sum(plEst.*log(plEst./pS2)); %Compute KL divergence
function pEst = getTheoreticalRayleighHist(x,Var,Nbins)
%x gives center of histogram bins. Function returns theoretical weights for
%a quantized version of a rayleigh dist at the points indicated in x. I.e.
%the integral of the Rayleigh probability density function over the
%histogram bins.
Rayleigh cdf: 1-exp(-x^2/(2*Var))
pEst = zeros(1, length(x));
\Delta = x(2) - x(1);
for m = 1:Nbins
    if m==1
        pEst(m) = 1 - \exp(-(x(m) + \Delta/2)^2 / (2 + Var));
    elseif m==Nbins
        pEst(m) = exp(-(x(m)-\Delta/2)^2/(2*Var));
    else
        pEst(m) = exp(-(x(m)-\Delta/2)^2/(2*Var)) - ...
```

exp($-\,(x\,(m)\,+\!\Delta/2)\,^2$ /(2*Var)); end

end

89
Appendix B

Proofs

B.1 CDF Based Kullback-Leibler Divergence Estimator

Starting from the definition of the KL divergence for continuous distributions given in (2.17), repeated here for convenience

$$D(f||g) = \mathbf{E}\left[\ln \frac{f(y)}{g(y)}\right].$$

Assume that the number of input samples N is large, from the law of large numbers (2.1) and (2.2)

$$D(f||g) = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \ln \frac{f(y_i)}{g(y_i)}$$
(B.1)

Without loss of generality, also assume the the received samples y_i are sorted in ascending order.

Recall from probability theory that the cumulative density function (CDF) of a random variable is defined as

$$F(y) = \int_{-\infty}^{y} f(t) \mathrm{d}t,$$

where f denotes the probability density function and F denotes the cumulative density function. Recall from the fundamental theorem of calculus that if F: $\mathbb{R} \to \mathbb{R}$ is defined on a closed interval on \mathbb{R} and f is continuous on this interval, F'(y) = f(y) [44]. Further recall from calculus that if a function $F : \mathbb{R} \to \mathbb{R}$ is differentiable at y and the limit exists [44]

$$F^{'}(y) = \lim_{\Delta y \to 0} \frac{F(y + \Delta y) - F(y)}{\Delta y}$$

By recalling that the input samples y_i are sorted in ascending order, applying the above results to (B.1) yields

$$D(f||g) = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \ln \frac{f(y_i)}{g(y_i)}$$

$$= \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \ln \frac{F'(y_i)}{G'(y_i)}$$

$$= \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \ln \frac{(F(y_i) - F(y_{i-1})) / \Delta y_i}{(G(y_i) - G(y_{i-1})) / \Delta y_i}$$

$$= \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \ln \frac{\Delta F(y_i)}{\Delta G(y_i)}$$

$$= \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \left(\ln \frac{\Delta F(y_i)}{\Delta G(y_i)} + \ln \frac{N}{N} \right)$$

$$= \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \left(\ln (N \Delta F(y_i)) - \ln (N \Delta G(y_i)) \right)$$
(B.2)

The samples y_i come from the distribution f by definition. Thus $F(y_i)$ will have a uniform distribution [43, 36] (This is intuitive as the CDF is a mapping CDF : $S \rightarrow [0, 1]$, where S is the support set of f.) Under this condition it can be shown that

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \ln \left(N \Delta F(y_i) \right) = -\gamma \tag{B.3}$$

where γ is the Euler-Mascheroni constant. This stems from the fact that $\Delta F(y_i)$ represents the difference between Poisson distributed stopping times, so $N\Delta F(y_i)$ has a unit exponential distribution [36]. Hence the mean value of $\ln N\Delta F(y_i)$ is

$$\mathbf{E} \left[\ln N \Delta F(y_i) \right] = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \ln N \Delta F(y_i)$$
$$= \int_0^\infty \exp\left(-y\right) \ln y \mathrm{d}y$$
$$\equiv -\gamma$$
(B.4)

Thus by applying (B.3), (B.2) reduces to

$$D(f||g) = -\gamma - \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \ln\left(N\Delta G(y_i)\right)$$
(B.5)

If f = g, $G(y_i)$ will also be uniform and $\lim_{N\to\infty} D(f||g) = -\gamma + \gamma = 0$. D(f||g) will also be asymptotically normal and have a variance

$$\sigma_D^2 = \frac{1}{N} \left(\frac{\pi^2}{6} - 1 \right) \tag{B.6}$$

The results on variance and asymptotic normality are given by applying theorem 5.1 in [43]. This yields, for large N the approximate estimator given in (5.9).

B.2 Fourier Transform of a Complex Exponential

$$\mathcal{F} \left[\exp\left(j2\pi f_{c}t\right) \right] = \mathcal{F} \left[\cos\left(2\pi f_{c}t\right) + j\sin\left(2\pi f_{c}t\right) \right] \\ = \frac{1}{2} \left(\delta[f + f_{c}] + \delta[f - f_{c}] + j^{2} (\delta[f + f_{c}] - \delta[f - f_{c}]) \right)$$
(B.7)
$$= \delta[f - f_{c}]$$

Appendix C

Derivation of Error Variance for Autocorrelation Estimation of AR(1) process

This section presents a derivation of the error variance in estimating the autocorrelation of a unit variance Autoregressive (AR) process of order one, in AWGN. The first part derives the error variance for the no noise case, and the second part extends the derivation to also include AWGN. Recall from [19], that the autocorrelation of a unit variance AR(1) process y[n] with parameter ρ is $r_{yy}[k] = \rho^k$.

This subsection presents the same analysis as above only for a rectangular window estimator

For a unit variance AR(1) process with parameter ρ , the estimation error in estimating the autocorrelation function is

$$\mathbf{E}\left[\epsilon_{r}^{2}[k]\right] = \frac{1}{N^{2}} \sum_{l=0}^{N-1} \sum_{m=0}^{N-1} \left(r_{yy}^{2}[m-l] + r_{yy}[k+m-l]r_{yy}[k-(m-l)]\right)$$

$$= \frac{1}{N^{2}} \sum_{l=0}^{N-1} \sum_{m=0}^{N-1} \left(\varrho^{2|m-l|} + \varrho^{|k+m-l|+|k-(m-l)|}\right)$$

$$= \frac{1}{N^{2}} \sum_{l=0}^{N-1} \left(\sum_{m=l}^{N-1} \varrho^{2(m-l)} + \sum_{m=0}^{l-1} \varrho^{2(l-m)}\right)$$

$$+ \frac{1}{N^{2}} \sum_{l=0}^{N-1} \left(\varrho^{-2l} \sum_{m=l+k}^{N-1} \varrho^{2m} + \varrho^{2k} \sum_{m=l-k+1}^{l+k-1} + \varrho^{2l} \sum_{m=0}^{l-k} \varrho^{-2m}\right) (C.1)$$

Solving the geometric series yields

$$\mathbf{E}\left[\left[\epsilon^{2}(n)\right]\right] = \frac{1}{N} \left(\frac{\varrho^{2} - \varrho^{-2} + 4N^{-1}\left(1 - \varrho^{2N}\right) + 2\varrho^{2k}\left(1 - \varrho^{-2}\right)}{\left(1 - \varrho^{2}\right)\left(1 - \varrho^{-2}\right)} + \varrho^{2k}(2k - 1)\right)$$
(C.2)

For an AR(1) process in AWGN, the estimation error can be derived as

$$\begin{split} \mathbf{E} \left[\hat{e}_{r}^{2}[k] \right] &= \\ &= \frac{1}{N^{2}} \sum_{l=0}^{N-1} \sum_{m=0}^{N-1} \left(r_{yy}^{2}[m-l] + r_{yy}[k+m-l]r_{yy}[k-(m-l)] \right) \\ &= \frac{1}{N^{2}} \sum_{l=0}^{N-1} \sum_{m=0}^{N-1} \left(\varrho^{|m-l|} + \sigma_{n}^{2}\delta[m-l] \right)^{2} \\ &+ \frac{1}{N^{2}} \sum_{l=0}^{N-1} \sum_{m=0}^{N-1} \left(\varrho^{|k+m-l|} + \sigma_{n}^{2}\delta[k+m-l] \right) \left(\varrho^{|k-(m-l)|} + \sigma_{n}^{2}\delta[k-(m-l)] \right) \\ &= \frac{1}{N^{2}} \sum_{l=0}^{N-1} \sum_{m=0}^{N-1} \left(\varrho^{2}[m-l] + \varrho^{|k+m-l|+|k-(m-l)|} \right) \\ &+ \frac{\sigma_{n}^{2}}{L_{2}^{2}} \sum_{l=0}^{N-1} \sum_{m=0}^{N-1} \left(2\varrho^{|m-l|} \delta[m-l] + \sigma_{n}^{2}\delta^{2}[m-l] \right) \\ &+ \frac{\sigma_{n}^{2}}{N^{2}} \sum_{l=0}^{N-1} \sum_{m=0}^{N-1} \left(\varrho^{|k+m-l|}\delta[k-(m-l)] \right) \\ &+ \frac{\varphi^{2}[k-(m-l)|\delta[k+m-l]}{N} \\ &+ \frac{\varphi^{2}[k-(m-l)|\delta[k+m-l]}{N} \\ &+ \frac{\sigma_{n}^{2}}{N^{2}} \sum_{l=0}^{N-1} \sum_{m=0}^{N-1} \left(\varrho^{|k+m-l|}\delta[k-(m-l)] \right) \\ &= \Psi + \frac{\sigma_{n}^{2}}{N^{2}} \sum_{l=0}^{N-1} \sum_{m=0}^{N-1} \left(\varrho^{|k+m-l|}\delta[k-(m-l)] + \varrho^{|k-(m-l)|}\delta[k+m-l] \right) \\ &= \Psi + \frac{\sigma_{n}^{2}}{N} \left(2 \left(2 + \varrho^{2k} \right) + \sigma_{n}^{2} (1 + \delta[k]) \right) \end{split}$$
(C.3)

where Ψ is the error variance for estimating the autocorrelation function of the unit variance AR(1) process without noise, found in (C.2). Substituting in for Ψ yields the final expression

$$\mathbf{E}\left[\epsilon_{r}^{2}[k]\right] = \frac{1}{N} \left(\frac{\varrho^{2} - \varrho^{-2} + 4N^{-1} \left(1 - \varrho^{2N}\right) + 2\varrho^{2k} \left(1 - \varrho^{-2}\right)}{(1 - \varrho^{2}) \left(1 - \varrho^{-2}\right)} + \varrho^{2k} (2k - 1) \right) \\ + \frac{\sigma_{n}^{2}}{N} \left(2 \left(2 + \varrho^{2k}\right) + \sigma_{n}^{2} (1 + \delta[k]) \right) \\ = \frac{1}{N} \left(\frac{\varrho^{2} - \varrho^{-2} + 4N^{-1} \left(1 - \varrho^{2N}\right) + 2\varrho^{2k} \left(1 - \varrho^{-2}\right)}{(1 - \varrho^{2}) \left(1 - \varrho^{-2}\right)} \right) \\ + \frac{1}{N} \left(\varrho^{2k} \left(2\sigma_{n}^{2} + 2k - 1 \right) + 4\sigma_{n}^{2} + \sigma_{n}^{4} (1 + \delta[k]) \right)$$
(C.4)

The above yields the final expression for the estimation error when estimating the autocorrelation function of a unit variance AR(1) process with parameter ρ in AWGN with noise variance σ_n^2 with a rectangular window utilizing N samples.