Compressive Power Spectral Analysis

Ph.D. Thesis

Dyonisius Dony Arianandа

Compressive Power Spectral Analysis

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof. ir. K.Ch.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op
donderdag, 19 maart 2015 om 15.00 uur

Door

Dyonisius Dony ARIANANDA
elektrotechnisch ingenieur (ir) geboren te Semarang, Jawa Tengah, Indonesie.

This dissertation has been approved by the
promotor: Prof. dr. ir. G.J.T. Leus

Composition of the doctoral committee:

Rector Magnificus
Prof. dr. ir. G.J.T. Leus promotor

Independent members:

Prof. dr. ir. A.-J. van der Veen	EEMCS, Technische Universiteit Delft
Prof. dr. ir. P.F.A. Van Mieghem	EEMCS, Technische Universiteit Delft
Prof. Dr. L. Jacques	Université catholique de Louvain, Belgium
Prof. Dr. R. López Valcarce	Universida de Vigo, Spain
Prof. Dr. V. Koivunen	Aalto University, Finland
Prof. Dr. Y.C. Eldar	Technion Israel Institute of Technology, Haifa, Israel
Prof. dr. K.G. Langendoen	EEMCS, Technische Universiteit Delft (reserve)

ISBN # 978-94-6186-441-3

Summary

At the heart of digital signal processing (DSP) are the sampling and quantization processes, which convert analog signals into discrete samples and which are implemented in the form of analog to digital converters (ADCs). In some recent applications, there is an increased demand for DSP applications to process signals having a very wide bandwidth. For such signals, the minimum allowable sampling rate is also very high and this has put a very high demand on the ADCs in terms of power consumption. Recently, the emergence of compressive sampling (CS) has offered a solution that allows us to reconstruct the original signal from samples collected from a sampling device operating at sub-Nyquist rate. The application of CS usually involves applying an additional constraint such as a sparsity constraint on the original signal. However, there are also applications where the signal to deal with has a high bandwidth (and thus sub-Nyquist rate sampling is still important) but where only the second-order statistics (instead of the original signal) are required to be reconstructed. In the latter case, depending on the characteristics of the signals, it might be possible to reconstruct the second-order statistics of the received analog signal from its sub-Nyquist rate samples without applying any additional constraints on the original signals. This idea is the key starting point of this thesis.

We first focus on time-domain wide-sense stationary (WSS) signals and introduce a method for reconstructing their power spectrum from their sub-Nyquist rate samples without requiring the signal or the power spectrum to be sparse. Our method is examined both in the time- and frequency-domain and the solution is computed using a simple least-squares (LS) approach, which produces a solution if the rank condition of the resulting system matrix is satisfied. To satisfy this rank condition, two options of sampling design are proposed, one of which is the so-called multi-coset sampling. It is show in this thesis that any of the so-called sparse

ruler can produce a multi-coset sampling design that guarantees the full rank condition of the system matrix, and thus the optimal compression is achieved by a minimal sparse ruler.

While the approach in the previous paragraph is related to time-domain signals, we could extend the discussion about the power spectrum reconstruction from sub-Nyquist rate samples in the context of the spatial-domain signal, which is defined as a sequence of outputs of the antennas in the antenna array at a particular time instant. Given the compressed spatial domain signals, which are obtained from the output of a uniform linear array (ULA) with some antennas turned off, of particular interest is to reconstruct the angular power spectrum, from which the direction of arrival (DOA) of the sources can generally be located. In this thesis, a method to estimate the angular power spectrum and the DOA of possibly fully correlated sources based on second-order statistics of the compressed spatial-domain signals is proposed by employing a so-called dynamic array which is built upon the so-called *underlying* ULA. In this method, we present the spatial correlation matrices of the output of the dynamic active antenna arrays at all time slots as a linear function of the spatial correlation matrix of the entire underlying uniform array and we solve for this last correlation matrix using LS. The required theoretical condition to ensure the full column rank condition of the system matrix is formulated and designs are proposed to satisfy this condition.

Next, we consider both spatio-angular and time-frequency domains and propose a compressive periodogram reconstruction method as our next contribution. We introduce the multibin model, where the entire band is divided into equal-size bins such that the spectra at two frequencies or angles, whose distance is at least equal to the bin size, are uncorrelated. This model results in a circulant structure in the so-called coset correlation matrix, which enables us to introduce a strong compression. We propose the sampling patterns based on a circular sparse ruler to guarantee the full column rank condition of the system matrix and to allow the LS reconstruction of the periodogram. We also provide a method for the case when the bin size is reduced such that the spectra at two frequencies or angles, whose distance is larger than the bin size, can still be correlated.

To combine frequency and DOA processing, we also introduce a compressive two-dimensional (2D) frequency- and angular-domain power spectrum reconstruction for multiple uncorrelated time-domain WSS signals received from different sources by a linear array of antennas. We perform spatial-domain compression by deactivating some antennas in an underlying ULA and time-domain compression by multi-coset sampling.

Finally, we propose a compressive cyclic spectrum reconstruction approach for wide-sense cyclostationary (WSCS) signals, where we consider sub-Nyquist rate samples produced by non-uniform sampling. This method is proposed after first observing that the block Toeplitz structure emerges in the WSCS signal correlation matrix. This structure is exploited to solve the WSCS signal correlation matrix by LS. The condition for the system matrix to have full column rank is provided and some possible non-uniform sampling designs to satisfy this full column rank condition are presented.

Based on all the works that have been done, we have found that focusing on reconstructing the statistical measure of the received signals has significantly relax the sampling requirements and the constraints on both the statistics and the signals themselves. Hence, we would like to conclude that, for given tasks of applications in hand, we should ask ourselves whether statistical measure reconstruction is sufficient since the answer for this question will likely to determine how we should collect the data from the observed phenomena. This underlines the importance of awareness on what kind of information is necessary and sufficient for the tasks in hand before conducting the sensing/sampling process.

Contents

List of Figures

List of Tables

Part I: Preamble

Chapter 1

Introduction

The main focus of this thesis is on the reconstruction of the second-order statistics from digital samples produced by compressive sampling a.k.a. sub-Nyquist-rate sampling. In this thesis, we use the term second-order statistics to refer to the auto- and cross-corelation function of the signals. However, we also use the term second-order statistics to refer to the power spectrum, which is the Fourier transform of the auto-correlation function of wide-sense stationary (WSS) signals, and also to refer to the cyclic spectrum of cyclostationary signals (see Chapter 7 for more details). Note that it has been known that compressive sampling offers substantial assistance in sampling rate reduction, which is important when we deal with signals having a very large bandwidth. In this chapter, we present the motivation of this thesis and provide an outline of our works.

1.1 Motivation

Digital signal processing has played a major role in the emergence of many applications that offer an improvement in the quality of human life. One crucial point in digital signal processing is of course the sampling and quantization process, which transforms real world analog signals into discrete samples through the use of an analog to digital converter (ADC). Some applications might require the reconstruction of the original analog signal from the digital samples or in other words, they require that a one-to-one mapping exists between the analog signal and its digital samples. According to the classical Shannon-Nyquist-Whittaker-Kotelnikov sampling theorem [2], [3], a band limited *real* signal $x(t)$, which is a signal having

$X(f) = 0$ for $f > f_{\max}$ and $f < -f_{\max}$ (Hertz) with $X(f)$ the frequency-domain representation of $x(t)$ and $f_{\max}$ a positive number, can be fully reconstructed from its samples $x(nT)$ if the sampling frequency $f_s = \frac{1}{T}$ satisfies $f_s = \frac{1}{T} \geq 2f_{\max}$, i.e., the sampling rate of a real analog signal should be at least twice the maximum frequency. When the signal bandwidth is very large, the sampling rate required by the Nyquist criterion (called Nyquist rate) that needs to be performed by the ADC is also very large, leading to a high power consumption. Finding a solution that allows us to disobey the Nyquist criterion and to sample the analog signal below the Nyquist rate is thus desirable since this will alleviate the strict requirements on the ADC. When the signal is sparse in a particular basis, one can consider the popular sampling theory known as compressive sampling (CS) [4, 5] in which the signal is linearly and randomly projected at sub-Nyquist rate leading to a limited number of measurements. Given these sub-Nyquist-rate measurements, the sparsity constraint provides a possibility to reconstruct the original analog signal with no or little information loss and this can be done by using any of the available sparse reconstruction methods (see for example, [6]).

What interests us is that there are applications where the sampling still needs to be done at sub-Nyquist rate (due to the high bandwidth of the signal of interest) but where the second-order statistics (instead of the original signal) are of interest. One application is, for instance, spectrum sensing for a cognitive radio network, which is a network where unlicensed radio systems opportunistically search for a currently unoccupied frequency band in the licensed spectrum and then borrow these discovered "white spaces" to establish a communication link. This spectrum sensing is continuously performed by these unlicensed systems since they have to monitor when the actual owners of the borrowed bands (called licensed users) become suddenly active, in which case the unlicensed radios have to vacate the spectrum. In this application, sampling the signal at sub-Nyquist rate is of interest since the spectral range that has to be sensed is generally very wide. However, note that the unlicensed radio systems are never interested in the original signal of the licensed users occupying the bands to be monitored. This implies that a power spectrum plot describing which frequency bands are occupied together with the amount of power in the occupied bands is more than enough and any efforts to reconstruct the original signal in this application will be overkill. Note that the Nyquist-rate criterion determining the minimum sampling rate is defined for signal reconstruction and not for reconstructing second-order statistics. This already indicates that the minimum sampling rate for reconstructing second-order statistics (without any additional constraints) might actually be lower than the Nyquist rate. In fact, this

is correct for WSS signals as we show in Chapters 2 and 3 of this thesis. Power spectrum reconstruction of WSS signals below the Nyquist rate is possible without any additional constraints on the original signal or the power spectrum. This is due to the fact that focusing on the second-order statistics allows us to gain more degrees of freedom. In fact, it might even possibly lead to more system equations than unknown parameters. Loosely speaking, in this thesis, the degrees of freedom refers to the number of system equations subtracted by the number of unknowns.

While the aforementioned discussion is related to time-domain signals, this discussion can also be applied to any other domain. In this thesis, we also consider *angular power spectrum* reconstruction from *spatial-domain signals*. What we define as a spatial-domain signal here is a collection of outputs of the antennas in the antenna array at a given time instant. Angular power spectrum information is important, for example, in direction of arrival (DOA) estimation of targets or in radio astronomy. In the latter application, an angular power spectral map is often constructed to describe the amount of power coming from radio sources in the sky at different directions. In our thesis, we restrict our concentration on a uniform linear array (ULA) of antennas receiving narrow-band far-field signals from multiple sources. Here, the sequence formed by the output of the antennas is perceived as digital samples of an analog spatial-domain signal. While applying the Fourier transform on the time-domain signal leads to its frequency-domain spectrum, applying the Fourier transform to the spatial-domain signal at the output of the ULA will lead to a spectrum in the angular domain but at a non-uniform grid of angles based on an inverse sinusoidal function. Similar to the time-domain signal case, where a high sampling rate is related to a high power consumption in the ADC, having more samples in the spatial domain also implies more resources, such as the number of active antennas and the related hardware receiver branches. As a result, performing a compression on the spatial-domain signal is generally desirable as it implies less spatial samples which is equivalent to a reduced number of receiver branches and/or antennas. Similar to the time-domain signal case, if the correlation of the spatial-domain samples only depends on the spatial lag and if we focus on reconstructing the angular power spectrum (which is the Fourier transform of the spatial-domain correlation) instead of the angular spectrum itself, we show later on in this thesis that the angular power spectrum can be recovered from the compressed spatial-domain samples without putting any sparsity constraint on the spatial-domain signal or on the angular power spectrum. In DOA estimation, when the distance between the sources and the ULA is much larger than the aperture of the ULA, the sources can be considered as point sources. When this is the case, it

is more common to use the output of the ULA, a.k.a. the spatial-domain samples, to perform a so-called line spectrum estimation and reconstruct a kind of a pseudo angular spectrum having sharp peaks that can be used to indicate the DOA of the sources. In some classical DOA estimation methods, the number of sources whose DOA can be estimated is generally smaller than the number of antennas in the ULA. However, DOA estimation methods that exploit second-order statistics of the sources or the spatial-domain samples could generally allow us to again gain more degrees of freedom and thus the number of sources whose DOA can be estimated can be larger than the number of antennas. These increased degrees of freedom can be obtained for example, by configuring the locations of the antennas based on a specific array geometry instead of a ULA when the sources are uncorrelated or by dynamically changing the array geometry configuration when the sources are correlated, as we will show later on in this thesis. At this stage, it is sufficient to say that what motivates us is how to explore the possibility to minimize the number of resources (antennas and/or receiver branches) for a given objective in estimating either the DOA and/or the complete angular power spectrum.

Apart from the above applications, there are some applications where the reconstruction of the second-order statistics is vital and where sub-Nyquist-rate sampling might be necessary, for example, in the case where the dimension of the covariance matrices is very large. These applications include portfolio selection, risk management, and asset pricing in the field of economics and finance (see [7]), machine learning (see [8]), and pattern recognition (see [9]). These applications are topics of future studies.

1.2 Outline and Contributions

Our thesis focuses on how to extract information out of second-order statistics from the sub-Nyquist-rate samples of the considered signal. A more detailed explanation about the outline and the contributions of our work is provided next.

Chapter 2:

In this chapter, we first provide a short overview of the classical compressive sampling problem in the context of compressive spectrum or signal reconstruction as well as parameter identification. We discuss the existing works and classify them based on how the compression is performed and what information is going to be reconstructed. By using the same classification, we then discuss some approaches

for compressively reconstructing the second-order statistics. However, for more details on the reconstruction approaches, we generally refer either to the following chapters or to some references.

Chapter 3:

This chapter introduces a compressive wideband power spectrum estimator for WSS signals sampled at sub-Nyquist rate, where any sparsity constraint on either the signal or the power spectrum is not required. This estimator is established by exploiting the Toeplitz structure in the WSS signal correlation matrix. We present our periodic sub-Nyquist sampling procedure, then examine the power spectrum reconstruction problem in the time domain and frequency domain, and propose three approaches labeled as the time-domain, alternative time-domain, and frequency-domain reconstruction approach. Two candidates for sub-Nyquist sampling implementation, namely complex Gaussian sampling and multi-coset sampling, are evaluated, where the latter can be related to the so-called sparse ruler problem. Our next contribution is the analysis of the statistical properties of the estimated power spectrum, where we calculate the mean, the covariance, and the analytical normalized mean squared error of the reconstructed power spectrum. Our power spectrum estimator can be developed into a power spectrum sensing module and a suitable detection threshold can be derived. This threshold can be computed by first refining our computed mean and covariance to the case when the received signal contains only circular complex zero-mean Gaussian i.i.d. noise.

This chapter has been published as

- D.D. Ariananda and G. Leus, "Compressive wideband power spectrum estimation", *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4775–4789, September 2012

Part of this chapter and some early results related to this chapter have also appeared in

- D.D. Ariananda and G. Leus, "Wideband power spectrum sensing using sub-Nyquist sampling", *Proc. of the 12th IEEE International Workshop on Signal Processing Advances in Wireless Communications (IEEE-SPAWC)*, pp. 101–105, San Francisco, California, June 2011

- D.D. Ariananda and G. Leus, "Compressive sampling for power spectrum estimation", *Proc. of WIC/IEEE SP Symposium on Information Theory and Signal Processing in the Benelux*, Brussel, Belgium, May 2011

- D.D. Ariananda, G. Leus, and Z. Tian, "Multi-coset sampling for power spectrum blind sensing", *Proc. of 17th International Conference on Digital Signal Processing*, Corfu, Greece, July 2011

- G. Leus and D.D. Arianandа, "Power spectrum blind sampling", *IEEE Signal Processing Letters*, vol. 18, no. 8, pp. 443–446, August 2011

Chapter 4:

While in Chapter 3, we focus on time-domain WSS signals and exploit the Toeplitz structure in their auto-correlation matrix to compressively estimate the power spectrum, in this chapter, we focus on spatial-domain signals at the output of a linear array receiving a signal from far field sources. Note that when the linear array is uniform and the sources are uncorrelated, the correlation matrix of the spatial-domain signal also has a Toeplitz structure, similar to the structure that is found in the time-domain correlation matrix in Chapter 3. This Toeplitz structure can also be exploited to estimate the angular power spectrum or even the DOA with some antennas in the ULA turned off and thus the number of required active antennas can be smaller than the number of sources (see our work in [10]). However, in this chapter, we advance this further and propose a new DOA estimation method for the case where the number of sources can be larger than the number of active receiving antennas and where the sources can be correlated. As the sources are correlated, no Toeplitz or other special structure emerges in the spatial correlation matrix and to solve this problem using only second-order statistics, we introduce the so-called dynamic (non-uniform) array of active antennas. This dynamic array is formed by considering an underlying uniform array and perform a periodic scanning on top of it, where one scanning period is split into several time slots and different sets of antennas are turned off in different time slots. Note that, we can have fewer active antennas than sources in each time slot. Mathematically, this is equivalent to adopting different spatial compression matrices for different time slots. The spatial correlation matrix of the underlying array and the spatial correlation matrices of the active antenna arrays for all time slots can be related by an overdetermined system of equations. The conditions for the system of equations to be full column-rank, which allows for a least squares (LS) reconstruction of the spatial correlation matrix of the underlying array, are also provided. We then introduce two greedy algorithms for dynamic array design that satisfies the aforementioned full rank condition of the overdetermined system. Next, we tailor the existing multiple signals classification (MUSIC) algorithm and spatial smoothing procedure to our approach

to estimate the DOAs of the possibly correlated sources from the reconstructed spatial correlation matrix of the underlying array. We also provide other options, where we express the reconstructed spatial correlation matrix of the underlying array as a linear function of the correlation matrix of the incoming signals at a grid of investigated angles. Depending on the grid resolution, this system of equations is solved using either LS or sparsity-regularized LS (possibly assisted by additional constraints).

This chapter has been published as

- D.D. Ariananda and G. Leus, "Direction of arrival estimation for more correlated sources than active sensors", *Elsevier Signal Processing*, vol. 93, no. 12, pp. 3435–3448, December 2013

Chapter 5:

Unlike the previous two chapters, where we treat time-domain signals and spatial-domain signals separately, we here focus on both domains and introduce an approach to reconstruct the angular-domain periodogram from spatial-domain signals received at different time indices and the frequency-domain periodogram from time-domain signals received at different wireless sensors. Different from the previous two chapters, the entire angular or frequency band is divided into equal-size bins and the bin size is configured such that, at two angles or frequencies separated by a distance of at least equal to the bin size, we have uncorrelated spectra. We discover that these problems in the two different domains result in a similar circulant structure in the so-called coset correlation matrix and thus a strong compression is possible. We are able to find the conditions for the resulting system matrix to have full column rank and relate the design of the spatial or temporal sampling patterns that achieve this full rank condition to the so-called circular sparse ruler. This achievable full rank condition of the system matrix allows for a simple LS reconstruction method. We then provide some analysis on the statistical properties of the compressively estimated periodogram, which includes a bias and variance analysis. For angular periodogram reconstruction, the proposed procedures assume that the received signals at different time instants have the same statistics and for frequency periodogram reconstruction, it is assumed that the signals received at different sensors also have the same statistics. In order to handle more general cases, we also propose a multi-cluster model (more details in Chapter 5). In a different situation when the received spectra at two angles or frequencies, whose distance is larger than the bin size, can still be correlated, the resulting coset correlation matrix

is generally not circulant anymore. In this case, a specific solution, which is also available in Chapter 5, is required.

This chapter is submitted as

- D.D. Arianandа, D. Romero and G. Leus, "Compressive periodogram reconstruction using uniform binning", *submitted to IEEE Transactions on Signal Processing, accepted with mandatory minor revision*

Part of this chapter has also appeared in

- D.D. Arianandа, D. Romero and G. Leus, "Compressive angular and frequency periodogram reconstruction for multiband signals", *Proc. of the 5th IEEE International Workshop on Computational Advances in Multi-sensor Adaptive Processing (CAMSAP)*, pp. 440–443, Saint Martin, French West-Indies, December 2013

Chapter 6:

In this chapter, we focus on multiple uncorrelated sources transmitting WSS signals and attempt to estimate the two-dimensional (2D) power spectrum in both frequency and DOA after performing sub-Nyquist sampling on the received signals in both the time and spatial domain. Using this 2D power spectrum, we can locate the operating frequency and the DOA of the sources despite sampling at sub-Nyquist rate and although we have fewer active antennas than sources. For this purpose, we integrate the frequency-domain power spectrum estimation approach of Chapter 3 and our work in [10] on compressive angular-domain power spectrum estimation into a single module. Note that we here do not adopt the angular power spectrum estimation approach of Chapter 4 since Chapter 4 focuses on correlated sources. We first consider a ULA as the underlying array where each antenna is connected to a receiver performing time-domain sub-Nyquist-rate sampling implemented using a multi-coset sampling device introduced in Chapter 3 leading to temporal compression. The compression in the spatial domain is then implemented by turning off some of the antennas in the underlying ULA. In our approach, both the spatial and temporal correlation functions between the resulting sub-Nyquist-rate samples at all receivers are computed. We are basically able to write the linear relationship between these correlation values with the 2D power spectrum. We present the conditions that ensure the full column rank of the system matrix, which allows for a simple LS reconstruction without applying any sparsity constraint on the signal

statistics. Once the 2D power spectrum is reconstructed, we can further estimate the DOAs of the sources by locating the peaks of the angular power spectrum.

This chapter has been published as

- D.D. Ariananda and G. Leus, "Compressive joint angular-frequency power spectrum estimations", *Proc. of the 21st European Signal Processing Conference (EUSIPCO 2013)*, Marrakech, Morocco, September 2013

Chapter 7:

In this chapter, we focus on estimating the cyclic spectrum of wide-sense cyclostationary (WSCS) signals from sub-Nyquist-rate samples collected using non-uniform sampling. Note that, unlike the correlation matrix of WSS signals which is the focus of Chapter 3, the correlation matrix of WSCS signals does not have a Toeplitz structure. However, we here decide to adopt the trick of [11] that sets the span of the random linear projection to an integer multiple of the cyclic period, which allows us to exploit the resulting block Toeplitz structure of the WSCS signal correlation matrix. Note that this chapter can also be related to Chapter 4, where we apply different spatial compression matrices in different time slots. Here, by setting the span of the random linear projection to an integer multiple of the cyclic period, we generally apply different temporal compression matrices for different cyclic periods in one span of the projection. This allows us to write the linear relationship between the WSCS signal correlation matrix and the correlations of the sub-Nyquist-rate samples as an overdetermined system. We discover the conditions for the system matrix to have full column rank allowing for a LS reconstruction of the WSCS signal correlation matrix from the correlations of the sub-Nyquist-rate samples. We also present the full column rank conditions for the system matrix when each of the sampling matrices is restricted to either an identity matrix or an empty matrix and relate the sampling matrix design for this special case to a sparse ruler. Next, we also evaluate the case when the support of the WSCS signal correlation is limited and propose a greedy algorithm for sampling matrix design that satisfies the full rank condition of the system matrix. For this limited support case, we again look at the case where each sampling matrix is restricted to either an identity matrix or an empty matrix and for this case, we are able to relate the sampling matrix design to a circular sparse ruler.

This chapter has been published as

- D.D. Ariananda and G. Leus, "Non-uniform sampling for compressive cyclic spectrum reconstruction", *Proc. of 2014 IEEE International Conference on*

Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, pp. 41–45, May 2014

Chapter 8:

In this chapter, we present the conclusions and provide possible future researches related to our work.

Next to the chapters included in this thesis, we also made additional contributions during the last four years. First, we have extended our power spectrum estimation approach of Chapter 3 into a cooperative scenario, where we have multiple sensors sensing the same WSS user signals passing through different fading channels on their way toward different sensors. In order to have a lower sampling rate per sensor compared to the one in the single sensor case of Chapter 3, we have exploited the cross-spectra between the compressive measurements at different sensors. As in the single sensor case, this can be done without requiring the power spectrum to be sparse. The extension into a cooperative scenario for the time-domain approach of Chapter 3 is published as

- D.D. Ariananda and G. Leus, "Cooperative compressive wideband power spectrum sensing", *Proc. of the 46th Asilomar Conference on Signal, Systems and Computers*, Pacific Grove, California, pp. 303–307, November 2012

while the extension for the alternative time-domain approach of Chapter 3 is published as

- D.D. Ariananda and G. Leus, "A study on cooperative compressive wideband power spectrum sensing", *Proc. of WIC/IEEE Symposium on Information Theory and Signal Processing in the Benelux*, Boekelo, The Netherlands, pp. 102–109, May 2012

The aforementioned works however build upon the knowledge of the channel state information (CSI). This fact has encouraged us to produce another cooperative compressive wideband power spectrum sensing approach that does not rely on the knowledge of the CSI while maintaining our goal to minimize the sampling rate per sensor. In this approach, we avoid the reliance on the knowledge of the CSI by not exploiting the cross-spectra between measurements at different sensors. As a result, we need to find a different way to obtain a lower sampling rate per sensor

than the one in the single sensor case of Chapter 3. In order to achieve this goal, we organize multiple wireless sensors into several groups, where different groups of sensors employ different compression schemes and where each group computes the temporal correlation estimates only at certain lags, which are different from group to group. The temporal correlation estimates at different lags are then collected from different groups of sensors by a fusion centre, which uses them to estimate the power spectrum. This work has been published as

- D.D. Arianandа, D. Romero, and G. Leus, "Cooperative compressive power spectrum estimation", *Proc. of 2014 IEEE 8th Sensor Array and Multichannel Signal Processing Workshop (SAM 2014)*, A Coruna, Spain, pp. 97–100, June 2014

As we have mentioned in the summary of Chapter 4, the Toeplitz structure that emerges in the time-domain correlation matrix of Chapter 3, also appears in the correlation matrix of the spatial-domain signal at the output of the ULA receiving far field signals from uncorrelated sources. We also have a work that exploits this Toeplitz structure and that estimates the angular power spectrum or the DOA of the sources with some antennas in the ULA deactivated based on the solution of the sparse ruler problem introduced in Chapter 3. The resulting active antenna configuration is similar to the minimum redundancy array (MRA) of [12] (though not always exactly the same). This work has been published as

- S. Shakeri, D.D. Arianandа, and G. Leus, "Direction of arrival estimation using sparse ruler array design", *Proc. of the 13th IEEE International Workshop on Signal Processing Advances in Wireless Communications (IEEE-SPAWC)*, Cesme, Turkey, pp. 525–529, June 2012

Another contribution in the field of DOA estimation involves the combination of the dynamic array of Chapter 4 with the so-called structured total least squares (structured TLS) approach to estimate the DOA of correlated sources where the number of sources can be larger than the number of sensors. There, we consider a possible perturbation on the pre-defined grid of investigated angles of arrival (see the above summary of Chapter 4). This work has been published as

- D.D. Arianandа and G. Leus, "Direction of arrival estimation of correlated signals using a dynamic linear array", *Proc. of the 46th Asilomar Conference on Signal, Systems and Computers*, Pacific Grove, California, pp. 2028–2035, November 2012

Chapter 2

Compressive Power Spectrum Estimation: An Overview

Compressive sampling or compressed sensing (CS) is often related to solving underdetermined systems of linear equations. In this thesis report however, we attempt to provide a different perspective and show that, depending on what kind of information we are going to reconstruct, compressive sampling can also be connected to solving determined systems of linear equations if we are able to increase the so-called degrees of freedom. In fact, it might even be connected to solving overdetermined systems. In this chapter, we will start by discussing the classical compressive sampling problem in the context of compressive spectrum or signal reconstruction as well as parameter identification [such as frequency or direction of arrival (DOA) estimation]. The latter can usually be related to identifying the non-zero support of the original vector or signal. We then show that, the degrees of freedom can be increased if we relax our objective and decide to reconstruct the power spectrum instead of the spectrum from compressive samples or if we use second-order statistics for parameter identification.

2.1 Preliminaries

Let us first focus on Nyquist-rate sampling, which, in this report, can be associated not only to time-domain samples produced by a time-domain Nyquist-rate sampler [temporal sampling (TS) case] but also to spatial-domain samples at the output of a uniform linear array (ULA) of antennas with half-wavelength spacing [spatial

sampling (SS) case]. To set the stage, in the TS case, we consider one or multiple wireless sensors sensing the same users signal. For both the TS and SS cases, let us then consider an $N \times 1$ complex vector

$$\mathbf{x}_t[k] = [x_t[kN], x_t[kN+1], \ldots, x_t[kN+N-1]]^T, \tag{2.1}$$

which illustrates the $(k+1)$-th block of N consecutive time-domain Nyquist-rate samples at sensor index t for the TS case or the output of the $(k+1)$-th block of N consecutive antennas in the ULA at time index t for the SS case. Note that both k and t here are integers given by $k = 0, 1, \ldots, K-1$ and $t = 0, 1, \ldots, T-1$, respectively. This also implies that t represents a discrete quantity. It is then clear that, for the TS case, K gives the total number of blocks of N consecutive time-domain samples and T is the total number of wireless sensors. Fig. 2.1 illustrates the TS case with $K = 3$, $T = 4$, and $N = 5$. Meanwhile, for the SS case, K gives the total number of blocks of N consecutive antennas in the ULA while T is the total number of considered time indices at each antenna. Fig. 2.2 illustrates the SS case again with $K = 3$, $T = 4$, and $N = 5$.

For the TS case, we might have one or multiple sensors collecting multiple blocks of N samples (periodic temporal sampling) or we might consider one or multiple sensors but only focus on one block of N samples in each sensor (single period temporal sampling). Similarly, for the SS case, we might have single period spatial sampling (when we consider only one block of N antennas and focus on the collected samples at either only one or multiple time indices t), or periodic spatial sampling (when we consider multiple blocks of N antennas and focus on the collected samples at either only one or multiple time indices t).

In the TS case, sampling large-bandwidth signals at Nyquist rate requires a high-rate analog-to-digital converter (ADC), which is power hungry. Recently, CS has emerged as a popular solution to alleviate the ADC requirements. While the details about CS theory can be found in [5, 13, 14], we here provide a brief explanation. Mathematically, the CS concept can be described by applying a random linear projection on the received Nyquist-rate samples $\mathbf{x}_t[k]$ leading to an $M \times 1$ measurement vector $\mathbf{y}_t[k]$, which can be written as

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

with $\mathbf{C}$ the so-called $M \times N$ CS or measurement matrix and with $M < N$. Since the problem is underdetermined, solving $\mathbf{x}_t[k]$ from the known $\mathbf{y}_t[k]$ and $\mathbf{C}$ generally does not result in a unique solution unless a specific constraint is introduced.

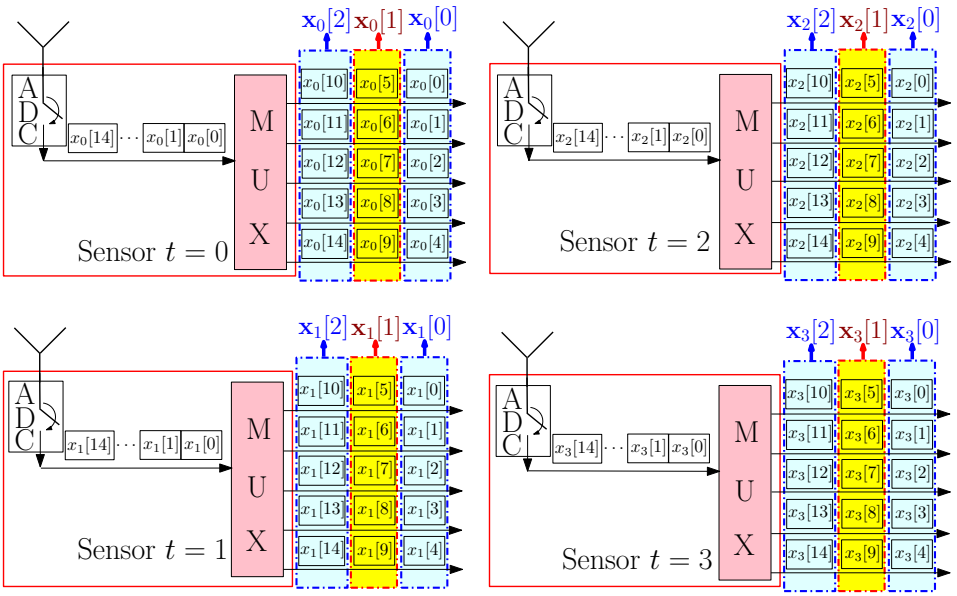


Figure 2.1: Illustration for the TS case. Here, we have $T = 4$ wireless sensors, which are assumed to sense the same user signal. Each sensor collects $K = 3$ blocks of $N = 5$ consecutive time-domain Nyquist-rate samples. Note that the multiplexer here is only used to show how (2.1) is formed.

The popular constraint that is introduced in the CS framework is a sparsity constraint. The vector $\mathbf{x}_t[k]$ is called an S-sparse vector (or a vector that has a sparsity order of S) if at least $N - S$ entries are exactly zero. Note however that we might not deal with signals that are themselves sparse, but which have a sparse representation in a particular basis [15]. For example, $\mathbf{x}_t[k]$ might not be sparse but it might be that, in the discrete Fourier transform (DFT) representation of $\mathbf{x}_t[k]$, only at most S DFT coefficents have non-zero value. In this case, we can still regard $\mathbf{x}_t[k]$ as an S-sparse signal. Most of the time, we encounter signals that are only approximately sparse instead of the ones that are truly sparse. These signals are often labeled as compressible signals, which means that they can be well-approximated by sparse signals [15]. For example, at least $N - S$ entries of $\mathbf{x}_t[k]$ (or its DFT) might have a magnitude smaller than $10^{-\gamma}$ with γ a positive number to indicate the accuracy level. In this case, we can regard $\mathbf{x}_t[k]$ as a compressible signal. Given above definition, if $\mathbf{x}_t[k]$ is a sparse signal or if it has a sparse representation in a particular basis, it is possible to obtain a unique solution of $\mathbf{x}_t[k]$ from $\mathbf{y}_t[k]$ although $M < N$ as long as $\mathbf{C}$ satisfies a specific requirement called the restricted isometry property

(RIP) [13] (see Appendix 2.B for a detailed definition of RIP). In this case, multiple sparse reconstruction techniques such as orthogonal matching pursuit [6], basis pursuit [16], and least absolute shrinkage and selection operator (LASSO) [17] are available for the reconstruction of $\mathbf{x}_t[k]$. Note that, in practice, the measurements $\mathbf{y}_t[k]$ are directly obtained from the analog version of $\mathbf{x}_t[k]$ through a so-called analog-to-information converter [18] instead of from $\mathbf{x}_t[k]$ itself.

Note that [13] suggests the use of i.i.d. random Gaussian or Bernoulli variables as the entries of $\mathbf{C}$. One implementation example for the former is available in [19] while one for the latter is available in [20]. However, note that (2.2) can also be used to model multi-coset sampling, a.k.a. periodic non-uniform sampling, discussed in [21]. In this case, $\mathbf{C}$ is a *selection matrix*, which contains M selected rows from the $N \times N$ identity matrix $\mathbf{I}_N$ and thus $\mathbf{y}_t[k]$ is obtained by selecting M entries of $\mathbf{x}_t[k]$ and discarding the remaining $N - M$ entries. More discussion on multi-coset sampling will be available in the next section.

Compressive sampling is applicable for the SS case as well. In the next section, we will focus on compressive spectrum reconstruction and parameter identification problems for both TS and SS cases. We will also consider several compressive sampling models for that particular problem.

2.2 Compressive Spectrum Reconstruction and Parameter Identification

2.2.1 Single Period Temporal or Spatial Compression

Signal or Spectrum Reconstruction

Let us start by focusing on signal or spectrum reconstruction and considering the case where we only have a single block of samples available, i.e., one value of k ($k = k'$) and where we also assume only one such measurement is available, i.e., one value of t ($t = t'$). We first concentrate on the TS case and consider compressive time-domain signal reconstruction or frequency-domain spectrum reconstruction by assuming that the received time-domain signal $\mathbf{x}_{t'}[k']$ has a sparse representation in the frequency domain (which is common for example, in spectrum sensing for cognitive radio). Then, $\mathbf{y}_{t'}[k']$ can be written as

$$\mathbf{y}_{t'}[k'] = \mathbf{C}\mathbf{x}_{t'}[k'] = \mathbf{C}\mathbf{F}_N^{-1}\check{\mathbf{x}}_{t'}[k'], \tag{2.3}$$

where $\mathbf{F}_N$ is the $N \times N$ DFT matrix and the $N \times 1$ vector $\check{\mathbf{x}}_{t'}[k'] = [X_{t'}[k', 0], X_{t'}[k', \frac{1}{N}], \ldots, X_{t'}[k', \frac{N-1}{N}]]^T$ is the DFT of $\mathbf{x}_{t'}[k']$ with $X_{t'}[k', f]$ the value of the

frequency domain representation at digital frequency f. Here, $\check{\mathbf{x}}_{t'}[k']$ is assumed sparse, i.e., its entries are non-zero only at a few positions. The model of (2.3) can also be used for the SS case where $\check{\mathbf{x}}_{t'}[k']$ gives the angular spectrum for the spatial signal $\mathbf{x}_{t'}[k']$. In all these cases, standard CS can be used to recover $\check{\mathbf{x}}_{t'}[k']$ (see for example the CS algorithms in [6, 22, 23] and in [17, 24, 25] for the noisy measurements).

Observe that only a single measurement vector (SMV) is collected in (2.3), which is $\mathbf{y}_{t'}[k']$. In CS, it is also possible to exploit multiple measurement vectors (MMVs), for example, when the signals received at different times and/or locations share the same non-zero support, i.e., they have exploitable joint sparsity structure. In the SS case, when MMVs are collected across different time indices t, i.e., $\{\mathbf{y}_t[k']\}_{t=0}^{T-1}$, we can form an $M \times T$ matrix $\mathbf{Y}[k'] = [\mathbf{y}_0[k'], \mathbf{y}_1[k'], \ldots, \mathbf{y}_{T-1}[k']]$ and write our CS model by taking (2.3) into account as

$$\mathbf{Y}[k'] = \mathbf{C}\mathbf{X}[k'] = \mathbf{C}\mathbf{F}_N^{-1}\check{\mathbf{X}}[k'], \tag{2.4}$$

with the $N \times T$ matrices $\mathbf{X}[k'] = [\mathbf{x}_0[k'], \mathbf{x}_1[k'], \ldots, \mathbf{x}_{T-1}[k']]$ and $\check{\mathbf{X}}[k'] = [\check{\mathbf{x}}_0[k'], \check{\mathbf{x}}_1[k'], \ldots, \check{\mathbf{x}}_{T-1}[k']]$. Recall that we here still focus on only one block of N consecutive antennas in our ULA model as we still have a single spatial period in our spatial sampling. Observe that, in case the signal sensed at a particular angle along different time indices t comes from the same user, the received angular domain components $\check{\mathbf{x}}_t[k']$ might still be different across different time indices t, for example, due to the existence of time-varying fading channels between the sources and the ULA and/or because we have different realizations of the user signals along index t (even if the statistics of the user signals do not change along t). However, $\{\check{\mathbf{x}}_t[k']\}_{t=0}^{T-1}$ usually share the same non-zero support. In this case, any joint sparse reconstruction method can be used to recover both the non-zero supports of all columns of $\check{\mathbf{X}}[k']$ or $\check{\mathbf{X}}[k']$ itself (see for example the CS algorithm of [1, 26, 27, 28]).

The model in (2.4) is also applicable for the TS case (only by adjusting the definition of t and k') when we have sensors $t = 0, 1, \ldots, T-1$, each of which collects one measurement vector $\mathbf{y}_t[k']$. One particular example occurs when we have a network of wireless sensors receiving user signals that pass through different fading channels on their way towards the different sensors. Observe that, although all sensors sense the same user signals, the received frequency domain components $\check{\mathbf{x}}_t[k']$ are not the same across different sensors t due to the different fading channels. Similar to the SS case however, they usually share the same non-zero support allowing the joint reconstruction of the non-zero supports of all columns of $\check{\mathbf{X}}[k']$

or $\check{\mathbf{X}}[k']$ itself. The distributed compressive spectrum sensing approach of [29] can be related to the model in (2.4) although [29] applies different CS matrices $\mathbf{C}$ for different sensors t.

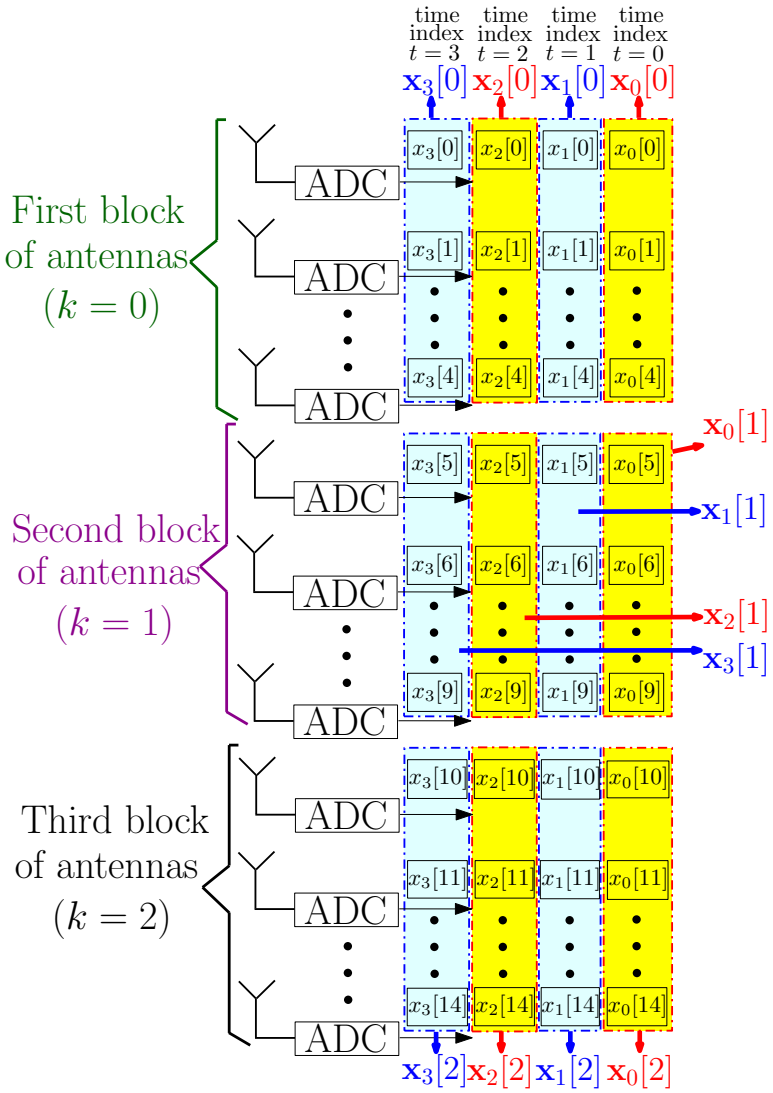


Figure 2.2: Illustration for the SS case. Here, we have a ULA of $KN = 15$ antennas, which can be split into $K = 3$ blocks of $N = 5$ consecutive antennas. At the output of the ADC connected to each antenna, digital samples are collected at $T = 4$ different time instants.

Parameter Identification

We now focus on parameter identification from compressive samples. For example, in the SS case, we might only need information about the angle of arrival of the signals instead of the entire map of the angular spectrum. When the sources whose DOAs are to be estimated can be classified as point sources, the parameter (in this case DOA) estimation problem boils down to a line spectrum estimation problem. By focusing on the DOA estimation of point sources using a ULA, let us rewrite $\mathbf{x}_{t'}[k']$ in (2.3) as

$$\mathbf{x}_{t'}[k'] = \mathbf{A}\mathbf{D}[k']\mathbf{s}_{t'}, \tag{2.5}$$

where $\mathbf{s}_{t'} = [s_{t',0}, s_{t',1}, \ldots, s_{t',I-1}]^T$ is the $I \times 1$ vector with $s_{t',i}$ one snapshot of the discrete source signal associated with the $(i+1)$-th point source and where $\mathbf{A}\mathbf{D}[k']$ is the array response matrix with the $N \times I$ matrix $\mathbf{A} = [\mathbf{a}(\varphi_0), \mathbf{a}(\varphi_1), \ldots, \mathbf{a}(\varphi_{I-1})]$, the $N \times 1$ vector $\mathbf{a}(\varphi_i) = [1, \phi(\varphi_i), \phi(\varphi_i)^2, \ldots, \phi(\varphi_i)^{N-1}]^T$, $\phi(\varphi_i) = \exp(j2\pi\varphi_i)$, φ_i given by $\varphi_i = 0.5\sin\theta_i$, θ_i the DOA of the $(i+1)$-th point source, and the $I \times I$ diagonal matrix $\mathbf{D}[k']$ having $\phi(\varphi_i)^{k'N}$ at its $(i+1)$-th diagonal element. As we only have one block of spatial samples (which is the block k'), we can include the phase terms of $\mathbf{D}[k']$ into $\mathbf{s}_{t'}$ to simplify the notation and rewrite (2.5) as

$$\mathbf{x}_{t'}[k'] = \mathbf{A}\mathbf{s}_{t'}[k']. \tag{2.6}$$

The spatial compression can then be performed on $\mathbf{x}_{t'}[k']$ as

$$\mathbf{y}_{t'}[k'] = \mathbf{C}\mathbf{x}_{t'}[k'] = \mathbf{C}\mathbf{A}\mathbf{s}_{t'}[k']. \tag{2.7}$$

While any type of matrix suggested in [13] can be used for $\mathbf{C}$ in (2.7), it is common to use a selection matrix for $\mathbf{C}$, which is equivalent to removing or deactivating some antennas in the ULA leading to a non-uniform linear array (NULA). However, it can be found in [30] that the use of a random matrix for $\mathbf{C}$ is also possible and in practice, this can be implemented by using analog phase shifters. Though the technique of [30] does not really reduce the number of active antennas, the number of receiver hardware is reduced leading to a reduced implementation budget (see [30] for more detailed information).

One of the DOA estimation approaches proposed in [31] can be perceived as following the above model although [31] technically does not mention any spatial compression and uses all the outputs of the ULA. The idea of [31] can be explained as follows. Note that the DOAs $\{\theta_i\}_{i=0}^{I-1}$ are generally unknown, which motivates [31] to introduce a grid of $Q \gg I$ investigated angles $\{\tilde{\theta}_q\}_{q=0}^{Q-1}$. In this

case, we can introduce the $N \times Q$ matrix $\tilde{\mathbf{A}}$ where the $(q+1)$-th column of $\tilde{\mathbf{A}}$ is given by $\mathbf{a}(\tilde{\varphi}_q)$. To accomodate this gridding, we can rewrite (2.7) as

$$\mathbf{y}_{t'}[k'] = \mathbf{C}\mathbf{x}_{t'}[k'] = \mathbf{C}\tilde{\mathbf{A}}\tilde{\mathbf{s}}_{t'}[k'], \tag{2.8}$$

where $\tilde{\mathbf{s}}_{t'}[k'] = [\tilde{s}_{t',0}[k'], \tilde{s}_{t',1}[k'], \ldots, \tilde{s}_{t',Q-1}[k']]^T$ is the $Q \times 1$ extended source vector with $\tilde{s}_{t',q}[k']$ the incoming signal at the investigated angle $\tilde{\theta}_q$. In [31], the reason to have a very large Q is to avoid grid mismatch, i.e., the desire is to have the actual DOAs $\{\theta_i\}_{i=0}^{I-1}$ to be on top or nearby some of the investigated angles $\{\tilde{\theta}_q\}_{q=0}^{Q-1}$. As Q is very large, we generally have $Q \gg N$ and $\tilde{\mathbf{A}}$ is generally a wide matrix. This is why the compression model in (2.8) can still be used to illustrate the framework in [31] without changing the actual concept. As the number of measurements is much less than Q, [31] applies a sparsity constraint on $\tilde{\mathbf{s}}_{t'}[k']$ in order to be able to recover $\tilde{\mathbf{s}}_{t'}[k']$ from $\mathbf{y}_{t'}[k']$. The location of the non-zero elements of $\tilde{\mathbf{s}}_{t'}[k']$ can then be used to estimate the actual DOAs. More details about this method can be found in [31]. When the number of grid points is still not sufficient enough to avoid grid mismatch, another approach that can be attempted is to introduce a kind of additive perturbation or error matrix on top of $\tilde{\mathbf{A}}$ in (2.8). More details on this approach can be found for example in [32, 33]. In [34], the compression model in (2.8) is further developed for DOA tracking instead of just DOA estimation.

Note that the aforementioned discussion about DOA estimation can also be repeated in the context of the identification of sinusoids (a.k.a. frequency estimation), i.e., when the received time-domain signal contains a sum of I sinusoids. In this case, φ_i represents the frequency of the $(i+1)$-th sinusoid.

For this case of parameter identification from compressive samples, it is also possible to exploit MMVs. For DOA estimation, one approach that exploits MMVs is also provided in [31]. In this case, we take (2.8) into account and write $\mathbf{Y}[k']$ in (2.4) as

$$\mathbf{Y}[k'] = \mathbf{C}\tilde{\mathbf{A}}\tilde{\mathbf{S}}[k'], \tag{2.9}$$

with $\tilde{\mathbf{S}}[k'] = [\tilde{\mathbf{s}}_0[k'], \tilde{\mathbf{s}}_1[k'], \ldots, \tilde{\mathbf{s}}_{T-1}[k']]$. Here, the vectors $\{\tilde{\mathbf{s}}_t[k']\}_{t=0}^{T-1}$ are sparse and they share the same non-zero support. The so-called ℓ_1 singular value decomposition (ℓ_1-SVD) is introduced by [31] to locate the locations of the non-zero support of all columns of $\tilde{\mathbf{S}}[k']$ in (2.9) and to find the DOA estimates. Other works that use the model in (2.9) can be found in [35] and [36]. The difference between [31] and [35, 36] is that [35] and [36] use the so-called joint ℓ_0 approximation (JLZA) and greedy block coordinate descent (GBCD), respectively, instead

of the ℓ_1-SVD to find the locations of the non-zero support of all columns of $\tilde{\mathbf{S}}[k']$ in (2.9). More details can be found in [31], [35] and [36]. For this MMV case, it is again possible that the number of grid points is still not sufficient enough to avoid grid mismatch and the use of sparse Bayesian inference is proposed in [37] to mitigate this problem.

Note that again, all the methods for the DOA estimation in the MMV case can theoretically be used for the identification of sinusoids. In the context of single period temporal compression, this implies that the multiple measurements are collected over a domain other than the time domain. One example is when we have multiple sensors sensing the same sinusoids and when measurements across the different sensors are collected.

2.2.2 Periodic Temporal or Spatial Compression

Signal or Spectrum Reconstruction

MMVs can also be collected along the domain where the compression is done. This leads to periodic temporal compression for the TS case and periodic spatial compression for the SS case. Mathematically, the CS model for both cases can be written as

$$\mathbf{Y}_{t'} = \mathbf{C}\mathbf{X}_{t'} = \mathbf{C}\mathbf{F}_N^{-1}\check{\mathbf{X}}_{t'}, \tag{2.10}$$

with the $M \times K$ matrix $\mathbf{Y}_{t'} = [\mathbf{y}_{t'}[0], \mathbf{y}_{t'}[1], \ldots, \mathbf{y}_{t'}[K-1]]$, the $N \times K$ matrix $\mathbf{X}_{t'} = [\mathbf{x}_{t'}[0], \mathbf{x}_{t'}[1], \ldots, \mathbf{x}_{t'}[K-1]]$ and the $N \times K$ matrix $\check{\mathbf{X}}_{t'} = [\check{\mathbf{x}}_{t'}[0], \check{\mathbf{x}}_{t'}[1], \ldots, \check{\mathbf{x}}_{t'}[K-1]]$. While (2.10) is easy to explain in the time domain, in the spatial domain, (2.10) implies that the same spatial compression is applied on each block of N consecutive antennas of the ULA described in Section 2.1 and the output $\mathbf{y}_{t'}[k]$ from all blocks k are then combined to form $\mathbf{Y}_{t'}$ in (2.10). The model in (2.10) is useful if the joint sparsity structure again exists along the columns of $\check{\mathbf{X}}_{t'}$ in which case we can again employ any joint sparse reconstruction approach to solve the underdetermined system in (2.10). For temporal compression, this boils down to joint frequency spectrum reconstruction over different sample blocks k while for spatial compression, this leads to joint angular spectrum reconstruction over different antenna blocks k.

We will now try to relate the compression model in (2.10) with multi-coset sampling, which is also used in the compressive spectrum reconstruction approach of [38]. Let us first define a $KN \times 1$ vector

$$\mathbf{x}_{t'} = \text{vec}(\mathbf{X}_{t'}) = [x_{t'}[0], x_{t'}[1], \ldots, x_{t'}[KN-1]]^T \tag{2.11}$$

with vec(.) an operation that stacks all columns of a matrix into a single vector. We can then label the collection of indices $\tilde{n}$ in $x_{t'}[\tilde{n}]$ that satisfy $\{\tilde{n}|\tilde{n} \in \{0,1,\ldots,KN-1\}, \tilde{n} \bmod N = n\}$ as the $(n+1)$-*th coset*, with $n = 0,1,\ldots,N-1$ and $\tilde{n} \bmod N$ the remainder of the integer division $\tilde{n}/N$. We can now see that the $(n+1)$-th row of $\mathbf{X}_{t'}$ actually contains the samples collected at the $(n+1)$-th coset. Multi-coset sampling is perceived as a periodic sampling technique that only collects samples at M out of N cosets available in the Nyquist-rate sampling. Hence, it is clear that when $\mathbf{C}$ is a selection matrix, (2.10) illustrates multi-coset sampling. The compressive spectrum reconstruction approach of [38] does not directly use (2.10). Instead, their approach can be perceived as performing a discrete-time Fourier transform (DTFT) on every row of $\mathbf{X}_{t'}$ for every digital frequency $\varphi \in [0, 1/N)$. The results of the DTFT operations are then stored into an $N \times 1$ vector $\bar{\mathbf{x}}_{t'}(\varphi)$ where the $(n+1)$-th entry of $\bar{\mathbf{x}}_{t'}(\varphi)$ gives the DTFT of the $(n+1)$-th row of $\mathbf{X}_{t'}$ at digital frequency φ. Since the multi-coset sampling in [38] activates only $M < N$ cosets, the compression model of [38] can generally be written as

$$\bar{\mathbf{y}}_{t'}(\varphi) = \mathbf{C}\bar{\mathbf{x}}_{t'}(\varphi), \;\; \varphi \in [0, 1/N), \tag{2.12}$$

with $\mathbf{C}$ a selection matrix. Let us now introduce the DTFT of $x_{t'}[\tilde{n}]$ stored in $\mathbf{x}_{t'}$ in (2.11) at frequency φ as $X_{t'}(\varphi)$. By introducing the $N \times 1$ vector $\tilde{\mathbf{x}}_{t'}(\varphi) = [X_{t'}(\varphi), X_{t'}(\varphi + \frac{1}{N}), \ldots, X_{t'}(\varphi + \frac{N-1}{N})]^T$ with $\varphi \in [0, 1/N)$, [38] can generally show the relationship between $\bar{\mathbf{x}}_{t'}(\varphi)$ and $\tilde{\mathbf{x}}_{t'}(\varphi)$ as

$$\bar{\mathbf{x}}_{t'}(\varphi) = \mathbf{F}_N^{-1}\tilde{\mathbf{x}}_{t'}(\varphi), \quad \varphi \in [0, 1/N). \tag{2.13}$$

The aim of [38] is to reconstruct $\tilde{\mathbf{x}}_{t'}(\varphi)$ in (2.13) from $\bar{\mathbf{y}}_{t'}(\varphi)$ in (2.12). To obtain a unique solution for $\tilde{\mathbf{x}}_{t'}(\varphi)$ and for all $\varphi \in [0, 1/N)$, [38] assumes that $\tilde{\mathbf{x}}_{t'}(\varphi)$ is sparse for each $\varphi \in [0, 1/N)$. This is ensured in [38] by assuming that the received signal has a multiband structure in the frequency domain, where the number of bands is restricted and the size of each band is smaller than $1/N$. More details about the reconstruction method used to recover $\tilde{\mathbf{x}}_{t'}(\varphi)$ from $\bar{\mathbf{y}}_{t'}(\varphi)$ for each $\varphi \in [0, 1/N)$ can be found in [38].

Note that the model in (2.12)-(2.13) can also be used in the context of spatial compression. This is true since the model in (2.12)-(2.13) is actually used by [39] in the SS case for angular spectrum reconstruction, where a multi-coset array (or periodic NULA) is employed. The details about the angular spectrum reconstruction approach used by [39] can be found in their paper. Also note that it is generally possible to combine the MMV model in the single period compression of Section 2.2.1 with the periodic compression of this section. For example, we might have multiple

sensors, each of which collects MMVs in the time-domain. Similarly, we can also activate $M < N$ antennas in every block of N consecutive antennas in the ULA leading to a periodic non-ULA, where every active antenna collects samples at multiple time indices. At this point, we will not go into the details about any of these possible combinations. However, we might discuss some of these combinations if necessary in the next section where we focus on the compressive reconstruction of the power spectrum or the second-order statistics.

Parameter Identification

Note that, while the frequency or sinusoid identification discussed in Section 2.2.1 can exploit the MMVs by using the single period temporal compression of Section 2.2.1 and by collecting measurements from multiple sensors, it is also possible to use periodic temporal compression and to form the MMVs from measurements collected at different time indices. The same is true for DOA estimation, i.e., we can collect MMVs from periodic spatial compression at a single time index. In this section, we do not provide any thorough discussion since the analysis for this section can straightforwardly be done based on the discussions in Section 2.2.1 and in the signal or spectrum reconstruction subsection in Section 2.2.2.

2.3 Exploitation of Second-Order Statistics

Observe that, in the signal/spectrum reconstruction or parameter (DOA or frequency) estimation approaches discussed in Section 2.2, the original signal/spectrum is required to have a sparse representation in a particular basis. Furthermore, if we for example consider (2.3) and assume that $\check{\mathbf{x}}_{t'}[k']$ is sparse then the dimension of the SMV $\mathbf{y}_{t'}[k']$, which is M, has to be larger than the number of non-zero entries in $\check{\mathbf{x}}_{t'}[k']$ (which is known as the degree of sparsity) [40]. When we apply this requirement on the model of (2.9) and consider the DOA estimation approach of [31], we can find that the number of actual sources whose DOAs are to be estimated (which is I) must be less than M.

Given compressive samples, we can show that when we focus on reconstructing the second-order statistics instead of the original signal/spectrum, the sparsity constraint can be removed from the signal/spectrum if the signal has a special structure in its second-order statistics. For example, let us consider (2.3) and compute the $M \times M$ auto-correlation matrix of $\mathbf{y}_{t'}[k']$ as

$$\mathbf{R}_{y_{t'}[k']} = E[\mathbf{y}_{t'}[k']\mathbf{y}_{t'}^H[k']] = \mathbf{C}\mathbf{R}_{x_{t'}[k']}\mathbf{C}^H, \tag{2.14}$$

with $\mathbf{R}_{x_{t'}[k']}$ the $N \times N$ auto-correlation matrix of $\mathbf{x}_{t'}[k']$. In the situation where $\mathbf{x}_{t'}[k']$ contains a sequence of stationary samples, $\mathbf{R}_{x_{t'}[k']}$ has a Toeplitz structure, which implies that, assuming $\mathbf{R}_{x_{t'}[k']}$ is not known, the number of unknowns is effectively $2N-1$ instead of N^2. Note that, despite the Toeplitz structure of $\mathbf{R}_{x_{t'}[k']}$, $\mathbf{R}_{y_{t'}[k']}$ generally does not have any special structure for a general matrix $\mathbf{C}$ so the number of equations is generally M^2. For this specific situation, stacking all columns of $\mathbf{R}_{y_{t'}[k']}$ into a single vector leads to

$$\text{vec}(\mathbf{R}_{y_{t'}[k']}) = (\mathbf{C}^* \otimes \mathbf{C})\text{vec}(\mathbf{R}_{x_{t'}[k']}) = (\mathbf{C}^* \otimes \mathbf{C})\mathbf{T}\mathbf{r}_{x_{t'}[k']}, \tag{2.15}$$

where the redundant information in $\mathbf{R}_{x_{t'}[k']}$ allows us to condense $\text{vec}(\mathbf{R}_{x_{t'}[k']})$ into the $(2N-1) \times 1$ vector $\mathbf{r}_{x_{t'}[k']}$, and where $\otimes$ denotes the Kronecker product operation (please see Appendix 2.A for the definition of the Kronecker product operation) and $\mathbf{T}$ is a special repetition matrix (see Section 3.3.2). Observe that the dimension of $\text{vec}(\mathbf{R}_{y_{t'}[k']})$ (which is M^2) can be larger than $2N-1$ despite $M < N$. If this is the case, we can reconstruct $\mathbf{r}_{x_{t'}[k']}$ from $\text{vec}(\mathbf{R}_{y_{t'}[k']})$ using least-squares (LS) under the full column rank condition of $(\mathbf{C}^* \otimes \mathbf{C})\mathbf{T}$ and without a sparsity constraint on either $\mathbf{r}_{x_{t'}[k']}$ or $\mathbf{x}_{t'}[k']$. Fig. 2.3 illustrates the overall concept about the reconstruction of the correlation of the stationary signals from compressive samples. Note that the power spectrum estimate can be computed by simply applying DFT on the reconstructed $\mathbf{r}_{x_{t'}[k']}$. Also recall that, in practice, the compressive measurements $\mathbf{y}_{t'}[k']$ are directly obtained from the analog version of $\mathbf{x}_{t'}[k']$ using the analog-to-information converter.

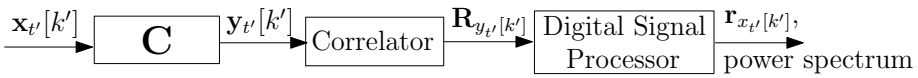


Figure 2.3: Conceptual illustration of second-order statistics reconstruction from compressive samples when the received signal is stationary.

Similarly, given compressive samples, we can also show for DOA estimation that exploiting second-order statistics might help us to increase the degrees of freedom such that the dimension of each measurement vector M can be smaller than the number of actual sources I if the sources have a special correlation structure. For example, let us consider (2.8) and write the correlation matrix $\mathbf{R}_{y_{t'}[k']}$ as

$$\mathbf{R}_{y_{t'}[k']} = \mathbf{C}\tilde{\mathbf{A}}\mathbf{R}_{\tilde{s}_{t'}[k']}\tilde{\mathbf{A}}^H\mathbf{C}^H, \tag{2.16}$$

with $\mathbf{R}_{\tilde{s}_{t'}[k']}$ the $Q \times Q$ auto-correlation matrix of $\tilde{\mathbf{s}}_{t'}[k']$. In the situation where the signals from different sources are uncorrelated, $\mathbf{R}_{\tilde{s}_{t'}[k']}$ is a diagonal matrix. For

this specific situation, cascading all columns of $\mathbf{R}_{y_{t'}[k']}$ into a single vector leads to

$$\text{vec}(\mathbf{R}_{y_{t'}[k']}) = ((\mathbf{C}\tilde{\mathbf{A}})^* \otimes (\mathbf{C}\tilde{\mathbf{A}}))\text{vec}(\mathbf{R}_{\tilde{s}_{t'}[k']}) = ((\mathbf{C}\tilde{\mathbf{A}})^* \odot (\mathbf{C}\tilde{\mathbf{A}}))\text{diag}(\mathbf{R}_{\tilde{s}_{t'}[k']}), \tag{2.17}$$

where $\odot$ denotes the Khatri-Rao product operation (please see Appendix 2.A for the definition of the Khatri-Rao product operation) and diag(.) gives the diagonal elements of a matrix. Observe that M^2 can be larger than the dimension of $\text{diag}(\mathbf{R}_{\tilde{s}_{t'}[k']})$, which is Q, despite $M < Q$. If $M^2 > Q$, we can reconstruct $\text{diag}(\mathbf{R}_{\tilde{s}_{t'}[k']})$ from $\text{vec}(\mathbf{R}_{y_{t'}[k']})$ in (2.17) using LS under the full column rank condition of $(\mathbf{C}\tilde{\mathbf{A}})^* \odot (\mathbf{C}\tilde{\mathbf{A}})$ and without a sparsity constraint on $\tilde{\mathbf{s}}_{t'}[k']$. By recalling that $I < Q$, the removal of the sparsity constraint from $\tilde{\mathbf{s}}_{t'}[k']$ and equivalently $\text{diag}(\mathbf{R}_{\tilde{s}_{t'}[k']})$ also implies that it is theoretically possible to have $M < I$ since $\text{diag}(\mathbf{R}_{\tilde{s}_{t'}[k']})$ can be reconstructed even if all of its entries are non-zero. Note however that having $M^2 > Q$ is quite unlikely as the number of grid points Q is required to be very large in order to avoid grid mismatch. But even if $M^2 < Q$, the idea of exploiting uncorrelatedness between the sources to increase the degrees of freedom is still useful as we will show later on, though it is performed by using a model slightly different from (2.17).

So far we have provided some examples using the models in (2.15) and (2.17). In practice, the expectation operation in the correlation matrix computation has to be approximated and this implies that we need MMVs because with an SMV, we can only roughly approximate $\mathbf{R}_{y_{t'}[k']}$, for instance, using the outer product of $\mathbf{y}_{t'}[k']$ leading to a poor approximation. Hence, the exploitation of the second-order statistics in the SMV case is not recommended. Next, we will look at the correlation reconstruction or the exploitation of the second-order statistics for parameter estimation in the context of MMVs. While many possible combinations between the types of compression and the types of information to be reconstructed might theoretically be possible, we will only discuss some of them that are related to our works in the following chapters.

2.3.1 Periodic Temporal or Spatial Compression: Correlation Reconstruction

Toeplitz Structure Exploitation

Let us start by considering (2.10) in the context of temporal compression and recalling (2.1). If the time-domain samples $x_t[\tilde{n}]$ collected in $\{\mathbf{x}_{t'}[k]\}_{k=0}^{K-1}$ in (2.10) form a stationary sequence, the correlation matrix between $\mathbf{x}_{t'}[k]$ and $\mathbf{x}_{t'}[k']$ de-

pends only on the block-lag $\kappa = k - k'$. Let us now focus on block-lag $\kappa = 0$ and write the auto-correlation matrix of $\mathbf{y}_{t'}[k]$ collected in $\mathbf{Y}_{t'}$ in (2.10) as

$$\mathbf{R}_{y_{t'}}[0] = E[\mathbf{y}_{t'}[k]\mathbf{y}_{t'}^H[k]] = \mathbf{C}E[\mathbf{x}_{t'}[k]\mathbf{x}_{t'}^H[k]]\mathbf{C}^H = \mathbf{C}\mathbf{R}_{x_{t'}}[0]\mathbf{C}^H. \tag{2.18}$$

The stationarity of $x_t[\tilde{n}]$ also implies that $\mathbf{R}_{x_{t'}}[0]$ has a Toeplitz structure, which means that we can again follow (2.15) to write

$$\text{vec}(\mathbf{R}_{y_{t'}}[0]) = (\mathbf{C}^* \otimes \mathbf{C})\text{vec}(\mathbf{R}_{x_{t'}}[0]) = (\mathbf{C}^* \otimes \mathbf{C})\mathbf{T}\mathbf{r}_{x_{t'}}[0], \tag{2.19}$$

where the redundant information in $\mathbf{R}_{x_{t'}}[0]$ is again condensed into the $(2N-1)\times 1$ vector $\mathbf{r}_{x_{t'}}[0]$, which can again be reconstructed from $\text{vec}(\mathbf{R}_{y_{t'}}[0])$ using LS subject to the full column rank condition of $(\mathbf{C}^* \otimes \mathbf{C})\mathbf{T}$. Due to the stationarity of $x_t[\tilde{n}]$, which is translated to the stationarity of $\mathbf{x}_{t'}[k]$ along k and by assuming the ergodicity of $x_t[\tilde{n}]$, in practice, $\mathbf{R}_{y_{t'}}[0]$ can be estimated using $\hat{\mathbf{R}}_{y_{t'}}[0] = \frac{1}{K}\mathbf{Y}_{t'}\mathbf{Y}_{t'}^H$ and LS is then applied to $\text{vec}(\hat{\mathbf{R}}_{y_{t'}}[0])$ instead of $\text{vec}(\mathbf{R}_{y_{t'}}[0])$ in (2.19). This leads to some errors since $\hat{\mathbf{R}}_{y_{t'}}[0]$ is actually given by $\hat{\mathbf{R}}_{y_{t'}}[0] = \mathbf{C}\hat{\mathbf{R}}_{x_{t'}}[0]\mathbf{C}^H$ with $\hat{\mathbf{R}}_{x_{t'}}[0] = \frac{1}{K}\mathbf{X}_{t'}\mathbf{X}_{t'}^H$, which has a Toeplitz structure only for $K \to \infty$. This model is the basis for our alternative time-domain (ATD) approach discussed in Chapter 3, where we generally suggest the use of a Gaussian sampling matrix or multi-coset sampling matrix based on a minimal sparse ruler (see Definition 7.3.2) for $\mathbf{C}$ in (2.19) to ensure the full column rank condition of $(\mathbf{C}^* \otimes \mathbf{C})\mathbf{T}$ in (2.19). This rank condition, however, can also be ensured by designing $\mathbf{C}$ based on either coprime sampling [41] or nested sampling [42]. Another method using multi-coset sampling proposed by [43] focuses on the frequency-domain representation and considers the correlation matrix of $\check{\mathbf{x}}_{t'}[k]$, which is collected in $\check{\mathbf{X}}_{t'}$ in (2.10). For a stationary signal, [43] claims that $\mathbf{R}_{\check{x}_{t'}}[0] = E[\check{\mathbf{x}}_{t'}[k]\check{\mathbf{x}}_{t'}^H[k]]$ is a diagonal matrix, which implies that they have less unknowns and allows them to reconstruct $\text{diag}(\mathbf{R}_{\check{x}_{t'}}[0])$ from $\text{vec}(\mathbf{R}_{y_{t'}}[0])$ in (2.19) using LS though they do not propose any specific $\mathbf{C}$ that ensures the full column rank condition of the resulting system matrix. Apart from our ATD approach, we also propose a more general time-domain (TD) approach in Chapter 3, where we basically also exploit the correlation matrix $\mathbf{R}_{y_{t'}}[\kappa]$ at block lags $|\kappa| > 0$. In addition to the TD and ATD approaches, Chapter 3 also discusses a related frequency-domain approach. In Chapter 3, the reconstructed correlation of the received signal is used to compute an estimate of its power spectrum.

Note that, if we consider (2.10) in the context of the SS case, the discussion in the previous paragraph is also theoretically applicable for the SS case when the

spatial-domain samples $x_t[\tilde{n}]$ collected in $\{\mathbf{x}_{t'}[k]\}_{k=0}^{K-1}$ in (2.10) also form a stationary sequence. In this case, a Toeplitz structure will emerge in the spatial correlation matrix $\mathbf{R}_{x_{t'}}[0]$ in (2.18). Here, we can use the multi-coset array of [39] using the minimal sparse ruler design suggested by the ATD approach in Chapter 3 to introduce compression while maintaining the identifiability of the spatial correlation matrix $\mathbf{R}_{x_{t'}}[0]$. Note that, one case that leads to a Toeplitz structure in the spatial correlation matrix $\mathbf{R}_{x_{t'}}[0]$ is when $x_t[\tilde{n}]$ is given by the output of the $(\tilde{n}+1)$-th antenna of a ULA receiving signals from uncorrelated point sources (see [41] for a detailed explanation).

Multiband Signals and Multibin Model

In [44], the model of (2.12) is exploited and, similar to [38], [44] also assumes that the received signal has a multiband structure in the frequency domain where the size of each band (in digital frequency units) is smaller than $1/N$. In addition, [44] also assumes that every different band corresponds to a different user signal and thus they are uncorrelated. Let us write the $M \times M$ correlation matrix of $\bar{\mathbf{y}}_{t'}(\varphi)$ in (2.12) as

$$\mathbf{R}_{\bar{y}_{t'}(\varphi)} = \mathbf{C}\mathbf{F}_N^{-1}\mathbf{R}_{\tilde{x}_{t'}(\varphi)}\mathbf{F}_N^{-H}\mathbf{C}^H, \;\; \varphi \in [0, 1/N), \tag{2.20}$$

where $\mathbf{R}_{\tilde{x}_{t'}(\varphi)}$ is the auto-correlation matrix of $\tilde{\mathbf{x}}_{t'}(\varphi)$ in (2.13) with $\tilde{\mathbf{x}}_{t'}(\varphi)$ not necessarily sparse. As different bands are uncorrelated and their size is smaller than $1/N$, it can be found that $\mathbf{R}_{\tilde{x}_{t'}(\varphi)}$ is a diagonal matrix for all $\varphi \in [0, 1/N)$. As a result, we can rewrite (2.20) as

$$\begin{aligned} \text{vec}(\mathbf{R}_{\bar{y}_{t'}(\varphi)}) &= ((\mathbf{C}\mathbf{F}_N^{-1})^* \otimes (\mathbf{C}\mathbf{F}_N^{-1}))\text{vec}(\mathbf{R}_{\tilde{x}_{t'}(\varphi)}) \\ &= ((\mathbf{C}\mathbf{F}_N^{-1})^* \odot (\mathbf{C}\mathbf{F}_N^{-1}))\text{diag}(\mathbf{R}_{\tilde{x}_{t'}(\varphi)}), \;\; \varphi \in [0, 1/N), \end{aligned} \tag{2.21}$$

with $(\mathbf{C}\mathbf{F}_N^{-1})^* \odot (\mathbf{C}\mathbf{F}_N^{-1})$ an $M^2 \times N$ matrix. Again, we can have $M^2 > N$ though $M < N$ and we can reconstruct $\text{diag}(\mathbf{R}_{\tilde{x}_{t'}(\varphi)})$ from $\text{vec}(\mathbf{R}_{\bar{y}_{t'}(\varphi)})$ using LS if $(\mathbf{C}\mathbf{F}_N^{-1})^* \odot (\mathbf{C}\mathbf{F}_N^{-1})$ has full column rank.

Note that [44] uses multi-coset sampling and proposes a sampling pattern that ensures the full column rank condition of $(\mathbf{C}\mathbf{F}_N^{-1})^* \odot (\mathbf{C}\mathbf{F}_N^{-1})$ but they do not look at a more general condition and they do not focus on the minimum possible compression rate. In contrast, we have found that $\mathbf{F}_N^{-1}\mathbf{R}_{\tilde{x}_{t'}(\varphi)}\mathbf{F}_N^{-H}$ in (2.20) is actually a circulant matrix. This circulant structure is exploited in Chapter 5, where we also adopt the model of (2.21) to examine the compressive reconstruction of the frequency-domain periodogram from time-domain signals received at different

wireless sensors. In Chapter 5, we will show that the full column rank condition of $(\mathbf{C}\mathbf{F}_N^{-1})^* \odot (\mathbf{C}\mathbf{F}_N^{-1})$ in (2.21) is ensured if the multi-coset sampling matrix $\mathbf{C}$ is designed based on a circular sparse ruler. In addition, we will also discuss in Chapter 5 that the multiband signal model discussed in [44] is actually only a special case of a more general multibin model. In the multibin model, we simply split the digital frequency band $\vartheta \in [0, 1)$ into N equal bins, each of them having a size of $1/N$. Then, we will show that the circulant structure in $\mathbf{F}_N^{-1}\mathbf{R}_{\tilde{x}_{t'}(\varphi)}\mathbf{F}_N^{-H}$ in (2.20) will also appear if the spectra at frequencies ϑ located at different bins are uncorrelated though the received signal does not have a multiband structure. More details on the frequency periodogram reconstruction using the multibin model can be found in Chapter 5.

Also note that the aforementioned discussion is also applicable for spatial correlation matrix reconstruction from periodic spatial compression. This is, for example, also discussed in Chapter 5 when we examine the compressive reconstruction of the angular-domain periodogram from spatial-domain signals received by a linear array of antennas at different time indices. Here, we first use the compression model of (2.12) in the context of periodic spatial compression as is done by [39]. This is implemented by perceiving the entire ULA model as multiple blocks of N consecutive antennas and activating only $M < N$ antennas out of every one block of these N consecutive antennas leading to a periodic NULA. We then also follow the correlation model of (2.20) by adjusting (2.20) to the context of spatial compression (t in $\mathbf{R}_{\bar{y}_t(\varphi)}$ then represents the time index and φ is related to angle). More details about the angular periodogram reconstruction from this scheme can be found in Chapter 5.

Cyclic Spectrum Reconstruction

Let us again return to the model of (2.18) in the context of temporal sampling and observe that $\mathbf{R}_{x_{t'}}[0]$ generally does not have an exploitable special structure if $x_t[\tilde{n}]$ collected in $\{\mathbf{x}_{t'}[k]\}_{k=0}^{K-1}$ does not form a stationary sequence. This problem is encountered in [45], which focuses on reconstructing the cyclic spectrum of a cyclostationary signal and has to assume sparsity on the cyclic spectrum. Let us now assume that $x_t[\tilde{n}]$ is a cyclostationary signal and consider $\mathbf{x}_{t'}$ in (2.11), which can also be written as $\mathbf{x}_{t'} = [\mathbf{x}_{t'}^T[0], \mathbf{x}_{t'}^T[1], \ldots, \mathbf{x}_{t'}^T[K-1]]^T$. If we follow the approach of [11], which sets the number of entries of $\mathbf{x}_{t'}[k]$, which is N, equal to the period of the cyclostationarity, we can find that the auto-correlation matrix of

$\mathbf{x}_{t'}$ is given by

$$\mathbf{R}_{x_{t'}} = \begin{bmatrix} \mathbf{R}_{x_{t'}}[0] & \mathbf{R}_{x_{t'}}[-1] & \dots & \mathbf{R}_{x_{t'}}[-K+1] \\ \mathbf{R}_{x_{t'}}[1] & \mathbf{R}_{x_{t'}}[0] & \dots & \mathbf{R}_{x_{t'}}[-K+2] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{x_{t'}}[K-1] & \mathbf{R}_{x_{t'}}[K-2] & \dots & \mathbf{R}_{x_{t'}}[0] \end{bmatrix}, \tag{2.22}$$

i.e., it has a block Toeplitz structure. This block Toeplitz structure can be exploited for compression by introducing K compression matrices $\{\mathbf{C}_k\}_{k=0}^{K-1}$ instead of a single compression matrix $\mathbf{C}$ and performing the compression as

$$\mathbf{y}_{t'}[\bar{k}K+k] = \mathbf{C}_k \mathbf{x}_{t'}[\bar{k}K+k],\ k = 0, 1, \dots, K-1. \tag{2.23}$$

where the size of $\mathbf{C}_k$ is $M_k \times N$. Note that, if we collect only K measurement vectors $\{\mathbf{y}_{t'}[k]\}_{k=0}^{K-1}$, we only have $\bar{k} = 0$ in (2.23). In this case, the temporal sampling is not temporally periodic anymore. Since, in practice, we have to approximate the expectation operation in the correlation matrix computation, we should collect multiple blocks of K measurement vectors. This implies that we again have periodic compression though the period of the temporal sampling is now K times the period of the compression that employs only a single compression matrix $\mathbf{C}$. The model in (2.23) is used in both [11] and Chapter 7 for compressive cyclic spectrum reconstruction with Chapter 7 focusing more on multi-coset sampling. The details about the reconstruction of the correlation matrix $\mathbf{R}_{x_{t'}}$ in (2.22) from $\{\mathbf{y}_{t'}[\bar{k}K+k]\}_{k,\bar{k}}$ in (2.23) and that of the cyclic spectrum from the reconstructed $\mathbf{R}_{x_{t'}}$ can be found in Chapter 7.

2.3.2 Single Period Temporal or Spatial Compression (Multiple Measurement Vectors): Parameter Identification

Uncorrelated Sources

Let us first consider (2.4) and (2.9) in the context of spatial compression and parameter identification, i.e., DOA estimation, but we focus on the true DOAs $\{\theta_i\}_{i=0}^{I-1}$ instead of the grid of investigated angles $\{\tilde{\theta}_q\}_{q=0}^{Q-1}$. In other words, we also consider (2.7) and write $\mathbf{Y}[k']$ as $\mathbf{Y}[k'] = \mathbf{CAS}[k']$ instead of (2.9) with $\mathbf{S}[k']$ similarly defined as $\tilde{\mathbf{S}}[k']$ in (2.9). We now assume that $\{\mathbf{s}_t[k']\}_{t=0}^{T-1}$ collected in $\mathbf{S}[k']$ is stationary along time index t, which also implies that $\{\mathbf{y}_t[k']\}_{t=0}^{T-1}$ and $\{\mathbf{x}_t[k']\}_{t=0}^{T-1}$ collected in $\mathbf{Y}[k']$ and $\mathbf{X}[k']$ in (2.4), respectively, are also stationary along t. As a result, the auto-correlation matrices of $\mathbf{y}_t[k']$ and $\mathbf{s}_t[k']$ are

the same for $t = 0, 1, \ldots, T-1$, i.e., $\mathbf{R}_{y_t[k']} = E[\mathbf{y}_t[k']\mathbf{y}_t^H[k']] = \mathbf{R}_{y[k']}$ and $\mathbf{R}_{s_t[k']} = E[\mathbf{s}_t[k']\mathbf{s}_t^H[k']] = \mathbf{R}_{s[k']}$. Based on (2.7) and (2.4), we can then write

$$\mathbf{R}_{y[k']} = \mathbf{C}\mathbf{R}_{x[k']}\mathbf{C}^H = \mathbf{C}\mathbf{A}\mathbf{R}_{s[k']}\mathbf{A}^H\mathbf{C}^H. \tag{2.24}$$

Note that, in practice, the stationarity of $\mathbf{y}_t[k']$ allows us to approximate $\mathbf{R}_{y[k']}$ as $\hat{\mathbf{R}}_{y[k']} = \frac{1}{T}\mathbf{Y}[k']\mathbf{Y}^H[k']$. Let us now assume that the sources, whose DOAs are to be estimated, are uncorrelated, which implies that $\mathbf{R}_{s[k']}$ in (2.24) is a diagonal matrix. We can then follow (2.17) and write

$$\text{vec}(\mathbf{R}_{y[k']}) = ((\mathbf{C}\mathbf{A})^* \odot (\mathbf{C}\mathbf{A}))\text{diag}(\mathbf{R}_{s[k']}). \tag{2.25}$$

At this point, we can again consider a grid of angles $\{\tilde{\theta}_q\}_{q=0}^{Q-1}$ as well as replace $\mathbf{A}$ and $\mathbf{R}_{s[k']}$ in (2.25) with $\tilde{\mathbf{A}}$ and $\mathbf{R}_{\tilde{s}[k']}$, respectively. In this case, we have a similar form to (2.17) and the discussion after (2.17) follows. However, it is also possible to consider a subspace algorithm like multiple signals classification (MUSIC) [46]. For this purpose, let us recall that $\mathbf{A}$ is the array response matrix of the considered single block of N antennas in our ULA model. Assuming that $\mathbf{C}$ is a selection matrix, we can observe that N_v distinct rows of $(\mathbf{C}\mathbf{A})^* \odot (\mathbf{C}\mathbf{A})$ actually provide the array response matrix of a virtual array (also called co-array) of N_v virtual antennas. Note that the number of virtual antennas N_v depends on the value of $\mathbf{C}$ and it is generally larger than the number of active antennas M. The upper bound for N_v is $N_v \leq M(M-1)/2$. We can then perceive the entries of $\text{vec}(\mathbf{R}_{y[k']})$ in (2.25) corresponding to the N_v distinct rows of $(\mathbf{C}\mathbf{A})^* \odot (\mathbf{C}\mathbf{A})$ as the outputs of the N_v virtual antennas. When this is the case, however, we also have to consider $\text{diag}(\mathbf{R}_{s[k']})$ in (2.25) as the new source vector and this can be problematic when we intend to apply MUSIC. The reason is the fact that $\text{diag}(\mathbf{R}_{s[k']})$ behaves like I fully coherent sources since $\text{diag}(\mathbf{R}_{s[k']})$ contains constants instead of random variables. More about some issues in the use of MUSIC for fully coherent sources can be found in Section 4.1 and the references therein. One way to solve this issue is provided by [47], which assumes that $\mathbf{s}_t[k']$ is quasi-stationary (instead of stationary) along time index t. In our model, this quasi-stationarity assumption can be interpeted as having $\{\mathbf{s}_t[k']\}_{t=0}^{\tau T-1}$ with $\mathbf{s}_t[k']$ only stationary within T consecutive time indices. Using this quasi-stationarity assumption, $\text{diag}(\mathbf{R}_{s[k']})$ in (2.25) varies with time and we can basically collect up to τ independent measurements. The spatial compression done in [47] and the following mathematical manipulation lead to a model that is equivalent to our model in (2.25) with $\mathbf{C}$ formed by the first M rows of $\mathbf{I}_N$, which implies that they have a ULA of M active antennas. As a result, in the

co-array domain, they have a virtual ULA of $N_v = 2M - 1$ virtual antennas and by using MUSIC, they can estimate the DOA of up to $I = 2M - 2$ sources, which is more than the number of physical active antennas M. To increase the degrees of freedom beyond what is achieved by [47], the papers [41] and [42] propose to select the M active antennas (which is equivalent to the design of $\mathbf{C}$) according to special geometries rather than using a ULA as is done by [47]. [42] proposes to use a nested array of active antennas while the selection of active antennas based on the coprime array is proposed in [41]. Another difference between the two approaches and the work of [47] is that both [41] and [42] assume that $\mathbf{s}_t[k']$ is stationary (instead of quasi-stationary) along time index t. After mathematical manipulations, both approaches of [41] and [42] also boil down to (2.25), and thus there should be a way to deal with the fact that $\text{diag}(\mathbf{R}_{s[k']})$ in (2.25) contains constants. The solution advocated by [41] and [42] is to use spatial smoothing (see [48]) before the use of MUSIC. One work by us in [10] also follows the work of [41] and [42] but we use the minimum redundancy array introduced in [12] instead of the nested or coprime array. Note that the approaches in [10], [41], and [42] can generally detect more sources than the one detected by [47] for a given number of M active antennas. More details can be found in the respective papers.

Note that all the above discussions can be repeated in the context of single period temporal compression (MMV case) and the identification of the frequencies of some sinusoids. In this case, the MMVs can be collected over the spatial domain, for example, over multiple wireless sensors. However, it is also possible in the identification of sinusoids to collect the MMVs over the time-domain, i.e., by using the periodic temporal compression of Section 2.3.1 (see [41] for example) though we do not provide any discussion on it here.

Correlated Sources

Let us now return to (2.24) and observe that $\mathbf{R}_{s[k']}$ is generally not a diagonal matrix if the sources, whose DOAs are to be estimated, are correlated. As a result, we will not have the model of (2.25) and instead, we will only have

$$\text{vec}(\mathbf{R}_{y[k']}) = (\mathbf{C}^* \otimes \mathbf{C})\text{vec}(\mathbf{R}_{x[k']}) = ((\mathbf{CA})^* \otimes (\mathbf{CA}))\text{vec}(\mathbf{R}_{s[k']}), \quad (2.26)$$

which is generally an underdetermined system. However, we can gain more equations by introducing T compression matrices $\{\mathbf{C}_t\}_{t=0}^{T-1}$ instead of a single compression matrix $\mathbf{C}$, as what we have done in Section 2.3.1 when we deal with cyclosta-

tionary signals. We can then write the compression similar to (2.23) as

$$\mathbf{y}_{\bar{t}T+t}[k'] = \mathbf{C}_t \mathbf{x}_{\bar{t}T+t}[k'],\ t = 0, 1, \ldots, T-1,\ \bar{t} = 0, 1, \ldots, \bar{T}-1. \qquad (2.27)$$

where the size of $\mathbf{C}_t$ is $M_t \times N$. Observe that the compression in (2.27) leads to the so-called dynamic linear array (DLA), which will be discussed in Chapter 4. Here, in different time indices t within T consecutive time indices, we activate different sets of $M < N$ antennas in the considered single block of N antennas in our underlying ULA model. Also note that (2.27) is only one possible implementation of the DLA since we can alter the compression matrix every two or more time indices instead of every time index t. Assuming that the number of introduced compression matrices remains T, this implies that the period of the spatial compression is at least equal to T times the period of the spatial compression that employs only a single compression matrix $\mathbf{C}$. Here, the auto-correlation matrix of $\mathbf{y}_{\bar{t}T+t}[k']$ for a given t is computed as $\mathbf{R}_{y_t[k']} = E[\mathbf{y}_{\bar{t}T+t}[k']\mathbf{y}^H_{\bar{t}T+t}[k']]$. In Chapter 4, it will be shown that we will first reconstruct the correlation matrix $\mathbf{R}_{x[k']}$ in (2.26) from $\{\mathbf{R}_{y_t[k']}\}_{t=0}^{T-1}$ before applying some DOA estimation methods on the reconstructed $\mathbf{R}_{x[k']}$. More details including how to approximate the expectation operation in the correlation matrix computation can be found in Chapter 4.

2.3.3 Joint Angular-Frequency Power Spectrum Estimation

In Chapter 6, we introduce an approach to compressively and jointly reconstruct both the frequency-domain and angular-domain power spectrum of the time-domain stationary signals. This is done by first considering only one block of N consecutive antennas in our ULA model and then activating only $M < N$ out of the N antennas leading to a spatial compression. The output of each antenna is then sent to the connected digital receiver which performs periodic temporal compression. More details about this work can be found in Chapter 6.

2.4 Prior Work on Power Spectrum Estimation from Sub-Nyquist-Rate Samples

Research on estimating the power spectrum from sub-Nyquist-rate samples has been going on since 1960 with the work by [49], which focused on random sampling of stationary signals. This paper showed that aliasing in the power spectrum estimate can generally be avoided if random sampling based on a Poisson distribution is used. This work has then been furthered by [50], which also evaluated Pois-

son sampling and showed that the power spectrum estimate is mean-square consistent for all positive values of the average sampling rate under weak smoothness conditions on the power spectral density. This work was then completed by [51], which derived the asymptotic bias and covariance of the estimates and discussed the influence of the spectral windows and the sampling rate on the performance of the estimates. It has been shown in [51] that the power spectrum estimate from samples taken at Poisson sampling instants is consistent under mild smoothness conditions of the power spectral density. In [52], it has been pointed out that the definition of the alias-free estimation in [49] does not necessarily imply that a consistent power spectrum estimate can be obtained from a finite number of samples. A new definition of alias-free sampling has then been introduced in [52] and various criteria for a sampling scheme to be alias-free have been developed. Note that the aforementioned works generally focus on random sampling, specifically Poisson sampling, where the spacing between sampling instants follows the exponential distribution. As pointed out by [53], two consecutive samples in Poisson sampling can be infinitely close since the value of the spacing between sampling instants can take any real value greater than 0. As a result, Poisson sampling can be practically challenging.

In the more recent work by [54], a method of estimating the power spectrum density of random ergodic signals is introduced. There are two main differences between the work of [54] and the works in [49, 50, 51, 52]. First, unlike the above works which rely on random sampling, a deterministic sampling scheme, i.e., a periodic non-uniform sampling is introduced in [54]. Second, while the above works and [54] all claim that they can reach an arbitrary low sampling rate, the non-uniform periodic sampling introduced in [54] introduces a lower limit on the allowed spacing between two consecutive sampling instants. These two differences theoretically show that the approach of [54] is more practical. However, [54] does not really explicitly and exactly mention the achievable lowest sampling rate. It appears that a lower sampling rate can be achieved by [54] if the period of the non-uniform sampling is increased (which implies a larger duration of the sampling time). When the duration of the sampling time and/or the period of the non-uniform sampling is fixed, it is not clear if the sampling pattern of [54] is the most optimal solution. Unlike [54], our periodic non-uniform sampling, a.k.a. multi-coset sampling, discussed in this thesis is based on sampling patterns designed according to a minimal (circular) sparse ruler. Under the constraint that the sampling instants are on top of the Nyquist-rate grid, our sampling pattern based on a minimal (circular) sparse ruler gives a more clear indication about the achievable minimum sampling

rate for a fixed period of non-uniform sampling (and thus also for a fixed duration of the sampling time).

Appendix

2.A Kronecker, Hadamard, and Khatri-Rao Products

Let us assume that the element of an $M \times N$ matrix $\mathbf{C}$ at the $(m+1)$-th row and the $(n+1)$-th column, with $m = 0, 1, \ldots, M-1$ and $n = 0, 1, \ldots, N-1$, is given by $[\mathbf{C}]_{m+1,n+1}$. The Kronecker product between the matrix $\mathbf{C}$ and an $\grave{M} \times \grave{N}$ matrix $\mathbf{G}$ is then given by

$$\mathbf{C} \otimes \mathbf{G} = \begin{bmatrix} [\mathbf{C}]_{1,1}\mathbf{G} & [\mathbf{C}]_{1,2}\mathbf{G} & \ldots & [\mathbf{C}]_{1,N}\mathbf{G} \\ [\mathbf{C}]_{2,1}\mathbf{G} & [\mathbf{C}]_{2,2}\mathbf{G} & \ldots & [\mathbf{C}]_{2,N}\mathbf{G} \\ \vdots & \vdots & \ddots & \vdots \\ [\mathbf{C}]_{M,1}\mathbf{G} & [\mathbf{C}]_{M,2}\mathbf{G} & \ldots & [\mathbf{C}]_{M,N}\mathbf{G} \end{bmatrix} \tag{2.28}$$

having a size of $M\grave{M} \times N\grave{N}$.

The Hadamard product between $\mathbf{C}$ and $\mathbf{G}$ only exists if $\mathbf{C}$ and $\mathbf{G}$ have the same size, i.e., $M = \grave{M}$ and $N = \grave{N}$. If this Hadamard product exists, it is given by

$$\mathbf{C} \circ \mathbf{G} = \begin{bmatrix} [\mathbf{C}]_{1,1}[\mathbf{G}]_{1,1} & [\mathbf{C}]_{1,2}[\mathbf{G}]_{1,2} & \ldots & [\mathbf{C}]_{1,N}[\mathbf{G}]_{1,N} \\ [\mathbf{C}]_{2,1}[\mathbf{G}]_{2,1} & [\mathbf{C}]_{2,2}[\mathbf{G}]_{2,2} & \ldots & [\mathbf{C}]_{2,N}[\mathbf{G}]_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ [\mathbf{C}]_{M,1}[\mathbf{G}]_{M,1} & [\mathbf{C}]_{M,2}[\mathbf{G}]_{M,2} & \ldots & [\mathbf{C}]_{M,N}[\mathbf{G}]_{M,N} \end{bmatrix} \tag{2.29}$$

having a size of $M \times N$.

The Khatri-Rao product between $\mathbf{C}$ and $\mathbf{G}$ only exists if $\mathbf{C}$ and $\mathbf{G}$ have the same number of columns, i.e., $N = \grave{N}$. By assuming that this condition is satisfied and by writing $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_N]$ and $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_N]$, with $\mathbf{c}_{n+1}$ and $\mathbf{g}_{n+1}$ the $(n+1)$-th column of $\mathbf{C}$ and $\mathbf{G}$, respectively, it is given by

$$\mathbf{C} \odot \mathbf{G} = [\mathbf{c}_1 \otimes \mathbf{g}_1, \mathbf{c}_2 \otimes \mathbf{g}_2, \ldots, \mathbf{c}_N \otimes \mathbf{g}_N], \tag{2.30}$$

which has a size of $M\grave{M} \times N$. Note that the Khatri-Rao product can be perceived as a column-wise Kronecker product.

2.B Restricted Isometry Property (RIP) of a Matrix

The RIP of a matrix, which is originally defined in [55], is provided in the following definition.

Definition 2.B.1. *Given an $M \times N$ matrix $\mathbf{C}$, with $M < N$ and with ℓ_2-normalized columns, we define the isometry constant $\delta_S \in (0, 1)$ as the smallest quantity such that*

$$(1 - \delta_S)||\mathbf{x}||_2^2 \leq ||\mathbf{C}\mathbf{x}||_2^2 \leq (1 + \delta_S)||\mathbf{x}||_2^2 \tag{2.31}$$

holds for all vectors $\mathbf{x}$ with maximum order of sparsity S (all vectors $\mathbf{x}$ having at most S non-zero entries). Then $\mathbf{C}$ is said to satisfy RIP of order S with a constant δ_S [13, 15, 56, 57].

The implication of Definition 2.B.1 is that if a matrix $\mathbf{C}$ obeys the RIP of order S (if δ_S is not too close to one), then $\mathbf{C}$ approximately preserves the Euclidean length of S-sparse vectors, which basically means that an S-sparse vectors cannot be in the null space of $\mathbf{C}$ [13, 57]. In this case, it is also mentioned in [56] that any submatrix of $\mathbf{C}$ having no more than S columns behaves like an orthonormal matrix.

Part II: Papers Included

Chapter 3

Compressive Wideband Power Spectrum Estimation

Dyonisius Dony Ariananda and Geert Leus

Abstract

In several applications, such as wideband spectrum sensing for cognitive radio, only the power spectrum (a.k.a. the power spectral density) is of interest and there is no need to recover the original signal itself. In addition, high-rate analog-to-digital converters (ADCs) are too power hungry for direct wideband spectrum sensing. These two facts have motivated us to investigate compressive wideband power spectrum sensing, which consists of a compressive sampling procedure and a reconstruction method that is able to recover the unknown power spectrum of a wide-sense stationary signal from the obtained sub-Nyquist rate samples. It is different from spectrum blind sampling (SBS), which aims at reconstructing the original signal instead of the power spectrum. In this paper, a solution is first presented based on a periodic sampling procedure and a simple least-squares reconstruction method. We evaluate the reconstruction process both in the time and frequency domain. Then, we examine two possible implementations for the compressive sampling procedure, namely complex Gaussian sampling and multi-coset sampling, although we mainly focus on the latter. A new type of multi-coset sampling is introduced based on the so-called minimal sparse ruler problem. Next, we analyze the statistical properties of the estimated power spectrum. The computation of the mean and the covariance of the estimates allows us to calculate the analytical normalized mean squared error (NMSE) of the estimated power spectrum. Further, when the received signal is assumed to contain only circular complex zero-mean Gaussian i.i.d. noise, the computed mean and covariance can be used to derive a suitable detection threshold. Simulation results underline the promising performance of our proposed approach. Note that all benefits of our method arise without putting any sparsity constraints on the power spectrum.

3.1 Introduction

In recent years, wideband spectrum estimation and sensing has become a popular topic in signal processing and telecommunications. A popular application is cognitive radio where unlicensed users have to sense a broad frequency range in order to locate the unoccupied licensed spectrum before establishing a communication link. One possible approach is to divide the entire wideband spectrum into a large number of narrowband channels followed by a channel-by-channel sequential sensing. However, this approach might introduce a significant amount of delay in the spectrum sensing process. In [58], a filter bank structure is introduced to perform multi-

channel spectrum sensing in the wideband regime. Similarly, [59, 60] optimize a bank of multiple narrowband detectors to improve the aggregate opportunistic throughput of a cognitive radio system by introducing the so-called multiband joint detection and multiband sensing-time-adaptive joint detection, respectively. Again, these methods are not efficient due to the need for a large number of bandpass filters. Another approach is to directly scan the wideband spectrum using a high-rate analog-to-digital converter (ADC), such as in [61], where wavelets are used to detect the edges or boundaries of the occupied bands. However, such high-rate ADCs consume a large amount of power [62].

To reduce the burden on the ADCs, many researches have been performed to exploit specific features of the spectrum (such as sparsity in the spectrum or the edge spectrum [63, 64, 20]). These specific properties allow for a reduction of the sampling rate compared to the Nyquist rate while maintaining perfect signal reconstruction when no noise is present. In [21], the so-called multi-coset sampling is examined and proposed to reduce the sampling rate when the considered multiband signals have a frequency support on a union of finite intervals. Given prior knowledge of the frequency support of the received signals, [21] has derived the condition for exact reconstruction as well as proposed an explicit reconstruction formula. Unfortunately, in many applications, such as cognitive radio, the frequency support is not known in advance and the multi-coset sampling approach proposed in [21] is not suitable. In order to solve this problem, [64, 38] proposed solutions for signal reconstruction based on multi-coset sampling without any prior knowledge about the frequency support of the original signal. Closely related ideas can also be found in [20], which discusses sub-Nyquist rate sampling for sparse multiband analog signals by means of a so-called modulated wideband converter, which consists of multiple branches, each of which employs a different periodic mixing function followed by low-pass filtering and low-rate uniform sampling. Since the objective of the methods discussed in [64, 20, 38] is to sample a signal with unknown frequency support at minimal rate and reconstruct the spectrum from the samples by exploiting spectrum sparsity, these approaches fall in the class of spectrum blind sampling (SBS). In these works, it has been found that the minimum average sampling rate for most signals is given by the Landau lower bound (as studied in [21]), which is equal to the Nyquist rate multiplied with the frequency occupancy ratio. However, in the worst case scenario, the minimum average sampling rate increases and is given by the minimum of twice the Landau lower bound and the Nyquist rate. Note that all of the above approaches can be cast into a compressive sampling framework where the signal reconstruction can be carried out by using your favorite

sparse recovery method such as the least absolute shrinkage and selection operator (LASSO) algorithm [17]. Also more classical methods can be adopted, such as the minimum variance distortionless response (MVDR) method [65], or multiple signal classification (MUSIC) [64].

All methods aforementioned concentrate on spectral estimation and aim at perfectly reconstructing the original signal. In fact, for spectrum sensing applications, only the power spectrum (a.k.a. the power spectral density), or equivalently, the auto-correlation function, needs to be recovered. Power spectrum estimation methods based on sub-Nyquist rate samples have been developed in [66, 67] by concentrating on the auto-correlation function instead of the original signal itself. In [66], the spectrum sensing approach proposed by [63], which exploits the embedded sparsity of the edge spectrum, is improved by taking advantage of the connection between the auto-correlation function of the compressive measurements and that of the Nyquist rate samples. Nevertheless, [66] assumes that the compressive measurements are wide-sense stationary, which is not true for most compressive sampling matrices. In [67], a compressive sampling framework is obtained by computing the output energy of a limited number of wideband filters to reconstruct the received energy in a large number of spectral bins. Unfortunately, [67] only exploits the output energy of each filter leading to an under-determined system of equations, while cross-correlations among the outputs of the different filters could also have been exploited. In [43], a power spectrum estimation method based on multi-coset sampling is proposed by exploiting the fact that a wide-sense stationary signal corresponds to a diagonal covariance matrix of the frequency domain representation of the signal. This observation is used in [43] to build an over-determined system of equations relating the frequency domain statistics of the compressive measurements with those of the signal, which is solvable by adopting a non-negative least-squares algorithm. Another method labeled as coprime sampling is provided by [41]. This method aims at estimating the frequencies of sinusoids buried in noise by exploiting two uniform sub-Nyquist samplers with sampling periods that are coprime multiples of the Nyquist period.

This paper concentrates on efficient power spectrum reconstruction and aims at designing effective periodic sub-Nyquist sampling procedures for this, also labeled as power spectrum blind sampling (PSBS) in [68]. Theoretically, this approach is able to perfectly reconstruct the unknown power spectrum of a wide-sense stationary signal using least-squares by exploiting the cross-correlations between the different outputs of the periodic sampling device. The least-squares algorithm requires some rank conditions to be satisfied, which will guide the actual implemen-

tation of the sampling device. In this paper, sampling techniques based on random modulating waveforms can be adopted, as used in [20], but the main focus will be on multi-coset approaches. A novel multi-coset sampling implementation is designed based on the minimal sparse ruler problem. The theoretical statistical properties of the estimated power spectrum are also investigated leading to the mean and the covariance of the estimated power spectrum, which is useful for formulating the normalized mean squared error (NMSE) analytically. Moreover, the detection threshold used to evaluate the presence or absence of a signal at a specific frequency can also be derived by assuming the received signal is merely circular complex zero-mean Gaussian i.i.d. noise. All the proposed schemes are compared via both analysis and simulations. In general, the developed sampling procedures can significantly decrease the sampling rate requirements by exploiting the spectral correlation properties without putting any sparsity constraints on the power spectrum.

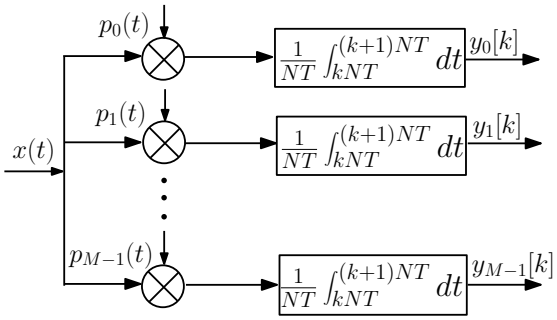


Figure 3.1: Illustration of the sample acquisition scheme, which modulates the received analog signal with M different periodic waveforms followed by an integrate-and-dump process.

3.2 System Model and Problem Statement

Let $x(t)$ be a wide-sense stationary analog signal, which is assumed to be complex-valued (e.g., the complex envelope of the observed real-valued signal) and bandlimited with bandwidth $1/T$ (which also indicates the Nyquist rate). We then consider a spectrum sensing application, where the task is to sense the power spectrum of $x(t)$. Fig. 3.1 depicts the employed sampling device, which can be regarded as one possible implementation of an analog to information converter (AIC) in a compressive sampling operation. However, note that this sampling device is capable

of modeling any AIC implementation, such as those proposed in [18, 69]. The considered sampling device has M branches, where the ith branch modulates the signal $x(t)$ with a possibly complex-valued periodic waveform $p_i(t)$ of period NT followed by an integrate-and-dump device with period NT (thus with rate equal to $1/N$ times the Nyquist rate). The output of the ith branch at the kth sampling index can thus be expressed as

$$y_i[k] = \frac{1}{NT}\int_{kNT}^{(k+1)NT} p_i(t)x(t)dt = \frac{1}{T}\int_{kNT}^{(k+1)NT} c_i(t-kNT)x(t)dt \quad (3.1)$$

where $c_i(t)$ yields a single period of $\frac{1}{N}p_i(t)$ i.e., $c_i(t) = \frac{1}{N}p_i(t)$ for $0 \leq t < NT$ and $c_i(t) = 0$ elsewhere. Assume now that $c_i(t)$ is a piecewise constant function having constant values in every interval of length T, i.e., $c_i(t) = c_i[-n]$ for $nT \leq t < (n+1)T$, where $n = 0, 1, \ldots, N-1$. Then, (3.1) can be rewritten as

$$\begin{aligned} y_i[k] &= \sum_{n=0}^{N-1} c_i[-n]\frac{1}{T}\int_{(kN+n)T}^{(kN+n+1)T} x(t)dt = \sum_{n=0}^{N-1} c_i[-n]x[kN+n] \\ &= \sum_{n=1-N}^{0} c_i[n]x[kN-n] \end{aligned} \quad (3.2)$$

where $x[n]$ can be viewed as the output of an integrate-and-dump process with period T (thus with rate equal to the Nyquist rate) applied to $x(t)$, which is not explicitly computed due to its high complexity. Note that the average sampling rate of this periodic sampler is given by the Nyquist rate multiplied by M/N and hence we will use $M < N$ to keep the complexity low. The presented sampling device is actually similar to the modulated wideband converter introduced in [20], where the values of $c_i[n]$ are randomly generated, e.g., adopting complex Gaussian sampling or random binary (from the set $\{\pm 1\}$) sampling. However, the sampler coefficients $c_i[n]$ can also be set to implement efficient multi-coset sampling, which will be discussed in more detail in Section 3.5.

Fig. 3.2 underlines the important fact that (3.2) can actually be perceived as a digital filtering operation of $x[n]$ by the filter $c_i[n]$ of length N followed by an N-fold decimation, i.e., $y_i[k] = z_i[kN]$, where

$$z_i[n] = c_i[n] \star x[n] = \sum_{m=1-N}^{0} c_i[m]x[n-m]$$

with $\star$ representing the convolution operator. This observation turns out to be useful for the reconstruction process.

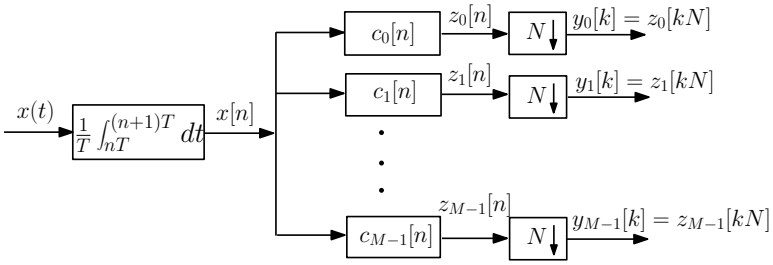


Figure 3.2: Digital interpretation of the sampling device of Fig. 3.1, consisting of a high-rate integrate-and-dump process, followed by a bank of M branches, where each branch consists of a digital filtering operation followed by a downsampling operation.

The goal of this paper is to reconstruct the power spectrum of $x(t)$ based on the obtained samples $\{y_i[k]\}_{i,k}$. Since $x[n]$ is obtained from $x(t)$ by an integrate-and-dump device operating at Nyquist-rate, the spectrum of $x[n]$ is given by a periodic extension of a slightly changed version of the spectrum of $x(t)$ without aliasing. As a result, the power spectrum of $x[n]$ is uniquely determined by the power spectrum of $x(t)$ and vice versa, and thus we will concentrate on reconstructing the power spectrum of $x[n]$ in this paper.

It is well-known that the power spectrum or power spectral density (PSD) of $x[n]$ is given by

$$P_x(\omega) = \sum_{n=-\infty}^{\infty} r_x[n] e^{-jn\omega}, \quad 0 \leq \omega < 2\pi$$

where $r_x[n]$ represents the auto-correlation function of $x[n]$, defined as $r_x[n] = E(x[m]x^*[m-n])$. Therefore, estimating the power spectrum $P_x(\omega)$ amounts to estimating the auto-correlation function $r_x[n]$. The major contribution of this work is that we will take advantage of all the M^2 different cross-spectra of $y_i[k]$ with $y_j[k]$ for $i, j = 0, 1, \ldots, M-1$, which will enable rate-compression without introducing any sparsity constraints on $x(t)$. Note that the cross-spectrum or cross spectral density (CSD) of $y_i[k]$ with $y_j[k]$ is given by

$$P_{y_i,y_j}(\omega) = \sum_{k=-\infty}^{\infty} r_{y_i,y_j}[k] e^{-jk\omega}, \quad 0 \leq \omega < 2\pi$$

where $r_{y_i,y_j}[k] = E(y_i[l]y_j^*[l-k])$ is the cross-correlation function of $y_i[k]$ with $y_j[k]$. These ensemble quantities $\{r_{y_i,y_j}[k]\}_{i,j}$ can be estimated by their sample averages, which in turn result in estimates of $\{P_{y_i,y_j}(\omega)\}_{i,j}$. In the following sections,

we will first describe a time-domain approach to reconstruct $r_x[n]$ given $r_{y_i,y_j}[k]$ for $i,j = 0,1,\ldots,M-1$. Next, a frequency-domain approach to estimate $P_x(\omega)$ given $P_{y_i,y_j}(\omega)$ for $i,j = 0,1,\ldots,M-1$ will be discussed.

3.3 Time-Domain Reconstruction Approach

3.3.1 Reconstruction Analysis

In this subsection, a method to reconstruct $r_x[n]$ given $r_{y_i,y_j}[k]$ for $i,j = 0,1,\ldots,M-1$ is presented. Since $y_i[k] = z_i[kN]$, the cross-correlation function of $y_i[k]$ with $y_j[k]$ can be expressed as the N-fold decimated version of the cross-correlation function of $z_i[n]$ with $z_j[n]$, as follows:

$$r_{y_i,y_j}[k] = E\left(y_i[l]y_j^*[l-k]\right) = E\left(z_i[lN]z_j^*[(l-k)N]\right) = r_{z_i,z_j}[kN]. \quad (3.3)$$

It is obvious that $r_{z_i,z_j}[n]$ can be written as

$$r_{z_i,z_j}[n] = r_{c_i,c_j}[n] \star r_x[n] = \sum_{m=-N+1}^{N-1} r_{c_i,c_j}[m]r_x[n-m] \quad (3.4)$$

where $r_{c_i,c_j}[n]$ is the "deterministic" cross-correlation function between $c_i[n]$ and $c_j[n]$:

$$r_{c_i,c_j}[n] = c_i[n] \star c_j^*[-n] = \sum_{m=1-N}^{0} c_i[m]c_j^*[m-n]. \quad (3.5)$$

From (3.3) and (3.4), we obtain:

$$r_{y_i,y_j}[k] = r_{z_i,z_j}[kN] = \sum_{m=-N+1}^{N-1} r_{c_i,c_j}[m]r_x[kN-m] = \sum_{l=0}^{1} \mathbf{r}_{c_i,c_j}^T[l]\mathbf{r}_x[k-l] \quad (3.6)$$

which is based on the following definitions:

$$\mathbf{r}_{c_i,c_j}[0] = [r_{c_i,c_j}[0], r_{c_i,c_j}[-1], \ldots, r_{c_i,c_j}[-N+1]]^T \quad (3.7)$$

$$\mathbf{r}_{c_i,c_j}[1] = [r_{c_i,c_j}[N], r_{c_i,c_j}[N-1], \ldots, r_{c_i,c_j}[1]]^T \quad (3.8)$$

$$\mathbf{r}_x[k] = [r_x[kN], r_x[kN+1], \ldots, r_x[(k+1)N-1]]^T. \quad (3.9)$$

By cascading the M^2 different cross-correlation functions $r_{y_i,y_j}[k]$, we obtain the $M^2 \times 1$ vector $\mathbf{r}_y[k] = [\ldots, r_{y_i,y_j}[k], \ldots]^T$, for $i,j = 0,1,\ldots,M-1$, which can

be derived from (3.6) as

$$\mathbf{r}_y[k] = \sum_{l=0}^{1} \mathbf{R}_c[l]\mathbf{r}_x[k-l] \tag{3.10}$$

where $\mathbf{R}_c[0]$ and $\mathbf{R}_c[1]$ are the $M^2 \times N$ matrices given by $\mathbf{R}_c[0] = [\dots, \mathbf{r}_{c_i,c_j}[0], \dots]^T$ and $\mathbf{R}_c[1] = [\dots, \mathbf{r}_{c_i,c_j}[1], \dots]^T$, respectively, for $i, j = 0, 1, \dots, M-1$.

Due to the bandlimitedness of $x[n]$, $\mathbf{r}_y[k]$ basically has unlimited support. However, in many practical situations, $\mathbf{r}_y[k]$ only has significant values within a range $-L \leq k \leq L$ and negligible values outside this range, where L is a design parameter that can be chosen as large as required. Hence, let us relax the bandlimitedness condition and assume that the support of $\mathbf{r}_y[k]$ is strictly limited to $-L \leq k \leq L$, which is a rather standard approach when computing a cross-spectrum from a cross-correlation function. Since from (3.10), it is clear that $\mathbf{r}_y[k]$ depends on both $\mathbf{r}_x[k]$ and $\mathbf{r}_x[k-1]$, one could think that under the above assumption the support of $\mathbf{r}_x[k]$ is limited to $-L-1 \leq k \leq L$, but that would mean that also $\mathbf{r}_y[-L-1]$ is non-zero. As a consequence, the support of $\mathbf{r}_x[k]$ should also be limited to $-L \leq k \leq L$. All these quantities can be collected into the following vectors:

$$\mathbf{r}_y = [\mathbf{r}_y^T[0], \mathbf{r}_y^T[1], \dots, \mathbf{r}_y^T[L], \mathbf{r}_y^T[-L], \dots, \mathbf{r}_y^T[-1]]^T \tag{3.11}$$

$$\mathbf{r}_x = [\mathbf{r}_x^T[0], \mathbf{r}_x^T[1], \dots, \mathbf{r}_x^T[L], \mathbf{r}_x^T[-L], \dots, \mathbf{r}_x^T[-1]]^T \tag{3.12}$$

where $\mathbf{r}_y$ has size $(2L+1)M^2 \times 1$ and $\mathbf{r}_x$ has size $(2L+1)N \times 1$. Let us further introduce two other important observations. First, based on the definition of $\mathbf{r}_x[k]$ in (3.9), the fact that the support of $\mathbf{r}_x[k]$ is limited to $-L \leq k \leq L$, and the complex conjugate symmetry in $r_x[n]$, it is clear that the support of $r_x[n]$ is limited to $-LN \leq n \leq LN$ and the last $N-1$ entries of $\mathbf{r}_x[L]$ are zero. Second, based on the definition of $\mathbf{r}_{c_i,c_j}[1]$ in (3.8) and the fact that the support of $r_{c_i,c_j}[n]$ is limited to $1-N \leq n \leq N-1$, it is clear that the first column of $\mathbf{R}_c[1]$ is zero. These two observations allow us to write the linear convolution in (3.10) as a circular convolution within $-L \leq k \leq L$, without any additional zero padding. Hence, we can eventually write the relation between $\mathbf{r}_y$ and $\mathbf{r}_x$ as

$$\mathbf{r}_y = \mathbf{R}_c\mathbf{r}_x \tag{3.13}$$

where $\mathbf{R}_c$ is the $(2L+1)M^2 \times (2L+1)N$ matrix given by

$$\mathbf{R}_c = \begin{bmatrix} \mathbf{R}_c[0] & & & & \mathbf{R}_c[1] \\ \mathbf{R}_c[1] & \mathbf{R}_c[0] & & & \\ & \mathbf{R}_c[1] & \mathbf{R}_c[0] & & \\ & & \ddots & \ddots & \\ & & & \mathbf{R}_c[1] & \mathbf{R}_c[0] \end{bmatrix}. \tag{3.14}$$

Note that (3.13) is solvable using least-squares (LS) if $\mathbf{R}_c$ has full column rank, which obviously requires $M^2 \geq N$.

The inverse problem of (3.13) can be further simplified by observing that $\mathbf{R}_c$ is a block circulant matrix with blocks of size $M^2 \times N$, which can easily be converted into a block diagonal matrix $\mathbf{Q}_c$ with blocks of size $M^2 \times N$. This can be carried out solely by using the $(2L+1)$-point (inverse) discrete Fourier transform ((I)DFT):

$$\mathbf{R}_c = (\mathbf{F}_{2L+1}^{-1} \otimes \mathbf{I}_{M^2})\mathbf{Q}_c(\mathbf{F}_{2L+1} \otimes \mathbf{I}_N)$$

where $\otimes$ represents the Kronecker product operation, $\mathbf{F}_{2L+1}$ is the $(2L+1) \times (2L+1)$ DFT matrix, and $\mathbf{Q}_c = \text{diag}\{\mathbf{Q}_c(0), \mathbf{Q}_c(2\pi\frac{1}{2L+1}), \ldots, \mathbf{Q}_c(2\pi\frac{2L}{2L+1})\}$ with $\mathbf{Q}_c(\omega)$ being the $M^2 \times N$ matrix spectrum of the $M^2 \times N$ matrix sequence $\{\mathbf{R}_c[k]\}_{k=0,1}$:

$$\mathbf{Q}_c(\omega) = \sum_{k=0}^{1} \mathbf{R}_c[k]e^{-jk\omega}. \tag{3.15}$$

Consequently, by defining the $(2L+1)N \times 1$ vector $\mathbf{q}_x$ and the $(2L+1)M^2 \times 1$ vector $\mathbf{q}_y$ as

$$\mathbf{q}_x = (\mathbf{F}_{2L+1} \otimes \mathbf{I}_N)\mathbf{r}_x \tag{3.16}$$

$$\mathbf{q}_y = (\mathbf{F}_{2L+1} \otimes \mathbf{I}_{M^2})\mathbf{r}_y \tag{3.17}$$

we can re-express (3.13) as

$$\mathbf{q}_y = \mathbf{Q}_c\mathbf{q}_x. \tag{3.18}$$

From (3.11), (3.12), (3.16), and (3.17), we can also write $\mathbf{q}_x$ and $\mathbf{q}_y$ as

$$\mathbf{q}_x = [\mathbf{q}_x^T(0), \mathbf{q}_x^T(2\pi\tfrac{1}{2L+1}), \ldots, \mathbf{q}_x^T(2\pi\tfrac{2L}{2L+1})]^T$$
$$\mathbf{q}_y = [\mathbf{q}_y^T(0), \mathbf{q}_y^T(2\pi\tfrac{1}{2L+1}), \ldots, \mathbf{q}_y^T(2\pi\tfrac{2L}{2L+1})]^T$$

where $\mathbf{q}_x(\omega)$ and $\mathbf{q}_y(\omega)$ are respectively the $N \times 1$ and $M^2 \times 1$ vector spectra of the $N \times 1$ and $M^2 \times 1$ vector sequences $\mathbf{r}_x[k]$ and $\mathbf{r}_y[k]$:

$$\mathbf{q}_y(\omega) = \sum_{k=-L}^{L} \mathbf{r}_y[k]e^{-jk\omega} \qquad \mathbf{q}_x(\omega) = \sum_{k=-L}^{L} \mathbf{r}_x[k]e^{-jk\omega}. \tag{3.19}$$

Here, $\mathbf{q}_y$ is obtained from the sample-averaged version of $\mathbf{r}_y[k]$, while $\mathbf{q}_x$, and thus $\mathbf{r}_x[k]$, is to be estimated. By using (3.19), we are able to rewrite (3.18) as a set of $2L+1$ matrix equations:

$$\mathbf{q}_y(2\pi\tfrac{l}{2L+1}) = \mathbf{Q}_c(2\pi\tfrac{l}{2L+1})\mathbf{q}_x(2\pi\tfrac{l}{2L+1}),\ l = 0, 1, \ldots, 2L. \tag{3.20}$$

If $\mathbf{Q}_c(2\pi\frac{l}{2L+1})$ has full column rank for $l = 0, 1, \ldots, 2L$, we can compute $\mathbf{q}_x(2\pi\frac{l}{2L+1})$ using LS from (3.20) for $l = 0, 1, \ldots, 2L$. Note that the above simplification unintentionally transformed the time-domain approach in some kind of frequency-domain approach. However, we can also directly start from the frequency domain, as indicated in Section 3.4. Having estimated $\mathbf{q}_x$, we can reconstruct $\mathbf{r}_x$ using (3.16) and then compute the $(2L+1)N \times 1$ power spectrum vector $\mathbf{s}_x$ as

$$\mathbf{s}_x = \mathbf{F}_{(2L+1)N}\mathbf{r}_x \tag{3.21}$$

where $\mathbf{s}_x = [P_x(0), P_x(2\pi\frac{1}{(2L+1)N}), \ldots, P_x(2\pi\frac{(2L+1)N-1}{(2L+1)N})]^T$ and $\mathbf{F}_{(2L+1)N}$ is the $(2L+1)N \times (2L+1)N$ DFT matrix.

3.3.2 Alternative Time-Domain Approach

In this subsection, we present an alternative yet different version of the time-domain reconstruction approach presented in Section 3.3.1. We start by rewriting (3.2) in matrix-vector notation

$$\mathbf{y}[k] = \mathbf{C}\mathbf{x}[k] \tag{3.22}$$

where the $M \times 1$ measurement vectors $\mathbf{y}[k]$ and the $N \times 1$ vector sequence $\mathbf{x}[k]$ are, respectively, defined as:

$$\mathbf{y}[k] = [y_0[k], y_1[k], \ldots, y_{M-1}[k]]^T \tag{3.23}$$

$$\mathbf{x}[k] = [x[kN], x[kN+1], \ldots, x[kN+N-1]]^T \tag{3.24}$$

while the $M \times N$ compressive sampling matrix $\mathbf{C}$ is given by:

$$\mathbf{C} = [\mathbf{c}_0, \mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_{M-1}]^T$$

with $\mathbf{c}_i = [c_i[0], c_i[-1], \ldots, c_i[1-N]]^T$.

Next, we compute the $M \times M$ auto-correlation matrix of $\mathbf{y}[k]$ in (3.23), which is given by $\mathbf{R}_y[0] = E\left(\mathbf{y}[k]\mathbf{y}^H[k]\right)$. If we also construct the $N \times N$ auto-correlation matrix of $\mathbf{x}[k]$ in (3.24) as $\mathbf{R}_x^a = E\left(\mathbf{x}[k]\mathbf{x}^H[k]\right)$, the relationship between $\mathbf{R}_x^a$ and $\mathbf{R}_y[0]$ can be expressed as:

$$\mathbf{R}_y[0] = \mathbf{C}\mathbf{R}_x^a\mathbf{C}^H \tag{3.25}$$

where the elements of $\mathbf{R}_x^a$ are given by $[\mathbf{R}_x^a]_{ij} = r_x[i-j] = r_x^*[j-i]$ due to the wide-sense stationary property of $x[n]$. While $\mathbf{R}_x^a$ has a Toeplitz structure, this is not the case for $\mathbf{R}_y[0]$ because the elements of $\mathbf{y}[k]$ in (3.23) are generally not wide-sense stationary due to the nature of the compressive sampling matrix $\mathbf{C}$. Therefore, it is theoretically possible to exploit all columns of $\mathbf{R}_y[0]$ to estimate one of the columns of $\mathbf{R}_x^a$ by first stacking all columns of $\mathbf{R}_y[0]$ into the $M^2 \times 1$ vector $\text{vec}(\mathbf{R}_y[0])$, which is nothing else than $\mathbf{r}_y[0]$ in (3.11). Here, $\text{vec}(.)$ is the operator that stacks all columns of a matrix in a large column vector. Based on (3.25), it is evident that $\mathbf{r}_y[0]$ in (3.11) can be written as:

$$\mathbf{r}_y[0] = \text{vec}(\mathbf{R}_y[0]) = (\mathbf{C}^* \otimes \mathbf{C})\text{vec}(\mathbf{R}_x^a). \tag{3.26}$$

Since all columns of $\mathbf{R}_x^a$ contain the same information, $\mathbf{R}_x^a$ can be condensed into the $(2N-1)\times 1$ vector $\mathbf{r}_x^a = [r_x[0], r_x[1], \ldots, r_x[N-1], r_x[1-N], \ldots, r_x[-1]]^T$, and we can write

$$\text{vec}(\mathbf{R}_x^a) = \mathbf{T}\mathbf{r}_x^a \tag{3.27}$$

where $\mathbf{T}$ is a special $N^2 \times (2N-1)$ repetition matrix with the i-th row of $\mathbf{T}$ given by the $((i-1+(N-2)\left\lfloor \frac{i-1}{N} \right\rfloor) \bmod (2N-1)+1)$-th row of the identity matrix $\mathbf{I}_{2N-1}$. By combining (3.26) and (3.27), we obtain:

$$\mathbf{r}_y[0] = (\mathbf{C}^* \otimes \mathbf{C})\mathbf{T}\mathbf{r}_x^a = \mathbf{R}_c^a \mathbf{r}_x^a \tag{3.28}$$

where the $M^2 \times (2N-1)$ matrix $\mathbf{R}_c^a = (\mathbf{C}^* \otimes \mathbf{C})\mathbf{T}$ is actually given by:

$$\mathbf{R}_c^a = [\mathbf{R}_c^a[0], \mathbf{R}_c^a[1]] \tag{3.29}$$

with $\mathbf{R}_c^a[0]$ equal to $\mathbf{R}_c[0]$ and $\mathbf{R}_c^a[1]$ obtained by removing the first column of $\mathbf{R}_c[1]$ (which actually has zero entries). If $\mathbf{R}_c^a$ has full column rank, it is possible to reconstruct the auto-correlation vector $\mathbf{r}_x^a$ from (3.28) using LS. Then, we can compute the power spectrum vector $\mathbf{s}_x$ as in (3.21) by replacing $\mathbf{r}_x$ with a zero-padded version of $\mathbf{r}_x^a$ denoted as $\tilde{\mathbf{r}}_x^a$, i.e., $\mathbf{s}_x = \mathbf{F}_{(2L+1)N}\tilde{\mathbf{r}}_x^a$. Note that as the time-domain approach only gives a valid power spectrum estimate when $r_x[n]$ has negligible correlation values above lag LN, where L can be freely selected, the alternative time-domain approach should only be preferred if $r_x[n]$ has negligible correlation values above lag $N-1$. Hence, for a fixed N, e.g., when the sampler is fixed, this is a clear disadvantage of the alternative approach.

Comparing the alternative approach to (3.13), we further observe that it is a different method that can not merely be obtained from the time-domain approach by setting $L = 0$ in (3.11) and (3.12). If we set $L = 0$ in (3.11), the support of

$\mathbf{r}_y[k]$ is limited to $k = 0$. However, as we explained in Section 3.3.1, the support of $\mathbf{r}_x[k]$ is then also limited to $k = 0$. Further, from the complex conjugate symmetry in $r_x[n]$, we can then conclude that setting $L = 0$ is only possible when $x[n]$ is assumed to be a white sequence, i.e., $r_x[n] = r_x[0]\delta[n]$. Hence, the minimum possible value of L in (3.11) and (3.12) is $L = 1$.

3.4 Frequency-Domain Reconstruction Approach

In this section, we develop a frequency-domain reconstruction approach. The reason for presenting this frequency-domain method is its tight connection to spectrum blind sampling (SBS) presented in [64, 20, 38]. SBS also starts from a frequency-domain viewpoint but focuses on spectrum reconstruction instead of power spectrum reconstruction.

Since $y_i[k] = z_i[kN]$, we can also write the cross-spectrum of $y_i[k]$ with $y_j[k]$ as an N-fold aliased version of the cross-spectrum of $z_i[n]$ and $z_j[n]$:

$$P_{y_i,y_j}(\omega) = \frac{1}{N}\sum_{n=0}^{N-1} P_{z_i,z_j}(\omega\tfrac{1}{N} + 2\pi\tfrac{n}{N}),\ 0 \le \omega < 2\pi. \tag{3.30}$$

It is well-known that $P_{z_i,z_j}(\omega)$ can be written as

$$P_{z_i,z_j}(\omega) = P_{c_i,c_j}(\omega)P_x(\omega),\ 0 \le \omega < 2\pi \tag{3.31}$$

where $P_{c_i,c_j}(\omega)$ is the "deterministic" cross-spectrum between $c_i[n]$ and $c_j[n]$:

$$P_{c_i,c_j}(\omega) = C_i(\omega)C_j^*(\omega),\ 0 \le \omega < 2\pi.$$

From (3.30) and (3.31), we can thus write

$$\begin{aligned} P_{y_i,y_j}(\omega) &= \frac{1}{N}\sum_{n=0}^{N-1} P_{c_i,c_j}(\omega\tfrac{1}{N} + 2\pi\tfrac{n}{N})P_x(\omega\tfrac{1}{N} + 2\pi\tfrac{n}{N}) \\ &= \mathbf{p}_{c_i,c_j}^T(\omega)\mathbf{p}_x(\omega),\ 0 \le \omega < 2\pi \end{aligned}$$

where we have that

$$\begin{aligned} \mathbf{p}_{c_i,c_j}(\omega) &= \frac{1}{N}[P_{c_i,c_j}(\omega\tfrac{1}{N}), \ldots, P_{c_i,c_j}(\omega\tfrac{1}{N} + 2\pi\tfrac{N-1}{N})]^T \\ \mathbf{p}_x(\omega) &= [P_x(\omega\tfrac{1}{N}), \ldots, P_x(\omega\tfrac{1}{N} + 2\pi\tfrac{N-1}{N})]^T. \end{aligned} \tag{3.32}$$

Note that reconstructing $\mathbf{p}_x(\omega)$ for $0 \le \omega < 2\pi$ is equivalent to reconstructing $P_x(\omega)$ for $0 \le \omega < 2\pi$.

Stacking the M^2 different cross-spectra $P_{y_i,y_j}(\omega)$ in the $M^2 \times 1$ vector $\mathbf{p}_y(\omega) = [\ldots, P_{y_i,y_j}(\omega), \ldots]^T$, for $i, j = 0, 1, \ldots, M-1$, we finally obtain

$$\mathbf{p}_y(\omega) = \mathbf{P}_c(\omega)\mathbf{p}_x(\omega),\ 0 \leq \omega < 2\pi \tag{3.33}$$

where $\mathbf{P}_c(\omega)$ is the $M^2 \times N$ matrix given by $\mathbf{P}_c(\omega) = [\ldots, \mathbf{p}_{c_i,c_j}(\omega), \ldots]^T$, for $i, j = 0, 1, \ldots, M-1$. Assuming that $\mathbf{P}_c(\omega)$ has full column rank for $0 \leq \omega < 2\pi$, we can solve (3.33) for $0 \leq \omega < 2\pi$ using LS.

The above approach allows us to estimate the power spectrum at any specific frequency, and as such does not require any limits on the support of $\mathbf{r}_y[k]$ and thus on the support of $r_x[n]$. However, to compute $\mathbf{p}_y(\omega)$ from $\mathbf{r}_y[k]$ in practice, the support of $\mathbf{r}_y[k]$ has to be truncated. So we could again, as before, relax the bandlimitness condition and assume that the support of $\mathbf{r}_y[k]$ and thus $\mathbf{r}_x[k]$ is limited to $-L \leq k \leq L$. In that case, to reconstruct the overall power spectrum $P_x(\omega)$, it suffices to compute $\mathbf{p}_x(\omega)$ for $\omega = 0, 2\pi\frac{1}{2L+1}, \ldots, 2\pi\frac{2L}{2L+1}$, which from (3.33) is completely determined by $\mathbf{p}_y(\omega)$ for $\omega = 0, 2\pi\frac{1}{2L+1}, \ldots, 2\pi\frac{2L}{2L+1}$. It can be shown that such an approach would be exactly equivalent to the special form of the time-domain approach presented in (3.20).

3.5 Minimal Sparse Ruler Sampling

To ensure the uniqueness of the LS solution of (3.13) and (3.20), many different implementations of the considered sampling procedure can be investigated. Although many types of random modulating waveforms can be studied [20], such as complex Gaussian sampling or random binary (from the set $\{\pm 1\}$) sampling, we mainly focus on multi-coset sampling in this paper. More specifically, we propose some new multi-coset implementations based on the so-called minimal sparse ruler problem, which we will label as minimal sparse ruler sampling.

Observing (3.2), multi-coset sampling can be implemented by simply setting for every branch i, one different entry of $c_i[n]$ to one and the others to zero, i.e., $c_i[n] = 1$ if $-n = n_i$ and $c_i[n] = 0$ if $-n \neq n_i$, where $n_i \neq n_j$ whenever $i \neq j$. Concisely, $c_i[n] = \delta[-n - n_i]$, where $n_i \neq n_j, \forall i \neq j$. This is actually identical to selecting M different rows from the identity matrix $\mathbf{I}_N$. However, note that this row selection cannot be random, because we need to deterministically guarantee the full column rank of $\mathbf{R}_c$ in (3.13) or equivalently of $\{\mathbf{Q}_c(2\pi\frac{l}{2L+1})\}_{l=0}^{2L}$ in (3.20). Observe that every row of $\mathbf{R}_c$ only contains a single one, which means that the full rank conditions can be fulfilled by ensuring that $\mathbf{R}_c$ has at least a single one in each of its columns. We can find from (3.5), (3.7) and (3.8) that when $\mathbf{R}_c[0]$ has a one in

the column corresponding to lag $-n$, $\mathbf{R}_c[1]$ has a one in the column corresponding to lag n. As a result, if the first $\lfloor \frac{N}{2} \rfloor + 1$ columns of $\mathbf{R}_c[0]$ all have at least a single one, then also the last $\lfloor \frac{N}{2} \rfloor$ columns of $\mathbf{R}_c[1]$ all have at least a single one, where $\lfloor x \rfloor$ represents the largest integer not greater than x. Hence, a sufficient condition to guarantee that all columns of $\mathbf{R}_c$ have at least a single one can be achieved by ensuring that the first $\lfloor \frac{N}{2} \rfloor + 1$ columns of $\mathbf{R}_c[0]$ all have at least a single one. Therefore, the problem we now like to solve is how to choose a proper combination of rows of $\mathbf{I}_N$ to generate the coefficients of $c_i[n]$ for $i = 0, 1, \ldots, M-1$, such that $\mathbf{R}_c[0]$ has at least a single one in each of its first $\lfloor \frac{N}{2} \rfloor + 1$ columns. Further, note that the aim is to keep the number of selected rows minimal, in order to minimize the number of branches M and thus to minimize the compression rate M/N.

Since $c_i[n] = \delta[-n - n_i]$, it is obvious from (3.5) that

$$r_{c_i,c_j}[n] = \delta[n + n_i - n_j] \tag{3.34}$$

which depends on the differences $n_i - n_j$. By introducing S as a set of M indices selected from $\{0, 1, \ldots, N-1\}$, representing the row indices of $\mathbf{I}_N$ selected by the multi-coset sampler, and Ω as the set of related index-differences, given by $\Omega = \{|n_i - n_j| \,|\forall n_i, n_j \in S\}$, the problem of constructing the sampler coefficients $\{c_i[n]\}_{i=0}^{M-1}$ becomes:

$$\min_{S} |S| \quad \text{s.t.} \quad \left\{0, 1, \ldots, \lfloor \tfrac{N}{2} \rfloor\right\} \subset \Omega \tag{3.35}$$

where $|S|$ represents the cardinality of the set S. While the solution of (3.35) can be found by exhaustive or greedy search procedures, one possible way to find a suboptimal solution of (3.35) is by reformulating the problem as a so-called minimal length-$\lfloor \frac{N}{2} \rfloor$ sparse ruler problem, which has been well-studied. This is done by introducing S' as a set of M indices selected from $\{0, 1, \ldots, \lfloor \frac{N}{2} \rfloor\}$ and Ω' as the set of related index-differences, given by $\Omega' = \{|n_i - n_j| \,|\forall n_i, n_j \in S'\}$, and by solving:

$$\min_{S'} \left|S'\right| \quad \text{s.t.} \quad \Omega' = \left\{0, 1, \ldots, \lfloor \tfrac{N}{2} \rfloor\right\}. \tag{3.36}$$

A sparse ruler with length $\lfloor \frac{N}{2} \rfloor$ can be regarded as a ruler having $k < \lfloor \frac{N}{2} \rfloor + 1$ distance marks $0 = n_0 < n_1 < \ldots < n_{k-1} = \lfloor \frac{N}{2} \rfloor$, but is still able to measure all integer distances from 0 up to $\lfloor \frac{N}{2} \rfloor$. The length-$\lfloor \frac{N}{2} \rfloor$ sparse ruler having k distance marks is called minimal if there is no length-$\lfloor \frac{N}{2} \rfloor$ sparse ruler having $k-1$ marks. The minimal sparse ruler problem has for instance been investigated in [70]. Many exact and approximate solutions for the sparse ruler problem have been pre-computed and tabulated. By making the connection between the

sparse ruler problem and our multi-coset design problem, the sampler coefficients $\{c_i[n]\}_{i=0}^{M-1}$ can be constructed using any known sparse ruler, which guarantees the full rank property of $\mathbf{R}_c$ and thus the uniqueness of the simple LS solution to power spectrum reconstruction.

For the alternative time-domain approach, the aim is to ensure the uniqueness of the LS solution of (3.28), which can be achieved if $\mathbf{R}_c^a$ in (3.28) has full column rank. In the case of multi-coset sampling, the full rank condition of $\mathbf{R}_c^a$ can be achieved if each column of $\mathbf{R}_c^a$ has at least a single one because we know that every row of $\mathbf{R}_c^a$ will only contain a single one by considering (3.5), (3.7), (3.8), and (3.29). Following the same analysis as in the previous paragraphs, the problem of constructing $\{c_i[n]\}_{i=0}^{M-1}$ while ensuring that $\mathbf{R}_c^a$ has at least a single one in each of its columns boils down to solving a minimal length-$(N-1)$ sparse ruler problem. For the same N, this obviously leads to a worse compression rate M/N than for the time-domain approach. However, while the minimal length-$\lfloor \frac{N}{2} \rfloor$ sparse ruler only provides a suboptimal solution for the time-domain approach, the minimal length-$(N-1)$ sparse ruler offers the minimum possible compression rate M/N for the alternative time-domain approach. This can easily be verified.

Table 3.1 shows some examples of minimal sparse ruler samplers for the time-domain (TD) and alternative time-domain (ATD) approaches.

Table 3.1: Examples of minimal sparse rulers (TD = time domain approach, ATD = alternative time domain approach)

N	length-$\lfloor \frac{N}{2} \rfloor$	M for length-$\lfloor \frac{N}{2} \rfloor$	M/N for TD	length-$(N-1)$	M for length-$(N-1)$	M/N for ATD
2	1	2	1	1	2	1
11	5	4	0.3636	10	6	0.5455
18	9	5	0.2778	17	7	0.3889
39	19	8	0.2051	38	11	0.2821
78	39	11	0.1410	77	15	0.1923
84	42	11	0.1310	83	16	0.1905
128	64	14	0.1094	127	20	0.1563

3.6 Estimation and Detection Performance

In this section, we evaluate the estimation and detection performance of the proposed power spectrum estimators. We first derive the mean and the covariance of

the estimated power spectrum in (3.21). Based on the derived mean and covariance, we then formulate the analytical normalized mean squared error (NMSE) of the estimate. To evaluate the detection performance, we assume the received sequence $x[n]$ only contains circular complex zero-mean Gaussian i.i.d. noise. Based on this assumption, we simplify the earlier derived mean and covariance of the estimator and we show that asymptotically, for a sufficient number of measurements, the estimated power spectrum is Gaussian as well. Based on these results, we can finally formulate the decision threshold for a constant false alarm rate.

3.6.1 Estimation Performance

Given K measurement vectors $\mathbf{y}[k]$, the unbiased estimate for $r_{y_i,y_j}[k]$ in (3.6) can be written as:

$$\hat{r}_{y_i,y_j}[k] = \frac{1}{K-|k|} \sum_{l=\max(0,k)}^{K-1+\min(0,k)} y_i[l] y_j^*[l-k] \tag{3.37}$$

where $\max(a,b)$ and $\min(a,b)$ gives the largest and smallest value of a and b, respectively.

Obviously, $E\left(\hat{r}_{y_i,y_j}[k]\right) = r_{y_i,y_j}[k]$ since $\hat{r}_{y_i,y_j}[k]$ is an unbiased estimate. Following Section 3.3.1, we can now compute the covariance matrix of the estimate $\hat{\mathbf{r}}_y$ for $\mathbf{r}_y$ in (3.11), which is given by the $(2L+1)M^2 \times (2L+1)M^2$ matrix $\mathbf{C}_{\hat{r}_y} = E\left(\hat{\mathbf{r}}_y \hat{\mathbf{r}}_y^H\right) - E\left(\hat{\mathbf{r}}_y\right) E\left(\hat{\mathbf{r}}_y^H\right)$. The elements of $\mathbf{C}_{\hat{r}_y}$ are given by:

$$\begin{aligned} \mathrm{Cov}(\hat{r}_{y_i,y_j}[k], \hat{r}_{y_w,y_v}[q]) &= E\left(\hat{r}_{y_i,y_j}[k] \hat{r}^*_{y_w,y_v}[q]\right) - E\left(\hat{r}_{y_i,y_j}[k]\right) E\left(\hat{r}^*_{y_w,y_v}[q]\right) \\ &= E\left(\hat{r}_{y_i,y_j}[k] \hat{r}^*_{y_w,y_v}[q]\right) - r_{y_i,y_j}[k] r^*_{y_w,y_v}[q] \end{aligned} \tag{3.38}$$

where $0 \leq i,j,w,v \leq M-1$, $-L \leq k,q \leq L$, and $E\left(\hat{r}_{y_i,y_j}[k] \hat{r}^*_{y_w,y_v}[q]\right)$ can be expressed as:

$$\begin{aligned} E\left(\hat{r}_{y_i,y_j}[k] \hat{r}^*_{y_w,y_v}[q]\right) &= E\left(\left(\frac{1}{K-|k|} \sum_{l=\max(0,k)}^{K-1+\min(0,k)} y_i[l] y_j^*[l-k]\right)\right. \\ &\left. \times \left(\frac{1}{K-|q|} \sum_{p=\max(0,q)}^{K-1+\min(0,q)} y_w^*[p] y_v[p-q]\right)\right) \\ &= \frac{1}{(K-|k|)(K-|q|)} \sum_{l=\max(0,k)}^{K-1+\min(0,k)} \sum_{p=\max(0,q)}^{K-1+\min(0,q)} \end{aligned}$$

$$E\left(y_i[l]y_j^*[l-k]y_w^*[p]y_v[p-q]\right). \tag{3.39}$$

Observe that the computation of $\mathbf{C}_{\hat{r}_y}$ is not trivial since it involves the computation of fourth order moments. Knowledge about the distribution of the received sequence $x[n]$ is thus required. For example, if $x[n]$ is Gaussian distributed, then the sequences $y_i[k]$ are also jointly Gaussian and the fourth order moments in (3.39) can be simplified as the sum of products of second order moments [71].

Based on the statistical properties of $\hat{\mathbf{r}}_y$, we can now compute the expected value and the covariance of the recovered power spectrum. Denote the estimated power spectrum for $\mathbf{s}_x$ in (3.21) by $\hat{\mathbf{s}}_x$. From (3.13) and (3.21), assuming that $\mathbf{R}_c$ has full column rank, the relationship between $\hat{\mathbf{s}}_x$ and $\hat{\mathbf{r}}_y$ is given by:

$$\hat{\mathbf{s}}_x = \mathbf{F}_{(2L+1)N}\mathbf{R}_c^{\dagger}\hat{\mathbf{r}}_y \tag{3.40}$$

where $\mathbf{R}_c^{\dagger} = (\mathbf{R}_c^H\mathbf{R}_c)^{-1}\mathbf{R}_c^H$. The expected value of $\hat{\mathbf{s}}_x$ is thus given by:

$$E\left(\hat{\mathbf{s}}_x\right) = \mathbf{F}_{(2L+1)N}\mathbf{R}_c^{\dagger}E\left(\hat{\mathbf{r}}_y\right) = \mathbf{F}_{(2L+1)N}\mathbf{R}_c^{\dagger}\mathbf{r}_y. \tag{3.41}$$

Correspondingly, we denote the $(2L+1)N \times (2L+1)N$ covariance matrix of $\hat{\mathbf{s}}_x$ by $\mathbf{C}_{\hat{s}_x}$, which can be written as:

$$\mathbf{C}_{\hat{s}_x} = \mathbf{F}_{(2L+1)N}\mathbf{R}_c^{\dagger}\mathbf{C}_{\hat{r}_y}(\mathbf{R}_c^{\dagger})^H\mathbf{F}_{(2L+1)N}^H \tag{3.42}$$

where the elements of $\mathbf{C}_{\hat{r}_y}$ are given by (3.38). Note that the variance of the elements of $\hat{\mathbf{s}}_x$ can be found on the diagonal of $\mathbf{C}_{\hat{s}_x}$.

It is well known that the NMSE of the estimated power spectrum $\hat{\mathbf{s}}_x$ is then given by:

$$\text{NMSE} = \frac{E\left(\|\hat{\mathbf{s}}_x - \mathbf{s}_x\|_2^2\right)}{\|\mathbf{s}_x\|_2^2} = \frac{\text{tr}\left(\mathbf{C}_{\hat{s}_x}\right) + \|E\left(\hat{\mathbf{s}}_x\right) - \mathbf{s}_x\|_2^2}{\|\mathbf{s}_x\|_2^2}, \tag{3.43}$$

where $\text{tr}(\cdot)$ is the trace operator. Since $\hat{\mathbf{s}}_x$ is a linear function of $\hat{\mathbf{r}}_y$ (see (3.40)) and $\hat{\mathbf{r}}_y$ is an unbiased estimate of $\mathbf{r}_y$ (see (3.37)), $\hat{\mathbf{s}}_x$ is an unbiased estimate of $\mathbf{s}_x$ as long as the support of the auto-correlation of the received signal $r_x[n]$ is limited to $-LN \leq n \leq LN$, which is the assumption we adopt in (3.12). When this is the case, the NMSE of $\hat{\mathbf{s}}_x$ is equal to $\text{tr}\left(\mathbf{C}_{\hat{s}_x}\right) / \|\mathbf{s}_x\|_2^2$.

Note that in (3.37), we have used an unbiased estimate $\hat{r}_{y_i,y_j}[k]$ of the cross-correlation $r_{y_i,y_j}[k]$ instead of a biased one. The reason for this can be explained as follows. Using an unbiased estimate instead of a biased one, the realness of the resulting power spectrum estimate $\hat{\mathbf{s}}_x$ is not jeopardized, but the positiveness

of $\hat{\mathbf{s}}_x$ can be affected. However, even when using a biased estimate, i.e., using a normalization factor $1/K$ instead of $1/(K-|k|)$ in (3.37), the positiveness of $\hat{\mathbf{s}}_x$ can generally not be guaranteed, in contrast to the Nyquist-rate sampling case. This is due to the fact that the pseudo-inverse of $\mathbf{R}_c$ is used in (3.40), which does not introduce an additional bias when transforming $\hat{\mathbf{r}}_y$ into $\hat{\mathbf{s}}_x$. That is the main reason why we started from an unbiased estimate from the beginning. We will come back to this issue in Section 3.7.

3.6.2 Constant False Alarm Rate (CFAR) Detection Performance

For the detection performance evaluation, let us assume that the received sequence $x[n]$ in Fig. 3.2 only contains circular complex zero-mean Gaussian i.i.d. noise with variance σ^2, i.e., $E\left(x[n]x^*[m]\right) = \sigma^2\delta[n-m]$ and $E\left(x[n]x[m]\right) = 0$ for all m and n. When this is the case, $E\left(\hat{r}_{y_i,y_j}[k]\right)$ can be simplified to

$$E\left(\hat{r}_{y_i,y_j}[k]\right) = r_{y_i,y_j}[k] = r_{c_i,c_j}[0]\sigma^2\delta[k]. \tag{3.44}$$

The elements of $\mathbf{C}_{\hat{r}_y}$ in (3.42) can then also be simplified according to Appendix 3.A and they are given by:

$$\mathrm{Cov}(\hat{r}_{y_i,y_j}[k], \hat{r}_{y_w,y_v}[q]) = \frac{\sigma^4 r_{c_i,c_w}[0] r^*_{c_j,c_v}[0]\delta[q-k]}{(K-|k|)}. \tag{3.45}$$

The expected value of $\hat{\mathbf{s}}_x$ is then given by (3.41) and the covariance matrix of $\hat{\mathbf{s}}_x$ by (3.42) where the elements of $E\left(\hat{\mathbf{r}}_y\right)$ are given by (3.44) and those of $\mathbf{C}_{\hat{r}_y}$ by (3.45).

Asymptotically, when the number of measurement vectors K is sufficiently large compared to L in (3.11), i.e. $L \ll K$, it can be shown that the Gaussian approximation is applicable for the distribution of each element of $\hat{\mathbf{s}}_x$, when $x[n]$ only contains circular complex zero-mean Gaussian i.i.d. noise. This evaluation on the asymptotic statistical distribution of $\hat{\mathbf{s}}_x$ is provided in Appendix 3.B.

Combining (3.42), (3.45), and the analysis in Appendix 3.B, we are now ready to evaluate the detection problem at frequencies $\omega = 0, 2\pi\frac{1}{(2L+1)N}, \ldots, 2\pi\frac{(2L+1)N-1}{(2L+1)N}$ under the asymptotic Gaussian behavior of $\hat{P}_x(\omega)$. Note that it is always possible to increase the number of grid points in the frequency domain by simply padding zeros to $\hat{\mathbf{r}}_x$, and thus it is always possible to evaluate the detection problem at frequencies other than the above specified frequencies. The detection problem is modeled as a selection between hypothesis $H_{1,\omega}$, which represents the occupancy of frequency ω, and hypothesis $H_{0,\omega}$, which represents the absence of a signal at frequency ω.

Here, we want to maximime the detection probability given the false alarm probability and thus we adopt the Neyman-Pearson theorem [72]. The decision rule is given by:

$$\hat{P}_x(\omega) \underset{H_{0,\omega}}{\overset{H_{1,\omega}}{\gtrless}} \gamma_\omega,$$

where γ_ω is the decision threshold for frequency ω. Let us denote the mean and variance of the noise power at frequency ω as $\mu_{n,\omega}$ and $\sigma^2_{n,\omega}$, respectively. Note that the value of $\mu_{n,\omega}$ and $\sigma^2_{n,\omega}$ for $\omega = 0, 2\pi\frac{1}{(2L+1)N}, \ldots, 2\pi\frac{(2L+1)N-1}{(2L+1)N}$ can be found as the elements of $E\left(\hat{\mathbf{s}}_x\right)$ in (3.41) and those of the diagonal of $\mathbf{C}_{\hat{s}_x}$ in (3.42) where the elements of $E\left(\hat{\mathbf{r}}_y\right)$ in (3.41) and $\mathbf{C}_{\hat{r}_y}$ in (3.42) are respectively given by (3.44) and (3.45). Due to the asymptotic Gaussian behavior of $\hat{P}_x(\omega)$, the false alarm probability at frequency ω, $p_{fa,\omega}$, given the threshold value γ_ω, is given by:

$$p_{fa,\omega} = Q\left(\frac{\gamma_\omega - \mu_{n,\omega}}{\sigma_{n,\omega}}\right),$$

where $Q(.)$ is the tail probability of the standard Gaussian distribution:

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right) du.$$

As a result, for a given false alarm probability $p_{fa,\omega}$, the decision threshold at frequency ω can be computed as:

$$\gamma_\omega = Q^{-1}\left(p_{fa,\omega}\right)\sigma_{n,\omega} + \mu_{n,\omega}. \tag{3.46}$$

3.6.3 Alternative Time-Domain Approach Case

In this subsection, we again evaluate the estimation and detection performance of the proposed power spectrum estimators but now for the alternative time-domain approach of Section 3.3.2. Note that this approach is not a special case of the time-domain approach and requires a separate analysis. We start the analysis on the estimation performance by considering $\hat{\mathbf{r}}_y[0]$ as an unbiased estimate of $\mathbf{r}_y[0]$ in (3.26). Note that the elements of $\hat{\mathbf{r}}_y[0]$ are simply given by (3.37) with $k = 0$ and it is evident that $E\left(\hat{\mathbf{r}}_y[0]\right) = \mathbf{r}_y[0]$. Next, we derive the covariance matrix of $\hat{\mathbf{r}}_y[0]$, which is given by the $M^2 \times M^2$ matrix $\mathbf{C}_{\hat{r}_y[0]} = E\left(\hat{\mathbf{r}}_y[0]\hat{\mathbf{r}}_y[0]^H\right) - E\left(\hat{\mathbf{r}}_y[0]\right)E\left(\hat{\mathbf{r}}_y[0]^H\right)$ whose elements are simply given by (3.38) with $k = q = 0$.

Based on the statistical properties of $\hat{\mathbf{r}}_y[0]$, we can compute the expected value and the covariance of $\hat{\mathbf{r}}^a_x$, which is the estimate of $\mathbf{r}^a_x$ in (3.28). Assuming that $\mathbf{R}^a_c$

in (3.28) has full column rank, the expected value of $\hat{\mathbf{r}}_x^a$ can be written as:

$$E\left(\hat{\mathbf{r}}_x^a\right) = \mathbf{R}_c^{a\dagger} E\left(\hat{\mathbf{r}}_y[0]\right) = \mathbf{R}_c^{a\dagger} \mathbf{r}_y[0]. \tag{3.47}$$

Correspondingly, we denote the $(2N-1) \times (2N-1)$ covariance matrix of $\hat{\mathbf{r}}_x^a$ by $\mathbf{C}_{\hat{r}_x^a}$, which can be expressed as

$$\mathbf{C}_{\hat{r}_x^a} = \mathbf{R}_c^{a\dagger} \mathbf{C}_{\hat{r}_y[0]} (\mathbf{R}_c^{a\dagger})^H \tag{3.48}$$

where $\mathbf{R}_c^{a\dagger} = (\mathbf{R}_c^{aH} \mathbf{R}_c^a)^{-1} \mathbf{R}_c^{aH}$. Let us denote the zero padded version of $\hat{\mathbf{r}}_x^a$ by $\hat{\tilde{\mathbf{r}}}_{\mathbf{x}}^{\mathbf{a}}$ and replace $E\left(\hat{\mathbf{r}}_x^a\right)$ and $\mathbf{C}_{\hat{r}_x^a}$ by the $(2L+1)N \times 1$ vector $E(\hat{\tilde{\mathbf{r}}}_{\mathbf{x}}^{\mathbf{a}})$ and the $(2L+1)N \times (2L+1)N$ matrix $\mathbf{C}_{\hat{\tilde{r}}_x^a}$. The mean and the covariance matrix of $\hat{\mathbf{s}}_x$ are then given by:

$$E\left(\hat{\mathbf{s}}_x\right) = \mathbf{F}_{(2L+1)N} E(\hat{\tilde{\mathbf{r}}}_{\mathbf{x}}^{\mathbf{a}}) \tag{3.49}$$

and

$$\mathbf{C}_{\hat{s}_x} = \mathbf{F}_{(2L+1)N} \mathbf{C}_{\hat{\tilde{r}}_x^a} \mathbf{F}_{(2L+1)N}^H, \tag{3.50}$$

respectively. Similar to the time-domain approach case, the NMSE of $\hat{\mathbf{s}}_x$ depends on its variance and bias. Since $\hat{\mathbf{s}}_x$ is a linear function of $\hat{\mathbf{r}}_y[0]$, which is an unbiased estimate of $\mathbf{r}_y[0]$, $\hat{\mathbf{s}}_x$ is also an unbiased estimate of $\mathbf{s}_x$ as long as the support of the auto-correlation of the received signal $r_x[n]$ is limited to $-N+1 \leq n \leq N-1$. When this is the case, the NMSE of $\hat{\mathbf{s}}_x$ is again equal to $\mathrm{tr}\left(\mathbf{C}_{\hat{s}_x}\right) / \left\|\mathbf{s}_x\right\|_2^2$.

For the detection performance evaluation, we again assume that the received sequence $x[n]$ in Fig. 3.2 only contains circular complex zero-mean Gaussian i.i.d. noise. When this is the case, the elements of $E\left(\hat{\mathbf{r}}_y[0]\right)$ in (3.47) are given by (3.44) with $k = 0$, and the elements of $\mathbf{C}_{\hat{r}_y[0]}$ in (3.48) are given by (3.45) with $k = q = 0$. Similar to the time-domain approach case, the Gaussian approximation is also applicable for the distribution of each element of $\hat{\mathbf{s}}_x$ under the alternative time-domain approach when $x[n]$ only contains circular complex zero-mean Gaussian i.i.d. noise and as long as K is sufficiently large. This is shown in Appendix 3.C. The derivation of the detection threshold given a fixed false alarm probability for this alternative time-domain approach case follows Section 3.6.2.

3.7 Additional Constraints

So far, we have assumed that the power spectrum $\mathbf{s}_x$ can be estimated without any additional constraints relying on the assumption that there are enough equations available. As we explained in Section 3.6.1, this estimated spectrum is real but not

necessarily positive. Hence, we could think about adding a positivity constraint to our reconstruction problem. To do so, let us first combine (3.16), (3.18), and (3.21) to produce

$$\mathbf{q}_y = \mathbf{\Phi}\mathbf{s}_x \tag{3.51}$$

where $\mathbf{\Phi} = \mathbf{Q}_c(\mathbf{F}_{2L+1} \otimes \mathbf{I}_N)\mathbf{F}^{-1}_{(2L+1)N}$ is of size $(2L+1)M^2 \times (2L+1)N$. The power spectrum estimate is then given by the solution of the following positivity-constrained least-squares problem:

$$\min_{\mathbf{s}_x} \|\hat{\mathbf{q}}_y - \mathbf{\Phi}\mathbf{s}_x\|_2^2 \text{ s.t. } \mathbf{s}_x \geq \mathbf{0} \tag{3.52}$$

where $\mathbf{0}$ is the $(2L+1)N \times 1$ vector containing only zeros and $\geq$ is a component-wise inequality. Similarly, if we know that the power spectrum is sparse, we could think about adding a sparsity constraint to our reconstruction problem. The power spectrum estimate is then given by the solution of the following sparsity-constrained least-squares problem:

$$\min_{\mathbf{s}_x} \|\hat{\mathbf{q}}_y - \mathbf{\Phi}\mathbf{s}_x\|_2^2 + \lambda \|\mathbf{s}_x\|_1 \tag{3.53}$$

where the weight $\lambda \geq 0$ balances the sparsity-bias tradeoff. And naturally, it is also possible to combine the positivity constraint with the sparsity constraint.

While this positivity and/or sparsity constraint might lead to more accurate power spectrum estimates, they also allow to solve the under-determined case, i.e., when $M^2 < N$, and as such allow to further reduce the sampling rate requirements. However, we decided not to include these constraints from the beginning, because of two reasons. First of all, the constrained least-squares problems are harder to solve than the unconstrained one, and thus lead to a higher computational complexity. And second, the unconstrained solution allows for an analytical performance evaluation (as carried out in Section 3.6), while this is not trivial for the constrained solutions, since we lose the linear relationship between the estimated power spectrum $\hat{\mathbf{s}}_x$ and the cross-correlation vector $\hat{\mathbf{r}}_y$.

Finally, note that a positive solution can always be obtained from the unconstrained solution, by simply setting the negative entries to zero. In that case, the NMSE derived in Section 3.6.1 can be viewed as an upper bound on the true NMSE while the optimal detection threshold (as well as the related detection and false alarm rate) at a specific frequency derived in Section 3.6.2 will not change at all.

3.8 Simulation Results

In this section, we present some simulation results describing the effectiveness of our proposed methods. In the first part, we examine the estimation performance of both the time-domain and alternative time-domain reconstruction approaches while the detection performance of both approaches is presented in the second part. We consider minimal sparse ruler sampling and complex Gaussian sampling, where the latter is merely considered to show that the proposed techniques also work for other samplers than multi-coset samplers. In this section, minimal sparse ruler sampling refers to the multi-coset sampling technique for which we define the sampler coefficients $c_i[n]$ by selecting the rows of an $N \times N$ identity matrix according to Section 3.5, and for which we acquire larger compression rates by randomly adding extra rows of the identity matrix to the already selected rows.

3.8.1 Estimation Performance

First of all, we evaluate the performance of our time-domain approach presented in Section 3.3.1. We consider a complex baseband signal $x[n]$ spanning the frequency bands $[-0.9\pi, -0.65\pi]$, $[0.1\pi, 0.35\pi]$, and $[0.55\pi, 0.8\pi]$. To generate this signal, we pass circular complex zero-mean Gaussian i.i.d. noise with variance σ^2 through a digital filter $h[n]$ of length $LN + 1$. As a result, the support of the true autocorrelation sequence $r_x[n]$ is limited to $-LN \leq n \leq LN$, as required by our theory, and it is given by

$$r_x[n] = h[n] \star h^*[-n] \star \sigma^2 \delta[n] \tag{3.54}$$

In this subsection, we take $L = 2$ and $N = 84$, and we vary the compression rate M/N. The motivation to fix N at $N = 84$ is computational complexity since a higher N will generally result in a higher complexity.

We now examine the proposed minimal sparse ruler sampling discussed in Section 3.5 and complex Gaussian sampling. Both estimates are computed using LS. In the first method, the coefficients of $c_i[n]$ for $i = 0, 1, \ldots, M-1$ are generated according to the length-42 minimal sparse ruler having $M = 11$ distance marks. This is equivalent to selecting the corresponding $M = 11$ rows from the first 43 rows of the identity matrix $\mathbf{I}_{84}$ leading to $M = 11$ branches in our sampling device. As a result, we have matrices $\mathbf{R}_c[0]$ and $\mathbf{R}_c[1]$ of size 121×84 in (3.10). We then implement the larger M/N cases by randomly adding additional rows of $\mathbf{I}_{84}$ to the already selected 11 rows. In the complex Gaussian sampling case, we simply vary the compression rate M/N from $11/84$ to 0.5. The coefficients of $c_i[n]$

for $i = 0, 1, \ldots, M-1$ are randomly generated according to a circular complex Gaussian distribution with zero mean and variance $1/M$. Note that we keep these coefficients fixed over the different simulation runs. In Fig. 3.3, the NMSE between the estimated power spectrum $\hat{\mathbf{s}}_x$ and the true one $\mathbf{s}_x$ is calculated for both the minimal sparse ruler sampling and complex Gaussian sampling. While no noise is considered in Fig. 3.3(a), random circular complex zero-mean Gaussian i.i.d. noise is introduced in Fig. 3.3(b), in such a way that the signal-to-noise ratio (SNR) in the active bands is given by 10 dB. Both the simulated and analytical NMSE are calculated for a varying number of measurement vectors (MVs) K (which is analogous to KN Nyquist rate samples) as an attempt to represent different sensing times. We also provide the simulated and analytical NMSE between the estimated power spectrum produced by Nyquist rate sampling and the true one for different sensing times as a benchmark. Here, the Nyquist rate estimate is obtained from the proposed multi-coset approach in Section 3.5 by setting $M = N$. From the figures, it is obvious that the quality of the estimation improves with M/N and it slowly converges towards that of the Nyquist rate. We can also see how the NMSE improves as the sensing time increases, which is to be expected. Observe that the proposed minimal sparse ruler sampling generally performs better than complex Gaussian sampling. If we compare the structure of $\mathbf{R}_c$ in (3.14) for complex Gaussian sampling with the one for minimal sparse ruler sampling, we can easily see that, for the minimal sparse ruler case, the columns of $\mathbf{R}_c$ are not only independent but also orthogonal. As a result, the condition number of $\mathbf{R}_c$ for minimal sparse ruler sampling is smaller than the one for complex Gaussian sampling. This issue might explain why the performance of minimal sparse ruler sampling is better than that of complex Gaussian sampling. Note also how the simulated NMSE is on top of the analytical NMSE for the minimal sparse ruler, complex Gaussian, and Nyquist rate sampling.

Next, we consider the performance of the alternative time-domain approach in Section 3.3.2 for both minimal sparse ruler and complex Gaussian sampling. The signal model that is used here has the same characteristics as the one used in the time-domain approach except for the fact that the filter used to generate the signal now has length N. Hence, the support of the true auto-correlation sequence $r_x[n]$ is now limited to $-N+1 \leq n \leq N-1$, as required by our theory. We again select $N = 84$. For the minimal sparse ruler sampling case, we first generate the coefficients $c_i[n]$ for $i = 0, 1, \ldots, M-1$ according to the length-83 minimal sparse ruler leading to $M = 16$ branches in our sampling device. This will result in a matrix $\mathbf{R}_c^a$ of size 256×167 in (3.28). We again implement the larger

M/N cases by randomly adding additional rows of $\mathbf{I}_{84}$ to the already selected 16 rows. For complex Gaussian sampling, we simply vary the compression rate M/N from $16/84$ to 0.5. Note however that complex Gaussian sampling is theoretically able to offer a smaller compression rate than the minimal sparse ruler counterpart while maintaining the full column rank property of $\mathbf{R}_c^a$ in (3.28) although it might not result in an acceptable performance. Similar to the time-domain reconstruction approach, we randomly generate the coefficients $c_i[n]$ according to a circular complex Gaussian distribution with zero mean and variance $1/M$ and keep these coefficients fixed over the different simulation runs. Both the simulated and analytical NMSE between the estimated power spectrum and the true one for the alternative time-domain approach are depicted in Fig. 3.4. The Nyquist rate based estimates in Fig. 3.4 are obtained from the multi-coset implementation of the alternative time-domain approach in Section 3.3.2 with $M = N$. In general, we find similar trends as the ones observed for the time-domain reconstruction approach (Fig. 3.3) with respect to the impact of the compression rate and sensing time, as well as the relative performance between the minimal sparse ruler and complex Gaussian sampling.

3.8.2 Detection Performance

In this subsection, we consider a complex baseband signal spanning a frequency band between 0.15π and 0.25π. Again, this signal is generated by passing circular complex zero-mean Gaussian i.i.d. noise through a digital filter $h[n]$ of length $LN + 1$ where L and N are set to $L = 2$ and $N = 84$, respectively. On top of this spectrum, we add circular complex zero-mean Gaussian i.i.d. noise such that a specific SNR is obtained in the active band. In order to simplify the analysis in the simulation study, the same filter $h[n]$ is used for both the time-domain and alternative time-domain approaches. This certainly results in a bias for the estimate $\hat{\mathbf{s}}_x$ under the alternative time-domain approach. The detection probability should be evaluated in the occupied band. Note however that the active band is not perfectly rectangular and there are two transition bands around the edges of the occupied band. Therefore, we decide to leave small guard bands around the two edges of the active band and evaluate the detection performance at the points in the band from 0.154π to 0.243π. Meanwhile, the false alarm probability is based on a band that is significantly far from the occupied band, namely from -0.92π to -0.68π. The SNR in the active band is varied from -2 dB to -5 dB while the compression rate is varied between $M/N = 0.26$ and $M/N = 0.5$. The false alarm

probability is set by determining the detection threshold at each frequency, which is computed according to (3.46). For simulation purposes, we try to vary the false alarm probability between $p_{fa} = 0.005$ and $p_{fa} = 0.1$, which is smaller than the value suggested by the IEEE 802.22 standard [73].

We first consider the detection performance of the time-domain approach for both the minimal sparse ruler and complex Gaussian sampling implementation, as shown in Figs. 3.5 and 3.6, respectively. Here, we calculate the false alarm and detection occurences at each frequency point separately based on the estimated power at that point. For a given band, we then combine the amount of false alarm and detection occurences at all frequency points within that band to calculate false alarm and detection probability. The sampler coefficients for the minimal sparse ruler and complex Gaussian sampling are generated in the same way as in Section 3.8.1. Two different numbers of measurement vectors K are simulated, i.e., $K = 1786$ and $K = 5952$. As it is obvious from the figures, minimal sparse ruler sampling generally has a better performance than complex Gaussian sampling. This somehow confirms the relative estimation performance between both approaches in the previous subsection. Both methods, however, have a good performance for all simulated M/N when $K = 5952$ and SNR=-2dB. For SNR=-5dB, the performance of minimal sparse ruler sampling still reaches an acceptable level as long as $K = 5952$ and $M/N \geq 0.4$.

Next, we evaluate the performance of the alternative time-domain approach for the two sampling techniques, depicted in Figs. 3.7 and 3.8. Again, the sampler coefficients for the minimal sparse ruler and complex Gaussian sampling are generated in the same way as in Section 3.8.1. We observe that the alternative time-domain approach suffers from a performance degradation compared to its time-domain approach counterpart and a poorer performance is found for the complex Gaussian sampling case. For $K = 5952$ and SNR=-2dB, however, the performance of the minimal sparse ruler sampling is still acceptable.

3.9 Conclusions

In this paper, we have developed a new approach for power spectrum estimation of wide-sense stationary signals based on samples produced by a sub-Nyquist sampling device. No sparsity constraints are required for this method. In general, the solution can be derived by solving simple LS problems, which are solvable as long as the rank condition of the corresponding system equations are satisfied. We have focused on multi-coset sampling where we cast the design of the sampling device as

a minimal sparse ruler problem. We have shown that any sparse ruler can produce a multi-coset sampling design that ensures the full rank condition of the formulated sampling problem, and thereby guarantees the uniqueness of the power spectrum estimate as the solution to a set of simple LS problems. Moreover, when minimal sparse rulers are employed, the resulting samplers approach the minimum sampling rate resulting in a strong compression. Finally, we have derived the mean and the variance of the estimated power spectrum. Based on these results, we are able to derive the analytical MSE of our power spectrum estimates. Moreover, by assuming that the received signal only contains circular complex zero-mean Gaussian i.i.d. noise, we are able to derive the detection threshold which is advantageous when we intend to adopt the Neyman-Pearson theorem for evaluating the detection performance. The simulation study shows that the performance of our proposed approach is quite acceptable in terms of both estimation and detection, therefore making it a promising candidate for power spectrum estimation and sensing of wide band signals. Cognitive radio is for instance one possible interesting application.

Appendix

3.A Derivation of Cov$(\hat{r}_{y_i,y_j}[k], \hat{r}_{y_w,y_v}[q])$ for CFAR Detection Performance Evaluation (Time-Domain Approach)

We start by considering Cov$(\hat{r}_{y_i,y_j}[k], \hat{r}_{y_w,y_v}[q])$ in (3.38) and $E\left(\hat{r}_{y_i,y_j}[k]\hat{r}^*_{y_w,y_v}[q]\right)$ in (3.39), which are valid for a general received signal $x[n]$. When the received signal $x[n]$ has a Gaussian distribution, $y_i[k]$ also has a Gaussian distribution. We can then adopt the following result for Gaussian random variables, which is proven in [74, 71]: If x_1, x_2, x_3, and x_4 are jointly complex or real Gaussian random variables then $E(x_1x_2x_3x_4) = E(x_1x_2)E(x_3x_4) + E(x_1x_3)E(x_2x_4) + E(x_1x_4)E(x_2x_3) - 2E(x_1)E(x_2)E(x_3)E(x_4)$. This allows us to rewrite $E\left(\hat{r}_{y_i,y_j}[k]\hat{r}^*_{y_w,y_v}[q]\right)$ in (3.39) as:

$$
\begin{aligned}
E\left(\hat{r}_{y_i,y_j}[k]\hat{r}^*_{y_w,y_v}[q]\right) = \frac{1}{(K-|k|)(K-|q|)} \sum_{l=\max(0,k)}^{K-1+\min(0,k)} \sum_{p=\max(0,q)}^{K-1+\min(0,q)} \\
\Big\{E\left(y_i[l]y^*_j[l-k]\right)E\left(y^*_w[p]y_v[p-q]\right) + E\left(y_i[l]y^*_w[p]\right)E\left(y^*_j[l-k]y_v[p-q]\right) \\
+E\left(y_i[l]y_v[p-q]\right)E\left(y^*_j[l-k]y^*_w[p]\right) \\
-2E\left(y_i[l]\right)E\left(y^*_w[p]\right)E\left(y^*_j[l-k]\right)E\left(y_v[p-q]\right)\Big\}
\end{aligned}
$$

$$= \frac{1}{(K-|k|)(K-|q|)} \sum_{l=\max(0,k)}^{K-1+\min(0,k)} \sum_{p=\max(0,q)}^{K-1+\min(0,q)} \left\{ r_{y_i,y_j}[k] r^*_{y_w,y_v}[q] \right.$$
$$+ r_{y_i,y_w}[l-p] r^*_{y_j,y_v}[l+q-k-p] + E\left(y_i[l] y_v[p-q]\right) E\left(y_j^*[l-k] y_w^*[p]\right)$$
$$\left. - 2E\left(y_i[l]\right) E\left(y_w^*[p]\right) E\left(y_j^*[l-k]\right) E\left(y_v[p-q]\right) \right\}. \quad (3.55)$$

For CFAR detection performance evaluation, we assume that the received signal $x[n]$ only contains circular complex zero-mean Gaussian i.i.d. noise. When this is the case, the third and last terms in (3.55) are zero since $E\left(y_i[l]\right) = \sum_{n=1-N}^{0} c_i[n] E\left(x[lN-n]\right) = 0$, and

$$E\left(y_i[l] y_v[p-q]\right) = \sum_{n=1-N}^{0} \sum_{m=1-N}^{0} c_i[n] c_v[m] E\left(x[lN-n] x[pN-qN-m]\right) = 0.$$

As a result, (3.55) can be rewritten as:

$$E\left(\hat{r}_{y_i,y_j}[k] \hat{r}^*_{y_w,y_v}[q]\right) = r_{y_i,y_j}[k] r^*_{y_w,y_v}[q] + \frac{1}{(K-|k|)(K-|q|)} \sum_{l=\max(0,k)}^{K-1+\min(0,k)}$$
$$\sum_{p=\max(0,q)}^{K-1+\min(0,q)} \left\{ r_{y_i,y_w}[l-p] r^*_{y_j,y_v}[l+q-k-p] \right\}. \quad (3.56)$$

By inserting (3.56) into (3.38), $\mathrm{Cov}(\hat{r}_{y_i,y_j}[k], \hat{r}_{y_w,y_v}[q])$ can be expressed as:

$$\mathrm{Cov}(\hat{r}_{y_i,y_j}[k], \hat{r}_{y_w,y_v}[q]) = \frac{1}{(K-|k|)(K-|q|)} \sum_{l=\max(0,k)}^{K-1+\min(0,k)} \sum_{p=\max(0,q)}^{K-1+\min(0,q)}$$
$$\left\{ r_{y_i,y_w}[l-p] r^*_{y_j,y_v}[l+q-k-p] \right\}. \quad (3.57)$$

If we then consider (3.44), we can observe in (3.57) that only the terms with $l = p$ and $k = q$ have a non-zero value. Hence, we have:

$$\mathrm{Cov}(\hat{r}_{y_i,y_j}[k], \hat{r}_{y_w,y_v}[q]) = \frac{r_{y_i,y_w}[0] r^*_{y_j,y_v}[0] \delta[q-k]}{(K-|k|)} = \frac{\sigma^4 r_{c_i,c_w}[0] r^*_{c_j,c_v}[0] \delta[q-k]}{(K-|k|)} \quad (3.58)$$

which is the result provided in (3.45).

3.B Evaluation of the Statistical Distribution of $\hat{\mathrm{s}}_x$ for Circular Complex Zero-mean Gaussian i.i.d. Noise $x[n]$ (Time-Domain Approach)

In order to evaluate the statistical distribution of $\hat{\mathrm{s}}_x$ for circular complex zero-mean Gaussian i.i.d. noise $x[n]$, we first rewrite (3.37) as:

$$\begin{aligned}\hat{r}_{y_i,y_j}[k] &= \frac{1}{K-|k|}\sum_{l=\max(0,k)}^{K-1+\min(0,k)} y_i[l]y_j^*[l-k] \\ &= \frac{1}{K-|k|}\sum_{l=\max(0,k)}^{K-1+\min(0,k)}\left\{\sum_{n=1-N}^{0} c_i[n]x[lN-n]\right\}\times \\ &\left\{\sum_{m=1-N}^{0} c_j^*[m]x^*[lN-kN-m]\right\} \\ &= \frac{1}{K-|k|}\sum_{n=1-N}^{0}\sum_{m=1-N}^{0} c_i[n]c_j^*[m]\sum_{l=\max(0,k)}^{K-1+\min(0,k)} \\ &x[lN-n]x^*[lN-kN-m]. \end{aligned} \tag{3.59}$$

For now, we only concentrate on the most inner summation in (3.59). Without loss of generality, we only consider $k \geq 0$ in the following analysis to simplify the writing. Note that the results of the analysis also hold for $k < 0$. We have three possible cases for $z = \sum_{l=k}^{K-1} x[lN-n]x^*[lN-kN-m]$

1) $n \neq m$: In this case, z is the sum of the product of two i.i.d. Gaussian distributions. Note also that the terms in the summation have identical distributions and they are also independent to one another. Hence, we can exploit the central limit theorem to assume that in this case, z has a Gaussian distribution for sufficiently large $K-k$.

2) $k = 0$ and $n = m$: In this case, z will have a chi-square distribution with $K-k$ degrees of freedom. The chi-square distribution will converge to a Gaussian distribution for a sufficiently large value of $K-k$.

3) $k \neq 0$ and $n = m$: Similar to the first case, z is now again the sum of the product of two i.i.d. Gaussian distributions. Even though the terms in the summation are identically distributed, they are generally dependent. As an example, the

summation for $k = 2$ and $K = 20$ can be written as:

$$\sum_{l=2}^{19} x[lN-n]x^*[(l-2)N-n] = x[2N-n]x^*[-n] + x[3N-n]x^*[N-n]$$
$$+ x[4N-n]x^*[2N-n] + \cdots + x[19N-n]x^*[17N-n]. \tag{3.60}$$

Note that the same sample $x[2N-n]$ has a contribution in the first and third terms in (3.60). In general, for all possible values of k, every single term has a statistical dependency on at most two other terms in the summation. In order to simplify the analysis, let us split z into two separate summations, that is:

$$\sum_{l=k}^{K-1} x[lN-n]x^*[lN-kN-n] =$$
$$\left\{ \sum_{s=1}^{\lfloor \frac{K-k}{2k} \rfloor} \sum_{l=0}^{k-1} x[((2s-1)k+l)N-n]x^*[((2s-2)k+l)N-n] \right.$$
$$\left. + \sum_{l=\left(2\lfloor \frac{K-k}{2k} \rfloor+1\right)k}^{\min\left(\left(2\lfloor \frac{K-k}{2k} \rfloor+2\right)k-1, K-1\right)} x[lN-n]x^*[lN-kN-n] \right\}$$
$$+ \left\{ \sum_{s=1}^{\lfloor \frac{K-k}{2k} \rfloor} \sum_{l=0}^{k-1} x[(2sk+l)N-n]x^*[((2s-1)k+l)N-n] \right.$$
$$\left. + \sum_{l=\left(2\lfloor \frac{K-k}{2k} \rfloor+2\right)k}^{K-1} x[lN-n]x^*[lN-kN-n] \right\}. \tag{3.61}$$

Note that we basically put every k consecutive terms in z together into one group where every term is a product of two i.i.d. Gaussian random variables. The first pair of brackets in (3.61) contains the odd groups whereas the second pair of brackets contains the even groups. Observe that the summation in each pair of brackets is the sum of independent and identically distributed terms. For instance, if we consider (3.61) for $k = 2$ and $K = 20$, we obtain:

$$\sum_{l=2}^{19} x[lN-n]x^*[(l-2)N-n] = \{x[2N-n]x^*[-n] + x[3N-n]x^*[N-n]$$
$$+ \cdots + x[18N-n]x^*[16N-n] + x[19N-n]x^*[17N-n]\}$$

$$+\{x[4N-n]x^*[2N-n]+x[5N-n]x^*[3N-n]+\dots$$
$$+x[16N-n]x^*[14N-n]+x[17N-n]x^*[15N-n]\}. \quad (3.62)$$

Note that, within each pair of brackets in (3.62), there is no sample that contributes to more than one term in the summation. This characteristic can also be found if we apply different values of k and K to (3.61). As a result, we can also exploit the central limit theorem as long as $\lfloor \frac{K-k}{2k} \rfloor k$ is sufficiently large. When this is the case, we have the summation of two terms, each of which has a Gaussian distribution. As a result, z can again be approximated by a Gaussian distribution for large $\lfloor \frac{K-k}{2k} \rfloor k$.

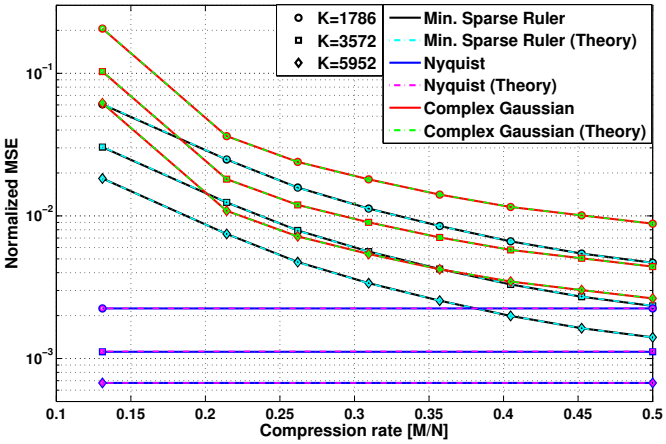
By covering all cases, we can conclude that the Gaussian approximation is applicable for the distribution of $\left\{\hat{r}_{y_i,y_j}[k]\right\}_{i,j=0}^{M-1}$ as long as $\lfloor \frac{K-k}{2k} \rfloor k$ is sufficiently large. From (3.11), we know that $\left\{\hat{r}_{y_i,y_j}[k]\right\}_{i,j=0}^{M-1}$ has a support limited to $-L \leq k \leq L$. As a result, the Gaussian assumption for $\left\{\hat{r}_{y_i,y_j}[k]\right\}_{i,j=0}^{M-1}$ is valid if $L \ll K$. By taking (3.11) and (3.40) into account, we can find that the distribution of each element of $\hat{\mathbf{s}}_x$ for the case of circular complex zero-mean Gaussian i.i.d. noise $x[n]$, is asymptotically Gaussian for a large K.

3.C Evaluation of the Statistical Distribution of $\hat{\mathbf{s}}_x$ for Circular Complex Zero-mean Gaussian i.i.d. Noise $x[n]$ (Alternative Time-domain Approach)

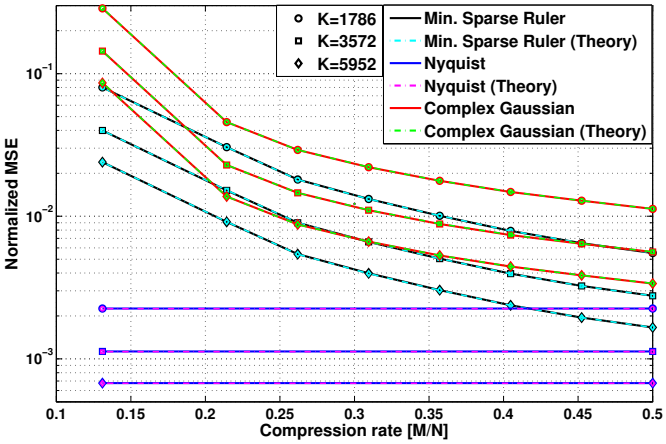
We investigate the statistical distribution of $\hat{\mathbf{s}}_x$ under the alternative time-domain approach by first considering the statistics of $\hat{\mathbf{r}}_y[0]$ in (3.47). Recall that the elements of $\hat{\mathbf{r}}_y[0]$ are $\hat{r}_{y_i,y_j}[0]$, which are given by (3.59) with $k=0$:

$$\hat{r}_{y_i,y_j}[0] = \frac{1}{K} \sum_{n=1-N}^{0} \sum_{m=1-N}^{0} c_i[n] c_j^*[m] \sum_{l=0}^{K-1} x[lN-n]x^*[lN-m]. \quad (3.63)$$

We only pay attention to the most inner summation in (3.63). When $x[n]$ only contains circular complex zero-mean Gaussian i.i.d. noise, we can observe that $z = \sum_{l=0}^{K-1} x[lN-n]x^*[lN-m]$ is the sum of i.i.d. random variables for both $m=n$ and $m \neq n$. As a result, for sufficiently large K, we are again able to exploit the central limit theorem to assume that z has a Gaussian distribution. By combining (3.47) - (3.50) and (3.63), we can conclude that for the case of circular complex zero-mean Gaussian i.i.d. noise $x[n]$, each element of $\hat{\mathbf{s}}_x$ is asymptotically Gaussian distributed for a large K under the alternative time-domain approach.

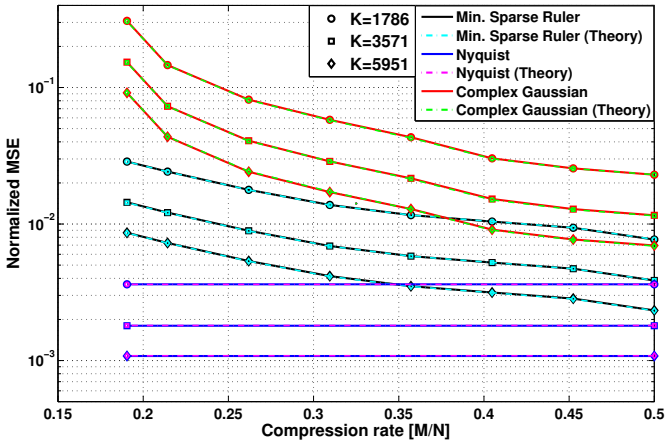


(a)

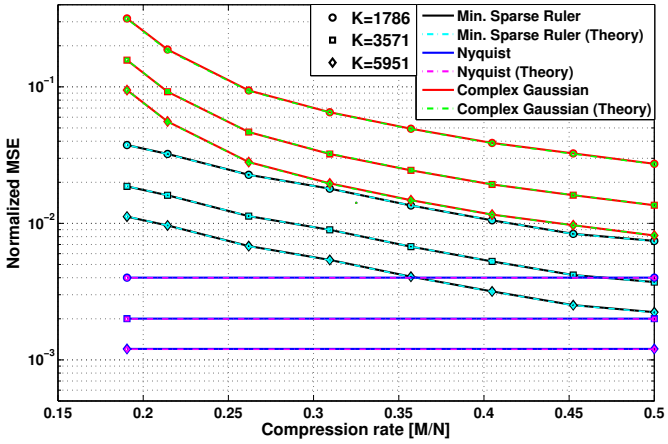


(b)

Figure 3.3: The normalized MSE between the estimated power spectrum (minimal sparse ruler and complex Gaussian sampling) and the true one for various numbers of MVs (K); (a) noise-free; (b) noisy (SNR=10 dB in active bands).

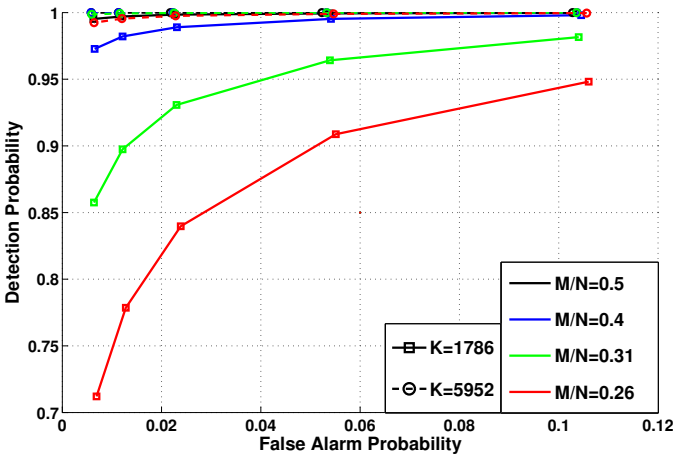


(a)

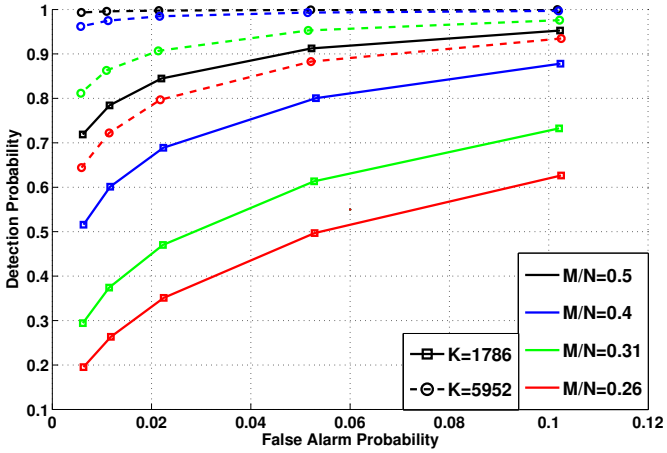


(b)

Figure 3.4: The normalized MSE between the estimated power spectrum (minimal sparse ruler and complex Gaussian sampling) based on the alternative time-domain approach and the true one for various numbers of MVs (K); (a) noise-free; (b) noisy (SNR=10 dB in active bands).

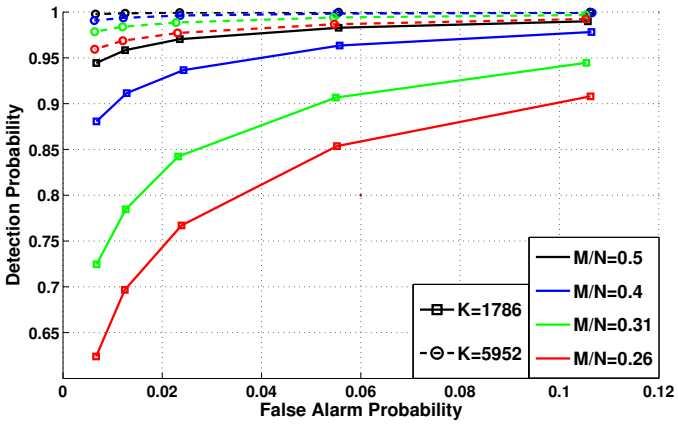


(a)

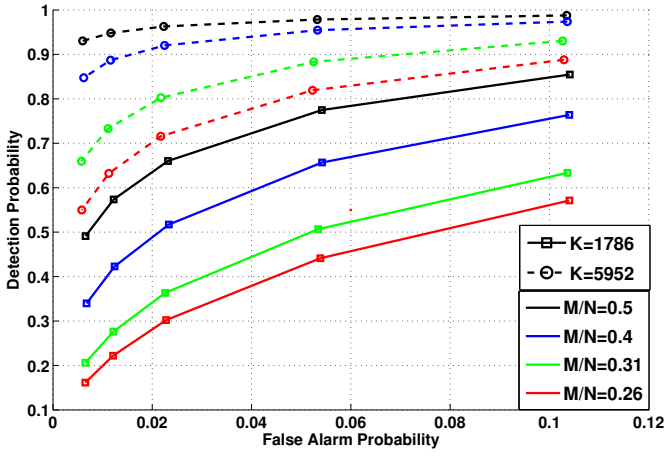


(b)

Figure 3.5: The detection performance of the proposed time-domain approach (minimal sparse ruler sampling) for various numbers of MVs (K) and M/N; (a) SNR = -2dB in active band; (b) SNR = -5dB in active band.

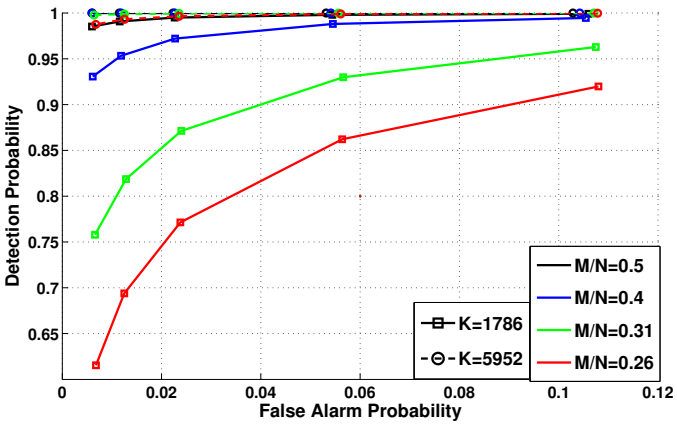


(a)

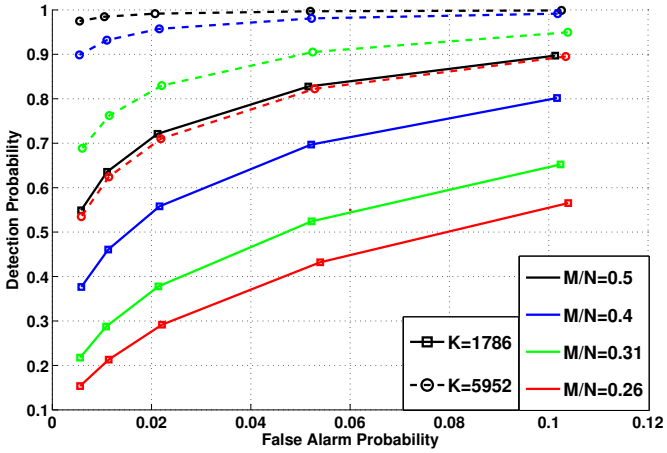


(b)

Figure 3.6: The detection performance of the proposed time-domain approach (complex Gaussian sampling) for various numbers of MVs (K) and M/N; (a) SNR = -2dB in active band; (b) SNR = -5dB in active band.

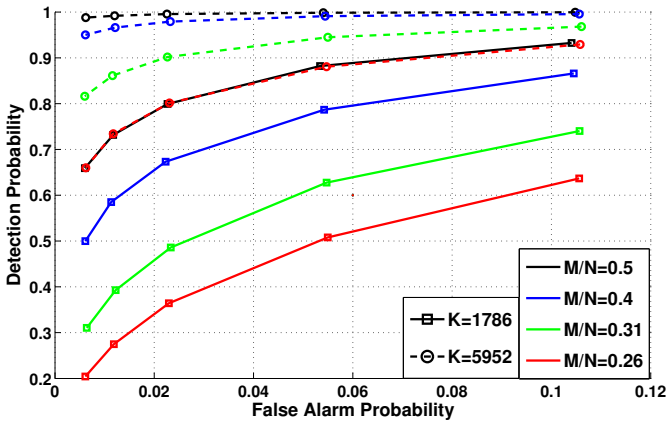


(a)

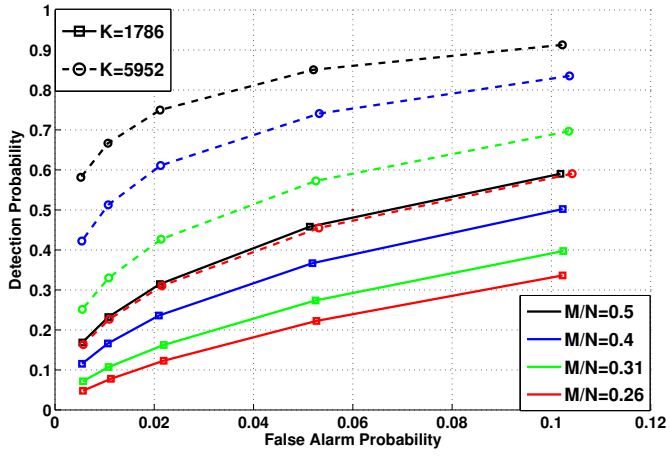


(b)

Figure 3.7: The detection performance of the proposed alternative time-domain approach (minimal sparse ruler sampling) for various numbers of MVs (K) and M/N; (a) SNR = -2dB in active band; (b) SNR = -5dB in active band.



(a)



(b)

Figure 3.8: The detection performance of the proposed alternative time-domain approach (complex Gaussian sampling) for various numbers of MVs (K) and M/N; (a) SNR = -2dB in active band; (b) SNR = -5dB in active band.

Chapter 4

Direction of Arrival Estimation for More Correlated Sources Than Active Sensors

Dyonisius Dony Ariananda and Geert Leus

Abstract

In this paper, a new direction of arrival (DOA) estimation method for more correlated sources than active receiving antennas is proposed. The trick to solve this problem using only second-order statistics is to consider a periodic scanning of an underlying uniform array, where a single scanning period contains several time slots and in different time slots different sets of antennas are activated leading to a dynamic non-uniform array with possibly less active antennas than sources in each time slot. We collect the spatial correlation matrices of the active antenna arrays for all time slots and are able to present them as a linear function of the spatial correlation matrix of the underlying array. We provide a necessary and sufficient condition for this system of equations to be full column-rank, which allows for a least squares (LS) reconstruction of the spatial correlation matrix of the underlying array. Some practical greedy algorithms are presented to design dynamic arrays satisfying this condition. In a second step, we use the resulting spatial correlation matrix of the underlying array to estimate the DOAs of the possibly correlated sources by spatial smoothing and MUSIC. Alternatively, we can express this matrix as a linear function of the correlation matrix of the sources (incoming signals) at a grid of investigated angles, and solve this system of equations using either LS or sparsity-regularized LS (possibly assisted by additional constraints), depending on the grid resolution compared to the number of antennas of the underlying array.

4.1 Context

In this section, we present the underlying model that will be used in the next sections and discuss related work on direction of arrival (DOA) estimation. Here, we restrict our attention to deterministic methods as well as stochastic methods exploiting up to second-order statistics. The presented approaches are further classified into methods that can handle correlated sources as well as methods that can handle more sources than sensors.

To set the stage, let us consider a uniform array of N antennas receiving K narrowband signals produced by possibly correlated sources. For simplicity, we assume in this paper that this uniform array of N antennas is one-dimensional, i.e., we adopt a uniform linear array (ULA), but our exposition can be extended to

higher dimensional arrays. The output of the ULA can be written as

$$\mathbf{x}(t) = \sum_{k=1}^{K} \mathbf{a}(\theta_k) s_k(t) + \mathbf{n}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) \tag{4.1}$$

where t is the time index, $\mathbf{x}(t)$ is the $N \times 1$ output vector containing the received signals at the N antennas, $\mathbf{n}(t)$ is the $N \times 1$ noise vector containing the noises at the N antennas, $\mathbf{s}(t) = [s_1(t), s_2(t), \ldots, s_K(t)]^T$ is the $K \times 1$ source vector with $s_k(t)$ the incoming signal from the k-th source with angle θ_k, and $\mathbf{A} = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \ldots, \mathbf{a}(\theta_K)]$ is the $N \times K$ array response matrix with $\mathbf{a}(\theta_k)$ the array response vector for the k-th source. If we consider the first element of the ULA as a reference point, we can express the array response vector as $\mathbf{a}(\theta_k) = [1, \phi(\theta_k)^d, \phi(\theta_k)^{2d}, \ldots, \phi(\theta_k)^{(N-1)d}]^T$, where d is the distance in wavelengths between two antennas and $\phi(\theta_k) = \exp(j2\pi \sin\theta_k)$. We always consider that $-\pi/2 \leq \theta_k < \pi/2$ and that $\{\theta_k\}_{k=1}^K$ contains different values. We generally assume that the impact of the wireless channel has been taken into account in $\mathbf{s}(t)$, that $\mathbf{n}(t)$ and $\mathbf{s}(t)$ are uncorrelated, and that the noises at the different antennas are mutually uncorrelated with variance σ_n^2, i.e., $E[\mathbf{n}(t)\mathbf{n}^H(t)] = \sigma_n^2 \mathbf{I}_N$, with $\mathbf{I}_N$ the $N \times N$ identity matrix. However, the incoming signals can possibly be correlated with correlation matrix $E[\mathbf{s}(t)\mathbf{s}^H(t)] = \mathbf{R}_s$. As a result, the spatial correlation matrix $E[\mathbf{x}(t)\mathbf{x}^H(t)] = \mathbf{R}_x$ can be written as

$$\mathbf{R}_x = \mathbf{A}\mathbf{R}_s\mathbf{A}^H + \sigma_n^2 \mathbf{I}_N. \tag{4.2}$$

To model the more general case of a non-uniform linear array (NULA), we can select M ($\leq N$) antennas from the above ULA of N antennas, which is referred to as the *underlying array* in the following. Defining $\mathbf{y}(t)$ as the $M \times 1$ output vector representing the received signals at the M selected active antennas, we obtain

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) = \sum_{k=1}^{K} \mathbf{b}(\theta_k) s_k(t) + \mathbf{m}(t) = \mathbf{B}\mathbf{s}(t) + \mathbf{m}(t), \tag{4.3}$$

where $\mathbf{C}$ is an $M \times N$ selection matrix containing M rows from $\mathbf{I}_N$, and where we further introduced $\mathbf{b}(\theta_k) = \mathbf{C}\mathbf{a}(\theta_k)$ and $\mathbf{B} = \mathbf{C}\mathbf{A}$ as the downsampled array response vector and matrix, respectively, which are both related to the set of M active antennas. Note that $\mathbf{m}(t)$ is the $M \times 1$ noise vector obtained as $\mathbf{m}(t) = \mathbf{C}\mathbf{n}(t)$, which has correlation matrix $E[\mathbf{m}(t)\mathbf{m}^H(t)] = \sigma_n^2 \mathbf{I}_M$. The spatial correlation matrix of the active antennas can then be written as

$$\mathbf{R}_y = E[\mathbf{y}(t)\mathbf{y}^H(t)] = \mathbf{C}\mathbf{R}_x\mathbf{C}^H = \mathbf{B}\mathbf{R}_s\mathbf{B}^H + \sigma_n^2 \mathbf{I}_M. \tag{4.4}$$

To retain the same aperture as the underlying ULA, we assume that $\mathbf{C}$ always selects the first and last antenna of the underlying ULA. We also assume that both the NULA and the underlying ULA introduce no spatial aliasing, which can be guaranteed by taking $d \leq 1/2$ and by designing $\mathbf{C}$ such that the indices of the selected antennas are coprime [75], which is true for most existing NULA designs.

Based on the above model for a NULA, we will now discuss a number of state-of-the-art DOA estimation methods that can either handle correlated sources or more sources than sensors. Note that modeling an NULA by selecting a subset of antennas from a ULA turns out to be useful to explain some of the following DOA estimation methods. Moreover, the dynamic array concept we will propose in this paper will also build upon such a model, as will be explained in Section 4.2.

4.1.1 Handling Correlated Sources

Depending on the characteristics of the sources $\mathbf{s}(t)$ and the number of sources, K, relative to the total number of active antennas in the array, M, it is possible to perform DOA estimation based on $\mathbf{R}_y$ in (4.4) using existing approaches. It is clear that the signal correlation matrix $\mathbf{R}_s$ in (4.4) is diagonal when the incoming signals are uncorrelated, is nondiagonal and full rank when the signals are partially correlated, and is nondiagonal and rank deficient when the signals are fully correlated (coherent) [48]. When $\mathbf{R}_s$ has full rank and $M > K$, MUSIC in [46] (or root-MUSIC in [76]) can be applied. For uncorrelated or mildly correlated incoming signals, MUSIC performs very well but for highly or fully correlated signals, $\mathbf{R}_s$ in (4.4) is close to or exactly singular and the MUSIC performance deteriorates. As discussed in [77], for an array that contains a sufficient number of translational equivalent subarrays (this is for instance the case for a ULA, i.e., when we do not perform antenna selection and $\mathbf{C} = \mathbf{I}_N$), the highly or fully correlated sources in $\mathbf{s}(t)$ can be handled by applying the spatial smoothing preprocessing scheme in [48, 78] to $\mathbf{R}_y$ in (4.4). This scheme leads to a spatially smoothed correlation matrix $\overline{\mathbf{R}}_y$ that can be expressed in terms of a full rank matrix $\overline{\mathbf{R}}_s$, which is a modified version of $\mathbf{R}_s$ in (4.4). Hence, MUSIC (or root-MUSIC) can now be applied to $\overline{\mathbf{R}}_y$. Interestingly, instead of using spatial smoothing and MUSIC to estimate the DOA of coherent incoming signals, [77] introduces a new method that relies on the evaluation of the distance between the investigated steering vectors and the subspace spanned by the columns of the product of the noise correlation matrix and the signal eigenvectors. This distance evaluation is based on the predefined distance metric called normalized distance functional.

For another class of approaches, a fine grid of investigated angles is defined in the angular domain and then the output of the array is expressed in terms of a linear combination of the steering vectors of these investigated angles. For the model in (4.3), this is equivalent to writing $\mathbf{y}(t)$ as

$$\mathbf{y}(t) = \sum_{q=1}^{Q} \mathbf{b}(\tilde{\theta}_q) s_{\tilde{\theta}_q}(t) + \mathbf{m}(t) = \tilde{\mathbf{B}}\tilde{\mathbf{s}}(t) + \mathbf{m}(t), \tag{4.5}$$

where $\tilde{\mathbf{s}}(t) = [s_{\tilde{\theta}_1}(t), s_{\tilde{\theta}_2}(t), \ldots, s_{\tilde{\theta}_Q}(t)]^T$ is the $Q \times 1$ extended source vector with $s_{\tilde{\theta}_q}(t)$ the incoming signal from the q-th investigated angle $\tilde{\theta}_q$, and $\tilde{\mathbf{B}} = [\mathbf{b}(\tilde{\theta}_1), \mathbf{b}(\tilde{\theta}_2), \ldots, \mathbf{b}(\tilde{\theta}_Q)]$ is the $M \times Q$ extended subsampled array response matrix with $\mathbf{b}(\tilde{\theta}_q)$ the subsampled array response vector for the q-th investigated angle $\tilde{\theta}_q$. As before, we always consider that $-\pi/2 \leq \tilde{\theta}_q < \pi/2$ and that $\{\tilde{\theta}_q\}_{q=1}^{Q}$ contains different values. It is important to note that $\{\tilde{\theta}_q\}_{q=1}^{Q}$ is known and might approximately contain the set of actual angles of arrival $\{\theta_k\}_{k=1}^{K}$ contained in $\mathbf{B}$ in (4.3), which is not known by the receiver. Based on (4.5) and defining the extended source correlation matrix as $E[\tilde{\mathbf{s}}(t)\tilde{\mathbf{s}}(t)^H] = \mathbf{R}_{\tilde{s}}$, we can also express $\mathbf{R}_y$ in (4.4) as

$$\mathbf{R}_y = \tilde{\mathbf{B}}\mathbf{R}_{\tilde{s}}\tilde{\mathbf{B}}^H + \sigma_n^2\mathbf{I}_M. \tag{4.6}$$

The DOA estimation approach proposed in [31] exploits the model described by (4.5) with $Q \gg M$. In this case, $\tilde{\mathbf{B}}$ has more columns than rows and as a result, the columns of $\tilde{\mathbf{B}}$ play the role of an overcomplete basis for $\mathbf{y}(t)$. To overcome this problem, [31] assumes that the coefficient vector with respect to this overcomplete basis is generally sparse. Further, [31] exploits multiple measurement vectors (MMV) by collecting data from multiple time indices, based on the assumption that the DOAs do not change within the duration of the sample acquisition. Hence, their data model is given by $\mathbf{Y} = \tilde{\mathbf{B}}\tilde{\mathbf{S}} + \mathbf{M}$, where $\mathbf{Y}$, $\mathbf{M}$, and $\tilde{\mathbf{S}}$ respectively stack $\mathbf{y}(t)$, $\mathbf{m}(t)$ and $\tilde{\mathbf{s}}(t)$ over different time indices in a row-wise fashion. Next, the so-called ℓ_1 singular value decomposition (ℓ_1-SVD) algorithm consisting of a dimensionality reduction of the MMV model as well as a mixed $\ell_{2,1}$-norm minimization is used to exploit the group sparsity of the columns of $\tilde{\mathbf{S}}$ in the MMV model. A closely related method can be found in [35], with the difference that a mixed $\ell_{2,0}$-norm approximation is used instead of a mixed $\ell_{2,1}$-norm to exploit the group sparsity of the columns of $\tilde{\mathbf{S}}$, leading to the so-called joint ℓ_0 approximation (JLZA) algorithm.

Other grid-based methods directly exploit (4.6) again with $Q \gg M$, such as the work of [79]. There, the columns of $\tilde{\mathbf{B}}$ play the role of overcomplete basis for each column of $\mathbf{R}_y$ and as before sparsity is assumed in the coefficient vector

corresponding to this basis. Since all columns of $\mathbf{R}_y$ have the same sparse structure with respect to this overcomplete basis, group sparsity is again exploited to estimate the DOA of the incoming signals leading to the so-called ℓ_1 sparse representation of array covariance vectors (ℓ_1-SRACV) algorithm. As the ℓ_1-SVD and JLZA algorithms, the ℓ_1-SRACV algorithm is robust to the correlation of the incoming signals. The model in (4.6) is also exploited in [80, 81], where a spatial correlation matching approach is considered. The resulting so-called sparse iterative covariance-based estimation (SPICE) method is derived assuming uncorrelated sources and sparsity of the sources in the angular domain, i.e., the extended source correlation matrix $\mathbf{R}_{\tilde{s}}$ in (4.6) is diagonal with only a few non-zero diagonal elements. Although SPICE has been derived based on uncorrelated sources, it has been shown to be robust against correlation.

4.1.2 Handling More Sources than Sensors

Most of the aforementioned methods can handle correlated sources, but they generally require more active antennas than sources ($M \geq K$). This is understandable for those methods involving the use of MUSIC, and it has also been shown to hold for the grid-based methods exploiting sparsity, which actually require $M \geq 2K$. For uncorrelated sources, on the other hand, some approaches have been proposed for DOA estimation when there are more sources than physical receiving antennas. One example in [82, 83] exploits the Caratheodory theorem and constructs the so-called augmented correlation matrix from the spatial correlation matrix. However, they rely on the exact knowledge of the spatial correlation matrix, which is unavailable in practice and has to be estimated from sample averaging. When this is the case, the augmented correlation matrix might not be positive semi-definite, i.e., it might not be a valid correlation matrix, thereby leading to a performance degradation. In [84], a complex algorithm has been introduced to convert the augmented correlation matrix into a valid positive semi-definite correlation matrix.

More popular techniques for uncorrelated sources exploit the fact that the source correlation matrix $\mathbf{R}_s$ is diagonal and rewrite (4.4) as

$$\text{vec}\left(\mathbf{R}_y\right) = \left(\mathbf{B}^* \odot \mathbf{B}\right)\text{diag}\left(\mathbf{R}_s\right) + \sigma_n^2 \text{vec}\left(\mathbf{I}_M\right) \tag{4.7}$$

where $\odot$ denotes the Khatri-Rao product operation and vec$(.)$ is the operator that cascades all columns of a matrix in a large column vector. Observe that the N_v distinct rows of $\mathbf{B}^* \odot \mathbf{B}$ provide the array response matrix of a virtual array (also known as co-array) of N_v virtual antennas receiving K virtual sources at the angles $\{\theta_k\}_{k=1}^K$ (note that generally $N_v > M$ and the upper bound is given by

$N_v \leq M(M-1)+1$). The problem now is that only a single measurement vector is available and thus the behavior of the model (4.7) is similar to an array receiving constant and hence fully coherent source signals, which is problematic for DOA estimation. A first method to solve this issue relies on the assumption that diag $(\mathbf{R}_s)$, and thus vec $(\mathbf{R}_y)$, is time-varying, which basically means that quasi-stationary sources are assumed and thus enough linearly independent measurement vectors can be obtained [47]. But it is clear that this method will fail in case of stationary sources, which was the starting point of this paper. A second technique again relies on gridding and exploits the model (4.6), which for uncorrelated sources can be written as

$$\text{vec}\,(\mathbf{R}_y) = (\tilde{\mathbf{B}}^* \odot \tilde{\mathbf{B}})\text{diag}\,(\mathbf{R}_{\tilde{s}}) + \sigma_n^2 \text{vec}\,(\mathbf{I}_M)\,. \tag{4.8}$$

If we select the angular resolution such that $Q = N_v$, we can solve (4.8) using ordinary least squares (LS) [10]. This is particularly interesting if we want to obtain an analytical performance analysis of the solution. However, when N_v is too small, this leads to a bad angular resolution and it will be difficult to estimate the DOA of off-grid sources. The other option is to take a fine grid of investigated angles, for which $Q > N_v$. In that case, we have to rely on the sparsity (possibly assisted by the positivity) of diag $(\mathbf{R}_{\tilde{s}})$ to solve the underdetermined problem (4.8), as was recently advocated in [85]. A final technique is based on constructing special array geometries of M antennas, through the design of the selection matrix $\mathbf{C}$ in (4.3), such that the overall virtual array of N_v antennas subsumes a ULA of N_u antennas, referred to as the virtual ULA, where generally $M < N_u \leq N_v$. Examples of such array designs are the two-level nested array [42], the coprime array [41], and the minimal sparse ruler array [10]. Under this virtual ULA, the DOAs can again be estimated using gridding and adopting the earlier mentioned LS [10] or sparsity-constrained LS (possibly assisted by a positivity constraint) [85]. Different from a general virtual array, $\tilde{\mathbf{B}}^* \odot \tilde{\mathbf{B}}$ now has a Vandermonde structure, which simplifies the implementation and analysis of these methods. Alternatively, it is now also possible to apply spatial smoothing and MUSIC based on the single available measurement vector from the virtual ULA.

4.1.3 Handling More Correlated Sources than Sensors

All the above methods either focus on correlated sources, but then the upper bound of the number of sources is lower than the number of antennas, or they focus on detecting more uncorrelated sources than sensors. To the best of our knowledge, there are no deterministic or stochastic methods using up to second-order statistics that

can handle more correlated sources than sensors. Only higher-order statistics have been exploited up to now to solve this problem (see e.g. [86]). In this paper, we will fill this gap, and introduce a new approach to tackle more correlated sources than *active* sensors. The paradigm shift introduced to reach this goal is to periodically change the selection of the M active antennas from the underlying ULA of N antennas. This will be explained in detail in the following sections.

4.2 Dynamic Array through Periodic Scanning

In this section, we introduce a novel dynamic array for DOA estimation of possibly correlated or even fully coherent signals when the number of sources is more than the number of active antennas at any given time. The underlying ULA consisting of N antennas discussed in Section 4.1 is used as an array of available antennas from which we activate only M antennas within a specific time slot, where the set of M activated antennas can differ from time slot to time slot. In this way, even though the number of required physical antennas is equal to N, the number of active antennas per time slot, and thus the number of hardware receiver branches, is reduced from N to M. This scheme reduces the power consumption without compromising the ability to locate the DOAs of the sources. In other words, the number of sources we will be able to detect is the same as if we had all N antennas from the underlying ULA available all the time. Note that instead of antenna switching, we can equivalently employ $M < N$ adjustable antennas to construct a dynamic array allowing us to alter the position of each antenna in every time slot.

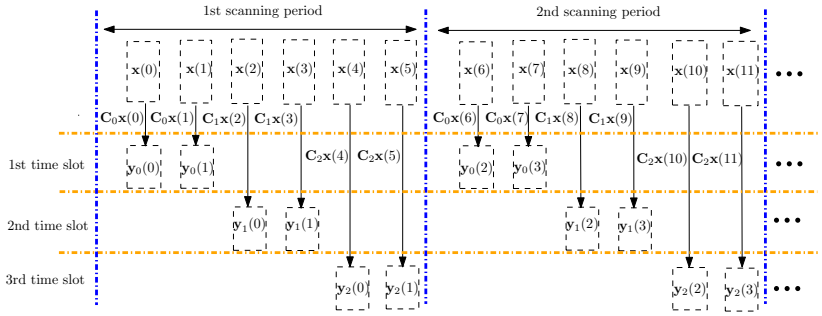


Figure 4.1: Description of the periodic scanning process where a single scanning period consists of L time slots. Here the number of time slots per scanning period is $L = 3$ and the number of samples per slot per antenna is given by $S = 2$.

Let us now discuss the conceived dynamic array in some more detail. We ba-

sically focus on some kind of periodic scanning with P scanning periods, where a single scanning period consists of L time slots, and where a single time slot consists of S samples (see Fig. 4.1). The set of M activated antennas within a single scanning period is different from time slot to time slot whereas the set of active antennas in the l-th time slot of different scanning periods is the same. Defining $\mathbf{x}(t)$ as the output vector of the underlying ULA of N antennas, as in Section 4.1, and assuming for simplicity that the sample period is 1, we introduce $\mathbf{y}_l(\tau)$ as the $M \times 1$ vector representing the output of the M active antennas of the linear array in the l-th time slot ($l = 0, 1, \ldots, L-1$), which is given by

$$\mathbf{y}_l(\tau) = \mathbf{C}_l \mathbf{x}(pLS + lS + s)$$

where $\lfloor \tau/S \rfloor = p$ ($p = 0, 1, \ldots, P-1$) indicates the scanning period index, $s = \tau - pS$ ($s = 0, 1, \ldots, S-1$) indicates the sample index within the l-th time slot of the p-th scanning period, and where the $M \times N$ matrix $\mathbf{C}_l$ is constructed by selecting M out of N rows from the identity matrix $\mathbf{I}_N$. Note that the indices of the M selected rows represent the indices of the M active antennas in the l-th time slot selected from the N available antennas in the underlying ULA. We are then able to compute the $M \times M$ spatial correlation matrix of $\mathbf{y}_l(\tau)$ as

$$\mathbf{R}_{y_l} = E\left[\mathbf{y}_l(\tau)\mathbf{y}_l(\tau)^H\right] = \mathbf{C}_l \mathbf{R}_x \mathbf{C}_l^T \tag{4.9}$$

where the second equality is due to the fact that $\mathbf{C}_l$ is a real matrix. Note that the expectation operation in (4.9) can be estimated by taking an average over PS time samples. Next, let us stack all columns of $\mathbf{R}_{y_l}$ into the $M^2 \times 1$ vector $\text{vec}(\mathbf{R}_{y_l})$. Based on (4.9), we can then express $\text{vec}(\mathbf{R}_{y_l})$ as

$$\mathbf{r}_{y_l} = \text{vec}(\mathbf{R}_{y_l}) = (\mathbf{C}_l \otimes \mathbf{C}_l)\text{vec}(\mathbf{R}_x) \tag{4.10}$$

where $\otimes$ denotes the Kronecker product operation. Finally, we can combine $\mathbf{r}_{y_l}$ in (4.10) for all time slots $l = 0, 1, \ldots, L-1$ into a single vector $\mathbf{r}_y$, which is given by $\mathbf{r}_y = [\mathbf{r}_{y_0}^T, \mathbf{r}_{y_1}^T, \ldots, \mathbf{r}_{y_{L-1}}^T]^T$. The relationship between $\mathbf{r}_y$ and $\mathbf{R}_x$ is then provided by

$$\mathbf{r}_y = \mathbf{\Psi}\text{vec}(\mathbf{R}_x) \tag{4.11}$$

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

$$\mathbf{\Psi} = [(\mathbf{C}_0 \otimes \mathbf{C}_0)^T, (\mathbf{C}_1 \otimes \mathbf{C}_1)^T, \ldots, (\mathbf{C}_{L-1} \otimes \mathbf{C}_{L-1})^T]^T. \tag{4.12}$$

This equation forms the basis of this paper and it allows for the reconstruction of $\text{vec}(\mathbf{R}_x)$ from $\mathbf{r}_y$, which will be discussed in the next section.

4.3 Reconstruction of Spatial Correlation Matrix $\mathbf{R}_x$

One option to solve (4.11) is using ordinary LS. If $\mathbf{\Psi}$ has full column rank, we then obtain

$$\text{vec}(\hat{\mathbf{R}}_x) = (\mathbf{\Psi}^T\mathbf{\Psi})^{-1}\mathbf{\Psi}^T\mathbf{r}_y. \tag{4.13}$$

Alternatively, we could add a positive semi-definite (p.s.d.) constraint on $\mathbf{R}_x$ in the aforementioned ordinary LS problem, leading to the following constrained LS reconstruction problem

$$\hat{\mathbf{R}}_x = \arg\min_{\mathbf{R}_x} \|\mathbf{r}_y - \mathbf{\Psi}\text{vec}(\mathbf{R}_x)\|_2^2 \text{ s.t. } \mathbf{R}_x \succeq \mathbf{0}_{N\times N}, \tag{4.14}$$

where $\mathbf{0}_{m\times n}$ denotes an $m \times n$ matrix containing only zeros. This could possibly alleviate the requirement of a full column rank $\mathbf{\Psi}$, but comes at the expense of a large computational complexity. Either way, it is still of great interest to design $\{\mathbf{C}_l\}_{l=0}^{L-1}$ such that $\mathbf{\Psi}$ has full column rank, which clearly requires $M^2L \geq N^2$. Observe that, since $M < N$, we can only have $M^2L \geq N^2$ if at least two time slots per scanning period are adopted, i.e., $L \geq 2$. In other words, a dynamic array through periodic scanning is indispensable for generating a full column rank $\mathbf{\Psi}$.

In the remainder of this section, we will first propose a necessary and sufficient condition for the periodic subsampling procedure to have a full column rank $\mathbf{\Psi}$ in (4.12). Next, we will develop some practical greedy approaches to design a periodic scanning scheme that satisfies this condition. Finally, we will discuss some trade-offs related to the design of the dynamic array.

4.3.1 Establishing Perfect Reconstruction

In order to simplify our analysis, let us start by introducing the following lemma.

Lemma 4.3.1. *$\mathbf{C}_l \otimes \mathbf{C}_l$ will have a one in the $[(i-1)N+j]$-th and $[(j-1)N+i]$-th columns if and only if $\mathbf{C}_l$ contains the i-th and j-th rows of the identity matrix $\mathbf{I}_N$.*

The proof of this lemma can be found in Appendix 4.A. Lemma 4.3.1 directly implies the following corollary.

Corollary 4.3.1.1. *If $\mathbf{C}_l$ is constructed by selecting M different rows of $\mathbf{I}_N$, the rows of $\mathbf{C}_l \otimes \mathbf{C}_l$ have a single one at exactly M^2 different positions. Out of the M^2 rows of $\mathbf{C}_l \otimes \mathbf{C}_l$, M rows are produced by the self-Kronecker product of every row of $\mathbf{C}_l$. On the other hand, every pair of two different rows of $\mathbf{C}_l$ contributes to*

two different rows of $\mathbf{C}_l \otimes \mathbf{C}_l$, each of which has a single one at a different position. Since we have $\binom{M}{2}$ possible combinations of two different rows, all Kronecker products between any two different rows of $\mathbf{C}_l$ lead to $M(M-1)$ rows of $\mathbf{C}_l \otimes \mathbf{C}_l$, all of which have a single one at a different position.

Let us now define Γ_l as the set of M indices selected from $\{1, 2, \ldots, N\}$ representing the rows of $\mathbf{I}_N$ that we use to construct $\mathbf{C}_l$. Then, the set of the M^2 indices of the columns of $\mathbf{C}_l \otimes \mathbf{C}_l$ that contain a one is provided by

$$\Omega_l = \{(i-1)N + j | i, j \in \Gamma_l\}. \tag{4.15}$$

Since every row of $\mathbf{C}_l \otimes \mathbf{C}_l$ has only a single one, it is clear from (4.12) that every row of $\mathbf{\Psi}$ also has only a single one. As a result, $\mathbf{\Psi}$ will have full column rank if and only if each of its columns has at least a single one, which from (4.12) and (4.15) is equivalent to

$$\bigcup_{l=0}^{L-1} \Omega_l = \left\{1, 2, \ldots, N^2\right\}. \tag{4.16}$$

This result leads to the following theorem.

Theorem 4.3.1. *$\mathbf{\Psi}$ in* (4.12) *has full column rank if and only if every possible combination of two antennas in the underlying ULA is active in at least one of the L possible time slots within a single scanning period.*

Proof. Note that the condition in Theorem 4.3.1 is equivalent to using every possible pair of two different rows of $\mathbf{I}_N$ in at least one of the L possible matrices $\{\mathbf{C}_l\}_{l=0}^{L-1}$. Based on Lemma 4.3.1 and Corollary 4.3.1.1, this will guarantee that the $[(i-1)N + j]$-th and the $[(j-1)N + i]$-th columns of $\mathbf{\Psi}$ have at least a single one for all $i, j \in \{1, \ldots, N\}$. This proves the sufficiency part of the theorem. In order to prove the necessity part, let us assume that $\mathbf{\Psi}$ in (4.12) has full column rank and that the a-th and b-th antennas in the underlying ULA are never simultaneously active in the same time slot. This equivalently means that none of the matrices $\{\mathbf{C}_l\}_{l=0}^{L-1}$ contains both the a-th and b-th rows of the identity matrix $\mathbf{I}_N$. According to Lemma 4.3.1 and using (4.12), it is then obvious that the $[(a-1)N + b]$-th and the $[(b-1)N + a]$-th columns of $\mathbf{\Psi}$ only contain zeros and thus $\mathbf{\Psi}$ does not have full column rank, which contradicts the initial assumption. This concludes the proof. □

Note that Theorem 4.3.1 automatically requires $M \geq 2$. Further note that since each row of $\mathbf{\Psi}$ only contains a single one and zeros elsewhere, the reconstruction of $\text{vec}(\mathbf{R}_x)$ using ordinary LS as in (4.13) is computationally easy to perform.

4.3.2 Greedy Dynamic Array Design

As we will discuss in Section 4.3.3, subject to the full column rank condition of $\mathbf{\Psi}$, it is generally not possible to minimize both the number of active antennas per time slot M and the number of time slots per scanning period L simultaneously. In this section, we discuss a greedy dynamic array design that aims to either minimize M given L or minimize L given M subject to (4.16).

Based on Theorem 4.3.1, we first try to minimize L given M subject to (4.16) by defining Λ as $\Lambda = \{(i,j)|i,j \in \{1,2,\ldots,N\}, i<j\}$ and Λ_l as the set of all possible combinations of two row indices of $\mathbf{I}_N$ that are used to construct $\mathbf{C}_l$, that is

$$\Lambda_l = \{(i,j)|i,j \in \Gamma_l, i<j\}.$$

Our task to minimize L given M subject to (4.16) can now be expressed as

$$\min_{L,\{\Gamma_l\}_{l=0}^{L-1}} L \text{ subject to } \bigcup_{l=0}^{L-1} \Lambda_l = \Lambda \text{ and } |\Gamma_l| = M, \forall l \tag{4.17}$$

where $|\Gamma_l|$ denotes the cardinality of the set Γ_l. The minimization problem in (4.17) is generally a non-trivial combinatorial problem. However, it is possible to find a lower bound for L. Note that $|\Lambda| = N(N-1)/2$ and $|\Lambda_l| = M(M-1)/2$. It is then clear that L is lower bounded by

$$L \geq \left\lceil \frac{|\Lambda|}{|\Lambda_l|} \right\rceil = \left\lceil \frac{N(N-1)}{M(M-1)} \right\rceil \tag{4.18}$$

where $\lceil x \rceil$ denotes the smallest integer not smaller than x. While the minimization problem in (4.17) is generally hard to solve, we propose a greedy algorithm to find a sub-optimal solution for L and $\{\Gamma_l\}_{l=0}^{L-1}$ given M subject to (4.16). This algorithm, called Algorithm 4.1, is described in Table 4.1 and its explanation is provided in Appendix 4.B.

Similarly, given a certain value of L, the minimization of M subject to (4.16) can be stated as

$$\min_{M,\{\Gamma_l\}_{l=0}^{L-1}} M \text{ subject to } \bigcup_{l=0}^{L-1} \Lambda_l = \Lambda \text{ and } |\Gamma_l| = M, \forall l.$$

In this case, the lower bound for M is given by

$$\frac{M^2 - M}{2} \geq \left\lceil \frac{|\Lambda|}{L} \right\rceil = \left\lceil \frac{N(N-1)}{2L} \right\rceil. \tag{4.19}$$

We also propose a greedy algorithm to find a sub-optimal solution for M and $\{\Gamma_l\}_{l=0}^{L-1}$ given L subject to (4.16). This algorithm, called Algorithm 4.2, is described in Table 4.2 and its explanation is provided in Appendix 4.C.

4.3.3 Trade-offs for Dynamic Array Design

While we want to achieve (4.16) to ensure the full column rank condition of $\mathbf{\Psi}$, we also want to keep the computational complexity low, the number of active antennas and hardware receiver branches M small, and the number of antenna reconfigurations L within a scanning period minimal. However, it turns out that simultaneously minimizing everything is not possible. Let us consider the following trade-offs.

Corollary 4.3.1.1 implies that, for each value of $l \in \{0, 1, \ldots, L-1\}$, only M^2 out of N^2 columns of $\mathbf{C}_l \otimes \mathbf{C}_l$ have at least a single non-zero element. Minimizing M will decrease the number of non-zero columns of $\mathbf{C}_l \otimes \mathbf{C}_l$. As a result, in order to ensure that all columns of $\mathbf{\Psi}$ in (4.12) have at least a single non-zero element, we need a larger L. This means that there is a trade-off between M and L for a given amount of time within a scanning period. Fig. 4.2 illustrates the trade-off between M and L for $N = 28$ where we vary M from 3 to 28. The value of L is computed by using the lower bound formula given by (4.18) to simplify the illustration. As a final remark on the trade-off between M and L, it is also important to observe from (4.11) that the size of $\mathbf{\Psi}$ to be inverted depends quadratically on M and only linearly on L.

Secondly, consider the relation between L, M, the number of scanning periods P, and the total number of received samples per time slot per antenna S. Recall that the larger PS the better the quality of the estimate of $\mathbf{R}_{y_l}$ in (4.9). If we have a fixed total sensing time (which implies a given PSL), a larger PS implies a smaller L, which in turn, also implies a larger number of antennas M that need to be activated in each time slot. Hence, for a given total sensing time PSL, we have a trade-off between M and the quality of the estimate of $\mathbf{R}_{y_l}$ in (4.9). This trade-off can also be illustrated by Fig. 4.2. For example, by fixing PSL to $PSL \approx \alpha$, we can compute for every value of PS the corresponding value of L as $L = \text{round}(\alpha/(PS))$ and relate this to a value of M from Fig. 4.2.

Alternatively, we might also require a certain quality for the estimate of $\mathbf{R}_{y_l}$ in (4.9) and fix PS. In this case, the trade-off now is between M and the total sensing time PSL. Again, we can use Fig. 4.2 to illustrate this trade-off. For example, by fixing PS to $PS \approx \beta$, we can compute for every value of PSL the corresponding value of L as $L = \text{round}(PSL/\beta)$ and relate this to a value of M

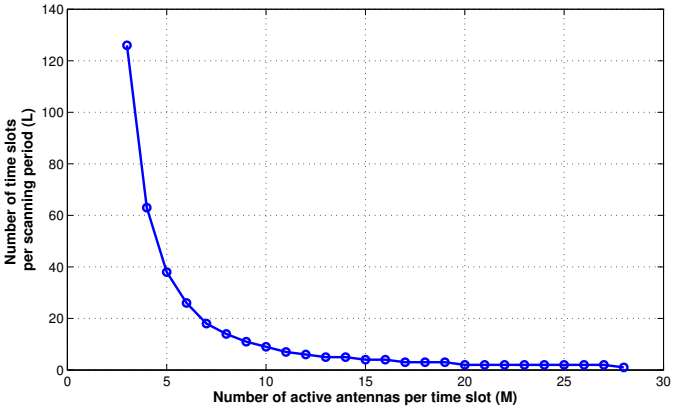


Figure 4.2: Illustration of the trade-off between the number of active antennas M and a lower bound on the number of time slots per scanning period L. Here, we have $N = 28$.

from Fig. 4.2.

4.4 Source Correlation Reconstruction and DOA Estimation

After the reconstruction of $\mathbf{R}_x$, any covariance-based method capable of handling correlated sources from Subsection 4.1.1 assuming no subsampling, i.e., $\mathbf{C} = \mathbf{I}_N$, can basically be used for DOA estimation. Examples are the spatial smoothing and MUSIC (or root-MUSIC) approach of [48, 78], the ℓ_1-SRACV method of [79], or possibly SPICE [80, 81]. In addition, we here also introduce some new covariance-based approaches which again rely on gridding the angular domain.

4.4.1 Least Squares Approach

The first possible approach is based on defining a grid of investigated angles in the angular domain and using a model similar to (4.6) but now with no subsampling, i.e., $\mathbf{C} = \mathbf{I}_N$. Using this model and taking (4.2) into account, we can write

$$\text{vec}\,(\mathbf{R}_x) = (\tilde{\mathbf{A}}^* \otimes \tilde{\mathbf{A}})\text{vec}\,(\mathbf{R}_{\tilde{s}}) + \sigma_n^2 \text{vec}\,(\mathbf{I}_N)\,. \tag{4.20}$$

where $\tilde{\mathbf{A}} = [\mathbf{a}(\tilde{\theta}_1), \mathbf{a}(\tilde{\theta}_2), \ldots, \mathbf{a}(\tilde{\theta}_Q)]$ is the $N \times Q$ extended array response matrix for the underlying ULA with $\mathbf{a}(\tilde{\theta}_q)$ the array response vector for the q-th investi-

gated angle $\tilde{\theta}_q$. From (4.20), we can reconstruct vec $(\mathbf{R}_{\tilde{s}})$ from vec$(\mathbf{R}_x)$ using LS, under the assumption that $\tilde{\mathbf{A}}$ in (4.20) has full column rank, which is only possible if $N \geq Q$. Assuming that $\tilde{\mathbf{A}}$ has full column rank, solving (4.20) using LS leads to

$$\text{vec}(\hat{\mathbf{R}}_{\tilde{s}}) = \left((\tilde{\mathbf{A}}^* \otimes \tilde{\mathbf{A}})^H (\tilde{\mathbf{A}}^* \otimes \tilde{\mathbf{A}})\right)^{-1} (\tilde{\mathbf{A}}^* \otimes \tilde{\mathbf{A}})^H \text{vec}\,(\mathbf{R}_x)\,. \tag{4.21}$$

Because of the Vandermonde structure of $\tilde{\mathbf{A}}$, it will always have full column rank if we take $N \geq Q$. However, in order to obtain a well-conditioned $\tilde{\mathbf{A}}$, we can take $N = Q$, use a half wavelength spacing ($d = 0.5$) for the underlying ULA in (4.1), and adopt an inverse sinusoidal angular grid where the investigated angles $\tilde{\theta}_q$ are defined as

$$\tilde{\theta}_q = \sin^{-1}\left(\frac{2}{Q}\left(q - 1 - \left\lceil \frac{Q-1}{2} \right\rceil\right)\right),\ q = 1, 2, \ldots, Q. \tag{4.22}$$

We can then easily derive that $\tilde{\mathbf{A}}$ is a permuted version of the inverse discrete Fourier transform (IDFT) matrix, which is a unitary matrix that does not introduce any noise enhancement when inverting the matrix. Furthermore, this also means that applying the inverse of $\tilde{\mathbf{A}}^* \otimes \tilde{\mathbf{A}}$ to a vector can easily be computed using fast Fourier transform (FFT) operations, leading to a complexity of order $N^2 \log N$. The diagonal of the computed estimate of the correlation matrix $\hat{\mathbf{R}}_{\tilde{s}}$ indicates the received power at the investigated angles $\{\tilde{\theta}_q\}_{q=1}^{Q}$. Therefore, diag$(\hat{\mathbf{R}}_{\tilde{s}})$ can be perceived as the angular power spectrum estimate. The estimates of the actual DOAs can be found by locating the peaks of this angular power spectrum estimate. The off-diagonal components of $\hat{\mathbf{R}}_{\tilde{s}}$, on the other hand, reveal the correlations between the signals at the different investigated angles.

4.4.2 Sparsity-Regularized Least Squares Approach

As the number of investigated angles Q reduces, the probability that the DOA of a particular point source k is not located on or nearby a grid point increases. Since the LS approach requires $N \geq Q$, when N is small, Q will be small and the angular resolution of the LS estimate is poor and it will be challenging to estimate the DOA of off-grid point sources. One way to mitigate this problem is to take a finer grid of investigated angles by allowing $Q > N$. When this is the case, the resulting $\tilde{\mathbf{A}}^* \otimes \tilde{\mathbf{A}}$ in (4.20) is a wide matrix and its columns play the role of an overcomplete basis for vec $(\mathbf{R}_x)$ in (4.20). In order to solve the resulting underdetermined problem, a popular idea is to assume that vec $(\mathbf{R}_{\tilde{s}})$ is generally sparse and formulate the

estimate of $\mathbf{R}_{\tilde{s}}$ as a solution of the sparsity-regularized LS problem:

$$\hat{\mathbf{R}}_{\tilde{s}} = \arg\min_{\mathbf{R}_{\tilde{s}}} \left\| \text{vec}\,(\mathbf{R}_x) - (\tilde{\mathbf{A}}^* \otimes \tilde{\mathbf{A}}) \text{vec}\,(\mathbf{R}_{\tilde{s}}) \right\|_2^2 + \lambda \left\| \text{vec}\,(\mathbf{R}_{\tilde{s}}) \right\|_1 \qquad (4.23)$$

where the weight $\lambda \geq 0$ balances the sparsity-bias tradeoff. This can be regarded as an extension of the method proposed in [85], which is designed to handle only uncorrelated sources.

In the noiseless case, it is interesting to discuss conditions on $\tilde{\mathbf{A}}^* \otimes \tilde{\mathbf{A}}$ that guarantee perfect reconstruction of vec $(\mathbf{R}_{\tilde{s}})$ in (4.23). If the number of pairs of correlated sources is given by K', then it is clear that vec$(\mathbf{R}_{\tilde{s}})$ is a $\tilde{K}$-sparse vector, i.e., it has at most $\tilde{K}$ non-zero elements where $\tilde{K} = K + 2K'$. The worst case occurs when $\tilde{K} = K^2$, i.e., each source is highly correlated to the other $K - 1$ sources. When we try to recover vec $(\mathbf{R}_{\tilde{s}})$ from vec $(\mathbf{R}_x)$ using ℓ_0-norm minimization, it is well known that the Kruskal rank of $\tilde{\mathbf{A}}^* \otimes \tilde{\mathbf{A}}$ should satisfy krank$(\tilde{\mathbf{A}}^* \otimes \tilde{\mathbf{A}}) \geq 2\tilde{K}$ [87]. Since we now have a wide matrix $\tilde{\mathbf{A}}$, krank$(\tilde{\mathbf{A}}^* \otimes \tilde{\mathbf{A}})$ = krank$(\tilde{\mathbf{A}})$ [88]. Based on this fact and by exploiting the Vandermonde structure of $\tilde{\mathbf{A}}$, it is clear that the necessary and sufficient condition that needs to be satified is given by $N \geq 2\tilde{K}$. For vec $(\mathbf{R}_{\tilde{s}})$ reconstruction using ℓ_1-norm minimization, the coherence property of $\tilde{\mathbf{A}}^* \otimes \tilde{\mathbf{A}}$ is of particular interest. Given the definition of coherence of $\tilde{\mathbf{A}}$ as $\gamma_{\tilde{\mathbf{A}}} = \max_{i \neq j} \frac{|\mathbf{a}(\tilde{\theta}_i)^H \mathbf{a}(\tilde{\theta}_j)|}{\|\mathbf{a}(\tilde{\theta}_i)\| \|\mathbf{a}(\tilde{\theta}_j)\|}$ [89], it is straightforward to show that $\gamma_{\tilde{\mathbf{A}}^* \otimes \tilde{\mathbf{A}}} = \gamma_{\tilde{\mathbf{A}}}$ [88]. As a result, in a noiseless scenario, a loose sufficient condition for a unique reconstruction of vec $(\mathbf{R}_{\tilde{s}})$ using ℓ_1-norm minimization is given by $\tilde{K} \leq \frac{1}{2}(1 + \frac{1}{\gamma_{\tilde{\mathbf{A}}}})$ [89].

4.4.3 Spatial Smoothing and MUSIC

Since we use a ULA of N antennas as our underlying array, it is also possible to apply the spatial smoothing procedure of [48] or [78] to the spatial correlation matrix estimate $\hat{\mathbf{R}}_x$ obtained from (4.13) or (4.14). Here, we opt to employ the forward-backward spatial smoothing (FBSS) introduced in [78], which is theoretically able to detect more correlated sources than that can be detected by the forward technique in [48]. Specifically, we divide the underlying ULA into N_s overlapping subarrays, each of which has N_a physical antennas, and compute the $N_a \times N_a$ spatially smoothed correlation matrix $\hat{\bar{\mathbf{R}}}_x$ as

$$[\hat{\bar{\mathbf{R}}}_x]_{i,j} = \frac{1}{2N_s} \sum_{i'=0}^{N_s-1} \left([\hat{\mathbf{R}}_x]_{i+i',j+i'} + [\hat{\mathbf{R}}_x^*]_{N+1-i-i',N+1-j-i'} \right) \qquad (4.24)$$

where $[\hat{\mathbf{R}}_x]_{i,j}$ represents the element of $\hat{\mathbf{R}}_x$ at the i-th row and the j-th column. We can then apply the MUSIC algorithm of [46] to the resulting $\hat{\bar{\mathbf{R}}}_x$ in order to

produce high resolution DOA estimates. It is important to note that a larger N_s implies a smaller N_a and vice versa, and that the maximum number of sources that can be detected by MUSIC after the FBSS preprocessing scheme is given by $\min(2N_s, N_a - 1)$ [78]. It is easy to show that for our underlying ULA, the optimal settings for N_s and N_a, which lead to the largest possible number of sources that can be detected, is given by $N_s = \lceil \frac{N}{3} \rceil$ with $N_a = N + 1 - \lceil \frac{N}{3} \rceil$.

4.5 Discussion

We first would like to underline some important issues with respect to the LS formulation for DOA estimation discussed in Section 4.4.1. Recall that in the LS formulation, the limited grid resolution in the angular domain might seriously affect the estimation of the DOA of point sources, especially when it is not located nearby the grid. When this is the case, we might expect MUSIC with spatial smoothing discussed in Section 4.4.3 to perform better than the LS approach. However, it is important to note that the LS approach also has its own merits. In fact, the gridding performed in the LS method aims at estimating the general angular power spectrum represented by the diagonal of the estimated signal correlation matrix $\mathbf{R}_{\tilde{s}}$, without any sparsity considerations (i.e., by allowing $Q \leq N$). In that sense, our LS approach can actually be interpreted as a conventional periodogram approach used for spectral estimation. Let us for instance consider the case where we have a source that is occupying a whole angular band (e.g., there are no point sources). We then have a conventional spectral estimation problem instead of a line spectrum estimation problem and there is a good chance for the LS approach to outperform MUSIC, which means that the considered LS angular power spectrum reconstruction method is reasonable in some cases. More details can be found in Section 4.6. Some extensions of MUSIC have been proposed in [90, 91] to estimate the DOA of spatially distributed sources but these approaches are not really designed to estimate an arbitrary smooth angular power spectrum.

Both the LS and the sparsity-regularized LS provide information about how the sources are correlated to each other. This information is available in the off-diagonal components of $\mathbf{R}_{\tilde{s}}$ and cannot be produced using the MUSIC approach. With the ability to use LS, sparsity-regularized LS, and MUSIC with spatial smoothing, our dynamic linear array approach has two features that complement each other. The gridding approach assisted with LS is very useful for a smooth angular power spectrum estimation while MUSIC with spatial smoothing is a more appropriate tool for the DOA estimation of point sources. In addition, subject to

the sparsity of the angular power spectrum, the sparsity-regularized LS approach is able to produce an estimate of a smooth angular power spectrum as well as accurate DOA estimates for point sources as long as the grid is sufficiently fine.

The fact that $\text{diag}(\hat{\mathbf{R}}_{\tilde{s}})$ provides the estimate of the received power at the investigated angles $\{\tilde{\theta}_q\}_{q=1}^{Q}$ can also be used as a motivation to add a positivity constraint on $\text{diag}(\mathbf{R}_{\tilde{s}})$ in the ordinary LS problem discussed in Section 4.4.1. This leads to the following constrained LS problem:

$$\hat{\mathbf{R}}_{\tilde{s}} = \arg\min_{\mathbf{R}_{\tilde{s}}} \left\| \text{vec}\,(\mathbf{R}_x) - (\tilde{\mathbf{A}}^* \otimes \tilde{\mathbf{A}})\text{vec}\,(\mathbf{R}_{\tilde{s}}) \right\|_2^2 \text{ s.t. } \text{diag}(\mathbf{R}_{\tilde{s}}) \geq \mathbf{0}_Q \quad (4.25)$$

with $\mathbf{0}_n$ denoting an $n \times 1$ vector containing only zeros. Furthermore, if a higher computational complexity is acceptable, we can even apply a p.s.d. constraint on $\mathbf{R}_{\tilde{s}}$, which means that the constrained LS problem in (4.25) now becomes

$$\hat{\mathbf{R}}_{\tilde{s}} = \arg\min_{\mathbf{R}_{\tilde{s}}} \left\| \text{vec}\,(\mathbf{R}_x) - (\tilde{\mathbf{A}}^* \otimes \tilde{\mathbf{A}})\text{vec}\,(\mathbf{R}_{\tilde{s}}) \right\|_2^2 \text{ s.t. } \mathbf{R}_{\tilde{s}} \succeq \mathbf{0}_{Q\times Q}. \quad (4.26)$$

Similarly, we can also add a positivity or p.s.d. constraint on $\text{diag}(\mathbf{R}_{\tilde{s}})$ in (4.23) for the sparsity-regularized LS approach discussed in Section 4.4.2.

Another interesting observation is that the LS and sparsity-regularized LS approaches can also be adapted to their one-step counterparts. Instead of first solving (4.11) to reconstruct $\mathbf{R}_x$ from $\mathbf{r}_y$ and then (4.20) to reconstruct $\mathbf{R}_{\tilde{s}}$ from $\mathbf{R}_x$, it is actually possible to reconstruct $\mathbf{R}_{\tilde{s}}$ directly from $\mathbf{r}_y$ in (4.11). This is performed by combining (4.11) and (4.20) and solving the resulting problem using a single LS or sparsity-regularized LS operation. More specifically, we can write

$$\mathbf{r}_y = \mathbf{\Psi}(\tilde{\mathbf{A}}^* \otimes \tilde{\mathbf{A}})\text{vec}\,(\mathbf{R}_{\tilde{s}}) + \sigma_n^2 \mathbf{\Psi}\text{vec}\,(\mathbf{I}_N) = \mathbf{G}\text{vec}\,(\mathbf{R}_{\tilde{s}}) + \sigma_n^2 \mathbf{\Psi}\text{vec}\,(\mathbf{I}_N) \quad (4.27)$$

where the $M^2L \times Q^2$ matrix $\mathbf{G}$ is given by

$$\mathbf{G} = [((\mathbf{C}_0\tilde{\mathbf{A}})^* \otimes (\mathbf{C}_0\tilde{\mathbf{A}}))^T, ((\mathbf{C}_1\tilde{\mathbf{A}})^* \otimes (\mathbf{C}_1\tilde{\mathbf{A}}))^T, \ldots, ((\mathbf{C}_{L-1}\tilde{\mathbf{A}})^* \otimes (\mathbf{C}_{L-1}\tilde{\mathbf{A}}))^T]^T. \quad (4.28)$$

From this equation, we can directly adopt a one-step LS or sparsity-regularized LS to reconstruct $\text{vec}\,(\mathbf{R}_{\tilde{s}})$ from $\mathbf{r}_y$. We expect that the corresponding reconstruction conditions, i.e., full column rank condition on $\mathbf{G}$ for LS and $\text{krank}(\mathbf{G}) \geq 2\tilde{K}$ or $\tilde{K} \leq \frac{1}{2}(1 + \frac{1}{\gamma_{\mathbf{G}}})$ conditions for sparsity-regularized LS, will be less strict than in the two-step approaches. However, since it is not directly clear what properties are required to obtain these conditions for the one-step approaches, we generally advocate to first solve (4.11) and then reconstruct $\mathbf{R}_{\tilde{s}}$ from (4.20) using LS or sparsity-regularized LS.

4.6 Numerical Study

In this section, we evaluate the proposed approaches using numerical experiments. In general, we run the proposed dynamic array through periodic scanning discussed in Section 4.2 and then reconstruct the spatial correlation matrix $\mathbf{R}_x$ from $\mathbf{r}_y$ in (4.11) using ordinary LS except for the last experiment where we use (4.14) to reconstruct $\mathbf{R}_x$. Next, given the estimate of $\mathbf{R}_x$, we evaluate the main source correlation matrix reconstruction and DOA estimation approaches elaborated in Section 4.4. For all experiments, we also apply the positivity constraint on the diagonal elements of $\mathbf{R}_{\tilde{s}}$ when we adopt the sparsity-regularized LS since adding this convex constraint does not incur a large computational cost. In general, we consider correlated sources as well as a spatially and temporally white noise, and we assume that the signals coming from different sources have equal power with the signal to noise ratio (SNR) defined with respect to the power of each signal at each antenna.

Under an SNR of 0 dB, we first conduct three experiments and examine the resulting LS, sparsity-regularized LS, and MUSIC angular power spectrum plots. In the first experiment, we consider a ULA of $N = 40$ antennas with half wavelength spacing as our underlying array and set the number of time slots per scanning period to $L = 28$. We intend to select the activated antennas in each time slot such that the number of active antennas per time slot M is minimal. Observe that, according to (4.19), the lower bound for M in this setting is given by $M \geq 8$. In this simulation study however, we run Algorithm 4.2 given in Table 4.2 for $N = 40$ and $L = 28$ in order to obtain a suboptimal solution for M and the indices of the antennas that are activated in each time slot, which is given by $\{\Gamma_l\}_{l=0}^{27}$. This results in $M = 10$ (which is larger than the lower bound) and produces the indices of the corresponding 10 active antennas in each of the 28 time slots. Note that the antenna array setup suggested by $\{\Gamma_l\}_{l=0}^{27}$ produced by Algorithm 4.2 leads to a full column rank 2800×1600 matrix $\mathbf{\Psi}$ in (4.12). The total number of time samples per time slot is $S = 1$ and the total number of scanning periods is $P = 57$ leading to a total number of time samples of $PSL = 1596$. For the LS approach, we set the number of grid points to $Q = N = 40$ and the investigated angles $\{\tilde{\theta}_q\}_{q=1}^{40}$ according to (4.22) in order to produce a well-conditioned matrix $\tilde{\mathbf{A}}^* \otimes \tilde{\mathbf{A}}$ in (4.20). The number of grid points for the sparsity-regularized LS approach is equal to $Q = 70$ and they are also set according to (4.22). For the sparsity-regularized LS, the weight λ in (4.23) is set to $\lambda = 2.88$. For the MUSIC approach, the FBSS preprocessing scheme is conducted by setting the number of subarrays to $N_s = 14$ and the

number of antennas per subarray to $N_a = 27$. We generate $K = 12$ sources with 9 degrees of separation, that is $\{\theta_k\}_{k=1}^{12} = \{-54^0, -45^0, \ldots, 45^0\}$. Note that the number of sources is more than the number of active antennas per time slot M. In order to investigate the performance of the proposed approach for the case when there is some correlation between the sources, the signal that arrives at angle θ_k is set to be exactly the same as the one arriving at direction θ_{k+6} leading to six pairs of fully correlated sources. The diagonal of $\hat{\mathbf{R}}_{\tilde{s}}$ recovered using ordinary LS and sparsity-regularized LS gives the angular power spectrum estimates at the investigated angles $\tilde{\theta}_q$ and is illustrated in Fig. 4.3. In this figure, the locations of the actual DOAs are indicated by vertical lines for simplicity. We then use the twelve highest peaks in the resulting angular power spectrum estimate to indicate the DOA estimates. We can see how the twelve correlated sources can generally be detected using ordinary LS since they are located nearby the LS grid points. Meanwhile, the sparsity-regularized LS produces a minor grid mismatch effect but has less power in the unoccupied angular band. The resulting MUSIC estimate is also illustrated in Fig. 4.3 and it generally outperforms both the LS estimate and sparsity-regularized LS estimate. It should be noted, however, that both the LS and the sparsity-regularized LS also provide information about the magnitude of the correlation between the signals at the different investigated angles $\tilde{\theta}_q$, which is provided in Fig. 4.4. As it is clear from the figure, both the power of the twelve sources and the magnitude of the cross-correlation between the sources are clearly identified by both the LS and the sparsity-regularized LS approaches.

In the second experiment, we consider a ULA of $N = 25$ antennas with half wavelength spacing as our underlying array. We activate $M = 5$ active antennas in each time slot and run Algorithm 4.1 given in Table 4.1, which produces $L = 36$ and the indices of the corresponding 5 active antennas in each of the 36 time slots. The grid points setting for the sparsity-regularized LS is the same as in the first experiment while for the LS approach, we have $Q = N = 25$ where $\{\tilde{\theta}_q\}_{q=1}^{25}$ is set based on (4.22). Here, the weight λ in (4.23) for the sparsity-regularized LS is set to $\lambda = 3.88$. For the MUSIC approach, the FBSS preprocessing scheme is conducted by setting the number of subarrays to $N_s = 9$ and the number of antennas per subarray to $N_a = 17$. We maintain $S = 1$ but have $P = 44$, which leads to a total number of time samples of $PSL = 1584$. We now generate three pairs of fully correlated sources leading to $K = 6$ sources having DOAs with 10 degrees of separation, that is $\{\theta_k\}_{k=1}^{6} = \{-2.97^0, 7.03^0, \ldots, 47.03^0\}$. Note that we again have more sources than active antennas per time slot M. Fig. 4.5 illustrates the angular power spectrum estimates. Again, the location of the actual DOAs is

indicated by vertical lines. Observe that for this realization, the accuracy of the DOA estimates produced by the sparsity-regularized LS is quite comparable to that of the MUSIC estimates though the sparsity-regularized LS approach introduces a significant amount of power in the unoccupied angular band. The ordinary LS DOA estimates, on the other hand, introduce a significant amount of grid mismatch due to a coarse grid of investigated angles. Fig. 4.6 describes the magnitude of the correlation between the signals at the different investigated angles $\tilde{\theta}_q$ estimated using the ordinary LS (bottom part) and the sparsity-regularized LS (top part) approaches. As it is clear from the figure, both the power of the six sources and the magnitude of the cross-correlation between the sources are better identified by the sparsity-regularized LS than by the ordinary LS approach.

All simulation settings for the third experiment are similar to those for the first experiment, but we now consider a continuous source from 30 to 40 degrees, which is simulated by generating 250 pairs of fully correlated sources with 0.02 degrees of separation. The parameter λ for the sparsity-regularized LS is set to $\lambda = 0.012$. The result is illustrated in Fig. 4.7. The MUSIC algorithm clearly fails for this continuous source scenario while the sparsity-regularized LS and LS approaches better reconstruct the continuous angular range where lower sidelobes and a better resolution are found for the sparsity-regularized LS.

The dynamic array, the FBSS, and the angular grid setting of the second experiment is now used in the fourth experiment to compute the root mean square error (RMSE) between the actual DOAs and the DOA estimates. The DOAs of three pairs of fully coherent sources are randomly generated between -60 and 60 degrees but with a fixed 10 degrees of separation. With $S = 1$ and SNR $= 0$dB, we first vary P from $P = 4$ to $P = 100$. Based on the resulting spectrum, we locate the six highest peaks. For every source, we compute the RMSE between the true DOA and the peak that is closest to this DOA, selected from the earlier determined six highest peaks. Fig. 4.8 illustrates the computed RMSE for the three approaches for a varying P. Observe how the performance of the ordinary LS approach is quite poor due to its limited grid resolution. The performance of the sparsity-regularized LS approach is much better than the ordinary LS approach but it tends to flatten at a particular level which is determined by the resolution of the 70 grid points used by this approach. Meanwhile, the MUSIC approach performs better than the two aforementioned approaches and its RMSE continues to decrease as P increases. A similar situation is also found in Fig. 4.9, where we fix the total number of scanning periods to $P = 7$ and vary the SNR. In this scenario, however, the performance of the sparsity-regularized LS has not yet hit the saturation point defined by the res-

olution of its 70 grid points. In fact, its RMSE continues to decrease as the SNR varies from -10 dB to 0 dB.

A much better performance in terms of RMSE for LS is found when we increase N and Q to $N = Q = 40$ as shown in Fig. 4.10. In this fifth experiment, we only focus on the LS and MUSIC approaches. Here, we have $M = 7$ and run Algorithm 4.1 to produce $L = 48$. We vary P from $P = 1$ to $P = 40$ and the SNR from -10 dB to 0 dB. The FBSS setting for MUSIC is the same as in the first experiment and there are four pairs of fully correlated sources (which again implies $K > M$) whose DOAs are generated in the same way as in the fourth experiment. Note how the performance of the ordinary LS approach is mainly dictated by the grid resolution.

In the last experiment, we investigate the impact of applying the p.s.d. constraint on the reconstructed $\mathbf{R}_x$ (see (4.14)) and focus on the performance of the MUSIC approach by considering different settings of the dynamic array. In general, we set the number of active antennas per time slot to $M = 3$ and examine three different dynamic array settings, i.e., $N = 17$, $N = 14$, and $N = 11$. Given $M = 3$, we execute Algorithm 4.1 in Table 4.1 for $N = 17$, $N = 14$, and $N = 11$, leading to $L = 47$, $L = 33$, and $L = 19$, respectively. For the FBSS process, it is important to note that for different values of N, we also have different optimal values of N_s and N_a. Here, we compute N_s and N_a as $N_s = \lceil \frac{N}{3} \rceil$ and $N_a = N + 1 - \lceil \frac{N}{3} \rceil$, respectively (as suggested in Section 4.4.3). With $S = 1$, we vary P between $P = 4$ and $P = 40$. On top of that, we also evaluate two different SNR values, i.e., SNR $= 0$dB and SNR $= -5$dB. We generate two pairs of fully correlated sources ($K = 4$) and compute the RMSE between the actual and the estimated DOAs by following the same procedure introduced in the fourth experiment. Fig. 4.11 illustrates the computed RMSE for this experiment. As expected, the performance of the MUSIC approach for all dynamic array settings gets worse for lower SNR. We can see that applying the p.s.d. constraint on the reconstructed $\mathbf{R}_x$ indeed improves the performance, especially for $N = 14$. It is interesting to observe that, for a given SNR, a larger performance degradation is experienced when we reduce $N = 14$ and $L = 33$ to $N = 11$ and $L = 19$, respectively, than when we reduce it from $N = 17$ and $L = 47$ to $N = 14$ and $L = 33$, respectively. In fact, when $N = 11$, we only have $N_s = 4$ subarrays of $N_a = 8$ antennas. According to the theoretical analysis in [78], this setting should still be able to estimate $\min(2N_s, N_a - 1) = 7$ correlated sources. However, in practice, we observe that there are a few occasions where there is only a small difference between the value of the fourth largest eigenvalue and that of the noise eigenvalues of the resulting

spatially smoothed matrix $\hat{\bar{\mathbf{R}}}_x$ in (4.24). As a result, the MUSIC approach might not be able to separate the fourth signal eigenvector from the noise subspace. Observe that, for $N = 11$, applying the p.s.d. constraint on the reconstructed $\mathbf{R}_x$ only offers a small improvement.

4.7 Conclusions

In this paper, we have developed a new method to estimate the DOA of possibly fully correlated sources based on second-order statistics by adopting a so-called dynamic array, which is formed by performing a periodic scanning of an underlying ULA having N available antennas. Here, different sets of M antennas are activated in different time slots. We first collect the spatial correlation matrices of the output of the antenna arrays for all time slots and present them as a linear function of the spatial correlation matrix $\mathbf{R}_x$. We then present the theoretical condition that needs to be satisfied to ensure the full column rank condition of the system matrix, which later allows us to reconstruct $\mathbf{R}_x$ using LS. Note that, apart from our dynamic array approach which allows for $\mathbf{R}_x$ reconstruction using LS, it is also possible to use low-rank matrix completion to reconstruct $\mathbf{R}_x$ under a finite number of measurements. This, for instance, is discussed in [92]. However, this topic is beyond the scope of our paper and we consider this as a possible topic for future research. Based on the estimated $\mathbf{R}_x$, we propose three different options. The first option is to define an angular grid of investigated angles where the number of grid points Q is less than or equal to the number of physical antennas N in the underlying array. This allows us to reconstruct the correlation matrix of the incident signals at the investigated angles using LS subject to the full column rank condition of the system matrix. Since the LS signal correlation matrix reconstruction is vulnerable to a grid mismatch effect due to the limited grid resolution, we propose a sparsity-regularized LS approach as the second option and increase the grid resolution by allowing $Q \gg N$. However, this option theoretically works well only when the actual angular power spectrum is sparse. The last option is to apply FBSS on the reconstructed $\mathbf{R}_x$ and use the MUSIC algorithm based on the spatially smoothed correlation matrix. This option might produce high resolution DOA estimates but does not provide information about how the sources are correlated to each other. In general, our dynamic array approach can estimate the DOAs of the impinging signals even when the number of correlated sources is larger than the number of active antennas per time slot. The simulation study has indicated that our method performs satisfactory even when some sources are fully coherent.

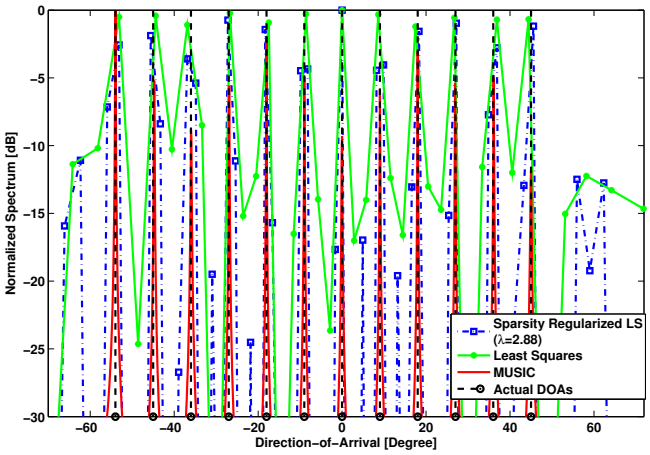


Figure 4.3: Normalized spectrum (dB) of the MUSIC, LS, and sparsity-regularized LS approaches versus DOA (degree) for the first experiment. We have $K = 12$ sources, $N = 40$, $L = 28$, $M = 10$, $P = 57$ and SNR=0 dB. For the LS and sparsity-regularized LS approaches, we have $Q = 40$ and $Q = 70$, respectively

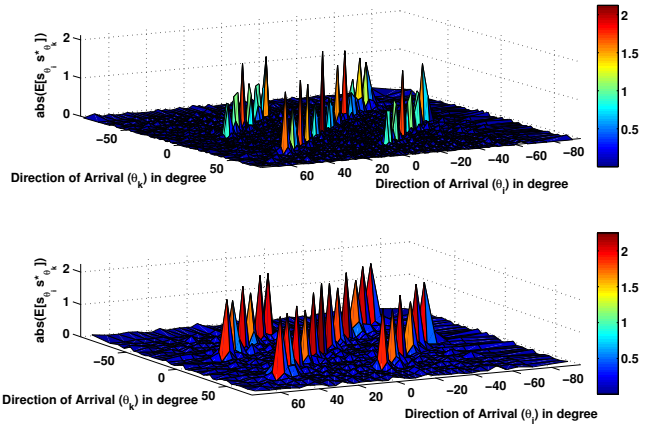


Figure 4.4: The magnitude of the elements of the estimated correlation matrix $\hat{\mathbf{R}}_{\tilde{s}}$ computed using the sparsity-regularized LS (top) and the LS (bottom) approaches for the first experiment. Here $K = 12$, SNR = 0 dB, $N = 40$, $M = 10$, $P = 57$, and $L = 28$. For the LS and sparsity-regularized LS approaches, we have $Q = 40$ and $Q = 70$, respectively.

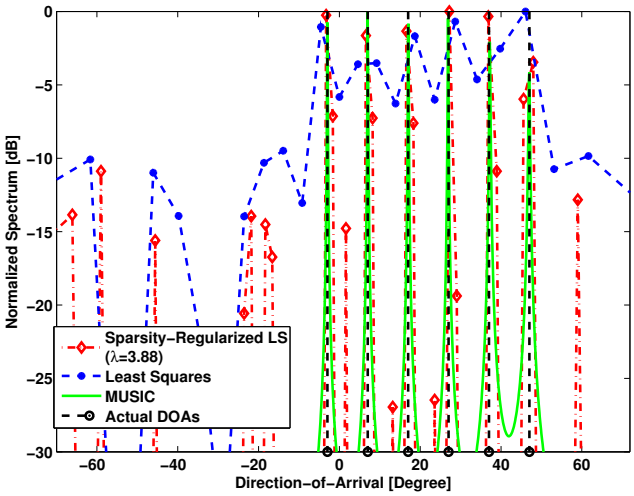


Figure 4.5: Normalized spectrum (dB) of the MUSIC, LS, and sparsity-regularized LS approaches versus DOA (degree) for the second experiment. We have $K = 6$ sources, $N = 25$, $L = 36$, $M = 5$, $P = 44$, and SNR = 0 dB. For the LS and sparsity-regularized LS approaches, we have $Q = 25$ and $Q = 70$, respectively.

Appendix

4.A Proof of Lemma 4.3.1

Let us assume that $\mathbf{e}_i^T$ and $\mathbf{e}_j^T$ are the i-th and the j-th rows of $\mathbf{I}_N$, respectively. It is easy to check that for any $i, j \in \{1, 2, \ldots, N\}$, the Kronecker product $\mathbf{e}_i^T \otimes \mathbf{e}_j^T$ results in a $1 \times N^2$ vector having a single one at the $[(i-1)N + j]$-th position. Correspondingly, $\mathbf{e}_j^T \otimes \mathbf{e}_i^T$ produces a $1 \times N^2$ vector having a single one at the $[(j-1)N + i]$-th position. As a result, if $\mathbf{C}_l$ has $\mathbf{e}_i^T$ and $\mathbf{e}_j^T$ as two of its rows, $\mathbf{C}_l \otimes \mathbf{C}_l$ will definitely have a one in the $[(i-1)N + j]$-th and $[(j-1)N + i]$-th columns. This proves the sufficiency part of the lemma. In order to prove the necessity part, let us assume that $\mathbf{C}_l \otimes \mathbf{C}_l$ has a one in the $[(i-1)N + j]$-th column but either the i-th row of $\mathbf{I}_N$, the j-th row of $\mathbf{I}_N$, or both of them are missing from $\mathbf{C}_l$. Further assume that the row of $\mathbf{C}_l \otimes \mathbf{C}_l$ having a one in the $[(i-1)N + j]$-th column is produced by the Kronecker product operation between two rows of $\mathbf{C}_l$ taken from the a-th and the b-th rows of $\mathbf{I}_N$, i.e., $\mathbf{e}_a^T \otimes \mathbf{e}_b^T$. Now, $\mathbf{e}_a^T \otimes \mathbf{e}_b^T$ results in a $1 \times N^2$ vector having a single one at the $[(a-1)N + b]$-th position. Therefore, it

is obvious that $a = i$ and $b = j$. In other words, the i-th and the j-th rows of $\mathbf{I}_N$ are not missing from $\mathbf{C}_l$, which is a contradiction. A similar proof applies for the row of $\mathbf{C}_l \otimes \mathbf{C}_l$ that has a single one in the $[(j-1)N + i]$-th column. This concludes the proof of this lemma.

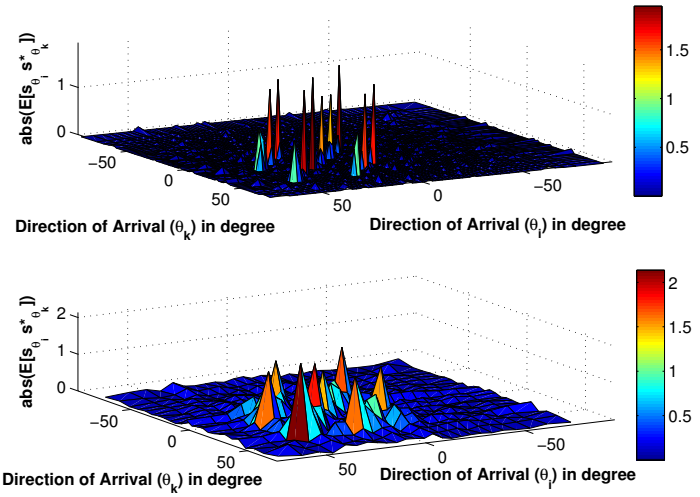


Figure 4.6: The magnitude of the elements of the estimated correlation matrix $\hat{\mathbf{R}}_{\tilde{s}}$ computed using the sparsity-regularized LS (top) and LS (bottom) approaches for the second experiment. Here $K = 6$, SNR = 0 dB, $N = 25$, $M = 5$, $P = 44$, and $L = 36$. For the LS and sparsity-regularized LS approaches, we have $Q = 25$ and $Q = 70$, respectively.

4.B Explanation for Algorithm 4.1 (see Table 4.1)

We use the indicator matrix $\mathbf{Z}^{(f)}$ in Table 4.1 to indicate whether a certain combination of two antennas has been activated in at least one of the first f time slots. Specifically, $[\mathbf{Z}^{(f)}]_{i,j} = 0$ implies that a combination of the i-th and the j-th antennas has never been simultaneously activated in the first f time slots whereas $[\mathbf{Z}^{(f)}]_{i,j} = 1$ indicates that the combination of the i-th and the j-th antennas has been simultaneously activated at least once in the first f time slots. Consequently, it is also obvious that $\mathbf{Z}^{(f)}$ is a symmetric matrix. Based on Theorem 4.3.1, our objective is to guarantee that every possible combination of two antennas in the underlying ULA is active in at least one of the L possible time slots within a scanning

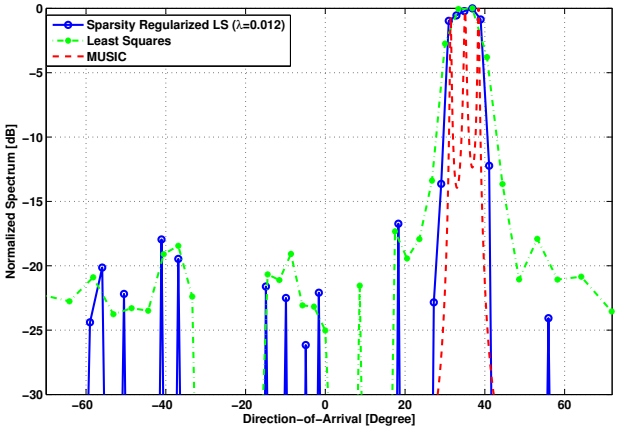


Figure 4.7: Normalized spectrum (dB) of the MUSIC, LS, and sparsity-regularized LS approaches versus DOA (degree) for the third experiment. Here we have $K = 500$ sources with DOAs between 30 and 40 degrees, SNR = 0 dB, $N = 40$, $M = 10$, and $L = 28$. For the LS and sparsity-regularized LS approaches, we have $Q = 40$ and $Q = 70$, respectively.

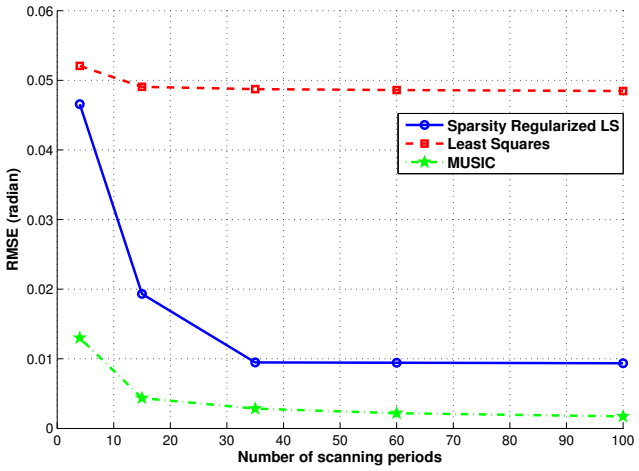


Figure 4.8: The performance of LS, sparsity-regularized LS and MUSIC DOA estimates for different scanning periods P. Here we have SNR $= 0$dB, $N = 25$, $M = 5$, $L = 36$, and $K = 6$ correlated sources whose DOAs are randomly generated with 10 degrees of separation. For the LS and sparsity-regularized LS approaches, we have $Q = 25$ and $Q = 70$, respectively.

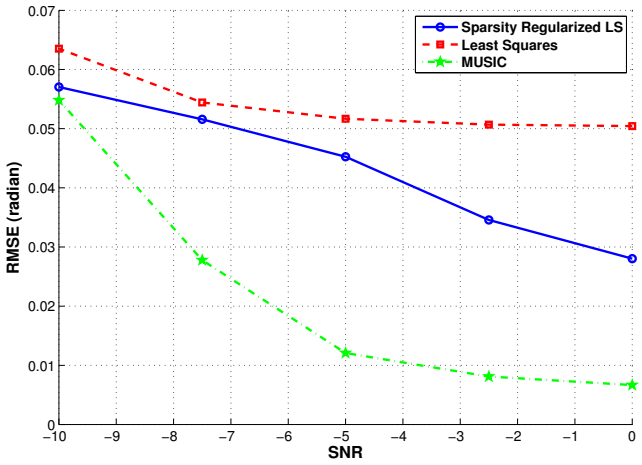


Figure 4.9: The performance of the LS, sparsity-regularized LS and MUSIC DOA estimates for different SNRs. Here we have $P = 7$, $N = 25$, $M = 5$, $L = 36$, and $K = 6$ correlated sources whose DOAs are randomly generated with 10 degrees of separation. For the LS and sparsity-regularized LS approaches, we have $Q = 25$ and $Q = 70$, respectively.

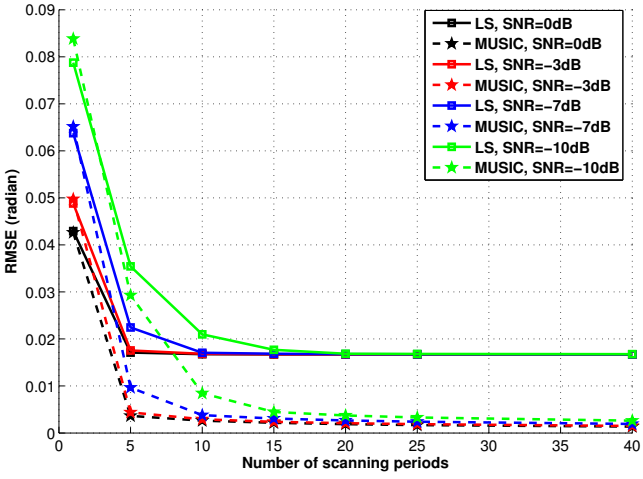


Figure 4.10: The performance of the LS and MUSIC DOA estimates for different SNRs and scanning periods P. Here we have $N = Q = 40$, $M = 7$, $L = 48$, and $K = 8$ correlated sources whose DOAs are randomly generated with 10 degrees of separation.

period. In other words, we are only interested in the off-diagonal components of $\mathbf{Z}^{(f)}$ and thus, we initialize $\mathbf{Z}^{(f)}$ with $\mathbf{Z}^{(0)} = \mathbf{I}_N$.

In general, Algorithm 4.1 consists of L main iterations indicated by a while loop in Table 4.1. The reason to use a while loop to implement the main iterations is due to the fact that L is unknown. The objective of the while loop is to choose M antennas for each time slot. One main iteration corresponds to the selection of M antennas for one time slot. The first task in the main iteration (see step 5 in Table 4.1) is to randomly select a combination of two antennas that has not been used in the previous time slot.

The task of the inner for loop in Table 4.1, which consists of $M-2$ iterations, is to choose the remaining $M-2$ antennas for the considered time slot. For each antenna selection, our aim is to maximize the number of conversion of zeros in $\mathbf{Z}^{(f)}$ to ones. This is done because we want to ensure that each antenna selection results in a maximum number of new combinations of two active antennas that have not been simultaneously used in the previous time slots.

Note that the main iterations will stop once every possible pair of two antennas has been selected for at least one time slot.

4.C Explanation for Algorithm 4.2 (see Table 4.2)

For Algorithm 4.2 in Table 4.2, we use a similar notation to the one used for Algorithm 4.1. However, in this algorithm, all time slots are considered simultaneously and thus we now use $\mathbf{Z}^{(f)}$ to indicate whether a certain combination of two antennas has been used as two of the first $f+1$ active antennas in any time slot. The task of the first for loop in Algorithm 4.2 is to select the first two antennas for each time slot. One iteration corresponds to the selection of the first two antennas for one time slot. Once we have selected the first two active antennas for each time slot, we proceed to the while loop indicated by steps 8-16 in Table 4.2. Here, one iteration of the while loop corresponds to the selection of one additional antennas for all time slots. The reason to use while loop here is due to the fact that M is not known. The inner for loop inside the while loop performs the selection of one additional antenna for one time slot in each iteration. As in Algorithm 4.1, each antenna selection aims to maximize the number of conversion of zeros in $\mathbf{Z}^{(f)}$ to ones.

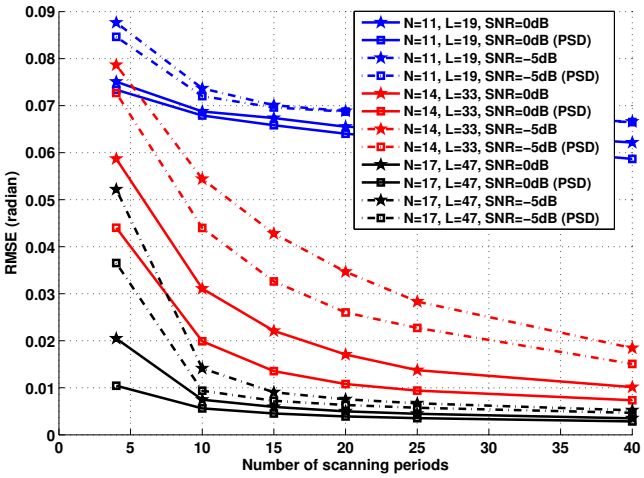


Figure 4.11: The impact of the p.s.d. constraint on the reconstructed $\mathbf{R}_x$. Here, we focus on the MUSIC DOA estimates for different settings of the dynamic array (different N and L) as well as different SNRs and scanning periods P using $M = 3$ active antennas. Here we have $K = 4$ correlated sources whose DOAs are randomly generated with 10 degrees of separation.

Table 4.1: Algorithm 4.1: A greedy algorithm to find a sub-optimal solution for L and $\{\Gamma_l\}_{l=0}^{L-1}$ given M subject to (4.16).

Algorithm 4.1	
1:	Introduce $\mathbf{Z}^{(f)}$ as an $N \times N$ indicator matrix at the f-th iteration and denote its element at the i-th row and the j-th column by $\left[\mathbf{Z}^{(f)}\right]_{i,j}$.
2:	Initialize $f = 0$ and $\mathbf{Z}^{(0)} = \mathbf{I}_N$.
3:	While $\mathbf{Z}^{(f)}$ has at least one zero entry do
4:	Set $f = f + 1$ and $\mathbf{Z}^{(f)} = \mathbf{Z}^{(f-1)}$.
5:	Randomly select $i, j \in \{1, 2, \ldots, N\}$ for which $\left[\mathbf{Z}^{(f)}\right]_{i,j} = 0$ and set $\Gamma_{f-1} = \{i, j\}$. Then also set both $\left[\mathbf{Z}^{(f)}\right]_{i,j}$ and $\left[\mathbf{Z}^{(f)}\right]_{j,i}$ to 1.
6:	for $\kappa = 1$ to $M - 2$ do
7:	Define a set $\Xi = \{1, 2, \ldots, N\} \setminus \Gamma_{f-1}$.
8:	Search in Ξ for the element g that satisfies: $g = \arg\min_{g' \in \Xi} \sum_{i' \in \Gamma_{f-1}} \left[\mathbf{Z}^{(f)}\right]_{i',g'}$.
9:	For all $i' \in \Gamma_{f-1}$ set $\left[\mathbf{Z}^{(f)}\right]_{i',g}$ and $\left[\mathbf{Z}^{(f)}\right]_{g,i'}$ to 1.
10:	Update Γ_{f-1} to $\Gamma_{f-1} = \Gamma_{f-1} \bigcup \{g\}$.
11:	end for
12:	end while
13:	The value of L is given by $L = f$ and the output of this algorithm is $\{\Gamma_l\}_{l=0}^{L-1}$.

Table 4.2: Algorithm 4.2: A greedy algorithm to find a sub-optimal solution for M and $\{\Gamma_l\}_{l=0}^{L-1}$ given L subject to (4.16)

Algorithm 4.2	
1:	Introduce $\mathbf{Z}^{(f)}$ as an $N \times N$ indicator matrix at the f-th iteration and denote its element at the i-th row and the j-th column by $\left[\mathbf{Z}^{(f)}\right]_{i,j}$.
2:	Initialize $f = 0$ and $\mathbf{Z}^{(0)} = \mathbf{I}_N$.
3:	Set $f = f + 1$ and $\mathbf{Z}^{(f)} = \mathbf{Z}^{(f-1)}$.
4:	for $l = 0$ to $L - 1$ do
5:	Randomly select $i, j \in \{1, 2, \ldots, N\}$ for which $\left[\mathbf{Z}^{(f)}\right]_{i,j} = 0$ and set $\Gamma_l = \{i, j\}$.
6:	Set both $\left[\mathbf{Z}^{(f)}\right]_{i,j}$ and $\left[\mathbf{Z}^{(f)}\right]_{j,i}$ to 1.
7:	end for
8:	While $\mathbf{Z}^{(f)}$ has at least one zero entry do
9:	Set $f = f + 1$ and then set $\mathbf{Z}^{(f)} = \mathbf{Z}^{(f-1)}$.
10:	for $\kappa = 0$ to $L - 1$ do
11:	Define a set $\Xi = \{1, 2, \ldots, N\} \setminus \Gamma_\kappa$.
12:	Search in Ξ for the element g that satisfies: $g = \arg\min_{g' \in \Xi} \sum_{i' \in \Gamma_\kappa} \left[\mathbf{Z}^{(f)}\right]_{i',g'}$.
13:	For all $i' \in \Gamma_\kappa$ set $\left[\mathbf{Z}^{(f)}\right]_{i',g}$ and $\left[\mathbf{Z}^{(f)}\right]_{g,i'}$ to 1.
14:	Update Γ_κ to $\Gamma_\kappa = \Gamma_\kappa \bigcup \{g\}$.
15:	end for
16:	end while
17:	The value of M is given by $M = f + 1$ and the output of this algorithm is $\{\Gamma_l\}_{l=0}^{L-1}$.

Chapter 5

Compressive Periodogram Reconstruction Using Uniform Binning

Dyonisius Dony Ariananda, Daniel Romero, and Geert Leus

Submitted to IEEE Transactions on Signal Processing (Accepted with Mandatory Minor Revision).

Abstract

In this paper, two problems that show great similarities are examined. The first problem is the reconstruction of the angular-domain periodogram from spatial-domain signals received at different time indices. The second one is the reconstruction of the frequency-domain periodogram from time-domain signals received at different wireless sensors. We split the entire angular or frequency band into uniform bins. The bin size is set such that the received spectra at two frequencies or angles, whose distance is equal to or larger than the size of a bin, are uncorrelated. These problems in the two different domains lead to a similar circulant structure in the so-called coset correlation matrix. This circulant structure allows for a strong compression and a simple least-squares reconstruction method. The latter is possible under the full column rank condition of the system matrix, which can be achieved by designing the spatial or temporal sampling patterns based on a circular sparse ruler. We analyze the statistical performance of the compressively reconstructed periodogram including bias and variance. We further consider the case when the bins are so small that the received spectra at two frequencies or angles, with a spacing between them larger than the size of the bin, can still be correlated. In this case, the resulting coset correlation matrix is generally not circulant and thus a special approach is required.

5.1 Introduction

The similarity between spectral analysis problems in the spatial-angular domain and in the time-frequency domain has attracted signal processing researchers since the 1970s. Direction of arrival (DOA) estimation and frequency identification of sinusoids are examples of such similar problems examined during that period [93]. The renewed interest in spectral analysis problems, especially due to the emergence of compressive sampling, has spurred reinvestigations on this similarity because, when time-domain or spatial-domain compression is introduced, this similarity can be exploited to tackle different problems using the same algorithmic approach.

This paper focuses on both the reconstruction of the angular-domain periodogram from far-field signals received by an antenna array at different time indices (problem P1) and that of the frequency-domain periodogram from the time-domain signals received by different wireless sensors (problem P2). It further underlines the similarity between P1 and P2. Unless otherwise stated, the entire angular or frequency band is divided into uniform bins, where the size of the bins is configured

such that the received spectra at two frequencies or angles, whose distance is equal to or larger than the size of a bin, are uncorrelated. In this case, the so-called coset correlation matrix will have a circulant structure, which allows the use of a periodic non-uniform linear array (non-ULA) in P1 and a multi-coset sampler in P2 in order to produce a strong compression.

Our work in P1 is motivated in part by [39], which attempts to reconstruct the angular spectrum from spatial-domain samples received by a non-ULA. Comparable works to [39] for P2 are [21] and [38], which focus on the analog signal reconstruction from its sub-Nyquist rate samples. However, the aim of [21, 38, 39] to reconstruct the original spectrum or signal leads to an underdetermined problem, which has a unique solution only if we add constraints on the spectrum such as a sparsity constraint. A less ambitious goal in the context of P2 is to reconstruct the power spectrum instead of the actual signal from sub-Nyquist rate samples. For wide-sense stationary (WSS) signals, this has been shown to be possible in [43] and [94] without applying a sparsity constraint on the power spectrum. Meanwhile, the work of [44] assumes the existence of a multiband signal where different bands are uncorrelated. In this case, the diagonal structure of the correlation matrix of the entries at different bands can be exploited. Note though that [44] does not focus on the strongest compression rate and uses frequency smoothing to approximate the correlation matrix computation as it relies on a single realization of the received signal. Comparable works to [94] in P1 are [10, 41, 42], which aim to estimate the DOA of uncorrelated point sources with fewer antennas than sources. This is possible because for uncorrelated point sources, the spatial correlation matrix of the received signals also has a Toeplitz structure. Hence, for a given ULA, we can deactivate some antennas but still manage to estimate the spatial correlation at all lags. For example, [42] and [41] suggest to place the active antennas based on a nested or coprime array, respectively, which results in a longer virtual array called the difference co-array (which is uniform in this case). As the difference co-array generally has more antennas and a larger aperture than the actual array, the degrees of freedom are increased allowing [42] and [41] to estimate the DOA of more uncorrelated sources than sensors. In a more optimal way, a uniform difference co-array can also be obtained by the minimum redundancy array (MRA) of [12], but the nested and coprime arrays present many advantages due to their algebraic construction. MRAs have been used in [10] to estimate the DOA of more uncorrelated sources than sensors, or more generally, to estimate the angular-domain power spectrum.

Unlike [39], our work for P1 focuses on the angular periodogram reconstruction

(similar to [10]). This allows us to have an overdetermined problem that is solvable even without a sparsity constraint on the angular domain. This is beneficial for applications that require only information about the angular periodogram and not the actual angular spectrum. Our work is also different from [10, 41, 42] as we do not exploit the Toeplitz structure of the spatial correlation matrix. As for P2, we focus on frequency periodogram reconstruction (unlike [21, 38]) but we do not exploit the Toeplitz structure of the time-domain correlation matrix (unlike [94]). On the other hand, the problem handled by [44] can be considered as a special case of P2 but, unlike [44], we aim for the strongest compression rate which is achieved by exploiting the circulant structure of the coset correlation matrix and solving the minimal circular sparse ruler problem. Moreover, unlike [44], we also exploit the signals received by different sensors to estimate the correlation matrix.

Also related to P2, a cooperative compressive wideband spectrum sensing scheme for cognitive radio (CR) networks is proposed in [29]. While [29] can reduce the required sampling rate per CR, its focus on reconstructing the spectrum or the spectrum support requires a sparsity constraint on the original spectrum. Unlike [29], [95] focuses on compressively estimating the power spectrum instead of the spectrum by extending [94] for a cooperative scenario. However, while the required sampling rate per sensor can be lowered without applying a sparsity constraint on the power spectrum, the exploitation of the cross-spectra between signals at different sensors in [95] requires the knowledge of the channel state information (CSI). Our approach for P2 does not require a sparsity constraint on the original periodogram (unlike [29]) and it does not require CSI since we are not interested in the cross-spectra between samples at different sensors (unlike [95]). In [96], each wireless sensor applies a threshold on the measured average signal power after applying a random wideband filter. The threshold output is then communicated as a few bits to a fusion centre, which uses them to recover the power spectrum by generalizing the problem in the form of inequalities. The achievable compression rate with such a system is not clear though, in contrast to what we will present in this paper.

In more advanced problems, such as cyclic spectrum reconstruction from sub-Nyquist rate samples of cyclostationary signals in [45, 97, 11] or angular power spectrum reconstruction from signals produced by correlated sources in [98], finding a special structure in the resulting correlation matrix that can be exploited to perform compression is challenging. A similar challenge is faced in Section 5.7, where we consider the case when we reduce the bin size such that the received spectra at two frequencies or angles with a spacing larger than the bin size can still

be correlated. As the resulting coset correlation matrix in this case is generally not circulant, we further develop the concepts originally introduced in [11] and [98] to solve our problem.

We now would like to summarize the advantages of our approach and highlight our contribution.

- We propose a compressive periodogram reconstruction approach, which does not rely on any sparsity constraint on the original signal or the periodogram. Moreover, it is based on a simple least-squares (LS) algorithm leading to a low complexity.

- In our approach, we also focus on the strongest possible compression that maintains the identifiability of the periodogram, which is shown to be related to a minimal circular sparse ruler.

- Our approach does not require any knowledge of the CSI.

- The statistical performance analysis of the compressively reconstructed periodogram is also provided.

- Our approach can also be modified to handle cases where the spectra in different bins are correlated.

This paper is organized as follows. The system model description (including the definition of the so-called coset correlation matrix) and the problem statement are provided in Section 5.2. Section 5.3 discusses the spatial (for P1) or temporal (for P2) compression as well as periodogram reconstruction using LS. Here, the condition for the system matrix to have full column rank and its connection to the minimal circular sparse ruler problem are provided. Section 5.4 shows how to approximate the expectation operation in the correlation matrix computation and summarizes the procedure to compressively estimate the periodogram. In Section 5.5, we provide an analysis on the statistical performance of the compressively reconstructed periodogram including a bias and variance analysis. Sections 5.2-5.5 assume that the received signals at different time instants (for P1) or at different sensors (for P2) have the same statistics. To handle more general cases, we propose a multi-cluster model in Section 5.6, which considers clusters of time indices in P1 or clusters of sensors in P2 and assumes that the signal statistics are only constant within a cluster. Another case is discussed in Section 5.7, where the received spectra at two frequencies or angles located at different predefined bins can still be

correlated. Some numerical studies are elaborated in Section 5.8 and Section 5.9 provides conclusions.

Notation: Upper (lower) boldface letters are used to denote matrices (column vectors). Given an $N \times N$ matrix $\mathbf{X}$, diag$(\mathbf{X})$ is an $N \times 1$ vector containing the main diagonal entries of $\mathbf{X}$. Given an $N \times 1$ vector $\mathbf{x}$, diag$(\mathbf{x})$ is an $N \times N$ diagonal matrix whose diagonal entries are given by the entries of $\mathbf{x}$.

5.2 System Model

5.2.1 Model Description and Problem Statement

We aim at estimating the following spectral representation of the power of a process $x[\tilde{n}]$:

$$\begin{aligned} P_x(\vartheta) &= \lim_{\tilde{N}\to\infty} E\left\{ \frac{1}{\tilde{N}} \left| \sum_{\tilde{n}=0}^{\tilde{N}-1} x[\tilde{n}] e^{-j\vartheta\tilde{n}} \right|^2 \right\} \\ &= \lim_{\tilde{N}\to\infty} E\left\{ \frac{1}{\tilde{N}} \left| X_{(\tilde{N})}(\vartheta) \right|^2 \right\}. \end{aligned} \tag{5.1}$$

Here, $x[\tilde{n}]$ represents either the spatial-domain process at the output of a ULA for P1 or the time-domain process sensed by a wireless sensor for P2. In addition, $X_{(\tilde{N})}(\vartheta)$ represents either the value of the angular spectrum at angle $\sin^{-1}(2\vartheta)$ for P1 or that of the frequency spectrum at frequency ϑ for P2, with $\vartheta \in [-0.5, 0.5)$. Note from [93] that, for a *WSS process* $x[\tilde{n}]$, $P_x(\vartheta)$ represents the *power spectrum*. To estimate $P_x(\vartheta)$ in (5.1), consider the $\tilde{N} \times 1$ complex-valued observation vectors $\mathbf{x}_t = [x_t[0], x_t[1], \ldots, x_t[\tilde{N}-1]]^T$, $t = 1, 2 \ldots, \tau$, where $x_t[\tilde{n}]$ represents the output of the $(\tilde{n}+1)$-th antenna in the ULA of $\tilde{N}$ half-wavelength spaced antennas at time index t for P1 or the $(\tilde{n}+1)$-th sample out of $\tilde{N}$ successive samples produced by the Nyquist-rate sampler at the t-th sensor for P2. To acquire an accurate Fourier interpretation, we assume a relatively large $\tilde{N}$, which is affordable for P2 and also realistic for P1, if we consider millimeter wave imaging applications where the antenna spacing is very small and thus the required aperture has to be covered by a large number of antennas [39]. Denote the discrete-time Fourier transform (DTFT) of $x_t[\tilde{n}]$ by $X_t(\vartheta)$. As $X_t(\vartheta)$ at $\vartheta \in [-0.5, 0)$ is a replica of $X_t(\vartheta)$ at $\vartheta \in [0.5, 1)$, we can focus on $X_t(\vartheta)$ in $\vartheta \in [0, 1)$.

Next, we divide the $\tilde{N}$ uniform grid points (that is, the antennas of the ULA for P1 or the indices of the Nyquist-rate samples for P2) into L non-overlapping blocks

of N uniform grid points. We collect all the $(n+1)$-th grid points from each of the L blocks and label this collection of grid points, i.e., $\{\tilde{n} \in \{0, 1, \ldots, \tilde{N} - 1\} | \tilde{n} \bmod N = n\}$, as the $(n+1)$-*th coset*, with $\tilde{n} \bmod N$ the remainder of the integer division $\tilde{n}/N$. In this paper, the *coset index* of the $(n+1)$-th coset is n. This procedure allows us to view the above uniform sampling as a multi-coset sampling [21] with N cosets. Consequently, the ULA of $\tilde{N}$ antennas in P1 can be regarded as N interleaved uniform linear subarrays (ULSs) [39] (which are the cosets) of L $(N\lambda/2)$-spaced antennas with λ the wavelength, whereas the $\tilde{N}$ time-domain samples in P2 can be considered as the output of a time-domain multi-coset sampler with L samples per coset. If we activate only the $(n+1)$-th coset, the spatial- or time-domain samples at index $\tilde{n}$ are given by

$$\bar{x}_{t,n}[\tilde{n}] = x_t[\tilde{n}] \sum_{l=0}^{L-1} \delta[\tilde{n} - (lN + n)], \; n = 0, 1, \ldots, N-1, \tag{5.2}$$

which can be collected into the $\tilde{N} \times 1$ vector $\bar{\mathbf{x}}_{t,n} = [\bar{x}_{t,n}[0], \bar{x}_{t,n}[1], \ldots, \bar{x}_{t,n}[\tilde{N}-1]]^T$. Observe that $\mathbf{x}_t = \sum_{n=0}^{N-1} \bar{\mathbf{x}}_{t,n}$. To show the relationship between the DTFT of $\bar{x}_{t,n}[\tilde{n}]$ and that of $x_t[\tilde{n}]$, we split $\vartheta \in [0, 1)$ into N equal-width bins and express the spectrum at the $(i+1)$-th bin ($i = 0, 1, \ldots, N-1$) as $X_{t,i}(\vartheta) = X_t\left(\vartheta + \frac{i}{N}\right)$ with ϑ now limited to $\vartheta \in [0, 1/N)$. As either the spatial or temporal sampling rate becomes $1/N$ times the Nyquist-rate when only the $(n+1)$-th coset is activated, the DTFT of $\bar{x}_{t,n}[\tilde{n}]$, denoted by $\bar{X}_{t,n}(\vartheta)$, is the sum of N aliased versions of $X_t(\vartheta)$ at N different bins. This is shown for $n = 0, 1, \ldots, N-1$ as [38]

$$\bar{X}_{t,n}(\vartheta) = \frac{1}{N} \sum_{i=0}^{N-1} X_{t,i}(\vartheta) e^{\frac{j2\pi ni}{N}}, \quad \vartheta \in [0, 1/N). \tag{5.3}$$

Collecting $\bar{X}_{t,n}(\vartheta)$, for $n = 0, 1, \ldots, N-1$, into the $N \times 1$ vector $\bar{\mathbf{x}}_t(\vartheta) = [\bar{X}_{t,0}(\vartheta), \bar{X}_{t,1}(\vartheta), \ldots, \bar{X}_{t,N-1}(\vartheta)]^T$ and introducing the $N \times 1$ vector $\mathbf{x}_t(\vartheta) = [X_{t,0}(\vartheta), X_{t,1}(\vartheta), \ldots, X_{t,N-1}(\vartheta)]^T$ allow us to write

$$\bar{\mathbf{x}}_t(\vartheta) = \mathbf{B}\mathbf{x}_t(\vartheta), \quad \vartheta \in [0, 1/N), \tag{5.4}$$

with the element of the $N \times N$ matrix $\mathbf{B}$ at the $(n+1)$-th row and the $(i+1)$-th column given by $[\mathbf{B}]_{n+1,i+1} = \frac{1}{N} e^{\frac{j2\pi ni}{N}}$.

We now assume the presence of K active users, consider the model in Fig. 5.1, and introduce the following definition (see also Fig. 5.1).

Definition 5.2.1. *We define the complex-valued zero-mean random processes $U_t^{(k)}(\vartheta)$ and $H_t^{(k)}(\vartheta)$ as*

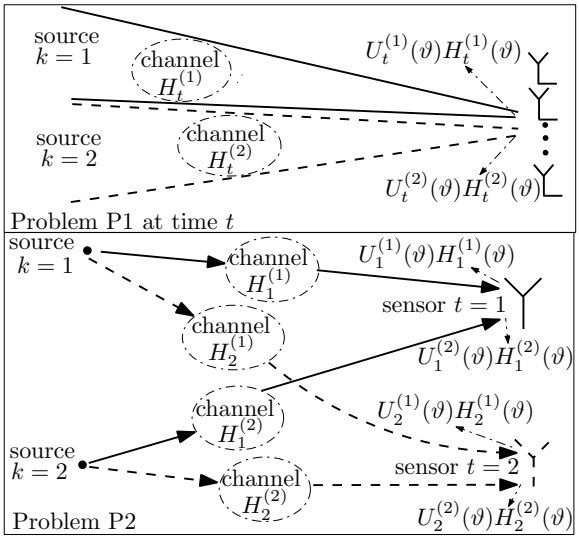


Figure 5.1: The system model for problems P1 and P2.

- *For P1, $U_t^{(k)}(\vartheta)$ is the source signal related to the k-th user received at time index t, which can depend on the DOA $\sin^{-1}(2\vartheta)$ due to scattering. For P2, it is the source signal related to the k-th user received at sensor t, which can vary with frequency ϑ due to power loading,*
- *$H_t^{(k)}(\vartheta)$ is the related channel response for the k-th user at time index t and DOA $\sin^{-1}(2\vartheta)$ (for P1) or at sensor t and frequency ϑ (for P2).*

Note from Fig. 5.1 that, theoretically, $U_t^{(k)}(\vartheta)$ is the only component observed by the ULA in P1 or by the sensors in P2 if no fading channel exists. Define $N_t(\vartheta)$ as the zero-mean additive white (both in ϑ and t) noise at DOA $\sin^{-1}(2\vartheta)$ and time index t (for P1) or at frequency ϑ and sensor t (for P2). By introducing $N_{t,i}(\vartheta) = N_t\left(\vartheta + \frac{i}{N}\right)$ and similarly also $H_{t,i}^{(k)}(\vartheta)$ as well as $U_{t,i}^{(k)}(\vartheta)$, we can then use Definition 5.2.1 to write $X_{t,i}(\vartheta)$ in (5.3) as

$$X_{t,i}(\vartheta) = \sum_{k=1}^{K} H_{t,i}^{(k)}(\vartheta)U_{t,i}^{(k)}(\vartheta) + N_{t,i}(\vartheta),\ \vartheta \in [0, 1/N). \tag{5.5}$$

Next, let us consider the following assumption.

Assumption 5.2.1. *$X_{t,i}(\vartheta)$ in* (5.5) *is an ergodic stochastic process along t.*

This ergodicity assumption requires that the statistics of $\mathbf{x}_t(\vartheta)$ in (5.4) do not change with t (a more general case is discussed in Section 5.6). Hence, we can

define the $N \times N$ correlation matrix of $\mathbf{x}_t(\vartheta)$ as $\mathbf{R}_x(\vartheta) = E[\mathbf{x}_t(\vartheta)\mathbf{x}_t^H(\vartheta)]$, for all t and $\vartheta \in [0, 1/N)$. The assumption that the statistics of $\mathbf{x}_t(\vartheta)$ do not vary with t is motivated for P1 when the signal received by the array is stationary in the time-domain. For P2, it implies that the statistics of the signal $\mathbf{x}_t$ received by different sensors t are the same. Observe from (5.5) that the element of $\mathbf{R}_x(\vartheta)$ at the $(i+1)$-th row and the $(i'+1)$-th column is given by

$$E[X_{t,i}(\vartheta)X_{t,i'}^*(\vartheta)] = E[|N_{t,i}(\vartheta)|^2]\delta[i-i'] + \sum_{k=1}^{K}\sum_{k'=1}^{K} E[U_{t,i}^{(k)}(\vartheta)U_{t,i'}^{(k')*}(\vartheta)]E[H_{t,i}^{(k)}(\vartheta)H_{t,i'}^{(k')*}(\vartheta)], \tag{5.6}$$

where we assume that the source signal $U_t^{(k)}(\vartheta)$, the noise $N_t(\vartheta)$, and the channel response $H_t^{(k)}(\vartheta)$ are mutually uncorrelated. We now consider the following remark.

Remark 5.2.1. *The diagonal of $\mathbf{R}_x(\vartheta)$, which is given by $\{E[|X_{t,i}(\vartheta)|^2]\}_{i=0}^{N-1}$ and which is independent of t, can be related to $P_x(\vartheta)$ in* (5.1)*. In practice, this expected value has to be estimated and Assumption 5.2.1 allows us to estimate $E[|X_{t,i}(\vartheta)|^2]$ using $\frac{1}{\tau}\sum_{t=1}^{\tau}|X_{t,i}(\vartheta)|^2$. We can then consider $\frac{1}{\tilde{N}\tau}\sum_{t=1}^{\tau}|X_{t,i}(\vartheta)|^2$ as a reasonable estimate for $P_x(\vartheta + \frac{i}{N})$ in* (5.1)*, for $\vartheta \in [0, 1/N)$. Here, $\frac{1}{\tilde{N}\tau}\sum_{t=1}^{\tau}|X_t(\vartheta)|^2$, for $\vartheta \in [0, 1)$, can be considered as the averaged periodogram (AP) of $x_t[\tilde{n}]$ over different time indices t in P1 or different sensors t in P2.*

Note that, even for the noiseless case, we can expect $X_{t,i}(\vartheta)$ in (5.5) to vary with t if either one (or both) of the following situations occurs.

- For P1, $U_t^{(k)}(\vartheta)$ varies with the time index t if the information that is being transmitted changes with time. For P2, it varies with the sensor index t where the signal is received if the sensors are not synchronized.
- For P1, $H_t^{(k)}(\vartheta)$ varies with the time index t if Doppler fading effects exist. For P2, it varies with the sensor index t where the signal is received, due to path loss, shadowing, and small-scale spatial fading effects.

We then consider the following remark.

Remark 5.2.2. *Recall that the size of the predefined bins in $\vartheta \in [0, 1)$ is a design parameter given by $\frac{1}{N}$, i.e., the inverse of the number of cosets. Using* (5.6)*, it is easy to find that $\mathbf{R}_x(\vartheta)$ is a diagonal matrix if either $E[U_t^{(k)}(\vartheta)U_t^{(k')*}(\vartheta')] = 0$ and/or $E[H_t^{(k)}(\vartheta)H_t^{(k')*}(\vartheta')] = 0$ for $|\vartheta' - \vartheta| \geq \frac{1}{N}$, with $\vartheta, \vartheta' \in [0, 1)$, and for all t, k, k'.*

One example for both P1 and P2 is when we have K non-overlapping active bands corresponding to K different users leading to a multiband structure in the ϑ-domain with either the K different users transmitting mutually uncorrelated source signals and/or the signals from the K different users passing through mutually uncorrelated wireless channels on their way to the receiver. If we denote the support of the k-th active band by $\mathcal{B}_k$ and its bandwidth by $\Lambda(\mathcal{B}_k) = \sup\{\mathcal{B}_k\} - \inf\{\mathcal{B}_k\}$, the condition in Remark 5.2.2 is then satisfied by setting N such that $\frac{1}{N} \geq \max_k \Lambda(\mathcal{B}_k)$. Note that such a choice is possible, especially for P2, as the channelization parameter for a communication network is usually known.

We focus on the case where $\mathbf{R}_x(\vartheta)$ is a diagonal matrix and define the so-called $N \times N$ *coset correlation matrix* as

$$\mathbf{R}_{\bar{x}}(\vartheta) = E[\bar{\mathbf{x}}_t(\vartheta)\bar{\mathbf{x}}_t^H(\vartheta)] = \mathbf{B}\mathbf{R}_x(\vartheta)\mathbf{B}^H, \ \ \vartheta \in [0, 1/N). \tag{5.7}$$

Observe that $\mathbf{R}_{\bar{x}}(\vartheta)$ is a circulant matrix when $\mathbf{R}_x(\vartheta)$ is a diagonal matrix since $\mathbf{B}$ is an inverse discrete Fourier transform (IDFT) matrix, as can be concluded from (5.4). Based on the aforementioned system model, we finally formulate our problem statement as follows:

Problem Statement: As an estimate of the spectral representation of the power $P_x(\vartheta)$ *in* (5.1) *(which is also the power spectrum when* $x[\tilde{n}]$ *in* (5.1) *is a WSS process), we aim to compressively reconstruct the AP of* $x_t[\tilde{n}]$ *in* (5.2) *over the index t, where we assume that* $x_t[\tilde{n}]$ *is ergodic along the index t and that its coset correlation matrix* $\mathbf{R}_{\bar{x}}(\vartheta)$ *has a circulant structure. We discuss the compression and the reconstruction in Section 5.3 and the estimation of the correlation matrix in Section 5.4.*

5.2.2 Interpretation of AP in Remark 5.2.1

How the AP in Remark 5.2.1 is interpreted with respect to $U_t^{(k)}(\vartheta)$ and $H_t^{(k)}(\vartheta)$ depends on which of the functions varies in t. For example, consider problem P2 and assume that only one user k can occupy a given frequency ϑ at a given time and that only $H_t^{(k)}(\vartheta)$ varies in t, i.e., $U_t^{(k)}(\vartheta) = U^{(k)}(\vartheta)$. For this example, we have from (5.5)

$$\begin{aligned}\frac{1}{\tilde{N}\tau}\sum_{t=1}^{\tau}|X_t(\vartheta)|^2 &= \frac{|U^{(k)}(\vartheta)|^2}{\tilde{N}}\sum_{t=1}^{\tau}\frac{|H_t^{(k)}(\vartheta)|^2}{\tau} + \sum_{t=1}^{\tau}\frac{|N_t(\vartheta)|^2}{\tilde{N}\tau} \\ &+ \sum_{t=1}^{\tau}\frac{2\mathrm{Re}(H_t^{(k)}(\vartheta)U^{(k)}(\vartheta)N_t^*(\vartheta))}{\tilde{N}\tau},\end{aligned} \tag{5.8}$$

where $\mathrm{Re}(x)$ gives the real component of x, the first term is the classical periodogram of the user signals $\frac{|U^{(k)}(\vartheta)|^2}{\tilde{N}}$ scaled by the averaged fading magnitude experienced at different channels $\frac{1}{\tau}\sum_{t=1}^{\tau}|H_t^{(k)}(\vartheta)|^2$, the second term is the AP of the noises at different sensors t, and the last term converges to zero as τ becomes larger due to the uncorrelatedness between the noise $N_t(\vartheta)$ and the channel response $H_t^{(k)}(\vartheta)$. The assumption that the statistics of $X_t(\vartheta)$ do not vary with t (as required by Assumption 5.2.1) implies that the statistics of the fading experienced by different sensors t are the same (e.g., they experience small-scale fading on top of the same path loss and shadowing).

As another example, consider problem P1 and assume that only one user k can occupy a given DOA $\sin^{-1}(2\vartheta)$ at a given time and that only $U_t^{(k)}(\vartheta)$ varies in t, i.e., $H_t^{(k)}(\vartheta) = H^{(k)}(\vartheta)$. For this example, we have from (5.5)

$$\frac{1}{\tilde{N}\tau}\sum_{t=1}^{\tau}|X_t(\vartheta)|^2 = |H^{(k)}(\vartheta)|^2\sum_{t=1}^{\tau}\frac{|U_t^{(k)}(\vartheta)|^2}{\tilde{N}\tau} + \sum_{t=1}^{\tau}\frac{|N_t(\vartheta)|^2}{\tilde{N}\tau} + \sum_{t=1}^{\tau}\frac{2\mathrm{Re}(U_t^{(k)}(\vartheta)H^{(k)}(\vartheta)N_t^*(\vartheta))}{\tilde{N}\tau}, \tag{5.9}$$

where the first term is the angular-domain AP of the user signals $\frac{1}{\tilde{N}\tau}\sum_{t=1}^{\tau}|U_t^{(k)}(\vartheta)|^2$ scaled by the magnitude of the time-invariant channel angular response $|H^{(k)}(\vartheta)|^2$, the second term is the angular-domain AP of the noise, and the last term again converges to zero as τ becomes larger due to the uncorrelatedness between $N_t(\vartheta)$ and $U_t^{(k)}(\vartheta)$.

5.3 Compression and Reconstruction

5.3.1 Spatial or Temporal Compression

As $\mathbf{R}_{\bar{x}}(\vartheta)$ in (5.7) is a circulant matrix, it is possible to condense its entries into an $N \times 1$ vector $\mathbf{r}_{\bar{x}}(\vartheta) = [r_{\bar{x}}(\vartheta,0), r_{\bar{x}}(\vartheta,1), \ldots, r_{\bar{x}}(\vartheta,N-1)]^T$ with $r_{\bar{x}}(\vartheta,(n-n') \bmod N) = E\left[\bar{X}_{t,n}(\vartheta), \bar{X}_{t,n'}^*(\vartheta)\right]$. We can then relate $\mathbf{r}_{\bar{x}}(\vartheta)$ to $\mathbf{R}_{\bar{x}}(\vartheta)$ as

$$\mathrm{vec}(\mathbf{R}_{\bar{x}}(\vartheta)) = \mathbf{T}\mathbf{r}_{\bar{x}}(\vartheta), \quad \vartheta \in [0, 1/N), \tag{5.10}$$

where $\mathbf{T}$ is an $N^2 \times N$ repetition matrix whose $(q+1)$-th row is given by the $\left(\left(q - \left\lfloor \frac{q}{N} \right\rfloor\right) \bmod N + 1\right)$-th row of the $N \times N$ identity matrix $\mathbf{I}_N$ and vec(.) is the operator that stacks all columns of a matrix into one column vector. The

possibility to condense the N^2 entries of $\mathbf{R}_{\bar{x}}(\vartheta)$ into the N entries of $\mathbf{r}_{\bar{x}}(\vartheta)$ facilitates compression by performing a spatial- or time-domain non-uniform periodic sampling (similar to [38]), in which only $M < N$ cosets are activated. Here, we use the set $\mathcal{M} = \{n_0, n_1, \ldots, n_{M-1}\}$, with $0 \leq n_0 < n_1 < \cdots < n_{M-1} \leq N-1$, to indicate the indices of the M active cosets. All values of $\bar{x}_{t,n}[\tilde{n}]$ in (5.2) are then collected and their corresponding DTFT $\bar{X}_{t,n}(\vartheta)$ in (5.3) is computed for all $n \in \mathcal{M}$. Stacking $\left\{\bar{X}_{t,n}(\vartheta)\right\}_{n\in\mathcal{M}}$ into the $M \times 1$ vector $\bar{\mathbf{y}}_t(\vartheta) = [\bar{X}_{t,n_0}(\vartheta), \bar{X}_{t,n_1}(\vartheta), \ldots, \bar{X}_{t,n_{M-1}}(\vartheta)]^T$ allows us to relate $\bar{\mathbf{y}}_t(\vartheta)$ to $\bar{\mathbf{x}}_t(\vartheta)$ in (5.4) as

$$\bar{\mathbf{y}}_t(\vartheta) = \mathbf{C}\bar{\mathbf{x}}_t(\vartheta), \quad \vartheta \in [0, 1/N), \tag{5.11}$$

where $\mathbf{C}$ is an $M \times N$ selection matrix whose rows are selected from the rows of $\mathbf{I}_N$ based on $\mathcal{M}$. Since $\mathbf{C}$ is real, the $M \times M$ correlation matrix of $\bar{\mathbf{y}}_t(\vartheta)$, for $\vartheta \in [0, 1/N)$, can be written as

$$\mathbf{R}_{\bar{y}}(\vartheta) = E[\bar{\mathbf{y}}_t(\vartheta)\bar{\mathbf{y}}_t^H(\vartheta)] = \mathbf{C}\mathbf{R}_{\bar{x}}(\vartheta)\mathbf{C}^T. \tag{5.12}$$

We then take (5.10) into account, cascade all columns of $\mathbf{R}_{\bar{y}}(\vartheta)$ into a column vector $\text{vec}(\mathbf{R}_{\bar{y}}(\vartheta))$, and write

$$\text{vec}(\mathbf{R}_{\bar{y}}(\vartheta)) = \mathbf{R}_c\mathbf{r}_{\bar{x}}(\vartheta), \quad \vartheta \in [0, 1/N), \tag{5.13}$$

where $\mathbf{R}_c = (\mathbf{C} \otimes \mathbf{C})\mathbf{T}$ is a real $M^2 \times N$ matrix and $\otimes$ denotes the Kronecker product operation.

5.3.2 Reconstruction

If $\mathbf{R}_c$ in (5.13) is a tall matrix ($M^2 \geq N$), which is possible despite $M < N$, and if it has full column rank, $\mathbf{r}_{\bar{x}}(\vartheta)$ in (5.13) can be reconstructed from $\text{vec}(\mathbf{R}_{\bar{y}}(\vartheta))$ using LS for all $\vartheta \in [0, 1/N)$. In addition, as long as the identifiability of $\mathbf{r}_{\bar{x}}(\vartheta)$ in (5.13) is preserved, we can also consider estimators other than LS (such as in [99]). To formulate a necessary and sufficient condition for the identifiability of $\mathbf{r}_{\bar{x}}(\vartheta)$ in (5.13) from $\text{vec}(\mathbf{R}_{\bar{y}}(\vartheta))$, let us review the concept of a circular sparse ruler defined in [100].

Definition 5.3.1. *A circular sparse ruler of length $N-1$ is a set $\mathcal{K} \subset \{0, 1, \ldots, N-1\}$ for which $\Omega(\mathcal{K}) = \{(\kappa - \kappa') \bmod N | \forall \kappa, \kappa' \in \mathcal{K}\} = \{0, 1, \ldots, N-1\}$. We call it minimal if there is no other circular sparse ruler of length $N-1$ with fewer elements.*

Detailed information about circular sparse rulers can be found in [100]. We can then use this concept to formulate the following theorem whose proof is available in [101].

Theorem 5.3.1. *$\mathbf{r}_{\bar{x}}(\vartheta)$ in (5.13) is identifiable from vec$(\mathbf{R}_{\bar{y}}(\vartheta))$, i.e., $\mathbf{R}_c$ has full column rank, if and only if $\mathcal{M}$ is a circular sparse ruler, i.e., $\Omega(\mathcal{M}) = \{0, 1, \ldots, N-1\}$. When this is satisfied, $\mathbf{R}_c$ contains all rows of $\mathbf{I}_N$.*

Our goal is to obtain the strongest possible compression rate M/N preserving the identifiability. This is achieved by minimizing the cardinality of the set $\mathcal{M}$, $|\mathcal{M}| = M$, under the condition that $\Omega(\mathcal{M}) = \{0, 1, \ldots, N-1\}$. This leads to a length-$(N-1)$ minimal circular sparse ruler problem, which can be written as

$$\min_{\mathcal{M}} |\mathcal{M}| \text{ s.t. } \Omega(\mathcal{M}) = \{0, 1, \ldots, N-1\}. \tag{5.14}$$

Solving (5.14) minimizes the compression rate M/N while maintaining the identifiability of $\mathbf{r}_{\bar{x}}(\vartheta)$ in (5.13).

Recall that, for P1, $\mathcal{M}$ indicates the indices of the $M < N$ active ULSs in our ULA, which will be referred to as the *underlying* array. Therefore, we have a periodic non-ULA of active antennas and $\mathcal{M}$ governs the location of the active antennas in each spatial period. When $\mathcal{M}$ is a solution of the minimal length-$(N-1)$ circular sparse ruler problem in (5.14), we can label the resulting non-ULA of active antennas as a *periodic circular MRA* and each of its spatial periods as a *circular MRA*. Similarly for P2, we can label the non-uniform sampling in each temporal period as *minimal circular sparse ruler sampling* and the entire periodic non-uniform sampling as *periodic minimal circular sparse ruler sampling* if the indices of the $M < N$ active cosets are given by the solution of (5.14).

Once $\mathbf{r}_{\bar{x}}(\vartheta)$ is reconstructed from vec$(\mathbf{R}_{\bar{y}}(\vartheta))$ in (5.13) using LS for $\vartheta \in [0, 1/N)$, we can use (5.10) to compute $\mathbf{R}_{\bar{x}}(\vartheta)$ from $\mathbf{r}_{\bar{x}}(\vartheta)$ and (5.7) to compute $\mathbf{R}_x(\vartheta)$ from $\mathbf{R}_{\bar{x}}(\vartheta)$ as $\mathbf{R}_x(\vartheta) = N^2\mathbf{B}^H\mathbf{R}_{\bar{x}}(\vartheta)\mathbf{B}$. As we have diag$(\mathbf{R}_x(\vartheta)) = [E[|X_{t,0}(\vartheta)|^2], E[|X_{t,1}(\vartheta)|^2], \ldots, E[|X_{t,N-1}(\vartheta)|^2]]^T$ with $\vartheta \in [0, 1/N)$, reconstructing diag$(\mathbf{R}_x(\vartheta))$ for all $\vartheta \in [0, 1/N)$ gives $E[|X_t(\vartheta)|^2]$ for all $\vartheta \in [0, 1)$.

5.4 Correlation Matrix Estimation

In practice, the expectation in (5.12) must be approximated. Here, we propose to approximate the expectation in (5.12) with the sample average over different time

indices t for P1 or sensors indices t for P2, i.e.,

$$\hat{\mathbf{R}}_{\bar{y}}(\vartheta) = \frac{1}{\tau}\sum_{t=1}^{\tau} \bar{\mathbf{y}}_t(\vartheta)\bar{\mathbf{y}}_t^H(\vartheta), \;\; \vartheta \in [0, 1/N), \tag{5.15}$$

where we recall that τ is either the total number of time indices or sensors from which the observations are collected. Observe that the $M \times M$ matrix $\hat{\mathbf{R}}_{\bar{y}}(\vartheta)$ is an unbiased estimate of $\mathbf{R}_{\bar{y}}(\vartheta)$ in (5.13). It is also a consistent estimate if Assumption 5.2.1 holds. We can then apply LS reconstruction on $\hat{\mathbf{R}}_{\bar{y}}(\vartheta)$ in (5.15) instead of $\mathbf{R}_{\bar{y}}(\vartheta)$ in (5.13). As a result, the procedure to compressively reconstruct the AP of $x_t[\tilde{n}]$ in (5.2) over the index t can be listed as

1. For $t = 1, 2, \ldots, \tau$, collect all values of $\bar{x}_{t,n}[\tilde{n}]$ in (5.2) and compute their corresponding DTFT $\bar{X}_{t,n}(\vartheta)$ in (5.3) for all $n \in \mathcal{M}$. We use them to form $\bar{\mathbf{y}}_t(\vartheta)$ in (5.11).

2. Compute $\hat{\mathbf{R}}_{\bar{y}}(\vartheta)$, for $\vartheta \in [0, 1/N)$, using (5.15).

3. Based on (5.13) and for $\vartheta \in [0, 1/N)$, we apply LS reconstruction on $\hat{\mathbf{R}}_{\bar{y}}(\vartheta)$ leading to
$$\hat{\mathbf{r}}_{\bar{x},LS}(\vartheta) = (\mathbf{R}_c^T\mathbf{R}_c)^{-1}\mathbf{R}_c^T \text{vec}(\hat{\mathbf{R}}_{\bar{y}}(\vartheta)). \tag{5.16}$$

4. Based on (5.10) and (5.7), for $\vartheta \in [0, 1/N)$, we compute $\text{vec}(\hat{\mathbf{R}}_{\bar{x},LS}(\vartheta)) = \mathbf{T}\hat{\mathbf{r}}_{\bar{x},LS}(\vartheta)$ and
$$\hat{\mathbf{R}}_{x,LS}(\vartheta) = N^2\mathbf{B}^H\hat{\mathbf{R}}_{\bar{x},LS}(\vartheta)\mathbf{B}. \tag{5.17}$$

5. Note that the $(i+1)$-th diagonal element of $\hat{\mathbf{R}}_{x,LS}(\vartheta)$, i.e., $[\text{diag}(\hat{\mathbf{R}}_{x,LS}(\vartheta))]_{i+1}$ is the LS estimate of the $(i+1)$-th diagonal element of $\mathbf{R}_x(\vartheta)$, which according to Remark 5.2.1 is given by $E[|X_{t,i}(\vartheta)|^2]$. Based on the definition of AP in Remark 5.2.1 and considering (5.15), we can then formulate the compressive AP (CAP) of $x_t[\tilde{n}]$ in (5.2) over the index t as
$$\hat{P}_{x,LS}(\vartheta + \frac{i}{N}) = \frac{1}{\tilde{N}}[\text{diag}(\hat{\mathbf{R}}_{x,LS}(\vartheta))]_{i+1}, \tag{5.18}$$
for $\vartheta \in [0, 1/N)$ and $i = 0, 1, \ldots, N-1$.

Note that, when reconstructing the CAP $\hat{P}_{x,LS}(\vartheta)$ in (5.18), we introduce additional errors with respect to the AP $\frac{1}{\tilde{N}\tau}\sum_{t=1}^{\tau}|X_t(\vartheta)|^2$ in Remark 5.2.1 (including the ones in (5.8) and (5.9)). This error emerges during the compression and the LS operation in (5.16). This issue will be discussed up to some extent in the next section.

5.5 Performance Analysis

5.5.1 Bias Analysis

The bias analysis of the CAP $\hat{P}_{x,LS}(\vartheta)$ in (5.18) with respect to $P_x(\vartheta)$ in (5.1) is given by the following theorem whose proof is available in Appendix 5.A.

Theorem 5.5.1. *For $\vartheta \in [0,1)$, the CAP $\hat{P}_{x,LS}(\vartheta)$ in* (5.18) *is an asymptotically (with respect to $\tilde{N}$) unbiased estimate of $P_x(\vartheta)$ in* (5.1)*.*

5.5.2 Variance Analysis

We start by recalling that the $(m+1)$-th element of $\bar{\mathbf{y}}_t(\vartheta)$ in (5.11) is given by $\bar{X}_{t,n_m}(\vartheta)$. By using (5.3), we can write the element of $\hat{\mathbf{R}}_{\bar{y}}(\vartheta)$ in (5.15) at the $(m+1)$-th row and the $(m'+1)$-th column, for $m, m' = 0, 1, \ldots, M-1$, as

$$[\hat{\mathbf{R}}_{\bar{y}}(\vartheta)]_{m+1,m'+1} = \frac{1}{N^2\tau} \sum_{t=1}^{\tau} \sum_{i=0}^{N-1} \sum_{i'=0}^{N-1} X_{t,i}(\vartheta) X^*_{t,i'}(\vartheta) e^{\frac{j2\pi(n_m i - n_{m'} i')}{N}}. \quad (5.19)$$

We continue to evaluate the covariance between the elements of $\hat{\mathbf{R}}_{\bar{y}}(\vartheta)$ in (5.19), which is not trivial for a general signal $x_t[\tilde{n}]$ in (5.2), as it involves the computation of fourth order moments. To get a useful insight, let us consider the case when the distribution of $x_t[\tilde{n}]$ in (5.2) (and thus also $X_{t,i}(\vartheta)$ in (5.19)) is jointly Gaussian. In this case, the fourth order moment computation is simplified by using the results in [71]: If x_1, x_2, x_3, and x_4 are jointly (real or complex) Gaussian random variables, we have $E[x_1x_2x_3x_4] = E[x_1x_2]E[x_3x_4] + E[x_1x_3]E[x_2x_4] + E[x_1x_4]E[x_2x_3] - 2E[x_1]E[x_2]E[x_3]E[x_4]$. Using this result, the covariance between the elements of $\hat{\mathbf{R}}_{\bar{y}}(\vartheta)$ in (5.19), when $x_t[\tilde{n}]$ in (5.2) is jointly Gaussian, can be shown to be

$$\begin{aligned}
&\mathrm{Cov}[[\hat{\mathbf{R}}_{\bar{y}}(\vartheta)]_{m+1,m'+1}, [\hat{\mathbf{R}}_{\bar{y}}(\vartheta)]_{a+1,a'+1}] = \frac{1}{N^4\tau^2} \sum_{t=1}^{\tau} \sum_{t'=1}^{\tau} \sum_{i=0}^{N-1} \sum_{i'=0}^{N-1} \\
&\sum_{b=0}^{N-1} \sum_{b'=0}^{N-1} e^{\frac{j2\pi(n_m i - n_{m'} i' - n_a b + n_{a'} b')}{N}} \left\{ E[X_{t,i}(\vartheta) X^*_{t',b}(\vartheta)] E[X^*_{t,i'}(\vartheta) X_{t',b'}(\vartheta)] + \right. \\
&\left. E[X_{t,i}(\vartheta) X_{t',b'}(\vartheta)] E[X^*_{t,i'}(\vartheta) X^*_{t',b}(\vartheta)] \right\}, \quad (5.20)
\end{aligned}$$

for $\vartheta \in [0, 1/N)$ and $m, m', a, a' = 0, 1, \ldots, M-1$, where we also assume that $x_t[\tilde{n}]$ in (5.2) has zero mean (see Definition 5.2.1).

Under the above assumptions, we introduce the $M^2 \times M^2$ covariance matrix $\mathbf{\Sigma}_{\hat{R}_{\bar{y}}}(\vartheta) = E[\text{vec}(\hat{\mathbf{R}}_{\bar{y}}(\vartheta))\text{vec}(\hat{\mathbf{R}}_{\bar{y}}(\vartheta))^H] - E[\text{vec}(\hat{\mathbf{R}}_{\bar{y}}(\vartheta))]E[\text{vec}(\hat{\mathbf{R}}_{\bar{y}}(\vartheta))^H]$, whose entry at the $(Mm'+m+1)$-th row and the $(Ma'+a+1)$-th column is given by $\text{Cov}[[\hat{\mathbf{R}}_{\bar{y}}(\vartheta)]_{m+1,m'+1}, [\hat{\mathbf{R}}_{\bar{y}}(\vartheta)]_{a+1,a'+1}]$ in (5.20). By recalling that $\mathbf{R}_c$ and $\mathbf{T}$ are real matrices, we can then compute the $N \times N$ covariance matrix of $\hat{\mathbf{r}}_{\bar{x},LS}(\vartheta)$ in (5.16) as

$$\mathbf{\Sigma}_{\hat{r}_{\bar{x},LS}}(\vartheta) = (\mathbf{R}_c^T\mathbf{R}_c)^{-1}\mathbf{R}_c^T\mathbf{\Sigma}_{\hat{R}_{\bar{y}}}(\vartheta)\mathbf{R}_c(\mathbf{R}_c^T\mathbf{R}_c)^{-1}, \tag{5.21}$$

and use (5.17) to introduce $\mathbf{\Sigma}_{\hat{R}_{x,LS}}(\vartheta)$ as the $N^2 \times N^2$ covariance matrix of $\text{vec}(\hat{\mathbf{R}}_{x,LS}(\vartheta))$, which can be written as

$$\mathbf{\Sigma}_{\hat{R}_{x,LS}}(\vartheta) = N^4(\mathbf{B}^T \otimes \mathbf{B}^H)\mathbf{T}\mathbf{\Sigma}_{\hat{r}_{\bar{x},LS}}(\vartheta)\mathbf{T}^T(\mathbf{B}^* \otimes \mathbf{B}), \tag{5.22}$$

for $\vartheta \in [0, 1/N)$. Recall from (5.18) that the CAP $\hat{P}_{x,LS}(\vartheta + \frac{i}{N})$, for $\vartheta \in [0, 1/N)$ and $i = 0, 1, \ldots, N-1$, is given by $\frac{1}{N}[\hat{\mathbf{R}}_{x,LS}(\vartheta)]_{i+1,i+1}$. It is then trivial to show that the variance of $\hat{P}_{x,LS}(\vartheta + \frac{i}{N})$ is given by

$$\text{Var}[\hat{P}_{x,LS}(\vartheta + \frac{i}{N})] = \frac{1}{\tilde{N}^2}[\mathbf{\Sigma}_{\hat{R}_{x,LS}}(\vartheta)]_{Ni+i+1,Ni+i+1}, \tag{5.23}$$

for $\vartheta \in [0, 1/N)$ and $i = 0, 1, \ldots, N-1$.

To get even more insight into this result, we consider a specific case in the next proposition whose proof is provided in Appendix 5.B.

Proposition 5.5.1. *When $x_t[\tilde{n}]$ in* (5.2) *contains only circular complex zero-mean Gaussian i.i.d. noise with variance σ^2, the covariance between the elements of $\hat{\mathbf{R}}_{\bar{y}}(\vartheta)$ in* (5.19)*, for $\vartheta \in [0, 1/N)$, is given by*

$$Cov[[\hat{\mathbf{R}}_{\bar{y}}(\vartheta)]_{m+1,m'+1}, [\hat{\mathbf{R}}_{\bar{y}}(\vartheta)]_{a+1,a'+1}] = \frac{L^2\sigma^4}{\tau}\delta[m-a]\delta[m'-a'], \tag{5.24}$$

for $m, m', a, a' = 0, 1, \ldots, M-1$.

It is clear from (5.24) that $\mathbf{\Sigma}_{\hat{R}_{\bar{y}}}(\vartheta)$ in (5.21) is then a diagonal matrix and we can find from (5.21)-(5.23) that $\text{Var}[\hat{P}_{x,LS}(\vartheta)] \propto \sigma^4$ or $\text{Var}[\hat{P}_{x,LS}(\vartheta)] \propto P_x^2(\vartheta)$. This observation can be related to a similar result found for the conventional periodogram estimate of white Gaussian noise sampled at Nyquist rate in [102].

5.5.3 Effect of the Compression Rate on the Variance

In this section, we focus on the impact of the compression rate M/N on the variance analysis by first defining an $N \times 1$ vector $\mathbf{w} = [w[0], w[1], \ldots, w[N-1]]^T$ containing binary entries, with $w[n] = 1$ if $n \in \mathcal{M}$ (i.e., the coset with index n is one of the M activated cosets) and $w[n] = 0$ if $n \notin \mathcal{M}$. In other words, the entries of $\mathbf{w}$ indicate which M out of the N cosets are activated. Let us then focus on (5.16) and consider the following remark.

Remark 5.5.1. *The same argument that leads to Theorem 5.3.1 (see Lemma 1 in [101]) shows that the rows of $\mathbf{R}_c$ are given by the $((g-f) \bmod N+1)$-th rows of $\mathbf{I}_N$, for all $f, g \in \mathcal{M}$. As a result, $\mathbf{R}_c^T\mathbf{R}_c$ is an $N \times N$ diagonal matrix. Denote the value of the κ-th diagonal element of $\mathbf{R}_c^T\mathbf{R}_c$ as γ_κ. We can then show that γ_κ is given by*

$$\gamma_\kappa = \sum_{n=0}^{N-1} w[(n+\kappa-1) \bmod N]w[n], \quad \kappa = 1, 2, \ldots, N. \tag{5.25}$$

The proof of (5.25) *is available in Appendix 5.C. Using* (5.25)*, we can also show that γ_κ gives the number of times the κ-th row of $\mathbf{I}_N$ appears in $\mathbf{R}_c$, i.e., the number of pairs (g, f) that lead to $(g-f) \bmod N+1 = \kappa$. As we have $|\mathcal{M}| = M$, we can find that $\sum_{\kappa=1}^{N} \gamma_\kappa = M^2$ and $\gamma_1 = M$.*

Using Remark 5.5.1, we then formulate the following theorem whose proof is available in Appendix 5.D.

Theorem 5.5.2. *When $x_t[\tilde{n}]$ in* (5.2) *contains only circular complex zero-mean Gaussian i.i.d. noise with variance σ^2, the variance of the CAP $\hat{P}_{x,LS}(\vartheta + \frac{i}{N})$ in* (5.23)*, for $\vartheta \in [0, 1/N)$ and $i = 0, 1, \ldots, N-1$, is given by*

$$Var[\hat{P}_{x,LS}(\vartheta + \frac{i}{N})] = \frac{\sigma^4}{M\tau} + \frac{\sigma^4}{\tau} \sum_{n=1}^{N-1} \frac{1}{\gamma_{n+1}}. \tag{5.26}$$

Note how (5.26) relates M and N to $\text{Var}[\hat{P}_{x,LS}(\vartheta)]$ for circular complex zero-mean Gaussian i.i.d. noise and $\vartheta \in [0, 1)$. Recalling from Remark 5.5.1 that $\sum_{n=1}^{N-1} \gamma_{n+1} = M^2 - M$, we can find that, for a given N, a stronger compression rate (smaller M/N) tends to lead to a larger $\text{Var}[\hat{P}_{x,LS}(\vartheta)]$. Based on (5.25) and (5.26), it is of interest to find the binary values of $\{w[n]\}_{n=0}^{N-1}$ (or equivalently the cosets $n_m \in \mathcal{M}$) that minimize $\text{Var}[\hat{P}_{x,LS}(\vartheta)]$ for a given M. This will generally lead to a non-convex optimization problem, which is difficult to solve, although

it is clear that the solution will force the values of $\{\gamma_{n+1}\}_{n=1}^{N-1}$ to be as equal as possible. Alternatively, we can also put a constraint on $\text{Var}[\hat{P}_{x,LS}(\vartheta + \frac{i}{N})]$ in (5.26) and find the binary values $\{w[n]\}_{n=0}^{N-1}$ that minimize the compression rate M/N. This however, will again lead to a non-convex optimization problem that is difficult to solve. Note that, although finding $\mathbf{w}$ that minimizes M/N for a given $\text{Var}[\hat{P}_{x,LS}(\vartheta)]$ in (5.26) or the one that minimizes $\text{Var}[\hat{P}_{x,LS}(\vartheta)]$ for a given M/N is not trivial, the solution will always have to satisfy the identifiability condition in Theorem 5.3.1. This is because we can show that if the identifiability condition is not satisfied, some γ_n in (5.26) will be zero and thus $\text{Var}[\hat{P}_{x,LS}(\vartheta)]$ in (5.26) will have an infinite value.

The analysis of the effect of M/N on $\text{Var}[\hat{P}_{x,LS}(\vartheta)]$ for a general Gaussian signal $x_t[\tilde{n}]$, however, is difficult since it is clear from (5.20) that $\text{Var}[\hat{P}_{x,LS}(\vartheta)]$ for this case depends on the unknown statistics of $x_t[\tilde{n}]$. This is also true for a more general signal.

5.5.4 Asymptotic Performance Analysis

We now discuss the asymptotic behaviour of the performance of the CAP $\hat{P}_{x,LS}(\vartheta)$. We start by noting that Assumption 5.2.1 ensures that $\hat{\mathbf{R}}_{\bar{y}}(\vartheta)$ in (5.15) is a consistent estimate of $\mathbf{R}_{\bar{y}}(\vartheta)$ in (5.13) i.e., $\hat{\mathbf{R}}_{\bar{y}}(\vartheta)$ converges to $\mathbf{R}_{\bar{y}}(\vartheta)$ as τ approaches ∞. As it is clear from (5.16) and (5.17) that $\hat{\mathbf{R}}_{x,LS}(\vartheta)$ is linearly related to $\hat{\mathbf{R}}_{\bar{y}}(\vartheta)$, it is easy to show that $\hat{\mathbf{R}}_{x,LS}(\vartheta)$ converges to $\mathbf{R}_x(\vartheta)$ in (5.7) as τ approaches ∞. This implies that the CAP $\hat{P}_{x,LS}(\vartheta + \frac{i}{N})$ in (5.18) also converges to $\frac{1}{\tilde{N}}[\text{diag}(\mathbf{R}_x(\vartheta))]_{i+1} = \frac{1}{\tilde{N}}E[|X_t(\vartheta+\frac{i}{N})|^2]$, for $\vartheta \in [0, 1/N)$ and $i = 0, 1, \ldots, N-1$, as τ approaches ∞. Since $x_t[\tilde{n}]$ in (5.2) is an observation of the true process $x[\tilde{n}]$ in (5.1), $\hat{P}_{x,LS}(\vartheta)$ will converge to $P_x(\vartheta)$ in (5.1) if both τ and $\tilde{N}$ (or L for a fixed N) approach ∞.

5.5.5 Complexity Analysis

Let us now compare the complexity of our CAP approach with an existing state-of-the-art approach to tackle similar problems. We compare our CAP approach with a method that reconstructs $X_t(\vartheta)$ (instead of the periodogram), for $\vartheta \in [0, 1)$ and all $t = 1, 2, \ldots, \tau$, from compressive measurements. The reconstruction of $\{X_t(\vartheta)\}_{t=1}^{\tau}$, for $\vartheta \in [0, 1)$, is performed by reconstructing $\{\mathbf{x}_t(\vartheta)\}_{t=1}^{\tau}$ in (5.4) from $\{\bar{\mathbf{y}}_t(\vartheta)\}_{t=1}^{\tau}$ in (5.11), for $\vartheta \in [0, 1/N)$, using the Regularized M-FOCUSS (RM-FOCUSS) approach of [1]. We then use the reconstructed $\{\mathbf{x}_t(\vartheta)\}_{t=1}^{\tau}$, for $\vartheta \in [0, 1/N)$, either to compute the periodogram or to compute the energy at $\vartheta \in$

$[0, 1)$ and to detect the existence of active user signals. Note that RM-FOCUSS is designed to treat $\{\bar{\mathbf{y}}_t(\vartheta)\}_{t=1}^{\tau}$, for each ϑ, as multiple measurement vectors (MMVs) and exploit the assumed joint sparsity structure in $\{\mathbf{x}_t(\vartheta)\}_{t=1}^{\tau}$.

Table 5.1 summarizes the computational complexity of CAP and RM-FOCUSS (see [1] for more details). Note that Table 5.1 only describes the computational complexity of RM-FOCUSS for a single iteration. The number of RM-FOCUSS iterations depends on the convergence criterion parameter (labeled as δ in [1]). Hence, we can argue that our CAP approach is simpler than RM-FOCUSS. Moreover, in RM-FOCUSS, we also need to determine a proper regularization parameter (labeled as λ in [1]), which is generally not a trivial task. Note that we also compare the detection performance of the two methods in the sixth experiment of Section 5.8.1. Note that the reconstruction of $\mathbf{x}_t(\vartheta)$ from $\bar{\mathbf{y}}_t(\vartheta)$ is also considered in [38] but it only considers the single-sensor case.

5.6 Multi-cluster Scenario

Recall that the ergodicity assumption on $\mathbf{x}_t(\vartheta)$ in Assumption 5.2.1 requires the statistics of $\mathbf{x}_t(\vartheta)$ to be the same along index t. Let us now consider the case where we have D clusters of τ time indices in P1 or of τ sensors in P2 such that $\mathbf{x}_t(\vartheta)$ is ergodic and its statistics do not change only along index t within a cluster. We can then consider Assumption 5.2.1 and the resulting case considered in Sections 5.2-5.5 as a special case of this multi-cluster scenario with $D = 1$. We introduce the correlation matrix of $\mathbf{x}_t(\vartheta)$ and $\bar{\mathbf{y}}_t(\vartheta)$ for all indices t belonging to cluster d as $\mathbf{R}_{x,d}(\vartheta)$ and $\mathbf{R}_{\bar{y},d}(\vartheta)$, respectively, with $d = 0, 1, \ldots, D-1$. We can then repeat all the steps of Sections 5.2-5.5 for each cluster. More precisely, we can follow (5.15) and define the estimate of $\mathbf{R}_{\bar{y},d}(\vartheta)$ as $\hat{\mathbf{R}}_{\bar{y},d}(\vartheta)$, which is computed by averaging the outer-product of $\bar{\mathbf{y}}_t(\vartheta)$ over indices t belonging to cluster d. Then, we apply (5.16)-(5.18) on $\hat{\mathbf{R}}_{\bar{y},d}(\vartheta)$ to obtain $\hat{\mathbf{R}}_{x,LS,d}(\vartheta)$ and the CAP for cluster d, i.e., $\hat{P}_{x,LS,d}(\vartheta)$. Also note that the bias and variance analysis in Section 5.5 is also valid for each cluster in this section.

We might then be interested in the averaged statistics over the clusters, i.e., $\sum_{d=0}^{D-1} \frac{\mathbf{R}_{x,d}(\vartheta)}{D}$. Since $\sum_{d=0}^{D-1} \frac{\hat{\mathbf{R}}_{\bar{y},d}(\vartheta)}{D}$ is a consistent estimate of $\sum_{d=0}^{D-1} \frac{\mathbf{R}_{\bar{y},d}(\vartheta)}{D}$, we can then consider the resulting $\sum_{d=0}^{D-1} \frac{\hat{\mathbf{R}}_{x,LS,d}(\vartheta)}{D}$ as a valid LS estimate of $\sum_{d=0}^{D-1} \frac{\mathbf{R}_{x,d}(\vartheta)}{D}$. Defining the theoretical spectral representation of the power at cluster d as $P_{x,d}(\vartheta)$, we can then apply Theorem 5.5.1 for each cluster to conclude that $\frac{1}{D}\sum_{d=0}^{D-1} \hat{P}_{x,LS,d}(\vartheta)$ is an asymptotically (with respect to $\tilde{N}$) unbiased esti-

mate of $\frac{1}{D}\sum_{d=0}^{D-1} P_{x,d}(\vartheta)$. This multi-cluster scenario is of interest for P2 when we have clusters of wireless sensors sensing user signals where the signal from each user experiences the same fading statistics (the same path loss and shadowing) on its way towards the sensors belonging to the same cluster. However, the fading statistics experienced by the signal between the user location and different clusters are not the same. For P1, the multi-cluster scenario implies that the array sensing time can be grouped into multiple clusters of time indices where the signal statistics do not vary along the time within the cluster but they vary across different clusters.

5.7 Correlated Bins

When the bin size is reduced by increasing N in (5.3), the received spectra at two frequencies or angles, which are separated by more than the size of the bin, might still be correlated. In this case, $\mathbf{R}_x(\vartheta)$ and $\mathbf{R}_{\bar{x}}(\vartheta)$ in (5.7) are respectively not a diagonal and circulant matrix anymore, and the temporal and spatial compression of Section 5.3.1 cannot be performed without violating the identifiability of $\mathbf{r}_{\bar{x}}(\vartheta)$ in (5.13). This section proposes a solution when this situation occurs under Assumption 5.2.1 and the single-cluster scenario (it does not apply to the multi-cluser scenario of Section 5.6). Let us organize τ indices t into several groups and write t as $t = pZ + z + 1$ with $p = 0, 1, \ldots, P-1$ and $z = 0, 1, \ldots, Z-1$, where Z and P represent the total number of groups and the number of indices belonging to a group, respectively. Writing $\bar{\mathbf{y}}_t(\vartheta)$ and $\bar{\mathbf{x}}_t(\vartheta)$ at $t = pZ + z + 1$ as $\bar{\mathbf{y}}_{p,z}(\vartheta)$ and $\bar{\mathbf{x}}_{p,z}(\vartheta)$, we can introduce for each z a compression similar to (5.11) as

$$\bar{\mathbf{y}}_{p,z}(\vartheta) = \mathbf{C}_z \bar{\mathbf{x}}_{p,z}(\vartheta), \quad \vartheta \in [0, 1/N), \tag{5.27}$$

where $\mathbf{C}_z$ is the $M \times N$ selection matrix for the z-th group of indices whose rows are also selected from the rows of $\mathbf{I}_N$. Next, we compute the correlation matrix of $\bar{\mathbf{y}}_{p,z}(\vartheta)$ in (5.27), i.e., $\mathbf{R}_{\bar{y}_z}(\vartheta) = E[\bar{\mathbf{y}}_{p,z}(\vartheta)\bar{\mathbf{y}}_{p,z}^H(\vartheta)]$, for $z = 0, 1, \ldots, Z-1$, as

$$\mathbf{R}_{\bar{y}_z}(\vartheta) = \mathbf{C}_z E[\bar{\mathbf{x}}_{p,z}(\vartheta)\bar{\mathbf{x}}_{p,z}^H(\vartheta)]\mathbf{C}_z^T = \mathbf{C}_z \mathbf{R}_{\bar{x}}(\vartheta)\mathbf{C}_z^T, \tag{5.28}$$

with $\mathbf{R}_{\bar{x}}(\vartheta) = E[\bar{\mathbf{x}}_{p,z}(\vartheta)\bar{\mathbf{x}}_{p,z}^H(\vartheta)]$, for all p, z, as Assumption 5.2.1 requires that the statistics of $\bar{\mathbf{x}}_t(\vartheta)$ do not vary with t.

Let us interpret the above model for problems P1 and P2. For P1, (5.27) implies that we split the array scanning time τ into P scanning periods, each of which consists of Z time slots. It is clear from (5.27) that, in different time slots per scanning

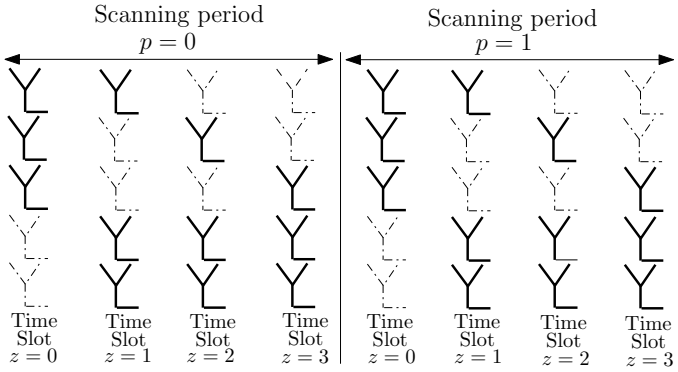


Figure 5.2: The DLA model used in problem P1 when the bins are correlated with $M = 3$, $N = 5$, $P = 2$, and $Z = 4$. Solid lines and dashed-dotted lines indicate active and inactive antennas, respectively.

period, different sets of M ULSs out of N available ULSs in the underlying ULA are activated leading to a dynamic linear array (DLA). This DLA model has actually been introduced in [98] though it is originally designed to estimate the DOA of more sources than active antennas, where the sources can be highly correlated. Here, the indices of the selected rows of $\mathbf{I}_N$ used to form $\mathbf{C}_z$ correspond to the indices of the active ULSs at time slot z, the set of M active ULSs in a given time slot z is the same across different scanning periods, and the number of received time samples per antenna in a time slot is one. Fig. 5.2 shows an example of this DLA model. For P2, (5.27) implies that τ sensors are organized into Z groups of P sensors, where the same sampling pattern is adopted by all sensors within the same group and where different groups employ different sampling patterns. The indices of the active cosets used by group z then correspond to the indices of the selected rows of $\mathbf{I}_N$ used to construct $\mathbf{C}_z$. Fig. 5.3 shows an example of the model for problem P2.

Since it turns out that the mathematical model in [98] is applicable for both P1 and P2, we can then follow [98], rewrite (5.28) for $z = 0, 1, \ldots, Z-1$ as

$$\mathbf{r}_{\bar{y}_z}(\vartheta) = \text{vec}(\mathbf{R}_{\bar{y}_z}(\vartheta)) = (\mathbf{C}_z \otimes \mathbf{C}_z)\text{vec}(\mathbf{R}_{\bar{x}}(\vartheta)),$$

combine $\mathbf{r}_{\bar{y}_z}(\vartheta)$ for all z into $\mathbf{r}_{\bar{y}}(\vartheta) = [\mathbf{r}^T_{\bar{y}_0}(\vartheta), \mathbf{r}^T_{\bar{y}_1}(\vartheta), \ldots, \mathbf{r}^T_{\bar{y}_{Z-1}}(\vartheta)]^T$, and write $\mathbf{r}_{\bar{y}}(\vartheta)$ as

$$\mathbf{r}_{\bar{y}}(\vartheta) = \mathbf{\Psi}\text{vec}(\mathbf{R}_{\bar{x}}(\vartheta)), \tag{5.29}$$

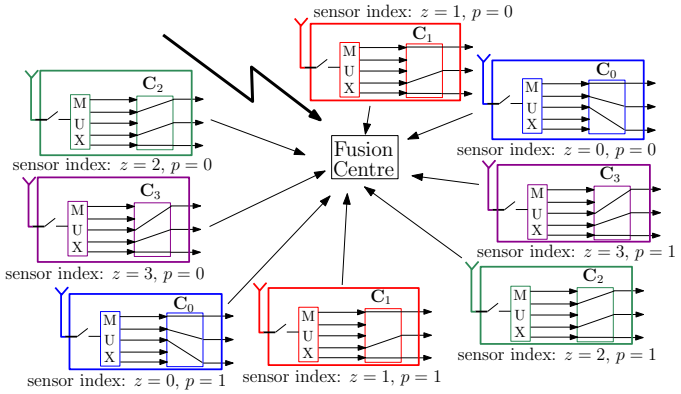


Figure 5.3: The model for problem P2 when the bins are correlated with $M = 3$, $N = 5$, $P = 2$, and $Z = 4$. For simplicity, we illustrate the multi-coset sampling as a Nyquist-rate sampling followed by a multiplexer and a switch that performs sample selection based on $\mathbf{C}_z$. Sensors in the same group have the same colour. For example, sensors in group $z = 0$ collect the samples at the cosets with coset indices 0,1, and 2.

with $\mathbf{\Psi}$ an $M^2Z \times N^2$ matrix given by

$$\mathbf{\Psi} = [(\mathbf{C}_0 \otimes \mathbf{C}_0)^T, \ldots, (\mathbf{C}_{Z-1} \otimes \mathbf{C}_{Z-1})^T]^T. \tag{5.30}$$

We can solve for $\text{vec}(\mathbf{R}_{\bar{x}}(\vartheta))$ from $\mathbf{r}_{\bar{y}}(\vartheta)$ in (5.29) using LS if $\mathbf{\Psi}$ in (5.30) has full column rank. It has been shown in [98] that $\mathbf{\Psi}$ has full column rank if and only if *each possible pair of two different rows* of $\mathbf{I}_N$ is simultaneously used in *at least one* of the matrices $\{\mathbf{C}_z\}_{z=0}^{Z-1}$. In P1, this implies that each possible combination of two ULSs in the underlying ULA should be active in at least one time slot per scanning period. In P2, this implies that each possible pair of two cosets (out of N possible cosets) should be simultaneously used by at least one group of sensors. Observe how the DLA model in Fig. 5.2 and the model in Fig. 5.3 satisfy this requirement. Once $\text{vec}(\mathbf{R}_{\bar{x}}(\vartheta))$ is reconstructed, we follow the procedure in Section 5.3.2 to reconstruct $\mathbf{R}_x(\vartheta) = E[\mathbf{x}_{p,z}(\vartheta)\mathbf{x}_{p,z}^H(\vartheta)]$ from $\mathbf{R}_{\bar{x}}(\vartheta)$.

In practice, to approximate the expectation operation in computing $\mathbf{R}_{\bar{y}_z}(\vartheta)$ in (5.28), we propose to take an average over $\bar{\mathbf{y}}_{p,z}(\vartheta)$ at different scanning periods p for P1 or at P sensors in group z for P2, i.e., $\hat{\mathbf{R}}_{\bar{y}_z}(\vartheta) = \frac{1}{P}\sum_{p=0}^{P-1} \bar{\mathbf{y}}_{p,z}(\vartheta)\bar{\mathbf{y}}_{p,z}^H(\vartheta)$. Introducing $\hat{\mathbf{r}}_{\bar{y}_z}(\vartheta) = \text{vec}(\hat{\mathbf{R}}_{\bar{y}_z}(\vartheta))$, the LS reconstruction is then applied to $\hat{\mathbf{r}}_{\bar{y}}(\vartheta) = [\hat{\mathbf{r}}_{\bar{y}_0}^T(\vartheta), \hat{\mathbf{r}}_{\bar{y}_1}^T(\vartheta), \ldots, \hat{\mathbf{r}}_{\bar{y}_{Z-1}}^T(\vartheta)]^T$.

5.8 Numerical Study

5.8.1 Uncorrelated Bins

In this section, we simulate the estimation and detection performance of the CAP approach for the uncorrelated bins case discussed in Sections 5.2-5.6. To keep the study general, in this section, we generally simulate the multi-cluster scenario of Section 5.6. In our first experiment, we consider problem P2 and have $\tilde{N} = 3060$, $L = 170$, and $N = 18$. Each sensor collects $M = 5$ samples out of every $N = 18$ possible samples based on a periodic length-17 minimal circular sparse ruler with $\mathcal{M} = \{0, 1, 4, 7, 9\}$. This is identical to forming a 5×18 matrix $\mathbf{C}$ in (5.11) by selecting the rows of $\mathbf{I}_{18}$ based on $\mathcal{M}$. The resulting $\mathbf{R}_c$ in (5.13) has full column rank and we have a compression rate of $M/N = 0.28$. We consider $K = 6$ user signals whose frequency bands are given in Table 5.2 together with the power at each band normalized by frequency. We generate these signals by passing six circular complex zero-mean Gaussian i.i.d. noise signals through different digital filters having 200 taps where the location of the unit-gain passband of the filter for each signal corresponds to the six different active bands. We set the variances of these noise signals based on the desired user signal powers in Table 5.2. We assume $D = 2$ clusters of $\tau = 100$ unsynchronized sensors, which means that, at a given point in time, different sensors observe different parts of the user signals. To simplify the experiment, the correlation between the different parts of the user signals observed by different sensors is assumed to be negligible such that they can be viewed as independent realizations of the user signals. The spatially and temporally white noise has a variance of $\sigma^2 = 7$ dBm. The signal of each user received by different sensors is assumed to pass through different and uncorrelated fading channels $H_t^{(k)}(\vartheta)$. Note however that the signal from a user received by sensors within the same cluster is assumed to suffer from the same path loss and shadowing. The amount of path loss experienced between each user and each cluster listed in Table 5.2 includes the shadowing to simplify the simulation. We simulate small-scale Rayleigh fading on top of the path loss by generating the channel frequency response based on a zero-mean complex Gaussian distribution with variance given by the path loss in Table 5.2. We assume flat fading in each band.

Fig. 5.4 shows the CAP of the faded user signals received at the sensors. As a benchmark, we provide the Nyquist-rate based AP (NAP), which is obtained when all sensors collect all the $\tilde{N}$ samples. With respect to the NAP, the degradation in the quality of the CAP is acceptable despite a strong compression, although more leakage is introduced in the unoccupied band. Next, we perform 1000 Monte Carlo

runs and vary the number of sensors per cluster τ, the noise variance at each sensor σ^2, and M/N (see Fig. 5.5). In Fig. 5.5, the compression rate of $M/N = 0.44$ is implemented by activating three extra cosets, i.e., $\{2, 12, 14\}$ (which we picked randomly). Fig. 5.5 shows the normalized mean square error (NMSE) of the CAP with respect to the NAP and indicates that increasing M/N by a factor of less than two significantly improves the estimation quality. Having more sensors τ also improves the estimation quality. Also observe that the compression introduces a larger NMSE for a larger noise power.

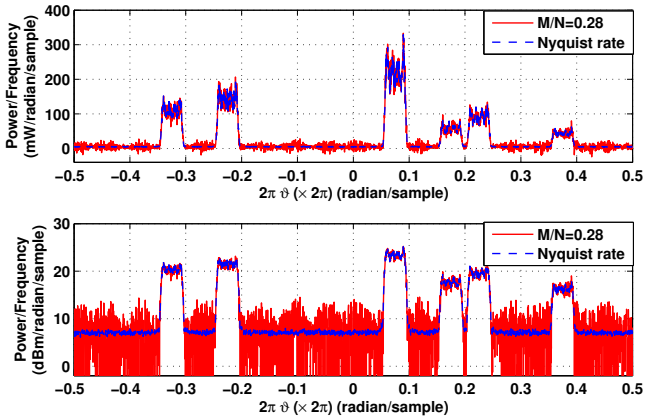


Figure 5.4: The CAP and the NAP of the faded user signals for the first experiment (unsynchronized sensors) as a function of frequency in a linear scale (top) and logarithmic scale (bottom).

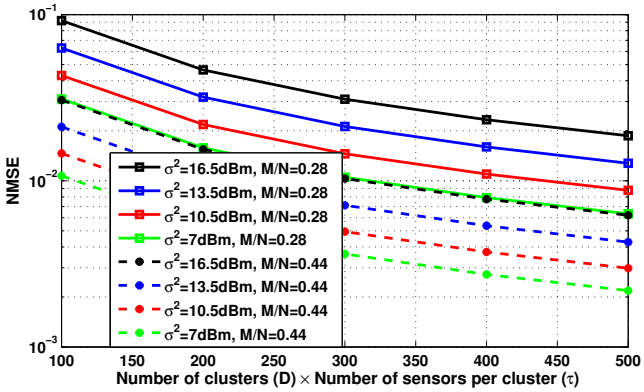


Figure 5.5: The NMSE between the CAP and the NAP for the first experiment (unsynchronized sensors).

We can also re-interpret the first experiment for problem P1. In P1, the first experiment implies that $M = 5$ ULSs (whose indices are indicated by $\mathcal{M}$) out of $N = 18$ ULSs are activated leading to a periodic circular MRA. Table 5.2 then gives the angular bands of the $K = 6$ user signals and the power for each band normalized by the angle. For P1, the first experiment also implies that each user transmits temporally independent signals and that the signals from different users k pass through statistically different and uncorrelated time-varying fading channels $H_t^{(k)}(\vartheta)$ on their way towards the receiving array. For each user k, the fading statistics remain constant within each cluster of time indices but the fading realization is temporally independent.

The second experiment uses the same setting as used in the first experiment (including Table 5.2). The only difference is that the sensors are now assumed to be synchronized. Fig. 5.6 depicts the CAP and the NAP of the faded user signals received at the sensors. Unlike in the unsynchronized sensors case (see Fig. 5.4), we now observe a significant variation in both the CAP and the NAP. This is because, when the sensors are synchronized, they observe the same part of the user signals. This means that, while the fading realization components in the received signals at different sensors are independent, the user signal components in the received signals at different sensors are fully correlated. Fig. 5.7 shows the NMSE of the CAP with respect to the NAP for the synchronized sensors case. In general, some trends found in the unsynchronized sensors case also appear here. Notice that the NMSE for the synchronized sensors case is smaller than the one for the unsynchronized sensors case since the quality of the NAP in the synchronized sensors case is also significantly worse than the one in the unsynchronized sensors case. Note that we can also re-interpret this second experiment for problem P1. This re-interpretation however, will make more sense, if we reverse the roles of $H_t^{(k)}(\vartheta)$ and $U_t^{(k)}(\vartheta)$. When this is the case, for P1, the second experiment implies that each user transmits temporally independent signals and that the signals from different users k pass through statistically different and uncorrelated *time-invariant* fading channels on their way towards the receiving array. Here, the statistics of the user signal are constant only within a cluster of time indices.

In the third experiment, we investigate the impact of varying the bin size (which is equivalent to varying N) and L for a given $\tilde{N}$ on the performance of the CAP approach. Let us consider the settings in the first experiment (i.e., we consider Table 5.2) except for the following. We now examine three different values of N, i.e., $N = 10$, $N = 14$, and $N = 18$ for a given $\tilde{N} = 3150$. For each value of N, we vary the compression rate M/N and examine the two sets of coset patterns

available in Table 5.3. We start from the minimal M/N offered by the minimal circular sparse ruler. Larger compression rates are implemented by selecting additional coset indices where the order of the selection is provided by the third column of Table 5.3. We fix the number of τ to $\tau = 76$ and perform 1000 Monte Carlo simulation runs for different noise variances (see Fig. 5.8). Fig. 5.8 illustrates the NMSE of the CAP with respect to the NAP for the two sets of coset patterns. Observe that varying N and L for a given $\tilde{N}$ does not really result in a clear trend in the estimation performance. While the performance of the CAP for $N = 10$ is worse than the one for the larger value of N, the performance of the CAP for $N = 14$ is better than the one for $N = 18$ for some values of M/N. Note that the NMSE also depends on the coset pattern that we select to implement a particular compression. At this point, we would like to mention that, as long as the bin size constraint in Remark 5.2.2 is satisfied, having a larger N is generally more advantageous as we will generally have a lower value of minimum M/N. This is because it can be found that, as N increases, the number of marks in the corresponding length-$(N-1)$ minimal circular sparse ruler (which is the minimum M) tends to be constant or to increase very slowly. As a result, the minimum compression rate M/N also generally (even though not monotonically) decreases with N.

In the next three experiments, we use the CAP to detect the existence of active user signals that suffer from fading channels and evaluate the detection performance. We start with the fourth experiment, where we again consider problem P2, $\tilde{N} = 3060$, $L = 170$, $N = 18$, and $M/N = 0.28$ (again by adopting $\mathcal{M} = \{0, 1, 4, 7, 9\}$). We now consider $D = 3$ clusters of τ unsynchronized sensors and $K = 3$ user signals (see their settings in Table 5.4), which are generated using the same procedure used in the first experiment. The amount of path loss (which includes shadowing) experienced between each user and each cluster is listed in Table 5.4. We then simulate a small-scale Rayleigh fading channel on top of it. We perform 5000 Monte Carlo runs and vary τ and σ^2 (see Fig. 5.9). We vary the detection threshold manually and out of the $\tilde{N} = 3060$ frequency points at which the CAP is reconstructed, we evaluate the resulting detection events at 363 frequency points in the active bands and the false alarm events at 363 frequency points in the bands that are far from the active bands, i.e., $[-0.77\pi, -0.53\pi]$. Here, we average the estimated power over every eleven subsequent frequency points ϑ and apply the threshold to these average values. The resulting receiver operating characteristic (ROC) is depicted in Fig. 5.9. Observe the acceptable detection performance of the CAP for the examined τ and σ^2 though the performance is slightly poor for $\tau = 17$ and $\sigma^2 = 14$ dBm. This detection performance demonstrates that

the proposed CAP can be used in a spectrum sensing application such as in a CR network.

The fifth experiment repeats the fourth experiment but for synchronized sensors. The ROC in Fig. 5.10 shows that the detection performance for the synchronized sensors case is worse than the one for the unsynchronized sensors case in Fig. 5.9 due to the significant variation in the CAP as shown in Fig. 5.6.

In the sixth experiment, we consider problem P2 and compare the detection performance of the spectrum sensing approach based on CAP with that of the one based on the RM-FOCUSS discussed in Section 5.5.5. To simulate the existence of a joint sparsity structure in $\{\mathbf{x}_t(\vartheta)\}_{t=1}^{\tau}$, we again use the settings in Table 5.4. However, we now only assume one cluster of $\tau = 50$ sensors where the amount of path loss experienced between each user and each sensor is set to -13 dB. The ROC for 5000 Monte Carlo runs and different M/N as well as σ^2 is illustrated in Fig. 5.11. Here, the compression rate of $M/N = 9/18$ is implemented by activating four extra cosets, i.e., 16, 8, 12, 13 (which we decide randomly), on top of the length-17 minimal circular sparse ruler. The M-FOCUSS convergence criterion parameter and the M-FOCUSS diversity measure parameter (labeled as p in [1]) are set to 0.001 and 0.8, respectively. Note that the latter setting follows the suggestion of [1]. To determine the M-FOCUSS regularization parameter, we first perform some experiments and examine ten different values of regularization parameters between 10^{-4} and 10. We then select the regularization parameter that leads to the smallest NMSE between the resulting compressive estimate of $\{|X_t(\vartheta)|^2\}_{t=1}^{\tau}$, for all the considered $\vartheta \in [0, 1)$, and the Nyquist-rate version. We finally decide to set the regularization parameter to 10 for the case of $M/N = 5/18$, to 0.01668 for the case of $M/N = 9/18$ and $\sigma^2 = 14$ dBm, and to 0.21544 for the case of $M/N = 9/18$ and $\sigma^2 = 11$ dBm (see Fig. 5.11). Observe from Fig. 5.11 that the spectrum sensing approach based on CAP has a better detection performance than the one based on signal/spectrum reconstruction using RM-FOCUSS. Recall that the approach of [1] requires the sparsity constraint on the vectors to be reconstructed (which are $\{\mathbf{x}_t(\vartheta)\}_{t=1}^{\tau}$). This implies that, if we have additional active users on top of the scenario used in the sixth experiment, the actual $\{\mathbf{x}_t(\vartheta)\}_{t=1}^{\tau}$ will have a smaller sparsity level. In this case, if we use the same compression rate M/N as the one used in the sixth experiment, the performance of RM-FOCUSS will be even worse.

5.8.2 Correlated Bins

In this section, we conduct the seventh experiment to evaluate the estimation performance of the CAP approach for the correlated bins case discussed in Section 5.7. Here, we consider problem P2, $\tilde{N} = 3080$, $L = 77$, $N = 40$, and $M = 14$ ($M/N = 0.35$). Recall from Section 5.7 that the mathematical model for the correlated bins case is similar to the one in [98]. Hence, to design the sampling matrices for all sensors, which are assumed to be synchronized, that ensure the full column rank of $\mathbf{\Psi}$ in (5.30), we use the algorithm of [98], which is originally designed to solve the antenna selection problem for estimating the DOA of highly correlated sources. This algorithm, which only offers a suboptimal solution for Z, suggests $Z = 12$ groups of $P = 25$ sensors where each group has a unique set of $M = 14$ active cosets. We consider $K = 2$ user signals whose setting is given in Table 5.5. To simulate the full correlation between all the frequency components within the band of the k-th user, we assume that the k-th user transmits exactly the same symbol at all these frequency components at each time instant. On its way toward the different sensors, the signal of the k-th user is assumed to pass through different and uncorrelated Rayleigh fading channels $H_t^{(k)}(\vartheta)$ but it suffers from the same path loss and shadowing, whose value is listed in Table 5.5. Again, we assume flat fading in each user band and have $\sigma^2 = 7$ dBm. Fig. 5.12 shows the CAP of the faded user signals using the correlated bins (CB) assumption. As a benchmark, we also provide the NAP and the CAP based on the uncorrelated bins (UB) assumption discussed in Sections 5.2-5.5, which is obtained by activating the same set of $M = 14$ cosets, i.e., $\mathcal{M} = \{0, 1, 2, 3, 4, 9, 10, 15, 16, 18, 20, 30, 33, 37\}$, in all sensors (leading to a full column rank matrix $\mathbf{R}_c$ in (5.13)). Observe that the quality of the CAP based on the UB assumption is extremely poor. On the other hand, with respect to the NAP, the degradation in the quality of the CAP based on the CB assumption is acceptable despite a significant variation in the unoccupied band. Next, we perform 1000 Monte Carlo runs and vary the number of sensors per group P, σ^2, and M/N (see Fig. 5.13). In Fig. 5.13, the compression rate of $M/N = 0.45$ is implemented by randomly activating four additional cosets on top of the already selected 14 cosets and the resulting sampling pattern is kept fixed throughout the entire Monte Carlo runs. Fig. 5.13 shows the NMSE of the CAP based on the CB assumption with respect to the NAP, which indicates that either increasing M/N or having more sensors per group P can significantly improve the estimation quality. Again, a larger NMSE is introduced for a larger noise power.

The interpretation of this seventh experiment for P1 is similar to the problem

discussed in [98]. For P1, this experiment is equivalent to having a ULA consisting of $N = 40$ ULSs, where the array scanning time is split into $P = 25$ scanning periods, each of which consists of $Z = 12$ time slots. In different time slots per scanning period, we activate different sets of $M = 14$ (out of $N = 40$) ULSs leading to a DLA. The interpretation will again make more sense if we reverse the roles of $H_t^{(k)}(\vartheta)$ and $U_t^{(k)}(\vartheta)$. When this is the case, for P1, the experiment implies that all users transmit temporally independent signals and that the signals from different users k pass through statistically different and uncorrelated time-invariant fading channels on their way towards the receiving array. As the signal received from the k-th user at different angles within its angular band is fully correlated, this can be related to a situation where the same symbol of the k-th user hits different scatterers (which play the role of the channel) before reaching the observing array. From the point of view of the array, the scattered versions of the symbol will be received from different angles within a particular angular band.

5.8.3 Circular Complex Gaussian Noise

The last experiment examines the performance of the CAP based on the UB assumption when the received signal $x_t[\tilde{n}]$ only contains circular complex zero-mean Gaussian spatially and temporally i.i.d. noise. Here, we have $\tilde{N} = 3060$, $L = 170$, $N = 18$, and $\sigma^2 = 7$ dBm. We perform 1000 Monte Carlo runs and vary τ (see Fig. 5.14). We compute the NMSE of the CAP with respect to the true power spectrum (since $x_t[\tilde{n}]$ in this case is clearly a WSS signal) and compare this NMSE obtained from the simulation with the analytical NMSE. Since it can be shown that, for circular complex Gaussian i.i.d. noise $x_t[\tilde{n}]$, $\hat{P}_{x,LS}(\vartheta)$ is an unbiased estimate of $P_x(\vartheta)$ even for finite $\tilde{N}$, the analytical NMSE only depends on the variance of $\hat{P}_{x,LS}(\vartheta)$ and it can be shown to be equal to $\frac{1}{\tau}(\frac{1}{M} + \sum_{n=1}^{N-1} \frac{1}{\gamma_{n+1}})$ by using (5.26). We start with $M/N = 0.28$ by using the cosets indexed by the length-17 minimal circular sparse ruler, i.e., $\mathcal{M} = \{0, 1, 4, 7, 9\}$, and then vary M/N. First, the cases of $M/N > 0.28$ are implemented by activating additional cosets based on Pattern 1 in Table 5.6. Then, we also test Pattern 2 and Pattern 3 as additional coset patterns to implement the case of $M/N = 0.5$. Observe in Fig. 5.14 how the analytical NMSE is on top of the simulated NMSE for all the evaluated M/N values. Also observe that, for $M/N = 0.5$, the three different coset patterns have led to different values of the NMSE depending on the resulting value of $\{\gamma_{n+1}\}_{n=1}^{N-1}$ in (5.26).

5.9 Conclusion and Future Work

This paper proposed a compressive periodogram reconstruction approach and considered both time-frequency and spatio-angular domains. In our model, the entire band is split into uniform bins such that the received spectra at two frequencies or angles, whose distance is equal to or larger than the size of a bin, are uncorrelated. In both considered domains, this model leads to a circulant coset correlation matrix, which allows us to perform a strong compression yet to present our reconstruction problem as an overdetermined system. When the coset patterns are designed based on a circular sparse ruler, the system matrix has full column rank and we can reconstruct the periodogram using LS. In a practical situation, our estimate of the coset correlation matrix is only asymptotically circulant. Hence, we also presented an asymptotic bias and variance analysis for the CAP. We further included a thorough variance analysis on the case when the received signal only contains circular complex zero-mean white Gaussian noise, which provides some useful insights in the performance of our approach. The variance analysis for a more general signal (i.e., a general Gaussian signal) has also been presented but it is not easy to interpret due to its dependence on the unknown statistics of the user signals. We also proposed a solution for the case when the bin size is decreased such that the received spectra at two frequencies or angles, with a spacing between them larger than the size of the bin, can still be correlated. Finally, the simulation study showed that the estimation performance of the evaluated approach is acceptable and that our CAP performs well when detecting the existence of the user signals suffering from fading channels.

As a future work, we are interested in the case when both problems P1 and P2 emerge simultaneously. In that case, we would consider a compressive linear array of antennas and a compressive digital recever unit per antenna leading to a two-dimensional (2D) digital signal. Our interest would then be to investigate if it is possible to perform compression in both the time and spatial domain and to jointly reconstruct the angular and frequency periodogram from the 2D compressive samples. To study that, we could follow an approach similar to [103], which assumes stationarity in both the time and spatial domain and exploits the existing Toeplitz structure in the correlation matrix.

Appendix

5.A Proof of Theorem 5.5.1

Recall that $\hat{\mathbf{R}}_{\bar{y}}(\vartheta)$ in (5.15) is an unbiased estimate of $\mathbf{R}_{\bar{y}}(\vartheta)$ in (5.12), i.e., $E[\hat{\mathbf{R}}_{\bar{y}}(\vartheta)] = \mathbf{R}_{\bar{y}}(\vartheta)$. Applying the expectation operator on (5.16) and (5.17), it is then clear that $\hat{\mathbf{r}}_{\bar{x},LS}(\vartheta)$ in (5.16) and $\hat{\mathbf{R}}_{x,LS}(\vartheta)$ in (5.17) are unbiased estimates of $\mathbf{r}_{\bar{x}}(\vartheta)$ in (5.13) and $\mathbf{R}_x(\vartheta)$ in (5.7), respectively, since $\mathbf{r}_{\bar{x}}(\vartheta)$ in (5.13) can perfectly be reconstructed from $\mathbf{R}_{\bar{y}}(\vartheta)$ using LS. Recall from Remark 5.2.1 that the $(i+1)$-th diagonal element of $\mathbf{R}_x(\vartheta)$ is equal to $E[|X_{t,i}(\vartheta)|^2]$. From (5.18), it is then obvious that the CAP $\hat{P}_{x,LS}(\vartheta + \frac{i}{N})$ is an unbiased estimate of $\frac{1}{N}E[|X_{t,i}(\vartheta)|^2]$. However, by taking (5.1) into account, we can observe that

$$\lim_{\tilde{N}\to\infty} \frac{1}{\tilde{N}} E[|X_{t,i}(\vartheta)|^2] = P_x(\vartheta + \frac{i}{N}), \;\; \vartheta \in [0, 1/N), \tag{5.31}$$

for $i = 0, 1, \ldots, N-1$, since $x_t[\tilde{n}]$ is a finite-length observation of the actual random process $x[\tilde{n}]$. Hence, by applying $\lim_{\tilde{N}\to\infty} E[\hat{P}_{x,LS}(\vartheta + \frac{i}{N})]$ and using (5.31), it is clear that $\hat{P}_{x,LS}(\vartheta + \frac{i}{N})$ is an asymptotically (with respect to $\tilde{N}$) unbiased estimate of $P_x(\vartheta + \frac{i}{N})$ in (5.1), for $\vartheta \in [0, 1/N)$ and $i = 0, 1, \ldots, N-1$. □

5.B Proof of Proposition 5.5.1

Note that for the specific case assumed in this proposition, we can rewrite (5.20) as

$$\begin{aligned}
&\mathrm{Cov}[[\hat{\mathbf{R}}_{\bar{y}}(\vartheta)]_{m+1,m'+1}, [\hat{\mathbf{R}}_{\bar{y}}(\vartheta)]_{a+1,a'+1}] \\
&= \frac{1}{N^4\tau^2} \sum_{t=1}^{\tau} \sum_{i=0}^{N-1} \sum_{i'=0}^{N-1} \sum_{b=0}^{N-1} \sum_{b'=0}^{N-1} e^{\frac{j2\pi(n_m i - n_{m'} i' - n_a b + n_{a'} b')}{N}} \times \\
&E[X_{t,i}(\vartheta) X^*_{t,b}(\vartheta)] E[X^*_{t,i'}(\vartheta) X_{t,b'}(\vartheta)],
\end{aligned} \tag{5.32}$$

where we also take the circularity of $x_t[\tilde{n}]$ into account. By using $\tilde{N} = LN$, we can find that $E[X_{t,i}(\vartheta) X^*_{t,b}(\vartheta)] = \sigma^2 \sum_{\tilde{n}=0}^{\tilde{N}-1} e^{j2\pi\tilde{n}(\frac{b-i}{N})} = \tilde{N}\sigma^2\delta[b-i]$, as it is clear from (5.32) that $b, i \in \{0, 1, \ldots, N-1\}$. Hence, we can simplify (5.32) as

$$\begin{aligned}
&\mathrm{Cov}[[\hat{\mathbf{R}}_{\bar{y}}(\vartheta)]_{m+1,m'+1}, [\hat{\mathbf{R}}_{\bar{y}}(\vartheta)]_{a+1,a'+1}] = \sum_{t=1}^{\tau} \sum_{i=0}^{N-1} \sum_{i'=0}^{N-1} \\
&\sum_{b=0}^{N-1} \sum_{b'=0}^{N-1} e^{\frac{j2\pi(n_m i - n_{m'} i' - n_a b + n_{a'} b')}{N}} \frac{L^2\sigma^4}{N^2\tau^2} \delta[b-i]\delta[i'-b']
\end{aligned}$$

$$= \frac{L^2\sigma^4}{N^2\tau} \sum_{i=0}^{N-1} e^{\frac{j2\pi i(n_m - n_a)}{N}} \sum_{i'=0}^{N-1} e^{\frac{j2\pi i'(n_{a'} - n_{m'})}{N}}$$
$$= \frac{L^2\sigma^4}{\tau} \delta[m-a]\delta[m'-a'], \;\; \vartheta \in [0, 1/N),$$

where the last equality is due to $n_m \in \{0, 1, \ldots, N-1\}$, for all m, and the fact that $n_m = n_a$ implies $m = a$. □

5.C Proof of (5.25)

First, by recalling that $\mathbf{R}_c = (\mathbf{C} \otimes \mathbf{C})\mathbf{T}$, we can write

$$\begin{aligned}\gamma_\kappa &= [\mathbf{R}_c^T \mathbf{R}_c]_{\kappa,\kappa} = [\mathbf{T}^T((\mathbf{C}^T\mathbf{C}) \otimes (\mathbf{C}^T\mathbf{C}))\mathbf{T}]_{\kappa,\kappa} \\ &= [\mathbf{T}^T(\text{diag}(\mathbf{w}) \otimes \text{diag}(\mathbf{w}))\mathbf{T}]_{\kappa,\kappa} \\ &= \sum_{n=0}^{N-1}\sum_{n'=0}^{N-1} [\mathbf{T}^T]_{\kappa, Nn+n'+1} [\text{diag}(\mathbf{w}) \otimes \text{diag}(\mathbf{w})]_{Nn+n'+1, Nn+n'+1} [\mathbf{T}]_{Nn+n'+1,\kappa}.\end{aligned} \tag{5.33}$$

Let us then recall that the $(q+1)$-th row of $\mathbf{T}$ is given by the $\left(\left(q - \left\lfloor \frac{q}{N} \right\rfloor\right) \bmod N + 1\right)$-th row of $\mathbf{I}_N$. We can then find that the $(\iota+1)$-th row of $\mathbf{T}^T$ contains ones at the $\{Nn + (n+\iota) \bmod N + 1\}_{n=0}^{N-1}$-th entries and zeros elsewhere. We can thus rewrite (5.33)

$$\begin{aligned}\gamma_\kappa &= \sum_{n=0}^{N-1} [\mathbf{w} \otimes \mathbf{w}]_{Nn+((n+\kappa-1) \bmod N)+1} = \sum_{n=0}^{N-1} [\mathbf{w}\mathbf{w}^T]_{((n+\kappa-1) \bmod N)+1, n+1} \\ &= \sum_{n=0}^{N-1} w[(n+\kappa-1) \bmod N] w[n],\end{aligned} \tag{5.34}$$

where we use $\mathbf{w}\mathbf{w}^T = \text{vec}^{-1}(\mathbf{w} \otimes \mathbf{w})$ in the second equation with $\text{vec}^{-1}(.)$ the inverse of the vec$(.)$ operation.

5.D Proof of Theorem 5.5.2

To simplify the discussion, we introduce the $N^2 \times 1$ vector

$$\hat{\boldsymbol{\rho}}_{\bar{x}}(\vartheta) = (\mathbf{C} \otimes \mathbf{C})^T \text{vec}(\hat{\mathbf{R}}_{\bar{y}}(\vartheta)). \tag{5.35}$$

From the definition of $\mathbf{C}$ in Section 5.3.1, it is clear that the $(Nf+g+1)$-th row of $(\mathbf{C}\otimes\mathbf{C})^T$ contains a single one at a certain entry and zeros elsewhere only if $f,g\in\mathcal{M}$, otherwise it contains zeros at all entries. Hence, we can write

$$[\hat{\boldsymbol{\rho}}_{\bar{x}}(\vartheta)]_{Nf+g+1}=0,\quad \text{if } f\notin\mathcal{M} \text{ or } g\notin\mathcal{M}. \tag{5.36}$$

When $f,g\in\mathcal{M}$, the $(Nf+g+1)$-th entry of $\hat{\boldsymbol{\rho}}_{\bar{x}}(\vartheta)$ is given by one of the entries of $\text{vec}(\hat{\mathbf{R}}_{\bar{y}}(\vartheta))$. Recall from Appendix 5.C that the $(\iota+1)$-th row of $\mathbf{T}^T$ contains ones at the $\{Nn+(n+\iota) \bmod N+1\}_{n=0}^{N-1}$-th entries and zeros elsewhere. As a result, we can use (5.16), (5.35), and Remark 5.5.1 to write the $(\iota+1)$-th entry of $\hat{\mathbf{r}}_{\bar{x},LS}(\vartheta)$ in (5.16) as

$$[\hat{\mathbf{r}}_{\bar{x},LS}(\vartheta)]_{\iota+1}=\frac{1}{\gamma_{\iota+1}}[\mathbf{T}^T\hat{\boldsymbol{\rho}}_{\bar{x}}(\vartheta)]_{\iota+1}=\frac{1}{\gamma_{\iota+1}}\sum_{n=0}^{N-1}[\text{vec}^{-1}(\hat{\boldsymbol{\rho}}_{\bar{x}}(\vartheta))]_{(n+\iota) \bmod N+1,n+1}, \tag{5.37}$$

with $\iota=0,1,\ldots,N-1$ and $\text{vec}^{-1}(\hat{\boldsymbol{\rho}}_{\bar{x}}(\vartheta))$ an $N\times N$ matrix.

At this stage, let us introduce the following definition.

Definition 5.D.1. *Define the collection of $[\text{vec}^{-1}(\hat{\boldsymbol{\rho}}_{\bar{x}}(\vartheta))]_{g'+1,f'+1}$ for $f',g'\in\{0,1,\ldots,N-1\}$ and all $((g'-f') \bmod N+1)=\kappa$ as the κ-th modular diagonal of $\text{vec}^{-1}(\hat{\boldsymbol{\rho}}_{\bar{x}}(\vartheta))$. Note that the first modular diagonal of $\text{vec}^{-1}(\hat{\boldsymbol{\rho}}_{\bar{x}}(\vartheta))$ is its main diagonal.*

We use Definition 5.D.1 to formulate the following lemma.

Lemma 5.D.1. *The κ-th modular diagonal of $\text{vec}^{-1}(\hat{\boldsymbol{\rho}}_{\bar{x}}(\vartheta))$ in* (5.37) *contain only γ_κ entries of $\text{vec}(\hat{\mathbf{R}}_{\bar{y}}(\vartheta))$ in* (5.35)*. The remaining $N-\gamma_\kappa$ entries of the κ-th modular diagonal of $\text{vec}^{-1}(\hat{\boldsymbol{\rho}}_{\bar{x}}(\vartheta))$ are equal to zeros. The summation in* (5.37) *then involves $N-\gamma_{\iota+1}$ zeros and only $\gamma_{\iota+1}$ out of M^2 entries of $\text{vec}(\hat{\mathbf{R}}_{\bar{y}}(\vartheta))$.*

Proof. Recall that, when $f,g\in\mathcal{M}$, the $(Nf+g+1)$-th entry of $\hat{\boldsymbol{\rho}}_{\bar{x}}(\vartheta)$ in (5.37) is given by one of the entries of $\text{vec}(\hat{\mathbf{R}}_{\bar{y}}(\vartheta))$. Since Remark 5.5.1 indicates that the number of pairs $g,f\in\mathcal{M}$ that lead to $(g-f) \bmod N+1=\kappa$ is equal to γ_κ, it is clear from Definition 5.D.1 that the κ-th modular diagonal of $\text{vec}^{-1}(\hat{\boldsymbol{\rho}}_{\bar{x}}(\vartheta))$ only contains γ_κ entries of $\text{vec}(\hat{\mathbf{R}}_{\bar{y}}(\vartheta))$. Equation (5.36) then confirms that the remaining $N-\gamma_\kappa$ entries of the κ-th modular diagonal of $\text{vec}^{-1}(\hat{\boldsymbol{\rho}}_{\bar{x}}(\vartheta))$ are equal to zero. Next, observe that the summation in (5.37) is the sum of all terms in the $(\iota+1)$-th modular diagonal of $\text{vec}^{-1}(\hat{\boldsymbol{\rho}}_{\bar{x}}(\vartheta))$. This can be found by applying Definition 5.D.1 on the column and row indices of $\text{vec}^{-1}(\hat{\boldsymbol{\rho}}_{\bar{x}}(\vartheta))$ in (5.37), i.e.,

$$((n+\iota) \bmod N-n) \bmod N+1=(n+\iota-n) \bmod N+1=\iota+1,$$

which exploits the property that $(\kappa \bmod N + \kappa') \bmod N = (\kappa + \kappa') \bmod N$. This concludes the proof. □

Let us now define $\mathbf{\Sigma}_{\hat{\rho}_{\bar{x}}}(\vartheta)$ as the $N^2 \times N^2$ covariance matrix of $\hat{\boldsymbol{\rho}}_{\bar{x}}(\vartheta)$ in (5.35), which can be written as $\mathbf{\Sigma}_{\hat{\rho}_{\bar{x}}}(\vartheta) = (\mathbf{C} \otimes \mathbf{C})^T \mathbf{\Sigma}_{\hat{R}_{\bar{y}}}(\vartheta)(\mathbf{C} \otimes \mathbf{C})$. First, recall (5.36) and that when $f, g \in \mathcal{M}$, the $(Nf+g+1)$-th entry of $\hat{\boldsymbol{\rho}}_{\bar{x}}(\vartheta)$ in (5.37) is given by one of the entries of $\mathrm{vec}(\hat{\mathbf{R}}_{\bar{y}}(\vartheta))$. By also recalling that, for circular complex Gaussian i.i.d. noise $x_t[\tilde{n}]$, $\mathbf{\Sigma}_{\hat{R}_{\bar{y}}}(\vartheta)$ is a diagonal matrix whose elements are given by (5.24), we can find that $\mathbf{\Sigma}_{\hat{\rho}_{\bar{x}}}(\vartheta)$ is also a diagonal matrix with its diagonal elements given by

$$[\mathrm{diag}(\mathbf{\Sigma}_{\hat{\rho}_{\bar{x}}}(\vartheta))]_{Nf+g+1} = \begin{cases} \frac{L^2\sigma^4}{\tau}, & \text{if } f, g \in \mathcal{M}. \\ 0, \text{ if } f \notin \mathcal{M} \text{ or } g \notin \mathcal{M}. \end{cases} \tag{5.38}$$

By taking (5.37), (5.38), and the diagonal structure of $\mathbf{\Sigma}_{\hat{\rho}_{\bar{x}}}(\vartheta)$ into account, we can then write the entry of $\mathbf{\Sigma}_{\hat{r}_{\bar{x},LS}}(\vartheta)$ in (5.21) at the $(\iota+1)$-th row and the $(\iota'+1)$-th column as

$$\begin{aligned} &\mathrm{Cov}[[\hat{\mathbf{r}}_{\bar{x},LS}(\vartheta)]_{\iota+1}, [\hat{\mathbf{r}}_{\bar{x},LS}(\vartheta)]_{\iota'+1}] = \frac{1}{\gamma_{\iota+1}\gamma_{\iota'+1}} \times \\ &\sum_{n=0}^{N-1}\sum_{n'=0}^{N-1} \left\{ [\mathbf{T}^T]_{\iota+1, Nn+n'+1} [\mathbf{\Sigma}_{\hat{\rho}_{\bar{x}}}(\vartheta)]_{Nn+n'+1, Nn+n'+1} [\mathbf{T}]_{Nn+n'+1, \iota'+1} \right\} \\ &= \frac{\delta[\iota - \iota']}{\gamma_{\iota+1}^2} \sum_{n=0}^{N-1} [\mathbf{\Sigma}_{\hat{\rho}_{\bar{x}}}(\vartheta)]_{Nn+((n+\iota) \bmod N)+1, Nn+((n+\iota) \bmod N)+1}, \end{aligned} \tag{5.39}$$

for $\iota, \iota' = 0, 1, \ldots, N-1$, which implies that $\mathbf{\Sigma}_{\hat{r}_{\bar{x},LS}}(\vartheta)$ is also a diagonal matrix for circular complex Gaussian i.i.d. noise $x_t[\tilde{n}]$. Recall from the proof of Lemma 5.D.1 that the summation in (5.37) is the sum of all terms in the $(\iota+1)$-th modular diagonal of $\mathrm{vec}^{-1}(\hat{\boldsymbol{\rho}}_{\bar{x}}(\vartheta))$. We can then observe that the summation in (5.39) is the sum of the variance of each term in the $(\iota+1)$-th modular diagonal of $\mathrm{vec}^{-1}(\hat{\boldsymbol{\rho}}_{\bar{x}}(\vartheta))$. Using Lemma 5.D.1 and (5.38), we can rewrite (5.39) as

$$\mathrm{Cov}[[\hat{\mathbf{r}}_{\bar{x},LS}(\vartheta)]_{\iota+1}, [\hat{\mathbf{r}}_{\bar{x},LS}(\vartheta)]_{\iota'+1}] = \frac{L^2\sigma^4}{\gamma_{\iota+1}\tau}\delta[\iota - \iota'], \tag{5.40}$$

for $\iota, \iota' = 0, 1, \ldots, N-1$. By considering (5.22) and noticing that $[\mathbf{B}^T \otimes \mathbf{B}^H]_{Ni+i'+1, Nn+n'+1} = \frac{1}{N^2} e^{-j\frac{2\pi}{N}(n'i'-ni)}$, let us rewrite $\mathrm{Var}[\hat{P}_{x,LS}(\vartheta + \frac{i}{N})]$ in (5.23), for $\vartheta \in [0, 1/N)$ and $i = 0, 1, \ldots, N-1$, as

$$\mathrm{Var}[\hat{P}_{x,LS}(\vartheta + \frac{i}{N})] = \frac{N^4}{\tilde{N}^2} \sum_{n=0}^{N-1}\sum_{n'=0}^{N-1}\sum_{\nu=0}^{N-1}\sum_{\nu'=0}^{N-1} \left\{ [\mathbf{B}^T \otimes \mathbf{B}^H]_{Ni+i+1, Nn+n'+1} \right.$$

$$\times[\mathbf{T}\boldsymbol{\Sigma}_{\hat{r}_{\bar{x},LS}}(\vartheta)\mathbf{T}^T]_{Nn+n'+1,N\nu+\nu'+1}[\mathbf{B}^*\otimes\mathbf{B}]_{N\nu+\nu'+1,Ni+i+1}\Big\}$$
$$=\frac{1}{L^2N^2}\sum_{n=0}^{N-1}\sum_{n'=0}^{N-1}\sum_{\nu=0}^{N-1}\sum_{\nu'=0}^{N-1}\Big\{e^{-j\frac{2\pi}{N}i(n'-n+\nu-\nu')}$$
$$\times[\mathbf{T}\boldsymbol{\Sigma}_{\hat{r}_{\bar{x},LS}}(\vartheta)\mathbf{T}^T]_{Nn+n'+1,N\nu+\nu'+1}\Big\}. \tag{5.41}$$

We now recall that the $(q+1)$-th row of $\mathbf{T}$ is given by the $\big(\big(q-\lfloor\frac{q}{N}\rfloor\big) \bmod N+1\big)$-th row of $\mathbf{I}_N$, exploit the diagonal structure of $\boldsymbol{\Sigma}_{\hat{r}_{\bar{x},LS}}(\vartheta)$ for circular complex Gaussian i.i.d. noise $x_t[\tilde{n}]$, and use (5.40) to write

$$\begin{aligned}&[\mathbf{T}\boldsymbol{\Sigma}_{\hat{r}_{\bar{x},LS}}(\vartheta)\mathbf{T}^T]_{Nn+n'+1,N\nu+\nu'+1}\\&=\frac{L^2\sigma^4}{\tau}\sum_{\iota=0}^{N-1}\frac{1}{\gamma_{\iota+1}}[\mathbf{T}]_{Nn+n'+1,\iota+1}[\mathbf{T}^T]_{\iota+1,N\nu+\nu'+1}\\&=\frac{L^2\sigma^4}{\tau}\frac{\delta[(n'-n) \bmod N-(\nu'-\nu) \bmod N]}{\gamma_{(n'-n) \bmod N+1}},\end{aligned} \tag{5.42}$$

for $n,n',\nu,\nu'=0,1,\ldots,N-1$. By inserting (5.42) into (5.41), the variance of $\hat{P}_{x,LS}(\vartheta+\frac{i}{N})$, for circular complex Gaussian i.i.d. noise $x_t[\tilde{n}]$ and $i=0,1,\ldots,N-1$, is given by

$$\begin{aligned}\text{Var}[\hat{P}_{x,LS}(\vartheta+\frac{i}{N})]&=\frac{1}{L^2N^2}\sum_{n=0}^{N-1}\sum_{n'=0}^{N-1}\frac{L^2\sigma^4N}{\tau\gamma_{(n'-n) \bmod N+1}}\\&=\frac{\sigma^4}{\tau}\sum_{n=0}^{N-1}\frac{1}{\gamma_{n+1}}=\frac{\sigma^4}{M\tau}+\frac{\sigma^4}{\tau}\sum_{n=1}^{N-1}\frac{1}{\gamma_{n+1}},\ \vartheta\in[0,1/N),\end{aligned}$$

where we use the last part of Remark 5.5.1 in the last equality. □

Table 5.1: Computational complexity of the CAP approach and the RM-FOCUSS of [1] for a given frequency point $\vartheta \in [0, 1/N)$.

CAP approach	
Computation steps	Computational complexity
Computation of $\hat{\mathbf{R}}_{\bar{y}}(\vartheta)$ in (5.15)	$\mathcal{O}(M^2\tau)$
Computation of $\mathbf{R}_c^T\mathbf{R}_c$ in (5.16)	$\mathcal{O}(N^2M^2)$
Inversion of $\mathbf{R}_c^T\mathbf{R}_c$ in (5.16)	$\mathcal{O}(N^3)$
Multiplication between $(\mathbf{R}_c^T\mathbf{R}_c)^{-1}$ and $\mathbf{R}_c^T\text{vec}(\hat{\mathbf{R}}_{\bar{y}}(\vartheta))$ in (5.16)	$\mathcal{O}(N^2M^2)+$ $\mathcal{O}(NM^2)$
Computation of (5.17) (recall that $\mathbf{B}$ in (5.17) is an IDFT matrix)	$\mathcal{O}(N \log N)$
Total	$\mathcal{O}(N^3) + \mathcal{O}(N^2M^2) + \mathcal{O}(M^2\tau)$
RM-FOCUSS of [1] (per iteration)	
Computation steps	Computational complexity
Computation of ℓ_2-norm of each row of an $N \times \tau$ matrix	$\mathcal{O}(N\tau)$
Multiplication between an $M \times N$ matrix and an $N \times N$ matrix	$\mathcal{O}(N^2M)$
Multiplication between an $M \times N$ matrix and an $N \times M$ matrix	$\mathcal{O}(NM^2)$
Inversion of an $M \times M$ matrix	$\mathcal{O}(M^3)$
Multiplication between an $N \times M$ matrix and an $M \times M$ matrix	$\mathcal{O}(NM^2)$
Multiplication between an $N \times M$ matrix and an $M \times \tau$ matrix	$\mathcal{O}(NM\tau)$
Multiplication between an $N \times N$ matrix and an $N \times \tau$ matrix	$\mathcal{O}(N^2\tau)$
Total	$\mathcal{O}(N^2M) + \mathcal{O}(M^3)+$ $\mathcal{O}(N^2\tau) + \mathcal{O}(NM^2) + \mathcal{O}(NM\tau)$

Table 5.2: The frequency band and the power of the users signal and the experienced path loss in the first, second, and third experiments.

User band (rad/sample)	Power/freq. (per rad/sample)	Path loss at cluster 1	Path loss at cluster 2
$[-0.69\pi, -0.61\pi]$	38 dBm	-17 dB	-19 dB
$[-0.49\pi, -0.41\pi]$	40 dBm	-20 dB	-18 dB
$[0.11\pi, 0.19\pi]$	34 dBm	-12 dB	-10 dB
$[0.31\pi, 0.39\pi]$	34 dBm	-16 dB	-18 dB
$[0.41\pi, 0.49\pi]$	32 dBm	-14 dB	-12 dB
$[0.71\pi, 0.79\pi]$	35 dBm	-18 dB	-20 dB

Table 5.3: The two sets of coset patterns used in the third experiment (comparison of different bin size).

First set of coset patterns		
N	Minimal circular sparse ruler indices	The order of the additional coset indices for implementing a larger compression rate
18	0, 1, 4, 7, 9	17, 2, 13, 12, 15, 6
14	0, 1, 2, 4, 7	10, 6, 12, 5
10	0, 1, 3, 5	8, 4
Second set of coset patterns		
N	Minimal circular sparse ruler indices	The order of the additional coset indices for implementing a larger compression rate
18	0, 1, 4, 7, 9	5, 2, 6, 17, 15, 14
14	0, 1, 2, 4, 7	12, 10, 13, 11
10	0, 1, 3, 5	4, 6

Table 5.4: The frequency band and the power of the user signals and the experienced path loss in the fourth and the fifth experiments.

User band (rad/sample)	Power/freq. (per rad/sample)	Path loss (in dB) at cluster		
		1	2	3
$[0.41\pi, 0.49\pi]$	25 dBm	-12	-13	-14
$[0.31\pi, 0.39\pi]$	25 dBm	-14.5	-13	-11.5
$[0.21\pi, 0.29\pi]$	25 dBm	-13.5	-13	-12.5

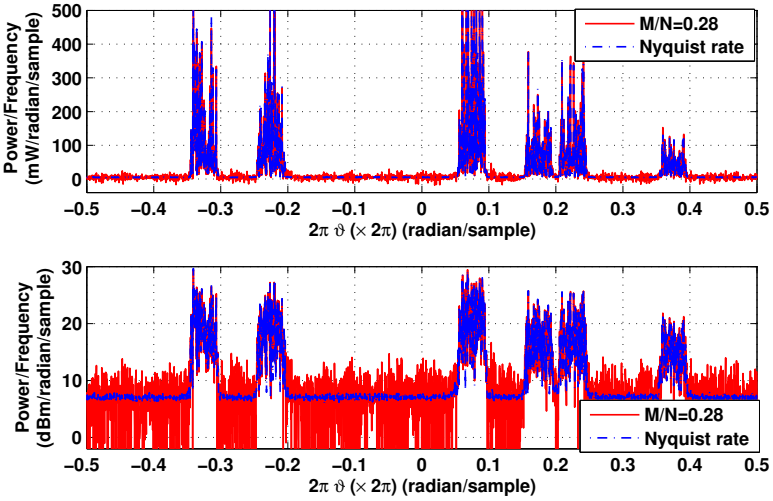


Figure 5.6: The CAP and the NAP of the faded user signals for the second experiment (synchronized sensors) as a function of frequency in a linear scale (top) and logarithmic scale (bottom).

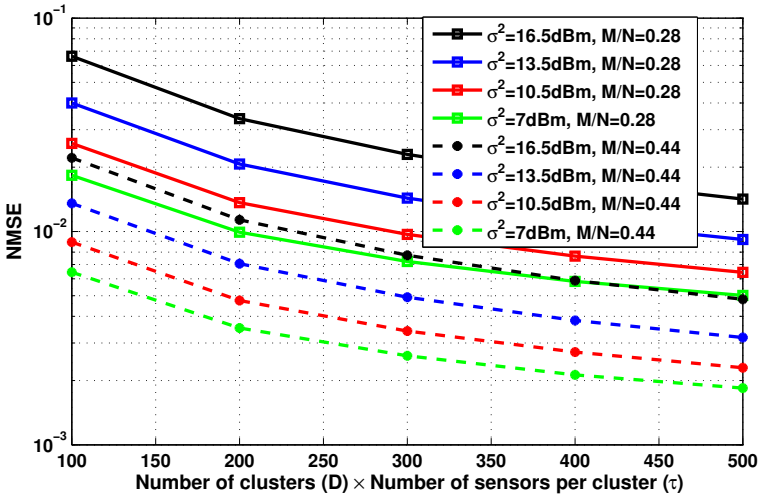
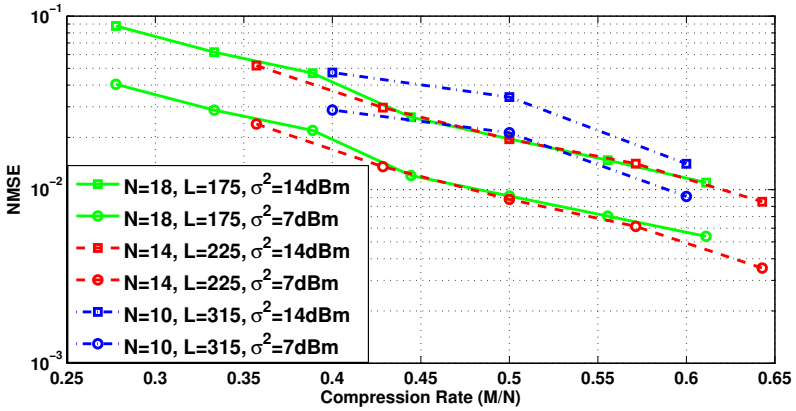
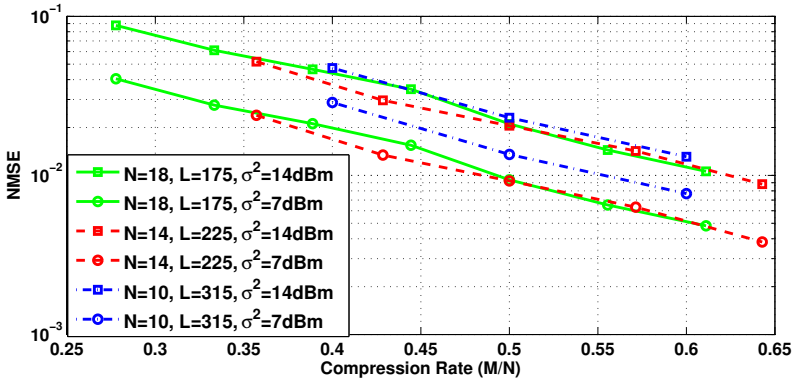


Figure 5.7: The NMSE between the CAP and the NAP for the second experiment (synchronized sensors).



(a)



(b)

Figure 5.8: The NMSE between the CAP and the NAP for the third experiment (comparison of different bin size); (a) using the first set of coset patterns (see Table 5.3); (b) using the second set of coset patterns.

Table 5.5: The frequency bands occupied by the users, their power, and the experienced path loss in the seventh experiment.

User band (rad/sample)	Power/freq. (per rad/sample)	Path loss
$[-0.88\pi, -0.2\pi]$	22 dBm	-6 dB
$[0.15\pi, 0.92\pi]$	25 dBm	-7 dB

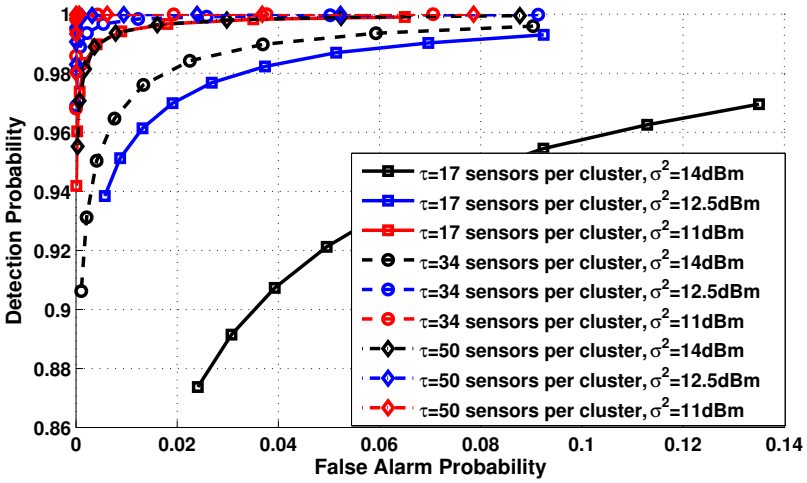


Figure 5.9: The resulting ROC when the CAP is used to detect the existence of the active user signals suffering from fading channels in the fourth experiment (unsynchronized sensors).

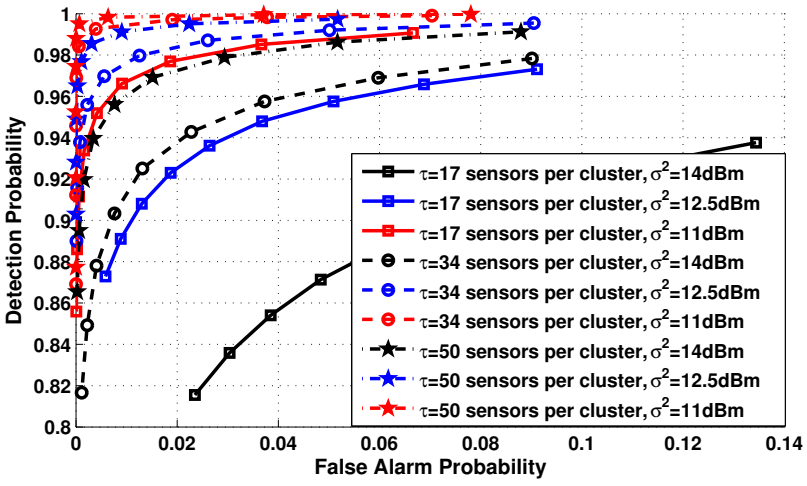


Figure 5.10: The resulting ROC when the CAP is used to detect the existence of the active user signals suffering from fading channels in the fifth experiment (synchronized sensors).

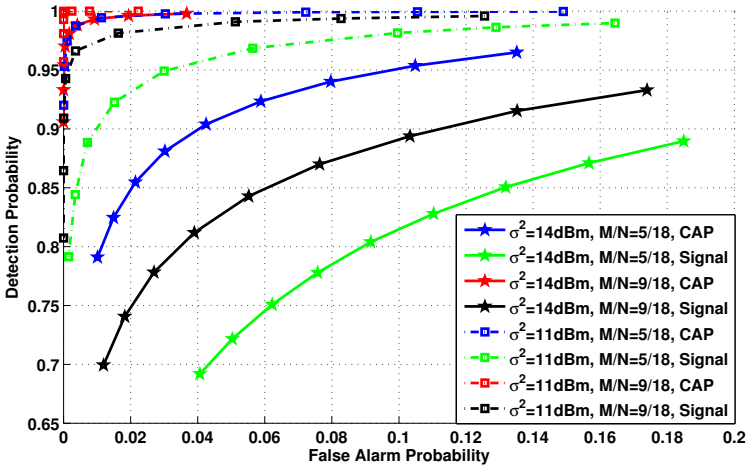


Figure 5.11: The resulting ROC of the detector when the CAP is used compared with the one when the compressive signal reconstruction using RM-FOCUSS of [1] is used (the sixth experiment).

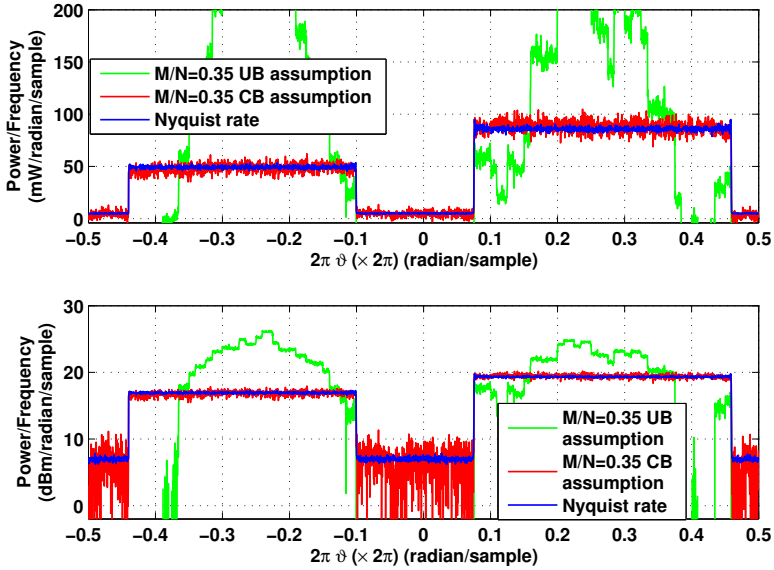


Figure 5.12: The CAP and the NAP of the faded user signals for the seventh experiment in Section 5.8.2 as a function of frequency in a linear scale (top) and logarithmic scale (bottom).

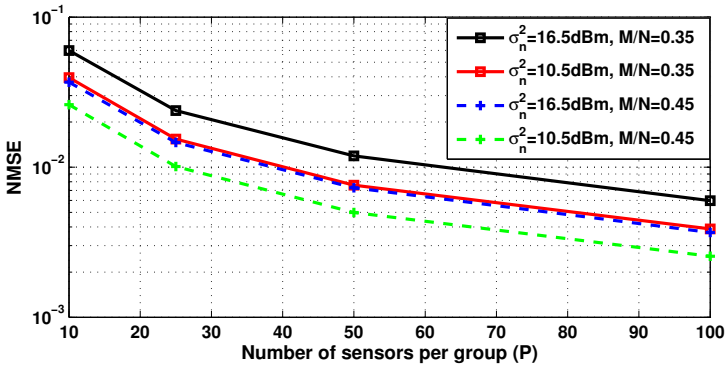


Figure 5.13: The NMSE between the CAP based on the correlated bins assumption and the NAP for the seventh experiment in Section 5.8.2.

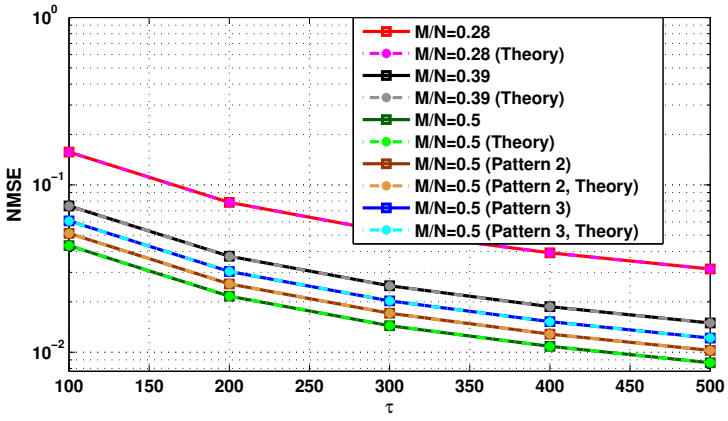


Figure 5.14: The simulated and analytical NMSE between the CAP and the true power spectrum when $x_t[\tilde{n}]$ only contains circular complex Gaussian i.i.d. noise. Unless mentioned otherwise, the cases of $M/N > 0.28$ are implemented by activating extra cosets based on Pattern 1.

Table 5.6: Three coset patterns to be added on top of the already selected minimal circular sparse ruler based coset indices for implementing $M/N > 0.28$ in Section 5.8.3.

Coset pattern	The order of the additional coset indices
Pattern 1	17, 11, 2, 6
Pattern 2	3, 5, 6, 8
Pattern 3	2, 3, 5, 6

Chapter 6

Compressive Joint Angular-Frequency Power Spectrum Estimation

Dyonisius Dony Ariananda and Geert Leus

Abstract

We introduce a new compressive power spectrum estimation approach in both frequency and direction of arrival (DOA). Wide-sense stationary signals produced by multiple uncorrelated sources are compressed in both the time and spatial domain where the latter compression is implemented by activating only some of the antennas in the underlying uniform linear array (ULA). We sample the received signal at every active antenna at sub-Nyquist rate, compute both the temporal and spatial correlation functions between the sub-Nyquist rate samples, and apply least squares to reconstruct the full-blown two-dimensional power spectrum matrix where the rows and columns correspond to the frequencies and the angles, respectively. This is possible under the full column rank condition of the system matrices and without applying any sparsity constraint on the signal statistics. Further, we can estimate the DOAs of the sources by locating the peaks of the angular power spectrum. We can theoretically estimate the frequency bands and the DOAs of more uncorrelated sources than active sensors using sub-Nyquist sampling.

6.1 Introduction

Compressive sampling and multi-coset sampling have drawn a lot of interest from the signal processing community due to the possibility to reconstruct a signal sampled at sub-Nyquist rate with no or little information loss under the constraint that the signal is sparse in a particular basis [5, 38]. All these works on sub-Nyquist sampling are important especially when it is needed to relax the requirements on the analog-to-digital converters (ADCs). For a wide-sense stationary (WSS) signal, it has also been shown that perfect reconstruction of its second-order statistics from sub-Nyquist rate samples is theoretically possible even without sparsity constraint [94]. This invention is important for some applications, such as wideband spectrum sensing for cognitive radio, where only perfect reconstruction of the temporal auto-correlation function is required instead of the signal itself. The principle of reconstructing the temporal auto-correlation function of a signal from the time-domain compressive measurements has in a dual form also been proposed in the spatial domain. Given a linear antenna array, [42] and [41] show that if the locations of the antennas are arranged according to a nested or coprime array, the spatial correlation values between the outputs of the antennas in the array can be used to generate the spatial correlation values between the outputs of the antennas in the virtual array or difference co-array (which is uniform in this case) which gener-

ally has more antennas and a larger aperture than the actual array. This enhances the degrees of freedom and allows [42] and [41] to estimate the direction of arrival (DOA) of more uncorrelated sources than sensors. The minimum redundancy array (MRA) of [12] can also be used to produce this feature but in a more optimal way. This has been exploited by [10] to perform compressive angular power spectrum reconstruction. The advantage offered by the nested and coprime arrays over the MRA however, is the possibility to derive a closed-form expression for the array geometry and the achievable number of correlation values in the resulting uniform difference co-array. In the aforementioned concept, the spatial compression is performed in the sense that we select a subset of antennas from a uniform linear array (ULA).

In this paper, we jointly reconstruct both the frequency-domain and angular-domain power spectrum using compressive samples. We use a ULA as the underlying array and activate only some of its antennas leading to a spatial-domain compression. The received signal at each active antenna is then sampled at sub-Nyquist-rate using multi-coset sampling. Next, we compute all the correlation values between the resulting sub-Nyquist rate samples at all active antennas both in the time domain and the spatial domain and use them to reconstruct the two-dimensional (2D) power spectrum matrix where each row gives the power spectrum in the frequency domain for a given angle and where each column contains the power spectrum in the angular domain for a given frequency. Further, we can estimate the DOA of the sources active at each frequency by locating the peaks in the angular power spectrum. This 2D power spectrum reconstruction can be done for more uncorrelated sources than active sensors without any sparsity constraint on the true power spectrum.

6.2 Preliminaries

First, consider a ULA having N_s antennas receiving signals from K uncorrelated WSS sources. We assume that the distance between the sources and the ULA is large enough compared to the length of the ULA and thus the wave incident on the ULA is assumed to be planar and the sources can be assumed as point sources. We also assume that the inverse of the bandwidth of the aggregated incoming signals is larger than the propagation delay across the ULA, which allows us to represent the delay between the antennas as a phase shift. Based on these assumptions, we can

write the ULA output as

$$\mathbf{x}(t) = \sum_{q=1}^{Q} \mathbf{a}(\theta_q) s_q(t) + \mathbf{n}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) \tag{6.1}$$

where $\mathbf{x}(t)$ is the $N_s \times 1$ output vector containing the received signal at the N_s antennas of the ULA, $\mathbf{n}(t)$ is the $N_s \times 1$ additive white Gaussian noise vector, $\mathbf{s}(t) = [s_1(t), s_2(t), \ldots, s_Q(t)]^T$ is the $Q \times 1$ extended source vector with $s_q(t)$ the incoming signal from the investigated angle θ_q, and $\mathbf{A} = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \ldots, \mathbf{a}(\theta_Q)]$ is the $N_s \times Q$ extended array manifold matrix with $\mathbf{a}(\theta_q)$ the $N_s \times 1$ array response vector containing the phase shifts experienced by $s_q(t)$ at each element of the ULA. Note that $\{\theta_q\}_{q=1}^{Q}$ is known and might only approximately contain the actual DOAs of the K sources. We generally assume that $\mathbf{n}(t)$ and $\mathbf{s}(t)$ are uncorrelated, that the impact of the wireless channel has been taken into account in $\mathbf{s}(t)$, and that the noises at different antennas are uncorrelated with variance σ_n^2, i.e., $E[\mathbf{n}(t)\mathbf{n}^H(t)] = \sigma_n^2 \mathbf{I}_{N_s}$, with $\mathbf{I}_{N_s}$ the $N_s \times N_s$ identity matrix. We consider the first element of the ULA as a reference point and express the array response vector $\mathbf{a}(\theta_q)$ as $\mathbf{a}(\theta_q) = [1, a(\theta_q)^d, a(\theta_q)^{2d}, \ldots, a(\theta_q)^{(N_s-1)d}]^T$, where $a(\theta_q) = \exp(j2\pi \sin(\theta_q))$ and d is the distance between two consecutive antennas in wavelengths, which is set to $d \leq 0.5$ in order to prevent spatial aliasing.

In order to simplify the further analysis, we introduce $\mathbf{x}[m] = \mathbf{x}(mT)$, $\mathbf{n}[m] = \mathbf{n}(mT)$, and $\mathbf{s}[m] = \mathbf{s}(mT)$ as a digital representation of $\mathbf{x}(t)$, $\mathbf{n}(t)$, and $\mathbf{s}(t)$, respectively, where $1/T$ is the Nyquist sampling rate at every ADC associated with each antenna. We then collect the output vectors $\mathbf{x}[m]$ at N_t consecutive sample indices into the $N_s \times N_t$ matrix $\mathbf{X}[n]$, for $n = 0, 1, \ldots, N_{n-1}$, as $\mathbf{X}[n] = [\mathbf{x}[nN_t], \mathbf{x}[nN_t + 1], \ldots, \mathbf{x}[(n+1)N_t - 1]]$ and write $\mathbf{X}[n]$ as

$$\mathbf{X}[n] = \mathbf{A}\mathbf{S}[n] + \mathbf{N}[n] \tag{6.2}$$

where $\mathbf{N}[n]$ is similarly defined as $\mathbf{X}[n]$ and the $Q \times N_t$ matrix $\mathbf{S}[n]$ is given by $\mathbf{S}[n] = [\mathbf{s}[nN_t], \mathbf{s}[nN_t + 1], \ldots, \mathbf{s}[(n+1)N_t - 1]]$. Let us also write the $Q \times 1$ vector $\mathbf{s}[nN_t + i]$ as $\mathbf{s}[nN_t + i] = [s_1[nN_t + i], s_2[nN_t + i], \ldots, s_Q[nN_t + i]]^T$ with $s_q[m] = s_q(mT)$ a digital representation of $s_q(t)$.

6.3 Time-Domain and Spatial-Domain Compression

In this section, we introduce the compression operations on the output matrix $\mathbf{X}[n]$ both in the spatial domain and time domain. The spatial-domain compression is

implemented by activating only M_s out of N_s available antennas in the ULA leading to a possibly non-ULA of less active antennas than sources. Further, in the receiver branches associated with the M_s active antennas, time-domain compression is performed by sampling the received analog signal at sub-Nyquist-rate using the multi-coset sampling principle discussed in [38], which can be implemented using the practical sampling device proposed in [94]. Here, the multi-coset sampling process is represented by selecting only M_t out of N_t time samples.

We first introduce the $M_s \times N_s$ spatial-domain selection matrix $\mathbf{C}_s$, which is formed by selecting M_s rows of $\mathbf{I}_{N_s}$. Here, the indices of the selected rows of $\mathbf{I}_{N_s}$ used to construct $\mathbf{C}_s$ correspond to the indices of the M_s active antennas selected from the N_s available antennas in the ULA. Based on (6.2), the $N_s \times N_t$ matrix $\mathbf{X}[n]$ is then compressed in the spatial-domain by $\mathbf{C}_s$ leading to the $M_s \times N_t$ matrix

$$\mathbf{Y}[n] = \mathbf{C}_s\mathbf{X}[n] \stackrel{\Delta}{=} \mathbf{B}\mathbf{S}[n] + \mathbf{M}[n] \tag{6.3}$$

where $\mathbf{Y}[n] = [\mathbf{y}[nN_t], \mathbf{y}[nN_t+1], \ldots, \mathbf{y}[(n+1)N_t-1]]$ with $\mathbf{y}[nN_t+l] = [y_1[nN_t+l], y_2[nN_t+l], \ldots, y_{M_s}[nN_t+l]]^T$, $\mathbf{B} = [\mathbf{b}(\theta_1), \mathbf{b}(\theta_2), \ldots, \mathbf{b}(\theta_Q)]$ is the $M_s \times Q$ array response matrix with $\mathbf{b}(\theta_q) = \mathbf{C}_s\mathbf{a}(\theta_q)$ the $M_s \times 1$ array response vector associated with the M_s activated antennas, the $M_s \times N_t$ matrix $\mathbf{M}[n]$ is given by $\mathbf{M}[n] = [\mathbf{m}[nN_t], \mathbf{m}[nN_t+1], \ldots, \mathbf{m}[(n+1)N_t-1]]$, and $\mathbf{m}[m]$ is the $M_s \times 1$ discrete noise vector given by $\mathbf{m}[m] = \mathbf{C}_s\mathbf{n}[m]$. Observe that $\mathbf{m}[m]$ generally has correlation matrix $E\left[\mathbf{m}[m]\mathbf{m}^H[m']\right] = \sigma_n^2\mathbf{I}_{M_s}\delta[m-m']$. The next step is to introduce the $M_t \times N_t$ time-domain selection matrix $\mathbf{C}_t$ formed by selecting M_t rows of the $N_t \times N_t$ identity matrix $\mathbf{I}_{N_t}$, and further compress $\mathbf{Y}[n]$ in (6.3) in the time domain, leading to the $M_s \times M_t$ matrix

$$\mathbf{Z}[n] = \mathbf{Y}[n]\mathbf{C}_t^T. \tag{6.4}$$

6.4 Power Spectrum Reconstruction

Denote the j-th row of $\mathbf{Z}[n]$ and $\mathbf{Y}[n]$ in (6.4) as $\mathbf{z}_j^T[n]$ and $\mathbf{y}_j^T[n]$, respectively, and write the $M_t \times 1$ vector $\mathbf{z}_j[n]$ in terms of its elements as $\mathbf{z}_j[n] = [z_{j,1}[n], z_{j,2}[n], \ldots, z_{j,M_t}[n]]^T$ and the $N_t \times 1$ vector $\mathbf{y}_j[n]$ as $\mathbf{y}_j[n] = [y_j[nN_t], y_j[nN_t+1], \ldots, y_j[(n+1)N_t-1]]^T$. This allows us to rewrite the time-domain compression in (6.4) in terms of the row vectors of $\mathbf{Z}[n]$ and $\mathbf{Y}[n]$, i.e.,

$$\mathbf{z}_j[n] = \mathbf{C}_t\mathbf{y}_j[n], \quad j = 1, 2, \ldots, M_s. \tag{6.5}$$

Using (6.5), our next step is to calculate the correlation matrix between $\mathbf{z}_i[n]$ and $\mathbf{z}_j[n]$ for all $i, j = 1, 2, \ldots, M_s$ as

$$\mathbf{R}_{z_i,z_j} = E\left[\mathbf{C}_t\mathbf{y}_i[n]\mathbf{y}_j[n]^H\mathbf{C}_t^H\right] = \mathbf{C}_t\mathbf{R}_{y_i,y_j}\mathbf{C}_t^H. \tag{6.6}$$

In practice, the expectation operator in (6.6) can be estimated by taking an average over N_n available matrices $\mathbf{Z}[n]$. After cascading all columns of $\mathbf{R}_{z_i,z_j}$ into the $M_t^2 \times 1$ vector vec$(\mathbf{R}_{z_i,z_j})$ and by taking into account the fact that $\mathbf{C}_t$ is a real matrix, we can express vec$(\mathbf{R}_{z_i,z_j})$ based on (6.6) as

$$\text{vec}(\mathbf{R}_{z_i,z_j}) = (\mathbf{C}_t \otimes \mathbf{C}_t)\text{vec}(\mathbf{R}_{y_i,y_j}) \tag{6.7}$$

where vec(.) is the operator that cascades all columns of a matrix into a single column vector and $\otimes$ represents the Kronecker product operation. Up to this stage, let us recall that $\{s_q(t)\}_{q=1}^{Q}$ in (6.1) are WSS processes since we have K WSS sources. Based on this fact, as well as (6.3), it is obvious that the elements of $\mathbf{y}_j[n]$ in (6.5) also form a WSS sequence. This means that the $N_t \times N_t$ matrix $\mathbf{R}_{y_i,y_j}$ in (6.7) has a Toeplitz structure allowing us to condense $\mathbf{R}_{y_i,y_j}$ into the $(2N_t - 1) \times 1$ vector $\mathbf{r}_{y_i,y_j} = [r_{y_i,y_j}[0], r_{y_i,y_j}[1], \ldots, r_{y_i,y_j}[N_t - 1], r_{y_i,y_j}[1 - N_t], \ldots, r_{y_i,y_j}[-1]]^T$ and write

$$\text{vec}(\mathbf{R}_{y_i,y_j}) = \mathbf{T}\mathbf{r}_{y_i,y_j} \tag{6.8}$$

where $\mathbf{T}$ is a special $N_t^2 \times (2N_t - 1)$ repetition matrix whose i-th row is given by the $((i - 1 + (N_t - 2)\left\lfloor \frac{i-1}{N_t} \right\rfloor) \bmod (2N_t - 1) + 1)$-th row of the identity matrix $\mathbf{I}_{2N_t-1}$. By combining (6.7) and (6.8), we obtain

$$\text{vec}(\mathbf{R}_{z_i,z_j}) = (\mathbf{C}_t \otimes \mathbf{C}_t)\mathbf{T}\mathbf{r}_{y_i,y_j} = \mathbf{R}_{c_t}\mathbf{r}_{y_i,y_j} \tag{6.9}$$

where $\mathbf{R}_{c_t} = (\mathbf{C}_t \otimes \mathbf{C}_t)\mathbf{T}$ is an $M_t^2 \times (2N_t - 1)$ matrix. Observe that it is possible to reconstruct $\mathbf{r}_{y_i,y_j}$ from vec$(\mathbf{R}_{z_i,z_j})$ in (6.9), for all $i, j = 1, 2, \ldots, M_s$, using least squares (LS) if $M_t^2 \geq 2N_t - 1$ and $\mathbf{R}_{c_t}$ has full column rank.

The next step is to figure out the relationship between $\{\mathbf{r}_{y_i,y_j}\}_{i,j=1}^{M_s}$ in (6.9) and the extended source matrix $\mathbf{S}[n]$ in (6.3). By taking into account the fact that every row of $\mathbf{Y}[n]$ and $\mathbf{S}[n]$ is a WSS sequence and the assumption that the extended source vector $\mathbf{s}[m]$ and the noise vector $\mathbf{m}[m]$ are uncorrelated, it is straightforward to find that the correlation matrix between $\mathbf{y}[nN_t + l]$ and $\mathbf{y}[nN_t + l']$ is given by

$$\mathbf{R}_y[l - l'] = \mathbf{B}\mathbf{R}_s[l - l']\mathbf{B}^H + \sigma_n^2\mathbf{I}_{M_s}\delta[l - l'] \tag{6.10}$$

for $l, l' = 0, 1, \ldots, N_t - 1$. Since the point sources are assumed to be uncorrelated, the elements of $\mathbf{s}[m]$ are also uncorrelated and thus the $Q \times Q$ matrix $\mathbf{R}_s[l - l']$ is a diagonal matrix. By exploiting this fact and stacking all columns of the $M_s \times M_s$ matrix $\mathbf{R}_y[l - l']$ in (6.10) into the $M_s^2 \times 1$ vector $\text{vec}(\mathbf{R}_y[l - l'])$, we obtain

$$\begin{aligned}\text{vec}(\mathbf{R}_y[l - l']) &= (\mathbf{B}^* \odot \mathbf{B})\text{diag}(\mathbf{R}_s[l - l']) \\ &+ \sigma_n^2 \text{vec}(\mathbf{I}_{M_s})\delta[l - l'], \quad l, l' = 0, 1, \ldots, N_t - 1,\end{aligned} \tag{6.11}$$

where $\odot$ represents the Khatri-Rao product operation. Let us now investigate the relationship between the elements of $\{\mathbf{r}_{y_i,y_j}\}_{i,j=1}^{M_s}$ in (6.9) and $\text{vec}(\mathbf{R}_y[l - l'])$ in (6.11). We can find that $\{\text{vec}(\mathbf{R}_y[l-l'])\}_{l,l'=0}^{N_t-1}$ is actually related to $\{\mathbf{r}_{y_i,y_j}\}_{i,j=1}^{M_s}$ as $\text{vec}(\mathbf{R}_y[l - l']) = [r_{y_1,y_1}[l - l'], r_{y_2,y_1}[l - l'], \ldots, r_{y_{M_s},y_{M_s}}[l - l']]^T$. Hence, we can use the elements of the reconstructed $\{\mathbf{r}_{y_i,y_j}\}_{i,j=1}^{M_s}$ in (6.9) to form $\{\text{vec}(\mathbf{R}_y[l - l'])\}_{l,l'=0}^{N_t-1}$ in (6.11) and then use them to reconstruct $\{\text{diag}(\mathbf{R}_s[l-l'])\}_{l,l'=0}^{N_t-1}$ in (6.11), which can be performed using LS if $M_s^2 \geq Q$ and $\mathbf{B}^* \odot \mathbf{B}$ has full column rank.

If we combine the $Q \times 1$ vectors $\{\text{diag}(\mathbf{R}_s[l-l'])\}_{l,l'=0}^{N_t-1}$ as $\bar{\mathbf{R}}_s = [\text{diag}(\mathbf{R}_s[0]), \text{diag}(\mathbf{R}_s[1]), \ldots, \text{diag}(\mathbf{R}_s[N_t - 1]), \text{diag}(\mathbf{R}_s[1 - N_t]), \ldots, \text{diag}(\mathbf{R}_s[-1])]$, we can observe that the q-th row of $\bar{\mathbf{R}}_s$ actually corresponds to the temporal auto-correlation of the incoming signal from the investigated angle θ_q, which can be written as $\mathbf{r}_{s_q}^T = [r_{s_q}[0], r_{s_q}[1], \ldots, r_{s_q}[N_t - 1], r_{s_q}[1 - N_t], \ldots, r_{s_q}[-1]]$. By defining $\mathbf{F}_{2N_t-1}$ as the $(2N_t-1)\times(2N_t-1)$ discrete Fourier transform (DFT) matrix, we can compute the power spectrum of $s_q[m]$ as $\mathbf{p}_{s_q} = \mathbf{F}_{2N_t-1}\mathbf{r}_{s_q}$, where $\mathbf{p}_{s_q}$ is the $(2N_t - 1) \times 1$ power spectrum vector of the incoming signal from the investigated angle θ_q. By combining $\{\mathbf{p}_{s_q}\}_{q=1}^{Q}$ into the $Q \times (2N_t - 1)$ matrix $\bar{\mathbf{P}}_s = [\mathbf{p}_{s_1}, \mathbf{p}_{s_2}, \ldots, \mathbf{p}_{s_Q}]^T$, we can write

$$\bar{\mathbf{P}}_s = \bar{\mathbf{R}}_s \mathbf{F}_{2N_t-1}. \tag{6.12}$$

Note that $\bar{\mathbf{P}}_s$ in (6.12) can be perceived as a 2D power spectrum matrix where every row of $\bar{\mathbf{P}}_s$ gives the power spectrum in the frequency-domain for a given investigated angle and every column of $\bar{\mathbf{P}}_s$ provides the power spectrum information in the angular domain for a given frequency.

6.5 Construction of the Compression Matrices

Recall that the 2D power spectrum matrix $\bar{\mathbf{P}}_s$ can be reconstructed from $\text{vec}(\mathbf{R}_{z_i,z_j})$ in (6.9), which contains the cross-correlations between the rows of the measurement matrix $\mathbf{Z}[n]$ in (6.4), by solving (6.9) and (6.11) using LS and then applying the DFT on the rows of the resulting matrix $\bar{\mathbf{R}}_s$. We now discuss the choice of

the selection matrix $\mathbf{C}_t$ and the extended array response matrix $\mathbf{B}$ that ensure the uniqueness of the LS solution of (6.9) and (6.11), respectively.

We first investigate the choice of $\mathbf{C}_t$ that results in a full column rank matrix $\mathbf{R}_{c_t}$. Since the rows of $\mathbf{C}_t$ and $\mathbf{T}$ in (6.9) are formed by selecting the rows of the identity matrix, it is clear that every row of both $\mathbf{C}_t \otimes \mathbf{C}_t$ and $\mathbf{T}$ only contains a single one and zeros elsewhere. This fact guarantees that each row of $\mathbf{R}_{c_t}$ has only a single one and thus, in order to ensure the full column rank condition of $\mathbf{R}_{c_t}$, we need to ensure that each column of it has at least a single one. This problem actually has been encountered and solved in [94] where the solution is to construct $\mathbf{C}_t$ by selecting the rows of $\mathbf{I}_{N_t}$ based on the so-called minimal length-$(N_t - 1)$ sparse ruler problem. In practice, this results in a multi-coset sampling procedure called the minimal sparse ruler sampling [94].

Next, we examine the choice of $\mathbf{B}$, which boils down to the selection of the activated antennas in the ULA and the investigated angles $\{\theta_q\}_{q=1}^{Q}$. Let us write $\mathbf{B}^* \odot \mathbf{B}$ in terms of $\{\mathbf{b}(\theta_q)\}_{q=1}^{Q}$ as

$$\mathbf{B}^* \odot \mathbf{B} = [\mathbf{b}^*(\theta_1) \otimes \mathbf{b}(\theta_1), \ldots, \mathbf{b}^*(\theta_Q) \otimes \mathbf{b}(\theta_Q)] \tag{6.13}$$

and $\mathbf{b}(\theta_q)$ in terms of $a(\theta_q)$ as

$$\mathbf{b}(\theta_q) = \left[a(\theta_q)^{d_1}, a(\theta_q)^{d_2}, \ldots, a(\theta_q)^{d_{M_s}}\right]^T \tag{6.14}$$

where d_i is the distance in wavelengths between the i-th *active* antenna and the reference antenna of the ULA defined in Section 6.2. It is clear from (6.13) and (6.14) that the q-th column of $\mathbf{B}^* \odot \mathbf{B}$ contains the elements $\exp\left(j(d_i - d_j)2\pi\sin(\theta_q)\right)$, for $i, j = 1, 2, \ldots, M_s$. While our task to find general design conditions to guarantee the full column rank of $\mathbf{B}^* \odot \mathbf{B}$ is not trivial, the following theorem suggests one possible way to achieve a full column rank $\mathbf{B}^* \odot \mathbf{B}$.

Theorem 6.5.1. *The matrix $\mathbf{B}^* \odot \mathbf{B}$ has full column rank if: 1) There exist Q distinct values of θ_q satisfying $-\frac{\pi}{2} < \{\theta_q\}_{q=1}^{Q} \leq \frac{\pi}{2}$, and 2) There exists an integer $N_v \geq Q$ such that $\{d_i - d_j\}_{i,j=1}^{M_s}$ contains an arithmetic sequence of N_v terms having a difference of $d \leq 0.5$ between each two consecutive terms.*

The proof of Theorem 6.5.1 can be found in Appendix 6.A. The second condition indicates that there exist N_v distinct rows from $\mathbf{B}^* \odot \mathbf{B}$ that form the array response matrix of a virtual ULA with N_v antennas, which can only be achieved for $N_v \leq 2N_s - 1$. This second condition also implies that we have more antennas in this virtual ULA than investigated angles. Some possible ways to satisfy Theorem 6.5.1

is to select the M_s active antennas from the N_s antennas in the ULA based on the MRA discussed in [12] (which also obeys the minimal sparse ruler problem [10]), the two-level nested array [42], or the coprime array [41]. For the MRA and the two-level nested array, Theorem 6.5.1 can be satisfied even for $N_v = 2N_s - 1$. Note that although the Q different values of θ_q can be chosen in an arbitrary fashion, they should not be too close to each other, since otherwise the resulting $\mathbf{B}^* \odot \mathbf{B}$ might be ill-conditioned. Theorem 6.5.1 also implies that the maximum number of detectable sources is upper bounded by $K \leq 2N_s - 1$ since we cannot detect more than Q sources. Apart from satisfying Theorem 6.5.1, another way to achieve a full column rank $\mathbf{B}^* \odot \mathbf{B}$ is suggested by Theorem 6.5.2.

Theorem 6.5.2. *The matrix $\mathbf{B}^* \odot \mathbf{B}$ has full column rank if: 1) $\{(d_i - d_j) \bmod \frac{Q}{2}\}_{i,j=1}^{M_s}$ has at least Q different values and 2) the grid of investigated angles $\{\theta_q\}_{q=1}^{Q}$ is designed based on the inverse sinusoidal angular grid where*

$$\theta_q = sin^{-1}\left(\frac{2}{Q}\left(q - 1 - \left\lceil \frac{Q-1}{2} \right\rceil\right)\right). \tag{6.15}$$

The proof for this theorem can be found in Appendix 6.B. Note that the first condition from Theorem 6.5.2 is less strict than the second condition from Theorem 6.5.1. A good option is to use a configuration satisfying Theorem 6.5.1 with $N_v = 2N_s - 1$ and $d = 0.5$, and to use (6.15) with $Q = 2N_s - 1$. This will not only ensure that the resulting $M_s^2 \times (2N_s - 1)$ matrix $\mathbf{B}^* \odot \mathbf{B}$ has full column rank but also that there exists a $(2N_s - 1) \times (2N_s - 1)$ submatrix from $\mathbf{B}^* \odot \mathbf{B}$ that forms a row-permuted version of the $(2N_s - 1) \times (2N_s - 1)$ inverse DFT matrix, meaning that $\mathbf{B}^* \odot \mathbf{B}$ is well-conditioned.

6.6 Numerical Study

In this section, we examine the proposed approach with some numerical study. We consider a ULA having $N_s = 36$ antennas as the underlying array and construct an MRA of active antennas by selecting the antenna indices based on the minimal length-35 sparse ruler problem discussed in [94, 10]. This leads to $M_s = 10$ activated antennas with $\{d_j\}_{j=1}^{10} = \{0, d, 4d, 10d, 16d, 22d, 28d, 30d, 33d, 35d\}$ where d is set to $d = 0.5$. The set of investigated angles $\{\theta_q\}_{q=1}^{Q}$ is set according to (6.15) with $Q = 2N_s - 1 = 71$. In the receiver branch corresponding to each active antenna, the time-domain compression rate of $M_t/N_t = 0.4048$ is obtained by setting $N_t = 84$ and $M_t = 34$. We construct the 34×84 selection matrix $\mathbf{C}_t$ by first solving the minimal length-83 sparse ruler problem which gives the indices of the 16

rows of $\mathbf{I}_{N_t}$ that have to be selected. The selection of these 16 rows will ensure that the resulting matrix $\mathbf{R}_{c_t}$ in (6.9) has at least a single one in each column. The additional 18 rows of $\mathbf{C}_t$ are then randomly selected from the remaining rows of $\mathbf{I}_{N_t}$ that have not been selected. We simulate the case when we have more sources than active antennas by generating $K = 12$ uncorrelated sources having DOAs with 9 degrees of separation, i.e., the set of DOAs is given by $\{-54^0, -45^0, \ldots, 45^0\}$. The sources produce complex baseband signals whose frequency bands are given in Table 6.1 and which are generated by passing circular complex zero-mean Gaussian i.i.d. noise with variance $\sigma^2 = 5$ into a digital filter of length $N_t = 84$ with the unit-gain passband of the filter for each source set according to Table 6.1. This will ensure that the true auto-correlation sequence for each source is limited to $-N_t + 1 \leq m \leq N_t - 1$. We assume a spatially and temporally white noise with variance $\sigma_n^2 = 5$ and set the number of measurement matrices $\mathbf{Z}[n]$ to $N_n = 5951$.

Table 6.1: The frequency band occupied by the sources

Source	Actual DOA	Occupied frequency band
1	-54^0	$[-0.275\pi, -0.2\pi]$
2	-45^0	$[-0.8\pi, -0.725\pi]$
3	-36^0	$[-0.35\pi, -0.275\pi]$
4	-27^0	$[0.35\pi, 0.425\pi]$
5	-18^0	$[0.875\pi, 0.95\pi]$
6	-9^0	$[0.05\pi, 0.125\pi]$
7	0^0	$[-0.95\pi, -0.875\pi]$
8	9^0	$[-0.65\pi, -0.575\pi]$
9	18^0	$[-0.425\pi, -0.35\pi]$
10	27^0	$[0.575\pi, 0.65\pi]$
11	36^0	$[0.125\pi, 0.2\pi]$
12	45^0	$[0.5\pi, 0.575\pi]$

Fig. 6.1 illustrates the estimate of the power spectrum as a function of the frequency and the investigated angles. It is clear that the 12 uncorrelated sources can generally be detected. We can find the DOA estimates by locating the peak of this spectrum though the actual DOAs might not fall on top of the defined investigated angles. For a given DOA estimate, we can locate the active frequency band of the corresponding source together with the value of the power spectrum estimate. The top view of Fig. 6.1, which is provided by Fig. 6.2, gives a much clearer picture of

the quality of the estimate. We can easily compare this figure with the data provided in Table 6.1. Observe that the estimate of the DOA, the power spectrum, as well as the active frequency band of the sources is quite satisfactory except for the sources with DOAs -9^0 and 9^0. For these two sources, it is apparent from Fig. 6.2 that the impact of the grid mismatch effect is quite significant and their power spectrum estimates seem to have been distributed among the two nearest grid points. Note that this 2D power spectrum estimate can be produced without applying any sparsity contraint on the true power spectrum, but can of course be improved if such a constraint is used.

Appendix

6.A Proof of Theorem 6.5.1

The second requirement of Theorem 6.5.1 implies that there exists a $Q \times Q$ matrix $\acute{\mathbf{B}} = [\acute{\mathbf{b}}(\theta_1), \acute{\mathbf{b}}(\theta_2), \ldots, \acute{\mathbf{b}}(\theta_Q)]$, which is a submatrix of $\mathbf{B}^* \odot \mathbf{B}$ in (6.13), that forms the array response matrix of a virtual ULA of Q antennas with $\acute{\mathbf{b}}(\theta_q)$ given by $\acute{\mathbf{b}}(\theta_q) = [a(\theta_q)^{\bar{d}}, a(\theta_q)^{\bar{d}+d}, \ldots, a(\theta_q)^{\bar{d}+(Q-1)d}]^T$, where $\bar{d}$ gives the distance between the first antenna in the virtual ULA and the reference antenna in the underlying ULA in Section 6.2. Hence, it is clear that $\acute{\mathbf{B}}$ is a column-wise Vandermonde matrix. From the well-known properties of a column-wise Vandermonde matrix, $\acute{\mathbf{B}}$ has full column rank due to the first requirement of Theorem 6.5.1 and since $d \leq 0.5$. It is then trivial to show that $\mathbf{B}^* \odot \mathbf{B}$ also has full column rank.

6.B Proof of Theorem 6.5.2

Based on (6.13) and (6.14) and the fact that the inverse sinusoidal angular grid in (6.15) is used, we can write $\mathbf{B}^* \odot \mathbf{B}$ in terms of its row vectors, i.e., $\mathbf{B}^* \odot \mathbf{B} = [\boldsymbol{\beta}(d_1 - d_1), \boldsymbol{\beta}(d_2 - d_1), \ldots, \boldsymbol{\beta}(d_{M_s} - d_{M_s})]^T$, with $\boldsymbol{\beta}(d_i - d_j)$ given by $\boldsymbol{\beta}(d_i - d_j) = [e^{j\frac{4\pi}{Q}(d_i - d_j)(-\lceil \frac{Q-1}{2} \rceil)}, \ldots, e^{j\frac{4\pi}{Q}(d_i - d_j)(-1)}, 1, e^{j\frac{4\pi}{Q}(d_i - d_j)}, \ldots$
$, e^{j\frac{4\pi}{Q}(d_i - d_j)(Q-1-\lceil \frac{Q-1}{2} \rceil)}]^T$. Observe that $\mathbf{B}^* \odot \mathbf{B}$ is a row-wise Vandermonde matrix since the elements of $\boldsymbol{\beta}(d_i - d_j)$ are ordered according to geometric progression. In order to ensure that $\mathbf{B}^* \odot \mathbf{B}$ has full column rank, we need Q distinct values of $\frac{4\pi}{Q}(d_i - d_j)$ modulo 2π which is guaranteed by the first requirement of Theorem 6.5.2.

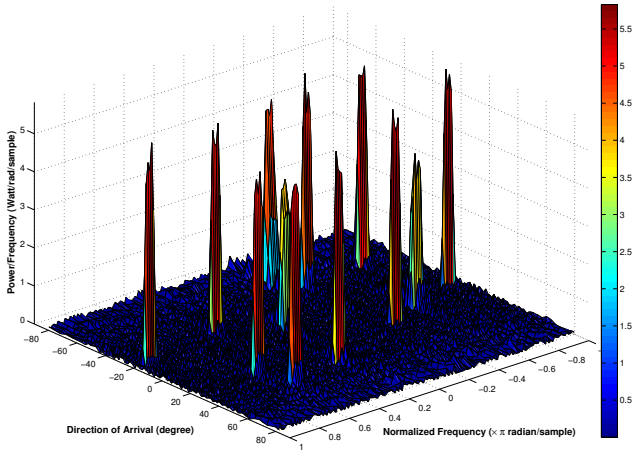


Figure 6.1: The power spectrum estimate (in watt/radian/sample) as a function of frequency (radian/sample) and angle (degree).

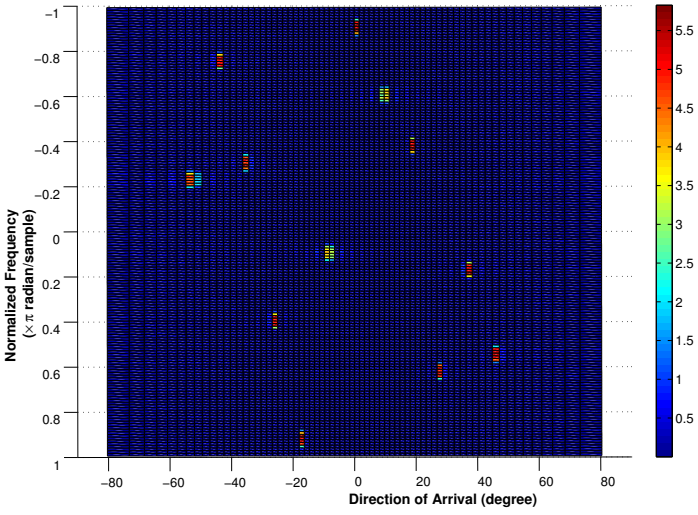


Figure 6.2: The top view of Fig. 6.1.

Chapter 7

Non-Uniform Sampling for Compressive Cyclic Spectrum Reconstruction

Dyonisius Dony Ariananda and Geert Leus

Abstract

We introduce a new cyclic spectrum estimation method for wide-sense cyclostationary (WSCS) signals sampled at sub-Nyquist rate using non-uniform sampling. We exploit the block Toeplitz structure of the WSCS signal correlation matrix and write the linear relationship between this matrix and the correlations of the sub-Nyquist rate samples as an overdetermined system. We find the condition under which the system matrix has full column rank allowing for least-squares reconstruction of the WSCS signal correlation matrix from the correlations of the compressive measurements. We also evaluate the case when the support of the WSCS signal correlation is limited and look at a special case where each selection matrix is restricted to either an identity matrix or an empty matrix. In the latter case, we can connect the full column rank condition of the system matrix with a circular sparse ruler.

7.1 Introduction and Related Works

Many researches have been done in the field of compressive sampling a.k.a. sub-Nyquist rate sampling due to the desire to relax the requirements on the analog-to-digital converters while maintaining the possibility for signal reconstruction with no or little information loss. This is possible under the constraint that the signal is sparse in a certain basis [5, 38]. Some applications, such as wideband spectrum sensing for cognitive radio networks, however, require perfect reconstruction of only the power spectrum or cyclic spectrum, instead of the signal itself. Perfect power spectrum reconstruction from sub-Nyquist rate samples has been shown to be possible for a wide-sense stationary (WSS) signal [94, 43] and for a multiband signal with uncorrelated spectra at different bands [44]. This can be performed even without applying a sparsity constraint on the actual power spectrum.

Since a stationary process can be perceived as a special case of a cyclostationary process, which is a process whose statistical characteristics vary periodically with time, the reconstruction of the power spectrum of a WSS signal can be treated as a special case of the reconstruction of the cyclic spectrum of a wide-sense cyclostationary (WSCS) signal. However, while compressively reconstructing the power spectrum of a WSS signal without applying a sparsity constraint on the power spectrum is possible due to the Toeplitz structure in the temporal correlation matrix, it is challenging to find a special structure in the WSCS signal correlation matrix that can be exploited to perform compression. This complicates the cyclic spectrum

reconstruction from sub-Nyquist rate samples of a WSCS signal and forces [45] to assume a sparse two-dimensional cyclic spectrum. A similar problem is found in [97], which focuses on only a real multiband signal and arrives at the assumption that the correlation matrix of the entries at different bands has nonzero values only at the diagonal and anti-diagonal elements. A different approach in [11] views the compressive measurements as random linear projections of the original signal and sets the span of the random linear projections equal to an integer multiple of the cyclic period. This allows [11] to exploit the block Toeplitz structure in the correlation matrix and to perform compression yet to present their reconstruction problem as an overdetermined system. The work of [11] does not focus on multicoset or non-uniform sampling and thus, it does not attempt to find the condition of the system matrix that allows for a least-squares (LS) solution for the overdetermined system. In this paper, we also set the span of the random linear projections equal to an integer multiple of the cyclic period but we focus only on non-uniform sampling. We express the correlations of the sub-Nyquist rate samples as a linear function of the correlation matrix of the corresponding Nyquist-rate samples. We find the condition for the system matrix to have full rank, which enables the LS reconstruction of the correlation matrix of the WSCS signal from the correlations of the compressive measurements. The cyclic spectrum can then be estimated from the reconstructed correlation matrix of the WSCS signal.

7.2 System Model and Compression

Let us consider a discrete WSCS signal $x[t]$, where the autocorrelation sequence $r_x[t,\tau] = E\{x[t]x^*[t-\tau]\}$ is periodic in t with a period of T. The cyclic autocorrelation sequence and the cyclic power spectrum of $x[t]$ are then respectively given by

$$\tilde{r}_x[f,\tau] = \frac{1}{T}\sum_{t=0}^{T-1} r_x[t,\tau]e^{-j2\pi f(t-\tau/2)/T}, \tag{7.1a}$$

$$s_x[f,\phi] = \sum_{\tau=-\infty}^{\infty} \tilde{r}_x[f,\tau]e^{-j2\pi\phi\tau}, \tag{7.1b}$$

with $\phi \in [0,1)$ the frequency and $f \in \{0,1,\ldots,T-1\}$ the cyclic frequency. Note that when $x[t]$ is produced by Nyquist-rate sampling at a rate of f_s Hz, f and ϕ correspond to an actual cyclic frequency of $f\frac{f_s}{T}$ Hz and an actual frequency of ϕf_s Hz, respectively. We cascade T consecutive samples $x[t]$ in $\mathbf{x}[n] = [x[nT], x[nT+$

$1], \ldots, x[nT+T-1]]^T$, which is a sequence of stationary vectors with $T \times T$ correlation matrix sequence $\mathbf{R}_x[k] = E\left\{\mathbf{x}[n]\mathbf{x}^H[n-k]\right\} = [r_x[t, kT+t-\tau]]_{t,\tau}$. Note that a one-to-one mapping exists between $r_x[t,\tau]$ and $\mathbf{R}_x[k]$. Next, we stack N consecutive $T \times 1$ stationary vectors $\mathbf{x}[n]$ into an $NT \times 1$ vector $\tilde{\mathbf{x}}[\tilde{n}]$ as $\tilde{\mathbf{x}}[\tilde{n}] = [\mathbf{x}^T[\tilde{n}N], \mathbf{x}^T[\tilde{n}N+1], \ldots, \mathbf{x}^T[\tilde{n}N+N-1]]^T$. The $NT \times NT$ correlation matrix of $\tilde{\mathbf{x}}[\tilde{n}]$ at lag 0 is then given by $\mathbf{R}_{\tilde{x}}[0] = E[\tilde{\mathbf{x}}[\tilde{n}]\tilde{\mathbf{x}}^H[\tilde{n}]]$, whose relationship to $\mathbf{R}_x[k]$ is given by

$$\mathbf{R}_{\tilde{x}}[0] = \begin{bmatrix} \mathbf{R}_x[0] & \mathbf{R}_x[-1] & \ldots & \mathbf{R}_x[-N+1] \\ \mathbf{R}_x[1] & \mathbf{R}_x[0] & \ldots & \mathbf{R}_x[-N+2] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_x[N-1] & \mathbf{R}_x[N-2] & \ldots & \mathbf{R}_x[0] \end{bmatrix} \tag{7.2}$$

which has a block Toeplitz structure. This allows us to perform a temporal compression by introducing an $\tilde{M} \times 1$ vector $\tilde{\mathbf{y}}[\tilde{n}] = [\mathbf{y}^T[\tilde{n}N], \mathbf{y}^T[\tilde{n}N+1], \ldots, \mathbf{y}^T[\tilde{n}N+N-1]]^T$, where $\mathbf{y}[\tilde{n}N+n]$ is an $M_n \times 1$ vector given by

$$\mathbf{y}[\tilde{n}N+n] = \mathbf{C}_n\mathbf{x}[\tilde{n}N+n],\ n = 0, 1, \ldots, N-1, \tag{7.3}$$

with $\tilde{M} = \sum_{n=0}^{N-1} M_n$ and $\mathbf{C}_n$ the $M_n \times T$ multi-coset sampling a.k.a. selection matrix whose rows are obtained by selecting the M_n rows of the $T \times T$ identity matrix $\mathbf{I}_T$. We can then write $\tilde{\mathbf{y}}[\tilde{n}]$ as

$$\tilde{\mathbf{y}}[\tilde{n}] = \tilde{\mathbf{C}}\tilde{\mathbf{x}}[\tilde{n}], \tag{7.4}$$

where $\tilde{\mathbf{C}}$ is an $\tilde{M} \times NT$ block diagonal matrix given by $\tilde{\mathbf{C}} = \text{diag}\{\mathbf{C}_0, \mathbf{C}_1, \ldots, \mathbf{C}_{N-1}\}$. As we will show in Section 7.3, for a certain value of n, it is possible that none of the rows of $\mathbf{I}_T$ is selected to form $\mathbf{C}_n$ (i.e., $M_n = 0$). In this case, for that particular n, $\mathbf{y}[\tilde{n}N+n] = \mathbf{C}_n = [\,]$, where $[\,]$ is an empty matrix, as none of the entries of $\mathbf{x}[\tilde{n}N+n]$ is selected. Note from (7.4) that we take $\tilde{M}$ random linear projections with a total span of NT (N times the period of the autocorrelation sequence $r_x[t,\tau]$) and attain compression by having $\tilde{M} < NT$.

Observe that $\mathbf{y}[\tilde{n}N+n]$ in (7.3) is generally not a sequence of stationary vectors since $\mathbf{C}_n$ is generally different for different values of n. However, we can obtain a sequence of stationary vectors by collecting $\mathbf{y}[\tilde{n}N+n]$ at different values of $\tilde{n}$ for a given n. We define the $M_n \times M_{n'}$ correlation matrix $\mathbf{R}_{y_{n,n'}} = E[\mathbf{y}[\tilde{n}N+n]\mathbf{y}^H[\tilde{n}N+n']]$, which can be written, by introducing $k = n - n'$, as

$$\mathbf{R}_{y_{n,n-k}} = \mathbf{C}_n\mathbf{R}_x[k]\mathbf{C}_{n-k}^H. \tag{7.5}$$

In practice, the computation of $\mathbf{R}_{y_{n,n-k}}$ in (7.5) must be approximated by taking an average over $\mathbf{y}[\tilde{n}N+n]$ at different indices $\tilde{n}$. We aim to reconstruct $\mathbf{R}_{\tilde{x}}[0]$ in (7.2), which is equivalent to reconstructing $\mathbf{R}_x[k]$ from $\{\mathbf{R}_{y_{n,n-k}}\}_{n=\max(0,k)}^{N-1+\min(0,k)}$ in (7.5) for $k = 0, \pm 1, \ldots, \pm(N-1)$. Let us now take the realness of $\mathbf{C}_n$ into account, stack the columns of $\mathbf{R}_{y_{n,n-k}}$ in (7.5) into an $(M_n M_{n-k}) \times 1$ vector $\text{vec}(\mathbf{R}_{y_{n,n-k}})$, and rewrite (7.5) as

$$\text{vec}(\mathbf{R}_{y_{n,n-k}}) = (\mathbf{C}_{n-k} \otimes \mathbf{C}_n)\text{vec}(\mathbf{R}_x[k]), \tag{7.6}$$

where $\text{vec}(.)$ is the operator that stacks all columns of a matrix into a column vector and $\otimes$ denotes the Kronecker product operation. In the event where we have either $\mathbf{C}_{n-k} = [\,]$ ($M_{n-k} = 0$) or $\mathbf{C}_n = [\,]$ ($M_n = 0$) in (7.6) for a particular n, the corresponding $\text{vec}(\mathbf{R}_{y_{n,n-k}})$ and $\mathbf{C}_{n-k} \otimes \mathbf{C}_n$ are given by $\text{vec}(\mathbf{R}_{y_{n,n-k}}) = \mathbf{C}_{n-k} \otimes \mathbf{C}_n = [\,]$. Stacking $\text{vec}(\mathbf{R}_{y_{n,n-k}})$ in (7.6) in ascending order of n, for all $n = \max(0,k), \ldots, N-1+\min(0,k)$, into a $\gamma_k \times 1$ vector $\mathbf{r}_{y,k}$ with $\gamma_k = \sum_{n=\max(0,k)}^{N-1+\min(0,k)} M_n M_{n-k}$, we can then express $\mathbf{r}_{y,k}$ as

$$\begin{aligned}\mathbf{r}_{y,k} &= \begin{bmatrix} \mathbf{C}_{\max(-k,0)} \otimes \mathbf{C}_{\max(0,k)} \\ \mathbf{C}_{\max(-k,0)+1} \otimes \mathbf{C}_{\max(0,k)+1} \\ \vdots \\ \mathbf{C}_{N-1+\min(-k,0)} \otimes \mathbf{C}_{N-1+\min(0,k)} \end{bmatrix} \text{vec}(\mathbf{R}_x[k]) \\ &= \mathbf{\Psi}_k \text{vec}(\mathbf{R}_x[k]), \quad k = 1-N, \ldots, N-2, N-1. \end{aligned} \tag{7.7}$$

7.3 Perfect Reconstruction

Observe that we can reconstruct $\text{vec}(\mathbf{R}_x[k])$ in (7.7) from $\mathbf{r}_{y,k}$ using LS as long as the $\gamma_k \times T^2$ matrix $\mathbf{\Psi}_k$ has full column rank. Hence, $\mathbf{R}_{\tilde{x}}[0]$ in (7.2) can be reconstructed from $\{\mathbf{r}_{y,k}\}_{k=1-N}^{N-1}$ using LS as long as $\{\mathbf{\Psi}_k\}_{k=1-N}^{N-1}$ all have full column rank. In order to simplify the discussion, we consider the following remark.

Remark 7.3.1. *The full column rank condition of $\mathbf{\Psi}_k$ can be achieved only if we have $\gamma_k \geq T^2$. In addition, observe in* (7.7) *that each row of $\mathbf{\Psi}_k$ has only a single one in one entry and zeros elsewhere since the rows of $\mathbf{C}_n$ in* (7.6) *are selected from the rows of $\mathbf{I}_T$. Hence, $\mathbf{\Psi}_k$ will have full column rank if and only if each of its columns has at least a single one.*

Let us now introduce the following definition.

Definition 7.3.1. *Define $\mathcal{C}_n$ as the set containing the indices of the rows of $\mathbf{I}_T$ used in $\mathbf{C}_n$. The set $\mathcal{C}_{n',n}$ is then defined as $\mathcal{C}_{n',n} = \{(i,j)|\forall i \in \mathcal{C}_{n'}, j \in \mathcal{C}_n\}$. Note that we generally have $\mathcal{C}_{n',n} \neq \mathcal{C}_{n,n'}$.*

We can then present the following lemma.

Lemma 7.3.1. *One row of $\mathbf{C}_{n'} \otimes \mathbf{C}_n$ will have a one in the $[(i-1)T+j]$-th entry and zeros elsewhere, if and only if $(i,j) \in \mathcal{C}_{n',n}$, i.e., $\mathbf{C}_{n'}$ contains the i-th row of $\mathbf{I}_T$ and $\mathbf{C}_n$ contains the j-th row of $\mathbf{I}_T$.*

Proof. The proof directly follows from the property of the Kronecker product operation. □

Based on Lemma 7.3.1, the full column rank condition of $\mathbf{\Psi}_k$ in (7.7) is provided by the following theorem.

Theorem 7.3.1. *$\mathbf{\Psi}_k$ in (7.7) has full column rank if and only if*

$$\Gamma_k \equiv \bigcup_{n=\max(0,k)}^{N-1+\min(0,k)} \mathcal{C}_{n-k,n} = \{(1,1),(1,2),\ldots,(1,T),(2,1),(2,2),\ldots,(T,T)\}. \tag{7.8}$$

Proof. Recall from Lemma 7.3.1 that if we have $(i,j) \in \mathcal{C}_{n-k,n}$, $\mathbf{C}_{n-k} \otimes \mathbf{C}_n$ will have a one in the $[(i-1)T+j]$-th column. We can then observe that satisfying (7.8) is equivalent to ensuring that the $[(i-1)T+j]$-th column of *at least one* of the matrices $\{\mathbf{C}_{n-k} \otimes \mathbf{C}_n\}_{n=\max(0,k)}^{N-1+\min(0,k)}$ contains a one in one entry and zeros elsewhere for all $i,j = 1,2,\ldots,T$. By taking the structure of $\mathbf{\Psi}_k$ in (7.7) into account, satisfying (7.8) also guarantees that every column of $\mathbf{\Psi}_k$ has at least a single one, which proves the sufficiency part of the theorem due to Remark 7.3.1. To prove the necessity part, assume that $\mathbf{\Psi}_k$ in (7.7) has full column rank but there is an (a,b) with $a,b \in \{1,2,\ldots,T\}$ such that $(a,b) \notin \Gamma_k$. Based on (7.8) and Lemma 7.3.1, this means that none of the matrices $\{\mathbf{C}_{n-k} \otimes \mathbf{C}_n\}_{n=\max(0,k)}^{N-1+\min(0,k)}$ has a one in the $[(a-1)T+b]$-th column. If we take the structure of $\mathbf{\Psi}_k$ in (7.7) into account, this implies that the $[(a-1)T+b]$-th column of $\mathbf{\Psi}_k$ only contain zeros. Using Remark 7.3.1, this indicates that the full column rank condition of $\mathbf{\Psi}_k$ is violated, which contradicts our initial assumption. □

Consider the case of $k = N-1$ (we have $\mathbf{\Psi}_{N-1} = \mathbf{C}_0 \otimes \mathbf{C}_{N-1}$ in (7.7)) and $k = -N+1$ (we have $\mathbf{\Psi}_{1-N} = \mathbf{C}_{N-1} \otimes \mathbf{C}_0$ in (7.7)). Observe that the size of both $\mathbf{\Psi}_{N-1}$ and $\mathbf{\Psi}_{1-N}$ is $M_0 M_{N-1} \times T^2$. For this specific case, we have the following theorem.

Theorem 7.3.2. *$\mathbf{\Psi}_{N-1}$ and/or $\mathbf{\Psi}_{1-N}$ will have full column rank if and only if $\mathbf{C}_0 = \mathbf{C}_{N-1} = \mathbf{I}_T$.*

Proof. The sufficiency part of this theorem is already clear as having $\mathbf{C}_0 = \mathbf{C}_{N-1} = \mathbf{I}_T$ leads to $\mathbf{\Psi}_{N-1} = \mathbf{\Psi}_{1-N} = \mathbf{I}_{T^2}$. The necessity part is shown for the full column rank condition of $\mathbf{\Psi}_{N-1}$ by considering Theorem 7.3.1, which requires $\mathcal{C}_{0,N-1} = \{(1,1),(1,2),\ldots,(1,T),(2,1),(2,2),\ldots,(T,T)\}$. This is identical to requiring $\mathcal{C}_0 = \mathcal{C}_{N-1} = \{1,2,\ldots,T\}$ due to Definition 7.3.1. The proof is concluded for $\mathbf{\Psi}_{N-1}$. The proof for $\mathbf{\Psi}_{1-N}$ follows the same steps. □

Since Theorem 7.3.2 requires us to have $\mathbf{C}_0 = \mathbf{C}_{N-1} = \mathbf{I}_T$, our task is to design $\{\mathbf{C}_n\}_{n=1}^{N-2}$ based on Theorem 7.3.1. One specific case occurs when we restrict $\mathbf{C}_n$ to either $\mathbf{C}_n = \mathbf{I}_T$ or $\mathbf{C}_n = [\,]$, for each $n = 1,2,\ldots,N-2$. In this case, $\mathbf{\Psi}_k$ in (7.7) must contain at least one self Kronecker product of $\mathbf{I}_T$ ($\mathbf{I}_T \otimes \mathbf{I}_T$) to preserve its full column rank condition. More precisely, for $k \in \{1-N,\ldots,N-1\}$, we need at least a pair of $n, n' \in \{0,1,\ldots,N-1\}$ with $n - n' = k$ and $\mathbf{C}_n = \mathbf{C}_{n'} = \mathbf{I}_T$ to ensure the full column rank condition of $\mathbf{\Psi}_k$. At this stage, let us review the concept of a linear sparse ruler discussed in [94, 100, 70].

Definition 7.3.2. *A length-$(N-1)$ linear sparse ruler is defined as a set $\mathcal{P} \subset \{0,1,\ldots,N-1\}$ such that $\{|p-p'||\forall p,p' \in \mathcal{P}\} = \{0,1,\ldots,N-1\}$. It is called minimal if no other linear sparse ruler of length $N-1$ exists with less elements.*

Using Definition 7.3.2, it is easy to show that the full column rank condition of $\{\mathbf{\Psi}_k\}_{k=1-N}^{N-1}$, when we restrict $\mathbf{C}_n$ to either $\mathbf{C}_n = \mathbf{I}_T$ or $\mathbf{C}_n = [\,]$ for each n, follows the following theorem.

Theorem 7.3.3. *Define W as the number of sampling matrices $\{\mathbf{C}_n\}_{n=0}^{N-1}$ that are set to $\mathbf{I}_T$, i.e., $\mathbf{C}_{n_w} = \mathbf{I}_T$, for $w = 0,1,\ldots,W-1$. The full column rank of all $\{\mathbf{\Psi}_k\}_{k=1-N}^{N-1}$ is ensured if and only if the set $\mathcal{W} = \{n_w | n_w \in \{0,1,\ldots,N-1\}, w = 0,1,\ldots,W-1\}$ is a linear sparse ruler.*

Under the constraint that we have either $\mathbf{C}_n = \mathbf{I}_T$ or $\mathbf{C}_n = [\,]$ for each $n = 0,1,\ldots,N-1$, it is of interest to obtain the strongest possible compression rate $\tilde{M}/NT$. This is equivalent to minimizing the cardinality of $\mathcal{W}$ in Theorem 7.3.3 under the condition that $\mathcal{W}$ is a length-$(N-1)$ linear sparse ruler. This boils down to a length-$(N-1)$ minimal linear sparse ruler problem [94, 100], whose solution minimizes $\tilde{M}/NT$ under the aforementioned constraint while maintaining the identifiability of $\{\text{vec}(\mathbf{R}_x[k])\}_{k=1-N}^{N-1}$ in (7.7). Once $\mathbf{R}_{\tilde{x}}[0]$ in (7.2) is reconstructed, we can reconstruct the cyclic autocorrelation sequence $\tilde{r}_x[f,\tau]$ from $\mathbf{R}_{\tilde{x}}[0]$ using (7.1a) and the cyclic power spectrum $s_x[f,\phi)$ from $\tilde{r}_x[f,\tau]$ using (7.1b).

7.4 Limited Correlation Support

Let us now assume that the support of $\mathbf{R}_x[k]$ in (7.5) is limited to $1-N \leq k \leq N-1$ and that the elements of $\mathbf{R}_x[k]$ are all close to 0 for $|k| \geq N$. In addition, unlike in the previous sections, we also exploit the correlation between $\tilde{\mathbf{y}}[\tilde{n}]$ in (7.4) and its neighboring blocks $\tilde{\mathbf{y}}[\tilde{n}+1]$ and $\tilde{\mathbf{y}}[\tilde{n}-1]$. While we have written $\mathbf{R}_{y_{n,n-k}}$ in (7.5) as $\mathbf{R}_{y_{n,n-k}} = E[\mathbf{y}[\tilde{n}N+n]\mathbf{y}^H[\tilde{n}N+n-k]]$ we can now also write it, for example, as $\mathbf{R}_{y_{n,n-k}} = E[\mathbf{y}[\tilde{n}N+n]\mathbf{y}^H[(\tilde{n}-1)N+n-k+N]]$ for $k>0$ or as $\mathbf{R}_{y_{n,n-k}} = E[\mathbf{y}[\tilde{n}N+n]\mathbf{y}^H[(\tilde{n}+1)N+n-k-N]]$ for $k<0$. By considering (7.3), we can now also rewrite (7.5) as

$$\mathbf{R}_{y_{n,n-k}} = \mathbf{C}_n \mathbf{R}_x[k] \mathbf{C}^H_{(n-k) \bmod N}, \tag{7.9}$$

with $n \bmod N$ the remainder of the integer division n/N. We stack the columns of $\mathbf{R}_{y_{n,n-k}}$ into $\text{vec}(\mathbf{R}_{y_{n,n-k}})$ as in Section 7.2 but we now cascade $\text{vec}(\mathbf{R}_{y_{n,n-k}})$, for all $n = 0, 1, \ldots, N-1$, in increasing order of n into a $\tilde{\gamma}_k \times 1$ vector $\tilde{\mathbf{r}}_{y,k}$ with $\tilde{\gamma}_k = \sum_{n=0}^{N-1} M_n M_{(n-k) \bmod N}$. We can then express $\tilde{\mathbf{r}}_{y,k}$ as

$$\begin{aligned}\tilde{\mathbf{r}}_{y,k} &= \begin{bmatrix} \mathbf{C}_{(-k) \bmod N} \otimes \mathbf{C}_0 \\ \mathbf{C}_{(1-k) \bmod N} \otimes \mathbf{C}_1 \\ \vdots \\ \mathbf{C}_{(N-1-k) \bmod N} \otimes \mathbf{C}_{N-1} \end{bmatrix} \text{vec}(\mathbf{R}_x[k]) \\ &= \tilde{\mathbf{\Psi}}_k \text{vec}(\mathbf{R}_x[k]), \quad k = 1-N, \ldots, N-2, N-1. \end{aligned} \tag{7.10}$$

We can again reconstruct $\text{vec}(\mathbf{R}_x[k])$ from $\tilde{\mathbf{r}}_{y,k}$ in (7.10) using LS where the condition that ensures a full column rank $\tilde{\mathbf{\Psi}}_k$ in (7.10) can be found by following the procedure in Section 7.3. By applying Remark 7.3.1 to $\tilde{\mathbf{\Psi}}_k$, using Definition 7.3.1 as well as Lemma 7.3.1, and following an analysis similar to the proof of Theorem 7.3.1, we can find that $\tilde{\mathbf{\Psi}}_k$ in (7.10) has full column rank if and only if

$$\begin{aligned}\tilde{\Gamma}_k \equiv \bigcup_{n=0}^{N-1} \mathcal{C}_{(n-k) \bmod N, n} = \{&(1,1),(1,2),\ldots,(1,T), \\ &(2,1),(2,2),\ldots,(T,T)\}. \end{aligned} \tag{7.11}$$

It is interesting to observe that, for this limited correlation support case, we do not have any condition similar to Theorem 7.3.2.

We again focus on the case where $\mathbf{C}_n$ is restricted to either $\mathbf{C}_n = \mathbf{I}_T$ or $\mathbf{C}_n = []$, for each $n = 0, 1, \ldots, N-1$. In order to maintain the full column rank condition

of $\tilde{\mathbf{\Psi}}_k$ in (7.10), $\tilde{\mathbf{\Psi}}_k$ must contain at least one self Kronecker product of $\mathbf{I}_T$, which is equivalent, for each $k \in \{1-N, \ldots, N-1\}$, to having at least a pair of $n, n' \in \{0, 1, \ldots, N-1\}$ with $(n - n') \bmod N = k$ and $\mathbf{C}_n = \mathbf{C}_{n'} = \mathbf{I}_T$. We now look at the concept of a circular sparse ruler discussed in [100].

Definition 7.4.1. *A circular sparse ruler of length $N-1$ is defined as a set $\mathcal{Q} \subset \{0, 1, \ldots, N-1\}$ such that $\{(q - q') \bmod N | \forall q, q' \in \mathcal{Q}\} = \{0, 1, \ldots, N-1\}$. It is called minimal if no other circular sparse ruler of length $N-1$ exists with less elements.*

Using Definition 7.4.1, it is obvious that the full column rank condition of $\{\tilde{\mathbf{\Psi}}_k\}_{k=1-N}^{N-1}$, when we restrict $\{\mathbf{C}_n\}_{n=0}^{N-1}$ to either $\mathbf{C}_n = \mathbf{I}_T$ or $\mathbf{C}_n = [\,]$, follows the following theorem.

Theorem 7.4.1. *Recall from Theorem 7.3.3 that $\mathcal{W} = \{n_w | n_w \in \{0, 1, \ldots, N-1\}, w = 0, 1, \ldots, W-1\}$ with $\{\mathbf{C}_{n_w}\}_{w=0}^{W-1} = \mathbf{I}_T$. The full column rank of all $\{\tilde{\mathbf{\Psi}}_k\}_{k=1-N}^{N-1}$ in (7.10) is guaranteed if the set $\mathcal{W}$ is a circular sparse ruler.*

Again, if we restrict $\mathbf{C}_n$ to either $\mathbf{C}_n = \mathbf{I}_T$ or $\mathbf{C}_n = [\,]$, for each $n = 0, 1, \ldots, N-1$, the best compression $\tilde{M}/NT$ is obtained by minimizing the cardinality of $\mathcal{W}$ in Theorem 7.4.1 under the condition that $\mathcal{W}$ is a length-$(N-1)$ circular sparse ruler, which boils down to a length-$(N-1)$ minimal circular sparse ruler problem [100].

7.5 Selection Matrix Construction

7.5.1 General Case

We focus on the formation of $\{\mathbf{C}_n\}_{n=0}^{N-1}$ for the general case discussed in Sections 7.2 and 7.3. Recall that Theorem 7.3.2 requires us to have $\mathbf{C}_0 = \mathbf{C}_{N-1} = \mathbf{I}_T$ to obtain full column rank $\mathbf{\Psi}_{N-1}$ and $\mathbf{\Psi}_{1-N}$. Also note from (7.7) that, for a larger k, less number of matrices $\{\mathbf{C}_n\}_{n=0}^{N-1}$ are contained in $\mathbf{\Psi}_k$. It is thus reasonable to construct $\{\mathbf{C}_n\}_{n=0}^{N-1}$ by evaluating the rank condition of $\mathbf{\Psi}_k$ starting from $k = N-1$ to $k = 0$. Note from (7.7) that the full column rank of $\mathbf{\Psi}_{-k}$ is ensured once $\mathbf{\Psi}_k$ has full column rank. Let us start by evaluating $\mathbf{\Psi}_{N-2} = [(\mathbf{C}_0 \otimes \mathbf{C}_{N-2})^T, (\mathbf{C}_1 \otimes \mathbf{C}_{N-1})^T]^T = [(\mathbf{I}_T \otimes \mathbf{C}_{N-2})^T, (\mathbf{C}_1 \otimes \mathbf{I}_T)^T]^T$. Assuming $\mathcal{C}_{N-2} = \{m_1, m_2, \ldots, m_{M_{N-2}}\}$ and $\mathcal{C}_1 = \{\tilde{m}_1, \tilde{m}_2, \ldots, \tilde{m}_{M_1}\}$, we have from Definition 7.3.1 $\mathcal{C}_{0,N-2} = \{(1, m_1), (1, m_2), \ldots, (1, m_{M_{N-2}}), (2, m_1), (2, m_2), \ldots, (T, m_{M_{N-2}})\}$ and $\mathcal{C}_{1,N-1} = \{(\tilde{m}_1, 1), (\tilde{m}_1, 2), \ldots, (\tilde{m}_1, T), (\tilde{m}_2, 1), (\tilde{m}_2, 2), \ldots, (\tilde{m}_{M_1}, T)\}$. Note from Theorem 7.3.1 that $\mathbf{\Psi}_{N-2}$ has full column rank if and

only if $\Gamma_{N-2} = \mathcal{C}_{0,N-2} \cup \mathcal{C}_{1,N-1}$ satisfies (7.8). Observe that this is possible only if we have at least one of $\mathbf{C}_{N-2}$ and $\mathbf{C}_1$ equal to $\mathbf{I}_T$. Once we set either $\mathbf{C}_{N-2}$ or $\mathbf{C}_1$ to $\mathbf{I}_T$, it is reasonable in this step to set the other one to $[\,]$ as we want to minimize the compression rate. We proceed to evaluate $\mathbf{\Psi}_{N-3} = [(\mathbf{C}_0 \otimes \mathbf{C}_{N-3})^T, (\mathbf{C}_1 \otimes \mathbf{C}_{N-2})^T, (\mathbf{C}_2 \otimes \mathbf{C}_{N-1})^T]^T = [(\mathbf{I}_T \otimes \mathbf{C}_{N-3})^T, (\mathbf{C}_1 \otimes \mathbf{C}_{N-2})^T, (\mathbf{C}_2 \otimes \mathbf{I}_T)^T]^T$. As we have set either $\mathbf{C}_{N-2}$ or $\mathbf{C}_1$ to $\mathbf{I}_T$ and the other one to $[]$, we have $\mathbf{C}_1 \otimes \mathbf{C}_{N-2} = []$ and $\mathbf{\Psi}_{N-3} = [(\mathbf{I}_T \otimes \mathbf{C}_{N-3})^T, (\mathbf{C}_2 \otimes \mathbf{I}_T)^T]^T$. Using the same analysis used when we considered $\mathbf{\Psi}_{N-2}$, we clearly have to set either $\mathbf{C}_{N-3}$ or $\mathbf{C}_2$ to $\mathbf{I}_T$. Again, it is reasonable to set the other matrix to $[\,]$.

At this stage, it can be found that we might face two possibilities with respect to the next considered $\mathbf{\Psi}_k$. First, we might be again required to set one of two selection matrices to $\mathbf{I}_T$ in order to ensure the full column rank of the next considered $\mathbf{\Psi}_k$. Second, we might have an option to set two selection matrices to $[\,]$ while maintaining the full column rank of the next considered $\mathbf{\Psi}_k$. These two possibilities are repeatedly faced for every considered $\mathbf{\Psi}_k$ as long as we set one selection matrix to $\mathbf{I}_T$ and the other one to $[\,]$, when the first one occurs, and we set both selection matrices to $[\,]$, when the second one occurs. If this procedure is followed, we can find that each of $\{\mathbf{C}_n\}_{n=0}^{N-1}$ is equal to either $\mathbf{I}_T$ or $[\,]$, which is the constraint that we considered when we formulated Theorem 7.3.3. Considering that this procedure is a reasonable way to minimize the compression rate, we suggest to design $\{\mathbf{C}_n\}_{n=0}^{N-1}$ by following Theorem 7.3.3 and minimizing the cardinality of $\mathcal{W}$ in Theorem 7.3.3, i.e., solving the minimal linear sparse ruler problem. Many solutions for the minimal linear sparse ruler problem have been tabulated.

7.5.2 Limited Correlation Support Case

Recall that we do not have any condition similar to Theorem 7.3.2 for the limited correlation support case. Hence, unlike in Section 7.5.1, it is reasonable for this case to expect that there might be better options than setting $\mathbf{C}_n$ either to $\mathbf{I}_T$ or $[]$ for all n, which leads to a circular sparse ruler based matrices design in Theorem 7.4.1. This motivates us to propose an algorithm, provided in Table 7.1, for designing $\{\mathbf{C}_n\}_{n=0}^{N-1}$ for the limited correlation support case in Section 7.4. The main for loop in step 4 of Table 7.1 decides on some rows of $\{\mathbf{C}_n\}_{n=0}^{N-1}$ while focusing on the full column rank condition of $\tilde{\mathbf{\Psi}}_k$. Only non-negative k is considered since $\tilde{\mathbf{\Psi}}_{-k}$ has full column rank once $\tilde{\mathbf{\Psi}}_k$ has full column rank. Each element of the indicator matrix $\mathbf{Z}^{(k)}$ is equal to either 0 or 1. $[\mathbf{Z}^{(k)}]_{i,j} = 1$ indicates that the $((i-1)T+j)$-th column of $\tilde{\mathbf{\Psi}}_k$ already has at least a single one. The while loop in step 5 checks if

all columns of the considered $\tilde{\mathbf{\Psi}}_k$ have at a least single one and, as long as this is not the case, it will iterate and add one row to one of the matrices $\{\mathbf{C}_n\}_{n=0}^{N-1}$. The inner for loop in steps 6-9 determines the indices of the candidate rows of $\mathbf{I}_T$ to be added to each of the matrices $\{\mathbf{C}_n\}_{n=0}^{N-1}$. However, steps 10-11 will decide that only one of the matrices $\{\mathbf{C}_n\}_{n=0}^{N-1}$ is going to be updated in each iteration of the while loop. Fig. 7.1 describes the achievable compression rate for the selection matrices designed using the greedy algorithm in Table 7.1. We run the algorithm 1000 times and pick the matrix offering the best compression for each N and T. Here, T is varied from $T = 18$ to $T = 30$ and N is varied from $N = 9$ to $N = 21$. We also plot the achievable compression rate for the selection matrices designed based on the minimal circular sparse ruler, which is independent of T since the minimal circular sparse ruler based $\mathbf{C}_n$ is set to either $\mathbf{I}_T$ or []. Observe that the minimal circular sparse ruler based selection matrices offer a stronger compression than the ones produced by the greedy algorithm. However, the minimal circular sparse ruler problem is a combinatorial problem whose solution has to be found using a brute force, which might be computationally infeasible for a large N. Moreover, the circular sparse ruler based selection matrices lead to many Nyquist-spaced samples and thus, the greedy algorithm might be more attractive for some applications.

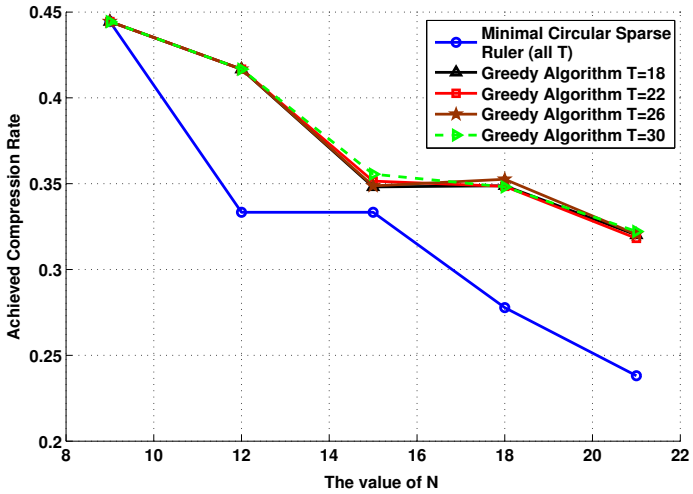


Figure 7.1: The achievable compression rate for the selection matrices designed using the greedy algorithm in Table 7.1 and those designed based on the minimal circular sparse ruler.

Table 7.1: A greedy algorithm to find a sub-optimal solution for $\{\mathbf{C}_n\}_{n=0}^{N-1}$ for limited correlation support case.

	Algorithm
1:	Introduce $\mathbf{Z}^{(k)}$ as a $T \times T$ indicator matrix with respect to $\tilde{\mathbf{\Psi}}_k$ and denote its element at the i-th row and the j-th column by $[\mathbf{Z}^{(k)}]_{i,j}$.
2:	For all $k = 0, \ldots, N-1$, initialize $\mathbf{Z}^{(k)} = \mathbf{0}_{T\times T}$ with $\mathbf{0}_{T\times T}$ a $T \times T$ matrix containing only zeros.
3:	For $k = N-1$, randomly select $i, j \in \{1, 2, \ldots, T\}$, and set $\mathcal{C}_0 = i$, $\mathcal{C}_{N-1} = j$, and $[\mathbf{Z}^{(N-1)}]_{i,j} = 1$.
4:	for $k = N-1$ to 0 in decreasing order do
5:	while $\mathbf{Z}^{(k)}$ has at least one zero entry do
6:	for $n = 0$ to $N-1$ do
7:	Define a set $\Xi = \{1, 2, \ldots, T\} \setminus \mathcal{C}_n$ and a function $f(g'_n, n)$ as $f(g'_n, n) = \sum_{i' \in \mathcal{C}_{(n-k) \bmod N}} (1 - [\mathbf{Z}^{(k)}]_{i', g'_n})$ $+ \sum_{i'' \in \mathcal{C}_{(n+k) \bmod N}} (1 - [\mathbf{Z}^{(k)}]_{g'_n, i''})$.
8:	Search in Ξ for the element g_n that satisfies: $g_n = \arg\max_{g'_n \in \Xi} f(g'_n, n)$, randomly pick one if we have multiple g_n, and set $h_n = f(g_n, n)$.
9:	end for
10:	Find $\bar{n}$ such that $h_{\bar{n}}$ is the maximum of $\{h_n\}_{n=0}^{N-1}$, randomly pick one if have multiple maxima $h_{\bar{n}}$, and update $\mathcal{C}_{\bar{n}}$ to $\mathcal{C}_{\bar{n}} = \mathcal{C}_{\bar{n}} \cup \{g_{\bar{n}}\}$.
11:	For all $i' \in \mathcal{C}_{(n-k) \bmod N}$ and $i'' \in \mathcal{C}_{(n+k) \bmod N}$, set $[\mathbf{Z}^{(k)}]_{i', g_{\bar{n}}}$ and $[\mathbf{Z}^{(k)}]_{g_{\bar{n}}, i''}$ to 1, respectively.
12:	end while
13:	if $k > 0$ do
14:	for $n = 0$ to $N-1$ do
15:	For all $i' \in \mathcal{C}_{(n-k+1) \bmod N}$, $i'' \in \mathcal{C}_{(n+k-1) \bmod N}$ and $j' \in \mathcal{C}_n$, set $[\mathbf{Z}^{(k-1)}]_{i', j'}$ and $[\mathbf{Z}^{(k-1)}]_{j', i''}$ to 1.
16:	end for
17:	end if
18:	end for

7.6 Conclusion

The knowledge of the cyclic period allows us to set the span of the random linear projections such that the block Toeplitz structure emerges in the WSCS signal correlation matrix. We have shown how to exploit this block Toeplitz structure, have introduced compression using non-uniform sampling, and have presented the reconstruction problem as an overdetermined system. We have presented the condition for the system matrix to have full column rank, which allows for LS reconstruction of WSCS signal correlation matrix. We considered two cases, the general case and the limited correlation support case. For the general case, we proposed the minimal linear sparse ruler based sampling matrices design as a reasonable way to approximately minimize the compression rate. For the limited correlation support case, we proposed a greedy algorithm to find a suboptimal solution for the sampling matrices, which might be more attractive in particular situations than the minimal circular sparse ruler based solution, though the latter appears to offer a stronger compression.

Chapter 8

Conclusions and Future Work

In this chapter, we provide the conclusions of the thesis and present some suggestions for future work.

8.1 Conclusions

In this thesis, we have formulated an approach for estimating the power spectrum of wide-sense stationary (WSS) signals based on sub-Nyquist rate samples without requiring any sparsity constraints. Our approach has been evaluated in both the time- and frequency-domain. Two candidates for sampling implementation, e.g., complex Gaussian sampling and multi-coset sampling, have been proposed to ensure that the rank condition of the resulting system equations is satisfied, which allows for a least-squares (LS) reconstruction of the power spectrum. We focus on the multi-coset sampling procedure, relate its design to a sparse ruler problem, and show that its design based on any sparse ruler can guarantee the full rank condition of the system equations.

As an extension of the above method, we have adopted a so-called dynamic array and formulated an approach to estimate the direction of arrival (DOA) of possibly fully correlated sources using second-order statistics. Here, we assume that we can have fewer active antennas than sources at a given time slot. In this method, the spatial correlation matrices of the output of the antenna arrays for all time slots are collected and they are presented as a linear function of the spatial correlation matrix of the entire underlying uniform array. We have developed the theoretical condition that needs to be satisfied to guarantee the full column rank condition of

the system matrix, which allows us to reconstruct the spatial correlation matrix of the entire underlying uniform array using LS. Based on this spatial correlation matrix, we have advocated three approaches to estimate the DOA of the sources, i.e., the LS approach (suitable for a smooth spectrum in the angular domain), the sparsity-regularized LS (suitable for a sparse angular power spectrum), and the MUSIC algorithm supported by spatial smoothing (suitable for point sources).

We have further proposed a compressive periodogram reconstruction approach where we consider both time-frequency and spatio-angular domains and where the entire band is split into equal-size bins such that the received spectra at two frequencies or angles, whose distance is equal to or larger than the size of a bin, are uncorrelated. We have found that this model leads to a circulant structure in the so-called coset correlation matrix, which allows for a strong compression. We have found that sampling patterns based on a circular sparse ruler ensure the full column rank condition of the system matrix, which allows for an LS reconstruction of the periodogram. We have also provided a solution for the case when the bin size is reduced such that, at two angles or frequencies separated by a distance larger than the bin size, we can still have correlated received spectra.

To merge frequency and angular power spectrum reconstruction, we have proposed an approach to compressively reconstruct the two-dimensional (2D) power spectrum as a function of frequency and DOA for multiple uncorrelated time-domain WSS signals received from different sources by a linear array of antennas. We have performed spatial-domain compression by activating only some antennas from an underlying uniform linear array and time-domain compression by applying sparse ruler sampling on the receivers associated with active antennas. The correlation between the sub-Nyquist rate samples at all receivers is calculated both in the time and spatial domain. We have been able to linearly relate these correlation values with the 2D power spectrum and to find the condition for the system matrices to have full column rank, which allows for an LS reconstruction of the 2D power spectrum. This approach is applicable even when we have more sources than active sensors (up to a certain limit) and even without any sparsity constraints on the true power spectrum.

Finally, we have proposed an approach to compressively reconstruct the cyclic spectrum of wide-sense cyclostationary (WSCS) signals from sub-Nyquist rate samples produced by non-uniform sampling. Once the block Toeplitz structure emerges in the WSCS signal correlation matrix, we have shown how the exploitation of this structure has allowed us to write the reconstruction problem as an overdetermined system. We have presented the condition that needs to be satis-

fied in order to have a full column rank system matrix, which allows for an LS reconstruction of the WSCS signal correlation matrix. We have also presented some possible non-uniform sampling designs to satisfy this full column rank condition.

Based on the overall thesis work, we have observed how concentrating on the reconstruction of the signal statistics (instead of the reconstruction of the original signals) have offered the increased degree of freedom, which finally leads to relaxed sampling requirements and alleviated constraints on both the signals and the signals' statistics. This implies that our works will be beneficial for applications where the information about signal statistics are sufficient, such as in spectrum sensing for cognitive radio applications, angular power spectrum map reconstruction for radio astronomy, and DOA estimation in radar applications. The final important point that can be concluded from this thesis is that it is important to have awareness on what information is necessary and sufficient for particular tasks or applications before we conduct the samples or data collection process, especially in this era of *big data*.

8.2 Suggestions for Future Work

Here, we present some suggestions for possible future research.

1. While, in Chapter 6, we have proposed an approach to compressively reconstruct the 2D power spectrum as a function of frequency and DOA for uncorrelated time-domain WSS signals received by a linear array of antennas, we still assume that the inverse of the bandwidth of the aggregated incoming signals is larger than the propagation delay across the ULA in order to allow us to represent the delay between the antennas as a phase shift. This indirectly implies that the overall signal bandwidth is not very wide, which contradicts the motivation for implementing sub-Nyquist sampling in the time-domain. It is thus interesting to refine the approach to handle signals with a very wide bandwidth and this demands us to leave the assumption that the propagation delay across the ULA is smaller than the inverse of the bandwidth of the aggregated incoming signals. One candidate approach that can be evaluated for this case is the coherent signal-subspace processing using focussing matrices of [104].

2. The dynamic array approach introduced in Chapter 4 to estimate the DOA of possibly fully correlated sources based on second-order statistics is only able to estimate the azimuth angle of the arrival of the sources. This is because

this dynamic array is based on a one-dimensional array as the underlying array. It would be interesting to find out how this dynamic array approach can be extended for a two-dimensional array so we can use it to estimate both the azimuth and elevation angles of arrival of correlated sources where the number of sources can be larger than the number of active antennas.

3. While the compressive wideband power spectrum estimation in Chapter 3 theoretically allows for power spectrum estimation of wideband WSS signals at very low sampling rates, it would be interesting to find out if this achievable minimum sampling rate can further be reduced if we only focus on detecting the existence of user signals in some particular bands instead of estimating the exact amount of power in those bands. In other words, it would be useful to develop a theory for wideband power spectrum detection.

4. The desire to reduce the burden on the analog-to-digital converter (ADC) when the signal to be sampled has a wide bandwidth is one reason for the emergence of compressive sampling (CS) a.k.a. sub-Nyquist sampling. However, the power consumption on an ADC is also dictated by the number of quantization levels used by the ADC. One possible future research should be related to the reconstruction of the power spectrum of WSS signals using a sub-Nyquist sampler with a very limited number of quantization levels.

5. While, in Chapter 7, we have shown how to compressively reconstruct the cyclic spectrum of WSCS signals from sub-Nyquist rate samples without any sparsity constraint, we still have to set the span of the random linear projection to an integer multiple of the cyclic period. It would be interesting to investigate how the span of the random linear projection can be related to the performance. Furthermore, it is also important to find the achievable compression rate if we only focus on cyclic feature detection instead of the reconstruction of the entire cyclic spectrum.

6. So far, the compressive power spectrum estimation approach in Chapter 3 is only implemented within the non-parametric power spectrum estimation framework. One possible future work will be on how to implement this compressive estimation approach within the parametric framework using some models such as autoregressive moving average (ARMA) or autoregressive (AR) models.

7. It is also interesting to find a way to elegantly integrate some classical windowing techniques into our compressive power spectrum estimation approach

like Hamming or Hann windowing. These are popular in conventional power spectrum estimation approaches like the correlogram.

Bibliography

[1] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2477–2488, July 2005.

[2] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Transaction on Signal Processing*, vol. 50, no. 6, pp. 1417–1428, June 2002.

[3] M. Unser, "Sampling-50 years after shannon," *Proceedings of the IEEE*, vol. 88, no. 4, pp. 569–587, Apr 2000.

[4] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, April 2006.

[5] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[6] J. A. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[7] J. Bai and S. Shi, "Estimating high dimensional covariance matrices and its applications," *Annals of Economics and Finance*, vol. 12, no. 2, pp. 199–215, 2011.

[8] Y. Zhang and J. Schneider, "Learning multiple tasks with a sparse matrix-normal penalty," *Advances in Neural Information Processing Systems*, vol. 23, pp. 2550–2558, 2010.

[9] C. E. Thomaz, *Maximum entropy covariance estimate for statistical pattern recognition: Ph.D Thesis*. London, United Kingdom: Department of Computing, Imperial College London, University of London, 2004.

[10] S. Shakeri, D. D. Ariananda, and G. Leus, "Direction of arrival estimation using sparse ruler array design," in *Proceeding of The 13th IEEE International Workshop on Signal Processing Advances in Wireless Communications*, June 2012, pp. 525–529.

[11] G. Leus and Z. Tian, "Recovering second-order statistics from compressive measurements," in *Proceeding of IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, December 2011, pp. 337–340.

[12] A. Moffet, "Minimum-redundancy linear arrays," *IEEE Transactions on Antennas and Propagation*, vol. 16, no. 2, pp. 172–175, March 1968.

[13] E. Candes and M. Wakin, "An introduction to compressive sampling," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 21–30, 2008.

[14] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, Feb. 2009. [Online]. Available: http://dx.doi.org/10.1137/060657704

[15] Y. C. Eldar and G. Kutyniok, *Compressed sensing: Theory and applications*. The Edinburgh Building, Cambridge, UK: Cambridge University Press, 2012.

[16] S. S. Chen, D. L. Donoho, Michael, and A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, 1998.

[17] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society, Series B*, vol. 58, no. 1, pp. 267–288, 1996.

[18] J. N. Laska, S. Kirolos, M. F. Duarte, T. S. Ragheb, R. G. Baraniuk, and Y. Massoud, "Theory and implementation of an analog-to-information converter using random demodulation," in *Proceeding of IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2007, pp. 1959–1962.

[19] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, "Beyond nyquist: Efficient sampling of sparse bandlimited signals," *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 520–544, Jan. 2010.

[20] 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*, vol. 4, no. 2, pp. 375–391, April 2010.

[21] R. Venkataramani and Y. Bresler, "Perfect reconstruction formulas and bound on aliasing error in sub-Nyquist nonuniform sampling of multiband signals," *IEEE Transactions on Information Theory*, vol. 46, no. 6, pp. 2173–2183, Sept. 2000.

[22] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, October 2004.

[23] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 600–616, March 1997.

[24] B. D. Rao, K. Engan, S. F. Cotter, J. Palmer, and K. Kreutz-Delgado, "Subset selection in noise based on diversity measure minimization," *IEEE Transactions on Signal Processing*, vol. 51, no. 3, pp. 760–770, March 2003.

[25] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis Elsevier*, vol. 26, no. 3, pp. 301–321, May 2009.

[26] M. Mishali and Y. C. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 4692–4702, October 2008.

[27] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation, Part I: Greedy pursuit," *Signal Processing Elsevier*, vol. 86, no. 3, pp. 572–588, March 2006.

[28] J. A. Tropp, "Algorithms for simultaneous sparse approximation, Part II: Convex relaxation," *Signal Processing Elsevier*, vol. 86, no. 3, pp. 589–602, March 2006.

[29] F. Zeng, C. Li, and Z. Tian, "Distributed compressive spectrum sensing in cooperative multihop cognitive networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 1, pp. 37–48, Feb. 2011.

[30] Y. Wang, G. Leus, and A. Pandharipande, "Direction estimation using compressive sampling array processing," in *IEEE/SP 15th Workshop on Statistical Signal Processing*, Sept. 2009, pp. 626–629.

[31] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[32] R. Jagannath, G. Leus, and R. Pribic, "Grid matching for sparse signal recovery in compressive sensing," in *Proceedings of the 9th European Radar Conference*, Nov. 2012, pp. 111–114.

[33] H. Zhu, G. Leus, and G. B. Giannakis, "Sparsity-cognizant total least squares for perturbed compressive sampling," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2002–2016, May 2011.

[34] A. Panahi and M. Viberg, "A novel method of DOA tracking by penalized least squares," in *Proceeding of IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, Dec. 2013, pp. 61–64.

[35] M. M. Hayder and K. Mahata, "Direction-of-arrival estimation using a mixed $\ell_{2,0}$ norm approximation," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4646–4655, Sept. 2010.

[36] X. Wei, Y. Yuan, and Q. Ling, "DOA estimation using a greedy block coordinate descent algorithm," *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6382–6394, Dec. 2012.

[37] Z. Yang, L. Xie, and C. Zhang, "Off-grid direction of arrival estimation using sparse Bayesian inference," *IEEE Transactions on Signal Processing*, vol. 61, no. 1, pp. 38–43, Jan. 2013.

[38] M. Mishali and Y. C. Eldar, "Blind multiband signal reconstruction: Compressed sensing for analog signals," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 993–1009, March 2009.

[39] J. D. Krieger, Y. Kochman, and G. W. Wornell, "Design and analysis of multi-coset arrays," in *Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 3781–3785.

[40] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4053–4085, Sept. 2011.

[41] P. Pal and P. P. Vaidyanathan, "Coprime sampling and the MUSIC algorithm," in *Proceeding of IEEE Digital Signal Processing and Signal Processing Education Workshop*, Jan. 2011, pp. 289–294.

[42] ——, "Nested arrays: a novel approach to array processing with enhanced degrees of freedom," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4167–4181, Aug. 2010.

[43] M. A. Lexa, M. E. Davies, J. S. Thompson, and J. Nikolic, "Compressive power spectral density estimation," in *Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2011)*, May 2011, pp. 3884–3887.

[44] C. P. Yen, Y. Tsai, and X. Wang, "Wideband spectrum sensing based on sub-Nyquist sampling," *IEEE Transactions on Signal Processing*, vol. 61, no. 12, pp. 3028–3040, June 2013.

[45] Z. Tian, Y. Tafesse, and B. M. Sadler, "Cyclic feature detection with sub-Nyquist sampling for wideband spectrum sensing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 1, pp. 58–69, Feb. 2012.

[46] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, March 1986.

[47] W. Ma, T. Hsieh, and C. Chi, "DOA estimation of quasi-stationary signals via Khatri-Rao subspace," in *Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing*, April 2009, pp. 2165–2168.

[48] T. J. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 4, pp. 806–811, Aug. 1985.

[49] H. S. Shapiro and R. A. Silverman, "Alias-free sampling of random noise," *Journal of the Society for Industrial and Applied Mathematics*, vol. 8, no. 2, pp. 225–248, June 1960.

[50] E. Masry and M.-C. C. Lui, "Discrete-time spectral estimation of continuous-parameter processes - A new consistent estimate," *IEEE Transactions on Information Theory*, vol. 22, no. 3, pp. 298–312, May 1976.

[51] E. Masry, "Poisson sampling and spectral estimation of continuous-time processes," *IEEE Transactions on Information Theory*, vol. 24, no. 2, pp. 173–183, March 1978.

[52] ——, "Alias-free sampling: An alternative conceptualization and its applications," *IEEE Transactions on Information Theory*, vol. 24, no. 3, pp. 317–324, May 1978.

[53] F. d. l. Huca Arce, *Compressive power spectral density estimation with non-uniform sampling: M.Sc Thesis*. Pamplona, Spain: Universidad Publica de Navarra, 2013.

[54] D. Qu and A. Tarczynski, "A novel spectral estimation method by using periodic non-uniform sampling," in *2007 Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers*, Nov. 2007, pp. 1134–1138.

[55] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[56] M. Elad, *Sparse and redundant representations: From theory to applications in signal and image processing*. Spring Street, New York, USA: Springer Science and Business Media, 2010.

[57] S. Gishkori, *Compressive Sampling for Wireless Communications: Ph.D Thesis*. Delft, The Netherlands: Delft University of Technology, 2014.

[58] M. Kim and J. Takada, "Efficient multi-channel wideband spectrum sensing technique using filter bank," in *Proceeding of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2009)*, Sept. 2009, pp. 1014–1018.

[59] Z. Quan, S. Cui, A. H. Sayed, and H. V. Poor, "Optimal multiband joint detection for spectrum sensing in cognitive radio networks," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1128–1140, March 2009.

[60] P. Paysarvi-Hoseini and N. C. Beaulieu, "Optimal wideband spectrum sensing framework for cognitive radio systems," *IEEE Transactions on Signal Processing*, vol. 59, no. 3, pp. 1170–1182, March 2011.

[61] Z. Tian and G. B. Giannakis, "A wavelet approach to wideband spectrum sensing for cognitive radios," in *Proceeding of IEEE International Conference on Cognitive Radio Oriented Wireless Networks and Communications*, June 2006.

[62] B. Le, T. W. Rondeau, J. H. Reed, and C. W. Bostian, "Analog-to-digital converters," *IEEE Signal Processsing Magazine*, vol. 22, no. 6, pp. 69–77, Nov. 2005.

[63] Z. Tian and G. B. Giannakis, "Compressed sensing for wideband cognitive radios," in *Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2007)*, vol. 4, April 2007, pp. IV/1357–IV/1360.

[64] Y. Bresler, "Spectrum-blind sampling and compressive sensing for continuous-index signals," in *Proceeding of Information Theory and Applications Workshop (ITA 2008)*, Jan. 2008, pp. 547–554.

[65] Y. Wang, A. Pandharipande, and G. Leus, "Compressive sampling based MVDR spectrum sensing," in *Proceeding of the International Workshop on Cognitive Information Processing (CIP 2010)*, June 2010, pp. 333–337.

[66] Y. L. Polo, Y. Wang, A. Pandharipande, and G. Leus, "Compressive wideband spectrum sensing," in *Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2009)*, April 2009, pp. 2337–2340.

[67] V. Havary-Nassab, S. Hassan, and S. Valaee, "Compressive detection for wide-band spectrum sensing," in *Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2010))*, March 2010, pp. 3094–3097.

[68] G. Leus and D. D. Ariananda, "Power spectrum blind sampling," *IEEE Signal Processing Letters*, vol. 18, no. 8, pp. 443–446, Aug. 2011.

[69] S. Kirolos, T. S. Ragheb, J. N. Laska, , M. F. Duarte, Y. Massoud, and R. G. Baraniuk, "Practical issues in implementing analog-to-information converters," in *Proceeding of the 6th International Workshop on System-on-Chip for Real-Time Applications*, Dec. 2006, pp. 141–146.

[70] J. Leech, "On the representation of 1,2,...,n by differences," *Journal of the London Mathematical Society*, vol. 31, pp. 160–169, April 1956.

[71] W. Bar and F. Dittrich, "Useful formula for moment computation of normal random variables with non-zero means," *IEEE Transactions on Automatic Control*, vol. 16, no. 3, pp. 263–265, June 1971.

[72] S. M. Kay, *Fundamentals of statistical signal processing, volume 2: detection theory*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1998.

[73] C. R. Stevenson, C. Cordeiro, E. Sofer, and G. Chouinard, "Functional requirements for the 802.22 WRAN standard, IEEE 802.22-05/0007r46," Sept. 2005.

[74] P. H. Janssen and P. Stoica, "On the expectation of the product of four matrix-valued Gaussian random variables," *IEEE Transactions on Automatic Control*, vol. 33, no. 9, pp. 867–870, Sept. 1988.

[75] P. P. Vaidyanathan and P. Pal, "Direct-MUSIC on sparse arrays," in *Proceeding of International Conference on Signal Processing and Communications (SPCOM)*, July 2012.

[76] A. J. Barabell, "Improving the resolution performance of eigenstructure-based direction-finding algorithm," in *Proceeding of IEEE International Conference on Acoustics, Speech, and Signal Processing*, April 1983, pp. 336–339.

[77] J. A. Cadzow, "A high resolution direction-of-arrival algorithm for narrow-band coherent and incoherent sources," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 7, pp. 965–979, July 1988.

[78] S. U. Pillai and B. H. Kwon, "Forward/backward spatial smoothing techniques for coherent signal identification," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 1, pp. 8–15, Jan. 1989.

[79] J. Yin and T. Chen, "Direction-of-arrival estimation using a sparse representation of array covariance vectors," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4489–4493, Sept. 2011.

[80] P. Stoica, P. Babu, and J. Li, "New method of sparse parameter estimation in separable models and its use for spectral analysis of irregularly sampled data," *IEEE Transactions on Signal Processing*, vol. 59, no. 1, pp. 35–47, Jan. 2011.

[81] ——, "SPICE: A sparse covariance-based estimation method for array processing," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 629–638, Feb. 2011.

[82] S. U. Pillai, Y. Bar-Ness, and F. Haber, "A new approach to array geometry for improved spatial spectrum estimation," *Proceeding of the IEEE*, vol. 73, no. 10, pp. 1522–1524, Oct. 1985.

[83] S. U. Pillai and F. Haber, "Statistical analysis of a high resolution spatial spectrum estimator utilizing an augmented covariance matrix," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 11, pp. 1517–1523, Nov. 1987.

[84] Y. I. Abramovich, D. A. Gray, A. Y. Gorokhov, and N. K. Spencer, "Positive-definite Toeplitz completion in DOA estimation for nonuniform linear antenna arrays - Part I: Fully augmentable arrays," *IEEE Transactions on Signal Processing*, vol. 46, no. 9, pp. 2458–2471, Sept. 1998.

[85] P. Pal and P. P. Vaidyanathan, "Correlation-aware techniques for sparse support recovery," in *Proceeding of IEEE Statistical Signal Processing (SSP) Workshop*, Aug. 2012, pp. 53–56.

[86] E. Gonen, M. C. Dogan, and J. M. Mendel, "Applications of cumulants to array processing: direction-finding in coherent signal environment," in *1994 Conference Record of the 28-th Asilomar Conference on Signals, Systems and Computers*, Nov. 1994, pp. 633–637.

[87] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," *Proceeding of National Academy of Sciences of the United States of America*, vol. 100, no. 5, pp. 2197–2202, March 2003.

[88] S. Jokar and V. Mehrmann, "Sparse solutions to underdetermined Kronecker product systems," *Elsevier Linear Algebra and its Applications*, vol. 431, no. 12, pp. 2437–2447, Dec. 2009.

[89] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representation in the presence of noise," *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.

[90] M. Bengtsson and B. Ottersten, "Low-complexity estimators for distributed sources," *IEEE Transactions on Signal Processing*, vol. 48, no. 8, pp. 2185–2194, Aug. 2000.

[91] Q. Wu, K. Wong, Y. Meng, and W. Read, "DOA estimation of point and scattered sources-vec-MUSIC," in *Proceeding of the IEEE 7th SP Workshop on Statistical Signal and Array Processing*, June 1994, pp. 365–368.

[92] A. Waters and V. Cevher, "Distributed bearing estimation via matrix completion," in *Proceeding of 2010 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, March 2010, pp. 2590–2593.

[93] P. Stoica and R. L. Moses, *Spectral analysis of signals*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2005.

[94] D. D. Ariananda and G. Leus, "Compressive wideband power spectrum estimation," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4775–4789, Sept. 2012.

[95] ——, "Cooperative compressive wideband power spectrum sensing," in *2012 Conference Record of the Forty-Sixth Asilomar Conference on Signals, Systems and Computers*, Nov. 2012, pp. 303–307.

[96] O. Mehanna and N. D. Sidiropoulus, "Frugal sensing: wideband power spectrum sensing from few bits," *IEEE Transactions on Signal Processing*, vol. 61, no. 10, pp. 2693–2703, May 2013.

[97] D. Cohen, E. Rebeiz, Y. C. Eldar, and D. Cabric, "Cyclic spectrum reconstruction and cyclostationary detection from sub-Nyquist samples," in *Proceeding of IEEE 14th Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2013, pp. 420–424.

[98] D. D. Arianandа and G. Leus, "Direction of arrival estimation for more correlated sources than active sensors," *Signal Processing Elsevier*, vol. 93, no. 12, pp. 3435–3448, Dec. 2013.

[99] D. Romero and G. Leus, "Wideband spectrum sensing from compressed measurements using spectral prior information," *IEEE Transactions on Signal Processing*, vol. 61, no. 24, pp. 6232–6246, Dec. 2013.

[100] ——, "Compressive covariance sampling," in *Proceeding of Information Theory and Applications Workshop (ITA 2013)*, Feb. 2013.

[101] D. D. Arianandа, D. Romero, and G. Leus, "Compressive angular and frequency periodogram reconstruction for multiband signals," in *Proceeding of IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, Dec. 2013, pp. 440–443.

[102] M. H. Hayes, *Statistical digital signal processing and modeling*. Hoboken, NJ, USA: John Wiley and Sons, Inc., 1996.

[103] D. D. Arianandа and G. Leus, "Compressive joint angular-frequency power spectrum estimation," in *Proc. the 21st European Signal Processing Conference*, Sept. 2013.

[104] H. Hung and M. Kaveh, "Focussing matrices for coherent signal-subspace processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 8, pp. 1272–1281, Aug. 1988.

Samenvatting

Bemonstering en kwantisatie die analoge signalen omzetten in digitale samples via een analoog-digitaalomzetter vormen het hart van digitale signaalverwerking. Er zijn recent heel wat signaalverwerkingstoepassingen waarbij signalen met een hoge bandbreedte moeten verwerkt worden. Dergelijke signalen vergen een hoge bemonsteringsfrequentie en analoog-digitaalomzetters met een hoog vermogen. *Compressive sensing* (CS) biedt een oplossing voor dit probleem en laat toe om een breedbandig signaal te reconstrueren op basis van samples verkregen aan een snelheid beneden de Nyquist snelheid. Om CS te kunnen toepassen is echter een extra voorwaarde nodig zoals de spaarsheid van het signaal. Maar er zijn ook toepassingen waarbij enkel de tweede-orde statistiek van het signaal moet gereconstrueerd worden (in plaats van het signaal zelf). In dat geval is het mogelijk om de tweede-orde statistiek te reconstrueren op basis van samples verkregen aan een snelheid beneden de Nyquist snelheid, zonder dat er extra voorwaarden nodig zijn. Dit idee vormt het startpunt van deze thesis.

Eerst richten we ons op tijdsdomein signalen die stationair zijn in de brede zin. We introduceren een methode die het frequentievermogenspectrum reconstrueert op basis van samples verkregen aan een bemonsteringssnelheid die lager is dan de Nyquist snelheid en dit zonder enige beperkingen op te leggen aan dit spectrum zoals spaarsheid. Onze methode is gebaseerd op de kleinste-kwadratenmethode en werkt zowel in het tijdsdomein als in het frequentiedomein. Enkel als de systeemmatrix van volle rang is wordt een oplossing verkregen. Om hieraan te kunnen voldoen stellen we twee bemonsteringsmethodes voor, waaronder de zogenaamde *multi-coset* methode. In deze thesis wordt aangetoond dat elke zogenaamde spaarse meetlat een *multi-coset* bemonstering genereert waarbij de gerelateerde systeemmatrix van volle rang is. De optimale compressie wordt dus bereikt met een mini-

male spaarse meetlat.

We kunnen de aanpak die wordt beschreven in de vorige paragraaf ook uitbreiden naar het spatiale domein. Hierbij zijn we vooral geïnteresseerd in het reconstrueren van het hoekvermogenspectrum op basis van een uniform lineair antennerooster waarbij een aantal antennas worden uitgeschakeld. Dit hoekvermogenspectrum is bijvoorbeeld belangrijk voor het schatten van de richtingen waaruit de signalen worden ontvangen. In deze thesis stellen we een methode voor om de richtingen van sterk gecorreleerde signalen te schatten op basis van een zogenaamde dynamisch antennrooster dat afgeleid is van het originele uniform lineair antennerooster.

Vervolgens bestuderen we de reconstructie van het periodogram op basis van gecomprimeerde signalen, en dit voor zowel het spatiale domein als het tijdsdomein. We introduceren een multisectie model waarbij de volledige band (in hoek of frequentie) wordt opgesplitst in uniforme secties zodanig dat het spectrum op twee hoeken of frequenties, die verder uit elkaar liggen dan de sectielengte, ongecorreleerd is. Zo een model geeft aanleiding tot een circulaire structuur in de systeemmatrix resulterend in een sterke compressie. We stellen bemonsteringspatronen voor die gebaseerd zijn op een circulaire spaarse meetlat en die een systeemmatrix opleveren met volle rang. Op die manier verkrijgen we een unieke oplossing voor het periodogram (in hoek of frequentie). We presenteren ook een methode voor het geval dat de sectielengte kleiner is zodanig dat het spectrum op twee hoeken of frequenties, die verder uit elkaar liggen dan de sectielengte, ook gecorreleerd kan zijn.

We stellen ook een gecomprimeerde tweedimensionale vermogenspectrumschatter voor waarbij zowel het frequentie- als het hoekvermogenspectrum wordt gereconstrueerd, en dit voor verschillende ongecorreleerde signalen die stationair zijn in brede zin en die worden ontvangen door een uniform linear antennerooster. We comprimeren in het spatiale domein door sommige antennes uit te schakelen en comprimeren in het tijdsdomein via *multi-coset* bemonstering.

Tot slot bestuderen we een reconstructiemethode voor het cyclisch vermogenspectrum van signalen die cyclostationair zijn in de brede zin en die bemonsterd worden aan een snelheid lager dan de Nyquist snelheid. Deze methode buit de blok-Toeplitz structuur uit van de correlatiematrix van het cyclostationaire signaal. Opnieuw wordt de kleinste-kwadratenmethode gebruikt waarbij de systeemmatrix van volle rang moet zijn. We stellen een aantal mogelijke bemonsteringstechnieken voor die aan deze voorwaarde voldoen.

Deze thesis bewijst dat de reconstructie van statistische informatie van een

gemeten signaal de bemonsteringsvoorwaarden beduidend verzwakt. Voor een bepaalde signaalverwerkingstaak moet men zich dus altijd afvragen of de reconstructie van statistische informatie niet voldoende is, omdat het antwoord op deze vraag zal bepalen hoe we de data moeten vergaren.

Propositions

1. The sampling process should take the considered signal processing task into account, especially in this era of big data.

2. While the power spectrum is always positive, the removal of this assumption in compressive power spectrum reconstruction is beneficial to carry out a performance analysis.

3. In a cognitive radio network, the proposed compressive power spectrum estimator is a more promising candidate for detecting spectral holes than a signal reconstruction approach exploiting the sparsity of the frequency occupancy.

4. If the goal of compressive power spectrum sensing is to reduce the power consumption in the analog-to-digital converters (ADCs), then not only reducing the number of samples but also reducing the number of quantization levels will be beneficial.

5. While sparsity-aware estimation tools show promising results in several applications, many features in nature are not sparse but have a more general structure. Hence, the development of structure-aware estimators will become important.

6. A great teacher is not the one that converts intelligent and highly talented students into achievers or award winners. Instead, it is the one that can help average or below average students to reach their potential.

7. Hard working but less talented people should be given more credit.

8. A person's attitude can be perceived as the statistics of a random variable and his/her actions as the realizations of this random variable. Assuming the statistics do not change, this person will end up where he/she belongs sooner or later.

9. Assume the objective is to minimize the number of people who dislike us, under the constraint that a certain level of human interaction is required. A good solution for this problem can be obtained by being low profile and by always trying to give more than we take.

10. Research can be successful or unsuccessful, and naturally requires freedom, passion and enjoyment. Unfortunately, society demands success, which takes away this freedom, passion and enjoyment.

These propositions are considered opposable and defendable, and as such have been approved by the supervisor prof.dr.ir. G.J.T. Leus.

Stellingen

1. Het bemonsteringsproces zou rekening moeten houden met de specifieke signaalverwerkingstaak, zeker in dit *big data* tijdperk.

2. Hoewel het vermogenspectrum altijd positief is, is het weglaten van deze veronderstelling in gecomprimeerde vermogenspectrumreconstructie nuttig om een performantieanalyse uit te voeren.

3. In een cognitief radionetwerk is de voorgestelde gecomprimeerde vermogenspectrumschatter een meerbelovende kandidaat om spectrale gaten te detecteren dan een signaalreconstructieaanpak die de spaarsheid van de frequentiebezetting uitbuit.

4. Als het reduceren van het vermogenverbruik van een analoog-digitaalomzetter het doel is van gecomprimeerde vermogenspectrumschatting, dan is het niet alleen nuttig om het aantal samples te reduceren maar ook het aantal kwantisatieniveaus.

5. Hoewel spaarse schattingstechnieken veelbelovende resultaten behalen in verschillende toepassingen, zijn vele eigenschappen in de natuur niet spaars maar hebben ze een meer algemene structuur. Vandaar dat de ontwikkeling van gestructureerde schatters alsmaar belangrijker wordt.

6. Een groot docent is niet diegene die intelligente en getalenteerde studenten omvormt tot strebers en prijzenwinnaars. Daarentegen, het is diegene die gemiddelde of ondermaatse studenten helpt om hun potentieel te bereiken.

7. Hardwerkende maar minder getalenteerde mensen zouden meer erkenning moeten krijgen.

8. Het gedrag van een persoon kan gezien worden als de statistiek van een stochastische variabele en zijn/haar akties als de realisaties van deze stochastische variabele. In de veronderstelling dat de statistiek niet verandert, zal deze persoon vroeger of later eindigen waar hij/zij thuis hoort.

9. Veronderstel dat we het aantal mensen die ons minachten willen minimaliseren, gegeven dat een zeker niveau van menselijke interactie nodig is. Een goede oplossing voor dit probleem kan bereikt worden door onopvallend te zijn en altijd te proberen om meer te geven dan te nemen.

10. Onderzoek kan succesvol zijn of niet en vergt natuurlijk ook vrijheid, passie en plezier. Jammer genoeg eist de maatschappij succes, wat deze vrijheid, passie en plezier wegneemt.

Deze stellingen worden opponeerbaar en verdedigbaar geacht en zijn als zodanig goedgekeurd door de promotor prof.dr.ir. G.J.T. Leus.

Acknowledgments

I would like to express my gratitude to all those people who have helped me in these last more than four and half years needed to complete this Ph.D. work

I would like to thank my supervisor, Professor Geert Leus, for all of his guidance and lessons throughout my Ph.D. Geert is the right professor for me. Since I cannot express it in words more than that sentence for all what he has done to coach me, I prefer to keep the other sentences to myself.

I would also like to thank Professor Alle-Jan van der Veen as the leader of Circuit and Systems group and for all the facilities that I could access during my Ph.D. work. I also thank him for his permission to put my computer in the laboratory and access it remotely. Finally, I also thank him for his role as one of the committee members.

I would like to thank my colleague Daniel Romero from the University of Vigo. His feedbacks and comments on some papers where he is the co-author are extremely useful and his ideas are always inspiring to increase the quality of our papers. I would also like to thank Professor Zhi Tian from Michigan Technological University, who has also given very useful comments and feedback in one of my papers. I would also like to thank some colleagues which have also been involved in my Ph.D. work. I thank Siavash Shakeri during his M.Sc extra project as well as Fernando de la Huca Arce and Zhang Ruijie during their M.Sc thesis. Even though their contributions cannot be directly reflected in this thesis, my cooperation with them during their projects has been very useful for me to better understand my Ph.D. problem.

I would like to thank all other members of my defence committee, in alphabetical order, Professor Laurent Jacques, Professor Piet Van Mieghem, Professor Roberto López Valcarce, Professor Visa Koivunen, and Professor Yonina C. Eldar.

I would like to thank them for reading my thesis and giving me extremely useful suggestions.

From the secretariat office, I would like to thank Minaksie Ramsoekh and the project support Rosario Salazar for their prompt help in many of the official matters. Minaksie deserves extra appreciation for her friendly gestures. It always makes me feel relaxed to know that she is there in her office. I also would like to thank Antoon Frehe for many technical helps during my Ph.D.

Next, I would like to show some appreciation to some of my roomates on 17th floor. I generally start with the people that I recognize earlier or the ones that share the room with me for a longer period of time. I would like to thank Shahzad Gishkori and I prefer to keep extra words for Shahzad to myself. The next one is Sina Maleki. His relaxed attitude always makes me relax as well. I am happy to have him to talk about football as we generally support the same team. I would like to express my gratitude to Sundeep Prabakhar Chepuri for providing his time for a lot of discussion (mostly non technical things) including a lot of general daily problems that I encountered during my time in the Netherlands. I would like to thank Venkat Roy for his generosity and his friendly attitude. I really enjoy playing badminton and having meals with him. The last one will be Raj Thilak Rajan, with whom I shared the office during my earlier time on the 17th floor. The discussion with him is enjoyable and always inspiring. Playing badminton with him is also a great experience.

I would like to thank all other colleagues in the signal processing group, Adib Sarijari, Andrea Simonetto (his sense of humour is really refreshing), Elvin Isufi, Georg Kail, Hamid Ramezani, Jorge Martinez, Millad Sardarabadi, Rocio Arroyo Valles (her advices are always appreciated), Seyran Khademi, Shahrzad Naghibzadeh, Yan Xie, and Yongchang Hu. I would like to thank my past colleagues Alon Amar, Hadi Jamali Rad, Mu Zhou, Nihan Cicek, Tao Xu, Toon van Waterschoot, Yiyin Wang, Yu Bi, and Zijian Tang. I had a quality time when they are/were around and got help from them in many ways.

Ph.D cannot be started without M.Sc and thus I would like to thank some people who assisted me a lot during that time. I thank Dr. Gerard Janssen for granting me M.Sc telecommunication grant and Dr. Homayoun Nikookar for being my M.Sc thesis supervisor. I would also like to thank Madan Kumar Lakshmanan as my M.Sc thesis mentor. Madan always motivates me since the first time I knew him up till today. I would like to thank my M.Sc colleagues, Vijay S. Rao and Ngo Minh Tri. I am happy that we are still frequently in touch until today. I also want to thank my senior, Bambang Soelistijanto in Indonesia as he is the one who informed me

about the chance for pursuing study at TU Delft.

I would also like to thank my Indonesian friends in Delft, Enny Kurniawati and her husband Juan, as well as Widita Budhysutanto. Thank you very much for always being available to reply to my questions in email.

Back in Indonesia, I would like to thank my father, my mother, my grandmother, my siblings and their families, and my relatives.

Curriculum Vitae

Dyonisius Dony Ariananda was born in Semarang, Central Java, Indonesia. He received the bachelors degree (cum laude) from Gadjah Mada University, Yogyakarta, Indonesia, in May 2000 and the M.Sc degree (cum laude) from Delft University of Technology, The Netherlands, in August 2009, both in electrical engineering. His M.Sc study in Delft University of Technology was partially assisted by telecommunication grant given by Department of Telecommunication, Faculty of EEMCS, Delft University of Technology. Apart from that, he was also granted the Australian Development Scholarship (ADS) by the Australian government to complete the M.S. degree in internetworking at the University of Technology Sydney, Australia, in July 2005. He has been a member of the junior education staff at the Informatics Engineering Department, Sanata Dharma University, Yogyakarta, from October 2000 to December 2003, and with the Electrical Engineering Department, Atma Jaya University, Jakarta, Indonesia, from January 2006 to July 2007. His research is mainly focused on signal processing for communication. During his Ph.D. study with the Circuit and System (CAS) group, Delft University of Technology, he worked on compressive spectral analysis and spectrum estimation. His Ph.D. study was fully funded by NWO-STW under the VICI program (project 10382).

Publications during Ph.D

Journals

- D.D. Arianandа and G. Leus, "Compressive Wideband Power Spectrum Estimation", *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4775–4789, Sept. 2012

- D.D. Arianandа and G. Leus, "Direction of Arrival Estimation for More Correlated Sources Than Active Sensors", *Elsevier Signal Processing*, vol. 93, no. 12, pp. 3435–3448, Dec. 2013

- G. Leus and D.D. Arianandа, "Power Spectrum Blind Sampling", *IEEE Signal Processing Letters*, vol. 18, no. 8, pp. 443–446, Aug. 2011

- D.D. Arianandа, D. Romero, and G. Leus, "Compressive Periodogram Reconstruction Using a Multibin Approach", *submitted to IEEE Transactions on Signal Processing (accepted with mandatory minor revisions)*

Conferences

- D.D. Arianandа, D. Romero, and G. Leus, "Cooperative Compressive Power Spectrum Estimation", *Proceeding of the 2014 IEEE 8th Sensor Array and Multichannel Signal Processing Workshop (SAM 2014)*, A Coruna, Spain, pp. 97–100, June 2014

- D.D. Arianandа and G. Leus, "Non-Uniform Sampling for Compressive Cyclic Spectrum Reconstruction", *Proceeding of the 2014 IEEE International Con-*

ference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, pp. 41–45, May 2014

- D.D. Arianandа, D. Romero, and G. Leus, "Compressive Angular and Frequency Periodogram Reconstruction for Multiband Signals", *Proceeding of the 5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP 2013)*, Saint Martin, French West Indies, pp. 440–443, Dec. 2013

- D.D. Arianandа and G. Leus, "Compressive Joint Angular-Frequency Power Spectrum Estimation", *Proceeding of the 21st European Signal Processing Conference (EUSIPCO 2013)*, Marrakech, Morocco, Sept. 2013

- D.D. Arianandа and G. Leus, "Direction of Arrival Estimation of Correlated Signals Using a Dynamic Linear Array", *2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, Pacific Grove, California, pp. 2028–2035, Nov. 2012

- D.D. Arianandа and G. Leus, "Cooperative Compressive Wideband Power Spectrum Sensing", *2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, Pacific Grove, California, pp. 303–307, Nov. 2012

- D.D. Arianandа and G. Leus, "A Study on Cooperative Compressive Wideband Power Spectrum Sensing", *Proceeding of WIC/IEEE SP Symposium on Information Theory and Signal Processing in the Benelux*, Boekelo, The Netherlands, pp. 102–109, May 2012

- D.D. Arianandа, G. Leus, and Z. Tian, "Multi-Coset Sampling for Power Spectrum Blind Sensing", *Proceeding of 2011 17th International Conference on Digital Signal Processing (DSP)*, Corfu, Greece, July 2011

- D.D. Arianandа and G. Leus, "Wideband Power Spectrum Sensing Using Sub-Nyquist Sampling", *Proceeding of The 12th IEEE International Workshop on Signal Processing Advances in Wireless Communications*, San Francisco, California, pp. 101–105, June 2011

- D.D. Arianandа and G. Leus, "Compressive Sampling for Power Spectrum Estimation", *Proceeding of WIC/IEEE SP Symposium on Information Theory and Signal Processing in the Benelux*, Brussels, Belgium, May 2011

• S. Shakeri, D.D. Ariananda and G. Leus, “Direction of Arrival Estimation Using Sparse Ruler Array Design”, *Proceeding of The 13th IEEE International Workshop on Signal Processing Advances in Wireless Communications*, Cesme, Turkey, pp. 525–529, June 2012

Glossary

Acronyms

2D	Two Dimensional
ADC	Analog to Digital Converter
AIC	Analog to Information Converter
AP	Averaged Periodogram
AR	Autoregressive
ARMA	Autoregressive Moving Average
ATD	Alternative Time Domain
CAP	Compressive Averaged Periodogram
CB	Correlated Bins
CFAR	Constant False Alarm Rate
CR	Cognitive Radio
CS	Compressive Sampling
CSD	Cross Spectral Density
CSI	Channel State Information
DFT	Discrete Fourier Transform

DLA	Dynamic Linear Array
DOA	Direction of Arrival
DTFT	Discrete Time Fourier Transform
FBSS	Forward Backward Spatial Smoothing
FC	Fusion Centre
FFT	Fast Fourier Transform
GBCD	Greedy Block Coordinate Descent
IDFT	Inverse Discrete Fourier Transform
JLZA	Joint ℓ_0 Approximation
LASSO	Least Absolute Shrinkage and Selection Operator
LS	Least Squares
MMV	Multiple Measurement Vectors
MRA	Minimum Redundancy Array
MUSIC	Multiple Signals Classification
MV	Measurement Vector
MVDR	Minimum Variance Distortionless Response
NAP	Nyquist-rate Based Averaged Periodogram
NMSE	Normalized Mean Squared Error
NULA	Non-Uniform Linear Array
p.s.d	positive semi-definite
PSBS	Power Spectrum Blind Sampling
PSD	Power Spectral Density
RIP	Restricted Isometry Property
RMSE	Root Mean Squared Error
ROC	Receiver Operating Characteristic

SBS	Spectrum Blind Sampling
SMV	Single Measurement Vector
SNR	Signal to Noise Ratio
SPICE	Sparse Iterative Covariance-Based Estimation
SRACV	Sparse Representation of Array Covariance Vectors
SS	Spatial Sampling
SVD	Singular Value Decomposition
TD	Time Domain
TS	Temporal Sampling
UB	Uncorrelated Bins
ULA	Uniform Linear Array
ULS	Uniform Linear Subarray
UWB	Ultra-Wideband
WSCS	Wide-Sense Cyclostationary
WSS	Wide-Sense Stationary

Notations

x	Scalar x
$\mathbf{x}$	Vector $\mathbf{x}$
$\mathbf{x}^T$	Transpose of vector $\mathbf{x}$
$\mathbf{x}^H$	Conjugate transpose of vector $\mathbf{x}$
$\mathbf{X}$	Matrix $\mathbf{X}$
$\mathbf{X}^{-1}$	Inverse of matrix $\mathbf{X}$
$\text{tr}(\mathbf{X})$	Trace of matrix $\mathbf{X}$ (sum of all diagonal entries of matrix $\mathbf{X}$)
$\mathbf{I}_N$	Identity matrix of size $N \times N$
$\otimes$	Kronecker product
$\odot$	Khatri-Rao product