Circuits and Systems
Mekelweg 4,
2628 CD Delft
The Netherlands
http://ens.ewi.tudelft.nl/

upna Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

M.Sc. Thesis

Compressive Power Spectral Density Estimation with Non-Uniform Sampling

Fernando de la Hucha Arce

Abstract

Wireless communications have experienced a fast growth over the last two decades, which is still going on nowadays. This fact, together with the current static spectrum allocation, has made spectral resources scarce and has left very few bandwidth to new applications. However, actual measurements of spectrum utilization show that many assigned bands are not being used at every location and time. A solution that allows secondary users to transmit information using available spectrum is cognitive radio. A cognitive radio transceiver performs spectrum sensing, that is, it estimates the power spectral density of the received signal and then reliably detects whether unused spectrum is available or not. The estimation has to be done over a wide frequency range in order to increase the probability of finding available spectrum. With uniform sampling, the scheme considered in classical sampling theory, estimation over a wide frequency band requires very high sampling rates, which for current analog-to-digital conversion technology means a high power consumption.

In this thesis, we study several methods for estimation of the power spectral density. They use non-uniform sampling schemes which allow reductions in the sampling rate, hence the word 'compressive'. These methods are least squares with hard limiting, maximum likelihood and a correlogram approach. The first two strategies use compressive sampling, while the last one employs random sampling. The performance of each method is studied by means of computer simulations. Besides, we have attempted to find new sampling matrices that guarantee the statistical identifiability of the power spectral density when compressive sampling is employed.

TU Delft

Faculty of Electrical Engineering, Mathematics and Computer Science — Delft University of Technology

Compressive Power Spectral Density Estimation with Non-Uniform Sampling

Thesis

submitted in partial fulfillment of the
requirements for the degree of

Telecommunication Engineering[1]

at

Universidad Pública de Navarra

by

Fernando de la Hucha Arce
born in Madrid, Spain

Advisor in TU Delft dr. Geert Leus

Mentor in TU Delft M.Sc. Dyonisius Dony Arianandа

Supervisor in UPNA dr. Miguel Ángel Gómez Laso

This work was performed as part of the ERASMUS programme in:

Circuits and Systems Group
Department of Microelectronics
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology

[1]The official name of the degree in Spain is Ingeniero de Telecomunicación.

Delft University of Technology
Department of
Microelectronics

Dated: 27-08-2013

Committee Members:

dr. Geert Leus

dr. ir. Rob Remis

M.Sc. Dyonisius Dony Arianandra

Abstract

Wireless communications have experienced a fast growth over the last two decades, which is still going on nowadays. This fact, together with the current static spectrum allocation, has made spectral resources scarce and has left very few bandwidth to new applications. However, actual measurements of spectrum utilization show that many assigned bands are not being used at every location and time. A solution that allows secondary users to transmit information using available spectrum is cognitive radio. A cognitive radio transceiver performs spectrum sensing, that is, it estimates the power spectral density of the received signal and then reliably detects whether unused spectrum is available or not. The estimation has to be done over a wide frequency range in order to increase the probability of finding available spectrum. With uniform sampling, the scheme considered in classical sampling theory, estimation over a wide frequency band requires very high sampling rates, which for current analog-to-digital conversion technology means a high power consumption.

In this thesis, we study several methods for estimation of the power spectral density. They use non-uniform sampling schemes which allow reductions in the sampling rate, hence the word 'compressive'. These methods are least squares with hard limiting, maximum likelihood and a correlogram approach. The first two strategies use compressive sampling, while the last one employs random sampling. The performance of each method is studied by means of computer simulations. Besides, we have attempted to find new sampling matrices that guarantee the statistical identifiability of the power spectral density when compressive sampling is employed.

Acknowledgments

First of all, I would like to thank my thesis advisor, Geert Leus, for giving me the opportunity to work in a challenging and interesting topic, and for his support during this year. His advice and positive attitude encouraged me to keep on working. I cannot forget the help of Dyonisius Dony Ariananda, who took the effort of helping me with my many questions, revised my reports and gave valuable advice on my work. Thanks a lot Dony for your support and patience. I would also like to thank my supervisor from the Public University of Navarra, Miguel Ángel Gómez Laso, who was very attentive and was always there whenever I needed advice. Besides, I wish to thank the people responsible of the ERASMUS programme, for making possible for some students to live a truly enriching experience in a different country.

Of course, I am very grateful to my family, without their help and support this adventure would not have been possible. Special thanks to my sister Ainhoa, her encouraging and cheerful nature always makes me smile. Many thanks to all the great people I have met in Delft, who made my life there wonderful. Special thanks to Andrea and Yuri, for all the shared moments. I cannot forget to give warm thanks to all my friends in Spain. You are too many to list here, but I am really lucky to have met all of you.

Finally, I would like to thank Funda, for her constant support and for our awesome moments together.

Fernando de la Hucha Arce
Delft
27-08-2013

Contents

Abstract iii

Acknowledgments v

1 Introduction 1
- 1.1 Motivation . . . 1
- 1.2 Organization of the thesis . . . 2

2 Overview of Spectral Estimation and Sampling 5
- 2.1 Sampling theory . . . 5
  - 2.1.1 Compressive Sampling . . . 6
  - 2.1.2 Multicoset Sampling . . . 7
  - 2.1.3 Random Sampling . . . 7
- 2.2 Spectral Estimation . . . 8
  - 2.2.1 The Autocorrelation Matrix . . . 9
  - 2.2.2 Statistical identifiability . . . 10

3 Least Squares Estimation with Hard Clipping 11
- 3.1 Least Squares Estimation of [1] . . . 11
- 3.2 Least Squares Estimation with Hard Clipping . . . 13
- 3.3 Simulation results . . . 16
- 3.4 Conclusions . . . 18

4 Maximum Likelihood Estimation 19
- 4.1 Statistical assumptions . . . 19
- 4.2 Derivation of the Maximum Likelihood Estimator . . . 21
- 4.3 Asymptotic Solutions to the ML Equations . . . 22
  - 4.3.1 Asymptotic Maximum Likelihood . . . 23
  - 4.3.2 Simplified Inverse Iteration Algorithm . . . 23
- 4.4 Complex data . . . 23
- 4.5 Simulation results . . . 24
- 4.6 Conclusions . . . 26

5 Sampling Matrix Design 29
- 5.1 Conditions for Full Column Rank . . . 29
- 5.2 Binary sampling matrices . . . 30
  - 5.2.1 Binary matrices of class $\mathbb{B}_1$ and concatenated sparse rulers . . . 32
  - 5.2.2 Binary matrices of class $\mathbb{B}_2$ . . . 33
- 5.3 Conclusions . . . 34

6 Spectral Estimation with Random Sampling 35
- 6.1 Additive Random Sampling . . . 35
- 6.2 Probability of Obtaining a Certain Lag . . . 36
- 6.3 Autocorrelation estimation . . . 38

6.4 Simulation results 39
6.5 Conclusions 41

7 Conclusions and Future Work **43**
7.1 Conclusions 43
7.2 Suggestions for Further Work 44

Bibliography **45**

List of Figures

3.1 System for LS estimation with a Hard Limiter 15
3.2 True PSD with $N = 63$, complex case 17
3.3 PSD estimates for LSE with and without hard limiting 17
3.4 NMSE vs Compression Rate for LS with a Hard Limiter 18

4.1 True PSD with $N = 63$, real case 25
4.2 PSD estimates using AML, SIIA and LSE 26
4.3 NMSE vs Compression Rate for MLE 27
4.4 NMSE vs SNR for MLE 27

6.1 True PSD with $N = 63$, complex case 40
6.2 PSD estimates with random sampling 40
6.3 NMSE vs Compression Rate for random sampling 41

1 Introduction

In this thesis, we study compressive spectral estimation using non-uniform sampling. Spectral estimation is the problem of determining, from a finite number of measurements, the distribution of the average power of a signal over frequency. The word 'compressive' refers to the rate at which these measurements are acquired. With uniform sampling, which is the most common type of sampling in signal processing, the rate is required to exceed the Nyquist rate in order to avoid aliasing, which prevents meaningful spectral estimation. Thus, over the years, research has been conducted in several schemes of non-uniform sampling that can avoid aliasing, even when the total sampling rate is below the Nyquist rate.

In this chapter, we motivate the need for studying compressive spectral estimation methods and explain the organization and content of subsequent chapters.

1.1 Motivation

One of the important problems in the field of signal processing is spectral estimation, which aims at extracting meaningful characteristics of a signal in the frequency domain. When the signal is modelled as a random process, it is often of interest to obtain its autocorrelation function or equivalently, under the assumption of wide sense stationarity, its power spectral density (PSD). This function describes the distribution of the average power of the signal of interest over frequency. In order to avoid aliasing, traditionally the power spectral density is estimated from samples taken from the signal at a rate higher than the Nyquist rate, which is twice the maximum frequency present in its spectrum, as proven in the Shannon-Whittaker-Kotelnikov[1] theorem [3]. The Nyquist rate sets a limit on the frequency ranges that can be estimated digitally, because sampling wideband signals would require very high sampling rates that nowadays are unfeasible for analog-to-digital converters, or would require a very high power consumption.

On pair of this, wireless communications have experienced a fast growth over the last two decades, which is still going on nowadays. The current regulatory laws allocate portions of the spectrum to licensed operators, who have to pay fees for its use. Thus, unlicensed applications are restricted to certain frequency bands, such as the well known ISM band 2.4 - 2.4835 GHz which is employed, among other technologies, by IEEE 802.11 (WiFi networks) and Bluetooth. As a consequence of the sustained growth of wireless communications, this static spectrum allocation has made spectral resources scarce and has left very few bandwidth to new applications. However, actual measurements of spectrum utilization show that many assigned bands are not being used at every location and time. Studies on spectrum use date back as far as 2002, with a report published by the Federal Communications Commission [4]. This report also discusses possible solutions that will improve accessibility to the spectrum and its efficient use.

[1] For historical details on the name, the first and second chapters of [2] provide a good overview.

A solution to the problems of spectrum scarcity and spectrum access that has been proposed by the scientific community is cognitive radio [5]. In this new approach to wireless communications, every device belonging to an unlicensed user scans the environment in order to detect occupied and unoccupied bands, and adjusts its communication parameters accordingly. In the case of detecting an emtpy band, the device can use the available spectrum to transmit information. By doing this, unlicensed users only transmit during the periods when the licensed users are inactive, so ideally harmful interference to them is avoided. In [6], cognitive radio is defined as an 'intelligent wireless communication system that is aware of its surrounding environment (i.e., outside world), and uses the methodology of understanding-by-building to learn from the environment and adapt its internal states to statistical variations in the incoming RF stimuli by making corresponding changes in certain operating parameters (e.g., transmit-power, carrier-frequency, and modulation strategy) in real-time, with two primary objectives in mind:

- highly reliable communications whenever and wherever needed;
- efficient utilization of the radio spectrum.'

In order to learn from the environment, a cognitive radio transceiver must perform spectrum sensing, that is, it must estimate the power spectral density of the received signal and then reliably detect whether unused spectrum is available or not. Estimation requires time and power to collect measurements from the incoming signal and process them, so one of the challenges is to estimate the PSD of a wideband signal spending as less time and power as possible. A major source of high power consumption in analog-to-digital converters is a large sampling rate [7], so research efforts have been made in spectral estimation methods with the aim of reducing the sampling rate necessary for wideband sensing. Thus, our work is developed with this goal in mind.

1.2 Organization of the thesis

Before describing the content of each chapter, we summarize our contributions. We have studied several estimation methods using two different sampling schemes at sub-Nyquist rates. These schemes are compressive sampling, which includes multicoset sampling as a particular case, and random sampling· For compressive sampling, we have studied maximum likelihood estimation and least squares estimation after hard limiting, and we have addressed the problem of designing sampling matrices that guarantee statistical identifiability of the power spectral density. For random sampling, we have studied a modification of the additive random sampling model by quantizing the sampling instants, so the scheme can be implemented in a synchronous digital system. The estimation method in this case uses the classical unbiased estimator of the autocorrelation. Besides, we addressed the problem of computing the probability of a certain lag being present, and we showed that the exact computation is a difficult problem.

Chapter 2: Overview of Spectral Estimation and Sampling In this chapter, we review sampling theory and spectral estimation, and we present a literature review on both topics. Regarding sampling theory, we explain the sampling schemes we will use throughout this work. Regarding spectral estimation, we review the definition of the power spectral

density and the properties of the autocorrelation matrix. Finally, we address the conditions for statistical identifiability, which are needed to guarantee that consistent estimation is possible. Within this section, we also explain the concept of a sparse ruler, which is important when multicoset sampling is used.

Chapter 3: Least Squares Estimation with Hard Clipping This chapter begins with an explanation of a PSD estimator proposed in [1]. The reason is that we study a modified version of this estimator that applies a hard limiter to the measurements. The use of hard limiters in the context of spectral estimation has already been addressed in the literature, but, to the best of our knowledge, it has not been combined together with compressive sampling. We describe our modified estimator, detailing its use within the compressive sampling framework for complex signals that come from a Gaussian process. This assumption is necessary in order to allow the recovery process, because this method relies on a mapping between the correlation of two real Gaussian processes and their hard limited versions. Finally, its performance is studied by means of computer simulations, and compared to the method in [1].

Chapter 4: Maximum Likelihood Estimation Here we study maximum likelihood (ML) estimation, which is a very popular estimation method due to its good statistical properties when the sample size is large. In the case where the received signal is Gaussian, we have derived the equations that yield the ML estimator for real signals. Because of the difficulty of obtaining an analytical solution and the high computational cost of the numerical methods that find the true solution, we investigate two algorithms which yield an approximate solution that asymptotically approaches the true one. We briefly consider under which conditions these algorithms can be used for complex signals, and finally we compare the performance of these two algorithms with the least squares estimator of [1].

Chapter 5: Sampling Matrix Design In this chapter, we focus on the problem of designing sampling matrices for compressive sampling that guarantee the statistical identifiability of the PSD. We begin by considering this problem in general, which boils down to constructing a basis where each element is the auto- or cross-correlation of two finite length sequences. Because of its difficulty, we have left open the general construction for further research. Then, we restrict the sequences to have only elements from the set $\{0, 1\}$, though we note that the sets $\{-1, 1\}$ and $\{-1, 0, 1\}$ may also be interesting. For multicoset sampling, we propose a design, based on the concatenation of sparse rulers of different lengths, that no longer has the minimum compression rate property, but allows more design flexibility. However, the statistical identifiability is not guaranteed. For binary matrices with two ones per row, we study the conditions for the statistical identifiability to be satisfied, but we have not been able to find sufficient conditions.

Chapter 6: Spectral Estimation with Random Sampling This chapter is devoted to random sampling, a scheme where the time between samples is a random variable. An important result from the literature is that consistent estimation is possible even for mean sampling rates lower than the Nyquist rate. First of all, we explain our modification of the classical model of additive random sampling by rounding each sampling instant to the closest instant in a Nyquist rate grid. This eliminates the possibility of two sampling instants

being infinitely close. Then, we study the probability of obtaining a certain lag, whose exact computation is difficult, as we show. Finally, we present a numerical study of the performance of the PSD estimator, which is based on the classical unbiased estimator of the autocorrelation.

Chapter 7: Conclusions and Future Work This chapter summarizes the main ideas from the thesis and gives suggestions for further work.

Overview of Spectral Estimation and Sampling

2

Spectral estimation aims at obtaining the distribution of power over frequency of a signal, which is sampled in order to process its measurements digitally. With classical sampling theory, the samples are taken uniformly at a rate higher than twice the maximum frequency present in the signal. The purpose is to avoid aliasing, which makes reconstruction impossible. Reducing partially or completely aliasing and decreasing the sampling rate are some of the main challenges in both spectral estimation and signal reconstruction. Research has been conducted at least since 1960 with the work of [8], and in the recent years renewed interest has come from the fields of cognitive radio [6] and compressive sampling [9]. In this chapter, we review sampling theory and some spectral estimation methods proposed in the literature. Besides, we highlight some of the concepts that we use in our work.

2.1 Sampling theory

Sampling theory is a field of mathematics dealing with methods to accurately represent a function from a countable number of its values. For functions over the real numbers, the problem can be stated as how to reconstruct the function $f(t)$, $\forall t \in \mathbb{R}$, from the set $\{f(t_n)\}_{n\in\mathbb{Z}}$. In signal processing, solutions to this problem are of key importance, as they provide a way to preserve the information of real world signals, which are analog, from measurements taken at certain instants.

One of the most celebrated theorems in this field is the Shannon-Whitakker-Kotelnikov theorem, which deals with perfect reconstruction of a bandlimited function from samples taken at equidistant instants. Shannon applied this result to communication theory in his work [3]. Mathematically, the reconstruction formula can be stated as

$$f(t) = \sum_{n=-\infty}^{\infty} f(nT_s)\,\mathrm{sinc}\left(\frac{t-nT_s}{T_s}\right), \tag{2.1}$$

where $\mathrm{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$, and T_s is the distance between sampling instants, called the sampling period. We note here that this formula has the form of a convolution with the impulse response of an ideal low-pass filter, which is non-causal, making it physically irrealizable. In practice, other interpolation methods are used. The theorem also provides sufficient conditions for the interpolation formula to hold. The first is that $f(t)$ is bandlimited, which means that its Fourier transform must vanish outside an interval $[-f_0, f_0]$. The second is that the sampling rate $f_s = \frac{1}{T_s}$ has to be at least twice higher than the maximum frequency f_0. The rate $f_s = 2f_0$ is called the Nyquist rate, and it is the minimum rate that guarantees perfect reconstruction. Satisying these two conditions avoids aliasing, that is, it is guaranteed that replicas of the spectrum of $f(t)$ do not overlap.

From a signal processing point of view, this theorem provides a simple way of taking samples from an analog signal, which can be stored and processed by digital means with ideally no loss of information. Sampling at equidistant instants $t_n = nT_s$ is called uniform sampling,

and it is the basis of digital signal processing. For more details, the reader can consult [10]. However, when the bandwidth is very high, satisfying the condition $f_s \geq 2f_0$ imposes a heavy burden on analog-to-digital converters, especially regarding power consumption, as studied in [7]. In certain applications, such as spectrum sensing, this is an important issue when the sensing task has to be done by devices powered by batteries. For instance, in the field of cognitive radio, unlicensed users need to estimate the distribution of power over frequency in order to detect inactive bands, which they can use for their own purposes, as we explained in Chapter 1.

Before the emergence of the cognitive radio idea, the desire to avoid aliasing led researchers to investigate new sampling schemes that are different from uniform sampling, such as random and multicoset sampling. More recently, several researchers proposed a new sampling framework, called compressed sensing [11, 12], which aims at perfect reconstruction of a high dimensional vector from samples that are linear functionals of the vector. Assuming that the vector is sparse in some basis, this will allow a sampling rate reduction by using sparse recovery methods. We will briefly explain the sampling schemes we consider in this work in the following sections.

2.1.1 Compressive Sampling

Compressive sampling, which is another name for compressed sensing, approaches sampling in a different way from (2.1). The signal of interest is assumed to be a finite dimensional vector that is k-sparse in some basis. This means that it can be represented by k non-zero coefficients in that basis. Instead of signal values, the measurements are inner products of the signal vector with a predefined set of vectors, which are grouped in a matrix called the sampling matrix. From these inner products, the goal is to reconstruct the signal vector using optimization techniques or greedy algorithms. These techniques find a solution to an underdetermined system of linear equations that has the same sparsity order k. An overview of compressive sampling is given in [9].

In our work, we use the same acquisition scheme of compressive sampling, but our reconstruction techniques are different, as we are interested in recovering statistical parameters. We consider a bandlimited signal $x(t)$ and we denote its Nyquist rate samples by the sequence $x[n]$. These samples can be grouped in the $N \times 1$ vector sequence $\mathbf{x}[k]$, which is given by

$$\mathbf{x}[k] = [x[kN], x[kN+1], \ldots, x[kN+N-1]]^T .$$

The compressive measurements are collected in the $M \times 1$ vectors $\mathbf{y}[k]$, given by

$$\mathbf{y}[k] = [y_0[k], y_1[k], \ldots, y_{M-1}[k]]^T .$$

The relationship between the measurements and the Nyquist rate samples of $x(t)$ can be represented mathematically by a sampling matrix $\mathbf{C}$, which establishes a linear relationship between them as follows

$$\mathbf{y}[k] = \mathbf{C}\,\mathbf{x}[k]. \tag{2.2}$$

We can express the $M \times N$ sampling matrix $\mathbf{C}$ in terms of its rows,

$$\mathbf{C} = [\mathbf{c}_0, \mathbf{c}_1, \ldots, \mathbf{c}_{M-1}]^T, \tag{2.3}$$

where $\mathbf{c}_i = [c_i[0], c_i[-1], \ldots, c_i[1-N]]^T$. Then we can also write the relationship between the measurements and the Nyquist rate samples as

$$y_i[k] = \sum_{n=1-N}^{0} c_i[n]\, x[kN - n]. \tag{2.4}$$

Thus, the measurements can be seen as the N-fold downsampled output of M non-causal digital filters whose impulse responses $c_i[n]$ have length N, and whose input is $x[n]$.

This acquisition model was used in [1], where they propose an analog implementation based on a bank of mixers and low-rate analog-to-digital converters, and they prove that its operation is equivalent to (2.4). In the i-th branch, the incoming signal $x(t)$ is multiplied by a piecewise constant function that encodes the values $c_i[n]$, and the resulting signal is sampled uniformly by an integrate-and-dump device working at a rate NT, where T is the Nyquist rate corresponding to the scanned bandwidth. As the device has M branches, the compression rate is $\frac{M}{N}$.

2.1.2 Multicoset Sampling

Multicoset sampling, also called periodic non-uniform sampling, is a scheme where a subset of the Nyquist rate samples is selected. It has been used for perfect reconstruction of multiband signals with reduced sampling rates and with partial or no knowledge of the bands locations in several works. In [13], the locations of the bands must obey a mathematical relationship to allow perfect reconstruction. A fully blind system that recovers the locations of the bands prior to reconstruction is proposed in [14]. The limitations of multicoset sampling to directly acquire multiband signals are addressed in [15], where they propose a new device called modulated wideband converter.

Let N and M be positive integers, and let $\mathcal{N}$ be a set of M different positive integers $\mathcal{N} = \{n_0, n_1, \ldots, n_{M-1}\}$, such that $n_m \leq N\,, \forall m$. The sampling instants in multicoset sampling are given by

$$\{n_m T + kNT\}_{m=0}^{M-1}\,, \quad k \in \mathbb{Z},$$

where $\frac{1}{T}$ is the Nyquist rate. The average sampling rate is given by $\frac{M}{NT}$, so if $M < N$ the scheme provides compression. For a given N, the set $\mathcal{N}$ characterizes which Nyquist rate samples are selected. The sampling scheme is periodic because the signal is divided in blocks of length N and the indexes kept are the same up to a shift by a multiple of N. Using the compressive sampling notation, multicoset sampling retains from $\mathbf{x}[k]$ only the samples indexed by $\mathcal{N}$. Thus, we can write $\mathbf{y}[k]$ as

$$\mathbf{y}[k] = [x[kN + n_0], x[kN + n_1], \ldots, x[kN + n_{M-1}]]^T\,.$$

The corresponding $M \times N$ sampling matrix $\mathbf{C}$ is the one whose rows contain a single one, in the positions given by $\mathcal{N}$, and $N-1$ zeros. This is equivalent to saying that $\mathbf{C}$ is a submatrix of the identity matrix $\mathbf{I}_N$ formed by selecting its rows according to the set $\mathcal{N}$.

2.1.3 Random Sampling

Random sampling is a type of sampling where the time between samples is a random variable following a specific distribution. This randomness may be caused by diverse factors, such as the physical process from which the signal is extracted, the measurement conditions, or the

will of the designer. In this last case, research regarding random sampling has been carried out at least since 1960, with the work presented in [8]. The authors prove that it is possible to avoid aliasing even at sub-Nyquist mean sampling rates for certain classes of sampling schemes. Sampling with a Poisson random process is treated in [16], and a more general approach is taken in [17] to obtain conditions for alias free consistent estimation. These papers also prove consistency for mean sampling rates less than the Nyquist rate. More recently, [18] analyzed several types of random sampling for spectral estimation, comparing their properties and their possible application to radio transceiver design. An application of spectral analysis of random sampling to communications is proposed in [19], where they focus on detection of the active bands rather than in estimation of the power spectral density.

In our work we focus on a particular model of random sampling called additive random sampling, which was first proposed in [8]. The samples in additive random sampling are taken at the instants

$$t_0 = 0,$$

$$t_n = t_{n-1} + \tau_n \quad n \in \{1, 2, \ldots, N-1\}.$$

The random variables τ_n are independent and identically distributed (i.i.d), which means that all intervals between consecutive samples are independent and have the same distribution. We will address Poisson and uniform sampling in Chapter 6.

2.2 Spectral Estimation

Spectral estimation deals with the problem of estimating the power spectral density of a random process from a finite number of measurements, typically noisy. Mathematically, if the received signal is modelled as a wide-sense stationary (WSS) random process $x(t)$, its autocorrelation function $r_x(t_1, t_2) = E\left[x(t_1)x^*(t_2)\right]$, where the superscript * represents complex conjugation, only depends on the time difference $\tau = t_1 - t_2$. Then, the well known Wiener-Khinchin theorem proves that the power spectral density $\Phi_x(\omega)$ and the autocorrelation $r_x(\tau)$ form a Fourier transform pair. The name power spectral density comes from the fact that

$$r_x(0) = E\left[x^2(t)\right] = \int_{-\infty}^{\infty} \Phi_x(\omega)\, d\omega. \tag{2.5}$$

This means that the average power of $x(t)$ can be found as the area under $\Phi_x(\omega)$.

If we assume that $x(t)$ is bandlimited with frequency support $[-f_0, f_0]$, it suffices to consider the discrete-time random process $x[n]$ obtained by sampling $x(t)$ at the Nyquist rate, which is $\frac{1}{T} = 2f_0$. The process $x[n]$ is also wide-sense stationary, and we can define its power spectral density $\phi_x(\Omega)$ using the discrete time Fourier transform as

$$\phi_x(\Omega) = \sum_{m=-\infty}^{\infty} r_x[m]\, e^{-j\Omega m}, \quad 0 \leq \Omega < 2\pi, \tag{2.6}$$

where $r_x[m]$ is the autocorrelation sequence of $x[n]$, defined by $r_x[m] = E\left(x[n]\, x^*[n-m]\right)$, and Ω is the angular frequency measured in rad/sample. More details on the power spectral density can be found in [20].

As we can see from equation (2.6), estimating $\phi_x(\Omega)$ can be done by estimating $r_x[m]$ and then taking its discrete time Fourier transform, which is usually done employing the

discrete Fourier transform. The methods for spectral estimation can be classified in two types, depending on whether they assume a particular signal model or not. Non-parametric methods estimate $r_x[m]$ without assuming any other condition than wide-sense stationarity of $x(t)$, while parametric methods use a signal model, such as autoregressive moving average, or multiple sinusoids. If the model is accurate for a particular situation, parametric methods yield better performance with shorter data records. However, they are sensitive to model mismatch. Several books that cover classical spectral estimation are available, and we refer to [21]. A survey on methods for spectral estimation with irregular sampling is the subject of [22].

Because in this work we are interested in wideband signals that cannot be assumed to belong to a particular model, we only study non-parametric techniques. These techniques will often use the properties of the autocorrelation matrix of a vector formed by N Nyquist rate samples from $x(t)$, which we defined as $\mathbf{x}[k]$. For this reason, we will review these properties in the next section.

2.2.1 The Autocorrelation Matrix

In order to estimate the autocorrelation sequence of the Nyquist rate samples $r_x[m]$, we can construct the $N \times N$ autocorrelation matrix of $\mathbf{x}[k]$, defined as $\mathbf{R}_x = E\left(\mathbf{x}[k]\mathbf{x}^H[k]\right)$. This matrix is Hermitian and positive semidefinite, and because of our assumption that $x[n]$ is a wide-sense stationary process, it has also a Toeplitz structure. When the process $x[n]$ has zero mean, $\mathbf{R}_x$ is also the covariance matrix. For details, the reader can consult [21]. The Hermitian Toeplitz structure means that the element $[\mathbf{R}_x]_{i,j}$ is given by

$$[\mathbf{R}_x]_{i,j} = r_x[i-j] = r_x^*[j-i].$$

Thus, the first column of $\mathbf{R}_x$ contains the values $r_x[0], r_x[1], \ldots, r_x[N-1]$, and the rest of the matrix contains the same elements or their conjugates. We can group the first column and row in the $(2N-1) \times 1$ vector

$$\mathbf{r}_x = [r_x[0], r_x[1], \ldots, r_x[N-1], r_x[1-N] \ldots, r_x[-1]]^T, \tag{2.7}$$

which is the vector of parameters we want to estimate. In the case where $x[n]$ is real, $\mathbf{R}_x$ is symmetric, and it suffices to consider the autocorrelation values at the positive lags, which we collect in the $N \times 1$ vector

$$\tilde{\mathbf{r}}_x = [r_x[0], r_x[1], \ldots, r_x[N-1]]^T. \tag{2.8}$$

Then, obtaining $\mathbf{r}_x$ can be done by concatenating $\tilde{\mathbf{r}}_x$ and its reversed version without the first element, i.e., $[r_x[N-1], r_x[N-2], \ldots, r_x[1]]^T$.

Now we consider the relationship between $\mathbf{R}_x$ and the $M \times M$ autocorrelation matrix of the compressive measurements $\mathbf{y}[k]$, which is defined as $\mathbf{R}_y = E\left(\mathbf{y}[k]\mathbf{y}^H[k]\right)$. Using (2.2), we can write

$$\mathbf{R}_y = \mathbf{C}\mathbf{R}_x\mathbf{C}^H. \tag{2.9}$$

This matrix is not Toeplitz for most matrices $\mathbf{C}$, which means that the measurement vectors $\mathbf{y}[k]$ are not WSS.

Once the estimate $\hat{\mathbf{r}}_x$ is calculated, the PSD estimate $\hat{\phi}_x$ is obtained by using the discrete Fourier transform (DFT). This can be written as

$$\hat{\phi}_x = \mathbf{F}_{2N-1}\hat{\mathbf{r}}_x,$$

where $\mathbf{F}_{2N-1}$ is the $(2N-1) \times (2N-1)$ DFT matrix.

We note here that we have constrained our reconstruction of $r_x[m]$ to the interval $[1-N, N-1]$. Thus, unless the process $x[n]$ has zero autocorrelation outside this interval, there is a truncation error in the PSD estimator. In practice, the value of N has to be selected according to prior knowledge of the rate of decay of the autocorrelation of the process.

2.2.2 Statistical identifiability

Statistical identifiability is the property of a statistical model that tells whether it is possible or not to consistently estimate the desired parameters. In particular, if the parameters are identifiable, then different values of the parameters give a different probability distribution for the data. For example, in the problem of estimating the variance σ^2 of Gaussian data, every different value of σ^2 gives a different Gaussian probability density function, and thus identifiability is satisfied.

In our case, without compression, the parameter vector $\mathbf{r}_x$ is identifiable, because every different value of $\mathbf{r}_x$ yields a different autocorrelation matrix $\mathbf{R}_x$. However, this is not necessarily the case after compression, because it might happen that different autocorrelation vectors $\mathbf{r}_{x,1}$, $\mathbf{r}_{x,2}$ yield the same matrix $\mathbf{R}_y$, depending on the sampling matrix $\mathbf{C}$. A study on the identifiability of compressive covariance matrix estimation was conducted in [23], where they study the general criterion for a sampling matrix $\mathbf{C}$ to satisfy identifiability, and they prove which matrices satisfy this condition for multicoset sampling.

The general condition states that, expressing $\mathbf{R}_x$ in a basis for the subspace of $N \times N$ Toeplitz positive semidefinite matrices, the linear transformation of this basis given by (2.9) has to yield an independent set of matrices, that is, it must preserve linear independence. For multicoset sampling, this boils down to choosing the set $\mathcal{N}$ according to a sparse ruler. This multicoset sampling design was first applied to compressive spectral estimation in [1, 24], where they prove that solving the minimal $(N-1)$-length sparse ruler problem yields the minimum compression rate for their algorithm while maintaining statistical identifiability. In the next section, we explain the concept of a sparse ruler.

2.2.2.1 Sparse Rulers

Intuitively, a sparse ruler of length N is a ruler with M marks that is still able to measure all integer distances up to N. To state this concept mathematically, we first define the difference set $\Delta(\mathcal{S})$ of a set $\mathcal{S}$ as

$$\Delta(\mathcal{S}) = \{d = s_i - s_j; \forall s_i, s_j \in \mathcal{S} \text{ s.t. } d \geq 0\}.$$

This is the set of all non-negative differences that can be obtained from elements of $\mathcal{S}$. Now we can state formally the concept of a sparse ruler.

Definition: A sparse ruler of length $N-1$ is a set $\mathcal{R} \subset \{0, 1, \ldots, N-1\}$ such that $\Delta(\mathcal{R}) = \{0, 1, \ldots N-1\}$. It is called minimal if no other sparse ruler of length $N-1$ exists with less elements.

The problem of sparse rulers has been studied in [25]. To the best of our knowledge, there is no quick procedure for finding minimal sparse rulers of a given length, making a brute force search necessary. This is more time consuming as N increases. However, several sparse rulers have been tabulated, making multicoset sampling based on sparse rulers feasible, at least for not very large values of N.

Least Squares Estimation with Hard Clipping

3

In this chapter, we study least squares estimation (LSE) of the autocorrelation sequence $r_x[m]$ using compressive sampling. We begin by explaining the method proposed in [1, 24], which derives a linear relationship between the cross-correlations of the compressive measurements and the autocorrelation sequence of the Nyquist rate samples. Afterwards, when the signal $x(t)$ is assumed to come from a Gaussian process, we propose a modification of this method that adds a hard limiter, a device that returns 1 when the input is greater than 0 and -1 when it is lower than 0. This means that we work with the sign of the samples and not with the samples themselves. In order to recover the autocorrelation of the received signal, we use least squares and a one-to-one mapping between the correlation of two real Gaussian processes and the correlation of their hard limited versions, which was derived in [26]. For complex signals, we show that the method can be applied by considering the real and imaginary parts separately. Finally, we compare the performance of both methods using computer simulations.

3.1 Least Squares Estimation of [1]

In this section, we provide an overview of the method proposed in [24] and also in [1] under the name of Alternative Time Domain method. The authors show that there exists a linear relationship between the cross-correlations of the measurements and the autocorrelation vector $\mathbf{r}_x$, and thus a least squares estimator can be used.

The first step is applying vectorization to equation (2.9), which yields

$$\operatorname{vec}(\mathbf{R}_y) = (\mathbf{C}^* \otimes \mathbf{C}) \operatorname{vec}(\mathbf{R}_x), \tag{3.1}$$

where $\operatorname{vec}(\cdot)$ is the operator that stacks all the columns of a matrix in one column vector and $\otimes$ represents the Kronecker product. Using the Hermitian Toeplitz structure of $\mathbf{R}_x$, we can write $\operatorname{vec}(\mathbf{R}_x)$ in terms of $\mathbf{r}_x$ as

$$\operatorname{vec}(\mathbf{R}_x) = \mathbf{T}\,\mathbf{r}_x, \tag{3.2}$$

where $\mathbf{T}$ is an $N^2 \times (2N-1)$ repetition matrix whose i-th row is given by the $((i-1+(N-2)\lfloor \frac{i-1}{N} \rfloor) \bmod (2N-1)+1)$-th row of the identity matrix $\mathbf{I}_{2N-1}$. The matrix $\mathbf{T}$ can also be written as a block matrix, as we consider in Chapter 5. Joining these equations we can write (3.1) as

$$\operatorname{vec}(\mathbf{R}_y) = \mathbf{\Theta}\,\mathbf{r}_x, \tag{3.3}$$

where the $M^2 \times (2N-1)$ matrix $\mathbf{\Theta}$ is given by

$$\mathbf{\Theta} = (\mathbf{C}^* \otimes \mathbf{C})\,\mathbf{T}, \tag{3.4}$$

The system of linear equations in (3.3) can be under-determined, if $M^2 < 2N-1$, or over-determined if $M^2 \geq 2N-1$, considering the determined case as part of the over-determined. As explained in [24], the under-determined case can be solved using one of the signal reconstruction techniques proposed in the compressive sensing literature, assuming that either the

power spectrum or the edge spectrum are sparse. For more details of this case, the reader can consult [27] and the references therein. However, we are more interested in the over-determined case as it allows a least squares solution, which is simpler and does not require sparsity assumptions. The best solution of (3.3) in terms of minimizing the Euclidean norm of the error is the least squares solution, which is given by

$$\hat{\mathbf{r}}_x = (\mathbf{\Theta}^H \mathbf{\Theta})^{-1} \mathbf{\Theta}^H \mathrm{vec}(\mathbf{R}_y). \tag{3.5}$$

The condition for the least squares solution to be unique is that $\mathbf{\Theta}$ has full column rank. Thus, on view of (3.4), $\mathbf{C}$ needs to be selected in a way that guarantees this property for $\mathbf{\Theta}$.

In [24], it is shown that each row of $\mathbf{\Theta}$ is the cross-correlation between two rows of $\mathbf{C}$. The cross-correlation of two deterministic sequences is defined as

$$r_{c_i,c_j}[m] = c_i[n] \star c_j^*[-n] = \sum_{n=1-N}^{0} c_i[n]\, c_j^*[n-m], \tag{3.6}$$

where $\star$ is the convolution operator. We note that the elements of $\mathbf{R}_y$ are $E\left\{y_i[k]\, y_j^*[k]\right\}$, which can be expressed as the cross-correlation sequence between $y_i[k]$ and $y_i[k]$ at lag 0. In general, the cross-correlation sequence r_{y_i,y_j} at any lag l is

$$r_{y_i,y_j}[l] = E\left\{y_i[k]\, y_j^*[k-l]\right\} = \sum_{s=1-N}^{N-1} r_{c_i,c_j}[s]\, r_x[lN-s], \tag{3.7}$$

where the last equality follows from (2.4), (3.6) and algebraic manipulations. Now, the j-th column of $\mathbf{R}_y$ can be written as

$$\mathbf{r}_{y_j}[0] = \left[r_{y_0,y_j}[0], r_{y_1,y_j}[0], \ldots, r_{y_{M-1},y_j}[0]\right]^T.$$

Then we can write

$$\mathrm{vec}(\mathbf{R}_y) = \left[\mathbf{r}_{y_0}^T[0], \mathbf{r}_{y_1}^T[0], \ldots, \mathbf{r}_{y_{M-1}}^T[0]\right]^T. \tag{3.8}$$

Taking into account (3.3) and setting $l = 0$ in (3.7), it is clear that each row of $\mathbf{\Theta}$, which we denote by $\mathbf{r}_{c_i,c_j}$, is given by

$$\mathbf{r}_{c_i,c_j} = \left[r_{c_i,c_j}[0], \ldots, r_{c_i,c_j}[1-N], r_{c_i,c_j}[N-1], \ldots, r_{c_i,c_j}[1]\right]^T.$$

With this notation, we can write

$$\mathbf{\Theta} = \left[\mathbf{r}_{c_0,c_0}, \ldots, \mathbf{r}_{c_{M-1},c_0}, \ldots, \mathbf{r}_{c_0,c_{M-1}}, \ldots, \mathbf{r}_{c_{M-1},c_{M-1}}\right]^T, \tag{3.9}$$

In the real case, it suffices to estimate $\tilde{\mathbf{r}}_x$, defined in (2.8), so a different $N^2 \times N$ matrix $\mathbf{T}_R$ is used to relate this vector with $\mathrm{vec}(\mathbf{R}_x)$. The matrix $\mathbf{T}_R$ can be written as a block matrix, where each block is an $N \times N$ matrix given by

$$\mathbf{\Xi}_0 = \mathbf{I}_N = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_N], \tag{3.10}$$

$$\mathbf{\Xi}_1 = [\mathbf{e}_2, \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{N-1}], \tag{3.11}$$

$$\mathbf{\Xi}_n = [\mathbf{e}_{n+1}, \mathbf{e}_n, \dots \mathbf{e}_2, \mathbf{e}_1, \dots, \mathbf{e}_{N-n}], \tag{3.12}$$

where $n \in \{2, \dots, N-1\}$, $\mathbf{I}_N$ is the identity matrix of order N, and $\{\mathbf{e}_j\}_{j=1}^N$ is the standard basis of $\mathbb{R}^N$. Now we can write $\mathbf{T}_R$ as

$$\mathbf{T}_R = [\mathbf{\Xi}_0, \mathbf{\Xi}_1, \dots, \mathbf{\Xi}_{N-1}]^T. \tag{3.13}$$

Then equation (3.3) changes to

$$\text{vec}(\mathbf{R}_y) = \mathbf{\Theta}_R \, \tilde{\mathbf{r}}_x, \tag{3.14}$$

where the $M^2 \times N$ matrix $\mathbf{\Theta}_R$ is given by

$$\mathbf{\Theta}_R = (\mathbf{C} \otimes \mathbf{C}) \, \mathbf{T}_R, \tag{3.15}$$

Now, using the even symmetry of $r_x[m]$, we can rewrite (3.7) as

$$r_{y_i,y_j}[0] = r_{c_i,c_j}[0] r_x[0] + \sum_{s=1}^{N-1} \left(r_{c_i,c_j}[s] + r_{c_i,c_j}[-s] \right) r_x[s]. \tag{3.16}$$

Thus, each row of $\mathbf{\Theta}_R$ is given by

$$\left[r_{c_i,c_j}[0], r_{c_i,c_j}[-1] + r_{c_i,c_j}[1], \dots, r_{c_i,c_j}[1-N] + r_{c_i,c_j}[N-1] \right].$$

Now we can see that the first columns of $\mathbf{\Theta}_R$ and $\mathbf{\Theta}$ are equal. For $m \in \{2, \dots, N\}$, the m-th column of $\mathbf{\Theta}_R$ is given by the sum of the m-th and the $(2N-m+1)$-th columns of $\mathbf{\Theta}$. Moreover, if $\mathbf{\Theta}$ has full column rank, then $\mathbf{\Theta}_R$ has full column rank as well, because each of its column vectors is the sum of two linearly independent vectors, and none of these vectors is repeated. Thus, any matrix $\mathbf{C}$ that guarantees full column rank for $\mathbf{\Theta}$ can be used for the real case as well.

Finally, [1, 24] provide conditions for $\mathbf{\Theta}$ to have full column rank in the case where multi-coset sampling is employed. As we explained in Section 2.2.2, when multi-coset sampling is employed, $\mathbf{C}$ is a binary matrix with a single one per row, and acts as a row selection matrix. As proven in [24], full column rank is achieved when every column of $\mathbf{\Theta}$ has a one, because there is necessarily a one in each row. This boils down to an $(N-1)$-length minimal sparse ruler problem.

3.2 Least Squares Estimation with Hard Clipping

In this section, we study a modification of the least squares estimation method explained in Section 3.1, where we add a hard limiter after the compressive sampling system. A limiter, or clipper, is a device that limits the amplitude of its input according to a predefined function. A hard limiter uses the sign function, that is, it returns 1 when the input is greater than 0 and -1 when it is lower than 0. This may be useful as the system would only need to store the sign of the samples instead of its value. Interestingly, there is a one to one mapping between the correlation of two real Gaussian processes and the correlation of their hard limited versions, as proven in [26]. This relationship can also be found in [20]. If the real input processes are denoted by $x(t)$ and $y(t)$, and their clipped versions by $x_c(t)$ and $y_c(t)$, their cross-correlation satisfies

$$r_{x_c,y_c}(\tau) = \frac{2}{\pi} \arcsin\left(\frac{r_{xy}(\tau)}{r_{xy}(0)} \right), \tag{3.17}$$

Thus, $r_{xy}(\tau)$ can be reconstructed from an estimate of $r_{x_c,y_c}(\tau)$, which is easy to compute as $x_c(t)$ and $y_c(t)$ are bipolar signals, and an estimate of $r_{xy}(0)$. The reconstruction is given by

$$\hat{r}_{xy}(\tau) = \hat{r}_{xy}(0)\sin\left(\frac{\pi}{2}\,\hat{r}_{x_c,y_c}(\tau)\right). \tag{3.18}$$

This method is called in the literature the polarity coincidence (PC) method, and was used in [28] to estimate the cross-correlation of the phase and quadrature components of a bandpass signal.

Now, we would like to apply the PC method to the compressive sampling framework. If we apply a hard limiter to each branch of the compressive sampling device[1], we would obtain the clipped measurement vector $\mathbf{z}[k]$, defined as

$$\mathbf{z}[k] = \text{sign}(\mathbf{y}[k]),$$

where the function sign$(\cdot)$ is applied element-wise, and we assume that $\mathbf{y}[k]$ is real. In the same fashion, we can define

$$\mathbf{s}[k] = \text{sign}(\mathbf{x}[k]).$$

However, we need to take into account that the measurement $y_i[k]$ is the inner product of $\mathbf{x}[k]$ and $\mathbf{c}_i$. Thus, the information about the sign of $x[n]$ is distorted, unless the sampling matrix $\mathbf{C}$ preserves the sign information. Mathematically, the required condition is given by

$$\text{sign}(\mathbf{y}[k]) = \text{sign}\left(\mathbf{C}\,\mathbf{x}[k]\right) = \mathbf{C}\,\text{sign}(\mathbf{x}[k]). \tag{3.19}$$

If the last equality holds, we can apply the same reasoning of Section 3.1 to the clipped measurements $\mathbf{z}[k]$ in order to recover the autocorrelation of the clipped Nyquist-rate samples $\mathbf{s}[k]$. Then, using (3.18), an estimate of $\mathbf{r}_x$ can be obtained from an estimate of the autocorrelation of $\mathbf{s}[k]$. Clearly, multicoset sampling satisfies (3.19), because it just selects certain samples from $\mathbf{x}[k]$. However, we need an estimate of $r_x[0]$, which cannot be recovered from the clipped measurements because amplitude information is lost after the hard limiter. Thus, we shall assume that the input signal is normalized, so it has $r_x[0] = 1$. This limits the application of our method, because the average power of the received signal has to be estimated separately by another means.

Besides, this method would only be able to deal with real data. In order to deal with complex data, we can estimate the autocorrelation of a complex signal by separating its real and imaginary parts, estimating their auto- and cross-correlation sequences separately, and then combining the estimates to get an estimate of the complex autocorrelation. Mathematically, the autocorrelation of a complex random sequence $f[n] = g[n] + jh[n]$ in terms of the correlations of its real and imaginary parts is given by

$$\begin{gathered} r_f[m] = E\left\{f[n]f^*[n-m]\right\} = \\ E\left\{(g[n] + j\,h[n])(g[n-m] - j\,h[n-m])\right\} = \\ E\left\{g[n]g[n-m]\right\} + E\left\{h[n]h[n-m]\right\} + j\left(E\left\{h[n]g[n-m]\right\} - E\left\{g[n]h[n-m]\right\}\right. \end{gathered} \tag{3.20}$$

We note here that for $m = 0$, the imaginary part vanishes[2]. Thus, when using equation (3.18), only $E\left\{g^2[n]\right\}$ and $E\left\{h^2[n]\right\}$ need to be estimated separately.

With these considerations, our reconstruction method can be expressed in the block diagram of Fig. 3.1. The sampling device collects K measurement vectors $\mathbf{y}[k]$. Applying

[1] In Section 2.1.1, we described the device proposed in [1] for implementing compressive sampling.

[2] Which we already knew because $r_f[m]$ has Hermitian symmetry, so $r_f[0]$ has to be real.

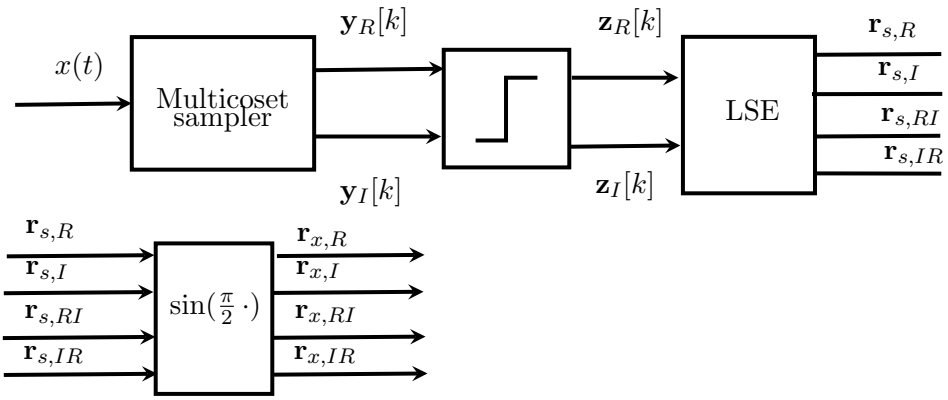


Figure 3.1: System for Least Squares estimation with a Hard Limiter

the sign function separately to their real and imaginary parts yields the vectors $\mathbf{z}_R[k]$ and $\mathbf{z}_I[k]$. We denote their autocorrelation matrices by $\mathbf{R}_{z_R}$ and $\mathbf{R}_{z_I}$, and their cross-correlation matrices by

$$\mathbf{R}_{z_{RI}} = E\left(\mathbf{z}_R[k]\,\mathbf{z}_I^T[k]\right),$$

$$\mathbf{R}_{z_{IR}} = E\left(\mathbf{z}_I[k]\,\mathbf{z}_R^T[k]\right) = \mathbf{R}_{z_{RI}}^T.$$

Now, equation (3.3) can be applied to find the $(2N-1)\times 1$ vectors $\mathbf{r}_{s,R}$, $\mathbf{r}_{s,I}$, $\mathbf{r}_{s,RI}$, $\mathbf{r}_{s,IR}$, which contain the values of the auto- and cross-correlations of the real and imaginary parts of $\mathbf{x}[k]$ after applying the hard limiter, indexed in the same way as $\mathbf{r}_x$ in (2.7). In the block diagram, these operations are performed within the block named LSE. We can write

$$\begin{aligned} \mathrm{vec}(\mathbf{R}_{z_R}) &= \mathbf{\Theta}\,\mathbf{r}_{s,R}, & \mathrm{vec}(\mathbf{R}_{z_I}) &= \mathbf{\Theta}\,\mathbf{r}_{s,I}, \\ \mathrm{vec}(\mathbf{R}_{z_{RI}}) &= \mathbf{\Theta}\,\mathbf{r}_{s,RI}, & \mathrm{vec}(\mathbf{R}_{z_{IR}}) &= \mathbf{\Theta}\,\mathbf{r}_{s,IR}. \end{aligned} \tag{3.21}$$

Once this step is completed, we can recover the correlation values of the Nyquist rate samples with equation (3.18). For the autocorrelation of the real part of $\mathbf{x}[k]$ this is written as

$$\mathbf{r}_{x,R} = \sin\left(\frac{\pi}{2}\,\mathbf{r}_{s,R}\right), \tag{3.22}$$

where the function $\sin(\cdot)$ is applied component-wise. The vectors $\mathbf{r}_{x,I}$, $\mathbf{r}_{x,RI}$ and $\mathbf{r}_{x,IR}$ are reconstructed in an analogous way.

Finally, they have to be combined using (3.20) to get the desired autocorrelation vector $\hat{\mathbf{r}}_x$, as follows,

$$\mathbf{r}_x = \mathbf{r}_{x,R} + \mathbf{r}_{x,I} + j\left(\mathbf{r}_{x,IR} - \mathbf{r}_{x,RI}\right). \tag{3.23}$$

3.3 Simulation results

In the next section, we compare the performance of our modified estimation method with the least squares estimator of [1] using computer simulations. The performance criterion is the normalized mean squared error (NMSE), which for $\hat{\boldsymbol{\phi}}_x$ is given by

$$\mathrm{NMSE}(\hat{\boldsymbol{\phi}}_x) = \frac{E\left(||\hat{\boldsymbol{\phi}}_x - \boldsymbol{\phi}_x||^2\right)}{||\boldsymbol{\phi}_x||^2}, \tag{3.24}$$

where $||\cdot||$ is the Euclidean norm. In order to calculate the MSE in practice, the expectation operator is substituted by the sample mean, resulting in the following expression:

$$\widehat{\mathrm{MSE}}(\hat{\boldsymbol{\phi}}_x) = \frac{1}{P}\sum_{i=1}^{P} ||\hat{\boldsymbol{\phi}}_x^{(i)} - \boldsymbol{\phi}_x||^2,$$

where $\hat{\boldsymbol{\phi}}_x^{(i)}$ is the i-th PSD estimate, and P is the number of estimates, which is set to 500 in our simulation. Dividing it by $||\boldsymbol{\phi}_x||^2$ we obtain the NMSE.

Besides, the estimates of the correlation matrices are used, which are the matrices $\mathbf{S}_{z_R}$, $\mathbf{S}_{z_I}$, $\mathbf{S}_{z_{RI}}$ and $\mathbf{S}_{z_{IR}}$. Considering that we collect, after the hard limiter, K measurement vectors in the matrix

$$\mathbf{Z} = [\mathbf{z}[0], \mathbf{z}[1], \ldots, \mathbf{z}[K-1]],$$

the matrix $\mathbf{S}_{Z_{RI}}$ is defined as

$$\mathbf{S}_{z_{RI}} = \frac{1}{K}\mathbf{Z}_R\mathbf{Z}_I^T,$$

where, as before, the subscripts $(\cdot)_R$ and $(\cdot)_I$ mean real and imaginary parts. The other matrices are defined in an analogous way.

In all the simulations, the compressive sampling scheme is multicoset based on a minimal sparse ruler of length 62, given by

$$\mathcal{R} = \{0, 1, 4, 5, 9, 18, 29, 35, 41, 47, 54, 57, 60, 62\}.$$

This yields a compression rate of $\frac{14}{63} = 0.22$. For higher compression rates, we select randomly rows of $\mathbf{I}_{63}$ and add them to the ones selected by the sparse ruler. The signals $x[n]$ considered are drawn from a complex circularly-symmetric Gaussian WSS random process whose PSD is obtained by filtering white Gaussian noise. Thus, the autocorrelation sequence $r_x[n]$ of the process is

$$r_x[n] = h[n] \star h^*[-n] \star \sigma_0^2\,\delta[n],$$

where $h[n]$ is the impulse response of the filter, σ_0^2 is the power of the white noise driving the filter and $\star$ represents convolution. Because our algorithms estimate the autocorrelation in the interval $[1-N, N-1]$, we will choose this support for $r_x[n]$. This is achieved using a filter $h[n]$ with length N, which is designed so that the PSD of the output process has active bands in $[-0.9\pi, -0.65\pi]$, $[0.1\pi, 0.35\pi]$ and $[0.55\pi, 0.8\pi]$, as shown in Fig. 3.2. The filter has not unit energy, so we use the value of $r_x(0)$ to normalize our PSD estimator, as required by our method.

In order to give a rough idea of what the estimates yielded by the algorithms we studied look like, one outcome of the estimators is shown in Fig. 3.3.

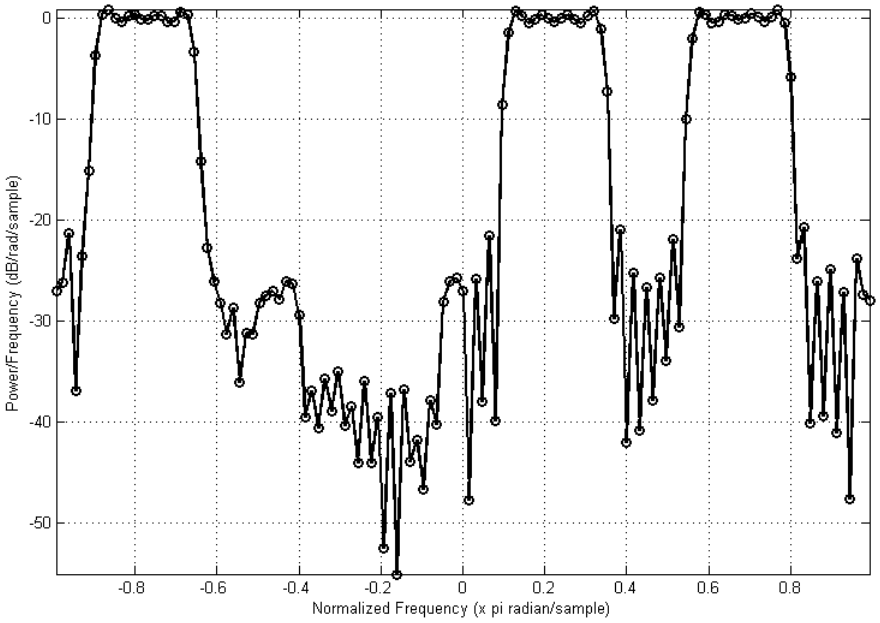


Figure 3.2: True power spectral density, $N = 63$, complex Gaussian process.

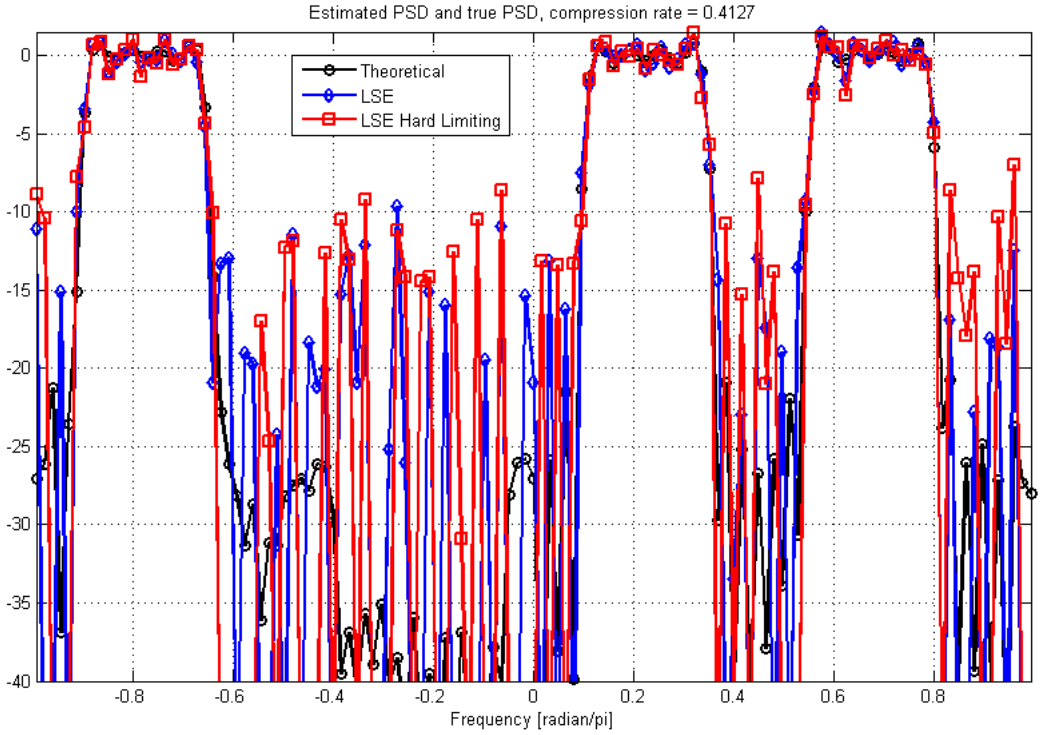


Figure 3.3: PSD estimates using the least squares methods with and without hard limiting. The compression rate is 0.413 and $K = 1024$ measurement vectors were used.

We want to study the influence of the compression rate in the NMSE, so we do not add

noise in this simulation. The number of measurement vectors collected is $K = 2048$. The results can be seen in Fig. 3.4, where we can see that the curves have a constant difference between them, which means that our method introduces an additional estimation error.

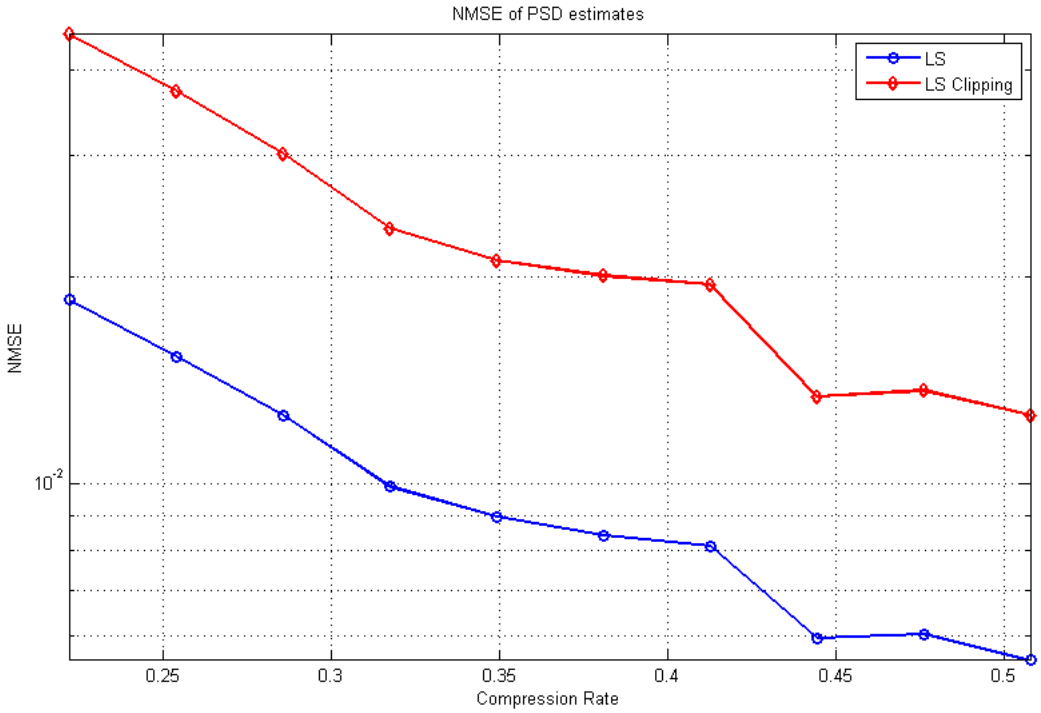


Figure 3.4: NMSE vs Compression Rate for Least Squares and Least Squares with a Hard Limiter.

3.4 Conclusions

In this chapter, we studied a modification of the least squares PSD estimator proposed in [1]. We add a hard limiter, which retains only the sign of the compressive measurements. This may be useful as the system that processes the samples only needs to store their sign. However, the average power of the received signal must be estimated by other means, because after the hard limiter the amplitude information is lost. By using a one-to-one mapping between the correlation of two real Gaussian processes and their versions after hard limitation, we can recover the autocorrelation of the received signal. For complex signals, the real and imaginary parts need to be considered separately, and the autocorrelation can be reconstructed from their auto- and cross-correlations. Finally, we compared the NMSE of our method and the least squares method of [1], and we saw that using the hard limiter introduces an additional error.

Maximum Likelihood Estimation

4

In this chapter, we investigate maximum likelihood estimation (MLE) of the autocorrelation sequence $r_x[m]$ using compressive sampling, when the signal $x(t)$ is assumed to come from a Gaussian process. This assumption is common in many estimation problems due to its mathematical tractability and the fact that several physical problems can be well approximated by the Gaussian distribution, owing to a group of mathematical theorems that fall under the name of Central Limit Theorems.

The problem of estimating $\mathbf{r}_x$ is a particular case of the general problem of estimating the covariance matrix of a random vector, when it is known that the matrix has a Toeplitz structure due to the wide-sense stationarity of $x[n]$. In the Gaussian case and without compression, this problem has been analyzed in [29], where the authors propose an algorithm to find the maximum likelihood estimator when independent outcomes of the random vector are available. More recently, [30] has studied maximum likelihood, using compressive sampling, to estimate the power of each band in a multiband signal, provided that the power spectral densities of the individual bands are known, or, equivalently, the associated covariance matrices. In our work, we study maximum likelihood estimation of the covariance matrix using compressive sampling without any prior information. Thus, our approach falls under the non-parametric methods for spectral estimation.

4.1 Statistical assumptions

In order to study maximum likelihood estimation, we must make an assumption about the statistical distribution of $x[n]$, which is assumed to be a zero-mean real Gaussian process. At the end of this section, we consider the case where $x[n]$ is complex valued. If $x[n]$ is a Gaussian process, every $\mathbf{x}[k]$ will be a random vector with multivariate Gaussian distribution. Its probability density function (pdf) is given by

$$f(\mathbf{x}[k]|\mathbf{R}_x) = \frac{1}{(2\pi)^{N/2}|\mathbf{R}_x|^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{x}^T[k]\mathbf{R}_x^{-1}\mathbf{x}[k]\right\} = \frac{1}{(2\pi)^{N/2}|\mathbf{R}_x|^{1/2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\mathbf{R}_x^{-1}\mathbf{x}[k]\mathbf{x}^T[k]\right)\right\}. \tag{4.1}$$

In order for the pdf to exist, $\mathbf{R}_x$ must be positive definite, which we will assume from now. Because $x[n]$ is a real process, $\mathbf{R}_x$ is symmetric instead of Hermitian, and can be written as $\mathbf{R}_x = E\left(\mathbf{x}[k]\mathbf{x}^T[k]\right)$. Owing to its Toeplitz structure, every element in the m-th lower diagonal is equal to $r_x[m]$ and, because of the symmetry, all the elements in the m-th upper diagonal are also equal to $r_x[m]$. We can use this property to express $\mathbf{R}_x$ as

$$\mathbf{R}_x = \sum_{m=0}^{N-1} r_x[m]\mathbf{B}_m, \tag{4.2}$$

where the matrices $\mathbf{B}_m$ form a basis for the subspace of square symmetric Toeplitz matrices of order N. These matrices have all zeros but in the m-th lower and upper diagonals, where they have ones. As an example, we show $\mathbf{B}_1$, which is given by

$$\mathbf{B}_1 = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ldots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Now we can characterize the statistics of $\mathbf{y}[k]$. Due to the linear relationship in (2.2) and the assumption that $\mathbf{x}[k]$ is Gaussian, it follows that $\mathbf{y}[k]$ is also an $M \times 1$ Gaussian random vector with zero mean and covariance matrix $\mathbf{R}_y = \mathbf{C}\mathbf{R}_x\mathbf{C}^T$. Using (4.2), we can express $\mathbf{R}_y$ as linear combination of matrices as follows

$$\mathbf{R}_y = \mathbf{C}\left(\sum_{m=0}^{N-1} r_x[m]\mathbf{B}_m\right)\mathbf{C}^T = \sum_{m=0}^{N-1} r_x[m]\,\mathbf{C}\mathbf{B}_m\mathbf{C}^T. \tag{4.3}$$

In order for the parameters $\{r_x[m]\}_{m=0}^{N-1}$ to remain identifiable, the set of matrices $\{\mathbf{C}\mathbf{B}_m\mathbf{C}^T\}_{m=0}^{N-1}$ must still be linearly independent. This restricts the choice of the sampling matrix $\mathbf{C}$, as we explained in Section 2.2.2.

Here, the likelihood function is the pdf of $\mathbf{y}[k]$ given particular values of the parameters we want to estimate. In this case, our parameter vector is $\tilde{\mathbf{r}}_x$. We can write the likelihood function as

$$f(\mathbf{y}[k]\,|\,\tilde{\mathbf{r}}_x) = \frac{1}{(2\pi)^{M/2}|\mathbf{R}_y|^{1/2}}\exp\left\{-\frac{1}{2}\mathrm{tr}\left(\mathbf{R}_y^{-1}\mathbf{y}[k]\mathbf{y}^T[k]\right)\right\}. \tag{4.4}$$

The vector $\tilde{\mathbf{r}}_x$ that maximizes this function for a given set of measurements $\mathbf{y}[k]$ is the maximum likelihood estimator. In view of (4.3), this is equivalent to find the values of $r_x[m]$ that make the best agreement between the matrix $\mathbf{R}_y$ and the observed values $\mathbf{y}[k]$. However, we note that so far we have considered the statistical properties of each measurement vector $\mathbf{y}[k]$ separately. As our sampling device operates in a periodic fashion, we can collect K measurement vectors in the $M \times K$ matrix $\mathbf{Y} = [\mathbf{y}[0], \mathbf{y}[1], \ldots, \mathbf{y}[K-1]]$.

Within the most common approach to obtain the maximum likelihood estimator of a covariance matrix, all the measurement vectors are assumed to be mutually independent, regardless of whether there is compression, like in our case, or not, like in [29]. However, the vectors $\mathbf{y}[k]$ come from one outcome of the random process $x[n]$, and are clearly not independent because they are correlated. If they were independent, the joint pdf of all the vectors, which we denote by $f\left(\{\mathbf{y}[k]\}_{k=0}^{K-1}\,|\,\tilde{\mathbf{r}}_x\right)$, would be the product of the individual pdfs

of each vector. We could write then

$$f\left(\{\mathbf{y}[k]\}_{k=0}^{K-1} \,|\, \tilde{\mathbf{r}}_x\right) = \prod_{k=0}^{K-1} f\left(\mathbf{y}[k] \,|\, \tilde{\mathbf{r}}_x\right) =$$

$$\prod_{k=0}^{K-1} \frac{1}{(2\pi)^{M/2}|\mathbf{R}_y|^{1/2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\mathbf{R}_y^{-1}\mathbf{y}[k]\mathbf{y}^T[k]\right)\right\} =$$

$$\frac{1}{(2\pi)^{KM/2}|\mathbf{R}_y|^{K/2}} \exp\left\{-\frac{1}{2}\sum_{k=0}^{K-1}\mathrm{tr}\left(\mathbf{R}_y^{-1}\mathbf{y}[k]\mathbf{y}^T[k]\right)\right\}. \tag{4.5}$$

By using the properties of the trace, we finally arrive to

$$f\left(\{\mathbf{y}[k]\}_{k=0}^{K-1} \,|\, \tilde{\mathbf{r}}_x\right) = \frac{1}{(2\pi)^{KM/2}|\mathbf{R}_y|^{K/2}} \exp\left\{-\frac{K}{2}\mathrm{tr}\left(\mathbf{R}_y^{-1}\mathbf{S}_y\right)\right\}, \tag{4.6}$$

where the matrix $\mathbf{S}_y$, commonly called the sample covariance matrix, is defined as

$$\mathbf{S}_y = \frac{1}{K}\sum_{k=0}^{K-1}\mathbf{y}[k]\mathbf{y}^T[k] = \frac{1}{K}\mathbf{Y}\mathbf{Y}^T, \tag{4.7}$$

This matrix performs averaging of the measurement vectors over time, and it has the desirable property of being a consistent estimator of $\mathbf{R}_y$. A consistent estimator of a parameter converges in some sense to the true value of the parameter as the sample size increases. This does not happen with $\mathbf{y}[k]\mathbf{y}^T[k]$ because, as the sample size increases, $\mathbf{y}[k]\mathbf{y}^T[k]$ clearly does not converge to $\mathbf{R}_y$ in any sense.

Now back to our case, the measurement vectors $\mathbf{y}[k]$ are not independent, so we cannot use equation (4.6), and the consistency of $\mathbf{S}_y$ might not hold. However, if we assume $\{y_i[k]\}_{i=0}^{M-1}$ to be ergodic, then this time average converges to the ensemble average $\mathbf{R}_y$ as $K \to \infty$, and this means that $\mathbf{S}_y$ is a consistent estimator of $\mathbf{R}_y$. The ergodicity of $\{y_i[k]\}_{i=0}^{M-1}$ implies that $x[n]$ has to be ergodic as well, which is reasonable in the context of the estimation of second-order statistics from signal measurements. For more details on ergodicity, we refer to [20].

Taking these facts into account, we can first derive the maximum likelihood estimator considering a single measurement vector $\mathbf{y}[k]$, and then, using the consistency of $\mathbf{S}_y$, we can obtain asymptotic solutions that reduce the computational complexity without critically decreasing the accuracy.

4.2 Derivation of the Maximum Likelihood Estimator

In order to maximize $f(\mathbf{y}[k] \,|\, \tilde{\mathbf{r}}_x)$, we need to set its gradient to zero. This task is easier if we take the natural logarithm, because it is a monotonically increasing function and thus the maximum will be attained for the same argument. Applying the logarithm yields

$$\ln f(\mathbf{y}[k] \,|\, \tilde{\mathbf{r}}_x) = -\frac{M}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{R}_y| - \frac{1}{2}\mathrm{tr}\left(\mathbf{R}_y^{-1}\mathbf{y}[k]\mathbf{y}^T[k]\right). \tag{4.8}$$

We first need to calculate the derivative of $\ln f(\mathbf{y}[k] \,|\, \tilde{\mathbf{r}}_x)$ with respect to each $r_x[n]$,

$$\frac{\partial \ln f(\mathbf{y}[k] \,|\, \tilde{\mathbf{r}}_x)}{\partial r_x[n]} = -\frac{1}{2}\frac{\partial \ln|\mathbf{R}_y|}{\partial r_x[n]} - \frac{1}{2}\frac{\partial}{\partial r_x[n]}\left\{\mathrm{tr}\left(\mathbf{R}_y^{-1}\mathbf{y}[k]\mathbf{y}^T[k]\right)\right\}.$$

We now have

$$\frac{\partial \ln|\mathbf{R}_y|}{\partial r_x[n]} = \frac{1}{|\mathbf{R}_y|}\frac{\partial |\mathbf{R}_y|}{\partial r_x[n]} = \frac{1}{|\mathbf{R}_y|}\mathrm{tr}\left\{\mathrm{Adj}\left(\mathbf{R}_y\right)\frac{\partial \mathbf{R}_y}{\partial r_x[n]}\right\} = \mathrm{tr}\left\{\mathbf{R}_y^{-1}\mathbf{C}\mathbf{B}_n\mathbf{C}^T\right\},$$

where we have used the fact that $\frac{\partial \mathbf{R}_y}{\partial r_x[n]} = \mathbf{C}\mathbf{B}_n\mathbf{C}^T$ as seen from (4.3). The second term is given by

$$\frac{\partial}{\partial r_x[n]}\left\{\mathrm{tr}\left(\mathbf{R}_y^{-1}\mathbf{y}[k]\mathbf{y}^T[k]\right)\right\} = \mathrm{tr}\left\{\frac{\partial}{\partial r_x[n]}\left(\mathbf{R}_y^{-1}\mathbf{y}[k]\mathbf{y}^T[k]\right)\right\} =$$
$$-\mathrm{tr}\left\{\mathbf{R}_y^{-1}\frac{\partial \mathbf{R}_y}{\partial r_x[n]}\mathbf{R}_y^{-1}\mathbf{y}[k]\mathbf{y}^T[k]\right\} = -\mathrm{tr}\left\{\mathbf{R}_y^{-1}\mathbf{C}\mathbf{B}_n\mathbf{C}^T\mathbf{R}_y^{-1}\mathbf{y}[k]\mathbf{y}^T[k]\right\},$$

where we have used $\frac{\partial \mathbf{A}^{-1}}{\partial t} = -\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial t}\mathbf{A}^{-1}$. The maximum will be found by setting each derivative to zero and solving the resulting system of equations. Using the derived results, we can write

$$\frac{\partial \ln f(\mathbf{y}[k]\,|\,\tilde{\mathbf{r}}_x)}{\partial r_x[n]} = -\frac{1}{2}\mathrm{tr}\left\{\mathbf{R}_y^{-1}\mathbf{C}\mathbf{B}_n\mathbf{C}^T\right\} + \frac{1}{2}\mathrm{tr}\left\{\mathbf{R}_y^{-1}\mathbf{C}\mathbf{B}_n\mathbf{C}^T\mathbf{R}_y^{-1}\mathbf{y}[k]\mathbf{y}^T[k]\right\}.$$

Setting this derivative to zero yields

$$\mathrm{tr}\left\{\mathbf{R}_y^{-1}\mathbf{C}\mathbf{B}_n\mathbf{C}^T\mathbf{R}_y^{-1}\mathbf{R}_y\right\} = \mathrm{tr}\left\{\mathbf{R}_y^{-1}\mathbf{C}\mathbf{B}_n\mathbf{C}^T\mathbf{R}_y^{-1}\mathbf{y}[k]\mathbf{y}^T[k]\right\}, \tag{4.9}$$

where we have introduced the factor $\mathbf{R}_y^{-1}\mathbf{R}_y$ in the first term, which does not affect the result. Expanding $\mathbf{R}_y$ by using (4.3), we arrive at

$$\sum_{m=0}^{N-1} r_x[m]\,\mathrm{tr}\left\{\mathbf{R}_y^{-1}\mathbf{C}\mathbf{B}_n\mathbf{C}^T\mathbf{R}_y^{-1}\mathbf{C}\mathbf{B}_m\mathbf{C}^T\right\} = \mathrm{tr}\left\{\mathbf{R}_y^{-1}\mathbf{C}\mathbf{B}_n\mathbf{C}^T\mathbf{R}_y^{-1}\mathbf{y}[k]\mathbf{y}^T[k]\right\}, \tag{4.10}$$

which has to hold for $\forall n \in \{0, 1, \ldots, N-1\}$. Solving this linear system of N equations with N unknowns would give our desired estimator $\hat{\tilde{\mathbf{r}}}_x$. However, there is no analytical solution known to this problem, so we need to search for numerical solutions, that generally trade off simplicity for accuracy.

4.3 Asymptotic Solutions to the ML Equations

Solving (4.10) numerically can lead to high computational complexity. As we stated at the end of Section 4.1, by using the consistency of $\mathbf{S}_y$ we can obtain solutions that converge asymptotically to the true ML solution. Calculating these solutions can reduce the computational complexity, at the expense of increasing the sensing time in order to make the sample size K large enough.

First, we notice that equation (4.10) includes the term $\mathbf{y}[k]\mathbf{y}^T[k]$, which is not a consistent estimator of $\mathbf{R}_y$. It is easy to see that if we replace $\mathbf{y}[k]\mathbf{y}^T[k]$ with a consistent estimator of $\mathbf{R}_y$, then (4.9) would hold when $K \to \infty$. Using this property, we can solve a modified version of (4.10), employing $\mathbf{S}_y$ instead of $\mathbf{y}[k]\mathbf{y}^T[k]$, and obtain a good approximation of the true solution. We can formulate two strategies that will approximate the ML solution in the case where K is large.

4.3.1 Asymptotic Maximum Likelihood

A very simple solution is substituting $\mathbf{R}_y^{-1}$ with $\mathbf{S}_y^{-1}$ in (4.10), which would be a good approximation as we know $\mathbf{S}_y \approx \mathbf{R}_y$ for large K. The linear system of equations to solve changes to

$$\sum_{m=0}^{N-1} \hat{r}_x[m] \operatorname{tr}\left\{\mathbf{S}_y^{-1}\mathbf{C}\mathbf{B}_n\mathbf{C}^T\mathbf{S}_y^{-1}\mathbf{C}\mathbf{B}_m\mathbf{C}^T\right\} = \operatorname{tr}\left\{\mathbf{S}_y^{-1}\mathbf{C}\mathbf{B}_n\mathbf{C}^T\right\}, \tag{4.11}$$

where $n \in \{0, 1, \ldots, N-1\}$. This requires the matrix $\mathbf{S}_y$ to be invertible, or equivalently, that it has full rank. For any real matrix $\mathbf{A}$ it holds that $\operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{A}\mathbf{A}^T)$, and thus requiring $\operatorname{rank}(\mathbf{S}_y) = M$ is equivalent to require that $\operatorname{rank}(\mathbf{Y}) = M$. A necessary but not sufficient condition is $K \geq M$, which is satisfied as we have assumed that K is large. Besides, due to the assumption of $x[n]$ being a Gaussian process, the columns of $\mathbf{Y}$ come from a continuous non-degenerate probability distribution, so they are linearly independent with probability one. This guarantees that $\mathbf{S}_y$ is full rank.

4.3.2 Simplified Inverse Iteration Algorithm

In order to find a better approximation, we can iterate over (4.10) by remembering that we use $\mathbf{S}_y$ instead of $\mathbf{y}[k]\mathbf{y}^T[k]$. This follows the work in [30], where they also choose the name of this algorithm, because it is a simplified version of the algorithm presented in [29].

The system of equations to solve in every iteration is

$$\sum_{m=0}^{N-1} \hat{r}_x^{(k+1)}[m] \operatorname{tr}\left\{(\hat{\mathbf{R}}_y^{(k)})^{-1}\mathbf{C}\mathbf{B}_n\mathbf{C}^T(\hat{\mathbf{R}}_y^{(k)})^{-1}\mathbf{C}\mathbf{B}_m\mathbf{C}^T\right\} = \operatorname{tr}\left\{(\hat{\mathbf{R}}_y^{(k)})^{-1}\mathbf{C}\mathbf{B}_n\mathbf{C}^T(\hat{\mathbf{R}}_y^{(k)})^{-1}\mathbf{S}_y\right\}, \ \forall n, \tag{4.12}$$

where $\hat{\mathbf{R}}_y^{(k)}, \hat{r}_x^{(k)}[m]$ are the estimates at iteration k of $\mathbf{R}_y$ and $r_x[m]$, respectively. For initialization, we set $\hat{\mathbf{R}}_y^{(0)} = \mathbf{S}_y$, and for $k \geq 1$ the relationship between $\hat{\mathbf{R}}_y^{(k)}$ and $\hat{r}_x^{(k)}[m]$ is given by equation (4.3). Note that $\hat{\mathbf{R}}_y^{(k)}$ has to be invertible in every iteration, and also that the initialization step is equal to the previous strategy we described, called Asymptotic Maximum Likelihood. The number of iterations is not fixed and can be performed indefinitely, as long as the invertibility condition is satisfied. A study of the convergence of this algorithm is not provided in this work. The computational complexity of SIIA is much greater than AML

4.4 Complex data

In order to deal with complex data, we have to modify our statistical asumptions about the random process $x[n]$. For a complex Gaussian random process or vector to have its distribution completely specified, we need to take into account the possible dependency between the real and imaginary parts of each element. In particular, if the real and imaginary parts of each element are independent and have the same variance, the complex random vector $\mathbf{x}[k]$ has a

circularly-symmetric complex Gaussian distribution [31], with probability density function

$$f\left(\mathbf{x}[k]\,|\,\mathbf{R}_x\right) = \frac{1}{\pi^N|\mathbf{R}_x|}\exp\left\{-\mathbf{x}^H[k]\mathbf{R}_x^{-1}\mathbf{x}[k]\right\} = \frac{1}{\pi^N|\mathbf{R}_x|}\exp\left\{-\mathrm{tr}\left(\mathbf{R}_x^{-1}\mathbf{x}[k]\mathbf{x}^H[k]\right)\right\}. \tag{4.13}$$

Note that the pdf is very similar to (4.1). However, as we discussed in Section 2.2.1, the matrix $\mathbf{R}_x$ is Hermitian Toeplitz, so we need to consider the subspace of square Hermitian Toeplitz matrices of order N. A basis for this subspace is given by the $N \times N$ matrices $\{\mathbf{B}_m\}_{m=0}^{N-1}$ that we used in (4.2), and the $N \times N$ matrices $\{\tilde{\mathbf{B}}_m\}_{m=1}^{N-1}$. These matrices have all zeros but in the m-th lower and upper diagonals, where they have the values of $-j$ and j respectively. As an example, we show $\tilde{\mathbf{B}}_1$, which is given by

$$\tilde{\mathbf{B}}_1 = \begin{bmatrix} 0 & j & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ -j & 0 & j & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & -j & 0 & j & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ldots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -j & 0 & j & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & -j & 0 & j \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & -j & 0 \end{bmatrix}.$$

Then we can write (4.2) for the complex case as

$$\mathbf{R}_x = \sum_{m=0}^{N-1}\alpha_m\mathbf{B}_m + \sum_{m=1}^{N-1}\beta_m\tilde{\mathbf{B}}_m, \tag{4.14}$$

where $r_x[m] = \alpha_m - j\beta_m$ and $\alpha_m, \beta_m \in \mathbb{R}\,\forall m \in \{0, 1, \ldots N-1\}$. Due to the Hermitian symmetry, $r_x[0]$ has to be real and thus $\beta_0 = 0$. Finding the MLE for this case has a very similar derivation to the one in Section 4.2. Thus, for this case we can use both AML and SIIA to estimate the power spectral density.

4.5 Simulation results

In this section, we illustrate the performance of the algorithms with numerical examples. We compare AML and SIIA with the method based on least squares proposed in [1] under the name of Alternative Time Domain approach. To avoid a high computational burden, we set the number of iterations of SIIA to 10. The performance criterion is the normalized mean squared error (NMSE), which we defined in (3.24). In this case, it is difficult to obtain it theoretically, so we estimate it with computer simulations. As we explained in Section 3.3, in order to calculate the MSE in practice, the expectation operator is substituted by the sample mean, resulting in the following expression:

$$\widehat{\mathrm{MSE}}(\hat{\boldsymbol{\phi}}_x) = \frac{1}{P}\sum_{i=1}^{P}||\hat{\boldsymbol{\phi}}_x^{(i)} - \boldsymbol{\phi}_x||^2,$$

where $\hat{\boldsymbol{\phi}}_x^{(i)}$ is the i-th PSD estimate, and P is the number of estimates, which is set to 500 in this simulation study. Dividing it by $||\boldsymbol{\phi}_x||^2$ leads to the NMSE.

In all the simulations, the employed compressive sampling scheme is multicoset sampling based on a minimal sparse ruler of length 62, given by

$$\mathcal{R} = \{0, 1, 4, 5, 9, 18, 29, 35, 41, 47, 54, 57, 60, 62\}.$$

This yields a compression rate of $\frac{14}{63} = 0.22$. For higher compression rates, we randomly select rows of $\mathbf{I}_{63}$ and add them to the ones selected by the sparse ruler, until the appropiate compression rate is achieved. The signals $x[n]$ considered are drawn from a real Gaussian WSS random process whose PSD is obtained by filtering white Gaussian noise. Thus, the autocorrelation sequence $r_x[n]$ of the process is

$$r_x[n] = h[n] \star h[-n] \star \sigma_0^2 \, \delta[n],$$

where $h[n]$ is the impulse response of the filter, σ_0^2 is the power of the white noise driving the filter and $\star$ represents convolution. Because our algorithms estimate the autocorrelation in the interval $[1 - N, N - 1]$, we will choose this support for $r_x[n]$. This is achieved by using a filter $h[n]$ with length N, which is designed so that the PSD of the output process has active bands in $[0.1\pi, 0.2\pi]$, $[0.45\pi, 0.65\pi]$ and $[0.8\pi, 0.9\pi]$, as shown in Fig. 4.1.

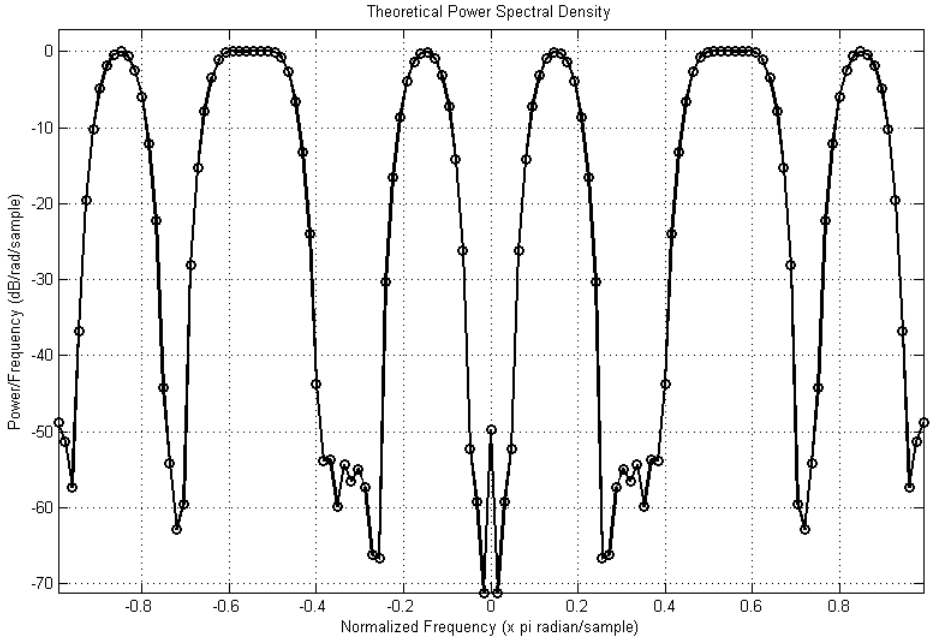


Figure 4.1: True power spectral density, $N = 63$, real Gaussian process.

In order to give a rough idea of what the estimates yielded by the algorithms we studied look like, one outcome of the estimators is shown in Fig. 4.2.

The influence of three factors is considered in the performance of the algorithms, i.e., the compression rate, the signal-to-noise ratio (SNR) and the sensing time. As an attempt to model different sensing times, we generated two different numbers K of vectors $\mathbf{x}[k]$, namely 1024 and 3663. Having K vectors corresponds to KN Nyquist rate samples. In order to evaluate the effect of compression rate and SNR, two separate simulations were conducted. The first computes the NMSE in the noiseless case for several compression rates. The resulting

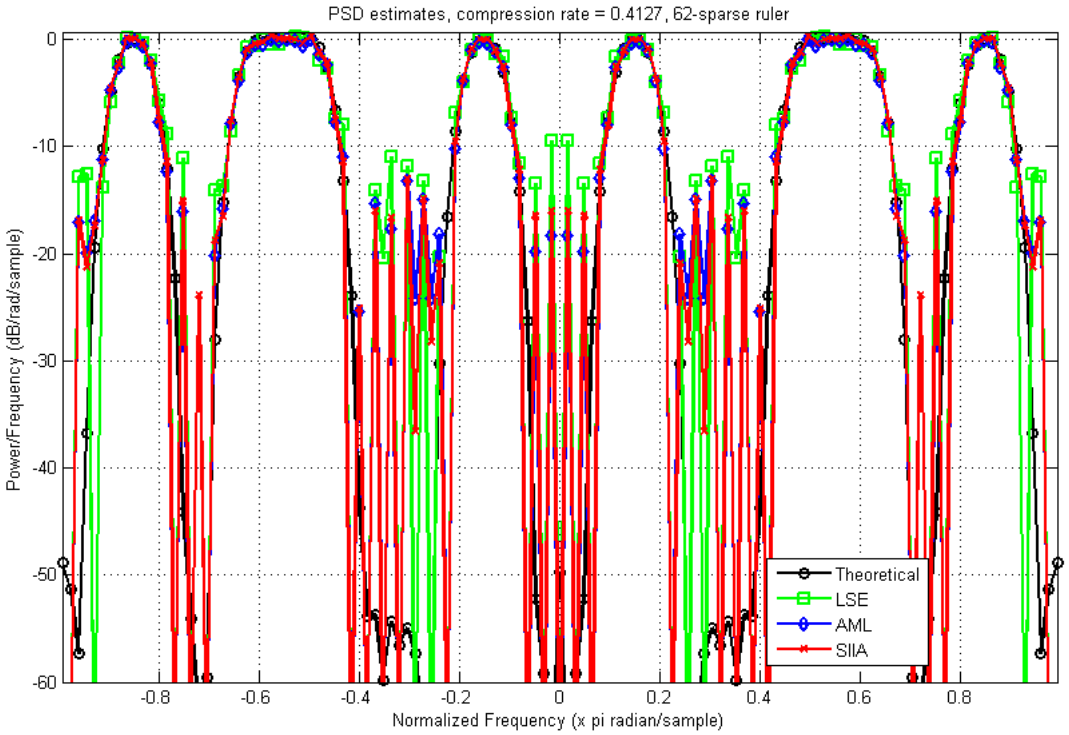


Figure 4.2: PSD estimates using AML, SIIA and LSE. The compression rate is 0.413 and $K = 1024$ measurement vectors were used.

graph can be seen in Fig. 4.3. The second fixes the compression rate to 0.349 and adds white Gaussian noise of different powers, so that the SNR changes in the range of 0 to 30 dB, as can be seen in Fig. 4.4.

As we can see, all three algorithms perform acceptably in terms of MSE. Both AML and SIIA perform better than Least Squares, which can be explained by the fact that our data come from a Gaussian population, and that the maximum likelihood estimation is asymptotically efficient, as shown in [32]. In Fig. 4.3, we can see that the performance of AML and SIIA is almost equal for a strong compression. The point where SIIA starts to perform better than AML is between 0.32 and 0.35. The difference in performance is less pronounced as the number of measurement vectors available K increases. In Fig. 4.4 we can see that, for moderately high SNRs, the performance of AML is the best of the three.

4.6 Conclusions

We studied Maximum Likelihood Estimation of the PSD using compressive sampling, where the data is assumed to be Gaussian. The true solution cannot be found analytically, and numerical methods may be very computationally intense. Thus, we turned to two algorithms that yield asymptotically the ML solution in order to reduce the computational cost. We evaluated their performance using computer simulations to obtain estimates of their mean squared error, and compared them to another existing method in the literature based on least squares [1].

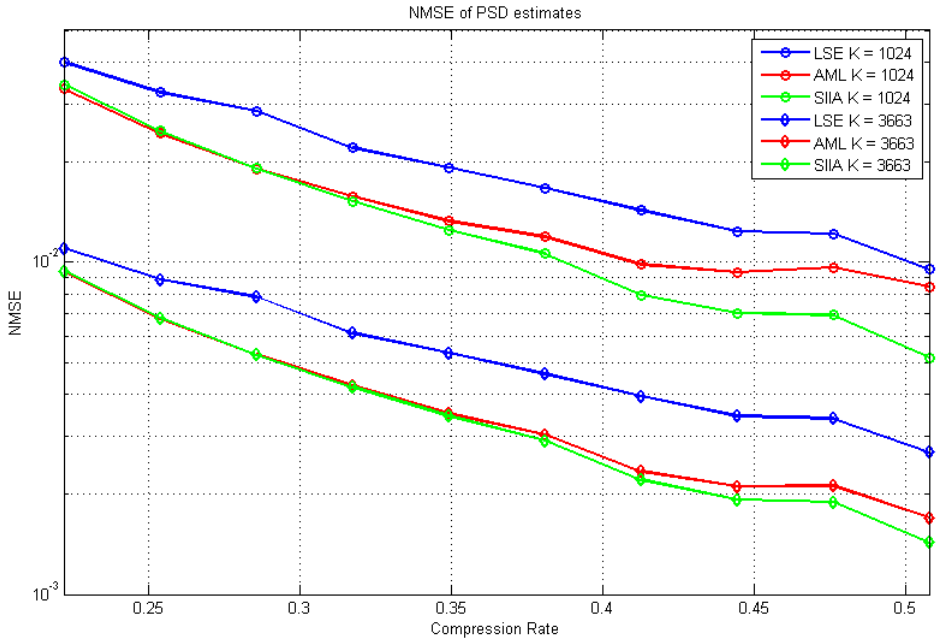


Figure 4.3: NMSE vs Compression Rate for AML, SIIA and LSE

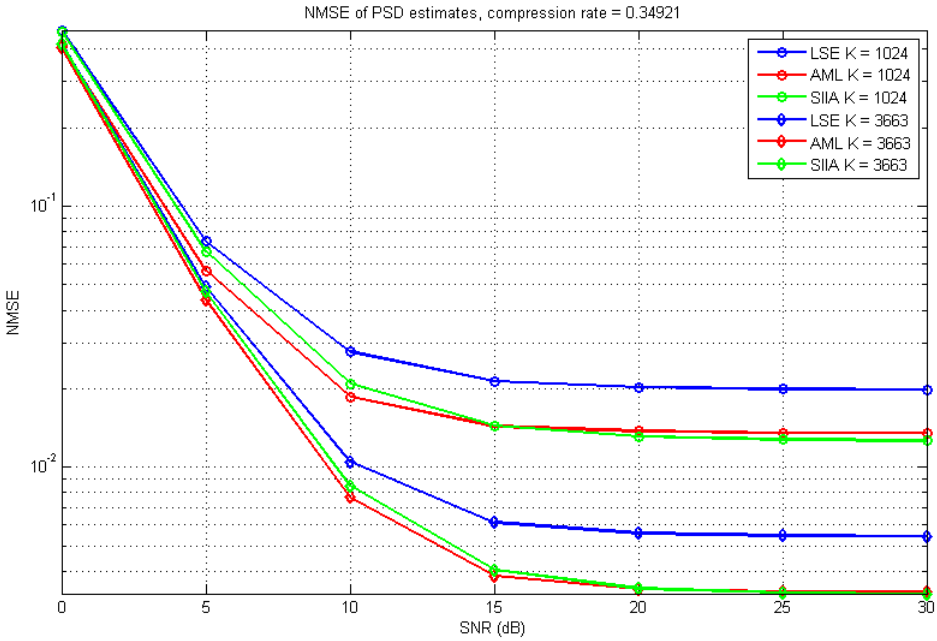


Figure 4.4: NMSE vs SNR for AML, SIIA and LSE

5 Sampling Matrix Design

In this chapter, we study the possibility of constructing new sampling matrices $\mathbf{C}$ that guarantee the statistical identifiability of the autocorrelation vector $\mathbf{r}_x$. Our criteria will be that the matrix $\mathbf{\Theta}$, which is used in the least squares estimator proposed in 3.1, has full column rank.

5.1 Conditions for Full Column Rank

We start by rewriting equation (3.9) again, where $\mathbf{\Theta}$ was expressed in terms of its rows as follows

$$\mathbf{\Theta} = [\mathbf{r}_{c_0,c_0}, \dots, \mathbf{r}_{c_{M-1},c_0}, \dots, \mathbf{r}_{c_0,c_{M-1}}, \dots, \mathbf{r}_{c_{M-1},c_{M-1}}]^T,$$

where $\mathbf{r}_{c_i,c_j} = \left[r_{c_i,c_j}[0], \dots, r_{c_i,c_j}[1-N], r_{c_i,c_j}[N-1], \dots, r_{c_i,c_j}[1]\right]^T$. Taking this into account, it seems easier to work with the rows in order to obtain conditions for the full column rank of $\mathbf{\Theta}$. As the column rank of a matrix is equal to the row rank, we need to guarantee that $2N-1$ linearly independent rows exist. In other words, a basis of $\mathbb{C}^{2N-1}$ must be obtained from the M^2 possible cross-correlations of the M sequences $\{c_i[n]\}_{i=0}^{M-1}$ that constitute the rows of the sampling matrix $\mathbf{C}$.

As we are working with vectors, it is convenient to express (3.6) in matrix form,

$$\begin{bmatrix} c_j^*[0] & c_j^*[-1] & c_j^*[-2] & \cdots & c_j^*[2-N] & c_j^*[1-N] \\ 0 & c_j^*[0] & c_j^*[-1] & \cdots & c_j^*[3-N] & c_j^*[2-N] \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & c_j^*[0] \\ c_j^*[1-N] & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ c_j^*[-1] & c_j^*[-2] & c_j^*[-3] & \cdots & c_j^*[1-N] & 0 \end{bmatrix} \begin{bmatrix} c_i[0] \\ c_i[-1] \\ c_i[-2] \\ \vdots \\ c_i[2-N] \\ c_i[1-N] \end{bmatrix} = \begin{bmatrix} r_{c_i,c_j}[0] \\ r_{c_i,c_j}[-1] \\ \vdots \\ r_{c_i,c_j}[1-N] \\ r_{c_i,c_j}[N-1] \\ \vdots \\ r_{c_i,c_j}[1] \end{bmatrix}.$$

More compactly we will write

$$\mathbf{T}_{c_j}^* \, \mathbf{c}_i = \mathbf{r}_{c_i,c_j}. \tag{5.1}$$

The matrix $\mathbf{T}_{c_j}^*$ is a $(2N-1) \times N$ Toeplitz matrix. Multiplying any $N \times 1$ vector $\mathbf{c}_i$ with this matrix yields the cross-correlation of the corresponding sequences $c_i[n]$ and $c_j[n]$. From its structure it can be seen that it has rank N as long as $c_j[n] \neq 0$ for some $n \in \{0, -1, \dots, 1-N\}$. In other words, it has rank N except when it is the zero matrix. Equation (5.1) implies that the vector $\mathbf{r}_{c_i,c_j}$ lies in the column space $\mathrm{Im}(\mathbf{T}_{c_j}^*)$ of this matrix. One basis for $\mathrm{Im}(\mathbf{T}_{c_j}^*)$ can be found using a singular value decomposition, yielding

$$\mathbf{T}_{c_j}^* = \mathbf{U}_j \mathbf{\Sigma}_j \mathbf{V}_j^H = \begin{bmatrix} \hat{\mathbf{U}}_j & \hat{\mathbf{U}}_j^{\perp} \end{bmatrix} \mathbf{\Sigma}_j \mathbf{V}_j^H, \tag{5.2}$$

where we have assumed that $\mathbf{T}_{c_j}^*$ has rank N. The columns of the $(2N-1) \times N$ matrix $\hat{\mathbf{U}}_j$ form an orthonormal basis for $\mathrm{Im}(\mathbf{T}_{c_j}^*)$, and $\hat{\mathbf{U}}_j^{\perp}$ is a $(2N-1) \times (N-1)$ matrix whose columns

form an orthonormal basis for $\mathrm{Ker}(\mathbf{T}_{c_j}^H)$, which is the orthogonal complement of the column space. Multiplying by $\mathbf{V}_j$ both sides of (5.2) allows us to write

$$\mathbf{T}_{c_j}^* \mathbf{V}_j = \mathbf{U}_j \mathbf{\Sigma}_j. \tag{5.3}$$

This means that the cross-correlation of a right singular vector with $\mathbf{c}_j$ gives its corresponding left singular vector scaled by its singular value.

In our problem, we have to select M vectors $\{\mathbf{c}_0, \mathbf{c}_1, \ldots, \mathbf{c}_{M-1}\}$ of dimension N. Their corresponding matrices $\mathbf{T}_{c_0}^*, \mathbf{T}_{c_1}^*, \ldots, \mathbf{T}_{c_{M-1}}^*$ have associated M subspaces of $\mathbb{C}^{2N-1}$, namely their column spaces, each of dimension N. Using (3.9) and (5.1), the M^2 rows of $\mathbf{\Theta}$ can be divided into M groups of M vectors, where each group belongs to one of the subspaces $\mathrm{Im}(\mathbf{T}_{c_j}^*)$. In order to obtain $2N-1$ linearly independent vectors, a necessary condition is that the union of all subspaces spans the complete vector space, that is,

$$\mathrm{Span}\left(\bigcup_{j=0}^{M-1} \mathrm{Im}(\mathbf{T}_{c_j}^*)\right) = \mathbb{C}^{2N-1}.$$

Besides, we would like the number of rows M of the sampling matrix to be as low as possible. To enforce this condition, we can require that no two subspaces $\mathrm{Im}(\mathbf{T}_{c_j})$, $\mathrm{Im}(\mathbf{T}_{c_i})$ are the same, because this would mean that all the rows of $\mathbf{\Theta}$ generated by $\mathbf{c}_i$ or $\mathbf{c}_j$ belong to the same subspace, and thus either $\mathbf{c}_i$ or $\mathbf{c}_j$ is not necessary. Thus, the size of $\mathbf{C}$ could be reduced by eliminating one of them. If at least one of the columns of $\hat{\mathbf{U}}_i$ is linearly independent from the columns of $\hat{\mathbf{U}}_j$, this condition is satisfied.

The problem is difficult because it is not clear how to select the vectors $\{\mathbf{c}_0, \mathbf{c}_1, \ldots, \mathbf{c}_{M-1}\}$ from the conditions given. One approach is to first select a basis for $\mathbb{C}^{2N-1}$, that we denote by $\{\theta_k\}_{k=1}^{2N-1}$, which would be the desired set of rows in $\mathbf{\Theta}$ that guarantees the full column rank property. However, without prior information about at least one vector $\mathbf{c}_j$, equation (5.1) can only be solved in the case $j = i$. Without loss of generality we can start with $j = 0$, that is, constructing $\mathbf{c}_0$. This case corresponds to finding a finite-length sequence whose autocorrelation is given. It is easily seen that this problem is non-linear, but can be solved by spectral factorization. However, this imposes that one vector θ_n must comply with the properties of the autocorrelation, which means having its maximum in lag $m = 0$ and Hermitian symmetry. Once this step is solved, the sequence $c_0[n]$ is known, and thus also the matrix $\mathbf{T}_{c_0}$. Equation (5.1) with $j = 0$ and $i = 1$ is now linear, and the sequence $c_1[n]$ can be found by solving

$$\mathbf{T}_{c_0}^* \mathbf{c}_1 = \boldsymbol{\theta}_k, \tag{5.4}$$

where $\boldsymbol{\theta}_k$ is the k-th vector of the basis we have previously chosen. However, as we said before, for an exact solution of this system of equations it must happen that $\boldsymbol{\theta}_k \in \mathrm{Im}(\mathbf{T}_{c_0}^*)$. If this is not the case for any $\boldsymbol{\theta}_k$, the best we can do is finding the projection of $\boldsymbol{\theta}_k$ onto $\mathrm{Im}(\mathbf{T}_{c_0}^*)$. However, once $c_1[n]$ is found, it is unclear what is the relationship between $\mathrm{Im}(\mathbf{T}_{c_1}^*)$ and $\mathrm{Im}(\mathbf{T}_{c_0}^*)$, and how to proceed further to find the remaining rows $\mathbf{c}_i$ of the sampling matrix.

5.2 Binary sampling matrices

In order to find a set of solutions, we can restrict our problem more. In this work, we will only consider binary matrices, that is, matrices whose entries are drawn from the set $\{0, 1\}$,

and in particular those that have a fixed number of ones per row. We will say that a binary matrix belongs to class $\mathbb{B}_b$ if it has exactly b ones in each row.

The motivation behind this choice is that the compressive sampling process in (2.3) can be implemented in the analog domain with the periodic sampling architecture proposed in [1], that we explained in 2.1.1. The sampling is conducted by mixing the incoming signal with a set of periodic piecewise continuous signals. These signals have a constant value at each interval $nT \leq t < (n+1)T$, where T is the Nyquist rate corresponding to the scanned bandwidth. The value in the i-th branch at a particular interval n is given by $c_i[-n]$. The generation of a periodic signal with only two possible values is easier than other more complicated signals. We note here that matrices whose elements belong to the sets $\{-1, 1\}$ or $\{-1, 0, 1\}$, named bipolar and ternary matrices respectively, can also be of interest.

In order to study the full column rank property of $\mathbf{\Theta}$, we will use its definition given by equation (3.4). We can omit the complex conjugate as these matrices only have real entries, and we will use the notation $c_{i,n} = c_i[-n]$ for simplicity. Recalling the definition of the Kronecker product we can write

$$\mathbf{C} \otimes \mathbf{C} = \begin{bmatrix} c_{0,0}\mathbf{C} & c_{0,1}\mathbf{C} & \cdots & c_{0,N-2}\mathbf{C} & c_{0,N-1}\mathbf{C} \\ c_{1,0}\mathbf{C} & c_{1,1}\mathbf{C} & \cdots & c_{1,N-2}\mathbf{C} & c_{1,N-1}\mathbf{C} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ c_{M-1,0}\mathbf{C} & c_{M-1,1}\mathbf{C} & \cdots & c_{M-1,N-2}\mathbf{C} & c_{M-1,N-1}\mathbf{C} \end{bmatrix}.$$

It is easily seen that if $\mathbf{C}$ is a binary, bipolar, or ternary matrix then $\mathbf{C} \otimes \mathbf{C}$ is, respectively, a binary, bipolar, or ternary matrix as well. We can also write $\mathbf{C} \otimes \mathbf{C}$ in terms of its rows as

$$[\mathbf{c}_0 \otimes \mathbf{c}_0, \mathbf{c}_0 \otimes \mathbf{c}_1, \ldots, \mathbf{c}_0 \otimes \mathbf{c}_{M-1}, \ldots, \mathbf{c}_{M-1} \otimes \mathbf{c}_0, \ldots, \mathbf{c}_{M-1} \otimes \mathbf{c}_{M-1}]^T. \tag{5.5}$$

As we can see, each row of $\mathbf{C} \otimes \mathbf{C}$ is given by $(\mathbf{c}_i \otimes \mathbf{c}_j)^T = \mathbf{c}_i^T \otimes \mathbf{c}_j^T$.

The repetition matrix $\mathbf{T}$, defined in (3.2), can be written as a block matrix, where each block is made of rows of $\mathbf{I}_{2N-1}$. In order to express $\mathbf{T}$ in a more convenient way, we will denote by $\{\mathbf{e}_j\}_{j=1}^{2N-1}$ the standard basis of $\mathbb{R}^{2N-1}$. We define the $(2N-1) \times N$ matrices $\mathbf{E}_n$ as

$$\mathbf{E}_0 = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_N], \tag{5.6}$$

$$\mathbf{E}_n = [\mathbf{e}_{2N-n}, \mathbf{e}_{2N-n+1}, \ldots \mathbf{e}_{2N-1}, \mathbf{e}_1, \ldots, \mathbf{e}_{N-n}], \tag{5.7}$$

with $n \in \{1, 2, \ldots, N-1\}$. Now we can write $\mathbf{T}$ as

$$\mathbf{T} = [\mathbf{E}_0, \mathbf{E}_1, \ldots, \mathbf{E}_{N-1}]^T. \tag{5.8}$$

The matrix $\mathbf{\Theta}$ is given by the product of $\mathbf{C} \otimes \mathbf{C}$ and $\mathbf{T}$, and thus each row of $\mathbf{\Theta}$ is a linear combination of the rows of $\mathbf{T}$ with coefficients given by the corresponding row of $\mathbf{C} \otimes \mathbf{C}$. Using (5.5) and (5.8) this is expressed as

$$\mathbf{r}_{c_j,c_i}^T = (\mathbf{c}_i^T \otimes \mathbf{c}_j^T)\,\mathbf{T} = \sum_{n=0}^{N-1} c_{i,n}\, \mathbf{c}_j^T \mathbf{E}_n^T. \tag{5.9}$$

It can be seen from (5.6) that the product $\mathbf{c}_j^T \mathbf{E}_0^T$ gives a $1 \times (2N-1)$ vector of the form $[\mathbf{c}_j^T\ \mathbf{0}_{N-1}^T]$, where $\mathbf{0}_{N-1}$ is the $(N-1) \times 1$ zero vector. From (5.7) it can be seen that the

rest of the products $\mathbf{c}_j^T \mathbf{E}_n^T$ are circular shifts of this vector by n positions to the left. For $n \in \{1, 2, \ldots, N-1\}$, we can write

$$\mathbf{c}_j^T \mathbf{E}_n^T = \left[c_{j,n}, c_{j,n+1}, \ldots, c_{j,N-1}, \mathbf{0}_{N-1}^T, c_{j,0}, \ldots, c_{j,n-1}\right]. \tag{5.10}$$

5.2.1 Binary matrices of class $\mathbb{B}_1$ and concatenated sparse rulers

These matrices have a single one per row, and correspond to multi-coset sampling matrices. An $M \times N$ matrix $\mathbf{C}$ belonging to this class is constructed by selecting M rows of the identity matrix $\mathbf{I}_N$. For them, equation (5.9) becomes

$$\mathbf{r}_{c_j,c_i}^T = \mathbf{c}_j^T \, \mathbf{E}_{n_i}^T,$$

where n_i is the position where the vector $\mathbf{c}_i$ has its only one. Because of the left circular shift introduced by $\mathbf{E}_{n_i}^T$, the position of the one in $\mathbf{r}_{c_j,c_i}$ is given by $n_i - n_j$. We recall here that the elements of $\mathbf{r}_{c_j,c_i}$ are indexed as $\mathbf{r}_{c_j,c_i} = \left[r_{c_j,c_i}[0], \ldots, r_{c_j,c_i}[1-N], r_{c_j,c_i}[N-1], \ldots, r_{c_j,c_i}[1]\right]^T$.

As proven in [24, 1], full column rank is achieved when every column of $\mathbf{\Theta}$ has a one, because there is necessarily a one in each row. If the compression rate desired is the minimum possible, the problem of designing $\mathbf{C}$ boils down to an $(N-1)$-length minimal sparse ruler problem. As this case is already studied, we refer to [1] for further details.

However, minimal sparse rulers cannot, to the best of our knowledge, be constructed by any quick procedure, making a brute force search over the set of length $(N-1)$ sparse rulers necessary. This has already been done, and several minimal sparse rulers have been tabulated. If the desired length is greater than the maximum length available, one can think on concatenating two or more rulers to increase the block length. This would increase the number of lags that can be estimated, thus widening the applicability of the estimation methods discussed in this work. Concatenating L minimal sparse rulers of the same length $N-1$ has already been studied in [23], proving that they are optimal in the sense that they preserve the identifiability of $\mathbf{r}_x$ while achieving the minimum compression rate. The sampling matrix associated with this concatenation would be $\bar{\mathbf{C}} = \mathbf{I}_L \otimes \mathbf{C}$, which is an $ML \times NL$ block diagonal matrix, where each block is the sampling matrix associated with an $(N-1)$-length minimal sparse ruler. With this method, the block sizes available are multiples of N. We are going to study the case of concatenating several sparse rulers of different lengths. By doing this, we allow more flexibility in the block length N, which can be expressed as the sum of the individual block sizes of the minimal sparse rulers selected.

Mathematically, concatenating L minimal sparse rulers means choosing the multi-coset set $\mathcal{P} = \mathcal{P}_0 \cup \mathcal{P}_1 \cup \cdots \cup \mathcal{P}_{L-1}$, defining $\mathcal{P}_l$ as the set

$$\mathcal{P}_l = \{r + \sum_{m=0}^{l-1} N_m, \; r \in \mathcal{R}_l\},$$

where $\mathcal{R}_l$ is a minimal sparse ruler of length $N_l - 1$. This means that $\mathcal{P}$ is formed by concatenating sparse rulers without overlap.

We define $M_l = |\mathcal{R}_l|$, that is, M_l is the cardinality of the l-th sparse ruler. As $\mathcal{P}_i \cap \mathcal{P}_j = \emptyset \; \forall i, j \in \{0, 1, \ldots, L-1\}$, the cardinality of $\mathcal{P}$ is $M_T = \sum_{l=0}^{L-1} M_l$, and the compression rate is given by $\frac{M_T}{N_T}$, defining $N_T = \sum_{l=0}^{L-1} N_l$. We note here that we have given up minimality,

because a minimal sparse ruler of length $N_T - 1$ will have less elements than M_T and thus it will offer a better compression. The sampling matrix in this case is given by

$$\bar{\mathbf{C}} = \begin{bmatrix} \mathbf{C}_0 & 0 & \cdots & 0 \\ 0 & \mathbf{C}_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{C}_{L-1} \end{bmatrix},$$

which is a block diagonal matrix where the l-th block is the matrix $\mathbf{C}_l$ associated with the l-th sparse ruler. This is again a matrix in $\mathbb{B}_1$, so the condition we need for statistical identifiabiliy is that its rows are selected from $\mathbf{I}_{N_T}$ according to a sparse ruler, or, equivalently, that the set $\mathcal{P}$ is a sparse ruler. Unfortunately, this is not always the case. A counterexample for length 20 is the concatenation of the following minimal sparse rulers of lengths 9 and 10 respectively,

$$\mathcal{R}_0 = \{0, 1, 4, 7, 9\},$$

$$\mathcal{R}_1 = \{0, 1, 2, 3, 6, 10\}.$$

The concatenated sparse ruler is the set $\{0, 1, 4, 7, 9, 10, 11, 12, 13, 16, 20\}$, and it is easily seen that the integers 14, 17 and 18 cannot be obtained as a difference of two of its elements. Thus, concatenating sparse rulers of different lengths does not always lead to the statistical identifiability of the autocorrelation values. However, checking if a certain sequence of sparse rulers is a sparse ruler is easy, so several options can be tried until a certain combination that satisfies the desired block length and the statistical identifiability property is found.

5.2.2 Binary matrices of class $\mathbb{B}_2$

These matrices have two ones per row. We denote the locations of the ones in the i-th row of $\mathbf{C}$ by the pair (n_{i1}, n_{i2}), where $n_{i1} < n_{i2}$. Besides, we assume that no two different rows have the same locations for their ones, as this would result in a rank-deficient matrix $\boldsymbol{\Theta}$.

In this case, equation (5.9) becomes

$$\mathbf{r}_{c_j,c_i}^T = \mathbf{c}_j^T \, \mathbf{E}_{n_{i1}}^T + \mathbf{c}_j^T \, \mathbf{E}_{n_{i2}}^T.$$

Due to the circular shifts, in the first term of the sum the ones will appear at the locations $n_{i1} - n_{j1}, n_{i1} - n_{j2}$, and in the second term they will appear at $n_{i2} - n_{j1}, n_{i2} - n_{j2}$. Thus, the number and locations of the non-zero elements of $\mathbf{r}_{c_j,c_i}$ can be divided in three cases.

1. The pairs $(n_{i1}, n_{i2}), (n_{j1}, n_{j2})$ have all different elements. In this case two things can happen. If the pairs satisfy $n_{i1} - n_{i2} = n_{j1} - n_{j2}$, then $n_{i1} - n_{j1} = n_{i2} - n_{j2}$ and we have a two at lag $m = n_{i1} - n_{j1}$ and two ones. Otherwise there are four ones and none of them in the first element, which corresponds to lag $m = 0$.

2. The pairs $(n_{i1}, n_{i2}), (n_{j1}, n_{j2})$ have one element in common. Then there will be four ones again, but one will appear in the first element, which corresponds to lag $m = 0$.

3. The pairs $(n_{i1}, n_{i2}), (n_{j1}, n_{j2})$ are equal, which can only happen in the case $i = j$, as we imposed that no two rows can have the same locations for their ones. This corresponds to the M rows corresponding to the autocorrelations $\mathbf{r}_{c_i,c_i}$. In this case there is a two at lag $m = 0$, and two ones at the locations $n_{i1} - n_{i2}$ and its opposite $n_{i2} - n_{i1}$.

For real signals it holds that $r_{c_j,c_i}[m] = r_{c_i,c_j}[-m]$, so the locations of the ones in $\mathbf{r}_{c_i,c_j}$ will be at the opposite lags of those in $\mathbf{r}_{c_j,c_i}$. If two pairs satisfy $n_{i1} + n_{i2} = n_{j1} + n_{j2}$, then $\mathbf{r}_{c_i,c_j} = \mathbf{r}_{c_j,c_i}$, as the locations of their ones will be equal. This can be seen considering the locations of the ones for $\mathbf{r}_{c_j,c_i}$, which are

- $n_{i1} - n_{j1} = n_{j2} - n_{i2} = a$
- $n_{i1} - n_{j2} = n_{j1} - n_{i2} = b$
- $n_{i2} - n_{j1} = n_{j2} - n_{i1} = -b$
- $n_{i2} - n_{j2} = n_{j1} - n_{i1} = -a$

Because in $\mathbf{r}_{c_i,c_j}$ the ones appear at opposite locations, it is clear that both vectors are equal. Taking this into account, we need to find sufficient conditions for $\mathbf{\Theta}$ to have full column rank. However, we have not been able to do it so far.

5.3 Conclusions

In this chapter, we have addressed the problem of the design of a sampling matrix $\mathbf{C}$ that preserves the statistical identifiability of the autocorrelation values. We took the approach of guaranteeing that the recovery matrix $\mathbf{\Theta}$ for the least squares estimator has full column rank. Because its rows are the auto and cross-correlations between every pair of rows of $\mathbf{C}$, the problem can be stated as how to find a basis of $\mathbb{C}^{2N-1}$ from the M^2 possible cross-correlations of the M sequences $\{c_i[n]\}_{i=0}^{M-1}$ that constitute the rows of the sampling matrix $\mathbf{C}$. We showed that the general problem is difficult, and we left it open for future research. In the last section, we considered binary matrices with a single one per row or a single pair of ones per row. In the first case, they correspond to multicoset sampling, which requires a sparse ruler pattern to satisfy the identifiability conditions. The possibility of concatenating minimal sparse rulers of different lengths is considered as a means of obtaining larger values of N using rulers that are already tabulated. This would avoid the problem of finding long minimal sparse rulers, which can only be done by brute force search. However, such a concatenation is not always a sparse ruler, and thus identifiability is not always satisfied. In the second case, we derived the possible locations of the elements of $\mathbf{\Theta}$. However, we failed to find sufficient conditions for identifiability, and we again left it open for future research.

Spectral Estimation with Random Sampling

6

In this chapter, we study spectral estimation with random sampling, a type of sampling we briefly described in Section 2.1.3. As we explained, random sampling includes every sampling scheme where the time between samples is a random variable following a specific distribution. In [8], random sampling was considered as an alternative to uniform sampling, with avoiding aliasing as the goal. The authors provided a definition of alias-free estimation and showed that several sampling schemes, based on the Poisson process, were alias-free, while some others were not. However, they did not study the effects of finite sensing time. The work [16] studied the properties of a PSD estimator based on a finite number of samples taken at instants following a Poisson process, and showed that consistent estimation is possible even for average sampling rates lower than the Nyquist rate. The same author points in [17] that the definition of alias-free estimation provided in [8] does not imply that consistent estimation can be obtained from a finite set of samples, and reformulates the definition, providing sufficient conditions for this new definition to hold. More recently, [18] studied the theoretical properties of several types of random sampling and applied them to radio receiver design.

In our work, we consider additive random sampling, which is a particular kind of random sampling that we explain in Section 6.1, and we modify it by imposing a minimum time between samples, which allows the sampling scheme to be implemented in synchronous digital circuitry. Spectral estimation is done by employing the classical unbiased estimator for the autocorrelation and taking its discrete Fourier transform. We also study the probability of a certain lag being present, which we show is a difficult problem.

6.1 Additive Random Sampling

Additive random sampling was first proposed in [8] in order to obtain alias-free estimators for the power spectral density. The samples in additive random sampling are taken at the instants

$$t_0 = 0,$$
$$t_n = t_{n-1} + \tau_n, \quad n \in \{1, 2, \dots\},$$

where the last equation can be written equivalently as

$$t_n = t_0 + \sum_{i=1}^{n} \tau_i \,, \quad n \in \{1, 2, \dots\}. \tag{6.1}$$

The random variables τ_n are independent and identically distributed (i.i.d), which means that all intervals between consecutive samples are independent and have the same distribution.

In Poisson sampling, the sampling instants form a Poisson process, which implies that the distribution of the interarrival times τ_n is the exponential distribution, with probability density function

$$f_{\tau_n}(t) = \lambda e^{-\lambda t}; \quad t \geq 0.$$

Here λ is the mean sampling rate , and its inverse is the mean time between samples $\mu = \frac{1}{\lambda}$.

As we can see, two consecutive samples are allowed to be infinitely close, because every τ_n can take any real value greater than 0. This is physically irrealizable, because any device taking measurements requires some time between one and the next. Thus, we have modified this scheme by rounding each sampling instant t_n to the nearest Nyquist rate sample nT, where $\frac{1}{T}$ is the Nyquist rate corresponding to the scanned bandwidth. By doing so, we can establish a relationship between the samples taken following our scheme and the Nyquist rate samples. Let us consider that we collect N samples from $x[n]$. Our sampling scheme selects the samples whose indexes are given by:

$$k_0 = 0,$$

$$k_n = \text{round}\left(k_0 + \sum_{i=1}^{n} \tau_i\right) \quad n \in \{1, 2, \dots\},$$

where round$(\cdot)$ is the operator that rounds its argument to the nearest integer and $\{\tau_i\}_{i=1}^{\infty}$ are i.i.d random variables with mean μ. In order to adapt to the finite size of $x[n]$, the maximum index selected cannot be greater than $N-1$ so we add the constraint

$$\max_n\{k_n\} \leq N - 1.$$

We note that the rounding gives rise to repeated indexes and this would introduce a bias in the autocorrelation estimator, because of some samples appearing more than once. In order to avoid this bias, we erase the repeated indexes leaving a set of indexes $\{k_0, k_1, \dots, k_{M-1}\}$ with the property

$$k_0 < k_1 < \ldots < k_{M-1}.$$

Thus, our modified random sampling scheme selects a subset of the Nyquist rate samples of size M, where each sample can only appear once. As a consequence of the randomness of the interval between consecutive samples, the number M is also random. If no elimination of repeated samples was performed the expected value of M would be $\frac{N}{\mu}$, as the average interval between consecutive samples is μ. However, it is always less due to this elimination. The compression rate is given by $\frac{M}{N}$.

6.2 Probability of Obtaining a Certain Lag

In order to estimate an autocorrelation function, which depends on the difference between instants, and not on the instants themselves, we need every integer lag to be present. This means that, for each lag, there must be at least a pair of sampling instants whose difference yields that lag. It suffices to have the lags given by $\{lT\}_{l=0}^{L-1}$, where $\frac{1}{T}$ is the Nyquist rate. For simplicity, we can consider $T = 1$ and the same results are valid using scaling by the appropriate value of T. We note here that as the maximum lag considered is $L-1$, there is a truncation error in our PSD estimator, as we mentioned at the end of Section 2.2.1.

In order to find the probability of obtaining a certain lag l, we have to consider the statistical properties of the set of random variables $\{S_{k,n} = t_{n+k} - t_n\}$, where $k \in \{0, 1, \dots, N-1\}$ and $n \in \{0, 1, \dots, N-1-k\}$. These random variables cover all the differences between arrival times and thus represent all the lags. They can also be seen as the waiting time for

the arrival of k samples, starting from the n-th sample arrival. Their probability distribution can be found by using (6.1),

$$S_{k,n} = t_{n+k} - t_n = t_0 + \sum_{i=1}^{n+k} \tau_i - t_0 - \sum_{i=1}^{n} \tau_i = \sum_{i=n+1}^{n+k} \tau_i. \tag{6.2}$$

As $S_{k,n}$ is the sum of k i.i.d. random variables, its distribution is independent of n. In the case of Poisson sampling, the distribution of the sum of k i.i.d. exponential random variables is an Erlang distribution with parameters (k, λ), whose probability density function is

$$f_k(t) = \frac{\lambda^k \, t^{k-1} \, e^{-\lambda t}}{(k-1)!}; \quad t \geq 0.$$

The next step is deriving the probability of obtaining each lag. We will consider that lag l is present if there is at least one pair of instants t_{n+k}, t_n whose difference $t_{n+k} - t_n$ belongs to the interval $\Lambda_l = [l - \frac{1}{2}, l + \frac{1}{2})$. This means that rounding the difference to the nearest integer gives l. We note here that the lag 0 is always present, because it is the difference between each instant and itself, so we consider only the intervals $\Lambda_1, \ldots, \Lambda_{M-1}$.

We will denote the event of lag l being present by A_l. Defining the set I_k as $I_k = \{0, 1, \ldots, N-1-k\}$, we can express A_l as

$$A_l \equiv \exists \, k \in I_0, n \in I_k : S_{k,n} \in \Lambda_l.$$

The probability of A_l can be written as

$$\Pr[A_l] = \Pr[\exists n \in I_1 : S_{1,n} \in \Lambda_l \cup \cdots \cup \exists n \in I_{N-1} : S_{N-1,n} \in \Lambda_l]. \tag{6.3}$$

This means that the probability of lag l happening is the probability of a certain difference $t_{n+1} - t_n$ yielding it, or a certain difference $t_{n+2} - t_n$, etc. Evaluating directly the probability of this union of events using the inclusion-exclusion principle seems tedious, so we try to find an easier way by taking into account the complementary event A_l^c, that is, the event of lag l not being present. It can be expressed as

$$A_l^c \equiv \forall \, k \in I_0, n \in I_k : S_{k,n} \notin \Lambda_l.$$

Because it is the complementary event of A_l, it follows that $\Pr[A_l] = 1 - \Pr[A_l^c]$ and

$$1 - \Pr[A_m^c] = 1 - \Pr[\forall n \in I_1 : S_{1,n} \notin \Lambda_l \cap \cdots \cap \forall n \in I_{N-1} : S_{N-1,n} \notin \Lambda_l]. \tag{6.4}$$

If these events were mutually independent, we could simply express the probability of their intersection as the product of the probability of each event. However, the random variables $S_{k,n}$ are not mutually independent. They satisfy two recursive relationships that follow from (6.2):

$$S_{k,n+1} = \sum_{i=n+2}^{n+1+k} \tau_i = S_{k,n} + \tau_{n+1+k} - \tau_{n+1}, \tag{6.5}$$

$$S_{k-1,n} = \sum_{i=n+1}^{n+k-1} \tau_i = S_{k,n} - \tau_{n+k}. \tag{6.6}$$

However, for each k the variables $S_{k,n}$ and $S_{k,n+k}$ are independent, because they represent the waiting times for k events in non-overlapping intervals. This can also be seen using (6.2),

$$S_{k,n} = \sum_{i=n+1}^{n+k} \tau_i; \quad S_{k,n+k} = \sum_{i=n+k+1}^{n+2k} \tau_i.$$

In the case $k = 1$, we have $S_{1,n} = t_{n+1} - t_n = \tau_{n+1}$ and we can compute the probability of obtaining lag l from a pair t_{n+1}, t_n,

$$\begin{aligned} \Pr[\exists n \in I_1 : S_{1,n} \in \Lambda_l] = 1 - \Pr[\forall n \in I_1 : S_{1,n} \notin \Lambda_l] = \\ 1 - \Pr\left[\bigcap_{n=0}^{N-2} S_{1,n} \notin \Lambda_l\right] = 1 - \prod_{n=0}^{N-2} \Pr[S_{1,n} \notin \Lambda_l]. \end{aligned} \tag{6.7}$$

The last equality follows from the fact that τ_n are independent random variables. As they are also identically distributed, $\Pr[S_{1,n} \notin \Lambda_l]$ is independent of n. We can define $p_{k,l} = \Pr[S_{k,n} \in \Lambda_l]$ and write

$$\Pr[S_{1,n} \notin \Lambda_l] = 1 - \Pr[S_{1,n} \in \Lambda_l] = 1 - \int_{l-\frac{1}{2}}^{l+\frac{1}{2}} f_{\tau_n}(t)\, dt = 1 - p_{1,l}.$$

Thus equation (6.7) can be simplified to

$$\Pr[\exists n \in I_1 : S_{1,n} \in \Lambda_l] = 1 - \prod_{n=0}^{N-2} \Pr[S_{1,n} \notin \Lambda_l] = 1 - (1 - p_{1,l})^{N-1} \tag{6.8}$$

As we can see, computing the probability in (6.4) is non-trivial, and so far we have not been able to advance more.

6.3 Autocorrelation estimation

The autocorrelation estimator we use is given by:

$$\hat{r}_x[l] = \frac{1}{N_l} \sum_{\forall k_n,\, (k_n+l) \in \mathcal{S}} x^*[k_n] x[k_n + l] \quad l \geq 0, \tag{6.9}$$

where $\mathcal{S}$ is the set of indexes given by the sampling process and N_l the number of samples available for lag l. For negative lags, we apply the Hermitian symmetry of the autocorrelation, $r_x^*[l] = r_x[-l]$. This estimator is unbiased as the scaling factor for each lag is the number of products used to compute it.

This computation is equivalent to applying a slightly changed version of the classical estimator given by

$$\hat{r}_x[l] = \frac{1}{N_l} \sum_{n=0}^{N-1-l} x^*[n] x[n + l] \quad l \geq 0,$$

to the signal $x[n]$ where the samples whose indexes do not belong to $\{k_n\}_{n=0}^{M-1}$ are set to zero. The difference is that we have used N_l instead of the usual $N - l$ because in general

$N_l \neq N - l$, and to keep the estimator unbiased the factor for each lag has to be the number of products used to compute it.

The number N_l determines the reliability of the estimator for each lag, as a larger N_l indicates more samples available and thus more averaging, bringing $\hat{r}_x[l]$ closer to is true value. In order to obtain a reliable estimator for the PSD, N_l should not change sharply as l grows, as all the values of $\hat{r}_x[l]$ are involved in computing the PSD estimate for each frequency. However, N_l has necessarily to decay with l, because we only collect a finite amount of samples.

The estimator in (6.9) can be applied up to lag $l = k_{M-1}$. In practice, we can collect more samples than the lags we want to estimate, in order to have a large enough N_l at each lag. We shall denote the maximum lag of interest by L. We can then group the autocorrelation estimates in the $(2L-1) \times 1$ vector $\hat{\mathbf{r}}_x$, defined as

$$\hat{\mathbf{r}}_x = [\hat{r}_x[0], \ldots, \hat{r}_x[L-1], \hat{r}_x[1-L], \ldots, \hat{r}_x[-1]\,]^T.$$

After obtaining the estimate for the autocorrelation, the WSS assumption on $x[n]$ and the Wiener-Khinchin theorem allow us to compute an estimate of the power spectral density by using the discrete Fourier transform:

$$\hat{\boldsymbol{\phi}}_x = \mathbf{F}_{2L-1}\, \hat{\mathbf{r}}_x,$$

where $\mathbf{F}_{2L-1}$ is the DFT matrix of size $(2L-1) \times (2L-1)$.

6.4 Simulation results

In this section, we show several graphs of the normalized MSE as a function of the compression rate. We compare modified Poisson sampling, which we described in Section 6.1, with a sampling scheme that selects M from the N samples of $x[n]$, and each sample has the same probability of being selected. The compression rate is $\frac{M}{N}$.

In all the simulations, the signal $x[n]$ considered is drawn from a complex circularly-symmetric Gaussian WSS random process whose PSD is obtained by filtering white Gaussian noise. Thus the autocorrelation sequence $r_x[n]$ of the process is

$$r_x[n] = h[n] \star h^*[-n] \star \sigma_0^2\, \delta[n],$$

where $h[n]$ is the impulse response of the filter, σ_0^2 is the power of the white noise driving the filter and $\star$ represents convolution. The filter we used has an impulse response $h[n]$ with length $N = 127$. It is designed so that the PSD of the output process has active bands in $[-0.9\pi, -0.65\pi]$, $[0.1\pi, 0.35\pi]$ and $[0.55\pi, 0.8\pi]$, as shown in Fig. 6.1.

For the μ parameter in our modified Poisson sampling, we choose the values 2.0, 2.7, 3.5, 4.2, 5.0, 5.7 and 6.3. In order to compare both schemes, the compression rates chosen for the scheme that selects M samples with equal probability are the same as those resulting in modified Poisson sampling.

In order to give a rough idea of what the estimates yielded look like, one outcome of the estimators is shown in Fig. 6.2.

The results are presented in Fig. 6.3, for three different number of signal lengths N, namely 20000, 50000 and 100000. As we can see, both sampling schemes yield the same NMSE. This can be explained by the fact that, if we sort in increasing order the outcomes of

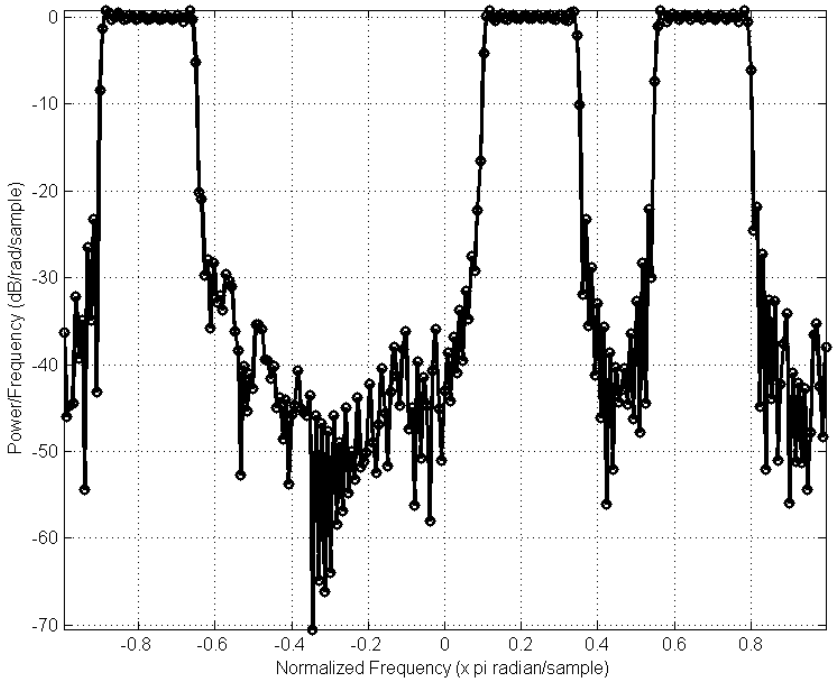


Figure 6.1: True power spectral density, $N = 127$, complex Gaussian process.

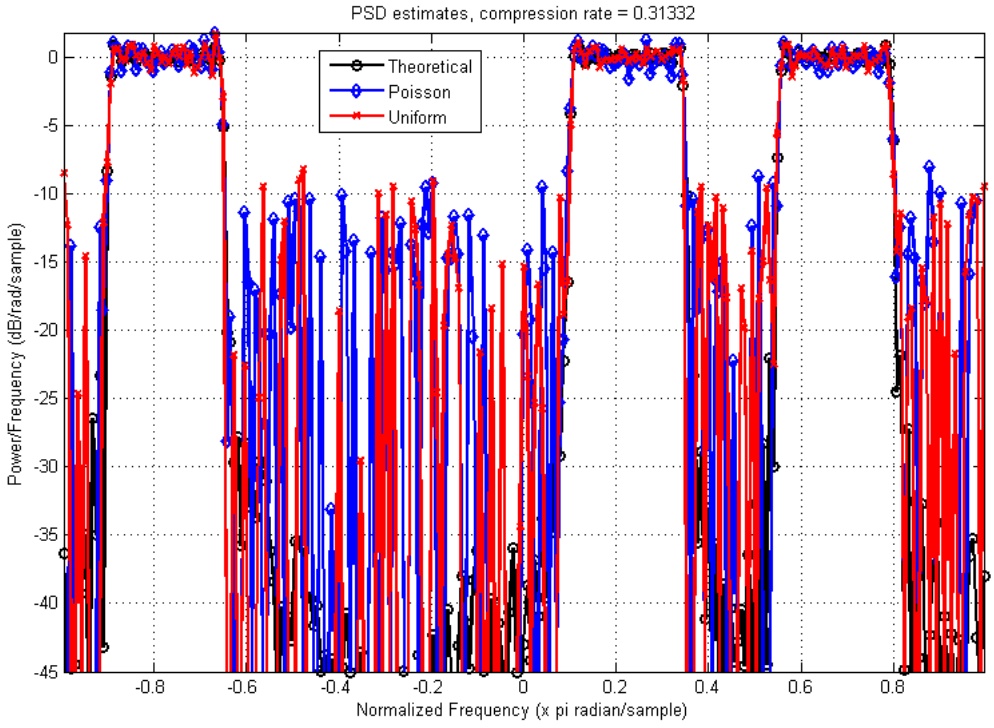


Figure 6.2: PSD estimates with random sampling. The compression rate is 0.31 and $N = 50000$ Nyquist rate samples were generated.

N i.i.d uniform random variables on the interval $[0, t]$, they represent N consecutive occurrence times of a Poisson process. For details, the reader can consult [33]. The sampling scheme that selects M samples with equal probability can be approximated by the outcomes of M i.i.d uniform random variables, sorted in ascending order. Thus, even if our modified Poisson sampling scheme is not exactly a Poisson process due to the rounding we introduce in the sampling instants, for large number of samples the relationship holds approximately.

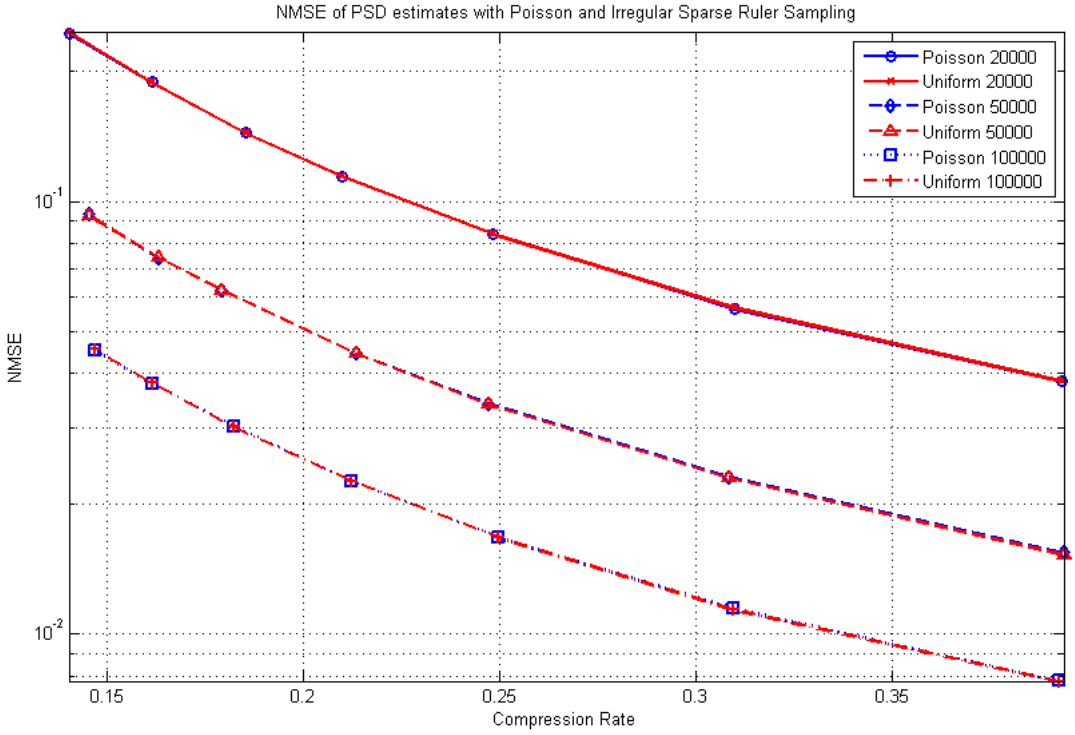


Figure 6.3: NMSE vs Compression Rate for random sampling.

6.5 Conclusions

In this chapter, we studied spectral estimation with random sampling, a type of sampling where the time between samples is a random variable. This scheme attracted attention from researchers as a mean to avoid aliasing, and because consistent estimation is possible even for mean average rates lower than the Nyquist rate. However, most schemes using a continuous positive random variable for the time between samples can require that two samples are infinitely close, making the method impossible to implement physically. Thus, we modified it by rounding each sampling instant to the closest instant in a Nyquist rate grid. This transforms the scheme into a selection of a subset of the Nyquist rate samples, where the samples are selected at random.

Using this modified version, we studied the probability of a certain lag being present as the difference between a pair of samples, and we showed that computing this probability exactly is a difficult problem, which is left open for further research. Besides, we studied the performance of the classical unbiased estimator for the autocorrelation applied to modified

Poisson sampling and to a sampling scheme where each sample has the same probaility of being selected. This was done by means of computer simulations, and they show that the NMSE is the same for both types of sampling, which can be explained by the statistical properties of the Poisson process.

7 Conclusions and Future Work

In this chapter, we summarize the work done in the thesis, draw the final conclusions, and suggest directions for further research.

7.1 Conclusions

We have studied several sampling schemes and estimation methods for the power spectral density that work at sub-Nyquist rates. Besides, we have attempted to find new sampling matrices that guarantee statistical identifiability when compressive sampling is employed.

In Chapter 3, we combined two existing methods, polarity coincidence and least squares, and applied them to spectral estimation with compressive sampling when the input signal is Gaussian. The proposed method works with the sign of the measurements instead of their value. The recovery stage, however, needs the average power of the signal to be estimated separately. The computer simulations showed that the mean squared error of our estimator is worse than the one using only least squares.

Maximum likelihood estimation for Gaussian signals with compressive sampling was investigated in Chapter 4, where we derived the conditions for the real case. They lead to a system of equations that has no closed form solution, and thus requires the use of numerical methods. From the literature, we know that finding the true maximum likelihood solution leads to a high computational cost, and thus we proposed two algorithms which yield approximate solutions that converge asymptotically to the true solution. Finally, we compared the performances of both algorithms and the least squares estimator of [1], using computer simulations. They showed that both ML algorithms perform better with regard to compression rate and additive white Gaussian noise. This can be explained by the good statistical properties of the ML estimator when the sample size is large.

Chapter 5 addresses the problem of designing sampling matrices for compressive sampling. In order for spectral estimation to be possible, its statistical identifiability has to be satisfied. Firstly, we studied the general problem of constructing a sampling matrix that preserves the identifiability, which can be stated as finding a set of M sequences of length N such that their M^2 auto- and cross-correlations contain a set of $2N-1$ linearly independent sequences. As we show, a general construction is difficult, and we leave open further research on it. Then we restrict the elements of the sampling matrix to be either 0 or 1, as the implementation of compressive sampling in hardware may be easier by using those matrices. In particular, we studied multicoset matrices, which have a single one per row, and binary matrices that have a single pair of ones per row. For multicoset matrices, their design has to be based on a sparse ruler in order to satisfy the identifiability condition. To the best of our knowledge, minimal sparse rulers of a certain length can only be found by brute force search, which is more costly as the length increases. The current approach is to use a concatenation of minimal sparse rulers that are already tabulated and have the same length. In order to offer a wider range of lengths, we propose a design that uses a concatenation of several minimal sparse rulers

of different lengths. The identifiability is possible if the resulting ruler is sparse, which is not always the case. Thus, sufficient conditions for a concatenation of sparse rulers to be a sparse ruler need to be found. Finally, we study binary matrices that have a single pair of ones per row, and depending on the positions of these pairs we find the locations of the non-zero elements in the recovery matrix for least squares. However, we have been unable to find conditions for this matrix to be full column rank, which is equivalent to preserving the statistical identifiability. We leave this line open for further research.

In Chapter 6, we studied spectral estimation with random sampling. We modified the random sampling model in order to be physically realizable by imposing a minimum distance between sampling intervals. Then, we attempted to compute, for a finite sample size, the probability of a certain lag to be present when random sampling is employed. This computation is shown to be difficult and we did not complete it. Finally, we compared the performances of two sampling schemes, based on Poisson sampling and on random selection of a subset of the collected samples, each one with equal probability. The estimator used is the classical unbiased estimator of the autocorrelation, and the performances were shown to be equal for the sample sizes considered, which may be explained by the properties of the Poisson process.

7.2 Suggestions for Further Work

Other statistical distributions In the estimation methods we have studied, we made the assumption that the signal of interest could be modelled as a Gaussian process. Mathematically, this is convenient because Gaussian processes are completely determined by their first and second order statistics (mean and autocovariance), and besides they lead to mathematically tractable problems. However, it may be of interest to study what happens when the signal has a different statistical distribution.

Sampling matrix design Here we left several lines open. First of all, a general design of a sampling matrix that satisfies the statistical identifiability property was addressed, but no algorithm or constructive procedure could be developed. In order to find an optimal sampling matrix, further research may be of interest. Later, we restricted our choices to binary matrices that have a fixed number of ones per row. When there is a single one in each row, it is already proven that the identifiability is guaranteed by using sparse rulers, but constructing them for an arbitrary length is a mathematically difficult problem. We suggested concatenating sparse rulers of different lengths, which are already tabulated, as means to avoid the problem of finding new rulers. However, sufficient conditions for such scheme to be a sparse ruler are still left open. Finally, for binary matrices with a single pair of ones per row, we left open the problem of finding sufficient conditions for the identifiability property to be satisfied. If these conditions can be found, it is interesting to see if the design of these matrices is easier than solving the sparse ruler problem for multicoset matrices.

Random sampling For random sampling, we left open the computation of the probability of obtaining a certain lag. If this was completed, a criterion could be developed for obtaining with high probability a sufficient amount of samples that guarantees that every lag of interest is present, and different statistical distributions could be compared under this idea.

Bibliography

[1] D.D. Arianandа and G. Leus. Compressive wideband power spectrum estimation. *IEEE Transactions on Signal Processing*, 60(9):4775–4789, 2012.

[2] F. Marvasti. *Non-uniform sampling: theory and practice*. Springer, 2001.

[3] C.E. Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949.

[4] Spectrum Efficiency Working Group. Report of the spectrum efficiency working group. Technical report, Federal Communications Commission, 2002.

[5] J. Mitola. *Cognitive radio: An integrated agent architecture for software defined radio*. PhD thesis, Royal Inst. Technol. (KTH), Stockholm, Sweden, 2000.

[6] Simon Haykin. Cognitive radio: brain-empowered wireless communications. *IEEE Journal on Selected Areas in Communications*, 23(2):201–220, 2005.

[7] Bin Le, T.W. Rondeau, J.H. Reed, and C.W. Bostian. Analog-to-digital converters. *Signal Processing Magazine, IEEE*, 22(6):69–77, 2005.

[8] Harold S. Shapiro and Richard A. Silverman. Alias-free sampling of random noise. *Journal of the Society for Industrial and Applied Mathematics*, 8(2):pp. 225–248, 1960.

[9] E.J. Candès and M.B. Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, 2008.

[10] J. G. Proakis and D. G. Manolakis. *Digital Signal Processing: Principles, Algorithms, And Applications, 4th ed.* Pearson Education, 2007.

[11] E.J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.

[12] D.L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[13] R. Venkataramani and Y. Bresler. Perfect reconstruction formulas and bounds on aliasing error in sub-Nyquist nonuniform sampling of multiband signals. *IEEE Transactions on Information Theory*, 46(6):2173–2183, 2000.

[14] M. Mishali and Y.C. Eldar. Blind multiband signal reconstruction: Compressed sensing for analog signals. *IEEE Transactions on Signal Processing*, 57(3):993–1009, 2009.

[15] M. Mishali and Y.C. Eldar. From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):375–391, 2010.

[16] E. Masry. Poisson sampling and spectral estimation of continuous-time processes. *IEEE Transactions on Information Theory*, 24(2):173–183, 1978.

[17] E. Masry. Alias-free sampling: An alternative conceptualization and its applications. *IEEE Transactions on Information Theory*, 24(3):317–324, 1978.

[18] J. J. Wojtiuk. *Randomised sampling for radio design*. PhD thesis, University of South Australia, 2000.

[19] B. I. Ahmad and A. Tarczynski. A SARS method for reliable spectrum sensing in multiband communication systems. *IEEE Transactions on Signal Processing*, 59(12):6008–6020, 2011.

[20] A. Papoulis and S. U. Pillai. *Probability, Random Variables and Stochastic Processes*. McGraw Hill, 2002.

[21] M. H. Hayes. *Statistical Digital Signal Processing and Modelling*. Johh Wiley & Sons, Inc., 1996.

[22] Prabhu Babu and Petre Stoica. Spectral analysis of nonuniformly sampled data: a review. *Digital Signal Processing*, 20(2):359 – 378, 2010.

[23] D. Romero and G. Leus. Compressive covariance sampling. In *Information Theory and Applications Workshop (ITA), 2013*, pages 1–8, 2013.

[24] D.D. Arianandа and G. Leus. Wideband power spectrum sensing using sub-Nyquist sampling. In *IEEE Inter. Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 101–105, 2011.

[25] John Leech. On the representation of 1, 2,. . . , n by differences. *Journal of the London Mathematical Society*, 1(2):160–169, 1956.

[26] J.H. Van Vleck and D. Middleton. The spectrum of clipped noise. *Proceedings of the IEEE*, 54(1):2–19, 1966.

[27] Y.L. Polo, Y. Wang, A. Pandharipande, and G. Leus. Compressive wide-band spectrum sensing. In *IEEE Inter. Conference in Acoustics, Speech and Signal Processing (ICASSP)*, pages 2337–2340, 2009.

[28] Ashok K. Gupta and William P. Birkemeier. Estimation of a quadrature cross- correlation detector by analog, polarity-coincidence, and relay methods (corresp.). *IEEE Transactions on Information Theory*, 30(6):859–865, 1984.

[29] J.P. Burg, D.G. Luenberger, and D.L. Wenger. Estimation of structured covariance matrices. *Proceedings of the IEEE*, 70(9):963–974, 1982.

[30] D. Romero, R. Lopez-Valcarce, and G.Leus. Compressive wideband spectrum sensing with spectral prior information. In *Proc. IEEE Inter. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4469–4473, 2013.

[31] Robert G. Gallager. Circularly-symmetric Gaussian random vectors. *preprint*, 2008. http://www.rle.mit.edu/rgallager/documents/CircSymGauss.pdf.

[32] S. M. Kay. *Fund. of Statistical Digital Signal Processing. vol. I: Estimation Theory*. "Prentice Hall", 1993.

[33] P. Van Mieghem. *Performance Analysis of Communications Networks and Systems*. Cambridge University Press, 2006.

Compressive Power Spectral Density Estimation with Non-Uniform Sampling

Fernando de la Hucha Arce

Introduction

Least Squares with Hard Limiting

Maximum Likelihood

Random Sampling

Compressive Power Spectral Density Estimation with Non-Uniform Sampling

Fernando de la Hucha Arce

27-08-2013

Index

1 Introduction

2 Least Squares with Hard Limiting

3 Maximum Likelihood

4 Random Sampling

Problem Statement

Spectral Density Estimation with Non-Uniform Sampling

Fernando de la Hucha Arce

Introduction

Least Squares with Hard Limiting

Maximum Likelihood

Random Sampling

- **Spectral Estimation**: the problem of estimating the power spectral density (PSD) $\mathbf{\Phi}_x(\omega)$ of a signal $x(t)$. The PSD is the Fourier transform of the autocorrelation $r_x(\tau) = E\left(x(t)x^*(t-\tau)\right)$, under the assumption that $x(t)$ is wide-sense stationary (WSS).
- **Sampling**: the problem of guaranteeing that either $x(t)$ or $\mathbf{\Phi}_x(\omega)$ can be recovered from measurements of $x(t)$ taken at certain instants $t_0, t_1, \ldots, t_N$.
- **Compression**: a reduction in the rate f_s used to acquire the measurements.
  Classical sampling theory requires $f_s \geq 2f_0$, assuming that $x(t)$ is bandlimited with maximum frequency f_0.

Spectrum Management

Spectral Density Estimation with Non-Uniform Sampling

Fernando de la Hucha Arce

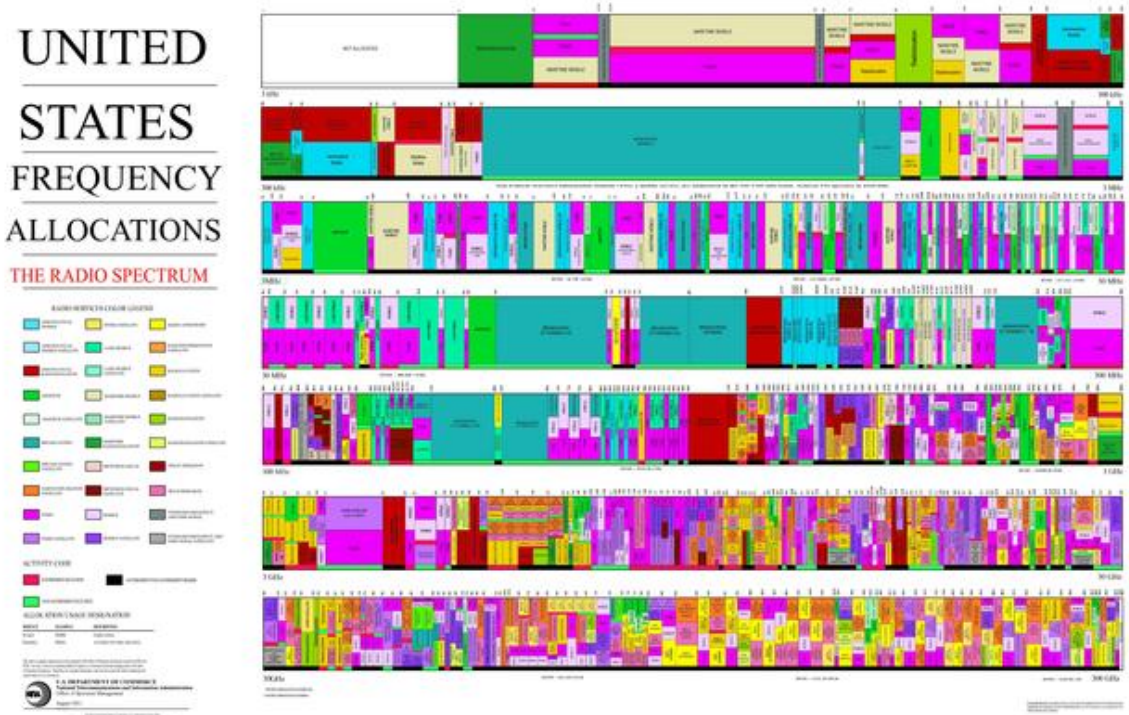
Introduction

Least Squares with Hard Limiting

Maximum Likelihood

Random Sampling

Current spectrum management laws allocate frequency bands in a static way.



Due to the growth of wireless communications, few bandwidth is left for new applications.

Spectrum Access and Cognitive Radio

Spectral Density Estimation with Non-Uniform Sampling

Fernando de la Hucha Arce

Introduction

Least Squares with Hard Limiting

Maximum Likelihood

Random Sampling

- Spectrum Utilization: Several studies[1] have shown that not every assigned band is being used at every location and time.
- Cognitive Radio: An approach to wireless communications where transceivers learn from the environment and adapt its parameters, in order to optimize spectrum utilization and provide reliable communications.
- Spectrum Sensing: The task of estimating the power spectral density of the incoming signal and detecting available spectrum.

[1] Such as this report from the FCC.

Motivation for our work

Spectral Density Estimation with Non-Uniform Sampling

Fernando de la Hucha Arce

Introduction

Least Squares with Hard Limiting

Maximum Likelihood

Random Sampling

For spectrum sensing, we know that:

- Wideband PSD estimation with classical sampling theory would require very large sampling rates, because of the constraint $f_s \geq 2f_0$.
- Very fast analog-to-digital converters have a high power consumption, which is undesirable for a communication device.

Thus, we seek to study methods that allow consistent estimation of the PSD with compression, that is

$$f_s < 2f_0.$$

Sampling

Spectral Density Estimation with Non-Uniform Sampling

Fernando de la Hucha Arce

Introduction

Least Squares with Hard Limiting

Maximum Likelihood

Random Sampling

The types of sampling we considered are:

- Uniform: $t_n = nT_s$, $n \in \mathbb{Z}$ and $T_s = \frac{1}{f_s}$.
- Compressive: $\mathbf{y}[k] = \mathbf{C}\,\mathbf{x}[k]$, where $\mathbf{C}$ is an $M \times N$ matrix with $M < N$, and

$$\mathbf{x}[k] = [x[kN], x[kN+1], \ldots, x[kN+N-1]]^T,$$

$$\mathbf{y}[k] = [y_0[k], y_1[k], \ldots, y_{M-1}[k]]^T.$$

  A special case is multicoset sampling, where $\mathbf{C}$ has a single one per row, and thus

$$\mathbf{y}[k] = [x[kN+n_0], x[kN+n_1], \ldots, x[kN+n_{M-1}]]^T.$$

- Random: $t_0 = 0, t_n = t_{n-1} + \tau_n$, $n \in \{1, 2, \ldots, \}$, where τ_n are independent and identically distributed (i.i.d) random variables.

Least Squares and Hard Limiting

Spectral Density Estimation with Non-Uniform Sampling

Fernando de la Hucha Arce

Introduction

Least Squares with Hard Limiting

Maximum Likelihood

Random Sampling

This method combines:

- Hard limiting: it works with the sign of the measurements instead of its values. The signal has to be Gaussian to allow the recovery of the autocorrelation of the measurements.
- Compressive sampling: it uses the compressive sampling model, but imposes a restriction on the allowed sampling matrices.
- Least squares: recovers the PSD using a linear least squares method developed in [1].

Least Squares

Spectral Density Estimation with Non-Uniform Sampling

Fernando de la Hucha Arce

Introduction

Least Squares with Hard Limiting

Maximum Likelihood

Random Sampling

- Autocorrelation matrices:
  - $\mathbf{R}_x = E\left(\mathbf{x}[k]\mathbf{x}^H[k]\right)$, $\mathbf{R}_y = E\left(\mathbf{y}[k]\mathbf{y}^H[k]\right)$.
  - Related by $\text{vec}(\mathbf{R}_y) = (\mathbf{C}^* \otimes \mathbf{C})\,\text{vec}(\mathbf{R}_x)$.
- Autocorrelation vector:
  - $\mathbf{r}_x = [r_x[0], r_x[1], \ldots, r_x[N-1], r_x[1-N] \ldots, r_x[-1]]^T$.
  - Due to the Hermitian Toeplitz structure of $\mathbf{R}_x$, it follows $\text{vec}(\mathbf{R}_x) = \mathbf{T}\,\mathbf{r}_x$, where $\mathbf{T}$ is a repetition matrix.
- Least Squares:
  - $\text{vec}(\mathbf{R}_y) = \mathbf{\Theta}\,\mathbf{r}_x$, which is an over-determined system of equations if $M^2 \geq 2N-1$.
  - The $M^2 \times (2N-1)$ recovery matrix $\mathbf{\Theta}$ is given by $\mathbf{\Theta} = (\mathbf{C}^* \otimes \mathbf{C})\,\mathbf{T}$.
  - A unique solution exists only if $\mathbf{\Theta}$ has full column rank.

Hard Limiting

The real input processes are denoted by $x(t)$ and $y(t)$, and their clipped versions by $x_c(t)$ and $y_c(t)$.

- Hard limiter: a device that returns 1 when the input is greater than 0 and -1 when it is lower than 0.
- Autocorrelation map: $r_{xy}(\tau) = r_{xy}(0) \sin\left(\frac{\pi}{2} r_{x_c,y_c}(\tau)\right)$. The value $r_{xy}(0)$ has to be estimated separately.
- Sign preservation:
  $\text{sign}(\mathbf{y}[k]) = \text{sign}(\mathbf{C}\,\mathbf{x}[k]) = \mathbf{C}\,\text{sign}(\mathbf{x}[k])$.
  If the last equality holds, the least squares method can be applied to the hard limited measurements.

Compressive Power Spectral Density Estimation with Non-Uniform Sampling

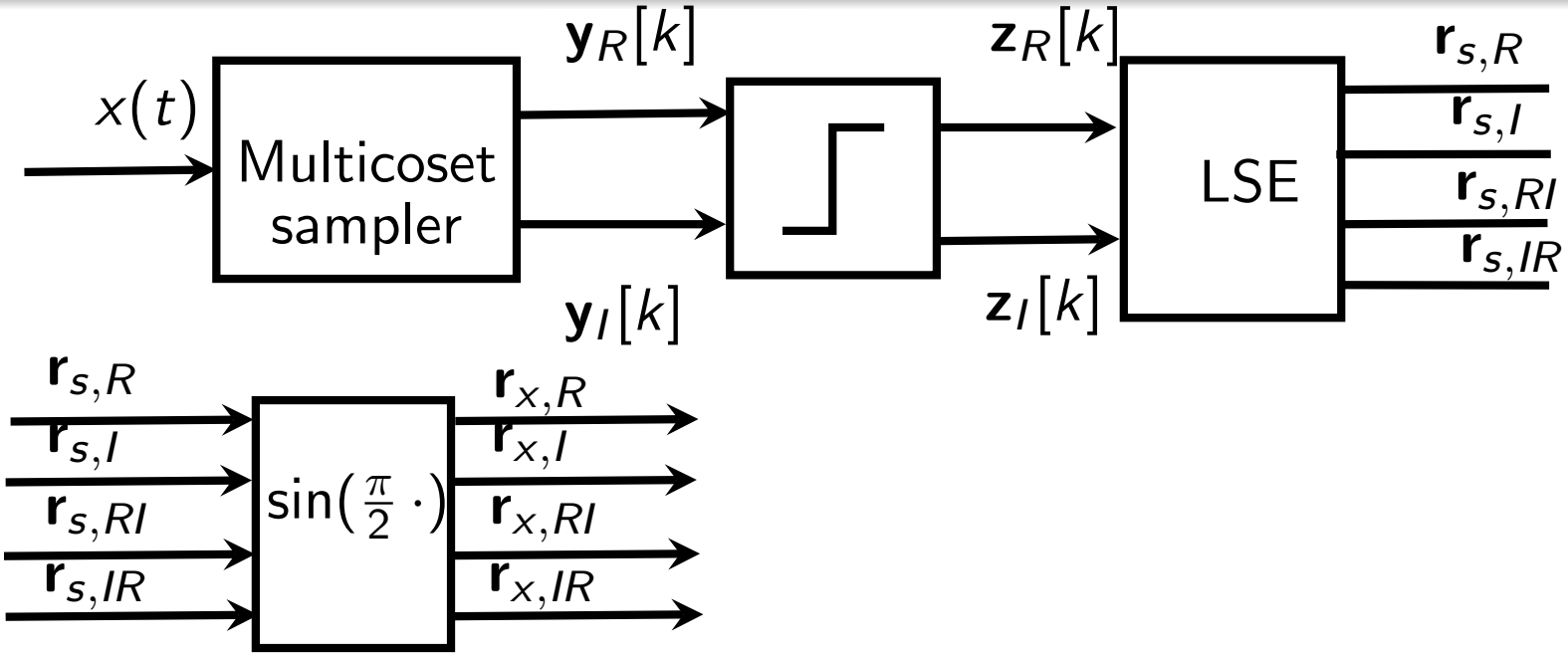
Fernando de la Hucha Arce

Introduction

Least Squares with Hard Limiting

Maximum Likelihood

Random Sampling



The block LSE applies the least squares method:

$$\text{vec}(\mathbf{R}_{z_R}) = \mathbf{\Theta}\,\mathbf{r}_{s,R}, \qquad \text{vec}(\mathbf{R}_{z_I}) = \mathbf{\Theta}\,\mathbf{r}_{s,I},$$
$$\text{vec}(\mathbf{R}_{z_{RI}}) = \mathbf{\Theta}\,\mathbf{r}_{s,RI}, \qquad \text{vec}(\mathbf{R}_{z_{IR}}) = \mathbf{\Theta}\,\mathbf{r}_{s,IR}.$$

The complex autocorrelation is then given by

$$\mathbf{r}_x = \mathbf{r}_{x,R} + \mathbf{r}_{x,I} + j\left(\mathbf{r}_{x,IR} - \mathbf{r}_{x,RI}\right).$$

Simulation results

Spectral Density Estimation with Non-Uniform Sampling

Fernando de la Hucha Arce

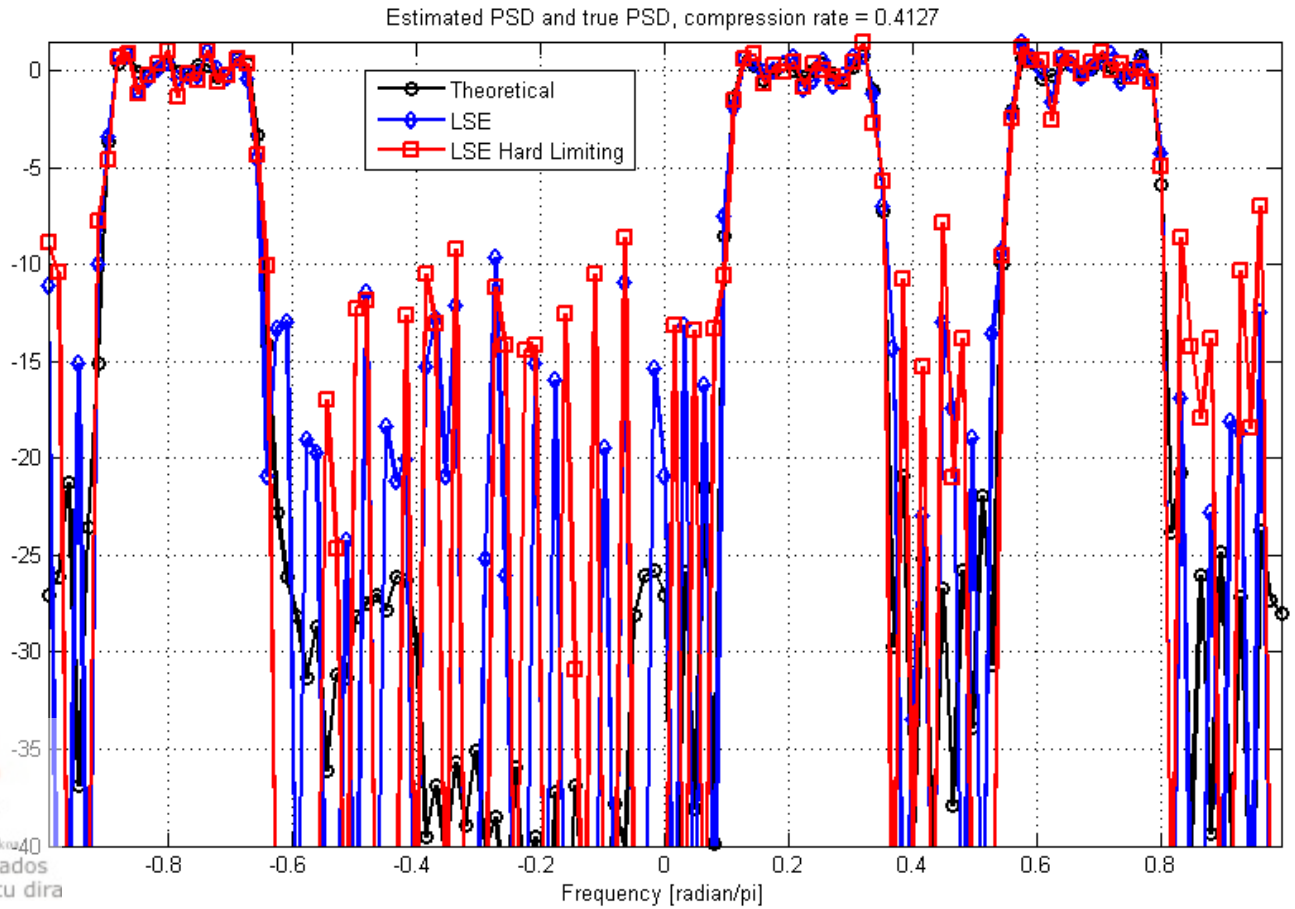
Introduction

Least Squares with Hard Limiting

Maximum Likelihood

Random Sampling

The sampling scheme employed is multicoset and the number of vectors generated is $K = 1024$.



Simulation results

Spectral Density Estimation with Non-Uniform Sampling

Fernando de la Hucha Arce

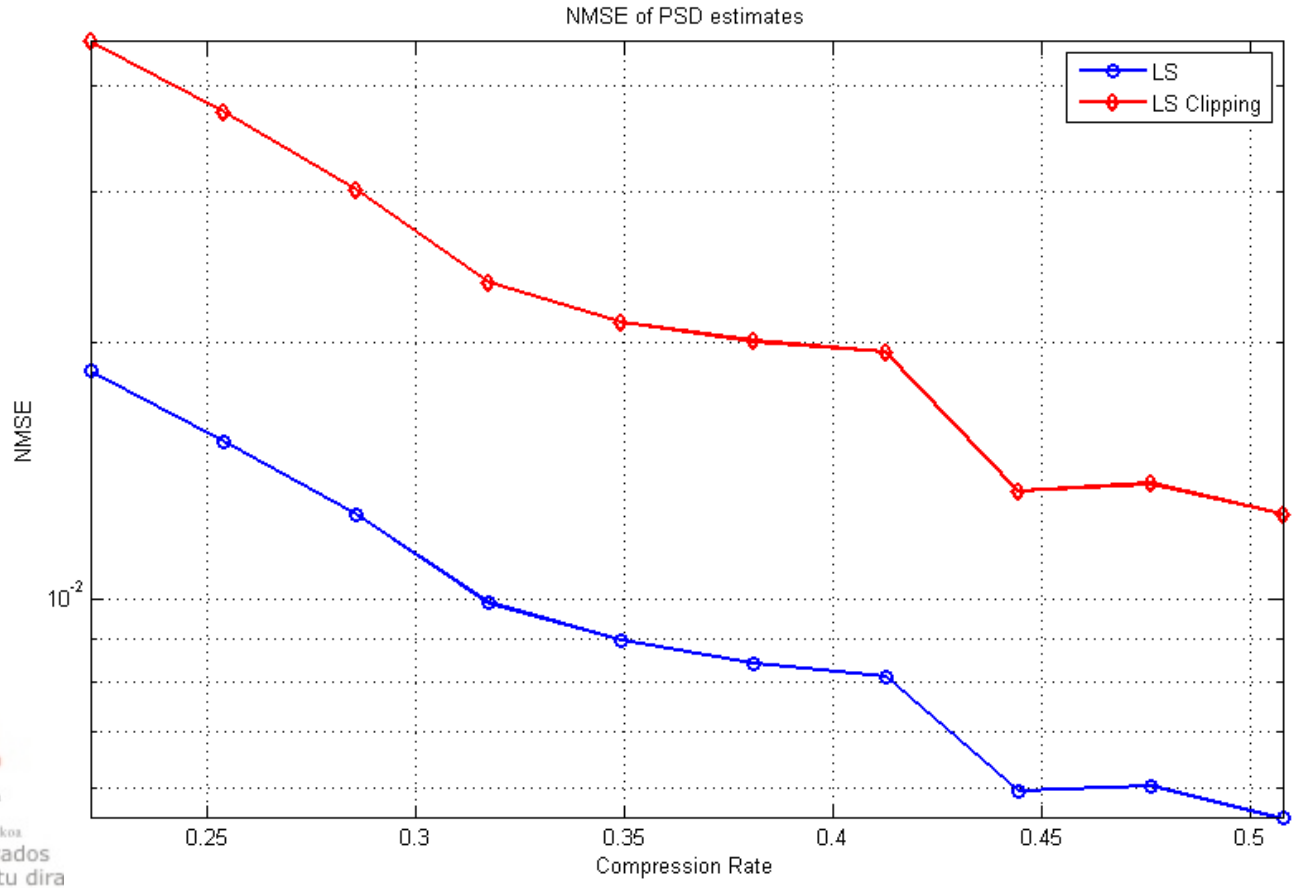
Introduction

Least Squares with Hard Limiting

Maximum Likelihood

Random Sampling

The mean squared error with hard limiting is worse than applying the least squares method without hard limiting.



Maximum Likelihood

Spectral Density Estimation with Non-Uniform Sampling

Fernando de la Hucha Arce

Introduction

Least Squares with Hard Limiting

Maximum Likelihood

Random Sampling

In this method, the signal $x(t)$ is assumed to come from a zero-mean real Gaussian process.
The distribution of $\mathbf{x}[k]$ is multivariate Gaussian, with pdf given by

$$f(\mathbf{x}[k]|\mathbf{R}_x) = \frac{1}{(2\pi)^{N/2}|\mathbf{R}_x|^{1/2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\mathbf{R}_x^{-1}\mathbf{x}[k]\mathbf{x}^T[k]\right)\right\}.$$

Because $x(t)$ is real, $\mathbf{R}_x$ is symmetric, and it suffices to estimate

$$\tilde{\mathbf{r}}_x = [r_x[0], r_x[1], \ldots, r_x[N-1]]^T$$

Compressive Power Spectral Density Estimation with Non-Uniform Sampling

Fernando de la Hucha Arce

Introduction

Least Squares with Hard Limiting

Maximum Likelihood

Random Sampling

Now, $\mathbf{R}_x$ can be expressed in a basis for the square symmetric Toeplitz matrices of order N,

$$\mathbf{R}_x = \sum_{m=0}^{N-1} r_x[m]\mathbf{B}_m.$$

These matrices have all zeros but in the m-th lower and upper diagonals, where they have ones. For example, $\mathbf{B}_1$ is given by

$$\mathbf{B}_1 = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ldots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & 0 \end{bmatrix}.$$

ML with compressive sampling

In compressive sampling, $\mathbf{y}[k]$ and $\mathbf{x}[k]$ are linearly related, and thus $\mathbf{y}[k]$ is also an $M \times 1$ Gaussian random vector with zero mean and covariance matrix

$$\mathbf{R}_y = \mathbf{C}\mathbf{R}_x\mathbf{C}^T = \sum_{m=0}^{N-1} r_x[m]\,\mathbf{C}\mathbf{B}_m\mathbf{C}^T.$$

The likelihood function is the pdf of $\mathbf{y}[k]$ given $\tilde{\mathbf{r}}_x$,

$$f(\mathbf{y}[k]\,|\,\tilde{\mathbf{r}}_x) = \frac{1}{(2\pi)^{M/2}|\mathbf{R}_y|^{1/2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\mathbf{R}_y^{-1}\mathbf{y}[k]\mathbf{y}^T[k]\right)\right\}.$$

The vector $\tilde{\mathbf{r}}_x$ that maximizes this function for a given set of measurements $\mathbf{y}[k]$ is the maximum likelihood estimator (MLE).

Compressive Power Spectral Density Estimation with Non-Uniform Sampling

Fernando de la Hucha Arce

Introduction

Least Squares with Hard Limiting

Maximum Likelihood

Random Sampling

Setting the gradient of $\ln f(\mathbf{y}[k] \,|\, \tilde{\mathbf{r}}_x)$ to zero yields

$$\operatorname{tr}\left\{\mathbf{R}_y^{-1}\mathbf{C}\mathbf{B}_n\mathbf{C}^T\mathbf{R}_y^{-1}\mathbf{R}_y\right\} = \operatorname{tr}\left\{\mathbf{R}_y^{-1}\mathbf{C}\mathbf{B}_n\mathbf{C}^T\mathbf{R}_y^{-1}\mathbf{y}[k]\mathbf{y}^T[k]\right\}.$$

which has to hold $\forall n \in \{0, 1, \ldots, N-1\}$.

- There is no known closed form solution for this system of equations, so numerical methods have to be employed.
- We can collect K measurement vectors in the $M \times K$ matrix $\mathbf{Y} = [\mathbf{y}[0], \mathbf{y}[1], \ldots, \mathbf{y}[K-1]]$.
- If we replace $\mathbf{y}[k]\mathbf{y}^T[k]$ with a consistent estimator of $\mathbf{R}_y$, the equations would hold when $K \to \infty$.
- If $x(t)$ is ergodic, then a consistent estimator of $\mathbf{R}_y$ is $\mathbf{S}_y = \frac{1}{K}\mathbf{Y}\mathbf{Y}^T$.

Asymptotic ML solutions

- Asymptotic Maximum Likelihood:

$$\sum_{m=0}^{N-1} \hat{r}_x[m] \operatorname{tr}\left\{\mathbf{S}_y^{-1}\mathbf{C}\mathbf{B}_n\mathbf{C}^T\mathbf{S}_y^{-1}\mathbf{C}\mathbf{B}_m\mathbf{C}^T\right\} = \operatorname{tr}\left\{\mathbf{S}_y^{-1}\mathbf{C}\mathbf{B}_n\mathbf{C}^T\right\},$$

$$\forall n \in \{0, 1, \ldots, N-1\}.$$

- Simplified Inverse Iteration Algorithm:

$$\sum_{m=0}^{N-1} \hat{r}_x^{(k+1)}[m] \operatorname{tr}\left\{(\hat{\mathbf{R}}_y^{(k)})^{-1}\mathbf{C}\mathbf{B}_n\mathbf{C}^T(\hat{\mathbf{R}}_y^{(k)})^{-1}\mathbf{C}\mathbf{B}_m\mathbf{C}^T\right\} =$$

$$\operatorname{tr}\left\{(\hat{\mathbf{R}}_y^{(k)})^{-1}\mathbf{C}\mathbf{B}_n\mathbf{C}^T(\hat{\mathbf{R}}_y^{(k)})^{-1}\mathbf{S}_y\right\},$$

$$\forall n \in \{0, 1, \ldots, N-1\}.$$

Simulation results

Spectral Density Estimation with Non-Uniform Sampling

Fernando de la Hucha Arce

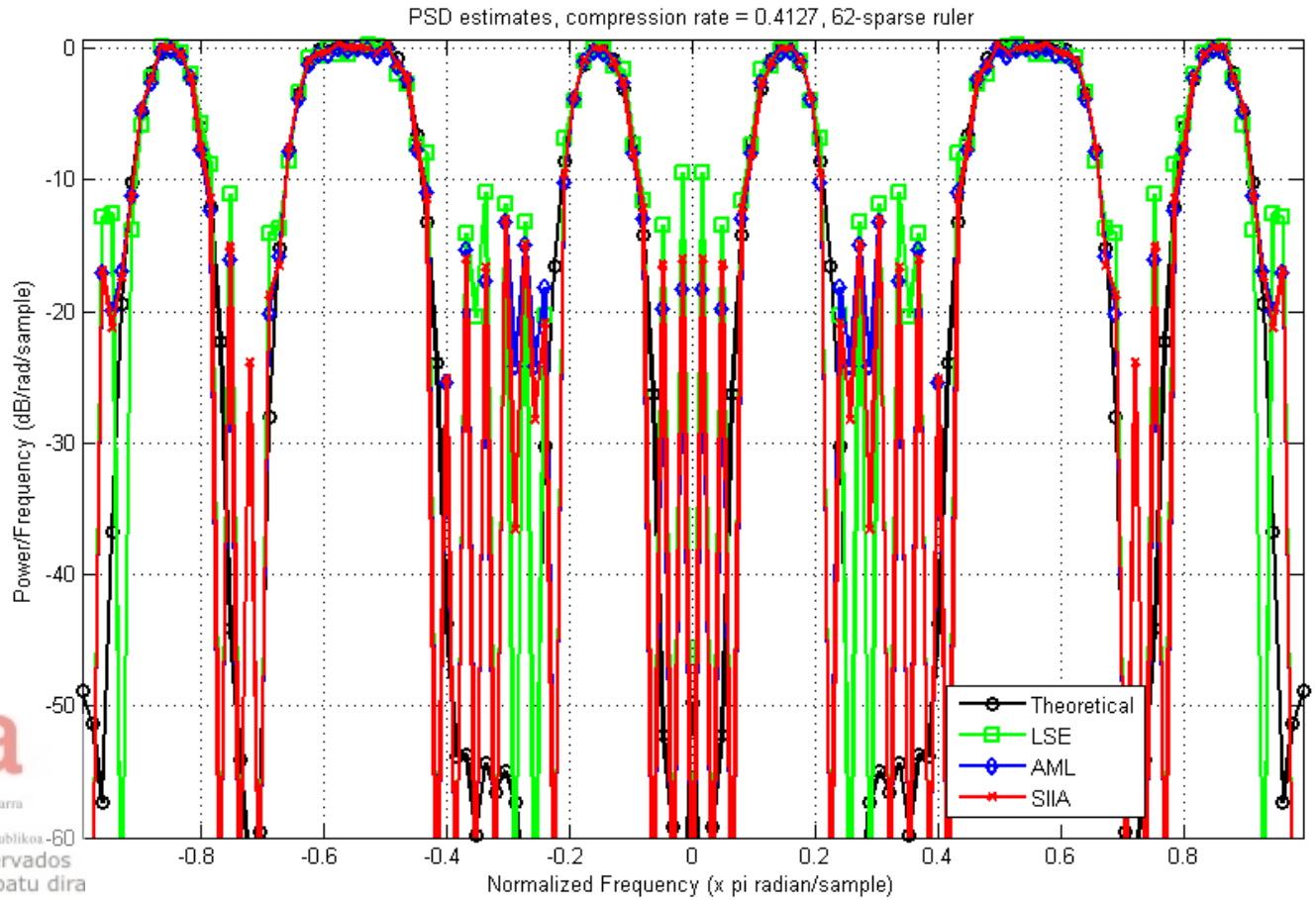
Introduction

Least Squares with Hard Limiting

Maximum Likelihood

Random Sampling

The sampling scheme employed is multicoset and the number of vectors generated is $K = 1024$.



Simulation results

Spectral Density Estimation with Non-Uniform Sampling

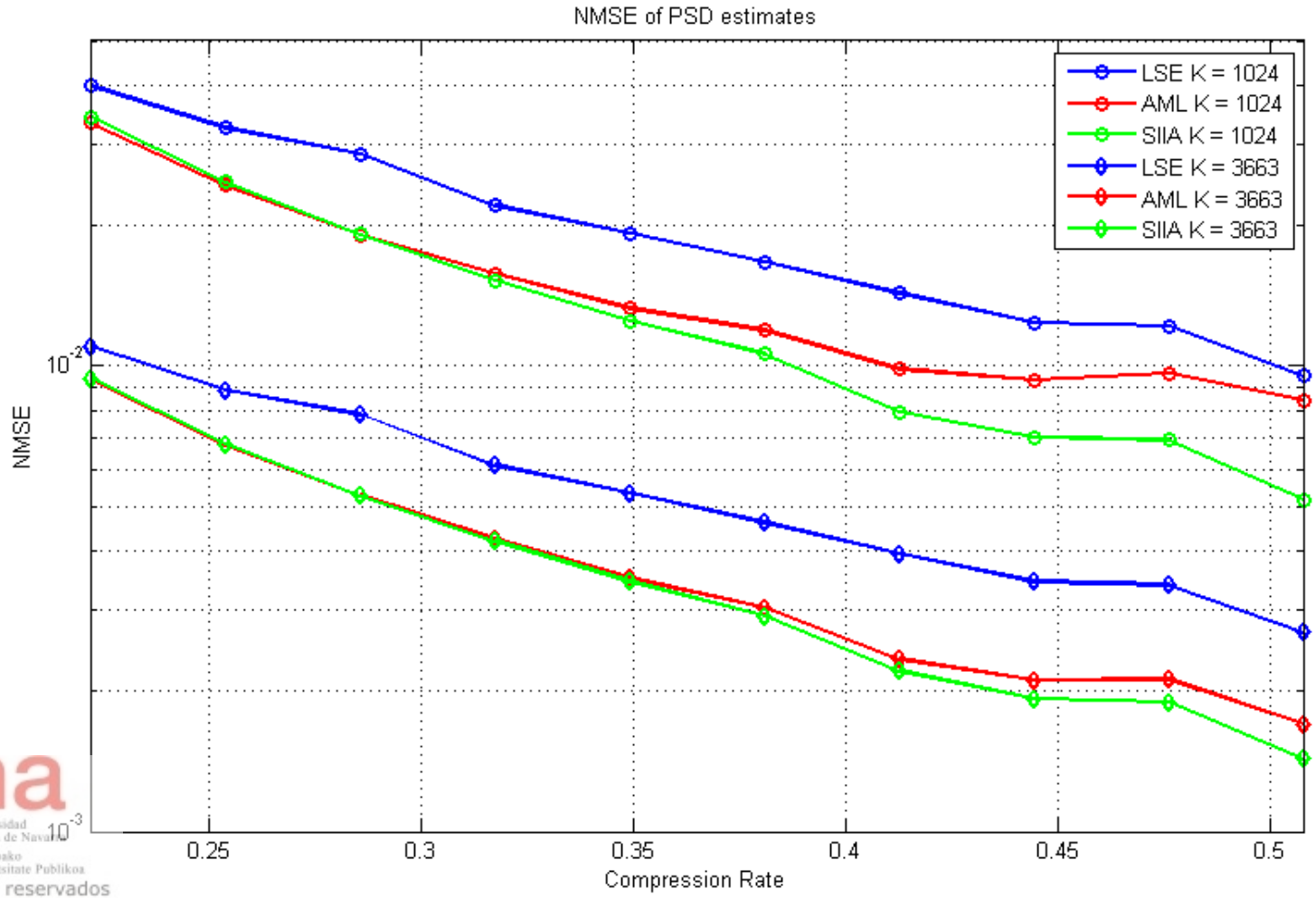
Fernando de la Hucha Arce

Introduction

Least Squares with Hard Limiting

Maximum Likelihood

Random Sampling



Simulation results

Spectral Density Estimation with Non-Uniform Sampling

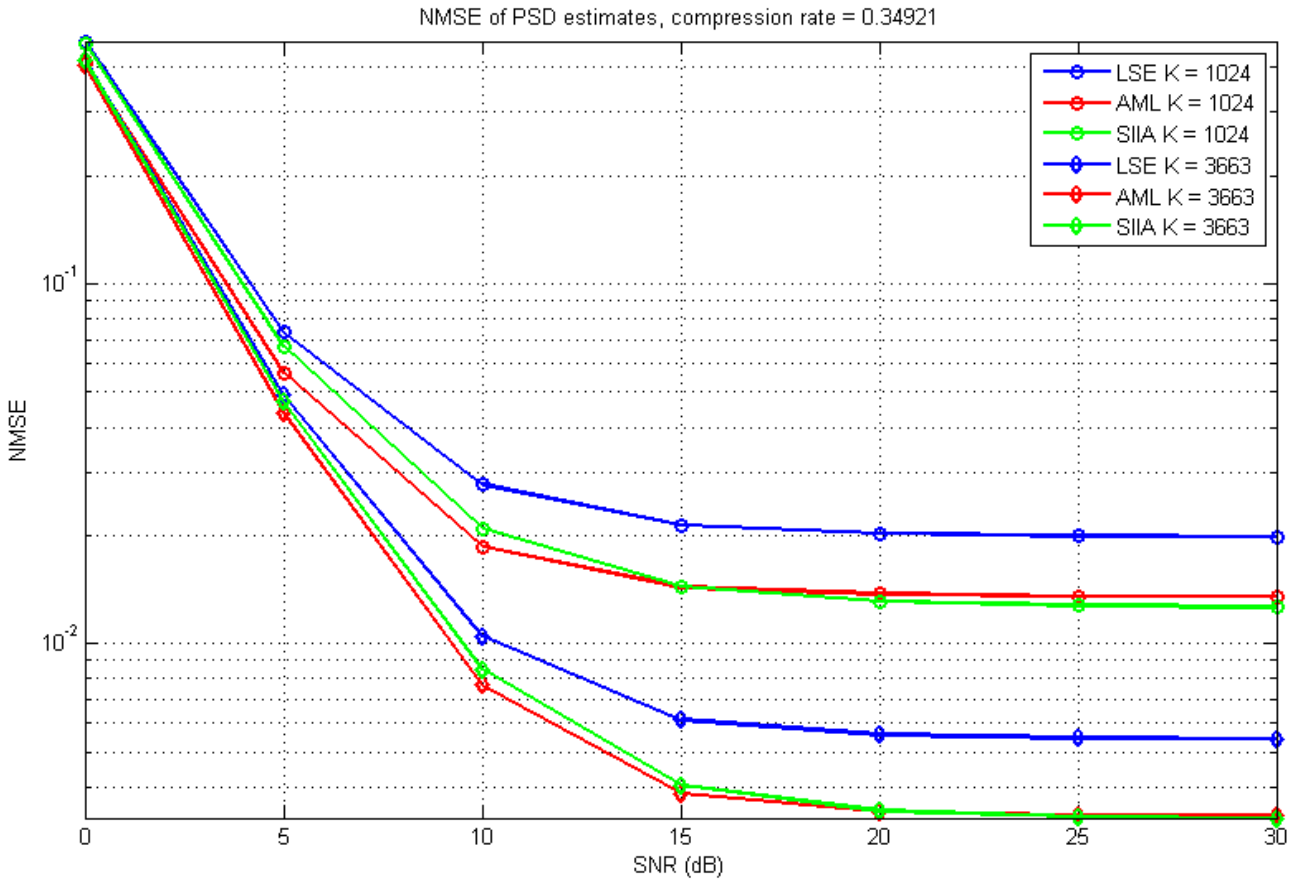
Fernando de la Hucha Arce

Introduction

Least Squares with Hard Limiting

Maximum Likelihood

Random Sampling



Both AML and SIIA perform better than Least Squares regarding compression rate and noise.

Random sampling

Spectral Density Estimation with Non-Uniform Sampling

Fernando de la Hucha Arce

Introduction

Least Squares with Hard Limiting

Maximum Likelihood

Random Sampling

In random sampling, the time between samples is a random variable:

$$t_0 = 0,$$

$$t_n = t_{n-1} + \tau_n, \quad n \in \{1, 2, \dots\}.$$

The random variables τ_n are independent and identically distributed (i.i.d). Equivalently, we can write

$$t_n = t_0 + \sum_{i=1}^{n} \tau_i\,, \quad n \in \{1, 2, \dots\}.$$

- Research in random sampling was started to avoid aliasing.
- Consistent estimation from a finite number of samples is possible, depending on the distribution of τ_n.
- This holds even with an average sampling rate lower than the Nyquist rate [2].

Modified random sampling

Spectral Density Estimation with Non-Uniform Sampling

Fernando de la Hucha Arce

Introduction

Least Squares with Hard Limiting

Maximum Likelihood

Random Sampling

- Regardless of the sampling rate, random sampling allows two samples to be infinitely close.
- In order to avoid this, we round each sampling instant to the nearest Nyquist rate instant,

$$k_0 = 0,$$

$$k_n = \text{round}\left(k_0 + \sum_{i=1}^{n} \tau_i\right), \quad n \in \{1, 2, \dots\},$$

- We collect a finite number of samples N, so

$$\max_n \{k_n\} \leq N - 1.$$

- Rounding allows repeated indexes, so we eliminate them, yielding

$$k_0 < k_1 < \ldots < k_{M-1}.$$

PSD estimation with random sampling

Spectral Density Estimation with Non-Uniform Sampling
Fernando de la Hucha Arce
Introduction
Least Squares with Hard Limiting
Maximum Likelihood
Random Sampling

We estimate first the autocorrelation by

$$\hat{r}_x[l] = \frac{1}{N_l} \sum_{\forall k_n,\,(k_n+l)\in\mathcal{S}} x^*[k_n]x[k_n+l], \quad l \geq 0,$$

where $\mathcal{S}$ denotes the set of sampling indexes.

- For negative lags we use the Hermitian symmetry: $r_x^*[l] = r_x[-l]$.
- The scaling factor is the number of products used to compute $\hat{r}_x[l]$, so this estimator is unbiased.
- In practice, we estimate $r_x[l]$ until lag L.

$$\hat{\mathbf{r}}_x = [\hat{r}_x[0], \ldots, \hat{r}_x[L-1], \hat{r}_x[1-L], \ldots, \hat{r}_x[-1]]^T.$$

- Then we can use the DFT to estimate the PSD:

$$\hat{\phi}_x = \mathbf{F}_{2L-1}\,\hat{\mathbf{r}}_x.$$

Compressive Power Spectral Density Estimation with Non-Uniform Sampling

Fernando de la Hucha Arce

Introduction

Least Squares with Hard Limiting

Maximum Likelihood

Random Sampling

We considered two random sampling schemes in our work

- Poisson sampling: the sampling instants form a Poisson process, so τ_n has an exponential distribution, with probability density function

$$f_{\tau_n}(t) = \lambda e^{-\lambda t}; \quad t \geq 0.$$

- Random permutation: this scheme selects M out of N Nyquist samples at random, and each sample has the same probability of being selected.

Simulation results

Spectral Density Estimation with Non-Uniform Sampling

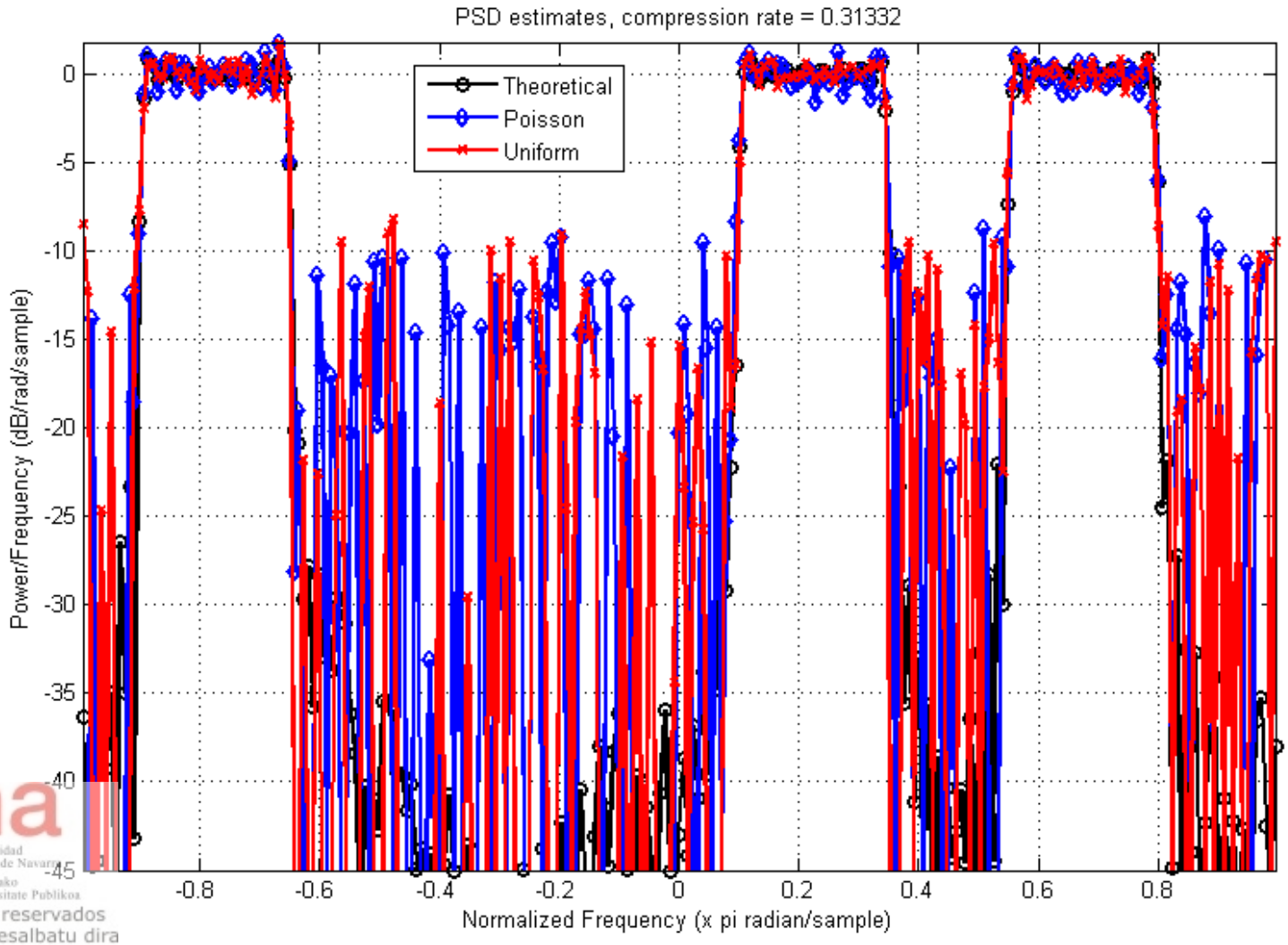
Fernando de la Hucha Arce

Introduction

Least Squares with Hard Limiting

Maximum Likelihood

Random Sampling



Compressive Power Spectral Density Estimation with Non-Uniform Sampling

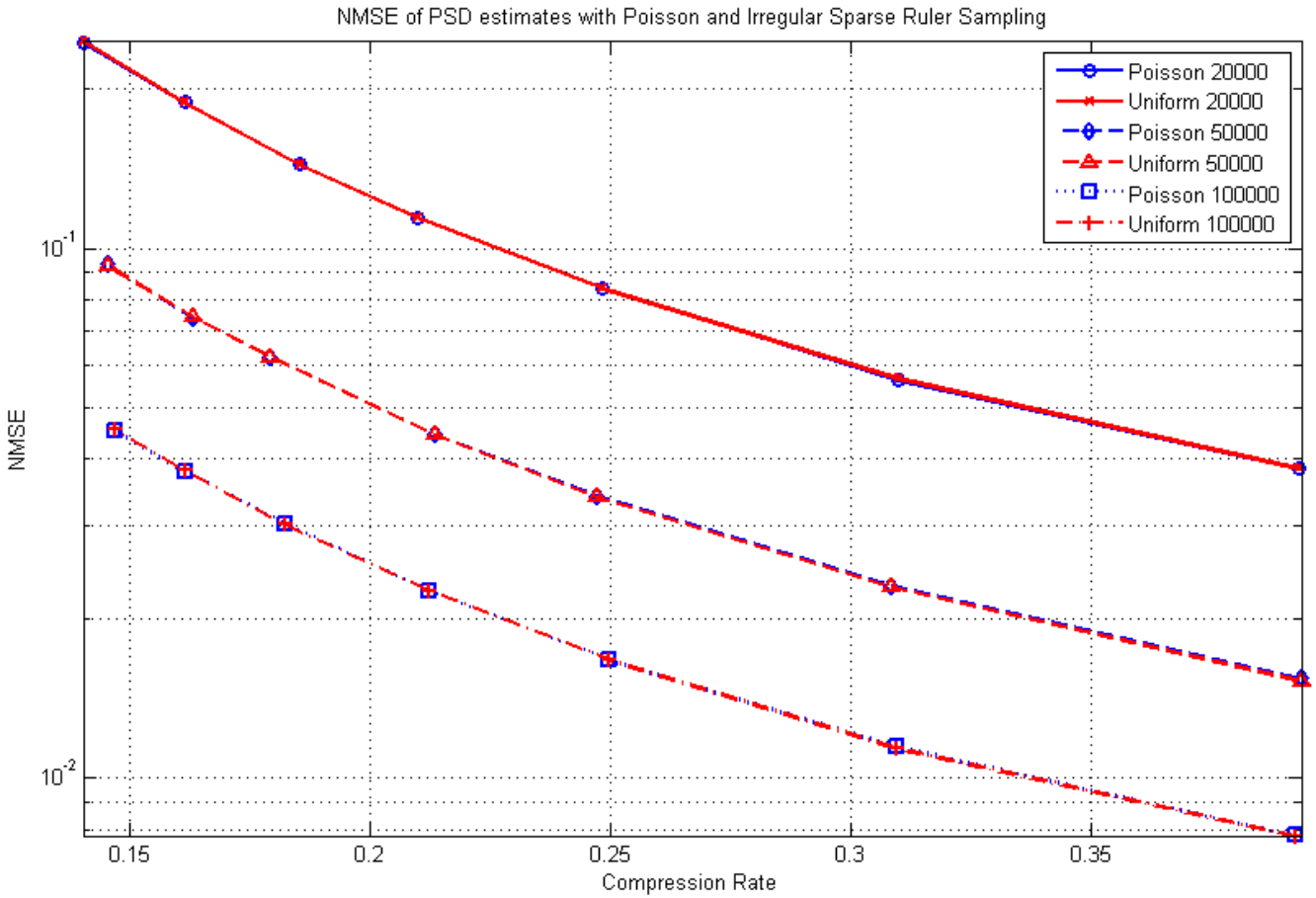
Fernando de la Hucha Arce

Introduction

Least Squares with Hard Limiting

Maximum Likelihood

Random Sampling



The occurrence times of N events of a Poisson process can be represented by N i.i.d. uniform random variables sorted in increasing order.

Conclusions

Spectral Density Estimation with Non-Uniform Sampling

Fernando de la Hucha Arce

Introduction

Least Squares with Hard Limiting

Maximum Likelihood

Random Sampling

- Spectral estimation with reduced sampling rates provides acceptable results.
- The methods trade off accuracy and computational complexity.
- The methods that showed the best performance require the assumption that the signal is Gaussian.

Compressive Power Spectral Density Estimation with Non-Uniform Sampling

Fernando de la Hucha Arce

Introduction

Least Squares with Hard Limiting

Maximum Likelihood

Random Sampling

THANKS!

Bibliography I

D.D. Arianandda and G. Leus,
Compressive Wideband Power Spectrum Estimation.
IEEE Transactions on Signal Processing, vol.60, no.9, pp. 4775-4789, September 2012

E. Masry,
Alias-Free Sampling: An Alternative Conceptualization and Its Applications,
IEEE Transactions on Information Theory, May 1978.