Atomic Norm decomposition for sparse model reconstruction applied to positioning and wireless communications

by:

Adrián Vega Delgado

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy in

Multimedia and Communications PhD program

Universidad Carlos III de Madrid

Director:

Dr. Matilde Pilar Sánchez Fernández

February 2022

Dedication

Gracias a mi familia, especialmente a mis padres, quienes con su confianza, compañía y apoyo incondicional me han impulsado a cumplir todas mis metas.

A todas las personas con las que he trabajado durante el doctorado. En especial a Antonia Tulino, un placer trabajar contigo y aprender de ti.

A Soraya, esta tesis posiblemente no habría visto la luz de no ser por tu empuje y apoyo en todo momento. Gracias por tu amor y tus ánimos siempre.

Y finalmente me gustaría expresar mi inmensa gratitud a mi directora Mati por su infinita paciencia y dedicación, por abrirme las puertas a la investigación y hacer que salga de cada reunión comprendiendo una cosa nueva más.

PUBLISHED AND SUBMITTED CONTENT

The following published papers are included as part of the thesis. They are not singled out by means of typography nor citation:

- A. V. Delgado, M. Sánchez-Fernández, L. Venturino and A. Tulino, "Super-Resolution in Automotive Pulse Radars," in IEEE Journal of Selected Topics in Signal Processing, vol. 15, no. 4, pp. 913-926, June 2021, doi: 10.1109/JSTSP.2021.3066126. URL: `https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9380368&isnumber=9448241`.

  This article, [1], is wholly included in the thesis as chapter 5: Super-resolution in automotive pulse radars.

- A. V. Delgado, M. Sanchez-Fernandez, J. Llorca and A. Tulino, "Feasible Transmission Strategies for Downlink MIMO in Sparse Millimeter-Wave Channels," in IEEE Communications Magazine, vol. 56, no. 7, pp. 49-55, July 2018, doi: 10.1109/MCOM.2018.1701022. URL: `https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8419178&isnumber=8419121`

  This article, [2], is wholly included in the thesis as chapter 6: Low complexity hybrid precoding strategies for MIMO in sparse mmWave channels.

OTHER RESEARCH MERITS

Next, the following research merits are not included in the document but are part of the research activity carried out during the PhD.

- Estefanía Crespo-Bardera, Adrián Vega Delgado, Aarón Garrido Martín, Alfonso Fernández-Durán, Matilde Sánchez-Fernández, "Textile Multiantenna Technology and Relaying Architectures for Emergency Networks," Wireless Communications and Mobile Computing, vol. 2019, Article ID 7292075, 7 pages, 2019. `https://doi.org/10.1155/2019/7292075` [3].

ABSTRACT

This thesis explores the recovery of sparse signals, arising in the wireless communication and radar system fields, via atomic norm decomposition. Particularly, we focus on compressed sensing gridless methodologies, which avoid the always existing error due to the discretization of a continuous space in on-grid methods. We define the sparse signal by means of a linear combination of so called atoms defined in a continuous parametrical atom set with infinite cardinality. Those atoms are fully characterized by a multi-dimensional parameter containing very relevant information about the application scenario itself. Also, the number of composite atoms is much lower than the dimension of the problem, which yields sparsity. We address a gridless optimization solution enforcing sparsity via atomic norm minimization to extract the parameters that characterize the atom from an observed measurement of the model, which enables model recovery. We also study a machine learning approach to estimate the number of composite atoms that construct the model, given that in certain scenarios this number is unknown.

The applications studied in the thesis lay on the field of wireless communications, particularly on MIMO mmWave channels, which due to their natural properties can be modeled as sparse. We apply the proposed methods to positioning in automotive pulse radar working in the mmWave range, where we extract relevant information such as angle of arrival (AoA), distance and velocity from the received echoes of objects or targets. Next we study the design of a hybrid precoder for mmWave channels which allows the reduction of hardware cost in the system by minimizing as much as possible the number of required RF chains. Last, we explore full channel estimation by finding the angular parameters that model the channel. For all the applications we provide a numerical analysis where we compare our proposed method with state-of-the-art techniques, showing that our proposal outperforms the alternative methods.

Keywords: atomic norm, compressed sensing, mmWave, massive multiple-input multiple-output (MIMO), sparse, automotive, pulse radar, superresolution, gridless, hybrid precoder, RF chain, channel estimation, AoD, AoA, machine learning.

Contents

List of Figures

List of Tables

Acronyms

5G fifth generation. 17, 59, 75

6G sixth generation. 17, 59

ADAS Advanced Driving Assistance System. 41

AN Atomic Norm. 9, 10, 42, 52, 82

AoA Angle of Arrival. 2, 8, 19, 20, 23, 26, 27, 41, 75, 76

AoD Angle of Departure. 2, 19, 20, 23, 27, 75, 76

BB baseband. 59, 60, 62, 63, 70

BS Base Station. xiv, 60–64, 70

CMOS Complementary Metal–Oxide–Semiconductor. 59

CNN Convolutional Neural Network. 75

CPI Coherent Processing Interval. 41, 43, 44

CSI Channel State Information. 61, 62, 68, 69

CSIT Channel State Information at the Transmitter. 61, 70

D Dimensional. ix, xiii, 2, 3, 8, 9, 11, 12, 18–23, 25, 30, 31, 42, 44, 45, 49, 50, 60, 71, 72, 82

DL Downlink. xi, xiv, xv, 59–61, 64–67, 70, 72, 73

DPC Dirty Paper Coding. 61

EMB Enhanced Mobile Broadband. 2

ESPRIT Estimation of Signal Parameters via Rotational Invariance Techniques. 41

FD fully-digital. xiv, 59, 63, 64, 68

GSM Global System for Mobile Communications. 17

HB Hybrid. xiv, 60, 63, 64, 70

IIC-AMF Adaptive Matched Filter with Iterative Interference Cancellation. 42, 55, 57

IoT Internet of Things. 17

LIS Large Intelligent Surfaces. 18

LOS line of sight. 60

LT Level Toeplitz. 10–14, 16, 26, 27, 30

MAE Mean Absolute Error. xiv, 55–58

MIMO Multiple-Input Multiple-Output. iv, ix, x, 2–4, 17, 18, 20, 22, 24, 26, 28, 59, 60, 62, 64–66, 68, 70–72, 74–76, 82

ML Machine Learning. 30

MLT Multi–Level Toeplitz. 10, 11, 81, 82

MMSE Minimum Mean Square Error. 64, 67, 68, 70

MMT Massive Machine Type. 2

MMV Multiple Measurement Vector. x, 24, 25, 27, 52, 69, 76

mmWave millimeter-wave. iv, xi, 1–4, 17, 18, 23–25, 27, 41, 42, 59, 60, 62, 64, 66, 68, 70, 72, 74, 75, 82

MRI Magnetic Resonance Imaging. 1

MSE Mean Square Error. xiii–xv, 33, 35, 37, 39, 53, 55–58, 62, 65–68, 70, 72, 73, 77, 79, 80

MU-MIMO Multi-User MIMO. xi, xiv, xv, 59–61, 64–66, 68, 70, 72, 73, 82

MUSIC MUltiple SIgnal Classification. 41, 70, 71, 75, 77, 79, 82

OMP Orthogonal Matching Pursuit. 70, 71, 75, 77, 82

PCSI Partial Channel State Information. 61

PRT Pulse Repetition Time. 43, 47, 54

PSD Positive Semidefinite. 11–13

RF Radio Frequency. 3, 4, 59, 60, 62–64, 68, 70, 71, 82

RL Rayleigh Length. 57

SINR Signal to Interference and Noise Ratio. 62, 65–68

SMV Single Measurement Vector. x, 24–27, 30, 69

SNR Signal to Noise Ratio. 41, 54

TDD Time Division Duplexing. 66

UL Uplink. xi, 64–67

URLL Ultra-Reliable Low-Latency. 2

Chapter 1

Introduction

Structurally constrained models with very limited degrees of freedom compared to the dimension of available data are present in many science and engineering fields. The applications where they arise go from signal processing [4]–[6], statistics [7], computer vision [8], machine learning [9] to neuroscience [10]. The sparsity of those models can be due to many factors, *e.g.* data gathering by means of sensor networks [11] where size and cost of the sensors yield sparsity, magnetic resonance imaging (MRI) where time for data acquisition is constrained, or wireless communications in millimeter-wave (mmWave) channels where the number of scatterers and thus propagation echoes is limited.

These models, known as sparse, are formulated by a linear combination of a set of so called atoms belonging to a known atom set. In this thesis we will focus on a specific set with infinite cardinality whose atoms are a Fourier transform fully characterized by a d-dimensional parameter $\mathbf{f} \in \mathbb{T}^d$. The linear combination is done over a small number of atoms if compared to the dimension of the available data, causing the model to be sparse. Due to the sparsity of the model, the measured observations of the linear combination of atoms need fewer samples than required by the Nyquist-Shannon sampling theorem [12] to recover the data. Reconstructing the sparse model from this particular underdetermined system is what is known as compressed sensing.

Existing methods for parameter extraction via compressed sensing can be classified in two main groups: on-grid methods and gridless methods. The former group searches for the parameters in a discrete space, forcing the estimates to lay on the grid. Thus, even with a very intensive sampling of the continuous space, there will always be an error between the estimates and the real values [13], [14]. Conversely, the latter operates directly on the continuous space, and therefore, avoids any errors caused by the grid. Gridless methods are usually more computationally expensive.

In this thesis we aim to a gridless solution that allows the recovery of the d-dimensional parameters that characterize the atoms in the continuous domain in which they are defined. Several works have studied these solutions but, initially they were only applied in the context of line spectral estimation [15]–[17] and then in Angle of Arrival (AoA) estimation in 1-dimensional (1-D) linear antenna arrays [18], [19] and non-uniform manifolds [20]. Only more recently, multidimensional scenarios have been considered in [21]–[23] and further elaborated in [24] for AoA estimation in uniform and non-uniform antenna array deployments. It is important to note that 3-dimensional (3-D) arrays, especially non-uniform arrays, have become very important in the present and future next generations of wireless communications with Enhanced Mobile Broadband (EMB), Massive Machine Type (MMT), or Ultra-Reliable Low-Latency (URLL) communications among others. Non-uniform arrays are much more common than uniform arrays since they can fit already existing spaces with other purposes [25]. The works in [26], [27] have studied non-uniform geometries with reference to a 1-D parameter space.

The applications studied in this work are related to MIMO wireless communications in the millimeter-wave range. The nature of propagation in this frequency range presents severe path loss if compared to traditional microwave channels, due to the $10\times$ increase in carrier frequency. This pathloss can be mitigated by means of MIMO deployments where multiple antenna arrays are placed both at transmitter and receiver sides. The aforementioned pathloss is seriously worsened with any obstacle that the signal may encounter. Therefore, it leads to sparse channels in terms of the number of propagation paths, specially if we compare that number to the full dimension of the MIMO scenario. The parameter $\mathbf{f}$ that characterizes the sparse model can contain, depending on the application, angular spread in the multidimensional propagation scenario, *i.e.* angle of departure (AoD) and angle of arrival (AoA); propagation delay or Doppler shift. We propose some techniques to recover the parameters enforcing sparsity using atomic norm minimization [24], [28], [29]. Also, we provide numerical analysis in each of the applications studied in the thesis.

1.1 Objectives and motivation

In wireless communications, and specially in mmWave channels there are plenty of scenarios where sparse models may apply. Furthermore, very relevant information about the scenario itself is contained in the parameters embedded in the model, such as, as mentioned above, angles of departure and arrival, delay or Doppler shift. Thus, a method capable of retrieving this information by measuring the model is of high interest and the main motivation of the thesis. Existing techniques usually can be divided into two different approaches, as introduced earlier. On the one hand, relying on a predefined grid to find the parameters of the model, yielding an exhaustive search over the grid, or dictionary, and a computational complexity that grows with the length of the dictionary. And on the other hand, gridless techniques aim to find the parameters of the model in a continuous space.

The objective of this thesis is to give some new results on the gridless recovery of multidimensional parameters in wireless communication and radar scenarios.

1.2 Contributions

The contributions of this thesis are:

- Validation via numerical analysis of the theoretical results presented in [24].
- Application of gridless methodology for parameter extraction in sparse models to the design of a low complexity precoder in mmWave channels. We limit as much as possible the number of radio-frequency (RF) chains needed in a hybrid precoder, enabling up to 3-dimensional MIMO propagation, whereas in the literature, hybrid precoders are mostly studied in the linear case (1D).
- Application to positioning by estimation of parameters such as direction, distance and velocity. Compared to state-of-the-art techniques based on grid, our gridless methods outperform them for equal computational complexity.
- Application to full channel estimation, including relevant parameter extraction such as angles of departure and arrival of propagation echoes. Propagation is studied in MIMO mmWave sparse channels.
- A machine learning based methodology to estimate the number of composite atoms in the sparse model. We leverage the prior knowledge of the system to enable the possibility of training a model offline and use its predictions online to have an estimation of the number of atoms. We explore and test several models.

1.3 Thesis structure

The rest of the document is organized as follows:

- **Chapter 2. Gridless methodology based on Atomic Norm for sparse model recovery**

  In this chapter, we present a general formulation for sparse model recovery. We define the parametric atom set based on the Fourier transform of some parameter $\mathbf{f}$ with which we will work throughout this research. Also, we present the optimization problems and associated resolvability conditions under which the parameters that define the atoms can be recovered.

- **Chapter 3. General model for MIMO millimeter-wave channels**

  We explore here MIMO propagation in mmWave channels and its sparse nature. We explain how different scenarios in wireless communications can fit into sparse models as the one presented in chapter 2.

- **Chapter 4. Machine Learning approach for estimation of number of composite atoms**

  Next, in this chapter, we present a machine learning approach to estimate the number of composite atoms that define the signal model. We particularize the scenario for a mmWave MIMO channel as defined in chapter 3 where the atoms can be also understood as propagation paths. We explore several models and present their performance.

- **Chapter 5. Super-resolution in automotive pulse radars**

  This and the following two chapters are applications of the type of sparse models studied in the thesis. In this first application we explore automotive pulse radars working in the mmWave range. This range provides the sparsity needed to model the received echoes as sparse. From this, and using the techniques presented in chapter 2 and the formulation in chapter 3 we extract relevant information about the echoes or targets such as angle of arrival, distance and velocity. Numerical analysis show that the obtained results with our proposed gridless method outperform those of state-of-the-art techniques.

- **Chapter 6. Low complexity hybrid precoding strategies for MIMO in sparse mmWave channels**

  The second application deals with low complexity hybrid precoder design in mmWave MIMO channels. The model is described and formulated as in chapter 3, with a very limited number of propagation paths compared to the dimension of the transmit and receive antenna arrays. The aim is to leverage the inherent sparsity of the channel to reduce as much as possible the number of needed RF chains yielding a reduction of the hardware cost at the transmitter. We use the gridless methodology presented in chapter 2 to extract relevant information that enables the low complexity design of hybrid precoders. We compare the results obtained with our gridless methodology which outperforms state-of-the-art grid methods.

- **Chapter 7. Full channel estimation using atomic norm**

  The last application presented in this work has to be with full channel estimation. Again we leverage the sparse nature of mmWave MIMO channels to extract relevant information about the parameters that characterize the channel, *i.e.* angle of departure and angle of arrival. With the extraction of these parameters we are able to estimate the channel. Some results are presented via numerical analysis and compared to state-of-the-art techniques where we show that our proposal outperforms the alternative methods.

- **Chapter 8. Conclusion and future work**

  Finally, in this last chapter we draw some conclusions and suggest some points where further research can be carried on in the future.

1.4 Notation

Scalars, vectors and matrices

Let $[\cdot]^\top$, $[\cdot]^\dagger$ and $(\cdot)^*$ denote, respectively, transpose, Hermitian and complex conjugate. $\|\cdot\|_\rho$ denotes the ℓ_ρ norm and $\mathrm{mod}\,(x, y)$ is the remainder of x divided by y, *i.e.* the y–modulus operator. $\mathbf{X}^{\mathrm{g}}$ is the generalized inverse of matrix $\mathbf{X}$. The inner product between $\mathbf{x}$ and $\mathbf{y}$ is denoted by $\mathbf{x}\cdot\mathbf{y}$. Let $\otimes$, $\odot$, and $\circ$ denote the Kronecker, the Khatri–Rao, and the Schur products, respectively. The operator $\mathrm{diag}(\mathbf{x})$ returns a diagonal matrix whose diagonal is the vector $\mathbf{x}$.

Sets

A set is defined by enumerating its elements or by comprehension (between curly brackets, $\{\cdot\}$). For any positive integer N, let $[N]$ be the set $\{1, 2, \ldots, N\}$. Given two integer numbers a and b, let $a : b$ be the set $\{a, a+1, \ldots, b\}$. $\mathbb{T}$ is the unit circle, *i.e.* the interval $[0, 1)$. For a given set $\mathcal{I}$, $|\mathcal{I}|$ denotes its cardinality, *i.e.* the number of elements in the set, and $\bar{\mathcal{I}}$ is the complement of $\mathcal{I}$.

Indexing

Given a matrix $\mathbf{X}$, let $\mathbf{X}^{(\mathcal{I})}$ and $\mathbf{X}_{(\mathcal{I})}$ denote respectively the submatrices resultant of selecting the rows and columns in the index set $\mathcal{I}$. Given a vector $\mathbf{x}$, $\mathbf{x}_{(\mathcal{I})}$ is the subvector resultant of selecting the elements in the index set $\mathcal{I}$.

Chapter 2

Gridless methodology based on Atomic Norm for sparse model recovery

We consider a continuous set of structured parametric vectors in $\mathbb{C}^J$ defined as:

$$\mathcal{A}_{\mathsf{J}} = \left\{ \mathbf{a}_{\mathsf{J}}(\mathsf{f}) = \frac{1}{\sqrt{J}} \left[e^{j2\pi \mathsf{j}_1^\top \mathsf{f}}, e^{j2\pi \mathsf{j}_2^\top \mathsf{f}}, \dots, e^{j2\pi \mathsf{j}_J^\top \mathsf{f}} \right]^\top : \ \mathsf{f} \in \mathbb{T}^d \right\}, \tag{2.1}$$

where j_n with $n \in [J]$ are vectors of dimension d as $\mathsf{j}_n = \left[\mathsf{j}_n^1, \dots, \mathsf{j}_n^d\right]^\top$ whose elements j_n^i take values $\mathsf{j}_n^i \in \{0, 1, \dots, \mathsf{J}_i - 1\}$ with $i \in [d]$. Thus,

$$\begin{bmatrix} \mathsf{j}_1^\top \\ \mathsf{j}_2^\top \\ \vdots \\ \mathsf{j}_J^\top \end{bmatrix} = \begin{bmatrix} \mathsf{j}_1^1 & \cdots & \mathsf{j}_1^d \\ \vdots & \ddots & \vdots \\ \mathsf{j}_J^1 & \cdots & \mathsf{j}_J^d \end{bmatrix} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \mathsf{J}_d - 1 \\ 0 & 0 & \cdots & 1 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \mathsf{J}_d - 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathsf{J}_1 - 1 & \mathsf{J}_2 - 1 & \cdots & \mathsf{J}_{d-1} - 1 & \mathsf{J}_d - 1 \end{bmatrix}.$$

Then, the structure of the so called *atoms* $\mathbf{a}_{\mathsf{J}}(\mathsf{f})$ belonging to the set $\mathcal{A}_{\mathsf{J}}$ is fully characterized by vector $\mathsf{J} = [\mathsf{J}_1, \mathsf{J}_2, \dots, \mathsf{J}_d]^\top$ containing the number of elements per dimension and the d-dimensional parameter $\mathsf{f} = [\mathsf{f}_1, \mathsf{f}_2, \dots, \mathsf{f}_d]^\top \in \mathbb{T}^d$.

This definition of the *atom set* $\mathcal{A}_{\mathsf{J}}$ yields a more compact representation where one single atom can be a concatenation of Kronecker products as

$$\begin{aligned}\mathbf{a}_{\mathsf{J}}(\boldsymbol{\mathsf{f}}) =&\frac{1}{\sqrt{J}}\left[e^{j2\pi\mathbf{j}_1^\top\boldsymbol{\mathsf{f}}}, e^{j2\pi\mathbf{j}_2^\top\boldsymbol{\mathsf{f}}}, \dots, e^{j2\pi\mathbf{j}_J^\top\boldsymbol{\mathsf{f}}}\right]^\top \\ =&\mathbf{a}_{\mathsf{J}}(\mathsf{f}_1)\otimes\mathbf{a}_{\mathsf{J}}(\mathsf{f}_2)\otimes\cdots\otimes\mathbf{a}_{\mathsf{J}}(\mathsf{f}_d) \\ =&\otimes_{i=1}^{d}\mathbf{a}_{\mathsf{J}}(\mathsf{f}_i),\end{aligned} \tag{2.2}$$

with $\mathbf{a}_{\mathsf{J}}(\mathsf{f}_i) = \frac{1}{\sqrt{\mathsf{J}_i}}\left[1, e^{j2\pi 1\mathsf{f}_i}, e^{j2\pi 2\mathsf{f}_i}, \dots, e^{j2\pi(\mathsf{J}_i-1)\mathsf{f}_i}\right]^\top \in \mathcal{A}_{\mathsf{J}_i}$.

This atom set arises in many scenarios, as it will be seen in next chapters, such as wireless propagation of electromagnetic waves or in d-D Fourier transform [4]–[6].

In this thesis we deal with signal models $\mathbf{x} \in \mathbb{C}^J$ which are a sparse linear combination of K atoms in $\mathcal{A}_{\mathsf{J}}$, and the sparsity is coming from the fact that K is very small if compared to the dimension of the data itself J, i.e $K \ll J$:

$$\mathbf{x} = \sum_{k=1}^{K} u_k\mathbf{a}_{\mathsf{J}}(\boldsymbol{\mathsf{f}}_k) = [\mathbf{a}_{\mathsf{J}}(\boldsymbol{\mathsf{f}}_1), \dots, \mathbf{a}_{\mathsf{J}}(\boldsymbol{\mathsf{f}}_K)]\,\mathbf{u} = \mathbf{A}_{\mathsf{J}}(\boldsymbol{\mathsf{f}}_{1:K})\mathbf{u} \tag{2.3}$$

where $u_k \in \mathbb{C}$, $\mathbf{a}_{\mathsf{J}}(\boldsymbol{\mathsf{f}}_k) \in \mathcal{A}_{\mathsf{J}}$ with $k \in K$, $\mathbf{A}_{\mathsf{J}}(\boldsymbol{\mathsf{f}}_{1:K}) = [\mathbf{a}_{\mathsf{J}}(\boldsymbol{\mathsf{f}}_1), \dots, \mathbf{a}_{\mathsf{J}}(\boldsymbol{\mathsf{f}}_K)]$ and $\boldsymbol{\mathsf{f}}_{1:K}$ represent the set of frequencies $\{\boldsymbol{\mathsf{f}}_1, \dots, \boldsymbol{\mathsf{f}}_K\}$, all of them assumed to be different and containing relevant information on the signal $\mathbf{x}$. Special attention must be given to the ordering in the Kronecker product in (2.2). Let the notation $\mathsf{J}_d \to \mathsf{J}_{d-1} \to \cdots \to \mathsf{J}_1$ denote an ordering such that first goes J_d, then one by one through J_{d-1}, to the last J_1. Note that this ordering induces a particular indexing in signal $\mathbf{x}$, which comes from the atoms $\mathbf{a}_{\mathsf{J}}(\boldsymbol{\mathsf{f}}_k)$, $k \in [K]$. The Kronecker ordering can be arbitrary, which implies that it can be properly chosen.

Particularly we are interested in measurement scenarios such that the measured signal $\mathbf{y} \in \mathbb{C}^S$ is a noisy sampled version of $\mathbf{x}$:

$$\mathbf{y} = \mathbf{B}\mathbf{x} + \mathbf{w} = \mathbf{B}\sum_{k=1}^{K} u_k\mathbf{a}_{\mathsf{J}}(\boldsymbol{\mathsf{f}}_k) + \mathbf{w} \tag{2.4}$$

where $\mathbf{w} \in \mathbb{C}^S$ is the measurement noise, and the so–called *sampling matrix* $\mathbf{B}$ is a binary fat matrix given by

$$\mathbf{B} = \left[\mathbf{I}_S|\mathbf{0}_{S\times(J-S)}\right]\boldsymbol{\Pi} \in \{0,1\}^{S\times J} \tag{2.5}$$

where $\boldsymbol{\Pi}$ is a permutation matrix and $S \leq J$. One very clear example of that measurement scenario is the received signal on an array of antennas where some of the antennas are not active (see chapter 3). In this scenario the parameter set $\boldsymbol{\mathsf{f}}_{1:K}$ would contain relevant information on the propagation characteristic of the electromagnetic waves, such as the AoA.

In summary, in this thesis we target the following problem, applied to different wireless propagation and positioning scenarios:

Problem 1. *Knowing the observed* ***noisy*** *vector* $\mathbf{y} = \mathbf{B}\mathbf{x}+\mathbf{w} = \mathbf{B}\sum_{k=1}^{K} u_k \mathbf{a}_{\mathbf{J}}(\mathsf{f}_k)+\mathbf{w} \in \mathbb{C}^S$ *and knowing the matrix that maps the linear measurements* $\mathbf{B} \in \mathbb{C}^{S\times J}$*, the objective is to* ***robustly recover***[1] *the signal data* $\mathbf{x}$*, the atoms* $\mathbf{a}_{\mathbf{J}}(\mathsf{f}_k)$ *for* $k \in [K]$ *alongside with the recovery of the parameter* f_k*.*

With this objective in mind the challenge is to gridlessly recover the signal model $\mathbf{x}$ and the atoms $\mathbf{a}_{\mathbf{J}}(\mathsf{f}_k)$ with their parameters f_k for $k \in [K]$, from the observed signal $\mathbf{y}$. For doing so, the model $\mathbf{x}$ needs to be sparse, *i.e.* it must follow (2.3). From that, a new challenge arises that is how to accomplish a sparse model on $\mathbf{x}$ by enforcing a very small number of degrees of freedom in the atom set $\mathcal{A}_{\mathbf{J}}$. Additionally, a procedure is needed to identify the atoms, $\{\mathbf{a}_{\mathbf{J}}(\mathsf{f}_1), \mathbf{a}_{\mathbf{J}}(\mathsf{f}_2), \ldots, \mathbf{a}_{\mathbf{J}}(\mathsf{f}_K)\}$, among the atom set $\mathcal{A}_{\mathbf{J}}$ or the frequency set $\{\mathsf{f}_1, \mathsf{f}_2, \ldots, \mathsf{f}_K\}$ that composes $\mathbf{x}$.

Next, we present two definition of norms [30], [31] and the corresponding associated optimization problems as a first approach to data signal $\mathbf{x}$ recovery.

2.1 Atomic norm definitions and associated optimization problems

We define two norms, analogously to ℓ_0 and ℓ_1, restricted to the span$\{\mathcal{A}_{\mathbf{J}}\} \subseteq \mathbb{C}^J$ of the atom set $\mathcal{A}_{\mathbf{J}}$. These norms will be denoted as atomic norms (AN) and are defined below.

Definition 1. *Given a J-D vector* $\mathbf{x}$*, its* ℓ_0*-AN (*$\ell_{\mathcal{A}_{\mathbf{J}},0}$*) in* $\mathcal{A}_{\mathbf{J}}$ *is defined as*

$$\|\mathbf{x}\|_{\mathcal{A}_{\mathbf{J}},0} = \inf_{\mathsf{f}\in\mathbb{T}^d, u_k\in\mathbb{C}} \left\{ K : \ \mathbf{x} = \sum_{k=1}^{K} u_k \mathbf{a}_{\mathbf{J}}(\mathsf{f}_k) \right\}.$$

□

Definition 2. *Given a J-D vector* $\mathbf{x}$*, its* ℓ_1*-AN (*$\ell_{\mathcal{A}_{\mathbf{J}},1}$*) in* $\mathcal{A}_{\mathbf{J}}$ *is defined as*

$$\|\mathbf{x}\|_{\mathcal{A}_{\mathbf{J}},1} = \inf_{\mathsf{f}\in\mathbb{T}^d, u_k\in\mathbb{C}} \left\{ \sum_{k=1}^{K} |u_k| : \ \mathbf{x} = \sum_{k=1}^{K} u_k \mathbf{a}_{\mathbf{J}}(\mathsf{f}_k) \right\}.$$

□

Note that the definition of the $\ell_{\mathcal{A}_{\mathbf{J}},0}$ enforces the sparsest linear combination of a signal vector $\mathbf{x}$ in the atom set $\mathcal{A}_{\mathbf{J}}$, therefore, given the measurement $\mathbf{y}$ in (2.4), we aim at targeting Problem 1 and recover $\mathbf{x}$ by solving [1]:

$$\min_{\mathbf{x}} (1-\eta) \|\mathbf{y} - \mathbf{B}\mathbf{x}\|_2^2 + \eta \|\mathbf{x}\|_{\mathcal{A}_{\mathbf{J}},0} \tag{P.1}$$

---

[1] A recovery scheme is said to be robust if the distance between recovered and true signals is only controlled by the noise power.

wherein we require that $\mathbf{Bx}$ be close (in the ℓ_2-norm) to $\mathbf{y}$, while forcing $\mathbf{x}$ to be a sparse linear combination of the atoms in $\mathcal{A}_{\mathbf{J}}$. Here, $\eta \in (0,1)$ is a parameter ruling the sparsity of the solution enforced by the $\ell_{\mathcal{A}_{\mathbf{J}},0}$ AN.

Similarly, using the definition of the $\ell_{\mathcal{A}_{\mathbf{J}},1}$ we still can enforce a linear combination of a signal vector $\mathbf{x}$ in the atom set $\mathcal{A}_{\mathbf{J}}$, although we cannot ensure that is the sparsest. Then to approach Problem 1 and recover $\mathbf{x}$ we also propose to solve as follows [1]:

$$\min_{\mathbf{x}}(1-\eta)\left\|\mathbf{y}-\mathbf{Bx}\right\|_2^2+\eta\left\|\mathbf{x}\right\|_{\mathcal{A}_{\mathbf{J}},1} \tag{P.2}$$

wherein we require that $\mathbf{Bx}$ be close (in the ℓ_2-norm) to $\mathbf{y}$, while forcing $\mathbf{x}$ to be a linear combination of the atoms in $\mathcal{A}_{\mathbf{J}}$ with the minimum ℓ_1 norm.

Unfortunately, (P.1) and (P.2) present some drawbacks that must be taken into consideration. Although the solution of any of both optimization problems would enforce respectivelly the sparsest solution in $\mathcal{A}_{\mathbf{J}}$ and the minimum ℓ_1 norm in $\mathcal{A}_{\mathbf{J}}$, we still need to address how to solve on the norms $\ell_{\mathcal{A}_{\mathbf{J}},0}$ and $\ell_{\mathcal{A}_{\mathbf{J}},1}$, also there is no insights on the atoms that build the model, *i.e.* the solution does not provide a procedure to recover the composite atoms.

Next, following a similar procedure as the one presented in [31] to solve optimization problems on ℓ_0 and ℓ_1 norms, we illustrate how the $\ell_{\mathcal{A}_{\mathbf{J}},0}$ can be computed by solving a rank minimization problem, and how the $\ell_{\mathcal{A}_{\mathbf{J}},1}$ can be computed by solving a nuclear norm minimization problem.

2.2 Rank and nuclear norm optimization to solve atomic norms

Given a signal as in (2.3) with atoms in $\mathcal{A}_{\mathbf{J}}$ we can find a so called a multi–level Toeplitz (MLT) matrix structure that arises as the signal covariance matrix [22]:

Definition 3. *Given the ordered vector* $\mathsf{J} = [\mathsf{J}_1, \dots, \mathsf{J}_d]^\top$ *with ordering* $\mathsf{J}_d \to \cdots \to \mathsf{J}_1$*, a* $J \times J$ *d-LT matrix* $\mathbf{T}_{\mathbf{J}}$ *is defined recursively as follows:*

$$\mathbf{T}_{\mathbf{J}} = \begin{bmatrix} \mathbf{T}_{0;\mathbf{J}_{2:d}} & \mathbf{T}_{1;\mathbf{J}_{2:d}} & \cdots & \mathbf{T}_{(\mathsf{J}_1-1);\mathbf{J}_{2:d}} \\ \mathbf{T}^\dagger_{1;\mathbf{J}_{2:d}} & \mathbf{T}_{0;\mathbf{J}_{2:d}} & \cdots & \mathbf{T}_{(\mathsf{J}_1-2);\mathbf{J}_{2:d}} \\ \vdots & & & \vdots \\ \mathbf{T}^\dagger_{(\mathsf{J}_1-1);\mathbf{J}_{2:d}} & \mathbf{T}^\dagger_{(\mathsf{J}_1-2);\mathbf{J}_{2:d}} & \cdots & \mathbf{T}_{0;\mathbf{J}_{2:d}} \end{bmatrix} \tag{2.6}$$

Thus, matrix $\mathbf{T}_{\mathbf{J}}$ *has* $\mathsf{J}_1 \times \mathsf{J}_1$ *blocks* $\mathbf{T}_{a;\mathbf{J}_{2:d}}$ *with* $0 \le a \le \mathsf{J}_1 - 1$*. Likewise, each* *block is defined in a recursive way having* $\mathsf{J}_2 \times \mathsf{J}_2$ *sub–blocks:*

$$\mathbf{T}_{a;\mathbf{J}_{2:d}} = \begin{bmatrix} \mathbf{T}_{a;0;\mathbf{J}_{3:d}} & \mathbf{T}_{a;1;\mathbf{J}_{3:d}} & \cdots & \mathbf{T}_{a;(\mathsf{J}_2-1);\mathbf{J}_{3:d}} \\ \mathbf{T}^{\dagger}_{a;1;\mathbf{J}_{3:d}} & \mathbf{T}_{a;0;\mathbf{J}_{3:d}} & \cdots & \mathbf{T}_{a;(\mathsf{J}_2-2);\mathbf{J}_{3:d}} \\ \vdots & & & \vdots \\ \mathbf{T}^{\dagger}_{a;(\mathsf{J}_2-1);\mathbf{J}_{3:d}} & \mathbf{T}^{\dagger}_{a;(\mathsf{J}_2-2);\mathbf{J}_{3:d}} & \cdots & \mathbf{T}_{a;0;\mathbf{J}_{3:d}} \end{bmatrix} \tag{2.7}$$

Again, each sub–block $\mathbf{T}_{a;b;\mathbf{J}_{3:d}}$ *with* $0 \leq b \leq \mathsf{J}_2 - 1$ *would be defined recursively with* $\mathsf{J}_3 \times \mathsf{J}_3$ *blocks. This procedure repeats until the last level, where each block is simply a* $\mathsf{J}_d \times \mathsf{J}_d$ *Hermitian Toeplitz matrix.*

For the 1-D scenario, *i.e.* $d = 1$, Definition 3 states that $\mathbf{T}_{\mathsf{J}_1}$ is a Hermitian Toeplitz matrix. Result from Carathéodory and Fejér [32] shows that a positive semidefinite (PSD) rank–deficient Toeplitz matrix with $\text{rank}\{\mathbf{T}_{\mathsf{J}_1}\} = r < \mathsf{J}_1$ can be uniquely decomposed as $\mathbf{T}_{\mathsf{J}_1} = \mathbf{A}_{\mathsf{J}_1}(\mathsf{f}_{1:r})\mathbf{P}\mathbf{A}^{\dagger}_{\mathsf{J}_1}(\mathsf{f}_{1:r})$, being $\mathbf{A}_{\mathsf{J}_1}(\mathsf{f}_{1:r})$ a $\mathsf{J}_1 \times r$ Vandermonde matrix whose r columns are atoms in the atom set $\mathcal{A}_{\mathsf{J}_1}$ with 1-D frequencies $\{\mathsf{f}_1, \ldots, \mathsf{f}_r\}$ and being $\mathbf{P}$ a diagonal matrix with real positive elements, *i.e.* in $\mathbb{R}^+$. Thus, if a matrix meets the conditions of being a PSD rank–deficient Toeptlitz matrix, it is at the same time the covariance matrix of a mixture of 1-D complex atoms from $\mathcal{A}_{\mathsf{J}_1}$. In the applications studied in this thesis one dimension is not enough so a challenge arises, that is, extending the result from [32] to d-D scenarios where $d > 1$. In addition, not only the result from [32] must be valid for multidimensional cases but it is also necessary a Vandermonde–like decomposition, parametrized by a d-D frequency set $\{\mathbf{f}_1, \ldots, \mathbf{f}_r\}$, that can be applied to a PSD rank–deficient MLT matrix. Similarly to the 1-D case, for the d-D case, if a matrix is PSD, rank–deficient and has a MLT structure, it is at the same time the covariance matrix of the mixture of r d-D atoms from $\mathcal{A}_{\mathbf{J}}$. This challenge was first introduced by [22] and [24] gave further extension to it.

This thesis will use the Vandermonde decomposition result for a multidimensional block Toeplitz matrix given in [24, Lemma 1, Remark 3]:

LEMMA 1. *Let* $\mathbf{T}_{\mathbf{J}} \in \mathbb{C}^{J \times J}$ *be a PSD d-LT matrix with rank* $r < J$. *If the rank of the* $\mathsf{J}_d \times \mathsf{J}_d$ *upper–left corner of* $\mathbf{T}_{\mathbf{J}}$ *is also equal to* r *and* $r < \mathsf{J}_d$*, then,* $\mathbf{T}_{\mathbf{J}}$ *can be uniquely decomposed via Algorithm 1, as* $\mathbf{T}_{\mathbf{J}} = \mathbf{A}_{\mathbf{J}}(\mathbf{f}_{1:r})\mathbf{P}\mathbf{A}^{\dagger}_{\mathbf{J}}(\mathbf{f}_{1:r})$*, with* $\mathbf{f}_{1:r} = \{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_r\} \in \mathbb{T}^{d \times r}$*, being a unique set of frequencies,* $\mathbf{A}_{\mathbf{J}}(\mathbf{f}_{1:r}) = [\mathbf{a}_{\mathbf{J}}(\mathbf{f}_1), \mathbf{a}_{\mathbf{J}}(\mathbf{f}_2), \ldots, \mathbf{a}_{\mathbf{J}}(\mathbf{f}_r)] \in \mathbb{C}^{J \times r}$ *with* $\mathbf{a}_{\mathbf{J}}(\mathbf{f}_k) \in \mathcal{A}_{\mathbf{J}}$*, being the atom matrix parametrized by the vector of frequencies* $\mathbf{f}_{1:r}$*, and* $\mathbf{P} = \text{diag}([p_1, \ldots, p_r])$*,* $p_k \in \mathbb{R}^+$ *with* $k \in [r]$. ☐

As mentioned before, the ordering of the Kronecker product in (2.2) and the induced indexing of the signal as in (2.3) deserve especial attention. Lemma 1 and its associated Algorithm 1 show the interesting relationship between that Kronecker ordering of the atoms decomposing the matrix and the nesting ordering of the d-LT matrix. The d-LT matrix is the input for Algorithm 1 which at the same time is the covariance matrix of the signal $\mathbf{x}$ whose indexing comes from that Kronecker ordering, and hence, both orderings are related. Let the ordering of the d-LT matrix be $\mathsf{J}_d \to \mathsf{J}_{d-1} \to \cdots \to \mathsf{J}_1$. This forces the output atoms from the decomposition in Algorithm 1 to follow an identical Kronecker structure $\mathsf{J}_d \to \mathsf{J}_{d-1} \to \cdots \to \mathsf{J}_1$.

Furthermore, Lemma 1 states that the uniqueness of the Vandermonde decomposition of the d-LT matrix $\mathbf{T}_{\mathbf{J}}$ is in fact given by the dimension of the innermost Toeplitz block of $\mathbf{T}_{\mathbf{J}}$. At least, it is $\min_i\{\mathsf{J}_i\}$, which would be equivalent to the result in [22, Theorem 1]. Given the aforementioned relation between the Kronecker structure of the signal and the nesting ordering in the corresponding d-LT matrix, and recalling that the Kronecker ordering is arbitrary and can be tailored, a proper ordering of the nesting structure in $\mathbf{T}_{\mathbf{J}}$ would yield an innermost block with dimension $\max_i\{\mathsf{J}_i\}$. This ordering will be called as *canonical* ordering and it gives a nesting structure in the d-LT matrix $\mathbf{T}_{\mathbf{J}}$ which will be called as *canonical* structure and will be the structure considered in this thesis.

DEFINITION 4. *A d-LT matrix is said to posses a canonical structure if the vector* $\mathbf{J} = [\mathsf{J}_1, \dots, \mathsf{J}_d]^\top$*, which contains the number of Fourier samples in each of the d dimensions, has an ordering such that* $\mathsf{J}_1 \le \mathsf{J}_2 \le \cdots \le \mathsf{J}_d$*. The matrix set of PSD canonical d-LT matrix is denoted as* $\mathcal{T}$*.*

We recall next the conditions and procedure to solve $\ell_{\mathcal{A}_{\mathbf{J}},0}$ and $\ell_{\mathcal{A}_{\mathbf{J}},1}$ in the noiseless case, using the results in [24], and from there accordingly reformulate (P.1) and (P.2).

Without loss of generality, the Kronecker ordering of the atoms in (2.2) will be always assumed to be such that $\mathsf{J}_1 \le \mathsf{J}_2 \le \cdots \le \mathsf{J}_d$ where $\mathsf{J}_i,\ i \in [d]$, is the number of samples in each of the d dimensions. Although this may not be the original situation, the reading of the elements of the data signal vector can be performed accordingly to enforce it. Note that this will ensure that the covariance matrix of the data signal has a canonical structure as defined in Definition 4 in every situation. Also, without loss or generality,

ASSUMPTION 1. *The d-D frequency vectors* f_k *with* $k \in [K]$ *are assumed to be identically distributed random vectors with independent components uniformly distributed on* $[0, 1)$ *such that* $\mathsf{f}_k^i \sim \mathcal{U}[0, 1)$*.*

DEFINITION 5. $\mathcal{B}^{(K)} \subset \{0,1\}^{S\times J}$ *denotes the subset of all binary sampling matrices as in* (2.5) *for which it exists a subset of rows of* $\mathbf{B} \in \{0,1\}^{S\times J}$*, say* $\mathcal{L} \subseteq [S]$ *with* $2K < |\mathcal{L}| \le S$ *such that:*

Condition 1. *Under Assumption 1,* $\mathbf{B}^{(\mathcal{L})}\mathbf{A}_{\mathbf{J}}(\mathsf{f}_{1:2K})$ *is injective as map from* $\mathbb{C}^{2K} \to \mathbb{C}^{|\mathcal{L}|}$*;*

Condition 2. *denoted by* $\mathsf{f}_{1:K}$ *a set of K frequencies satisfying Assumption 1 and by* $\mathsf{f}^\diamond_{1:r^\diamond}$ *an arbitrary set of* $r^\diamond \le K$ *frequencies such that for any* $k \in [K]$

and $i \in [r^\diamond]$, $\mathbf{f}_i^\diamond \neq \mathbf{f}_k$, $\operatorname{rank}\left\{\mathbf{B}^{(\mathcal{L})}\mathbf{A}_{\mathsf{J}}(\mathbf{f}^\diamond_{1:K+r^\diamond})\right\} = K + \operatorname{rank}\left\{\mathbf{B}^{(\mathcal{L})}\mathbf{A}_{\mathsf{J}}(\mathbf{f}^\diamond_{1:r^\diamond})\right\}$, *with* $\mathbf{f}^\diamond_{1:K+r^\diamond} = \left[\mathbf{f}_{1:K}, \mathbf{f}^\diamond_{1:+r^\diamond}\right]$. □

Definition 5 sets the mathematical conditions under which the sampling matrix $\mathbf{B}$ allows unique recovery of the data signal. Next we provide the Theorem that states the equivalence between optimizing for $\ell_{\mathcal{A}_{\mathsf{J}},0}$ and a rank minimization problem.

THEOREM 1. *[24, Th. 1] Consider a measured signal* $\mathbf{y} = \mathbf{B}\mathbf{x} = \mathbf{B}\sum_{k=1}^{K} u_k \mathbf{a}_{\mathsf{J}}(\mathbf{f}_k)$ *where the atoms are defined in the atom set domain* $\mathcal{A}_{\mathsf{J}}$. *Under Assumption 1, provided that* $K < \max_{i\in[d]}\{\mathsf{J}_i\}$, *if* $\mathbf{B} \in \mathcal{B}^{(K)}$ *as in Definition 5, the sparse model* $\mathbf{x}$ *in* (2.3) *and its associated atoms/frequencies* $\mathbf{f}_{1:K} \in \mathbb{T}^{d\times K}$ *can be uniquely identified by solving the optimization problem*

$$\begin{aligned} \min_{r,\mathbf{x}\in\mathbb{C}^J,\mathbf{T}_{\mathsf{J}}\in\mathcal{T}} \quad & \operatorname{rank}\{\mathbf{T}_{\mathsf{J}}\} \\ \text{s.t.} \quad & \begin{bmatrix} \mathbf{T}_{\mathsf{J}} & \mathbf{x} \\ \mathbf{x}^\dagger & r \end{bmatrix} \succeq 0 \quad \mathbf{y} = \mathbf{B}\mathbf{x} \end{aligned} \tag{P.3}$$

where $\mathcal{T} \subseteq \mathbb{C}^{J\times J}$ *denotes the set of all canonical PSD d-LT matrices (Definition 4). The frequencies* $\mathbf{f}_{1:K} \in \mathbb{T}^{d\times K}$ *can be obtained by Vandermonde decomposition, via Algorithm 1, of the d-LT matrix* $\mathbf{T}_{\mathsf{J}}$ *solution of* (P.3).

It is interesting to note that the constraint $\begin{bmatrix} \mathbf{T}_{\mathsf{J}} & \mathbf{x} \\ \mathbf{x}^\dagger & r \end{bmatrix} \succeq 0$ implies that $\mathbf{x}$ lies in the column span of $\mathbf{T}_{\mathsf{J}}$ according to the Schur complement lemma. Given that the optimal solution of (P.3) also ensures unique multi–level Vandermonde decomposition of the optimal $\mathbf{T}_{\mathsf{J}}$ ($\mathbf{T}^\circ_{\mathsf{J}}$) as in Lemma 1, consequently, the optimal vector $\mathbf{x}^\circ$ solving (P.3) must be a linear combination of the atoms taken from $\mathcal{A}_{\mathsf{J}}$, possesing by construction the same nesting order of the sub–blocks composing $\mathbf{T}^\circ_{\mathsf{J}}$. Also $\operatorname{rank}\{\mathbf{T}^\circ_{\mathsf{J}}\} = r^\circ = K$.

To solve $\ell_{\mathcal{A}_{\mathsf{J}},1}$ as a nuclear norm minimization problem we recall next Theorem:

THEOREM 2. *[24, Th. 3] Consider a measured signal* $\mathbf{y} = \mathbf{B}\mathbf{x} = \mathbf{B}\sum_{k=1}^{K} u_k \mathbf{a}_{\mathsf{J}}(\mathbf{f}_k)$ *where the atoms are defined in the atom set domain* $\mathcal{A}_{\mathsf{J}}$. *The sparse model* $\mathbf{x}$ *in* (2.3) *and its associated atoms/frequencies* $\mathbf{f}_{1:K} \in \mathbb{T}^{d\times K}$ *can be uniquely identified by solving the optimization problem*

$$\begin{aligned} \min_{t,\mathbf{x}\in\mathbb{C}^J,\mathbf{T}_{\mathsf{J}}\in\mathcal{T}} \quad & \frac{1}{2}t + \frac{1}{2}\operatorname{Tr}\{\mathbf{T}_{\mathsf{J}}\} \\ \text{s.t.} \quad & \begin{bmatrix} \mathbf{T}_{\mathsf{J}} & \mathbf{x} \\ \mathbf{x}^\dagger & t \end{bmatrix} \succeq 0 \quad \mathbf{y} = \mathbf{B}\mathbf{x} \end{aligned} \tag{P.4}$$

if under Assumption 1, and with $\mathbf{B} \in \mathcal{B}^{(K)}$ *as in Definition 5, we have that* i) $K < \max_{i\in[d]}\{\mathsf{J}_i\}$ *and that* ii) *the optimal solution to* (P.4) *denoted by* $\left(t^\circ, \mathbf{x}^\circ_{\ell_{\mathcal{A}_{\mathsf{J}},1}}, \mathbf{T}^\circ_{\mathsf{J}}\right)$, *also satisfies* $r^\circ_{\ell_{\mathcal{A}_{\mathsf{J}},1}} < \max_{i\in[d]}\{\mathsf{J}_i\}$.

Again, for the nuclear norm optimization problem, the frequencies $\mathsf{f}_{1:K} \in \mathbb{T}^{d\times K}$ can be obtained by Vandermonde decomposition, via Algorithm 1, of the d-LT matrix $\mathbf{T}_{\mathsf{J}}$ solution of (P.4).

With these two results we can reformulate (P.1) and (P.2) as follows:

$$\begin{aligned} &\min_{\mathbf{x}}(1-\eta)\left\|\mathbf{y}-\mathbf{B}\mathbf{x}\right\|_2^2 + \eta \mathrm{rank}\{\mathbf{T}_{\mathsf{J}}\} \\ &\qquad\qquad\qquad \text{s.t.} \begin{bmatrix} \mathbf{T}_{\mathsf{J}} & \mathbf{x} \\ \mathbf{x}^{\dagger} & r \end{bmatrix} \succeq 0 \end{aligned} \tag{P.5}$$

$$\begin{aligned} &\min_{\mathbf{x}}(1-\eta)\left\|\mathbf{y}-\mathbf{B}\mathbf{x}\right\|_2^2 + \eta\left(\frac{1}{2}t + \frac{1}{2}\mathrm{Tr}\left\{\mathbf{T}_{\mathsf{J}}\right\}\right) \\ &\qquad\qquad\qquad \text{s.t.} \begin{bmatrix} \mathbf{T}_{\mathsf{J}} & \mathbf{x} \\ \mathbf{x}^{\dagger} & t \end{bmatrix} \succeq 0 \end{aligned} \tag{P.6}$$

These two optimization problems, particularly the convex relaxation formulation in (P.6), will be used along this thesis to solve and recover relevant parameters in signals constructed as a sparse linear combination of atoms in $\mathcal{A}_{\mathsf{J}}$. Similarly to what was done in (P.3) and (P.4) once we have the optimal solution to (P.6), the frequencies $\mathsf{f}_{1:K} \in \mathbb{T}^{d\times K}$ can be obtained by Vandermonde decomposition, via Algorithm 1, of the d-LT matrix $\mathbf{T}_{\mathsf{J}}$ solution.

2.2.1 Sampling matrices with injectivity properties

The conditions in Definition 5 may appear difficult to be achieved, however, they can be easily met for some practical sampling schemes $\mathbf{B}$. For that, we need to find the largest non-trivial atom with structure in $\mathcal{A}_{\mathsf{S}}$ with $\mathsf{S} = [\mathsf{S}_1, \ldots \mathsf{S}_d]$ inside the sampled atom $\mathbf{B}\mathbf{a}_{\mathsf{J}}(\mathsf{f})$, or equivalently a permutation matrix $\boldsymbol{\Pi}_o$, such that

$$\mathbf{B}\mathbf{a}_{\mathsf{J}}(\mathsf{f})\boldsymbol{\Pi}_o = \begin{bmatrix} e^{j2\pi\boldsymbol{\Delta}\mathsf{f}}\mathbf{a}_{\mathsf{S}}(\mathbf{g}) \\ \mathbf{q} \end{bmatrix} \tag{2.8}$$

where $\mathbf{a}_{\mathsf{S}}(\mathbf{g}) \in \mathcal{A}_{\mathsf{S}}$ with $\mathsf{J} \succeq \mathsf{S}$, $\mathbf{g}$ is any permutation of $\left[\alpha_1\mathsf{f}^1, \ldots, \alpha_d\mathsf{f}^d\right]^\top$ with α_i a proper positive integer, $\boldsymbol{\Delta} = [\Delta_1, \ldots, \Delta_d]$ and $\Delta_i \in [\mathsf{J}_i]$, $i \in [d]$, $\mathbf{q} \in \mathbb{C}^{J-\prod_{i=1}^{d}\mathsf{S}_i}$. We define the following parameters $L_S = \sum_{i=1}^{d}\mathsf{S}_i$ and $N_S = \prod_{i=1}^{d}\mathsf{S}_i$, and we identify a non-trivial atom inside $\mathbf{B}\mathbf{a}_{\mathsf{J}}(\mathsf{f})$ only in those cases where $L_S \geq (d+1)$ or $N_S \geq 2$.

If we have found the non-trivial atom inside $\mathbf{B}\mathbf{a}_{\mathsf{J}}(\mathsf{f})$, to meet Definition 5 conditions we need that $L_S - (d-1) \geq 2K$ [24, Col. 1]. Also therein we can find a conjecture, under which Definition 5 conditions still holds if $N_S > 2K$.

Along the results presented in this thesis, we will make sure that these conditions are met in order to ensure perfect frequency recovery in the noiseless case as a prerequisite for accurate parameter estimation in the presence of noise.

2.3 Computational complexity

According to [33], an interior point method to solve a semidefinite optimization problem requires

$$O\left(m(n^3 + mn^2 + m^2)\sqrt{n}\log(1/\epsilon)\right) \tag{2.9}$$

arithmetic operations, where n is the dimension of the variable to be optimized, m is the number of constraints and ϵ is the accuracy parameter.

Using (2.9), the complexity of either (P.6) or of any convex combination of [22, Eq. 45] with ℓ_2 norm ([22, Eq. 53]) is obtained upon setting $n = J$ and $m = 1$. The complexity in all these scenarios is said to be given by $O(J^{3.5})$.

Algorithm 1 Vandermonde decomposition of a d-LT matrix

```
Definitions:
  $\mathbf{\mathsf{J}} = [\mathsf{J}_1, \dots, \mathsf{J}_d]$ and $\mathbf{\mathsf{J}}_{t:d} = [\mathsf{J}_t, \dots, \mathsf{J}_d]$.
  $\mathbf{A}_{\mathbf{\mathsf{J}}_{t:d}}(\mathbf{\mathsf{f}}_{1:r}^{t:d}) = \odot_{s=t}^{d} \mathbf{A}_{\mathsf{J}_s}(\mathbf{\mathsf{f}}_{1:r}^{s})$ with $t \in [d]$ and:
    $\mathbf{\mathsf{f}}_{1:r}^{t:d} = [\mathbf{\mathsf{f}}_1^{t:d}, \dots, \mathbf{\mathsf{f}}_r^{t:d}]$, being $\mathbf{\mathsf{f}}_k^{t:d} = [\mathsf{f}_k^t, \dots, \mathsf{f}_k^d]^\top$.
    $\mathbf{A}_{\mathsf{J}_s}(\mathbf{\mathsf{f}}_{1:r}^{s}) = [\mathbf{a}_{\mathsf{J}_s}(\mathsf{f}_1^s), \dots, \mathbf{a}_{\mathsf{J}_s}(\mathsf{f}_r^s)]$, being $\mathbf{a}_{\mathsf{J}_s}(\mathsf{f}_k^s) = \left[1, e^{j2\pi \mathsf{f}_k^s}, \dots, e^{j2\pi(\mathsf{J}_s - 1)\mathsf{f}_k^s}\right]$.
Input: $\mathbf{T}_{\mathbf{\mathsf{J}}}$ with $\text{rank}\{\mathbf{T}_{\mathbf{\mathsf{J}}}\} = r$.
1) Obtain the Cholesky decomposition of $\mathbf{T}_{\mathbf{\mathsf{J}}} = \mathbf{C}_{1:d}\mathbf{C}_{1:d}^\dagger$.
Decomposing in the spatial $i$ dimension
for $i = 1, ..., d-1$ do
  2) Define the sets $\mathcal{I}_i = \{1, \dots, (\mathsf{J}_i - 1)\prod_{s=i+1}^{d} \mathsf{J}_s\}$ and $\mathcal{I}_i^+ = \{1 + \prod_{s=i+1}^{d} \mathsf{J}_s, \dots, \prod_{s=i}^{d} \mathsf{J}_s\}$ and $\mathcal{I}_{i+1} = \{1, \dots, \prod_{s=i+1}^{d} \mathsf{J}_s\}$.
  3) Find the $r \times r$ $\mathbf{U}_i$ unitary matrix such that $\mathbf{C}_{i:d}^{(\mathcal{I}_i)}\mathbf{U}_i = \mathbf{C}_{i:d}^{(\mathcal{I}_i^+)}$.
  4) Obtain the eigen–decomposition $\mathbf{U}_i = \mathbf{K}_i\mathbf{J}_i\mathbf{K}_i^\dagger$ having that $\mathbf{J}_i = \text{diag}\left(\left[e^{j2\pi \mathsf{f}_1'^i}, \dots, e^{j2\pi \mathsf{f}_r'^i}\right]\right)$.
  5) Set the $\prod_{s=i+1}^{d} \mathsf{J}_s \times r$ matrix $\mathbf{C}_{(i+1):d} = \mathbf{C}_{i:d}^{(\mathcal{I}_i+1)}$.
end for
6) Define the sets $\mathcal{I}_d = \{1, \dots, \mathsf{J}_{d-1}\}$ and $\mathcal{I}_d^+ = \{2, \dots, \mathsf{J}_d\}$.
7) Find the $r \times r$ $\mathbf{U}_d$ Unitary matrix such that $\mathbf{C}_d^{(\mathcal{I}_d)}\mathbf{U}_d = \mathbf{C}_d^{(\mathcal{I}_d^+)}$.
8) Obtain the eigen–decomposition $\mathbf{U}_d = \mathbf{K}_d\mathbf{J}_d\mathbf{K}_d^\dagger$ having that $\mathbf{J}_d = \text{diag}\left(\left[e^{j2\pi \mathsf{f}_1'^d}, \dots, e^{j2\pi \mathsf{f}_r'^d}\right]\right)$.
Frequency pairing
9) Set $\mathbf{\mathsf{f}}_{1:r}^{d} = \mathbf{\mathsf{f}}_{1:r}'^{d} = [\mathsf{f}_1'^d, \dots, \mathsf{f}_r'^d]$
for $i = d : d-1 : 1$ do
  10) Get $\mathbf{P}_i = \left(\mathbf{A}_{\mathbf{\mathsf{J}}_{i:d}}\left(\mathbf{\mathsf{f}}_{1:r}^{i:d}\right)\right)^{\mathrm{g}} \mathbf{C}_{i:d}\mathbf{C}_{i:d}^\dagger \left(\mathbf{A}_{\mathbf{\mathsf{J}}_{i:d}}^\dagger\left(\mathbf{\mathsf{f}}_{1:r}^{i:d}\right)\right)^{\mathrm{g}}$
  11) Get $\mathbf{O}_i = 1/\sqrt{\mathbf{P}_i}\left(\mathbf{A}_{\mathbf{\mathsf{J}}_{i:d}}\left(\mathbf{\mathsf{f}}_{1:r}^{i:d}\right)\right)^{\mathrm{g}} \mathbf{C}_{i:d}$.
  12) Define the paired set

        $$\mathbf{\mathsf{f}}_{1:r}^{(i-1):d} = \left[\left(\mathbf{\mathsf{f}}_{1:r}'^{i-1}\right)^\top, \left(\mathbf{\mathsf{f}}_{1:r}^{i:d}\mathbf{O}_i\right)^\top\right]^\top.$$

end for
Output: The recovered paired frequencies $\mathbf{\mathsf{f}}_{1:r}$, the atom set $\mathbf{A}_{\mathbf{\mathsf{J}}}(\mathbf{\mathsf{f}}_{1:r}) = [\mathbf{a}_{\mathbf{\mathsf{J}}}(\mathbf{\mathsf{f}}_1), \dots, \mathbf{a}_{\mathbf{\mathsf{J}}}(\mathbf{\mathsf{f}}_r)]$ and the $r \times r$ diagonal matrix $\mathbf{P}$ such that $\mathbf{T}_{\mathbf{\mathsf{J}}} = \mathbf{A}_{\mathbf{\mathsf{J}}}(\mathbf{\mathsf{f}}_{1:r})\mathbf{P}\mathbf{A}_{\mathbf{\mathsf{J}}}^\dagger(\mathbf{\mathsf{f}}_{1:r})$.
```

Chapter 3

General model for MIMO millimeter-wave channels

New paradigms in wireless communications, fifth (5G) and sixth (6G) generations, demand a huge increase in the systems capacity as well as other requirements such as multidimensional space awareness or ubiquitous connectivity. This has led to the development and leverage of a series of technologies to meet those needs, among which we can find Multiple-Input Multiple-Output (MIMO), carrier aggregation, advanced channel coding and interference coordination, which are used to deploy the antenna arrays and achieve the required capacity while meeting the other requirements. Large antenna arrays have been widely explored [34] and are usually deployed embedded in already existing spaces that have other additional usages, such as surfaces [25], [35] or smart surfaces [36]. However, despite the importance of the aforementioned technologies, the current spectrum is saturated which is incompatible with the needs of the new paradigms. Therefore, it has led researchers and technicians to explore alternative and underused bands.

The GSM Association identifies three bands, or frequency ranges, that meet the needs imposed by the novel paradigms in wireless communications [37]. Starting from lower frequency to higher frequency, the first proposed range is the band below 1 GHz. This band is aimed to give wide coverage and support in all areas no matter they are urban, suburban or rural. Additionally, low rate applications such as Internet of Things (IoT) are also eligible to operate within this band. Next, the frequencies from 1 to 6 GHz are intended to give service to emerging utilities such as intelligent transport including self-driving cars. Last, frequencies above 6 GHz and specially the millimeter-wave (mmWave) range, which goes from 30 GHz to 300 GHz [38], [39], will support ultra-high broadband services, *e.g.* augmented reality, or high resolution positioning systems. The latter range is the one that meets the requirements mentioned above and will be studied in this work.

Although it may seem that the use of mmWave bands has only advantages, the truth is that its nature present severe propagation drawbacks which implies many challenges. One well-known characteristic of propagation in mmWave channels is the extremely high free-space path loss due to 10× increase in carrier frequency if compared to traditional microwave systems. Fortunately, the very same reason makes multidimensional antenna arrays achievable given that the increase in carrier frequency implies a decrease in wavelength which allows to deploy smaller antennas creating small sized arrays. Those arrays can then mitigate the inherent path loss in mmWave channels via MIMO multiplexing and diversity gains.

The aforementioned path loss experienced in mmWave band systems is worsened by any obstacle the signal may encounter, *e.g.* rain, foliage, buildings, walls, specular reflections or scattering, etc, and also by other phenomena such as atmospheric absorption. For the sake of illustration, at 30 GHz (mmWave range lower bound) and at a distance of 100 m, the path loss can be over 102 dB which is a pronounced loss. Furthermore, frequencies around 60 GHz and 180 GHz suffer from atmospheric absorption effects which may cause an extra attenuation of 10 dB/km. All of this often results in a small number of propagation paths between both ends of the communication, *i.e.* low angular spreads both in azimuth and elevation [39] and delay-spreads below 10 ns in outdoor urban environments [38], [40], and below 20 ns in indoor environments [41]. The scarcity of propagation paths between the transmitter side and the receiver side leads to sparse channel models described in [42].

Thus, large MIMO arrays can be leveraged to mitigate all the path loss effects inherent to mmWave systems. Additionally, due to the sparsity of the model, low complexity designs can be achieved.

The aim of this chapter is to define a general model for a point-to-point MIMO mmWave propagation channel system with M transmit antennas and N receive antennas. The model can be easily tailored for multi-user scenarios, if needed. All the applications studied in this thesis will share the channel model defined herein, which will be particularized for each of the scenarios. The geometry of MIMO arrays vary within a wide range, they can go from *i)* uniform arrays where antennas are deployed with a fixed separation in each dimension, *ii)* they can have a “uniform array”-like shape with some missing elements, *e.g.* 3-D arrays usually do not deploy antennas in the inside of the volume, or *iii)* they can have fully arbitrary shapes along the 3-D space, as for example Large Intelligent Surfaces (LIS) [36] or conformal array deployments [3]. Having uniform arrays in 3-D antenna deployments is rare given that they typically will have to adapt to surfaces. In this thesis, non-uniform arrays with a “uniform array”-like shape are studied, where basically the actual array is a sampled version of a virtual or larger uniform array.

3.1 Steering vectors for multidimensional arrays

Propagation in wireless communications is partly characterized by the transmit and receive antenna steering vectors $\mathbf{t}_M(\vartheta, \phi) \in \mathbb{C}^M$ and $\mathbf{r}_N(\theta, \varphi) \in \mathbb{C}^N$, that respectively depend on the angle of departure (AoD) azimuth and elevation $(\vartheta, \phi) \in \left([0, 2\pi], \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]\right)$ and on the angle of arrival (AoA) azimuth and elevation $(\theta, \varphi) \in \left([0, 2\pi], \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]\right)$ and provide information about the phase response of the array for every propagation angular direction. The structure of the transmit and receive steering vectors depend on the number of antennas M and N and on the normalized positions of the radiating elements, denoted by $\mathbf{q}_m = [q_m^x, q_m^y, q_m^z]^\top$ with $m \in [M]$ for the transmit elements and $\mathbf{v}_n = [v_n^x, v_n^y, v_n^z]^\top$ with $n \in [N]$ for the receive elements in the 3-D space.

Antenna array deployments have at most 3 dimensions, *i.e.* $d \in \{1, 2, 3\}$ being d the number of dimensions. A linear array is the 1-D case where only propagation in the azimuth direction is relevant. The 2-D scenario is a planar array where hemispherical propagation is significant and finally, in the 3-D case, spherical propagation is considered. The coordinate system is arbitrary and thus the axes can be placed properly. Without loss of generality, the coordinate axes are defined such that in 1-D deployments, only the Z axis is considered, in 2-D, Y and Z axes are considered and in 3-D the three X, Y and Z axes are accounted. Thus, the antenna positions for a d-D deployment are

$$\begin{aligned}
\mathbf{q}_m &= \begin{cases} [q_m^z]^\top & \text{if } d_t = 1 \\ [q_m^y, q_m^z]^\top & \text{if } d_t = 2 \\ [q_m^x, q_m^y, q_m^z]^\top & \text{if } d_t = 3 \end{cases} \\
\mathbf{v}_n &= \begin{cases} [v_n^z]^\top & \text{if } d_r = 1 \\ [v_n^y, v_n^z]^\top & \text{if } d_r = 2 \\ [v_n^x, v_n^y, v_n^z]^\top & \text{if } d_r = 3 \end{cases}
\end{aligned} \tag{3.1}$$

where d_t and d_r are the particularized number of dimensions at the transmitter and receiver respectively.

Let $\mathbf{c}(\alpha, \varepsilon)$ be the general propagation vector in a d-D deployment, either at the transmitter or the receiver side with azimuth angle α and elevation angle ε:

$$\mathbf{c}(\alpha, \varepsilon) = \begin{cases} [c_1^z(\alpha, \varepsilon)]^\top = 2\pi \left[\cos\alpha\right]^\top & \text{if } d = 1 \\ [c_1^y(\alpha, \varepsilon), c_2^z(\alpha, \varepsilon)]^\top = 2\pi \left[\sin\alpha \sin\varepsilon,\ \cos\alpha\right]^\top & \text{if } d = 2 \\ [c_1^x(\alpha, \varepsilon), c_2^y(\alpha, \varepsilon), c_3^z(\alpha, \varepsilon)]^\top = 2\pi \left[\sin\alpha \cos\varepsilon,\ \sin\alpha \sin\varepsilon,\ \cos\alpha\right]^\top & \text{if } d = 3 \end{cases} \tag{3.2}$$

where $\alpha \in [0, 2\pi]$ and $\varepsilon \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. Note that a linear or a planar array can also be represented as a 3-D structure but the normalized positions of the m-th antenna would be, for the linear array, $\mathbf{q}_m = [0, 0, q_m^z]^\top$, *i.e.* only has antennas in the Z axis, and for the planar array, $\mathbf{q}_m = [0, q_m^y, q_m^z]^\top$, *i.e.* Y and Z axes.

The corresponding phases seen by the m-th and n-th antennas are given by:

$$\begin{aligned} \Psi_m(\vartheta, \phi) &= \mathbf{q}_m^\top \mathbf{c}(\vartheta, \phi) \\ \Phi_n(\theta, \varphi) &= \mathbf{v}_n^\top \mathbf{c}(\theta, \varphi) \end{aligned} \tag{3.3}$$

Thence, the steering vectors respectively at the transmitter and receiver can be represented as

$$\begin{aligned} \mathbf{t}_M(\vartheta, \phi) &= \left[e^{j\Psi_1(\vartheta,\phi)}, \dots, e^{j\Psi_M(\vartheta,\phi)}\right]^\top \\ \mathbf{r}_N(\theta, \varphi) &= \left[e^{j\Phi_1(\theta,\varphi)}, \dots, e^{j\Phi_N(\theta,\varphi)}\right]^\top \end{aligned} \tag{3.4}$$

The steering vectors in (3.4) can be simplified for uniform arrays scenarios. Furthermore, the steering vector of a general non-uniform array can be equivalently represented by a sampled version of the steering vector of a virtual uniform array via a proper sampling/sensing measuring matrix. This will be shown later in the following sections.

3.2 Steering vector for uniform d-D arrays

Uniform d-D antenna arrays deploy equally spaced radiating elements along all the d-D space. Thus, the number of total antenna elements at each side of the communication link is the product of the number of antennas in each dimension, which we name respectively M_i with $i \in [d_t]$ and N_i with $i \in [d_r]$, *i.e.* $M_u = \prod_{i=1}^{d_t} \mathsf{M}_i$ at the transmitter and $N_u = \prod_{i=1}^{d_r} \mathsf{N}_i$ at the receiver, where d_t and d_r are the number of dimensions at the transmitter and receiver, respectively. The uniform structures can be identified with vectors $\boldsymbol{\mathsf{M}} = [\mathsf{M}_1, \dots, \mathsf{M}_{d_t}]^\top$ and $\boldsymbol{\mathsf{N}} = [\mathsf{N}_1, \dots, \mathsf{N}_{d_r}]^\top$ respectively for both transmitter and receiver. In uniform deployments the antennas take fixed positions based on the uniform spacing, which are given by $\mathbf{q}_m = \left[\mathsf{m}_m^1 \delta^1, \dots, \mathsf{m}_m^{d_t} \delta^{d_t}\right]^\top$ with $\mathsf{m}_m^i \in \{0, 1, \dots \mathsf{M}_i - 1\}$ and δ^i being the antenna spacing in dimension i, with $i \in [d_t]$, at the transmitter side for the m-th antenna. Similarly at the receiver, the position of the n-th antenna is given by $\mathbf{v}_n = \left[\mathsf{n}_n^1 \rho^1, \dots, \mathsf{n}_n^{d_r} \rho^{d_r}\right]^\top$ with $\mathsf{n}_n^i \in \{0, 1, \dots \mathsf{N}_i - 1\}$ and ρ^i being the antenna spacing in dimension i, with $i \in [d_r]$.

Next, consider the normalized frequency vectors $\boldsymbol{\mathsf{g}} \in \mathbb{T}^{d_t}$ and $\boldsymbol{\mathsf{f}} \in \mathbb{T}^{d_r}$ that contain, respectively, the information on the AoD (transmitter) and AoA (receiver) azimuth and elevation. Note that the elements of $\boldsymbol{\mathsf{g}}$ and $\boldsymbol{\mathsf{f}}$ are defined within the unit circle $\mathbb{T}$, so the 1-modulus operator is computed in each of the dimensions as

$$\begin{aligned} \boldsymbol{\mathsf{g}} = [\mathsf{g}^1, \dots, \mathsf{g}^{d_t}]^\top &= \left[\mathrm{mod}\left(\delta_1 c_1(\vartheta, \phi), 1\right), \dots, \mathrm{mod}\left(\delta_{d_t} c_{d_t}(\vartheta, \phi), 1\right)\right]^\top \\ \boldsymbol{\mathsf{f}} = [\mathsf{f}^1, \dots, \mathsf{f}^{d_r}]^\top &= \left[\mathrm{mod}\left(\rho_1 c_1(\theta, \varphi), 1\right), \dots, \mathrm{mod}\left(\rho_{d_r} c_{d_r}(\theta, \varphi), 1\right)\right]^\top \end{aligned} \tag{3.5}$$

where c_i is the i-th element of the generic phase response vector $\mathbf{c}$ defined in (3.2).

The normalized positions of the m-th and n-th antennas are defined as $\boldsymbol{\mathsf{m}}_m = \left[\mathsf{m}_m^1, \ldots, \mathsf{m}_m^{d_t}\right]^\top$ and $\boldsymbol{\mathsf{n}}_n = \left[\mathsf{n}_n^1, \ldots, \mathsf{n}_n^{d_r}\right]^\top$. Then, developing the general steering vectors in (3.4), the steering vectors $\boldsymbol{\mathsf{t}}_{\boldsymbol{\mathsf{M}}}(\boldsymbol{\mathsf{g}})$ and $\mathbf{r}_{\boldsymbol{\mathsf{N}}}(\boldsymbol{\mathsf{f}})$, labelled respectively using the vectors $\boldsymbol{\mathsf{M}}$ and $\boldsymbol{\mathsf{N}}$ that identify the uniform structure, are obtained as

$$\begin{aligned}
\boldsymbol{\mathsf{t}}_{\boldsymbol{\mathsf{M}}}(\boldsymbol{\mathsf{g}}) &= \frac{1}{\sqrt{M_u}}\left[e^{j2\pi\boldsymbol{\mathsf{g}}^\top\boldsymbol{\mathsf{m}}_1}, e^{j2\pi\boldsymbol{\mathsf{g}}^\top\boldsymbol{\mathsf{m}}_2}, \ldots, e^{j2\pi\boldsymbol{\mathsf{g}}^\top\boldsymbol{\mathsf{m}}_M}\right]^\top \\
&= \otimes_{i=1}^{d_t}\boldsymbol{\mathsf{t}}_{\mathsf{M}_i}(\mathsf{g}^i) \\
\mathbf{r}_{\boldsymbol{\mathsf{N}}}(\boldsymbol{\mathsf{f}}) &= \frac{1}{\sqrt{N_u}}\left[e^{j2\pi\boldsymbol{\mathsf{f}}^\top\boldsymbol{\mathsf{n}}_1}, e^{j2\pi\boldsymbol{\mathsf{f}}^\top\boldsymbol{\mathsf{n}}_2}, \ldots, e^{j2\pi\boldsymbol{\mathsf{f}}^\top\boldsymbol{\mathsf{n}}_N}\right]^\top \\
&= \otimes_{i=1}^{d_r}\mathbf{r}_{\mathsf{N}_i}(\mathsf{f}^i)
\end{aligned} \tag{3.6}$$

where, for the i-th dimension, $\boldsymbol{\mathsf{t}}_{\mathsf{M}_i}(\mathsf{g}^i) = \frac{1}{\sqrt{\mathsf{M}_i}}\left[1, e^{j2\pi\mathsf{g}^i}, e^{j2\pi 2\mathsf{g}^i}, \ldots, e^{j2\pi(\mathsf{M}_i-1)\mathsf{g}^i}\right]^\top$ at the transmitter side, and $\mathbf{r}_{\mathsf{N}_i}(\mathsf{f}^i) = \frac{1}{\sqrt{\mathsf{N}_i}}\left[1, e^{j2\pi\mathsf{f}^i}, e^{j2\pi 2\mathsf{f}^i}, \ldots, e^{j2\pi(\mathsf{N}_i-1)\mathsf{f}^i}\right]^\top$ at the receiver.

3.3 Sparse MIMO channel model with non-uniform d-D antenna deployment

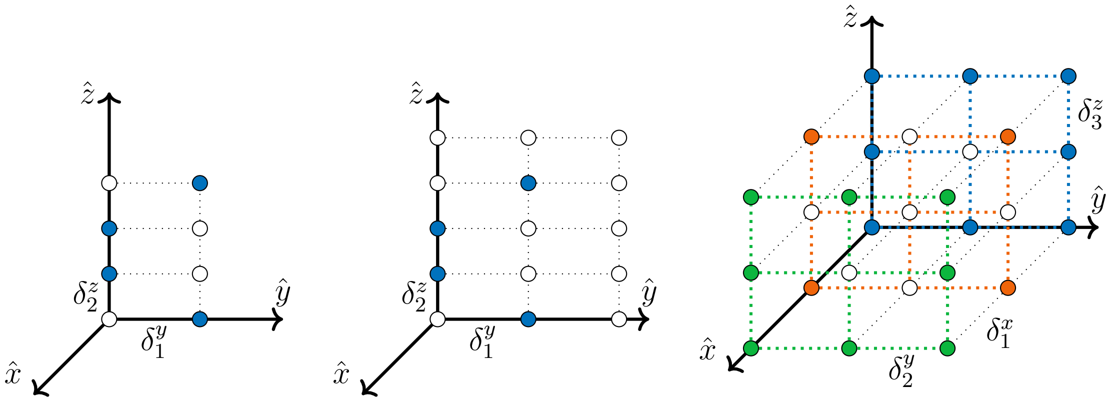
As mentioned above, we focus on non-uniform antenna deployments that can be represented by a sampled version of a so called virtual uniform array via a proper sampling/sensing measuring matrix. Thus, a non-uniform deployment is a more general scenario than a uniform array.

Let the virtual uniform array at the transmitter be characterized by the steering vector $\boldsymbol{\mathsf{t}}_{\boldsymbol{\mathsf{M}}}(\boldsymbol{\mathsf{g}}) \in \mathbb{C}^{M_u}$, where $\boldsymbol{\mathsf{M}}$ is the vector containing the number of antennas in each of the dimensions and M_u is the total number of transmit antennas in the virtual uniform array. Equivalently, let $\mathbf{r}_{\boldsymbol{\mathsf{N}}}(\boldsymbol{\mathsf{f}}) \in \mathbb{C}^{N_u}$ be the steering vector of the virtual uniform array at the receiver, where $\boldsymbol{\mathsf{N}}$ is the vector containing the number of antennas in each of the dimensions and N_u is the total number of receive antennas in the virtual uniform array. Thus, the sampled non-uniform antenna deployment with M transmit antennas and N receive antennas is characterized by the following two steering vectors, respectively for the transmitter and receiver:

$$\begin{aligned}
\mathbf{t}_M(\boldsymbol{\mathsf{f}}) &= \mathbf{B}_t\boldsymbol{\mathsf{t}}_{\boldsymbol{\mathsf{M}}}(\boldsymbol{\mathsf{g}}) \\
\mathbf{r}_N(\boldsymbol{\mathsf{f}}) &= \mathbf{B}_r\mathbf{r}_{\boldsymbol{\mathsf{N}}}(\boldsymbol{\mathsf{f}})
\end{aligned} \tag{3.7}$$

where $\mathbf{B}_t \in \{0,1\}^{M\times M_u}$ and $\mathbf{B}_r \in \{0,1\}^{N\times N_u}$ are binary fat matrices, *i.e.* $M \leq M_u$ and $N \leq N_u$, whose element b_{ij} equals 1 if the uniform array element $j \in [M_u]$ (or $j \in [N_u]$ for the receiver) is included as the sampled array element $i \in [M]$ (or $i \in [N]$ for the receiver), and $b_{ij} = 0$ otherwise. Thus, sensing matrices $\mathbf{B}_t \in \{0,1\}^{M\times M_u}$ and

$\mathbf{B}_r \in \{0,1\}^{N \times N_u}$ must effectively remove antenna elements from the virtual uniform array and must ensure that the elements from the uniform array are mapped only once into the sampled array. The sensing or sampling matrices must then have at most one 1 per column and row, having all-zero columns in the positions of the elements that are not included in the sampled array. This yields some restrictions to the sensing matrices: $\mathbf{B}_t = \left[\mathbf{I}_M, \mathbf{0}_{M\times(M_u-M)}\right]\mathbf{\Pi}$, with $\mathbf{\Pi}$ being a permutation matrix, all being equivalent at the receiver. Note that this definition is also valid for uniform structures where the sampling matrices are identity matrices, *i.e.* $\mathbf{B}_t = \mathbf{I}_M$ for the transmitter array and $\mathbf{B}_r = \mathbf{I}_N$ for the receiver deployment, given that all antennas from the virtual uniform array are included into the real array.



(a) (b) (c)

Fig. 3.1. Examples of non-uniform deployments sampled from virtual uniform arrays at the transmitter side. (a) A 2-D deployment with $M = 4$ antennas sampled from a virtual uniform array of $\mathbf{M} = [2,4]$. (b) A 2-D deployment with $M = 4$ antennas sampled from a virtual uniform array of $\mathbf{M} = [3,4]$. (c) A 3-D deployment with $M = 20$ antennas sampled from a virtual uniform array of $\mathbf{M} = [3,3,3]$. Filled antennas are sampled and blank antennas are filtered out from the virtual uniform array.

In Fig. 3.1 there are some examples of non-uniform deployments sampled from a virtual uniform array where sampled antennas are filled whereas removed or filtered antennas are blank. The examples are at the transmitter side, but examples of receive deployments would be equivalent. Note how different uniform arrays can be sampled accordingly to obtain the same sampled non-uniform deployment as it is the case for Fig. 3.1(a) and Fig. 3.1(b). In Fig. 3.1(a), a virtual uniform 2-D structure with $\mathbf{M} = [2,4]$ is sampled to obtained non-uniform deployment with

$M = 4$ antennas. The sampling procedure is the following:

$$\mathbf{t}_M(\mathsf{g}) = \mathbf{B}_t \mathbf{t}_{\mathsf{M}}(\mathsf{g})$$

$$\underbrace{\begin{bmatrix} e^{j2\pi \mathsf{f}^z} \\ e^{j2\pi 2\mathsf{f}^z} \\ e^{j2\pi \mathsf{f}^y} \\ e^{j2\pi(\mathsf{f}^y+3\mathsf{f}^z)} \end{bmatrix}}_{\mathbf{t}_M(\mathsf{g})} = \underbrace{\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{B}_t} \underbrace{\begin{bmatrix} 1 \\ e^{j2\pi \mathsf{f}^z} \\ e^{j2\pi 2\mathsf{f}^z} \\ e^{j2\pi 3\mathsf{f}^z} \\ e^{j2\pi \mathsf{f}^y} \\ e^{j2\pi(\mathsf{f}^y+\mathsf{f}^z)} \\ e^{j2\pi(\mathsf{f}^y+2\mathsf{f}^z)} \\ e^{j2\pi(\mathsf{f}^y+3\mathsf{f}^z)} \end{bmatrix}}_{\mathbf{t}_{\mathsf{M}}(\mathsf{g})} \tag{3.8}$$

For the applications studied in this thesis a narrowband stationary mmWave propagation scenario with M antennas at the transmitter and N antennas at the receiver is considered [42]. Letting d_t and d_r be the number of dimensions respectively at transmitter and receiver, where both $d_t \leq 3$ and $d_r \leq 3$, the channel matrix $\mathbf{H} \in \mathbb{C}^{N \times M}$ is then modeled as

$$\begin{aligned} \mathbf{H} &= \sum_{k=1}^{K} \gamma_k \mathbf{r}_N(\mathsf{f}_k) \mathbf{t}_M^\dagger(\mathsf{g}_k) \\ &= \mathbf{R}_N(\mathsf{f}_{1:K}) \mathbf{G} \mathbf{T}_M^\dagger(\mathsf{g}_{1:K}) \\ &= \mathbf{B}_r \mathbf{R}_{\mathsf{N}}(\mathsf{f}_{1:K}) \mathbf{G} \mathbf{T}_{\mathsf{M}}^\dagger(\mathsf{g}_{1:K}) \mathbf{B}_t^\dagger \\ &= \mathbf{B}_r \mathbf{H}_u \mathbf{B}_t^\dagger \end{aligned} \tag{3.9}$$

where K is the number of propagation paths, $\mathsf{g}_k \in \mathbb{T}^{d_t}$ and $\mathsf{f}_k \in \mathbb{T}^{d_r}$ are the normalized frequency vectors that respectively contain the information on the AoD and AoA of the k-th propagation path, $\gamma_k \in \mathbb{C}$ is the gain of the k-th multipath, $\mathbf{G} = \mathrm{diag}([\gamma_1, \ldots, \gamma_K]^\top) = \mathrm{diag}(\boldsymbol{\gamma}) \in \mathbb{C}^{K \times K}$, $\mathbf{T}_M(\mathsf{g}_{1:K}) = [\mathbf{t}_M(\mathsf{g}_1), \ldots, \mathbf{t}_M(\mathsf{g}_K)] \in \mathbb{C}^{M \times K}$ and $\mathbf{R}_N(\mathsf{f}_{1:K}) = [\mathbf{r}_N(\mathsf{f}_1), \ldots, \mathbf{r}_N(\mathsf{f}_K)] \in \mathbb{C}^{N \times K}$ are the matrices containing the steering vectors of the sampled array at the transmitter and receiver, respectively. Similarly, the steering vectors of the virtual uniform array at the transmitter and receiver are $\mathbf{T}_{\mathsf{M}}(\mathsf{g}_{1:K}) = [\mathbf{t}_{\mathsf{M}}(\mathsf{g}_1), \ldots, \mathbf{t}_{\mathsf{M}}(\mathsf{g}_K)] \in \mathbb{C}^{M_u \times K}$ and $\mathbf{R}_{\mathsf{N}}(\mathsf{f}_{1:K}) = [\mathbf{r}_{\mathsf{N}}(\mathsf{f}_1), \ldots, \mathbf{r}_{\mathsf{N}}(\mathsf{f}_K)] \in \mathbb{C}^{N_u \times K}$, and $\mathbf{B}_t \in \{0,1\}^{M \times M_u}$ and $\mathbf{B}_r \in \{0,1\}^{N \times N_u}$ are the sensing matrices that lead to non-uniform array antenna deployments.

Channel matrix $\mathbf{H}$ can be vectorized as

$$\begin{aligned}
\mathbf{h} &= \operatorname{vec}(\mathbf{H}) \\
&= (\mathbf{B}_t \otimes \mathbf{B}_r) \operatorname{vec}(\mathbf{H}_u) \\
&= (\mathbf{B}_t \otimes \mathbf{B}_r) \sum_{k=1}^{K} \gamma_k \mathbf{t}^*_{\boldsymbol{\mathsf{M}}}(\boldsymbol{\mathsf{g}}_k) \otimes \mathbf{r}_{\boldsymbol{\mathsf{N}}}(\boldsymbol{\mathsf{f}}_k) \\
&= (\mathbf{B}_t \otimes \mathbf{B}_r) \sum_{k=1}^{K} \gamma_k \mathbf{u}_{\boldsymbol{\mathsf{L}}}(\boldsymbol{\mathsf{l}}_k) \\
&= (\mathbf{B}_t \otimes \mathbf{B}_r) \mathbf{U}_{\boldsymbol{\mathsf{L}}}(\boldsymbol{\mathsf{l}}_{1:K}) \boldsymbol{\gamma} \\
&= (\mathbf{B}_t \otimes \mathbf{B}_r) \mathbf{h}_u
\end{aligned} \tag{3.10}$$

where $d_l = d_t + d_r$, $\boldsymbol{\mathsf{L}} = [\boldsymbol{\mathsf{M}}^\top, \boldsymbol{\mathsf{N}}^\top]^\top = [\mathsf{M}_1, \dots, \mathsf{M}_{d_t}, \mathsf{N}_1, \dots, \mathsf{N}_{d_r}]^\top = [\mathsf{L}_1, \dots, \mathsf{L}_{d_l}]^\top$, $\boldsymbol{\mathsf{l}}_k = \left[-\boldsymbol{\mathsf{g}}_k^\top, \boldsymbol{\mathsf{f}}_k^\top\right]^\top \in \mathbb{T}^{d_l}$, and considering $L_u = M_u N_u$, then $\mathbf{h}_u = \sum_{k=1}^{K} \gamma_k \mathbf{u}_{\boldsymbol{\mathsf{L}}}(\boldsymbol{\mathsf{l}}_k) \in \mathbb{C}^{L_u}$.

3.4 Measurement scenarios

The different challenges that arise in each of the applications studied in this thesis differ mainly on both the information that is accessible and the information that has to be recovered. Every application shares the definition of the MIMO mmWave channel posed in (3.9) where the sampling matrices $\mathbf{B}_t \in \{0,1\}^{M \times M_u}$ and $\mathbf{B}_r \in \{0,1\}^{N \times N_u}$ are used to sample a uniform structure into a non-uniform antenna deployment as depicted in Fig. 3.1.

3.4.1 Single Measurement Vector

Let a signal vector $\mathbf{s} \in \mathbb{C}^M$ be transmitted over a channel $\mathbf{H} \in \mathbb{C}^{N \times M}$ with M antennas at the transmitter and N antennas at the receiver. In this case, only one vector $\mathbf{s} \in \mathbb{C}^M$ is transmitted and thus, only one observation is received and processed, which yields to a Single Measurement Vector (SMV) scenario where the received measurement vector is given by

$$\begin{aligned}
\mathbf{y} &= \mathbf{H}\mathbf{s} + \mathbf{w} \\
&= \mathbf{B}_r \mathbf{R}_{\boldsymbol{\mathsf{N}}}(\boldsymbol{\mathsf{f}}_{1:K}) \mathbf{G} \mathbf{T}^\dagger_{\boldsymbol{\mathsf{M}}}(\boldsymbol{\mathsf{g}}_{1:K}) \mathbf{B}_t^\dagger \mathbf{s} + \mathbf{w}
\end{aligned} \tag{3.11}$$

where $\mathbf{w} \in \mathbb{C}^N$ is additive white circular-symmetric complex Gaussian noise with independent zero-mean components and σ_w^2 variance.

3.4.2 Multiple Measurement Vector

Next, consider the approach of Multiple Measurement Vector (MMV) where a set of P observations is received $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_P] \in \mathbb{C}^{N \times P}$ and processed jointly after a signal sequence $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_P] \in \mathbb{C}^{M \times P}$ is transmitted.

$$\begin{aligned} \mathbf{Y} &= \mathbf{HS} + \mathbf{W} \\ &= \mathbf{B}_r \mathbf{R}_{\boldsymbol{\mathsf{N}}}(\boldsymbol{\mathsf{f}}_{1:K}) \mathbf{G} \mathbf{T}^{\dagger}_{\boldsymbol{\mathsf{M}}}(\boldsymbol{\mathsf{g}}_{1:K}) \mathbf{B}_t^{\dagger} \mathbf{S} + \mathbf{W} \end{aligned} \tag{3.12}$$

where $\mathbf{W} \in \mathbb{C}^{N \times P}$ is additive white circular-symmetric complex Gaussian noise with independent zero-mean components and σ_w^2 variance.

From the vectorized version of the channel in (3.10), the received MMV signal can be vectorized as

$$\begin{aligned} \mathbf{y} &= \operatorname{vec}(\mathbf{Y}) \\ &= \left(\mathbf{S}^{\top} \otimes \mathbf{I}_N\right) \left(\mathbf{B}_t \otimes \mathbf{B}_r\right) \mathbf{U}_{\boldsymbol{\mathsf{L}}}(\boldsymbol{\mathsf{l}}_{1:K}) \boldsymbol{\gamma} + \mathbf{w} \\ &= \left(\mathbf{S}^{\top} \otimes \mathbf{I}_N\right) \left(\mathbf{B}_t \otimes \mathbf{B}_r\right) \mathbf{h}_u + \mathbf{w} \end{aligned} \tag{3.13}$$

where $\mathbf{w} = \operatorname{vec}(\mathbf{W}) \in \mathbb{C}^{PN}$ and $\mathbf{h}_u \in \mathbb{C}^{L_u}$ is the vectorized parametric channel.

3.5 Applications

This section will cover the challenges that arise in the studied applications. From SMV and MMV formulas in (3.11), (3.12) and (3.13), the specific problems for each of the applications will be derived.

The steering vector formulation for a uniform d-D structure can be generalized from that on (3.6). Note that if the array is uniform, the steering vector, either at the transmitter or at the receiver if fully characterized by only two parameters: the vector containing the number of antennas in each dimension and the frequency parameters. Thus, at transmitter side, $\mathbf{t}_{\boldsymbol{\mathsf{M}}}(\boldsymbol{\mathsf{g}}_k)$ is fully characterized by the vector $\boldsymbol{\mathsf{M}}$ and the frequency parameters $\boldsymbol{\mathsf{g}}_k$, with $k \in [K]$ and, in the same way, at receiver side, $\mathbf{r}_{\boldsymbol{\mathsf{N}}}(\boldsymbol{\mathsf{f}}_k)$ is fully characterized by vector $\boldsymbol{\mathsf{N}}$ and the frequency parameters $\boldsymbol{\mathsf{f}}_k$, with $k \in [K]$. This allows to redefine a generic steering vector which is valid for both ends of the communication:

$$\mathbf{a}_{\boldsymbol{\mathsf{J}}}(\boldsymbol{\mathsf{f}}_k) = \frac{1}{\sqrt{J}} \begin{bmatrix} e^{j2\pi \boldsymbol{\mathsf{j}}_1^{\top} \boldsymbol{\mathsf{f}}_k} \\ \vdots \\ e^{j2\pi \boldsymbol{\mathsf{j}}_J^{\top} \boldsymbol{\mathsf{f}}_k} \end{bmatrix} \tag{3.14}$$

where $\boldsymbol{\mathsf{J}} = [\mathsf{J}_1, \ldots, \mathsf{J}_d]^{\top}$ is a generic vector containing the number of antennas in each of the d dimensions and $J = \prod_{i=1}^{d} \mathsf{J}_i$. The generic vector $\boldsymbol{\mathsf{J}}$ can be particularized as $\boldsymbol{\mathsf{M}}$ for the transmitter and as $\boldsymbol{\mathsf{N}}$ for the receiver. Also, consider the matrix $\mathbf{A}_{\boldsymbol{\mathsf{J}}}(\boldsymbol{\mathsf{f}}_{1:K}) = [\mathbf{a}_{\boldsymbol{\mathsf{J}}}(\boldsymbol{\mathsf{f}}_1), \ldots, \mathbf{a}_{\boldsymbol{\mathsf{J}}}(\boldsymbol{\mathsf{f}}_K)]$ as the concatenation of all K steering vectors.

3.5.1 Applications using SMV

Application to automotive pulse radar in mmWave (see chapter 5) and to low complexity hybrid precoder design in mmWave (see chapter 6) when no complexity constrain is imposed can be considered as a SMV scenario, where we basically have an observation as

$$\mathbf{y} = \mathbf{B}\mathbf{A}_{\boldsymbol{\mathsf{J}}}(\boldsymbol{\mathsf{f}}_{1:K})\mathbf{u} + \mathbf{w} \tag{3.15}$$

where the different parameters $\mathbf{B}$, $\boldsymbol{\mathsf{J}}$, $\boldsymbol{\mathsf{f}}$ and $\mathbf{u}$ will be particularized for each applications, and $\mathbf{w}$ is Gaussian noise with independent zero-mean and σ_w^2 variance.

Automotive pulse radar

In the case of pulse radar we want to extract relevant information at the receiver from an unknown signal $\mathbf{s} \in \mathbb{C}^M$ which is transmitted over the parametric channel $\mathbf{H} \in \mathbb{C}^{N \times M}$. This yields a received measurement signal $\mathbf{y} \in \mathbb{C}^N$ as in (3.11). Processing the observed vector $\mathbf{y} \in \mathbb{C}^N$, the objective is to extract the unknown parameters of the channel $\{\boldsymbol{\mathsf{f}}_1, \ldots, \boldsymbol{\mathsf{f}}_K\}$, which contain information about the AoA azimuth and elevation angles (θ, φ) of the received signal, and also delay and Doppler shift which allow an estimate of distance and velocity.

Since only the reception parameters are to be recovered, from (3.11) the observed vector $\mathbf{y} \in \mathbb{C}^N$ can be reformulated as:

$$\begin{aligned}
\mathbf{y} &= \mathbf{B}_r\mathbf{R}_{\boldsymbol{\mathsf{N}}}(\boldsymbol{\mathsf{f}}_{1:K})\underbrace{\mathbf{G}\mathbf{T}_{\boldsymbol{\mathsf{M}}}^{\dagger}(\boldsymbol{\mathsf{g}}_{1:K})\mathbf{B}_t^{\dagger}\mathbf{s}}_{\mathbf{u}} + \mathbf{w} \\
&= \mathbf{B}_r\mathbf{R}_{\boldsymbol{\mathsf{N}}}(\boldsymbol{\mathsf{f}}_{1:K})\mathbf{u} + \mathbf{w} \\
&= \mathbf{B}_r\mathbf{A}_{\boldsymbol{\mathsf{N}}}(\boldsymbol{\mathsf{f}}_{1:K})\mathbf{u} + \mathbf{w}
\end{aligned} \tag{3.16}$$

where the unknown information on the transmission parameters $\{\boldsymbol{\mathsf{g}}_1, \ldots, \boldsymbol{\mathsf{g}}_K\}$, the gain matrix $\mathbf{G} \in \mathbb{C}^{K \times K}$ and transmitted signal $\mathbf{s} \in \mathbb{C}^M$ is gathered into the unknown vector $\mathbf{u} \in \mathbb{C}^K$. Thus, the known information in (3.16) is the shape of the array, which includes the knowledge of vector $\boldsymbol{\mathsf{N}}$ and matrix $\mathbf{B}_r \in \{0,1\}^{N \times N_u}$, the observable received vector $\mathbf{y} \in \mathbb{C}^N$ and the structure of the steering vectors of the virtual uniform array $\mathbf{A}_{\boldsymbol{\mathsf{N}}}(\boldsymbol{\mathsf{f}}_{1:K}) \in \mathbb{C}^{N_u \times K}$. Conversely, the unkowns are the vector $\mathbf{u} \in \mathbb{C}^K$ containing the information on the gains and transmission, the noise $\mathbf{w} \in \mathbb{C}^N$ and the parameters/frequencies of the channel $\{\boldsymbol{\mathsf{f}}_1, \ldots, \boldsymbol{\mathsf{f}}_K\} \in \mathbb{T}^{d_r}$ which contains information on the AoA. distance and velocity and are the input of the steering vector matrix.

By solving (P.5) or (P.6) the d_r-LT matrix $\mathbf{T}_{\boldsymbol{\mathsf{N}}}$ is obtained. From that, Algorithm 1 is applied on $\mathbf{T}_{\boldsymbol{\mathsf{N}}}$ to recover the channel parameters $\{\boldsymbol{\mathsf{f}}_1, \ldots, \boldsymbol{\mathsf{f}}_K\}$ which contain the relevant information about the received signal.

Low complexity hybrid precoder design

For the hybrid precoder design, we explore the structure of an optimal precoder unveiling a measurement of a linear combination of atoms, due to the nature of the mmWave channel in (3.9). This allows us to design a hybrid precoder following a set of SMV where the observations are the columns of the optimal precoder, which we assume accessible, and the objective is to retrieve the parameters that characterize the atoms so that a hybrid precoder can be built that approximates as much as possible to the optimal.

3.5.2 Applications using MMV

For the applications mentioned above we also study MMV scenarios that arise when we impose a complexity constrain. For example, in the radar application, we may want to divide the samples in one of the dimensions into several groups generating a MMV with which the optimization proposed in (P.6) is adapted in chapter 5 to solve for all the measurements at once. Similarly, for the hybrid precoder application we may want to find the parameters that characterize the atoms jointly for all the columns of the precoder at once. In this case, we have a MMV, *i.e.* one measurement vector for each column of the precoder, and also (P.6) is adapted in chapter 6 for the multiple measurement case.

Full Channel estimation using atomic norm

The last application of the thesis also follows a MMV scenario. Here, a known sequence of pilots $\mathbf{P} \in \mathbb{C}^{M \times P}$ is transmitted over the channel and a set of observations is received as $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_P] \in \mathbb{C}^{N \times P}$. Using the formulation in (3.13) of the vectorized received signal we have that

$$\begin{aligned}
\mathbf{y} &= \operatorname{vec}(\mathbf{Y}) \\
&= \left(\mathbf{P}^\top \otimes \mathbf{I}_N\right) (\mathbf{B}_t \otimes \mathbf{B}_r)\, \mathbf{U}_{\mathsf{L}}(\mathsf{l}_{1:K})\boldsymbol{\gamma} + \mathbf{w} \\
&= \mathbf{Q}\mathbf{U}_{\mathsf{L}}(\mathsf{l}_{1:K})\boldsymbol{\gamma} + \mathbf{w} \\
&= \mathbf{Q}\mathbf{A}_{\mathsf{L}}(\mathsf{l}_{1:K})\boldsymbol{\gamma} + \mathbf{w} \\
&= \mathbf{Q}\mathbf{h}_u + \mathbf{w}
\end{aligned} \tag{3.17}$$

where $\mathbf{Q} = \left(\mathbf{P}^\top \otimes \mathbf{I}_N\right)(\mathbf{B}_t \otimes \mathbf{B}_r) \in \mathbb{C}^{NP \times M_u N_u}$ is a known matrix containing the information on the transmitted pilots and on the non-uniform array structures. By solving (P.5) or (P.6) and then applying Algorithm 1 on the obtained d_r-LT matrix $\mathbf{T}_{\mathsf{L}}$, the channel parameters $\{\mathsf{l}_1, \ldots, \mathsf{l}_K\}$ which contain the information on the AoD and AoA can be recovered alongside with the channel fading coefficients $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_K]^\top$ which allows full channel estimation.

Chapter 4

Machine Learning approach for estimation of number of composite atoms

Ideally, in the noiseless case, the rank of the data signal covariance matrix $\mathbf{T_J} \in \mathbb{C}^{J \times J}$, which results from the optimization problems (P.3) and (P.4), should be equal either to the number of composite atoms K if $K < J$, or equal to J in the other case. Furthermore, given the conditions of Theorems 1 and 2 we seek that the matrix is rank deficient having $\text{rank}\{\mathbf{T_J}\} = r = K < J$. With that we would ensure unique signal and frequency recovery. However, there are several scenarios where this rank condition would not hold:

- In the noiseless case, for the convex relaxation approach (P.4) and depending on the frequency realization: indeed, depending on the frequency realization (rarely), solving (P.4) provide a solution such that the number of relevant eigenvalues of $\mathbf{T_J}$ are larger than K and therefore $\text{rank}\{\mathbf{T_J}\} > K$. This would be an scenario where unique recovery of the frequencies and signal cannot be assured, and that is the reason why a rank condition is explicitly added in Th. 2. If all conditions hold in Th. 2, we can assure that $\text{rank}\{\mathbf{T_J}\} = r = K < J$ [24].

- In the noiseless case, both for (P.3), (P.4): numerical inaccuracies of the solver used (in our case in Matlab) lead to $\text{rank}\{\mathbf{T_J}\} = J$. However, from observation, there are clearly K eigenvalues different from zero, and the rest are extremely close to zero.

- The presence of noise provide $\text{rank}\{\mathbf{T_J}\} = J$, despite there maybe a number close to K of relevant eigenvalues from the matrix decomposition.

Note that in any of these scenarios, a problem arises in the sense that $\text{rank}\{\mathbf{T_J}\}$, along with the matrix $\mathbf{T_J}$ itself, are the input to Algorithm 1 and will tell the algorithm how many frequencies f_k, $k \in [\text{rank}\{\mathbf{T}_{\mathsf{J}}\}]$ must recover. Therefore, it is of high interest to have the best estimation of K taking only into account the most relevant eigenvectors from the decomposition of $\mathbf{T}_{\mathsf{J}}$.

The rank of a matrix can be defined as the number of non-zero eigenvalues with the i-th eigenvalue denoted as $\lambda_i \in \mathbb{R}$, $i \in [J]$. Thus, although the number of non-zero eigenvalues is expected to be equal to K, in the aforementioned scenarios we may find that matrix $\mathbf{T}_{\mathsf{J}}$ is full-rank, *i.e.* $\text{rank}\{\mathbf{T}_{\mathsf{J}}\} = J$, which will make impossible for Algorithm 1 to decompose the exact number of desired parameter sets. Fortunately, in those scenarios where the first K eigenvalues $\{\lambda_1, \ldots, \lambda_K\}$ are dominant, *i.e.* much larger than the rest, and eigenvalues $\{\lambda_{K+1}, \ldots, \lambda_J\}$ are close to 0, we can consider the latest eigenvalues as spurious and therefore discard them.

In this chapter, a Machine Learning (ML) approach is proposed to identify the relevant eigenvalues. Basically, the problem consists in estimating the number of dominant eigenvalues from the set of eigenvalues by training a model that learns when an eigenvalue is dominant and when an eigenvalue can be considered as spurious. Thus, the output of that model, $\hat{K}$, is an estimate of the number of composite atoms.

4.1 Problem setup and data gathering

For this problem, we will consider a Single Measurement Vector (SMV) as in (3.16) where we receive a number of echoes in a d-D antenna deployment with J antennas and J_x, J_y and J_z antennas in the abscissa, ordinate and applicate axes respectively which will be vectorized as $\mathsf{J} = [\mathsf{J}_x, \mathsf{J}_y, \mathsf{J}_z]^\top = [\mathsf{J}_1, \ldots, \mathsf{J}_d]^\top$. Without loss of generality, we assume a canonical ordering as explained in chapter 2 such that $\mathsf{J}_1 \leq \cdots \leq \mathsf{J}_d$.

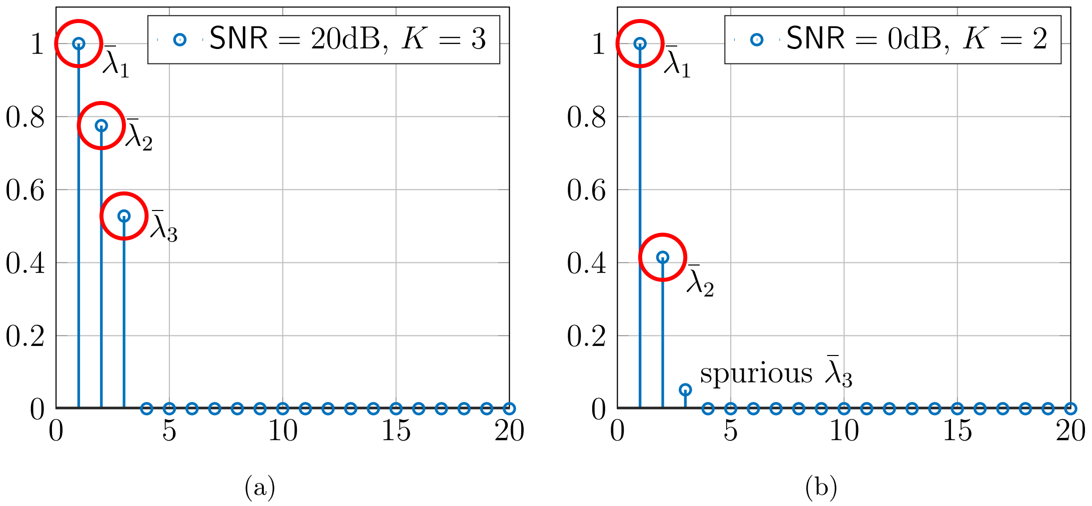
We assume a uniform antenna deployment with which the received signal can be formulated as

$$\mathbf{y} = \mathbf{A}_{\mathsf{J}}(\mathsf{f}_{1:K})\mathbf{u} + \mathbf{w}, \tag{4.1}$$

where the unknowns are the number of echoes K, the frequency set $\mathsf{f}_{1:K}$, the signal $\mathbf{u} \in \mathbb{C}^K$ and, of course, the noise $\mathbf{w} \in \mathbb{C}^J$.

From this measurement, the d-LT matrix $\mathbf{T}_{\mathsf{J}} \in \mathbb{C}^{J \times J}$ is obtained via the convex optimization problem in (P.6), and we assume that conditions in Theorem 2 apply. Also, due to the canonical ordering imposed in vector J, this matrix will have a Kronecker structure such that $\mathsf{J}_d \to \mathsf{J}_{d-1} \to \cdots \to \mathsf{J}_1$.

Thence, matrix $\mathbf{T}_{\mathsf{J}} \in \mathbb{C}^{J \times J}$ will be decomposed to extract its eigenvalues. In favorable scenarios with sufficient SNR, there will be K dominant eigenvalues and the rest being very close to 0. Conversely, for non favorable scenarios with very limited SNR the spurious eigenvalues can be very close to the dominant ones making almost impossible to distinguish them.



(a)

(b)

Fig. 4.1. Examples of the first 15 ordered and normalized eigenvalues of $\mathbf{T}_{\mathsf{J}}$ with $\mathsf{J} = [3,5,6]^\top$ for (a) a favorable scenario with $K = 3$ and $\mathsf{SNR} = 20\text{dB}$ and (b) a non favorable scenario with $K = 2$ and $\mathsf{SNR} = 0\text{dB}$.

In Fig. 4.1, examples of the 15 first normalized eigenvalues of $\mathbf{T}_{\mathsf{J}}$, with $\mathsf{J} = [3,5,6]^\top$, are depicted for two very different scenarios. Normalized eigenvalues are denoted as $\bar{\lambda}_i$ with $i \in [J]$, and they standardize the dataset forcing the first normalized eigenvalue to be always $\bar{\lambda}_1 = 1$. Fig. 4.1(a) is a favorable scenario with $K = 3$ and $\mathsf{SNR} = 20\text{dB}$ whereas Fig. 4.1(b) is a non favorable scenario with $K = 2$ and $\mathsf{SNR} = 0\text{dB}$. Note how in Fig. 4.1(a) we can clearly identify 3 dominant eigenvalues corresponding to the actual value of $K = 3$, however, in Fig. 4.1(b), we see two clear dominant eigenvalues but we also see a third one which is greater than 0. At simple sight, one can identify the dominant $\bar{\lambda}_1$ and $\bar{\lambda}_2$ in Fig. 4.1(b) and discard the spurious $\bar{\lambda}_3$, but doing it automatically is of high interest and require a machine that learns from a set of examples. It is important to note that in both scenarios, $\bar{\lambda}_{4:J}$ are close but not exactly 0, causing matrix $\mathbf{T}_{\mathsf{J}}$ to be full-rank. Therefore, the objective is to successfully identify the number of dominant eigenvalues, $\hat{K}$, given a complete set of eigenvalues of $\mathbf{T}_{\mathsf{J}}$. Thence, with eigenvalues $\bar{\lambda}_{1:\hat{K}}$, matrix $\mathbf{T}_{\mathsf{J}}$ is truncated accordingly before plugging it into Algorithm 1 along with the estimate $\hat{K}$ itself.

4.2 Data preparation

The first step is to gather sufficient amount of data to train a Machine Learning model. For that we generate synthetic data by running a set of simulations for a 3-D scenario with uniform deployment at the receiver with $J = 90$ receive antennas and $\mathsf{J} = [\mathsf{J}_x, \mathsf{J}_y, \mathsf{J}_z]^\top = [\mathsf{J}_1, \mathsf{J}_2, \mathsf{J}_3]^\top = [3,5,6]^\top$, *i.e.* $\mathsf{J}_x = 3$, $\mathsf{J}_y = 5$ and $\mathsf{J}_z = 6$, as the

Variable	**Set of values**
K	$\{1, 2, 3, 4, 5\}$
SNR (dB)	$\{-10, -3, -2, -1, 0, 1, 2, 3, 5, 10, 15, 20, 25, 30, 50, 75, 100, 500\}$
$\mathbf{f}_{1:K}$ and $\mathbf{u}$ samples	100 different samples
Noise $\mathbf{w}$ samples	100 different samples

TABLE 4.1. SCENARIO PARAMETERS FOR DATA GATHERING.

one used in Fig. 4.1. The simulations run different scenarios with the parameters described in Table 4.1. Therefore, for each SNR value we have 50000 examples which yields a dataset with a total of 900000 examples. The model should be able to discern the dominant $\lambda_{1:\hat{K}}$ in a varied set of conditions which are covered with these simulation parameters. Note that the array structure at the receiver, J, is known and is also assumed, without loss of generality, to possess a canonical ordering. With this information we know that the receiver can only decompose at max 5 propagation paths, given that $\max_i\{\mathsf{J}_i\} = 6$. Thus, the estimated number of echoes $\hat{K}$ is bounded in the set $\{1, 2, 3, 4, 5\}$ which is the reason why only $K \in [5]$ are simulated.

Also, as mentioned previously, the set of eigenvalues is normalized so that $\bar{\lambda}_1 = 1$. This prevents having undesired variance among different conditions and standardizes the sets of eigenvalues helping the models learn the patterns.

Once the data is gathered we have sufficient information to train a supervised model. The input data for the model is denoted as $\mathbf{X}_{\text{in}}$ and the output or labeled data, $\mathbf{y}_{\text{out}}$:

- $\mathbf{X}_{\text{in}} \in \mathbb{R}^{900000 \times 90}$ are the examples of the set of $J = 90$ eigenvalues of matrix $\mathbf{T}_{\mathsf{J}}$ for each of the simulation runs.
- $\mathbf{y}_{\text{out}} \in [5]^{900000}$ is the ground truth for the model, *i.e.* the true K for each of the simulation runs.

Since the structure of the antenna array is known, it is important to note that the model can be trained offline with this data. Then, predictions can be made later in an online manner with an already prepared model for the specific antenna deployment.

This problem can be modeled as a regression scenario where we try to approximate a mapping function (f) that maps input data ($\mathbf{X}_{\text{in}}$) to output data ($\mathbf{y}_{\text{out}}$) where $\mathbf{y}_{\text{out}}$ takes discrete values. Note that for regression models, the output $\hat{y}_{\text{out}}$ would be a single number which will have to be approximated to the closest value in the set $\{1, 2, 3, 4, 5\}$ to find the final solution, *i.e.*

$$\begin{aligned}\hat{K} =& \arg \min_{i \in [\max_l\{\mathrm{J}_l\}-1]} \|\hat{y}_{\text{out}} - i\|_2^2 \\ =& \arg \min_{i \in [5]} \|\hat{y}_{\text{out}} - i\|_2^2 .\end{aligned} \tag{4.2}$$

However, since the structure of the receive array is known and thence we know the maximum number of echoes that our system can detect, *i.e.* $K \in [5]$, the problem can also be approached as a multiclass or multinomial classification. We can transform the output into 5 different classes and try to approximate a function f that maps input data $\mathbf{X}_{\text{in}}$ to one single output class in $\mathbf{y}_{\text{out}}$. Thus, this is one of those special cases where regression and classification models overlap. We will compare the performance of a number of models, both regression and classification, to finally propose one of them to solve the problem.

With the data collected, the split in train and test sets is done carefully in a way that it is made independently for every SNR value in Table 4.1. This is done so that both train and test sets have examples of every noise scenario. Therefore, once the split is done separately for every value of SNR, the pieces are joined together to have a single train dataset ($\mathbf{X}_{\text{train}}$ and $\mathbf{y}_{\text{train}}$) and a single test dataset ($\mathbf{X}_{\text{test}}$ and $\mathbf{y}_{\text{test}}$). The split is done with percentages 80% training data and 20% test data.

4.3 Proposed models

Next, a set of models are tested and compared in terms of their performance. The training is done following a 3-fold cross-validation over the training set to avoid overfitting and performing a grid search of hyperparameters in those models that accepts them, selecting the best option.

First, we have a set of regression models:

- **Linear regression**, where the mean square error of a linear function that maps $\mathbf{X}_{\text{train}}$ to $\mathbf{y}_{\text{train}}$ is minimized [43].
- **XGBoost Regressor** or Extreme Gradient Boosting for regression. This is an ensemble based on decision trees where the MSE is minimized using gradient descent [44]. The hyperparameter tuning in training/validation stage yielded 20 estimators, a maximum tree depth of 4 and a learning rate of 0.1.
- **Fully connected neural network regression**. We test several topologies with different number of hidden layers with different sizes as depicted in Fig. 4.2(a) where the output layer is activated with a linear function. The best topology found in training/validation stage has 3 hidden layers of sizes 200, 50 and 10.

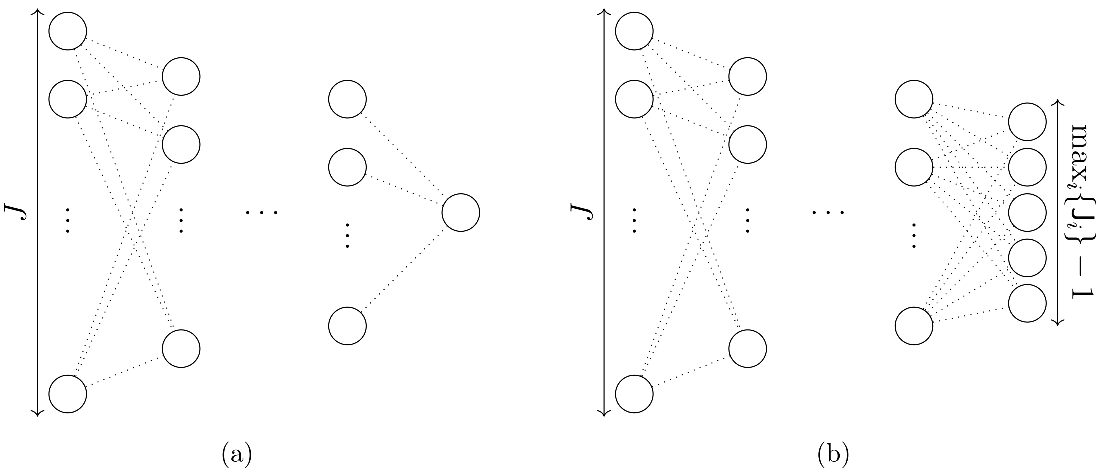


Fig. 4.2. Fully connected neural network topologies. In (a) a regression network with J inputs and only one output and in (b) a multiclass classification with J inputs and 5 outputs activated with a softmax function.

And then a set of classification models:

- **Multinomial Naive Bayes classification**. A very computational efficient algorithm based on Bayes theorem. Often used as a benchmark due to its efficiency [45].

- **XGBoost Classification** or Extreme Gradient Boosting for classification. This is an ensemble based on decision trees where the accuracy is minimized using gradient descent. This works very similarly to the XGBoost Regressor [44]. The hyperparameter tuning in training/validation stage yielded 20 estimators, a maximum tree depth of 4 and a learning rate of 0.1.

- **Multinomial logistic regression**. An extension to Logistic Regression, where binary classification is supported via a binomial . This predicts a multinomial probability, *i.e.* one probability value for each of the output classes [46]. The hyperparameters for the model are: $\ell 2$ penalty term, and a L-BFGS solver [47].

- **Fully connected neural network classification**. As in Fig. 4.2(b) where the output layer is activated with a softmax function with 5 different outputs (one per each class) which tell the probability of the input to belong to each of the classes. Several topologies are tested in training/validation stage. The topology giving the best results is the same as for the regression neural network, except for the output layer (5 different outputs) and the activation function in the output layer.

	MSE		
Model	0dB	5dB	25dB
Linear regression	1.44	0.86	0.59
XGBoost Regressor	1.29	0.81	0.65
Neural Network regression	1.2	0.49	0.1

TABLE 4.2. REGRESSION MODELS PERFORMANCE

4.4 Results

4.4.1 Regression models

The comparison on the performance of the regression models is shown in Table 4.2 where we calculate the MSE for three different SNRs values. With these three values, we have an insightful view on the behavior of the models under different noisy scenarios. The MSE is calculated as

$$\epsilon_{\text{MSE}} = \frac{1}{n}\sum_{i}^{n}\left|K - \hat{K}\right|^2 \tag{4.3}$$

where n is the number of examples in the test dataset, K is the actual number of atoms and $\hat{K}$ is the estimate of the model.

The best regression model, based on the MSE, is the neural network regression.

4.4.2 Classification models

Finally, we discuss here the performance of the classification models. In Table 4.3 the comparison on the performance of the models is shown for three different SNR values where we calculate the overall accuracy and the MSE. Note that calculating the MSE in a classification problem is not common at all, but, since this is a very special case of classification where the labels have also an actual numerical meaning we can calculate it following (4.3). In Figs 4.3-4.6 we depict the confusion matrices for the classifiers for the same three values of SNR. Also, in Fig. 4.7(b) we plot the performance (accuracy and MSE) in terms of SNR.

For multinomial classification problems, the typical metrics for binary classification cannot be used, except for the accuracy that basically is the rate of correct predictions over all the predicted set. However, the problem can be converted into a set of binary classifications where we calculate the metrics evaluating one *vs* the rest, *e.g.* we can consider that the positive class is $K = 1$ and the negative class $K \neq 1$, and then repeat the process for the 5 classes and average over all the One *VS* Rest metrics. Thence, considering TP as the number of true positives, TN as the number of true negatives, FP as the number of false positives and FN as the number of false negatives we calculate the following metrics:

- **Accuracy**. This is the most intuitive metric. The ratio of observations predicted correctly over the total of observations:

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{4.4}$$

- **Precision**. Among every observation that the model predicted positive ($TP + FP$), how many did the model predicted correctly. A typical example where this metric is useful is anti-spam filters where the model has to be very sure that if a mail is labelled as spam, it really is.

$$precision = \frac{TP}{TP + FP} \tag{4.5}$$

- **Recall**. Among all observations that are actually positive ($TP + FN$), how many did the model predicted correctly. A typical example where this metric is useful is in health related issues where the illness of an individual needs to be detected precisely.

$$recall = \frac{TP}{TP + FN} \tag{4.6}$$

- **F1-score**. This is a balance between precision and recall metrics and is calculated as

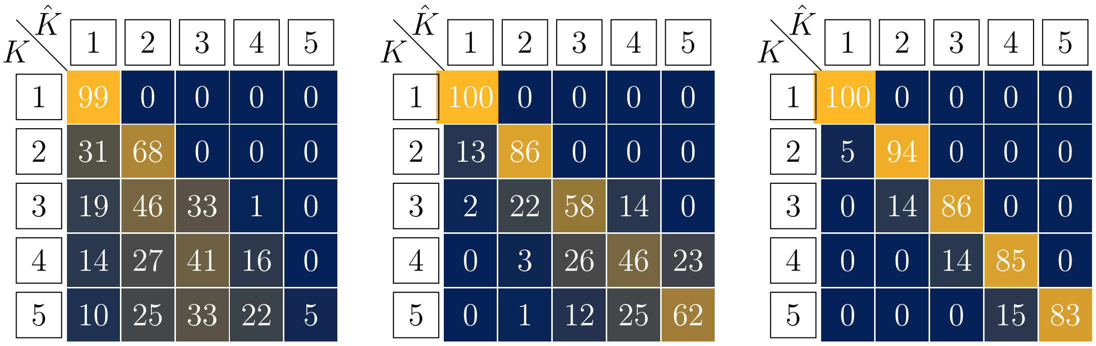
$$\text{f1-}score = \frac{2 \times recall \times precision}{recall + precision} \tag{4.7}$$

- **AUROC** (Area Under the Receiver Operating Curve). The ROC curve plots the probability of having false positive against the probability of true positive and the AUROC is the area under that curve which is used as a summary of the curve itself [48]. A value of $0.5 < \text{AUROC} < 1$ denotes a high chance that the model predicts the labels correctly, with 0.5 meaning that the model makes random guesses and a value of 1 when the model labels the observations perfectly. Values less than 0.5 may be a sign that there is enough information to train a good model but the labelling or the training went wrong.

In the lights of the results, we choose the neural network classifier as the best model among both the regressors and classifiers, according to its confusion matrices in Fig. 4.6, its overall performance in Table 4.3 and Fig. 4.7(b), and the One *VS* Rest results in Fig. 4.8(d).

	Accuracy			**MSE**		
Model	0dB	5dB	25dB	0dB	5dB	25dB
Naive Bayes classifier	44.75%	70.85%	89.85%	2	0.44	0.1
Logistic regression	47.84%	70.28%	94.86%	1.72	0.53	0.05
XGBoost classifier	48.79%	64.79%	95.57%	1.58	0.47	0.04
Neural Network classifier	48.7%	72.31%	97.47%	1.54	0.18	0.02

TABLE 4.3. CLASSIFICATION MODELS PERFORMANCE



K \ $\hat{K}$	1	2	3	4	5
1	99	0	0	0	0
2	31	68	0	0	0
3	19	46	33	1	0
4	14	27	41	16	0
5	10	25	33	22	5

(a) SNR = 0dB

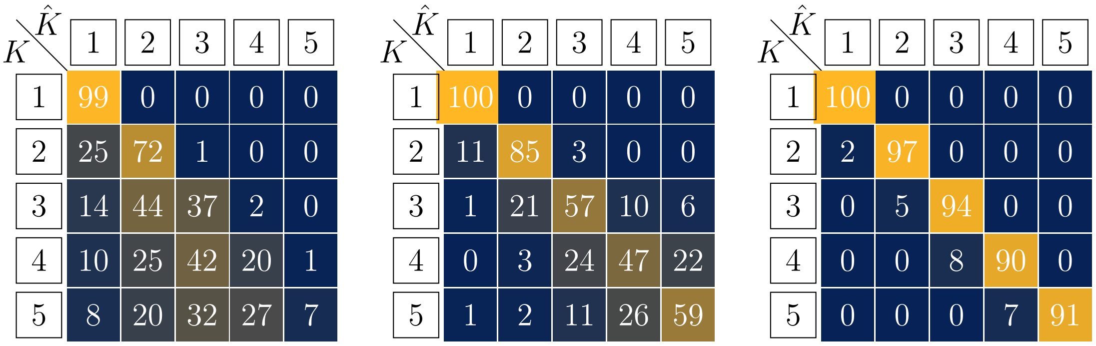
K \ $\hat{K}$	1	2	3	4	5
1	100	0	0	0	0
2	13	86	0	0	0
3	2	22	58	14	0
4	0	3	26	46	23
5	0	1	12	25	62

(b) SNR = 5dB

K \ $\hat{K}$	1	2	3	4	5
1	100	0	0	0	0
2	5	94	0	0	0
3	0	14	86	0	0
4	0	0	14	85	0
5	0	0	0	15	83

(c) SNR = 25dB

Fig. 4.3. Normalized (%) confusion matrices for Multinomial Naive Bayes classifier for different SNRs. In (a) for SNR = 0dB, in (b) for SNR = 5dB and in (c), SNR = 25dB.



K \ $\hat{K}$	1	2	3	4	5
1	99	0	0	0	0
2	25	72	1	0	0
3	14	44	37	2	0
4	10	25	42	20	1
5	8	20	32	27	7

(a) SNR = 0dB

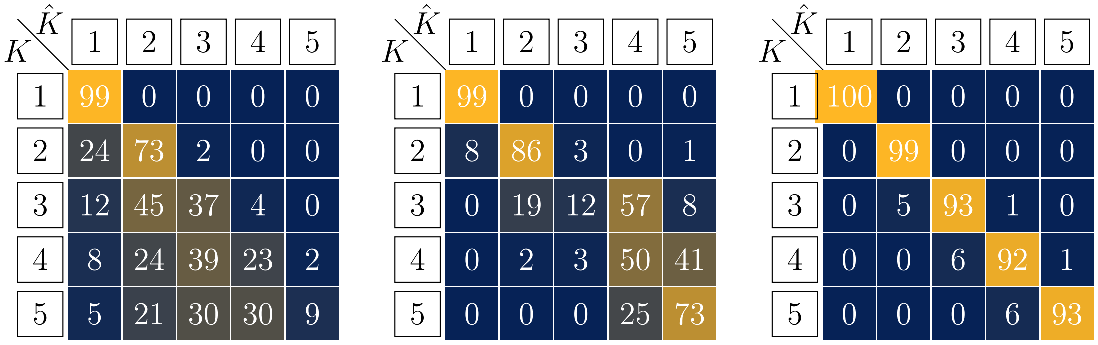
K \ $\hat{K}$	1	2	3	4	5
1	100	0	0	0	0
2	11	85	3	0	0
3	1	21	57	10	6
4	0	3	24	47	22
5	1	2	11	26	59

(b) SNR = 5dB

K \ $\hat{K}$	1	2	3	4	5
1	100	0	0	0	0
2	2	97	0	0	0
3	0	5	94	0	0
4	0	0	8	90	0
5	0	0	0	7	91

(c) SNR = 25dB

Fig. 4.4. Normalized (%) confusion matrices for Logistic regression for different SNRs. In (a) for SNR = 0dB, in (b) for SNR = 5dB and in (c), SNR = 25dB.



K \ $\hat{K}$	1	2	3	4	5
1	99	0	0	0	0
2	24	73	2	0	0
3	12	45	37	4	0
4	8	24	39	23	2
5	5	21	30	30	9

(a) SNR = 0dB

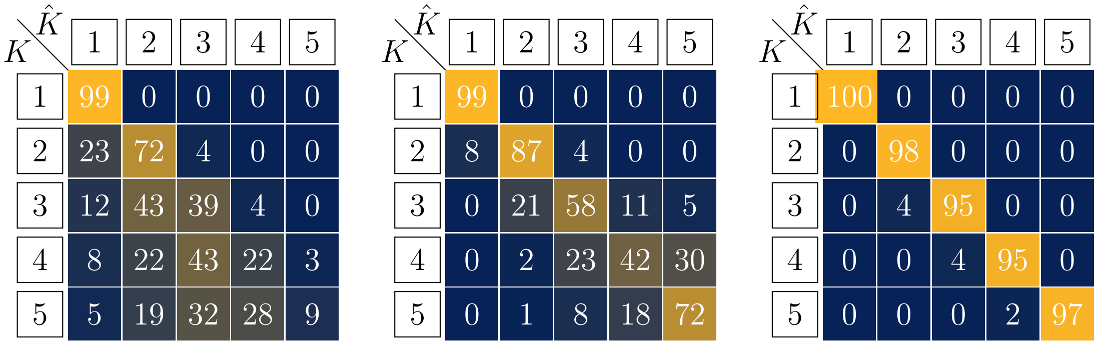
K \ $\hat{K}$	1	2	3	4	5
1	99	0	0	0	0
2	8	86	3	0	1
3	0	19	12	57	8
4	0	2	3	50	41
5	0	0	0	25	73

(b) SNR = 5dB

K \ $\hat{K}$	1	2	3	4	5
1	100	0	0	0	0
2	0	99	0	0	0
3	0	5	93	1	0
4	0	0	6	92	1
5	0	0	0	6	93

(c) SNR = 25dB

Fig. 4.5. Normalized (%) confusion matrices for XGBoost classifier for different SNRs. In (a) for SNR = 0dB, in (b) for SNR = 5dB and in (c), SNR = 25dB.



K \ $\hat{K}$	1	2	3	4	5
1	99	0	0	0	0
2	23	72	4	0	0
3	12	43	39	4	0
4	8	22	43	22	3
5	5	19	32	28	9

(a) SNR = 0dB

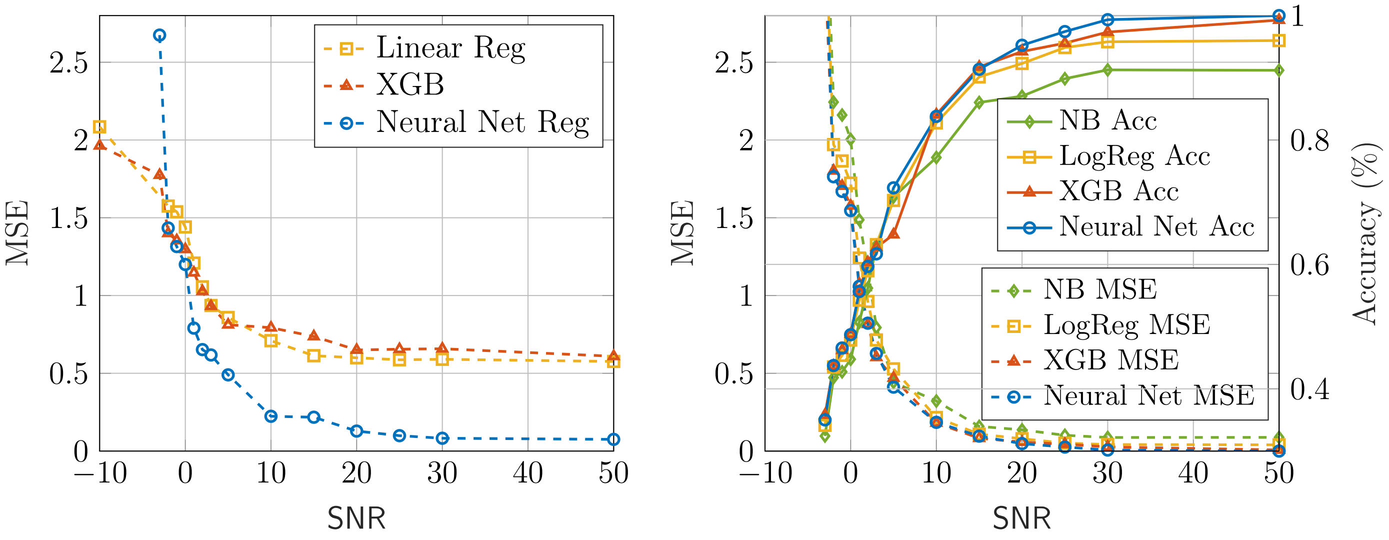
K \ $\hat{K}$	1	2	3	4	5
1	99	0	0	0	0
2	8	87	4	0	0
3	0	21	58	11	5
4	0	2	23	42	30
5	0	1	8	18	72

(b) SNR = 5dB

K \ $\hat{K}$	1	2	3	4	5
1	100	0	0	0	0
2	0	98	0	0	0
3	0	4	95	0	0
4	0	0	4	95	0
5	0	0	0	2	97

(c) SNR = 25dB

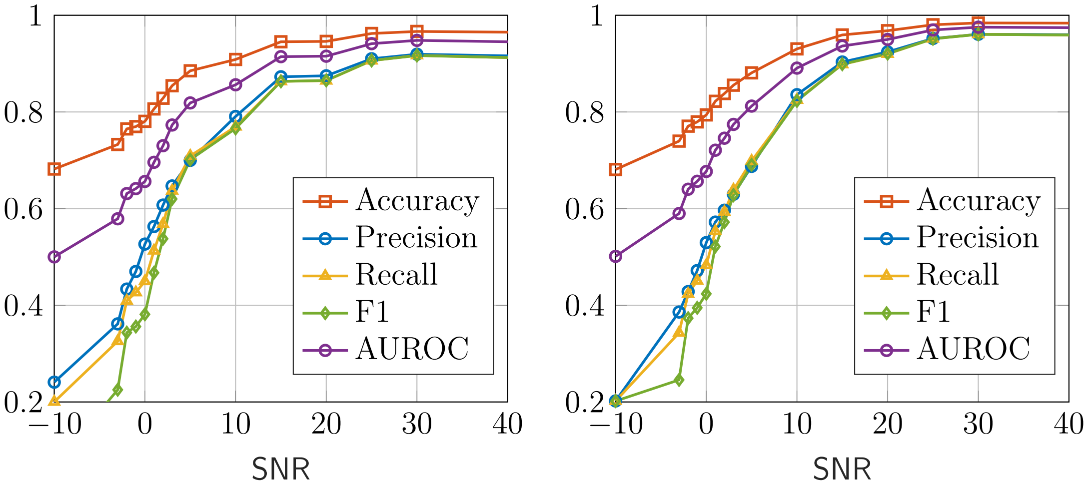
Fig. 4.6. Normalized (%) confusion matrices for Neural network classifier for different SNRs. In (a) for SNR = 0dB, in (b) for SNR = 5dB and in (c), SNR = 25dB.



(a)

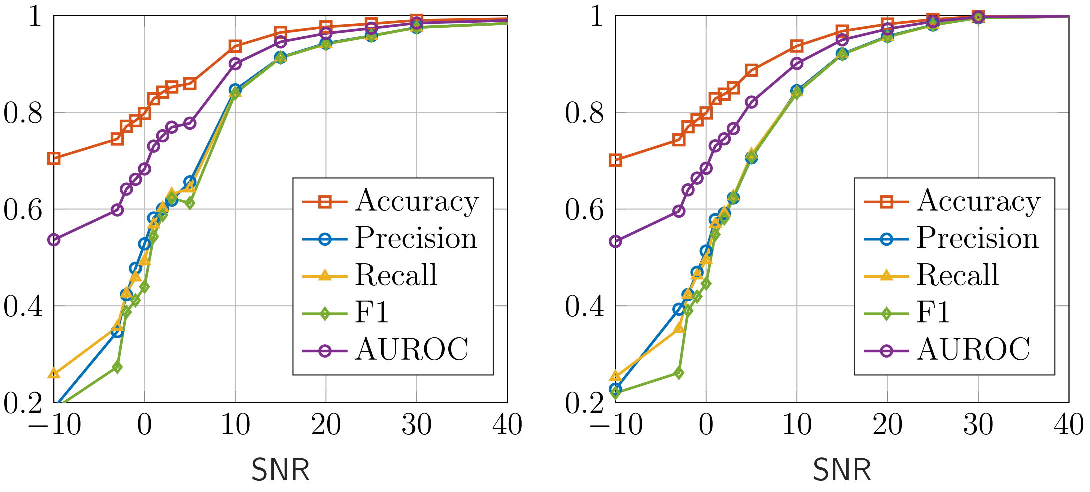
(b)

Fig. 4.7. Comparison of the proposed models. In (a), MSE is depicted in terms of SNR for the regression models and in (b), accuracy and MSE in terms of SNR are plotted for classification models.



(a) Multinomial Naive Bayes

(b) Multinomial Logistic Regression



(c) XGBoost classifier

(d) Neural network classifier

Fig. 4.8. One *VS* Rest metrics in terms of SNR for the classification models. (a) for Multinomial Naive Bayes, (b) for Multinomial Logistic Regression, (c) for XGBoost Classifier and (d) for the neural network classifier.

Chapter 5

Super-resolution in automotive pulse radars

Millimeter-wave (mmWave) radars are one of the key technologies for advanced driver assistant systems (ADASs) and eventually for future self–driving cars [49]–[53]. This adoption is made possible given recent hardware developments, which allow building cost-effective wideband transceivers equipped with a large number of antennas and powerful digital processing units [54], [55]. Automotive radars currently operate in two frequency bands across the globe, namely, 24.25$\div$26.65 GHz (expected to be phased out by 2022) and 76 $\div$ 81 GHz [55]–[57], and are already employed in adaptive cruise control with stop-and-go, automatic emergency brake, lane change and park assistance, cross traffic alert, blind spot detection, and collision warning [58]. Differently from other optical and ultrasound sensors, mmWave radars can operate under adverse weather and light conditions and enable the simultaneous measurement of the delay, radial velocity, and azimuth and elevation angles of arrival (AoA) of prospective targets/obstacles within a single Coherent Processing Interval (CPI), thus allowing the construction of a four dimensional image of the inspected region.

This image is obtained by using adequate signal processing techniques to recover the echoes generated by multiple prospective targets. Classical approaches for delay and Doppler estimation rely on the correlation properties of the probing signal and achieve a resolution approximately equal to the inverse of the signal bandwidth and duration, respectively, also known as Rayleigh limit. Similarly, digital beamforming is commonly used for direction of arrival estimation, whose angular resolution mainly depends upon the number of antennas deployed. More advanced algorithms may obtain a resolution beyond the Rayleigh limit (usually referred to as super-resolution) for a sufficiently large SNR at the price of an increased complexity. Unfortunately, well-established solutions based on the use of the sample covariance matrix, such as the Capon's method [59], the multiple signal classification (MUSIC) [60], the estimation of signal parameters via rotational invariance techniques (ESPRIT) [61],[62], and their subsequent extensions, may hardly be employed in automotive applications. Indeed, the surrounding environment is rapidly changing over time following the vehicle speed and the dynamics of the vehicle surroundings, limiting the number of valid snapshots that can be processed since in most cases the scenario will change even at every new snapshot. Also, the performance of these algorithms may degrade when the number of targets is unknown. Recently, an adaptive matched filter with Adaptive Matched Filter with Iterative Interference Cancellation (IIC-AMF) has been derived in [63], which extracts the prospective echoes one-by-one, after removing the interference caused by the previously-detected (stronger) signal components. This approach perform well on a single snapshot and do not require prior knowledge of the model order; its major limitation is the need of a search-grid, whereby not only we will face off-grid losses, but more importantly its implementation may become unaffordable in a multi-dimensional parameter space due to the exponential increase of required grid points.

To overcome the limitations of the above methods, we propose to address gridless techniques based on the atomic norm minimization to enforce sparsity in the continuous parameter domain [24], [28], [29] with the constraint that we need to address non-uniform antenna deployments that for sure will be present in automotive applications since the radar is likely to be integrated in the car body and massive antenna deployments may be forced to match the curvature/volume of the available surfaces/cavities.

In this chapter, we consider a wideband mmWave pulse radar and study the problem of super-resolving the echoes generated by multiple prospective targets. We first derive the signal model and show that the sum of the received echoes can be represented as a sparse linear combination of atoms as in (2.4); the model unveils the four dimensional environment image with delay and Doppler dimensions as well as a non-uniform 3-D receive array for azimuth and elevation measurement. Then, the positioning parameters are recovered by leveraging the procedure presented in chapter 2 and in [24]. Furthermore, since this problem has a complexity scaling up with the dimension of the observed data vector, we propose and study a reduced-complexity approximation. In addition to the convex relaxation of the ℓ_0-AN in terms of the ℓ_1 atomic norm (ℓ_1-AN) [21], [30], we parse the data samples acquired along one (or more) dimension(s) into multiple reduced-sized groups sharing the same (up to the phase of the weighting coefficients) atomic decomposition. Sufficient conditions for unique signal recovery in the noiseless case are provided and discussed in chapter 2, while the performance achievable in additive Gaussian noise are assessed by means of numerical examples.

5.1 Radar system description

The radar is considered to be operating at the carrier frequency f_o and equipped with a wide-beam illuminator and a co-located receive array. The designed system aims at determining of the number of observed echoes and of their absolute amplitudes and location parameters (delay, Doppler shift, azimuth, and elevation), assuming that each echo is generated by a point-like target.

5.1.1 Probing signal

The probing signal is composed of pulse trains in order to provide reliable reference time for delay measurement. Up to a scaling factor accounting for the transmit energy (included next in the target response), the baseband signal emitted over a CPI with duration T_{cpi} is [64]

$$\begin{aligned} s(t) &= \sum_{l=0}^{\mathsf{J}_\nu-1} \underbrace{\sum_{p=0}^{C} c_l(p)\psi(t-pT_c-lT)}_{g_l(t-lT)} \\ &= \sum_{l=0}^{\mathsf{J}_\nu-1} g_l(t-lT) \end{aligned} \tag{5.1}$$

where

- T is the Pulse Repetition Time (PRT), which controls both the unambiguous delay and Doppler intervals;
- $\mathsf{J}_\nu \geq 1$ is the number of processed pulses, which determines the length of the CPI, say $T_{\text{cpi}} = \mathsf{J}_\nu T$;
- $g_l(t)$ is the l-th pulse, composed of C subpulses (also called chips), where C is a positive integer;
- $\mathbf{c}_l = [c_l(0), \dots, c_l(C-1)]^\top$ is the fast-time code of the l-th pulse, normalized to have $\|\mathbf{c}_l\|_2^2 = C$;
- $\psi(t)$ is the subpulse waveform, assumed to have unit-energy and support in $[0, T_\psi]$;
- T_c is the subpulse repetition time.

In the following we denote by $T_g = (C-1)T_c+T_\psi$ the duration of each pulse. The pulse $g_l(t)$, depending on the fast-time codes, determine the occupied bandwidth, say B, of the probing signal. In this work, we propose to set $\mathbf{c}_l$ equal for all pulses following a Barker sequence [64], [65]. Finally, the CPI is limited by the target mobility, as otherwise delay and/or Doppler and/or angle migration may occur. For example, let $\upsilon_{\max} > 0$ and $\dot{\upsilon}_{\max} > 0$ be the maximum radial velocity and acceleration; then, delay and Doppler migration can be neglected if

$$\frac{2\upsilon_{\max} T_{\text{cpi}}}{c_o} \ll \frac{1}{B}, \quad \frac{2\dot{\upsilon}_{\max} T_{\text{cpi}}}{\lambda_o} \ll \frac{1}{T_{\text{cpi}}} \tag{5.2}$$

respectively.

5.1.2 Received signal

We first consider a 3-D uniform array at the radar receiver and denote by $\mathsf{J}_x \geq 1$, $\mathsf{J}_y \geq 1$, and $\mathsf{J}_z \geq 1$ the number of elements along each of the spatial dimensions, and by ρ^x, ρ^y, and ρ^z the corresponding spacing normalized by λ_o. The signal received by the n-th antenna, positioned at $\lambda_o[x_n\rho^x, y_n\rho^y, z_n\rho^z]$, with $x_n \in \{0, \dots, \mathsf{J}_x - 1\}$, $y_n \in \{0, \dots, \mathsf{J}_y - 1\}$, $z_n \in \{0, \dots, \mathsf{J}_z - 1\}$, and $n \in [\mathsf{J}_x\mathsf{J}_y\mathsf{J}_z]$, can be modeled as

$$r_n(t) = \sum_{k=1}^{K} h_{k,n}(\theta_n, \varphi_n) e^{-j2\pi\nu_k t} s(t - \tau_k) + w_n(t) \tag{5.3}$$

where

- $K \geq 0$ is the number of observed echoes;
- $\tau_k \in [\tau_{\min}, \tau_{\max})$ is the delay of the k-th echo, with $0 < \tau_{\min} < \tau_{\max}$; we assume $\tau_{\max} < T - T_g$, so that the echoes produced by the current pulse are entirely received before the transmission of the next pulse; for half-duplex operation, we also require $\tau_{\min} > T_g$;
- $\nu_k \in [-\nu_{\max}, \nu_{\max})$ is the Doppler shift of the k-th echo, with $\nu_{\max} = 2\upsilon_{\max}/\lambda_o$;
- $h_{k,n}(\theta_k, \varphi_k) \in \mathbb{C}$ is the channel response of the k-th echo at the n-th antenna:

  $$h_{k,n}(\theta_k, \varphi_k) = \gamma_k e^{j2\pi(\mathsf{f}_k^x x_n + \mathsf{f}_k^y y_n + \mathsf{f}_k^z z_n)} \tag{5.4}$$

  where $\gamma_k \in \mathbb{C}$ is the target response or gain of the k-th multipath and

  $$\begin{aligned} \mathsf{f}_k^x &= \operatorname{mod}(\rho^x \sin(\theta_k)\cos(\varphi_k), 1) \\ \mathsf{f}_k^y &= \operatorname{mod}(\rho^y \sin(\theta_k)\sin(\varphi_k), 1) \\ \mathsf{f}_k^z &= \operatorname{mod}(\rho^z \cos(\theta_k), 1) \end{aligned} \tag{5.5}$$

  are the frequencies along the dimensions of the array, respectively, with θ_k and φ_k being azimuth and elevation angles of arrival, respectively, defined similarly as in (3.5)[2].

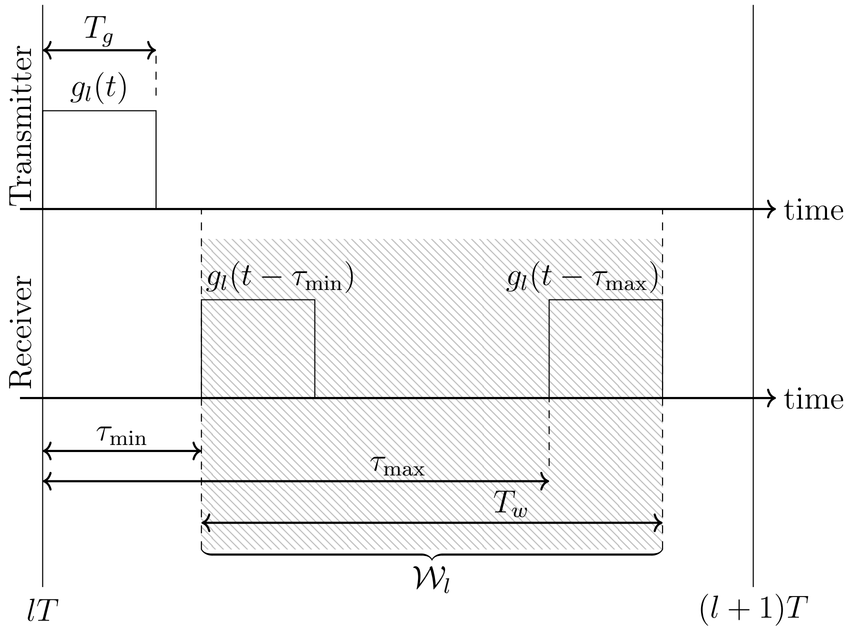


Fig. 5.1. Example of processing window in the l-th pulse interval.

- $w_n(t)$ is the additive noise, modeled as a complex circularly-symmetric Gaussian process with independent zero-mean components and σ_w^2 variance.

As illustrated in Fig. 5.1, in the l-th pulse interval we process the signal received in the following window

$$\mathcal{W}_l = \Big[lT + \lfloor \tau_{\min}/T_c \rfloor T_c, lT + \lceil (\tau_{\max} + T_g)/T_c \rceil T_c\Big] \tag{5.6}$$

with duration given by T_w.

Also, consider the set of projection signals $\{\chi_{l,m}(t)\}_{m=-\infty}^{+\infty}$

$$\chi_{l,m}(t) = \begin{cases} \frac{1}{\sqrt{T_w}} e^{-j2\pi \frac{m}{T_w} t}, & t \in \mathcal{W}_l \\ 0, & t \notin \mathcal{W}_l. \end{cases} \tag{5.7}$$

which we use for further processing the received signal $r_n(t)$ in the n-th antenna by projecting it onto each of the signals $\chi_{l,n}(t)$. In particular, we have the following

---

[2] Note that, compared to the channel model in chapter 3 in (3.9), here $h_{k,n}(\theta_k, \varphi_k)$ represents the channel response from the k-th target as a single element radiating point, *i.e.* $\mathsf{M} = [1]$, to a 3-D uniform array $\mathsf{N} = [\mathsf{J}_x, \mathsf{J}_y, \mathsf{J}_z]^\top$. Then, for the k-th target, the channel matrix would correspond to:

$$\mathbf{H}_k = \begin{bmatrix} h_{k,1}(\theta_k, \varphi_k) \\ h_{k,2}(\theta_k, \varphi_k) \\ \vdots \\ h_{k,\mathsf{J}_x\mathsf{J}_y\mathsf{J}_z}(\theta_k, \varphi_k) \end{bmatrix} = \mathbf{r}_{\mathsf{N}}\left(\begin{bmatrix} \mathsf{f}_k^x \\ \mathsf{f}_k^y \\ \mathsf{f}_k^z \end{bmatrix}\right)\gamma_k$$

resulting samples

$$\begin{aligned} y_{n,l,m} =& \int_{\mathcal{W}_l} r_n(t)\chi^*_{l,m}(t)dt \\ =&\sqrt{\frac{1}{T_w}}\sum_{k=1}^{K} h_{k,n}(\theta_k,\varphi_k)\sum_{p=0}^{C-1} c_l(p)\int_{\mathcal{W}_l}\psi(t-\tau_k-pT_c-lT)e^{j2\pi\left(\frac{m}{T_w}-\nu_k\right)t}dt \\ &+\frac{1}{\sqrt{T_w}}\int_{\mathcal{W}_l} w_n(t)e^{j2\pi\frac{m}{T_w}t}dt \\ =&\sqrt{\frac{1}{T_w}}\sum_{k=1}^{K} h_{k,n}(\theta_k,\varphi_k)\sum_{p=0}^{C-1} c_l(p)\Psi\left(\frac{m}{T_w}-\nu_k\right)e^{j2\pi\left(\frac{m}{T_w}-\nu_k\right)(\tau_k+pT_c+lT)}+w_{n,l,m} \end{aligned} \tag{5.8}$$

where $\Psi(\beta)$ is the Fourier Transform of $\psi(t)$:

$$\Psi(\beta)=\int_0^{T_\psi}\psi(t)e^{j2\pi\beta t}dt \tag{5.9}$$

and $w_{n,m,l}$ is the projection of $w_n(t)$ onto $\chi_{l,n}(t)$.

Notice now that for typical system parameters we have

$$\nu_k/B \ll 1 \tag{5.10}$$

$$2\pi\nu_k T_g \ll 1; \tag{5.11}$$

for example, if $f_o = 79$ GHz, $B = 4$ GHz, $T_c \simeq 1/B$, $C = 10$, and $T_\psi = 6T_c$, even assuming $\nu_k = 80$ KHz (corresponding to a relative radial velocity of 546 Km/h, which is beyond that encountered in most automotive applications), we have $\nu_k/B \simeq 2\cdot 10^{-5}$ and $2\pi\nu_k T_g \simeq 1.9\cdot 10^{-3}$. Exploiting (5.10) and (5.11), we can write

$$\Psi\left(\frac{m}{T_w}-\nu_k\right)\simeq\Psi\left(\frac{m}{T_w}\right) \tag{5.12}$$

$$e^{j2\pi\nu_k pT_c}\simeq 1,\quad p=0,\dots,C-1 \tag{5.13}$$

whereby $y_{n,m,l}$ in (5.8) simplifies to

$$\begin{aligned} y_{n,l,m} =& \underbrace{\sqrt{\frac{1}{T_w}}\Psi\left(\frac{m}{T_w}\right)\left(\sum_{p=0}^{C-1} c_l(p)e^{j2\pi\frac{m}{M}p}\right)e^{j2\pi\frac{mlN}{M}}}_{d_{l,m}}\sum_{k=1}^{K} h_{k,n}(\theta_k,\varphi_k)e^{j2\pi\mathsf{f}^\tau_k m}e^{j2\pi\mathsf{f}^\nu_k l}+w_{n,m,l} \\ =&d_{l,m}\sum_{k=1}^{K}\beta_k e^{j2\pi\mathsf{f}^x_k x_n}e^{j2\pi\mathsf{f}^y_k y_n}e^{j2\pi\mathsf{f}^z_k z_n}e^{j2\pi\mathsf{f}^\tau_k m}e^{j2\pi\mathsf{f}^\nu_k l}+w_{n,m,l} \end{aligned} \tag{5.14}$$

where we have absorbed the irrelevant term $e^{-j2\pi\nu_k\tau_k}$ into the phase of the target response in (5.4) defining the new term $\beta_k = e^{-j2\pi\nu_k\tau_k}\gamma_k$, and we have defined the delay and Doppler frequencies

$$\mathsf{f}^\tau_k = \mathrm{mod}\left(\frac{\tau_k}{T_w},1\right) \tag{5.15}$$

$$\mathsf{f}^\nu_k = \mathrm{mod}\left(-\nu_k T,1\right) \tag{5.16}$$

respectively.

- The complex coefficient $d_{l,m}$ is independent of the unknown parameters of the observed echoes.
- Since $\tau_{\max} - \tau_{\min} < T_w$ by construction, f_k^τ uniquely specifies the delay of the k-th target.
- If $\nu_{\max} \leq 1/(2T)$, f_k^ν uniquely specifies the Doppler shift of the k-th target.

The number of non trivial samples $\{y_{n,l,m}\}_{m=-\infty}^{\infty}$ in m is deeply linked to the product between the signal bandwidth and the length of the processing window. Defining $F = \lfloor BT_w/2 \rfloor$ we consider as significant the samples $\{y_{n,l,m}\}_{m=-F}^{F} \forall n, l$ and define $\mathsf{J}_\tau = 2F + 1$. However, for the n-th antenna and l-th PRT, computing and processing all the Fourier coefficients is not feasible in practice due to complexity constraints. Here we may choose $F < \lfloor BT_w/2 \rfloor$ to limit the complexity, as it shown in section 5.2.2.

5.1.3 Vector representation of the received signal

In total, the number of received samples is $J = \mathsf{J}_x\mathsf{J}_y\mathsf{J}_z\mathsf{J}_\nu\mathsf{J}_\tau$, being each sample $y_{n,(l-1),(m-(F+1))}$ with $n \in [\mathsf{J}_x\mathsf{J}_y\mathsf{J}_z]$, $l \in [\mathsf{J}_\nu]$ and $m \in \mathsf{J}_\tau$. The set of dimensions $\{\mathsf{J}_x, \mathsf{J}_y, \mathsf{J}_z, \mathsf{J}_\nu, \mathsf{J}_\tau\}$ can be properly ordered defining the ordered vector $\mathbf{J} = [\mathsf{J}_1, \mathsf{J}_2, \mathsf{J}_3, \mathsf{J}_4, \mathsf{J}_5]^\top = \mathbf{\Pi}[\mathsf{J}_x, \mathsf{J}_y, \mathsf{J}_z, \mathsf{J}_\nu, \mathsf{J}_\tau]^\top$, where $\mathbf{\Pi}$ is a $d \times d$ permutation matrix, with in this case $d = 5$, which sorts the dimensions in ascending order such that $\mathsf{J}_1 \leq \mathsf{J}_2 \leq \cdots \leq \mathsf{J}_5$. Note that the received samples can be piled arbitrarily and if done properly, the composite atoms posses a Kronecker ordering such that $\mathsf{J}_5 \rightarrow \mathsf{J}_4 \rightarrow \cdots \rightarrow \mathsf{J}_1$, with sampling dimensions following $\mathsf{J}_1 \leq \mathsf{J}_2 \leq \cdots \leq \mathsf{J}_5$. The irrelevant scaling coefficient $e^{-j2\pi \mathsf{f}_k^\tau F} \approx 1$ can be embedded into the recently created term β_k, defining the term $u_k = e^{-j2\pi \mathsf{f}_k^\tau F} e^{j2\pi\nu_k\tau_k}\gamma_k$. Then, the resulting vector from piling the received samples in a proper way is

$$\mathbf{y} = \mathbf{D}\sum_{k=1}^{K} u_k \mathbf{a}_{\mathbf{J}}(\mathbf{f}_k) + \mathbf{w} \tag{5.17}$$

where $\mathbf{a}_{\mathbf{J}}(\mathbf{f}_k) \in \mathcal{A}_{\mathbf{J}}$, $\mathbf{f} = [\mathsf{f}_1, \mathsf{f}_2, \mathsf{f}_3, \mathsf{f}_4, \mathsf{f}_5]^\top = \mathbf{\Pi}[\mathsf{f}_x, \mathsf{f}_y, \mathsf{f}_z, \mathsf{f}_\nu, \mathsf{f}_\tau]^\top$, $\mathbf{D} \in \mathbb{C}^{J\times J}$ is a diagonal matrix resulting from piling coefficients $d_{l,m}$ with the same piling structure as for the received samples $y_{n,(l-1),(m-(F+1))}$, and the noise vector $\mathbf{w} \in \mathbb{C}^J$ is obtained in a similar way, piling the noise coefficients $w_{n,m,l}$. Note that the Kronecker ordering of the composite atoms allows that

$$\begin{aligned} \mathbf{a}_{\mathbf{J}}(\mathbf{f}_k) =& \mathbf{a}_{\mathsf{J}_1}(\mathsf{f}_k^1) \otimes \mathbf{a}_{\mathsf{J}_2}(\mathsf{f}_k^2) \otimes \mathbf{a}_{\mathsf{J}_3}(\mathsf{f}_k^3) \otimes \mathbf{a}_{\mathsf{J}_4}(\mathsf{f}_k^4) \otimes \mathbf{a}_{\mathsf{J}_5}(\mathsf{f}_k^5) \\ =& \otimes_{i=1}^{d} \mathbf{a}_{\mathsf{J}_i}(\mathsf{f}_k^i) \end{aligned} \tag{5.18}$$

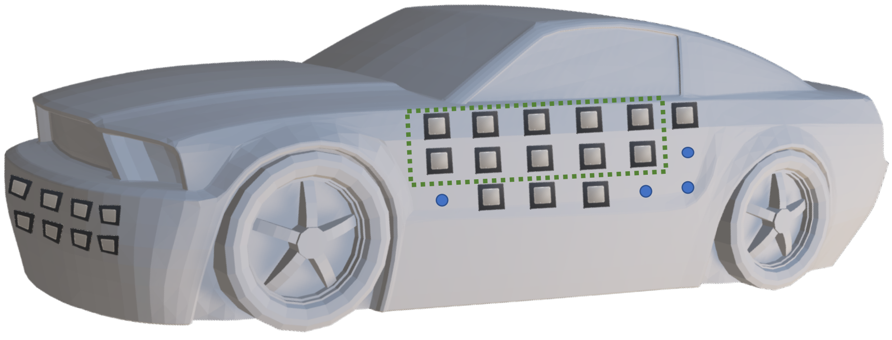


Fig. 5.2. Example of a uniform (front) and a non-uniform (side) planar array deployed on the body of a car; for ease of explanation, the curvature of the involved surfaces is neglected. The blue dots are the missing antennas of the virtual uniform array (see Sec. 5.1.4 for details). The green dotted rectangle identifies the uniform subarray, contained in the side array, with the largest sum of the elements along each dimension (see Sec. 2.2.1 for details).

with $d = 5$.

The piled vector $\mathbf{y} \in \mathbb{C}^J$ in (5.17) is a measurement vector of a sparse model with atoms from the atom set $\mathcal{A}_{\boldsymbol{\mathsf{J}}}$. That vector can also be represented in an enlarged virtual atom domain. To achieve a virtual enlargement of the atom set $\mathcal{A}_{\boldsymbol{\mathsf{J}}}$ to an enlarged atom set $\mathcal{A}_{\boldsymbol{\mathsf{E}}}$, with $\boldsymbol{\mathsf{E}} = [\mathsf{E}_1, \mathsf{E}_2, \mathsf{E}_3, \mathsf{E}_4, \mathsf{E}_5]^\top$, $\boldsymbol{\mathsf{E}} \succeq \boldsymbol{\mathsf{J}}$, $E = \prod_{i=1}^{d} \mathsf{E}_i$ and $\mathsf{E}_i > \mathsf{J}_i$ for at least 1 dimension $i \in [d]$, we make use of a matrix $\mathbf{B}^e \in \{0,1\}^{J \times E}$ such that

$$\begin{aligned}\mathbf{y} =& \mathbf{D} \sum_{k=1}^{K} u_k \mathbf{a}_{\boldsymbol{\mathsf{J}}}(\boldsymbol{\mathsf{f}}_k) + \mathbf{w} \\ =& \mathbf{D} \sum_{k=1}^{K} u_k \mathbf{B}^e \mathbf{a}_{\boldsymbol{\mathsf{E}}}(\boldsymbol{\mathsf{f}}_k) + \mathbf{w} \\ =& \mathbf{B} \sum_{k=1}^{K} u_k \mathbf{a}_{\boldsymbol{\mathsf{E}}}(\boldsymbol{\mathsf{f}}_k) + \mathbf{w}\end{aligned} \tag{5.19}$$

where the enlarged atoms $\mathbf{a}_{\boldsymbol{\mathsf{E}}}(\mathsf{f}_k) \in \mathcal{A}_{\boldsymbol{\mathsf{E}}}$ and $\mathbf{B} = \mathbf{D}\mathbf{B}^e \in \mathbb{C}^{J \times E}$. Note that given the diagonal structure of matrix $\mathbf{D} \in \mathbb{C}^{J \times J}$ and the structure of sampling matrix $\mathbf{B}^e \in \{0, 1\}$, *i.e.* $\mathbf{B}^e = \left[\mathbf{I}_J, \mathbf{0}_{J \times (E-J)}\right] \mathbf{\Pi}$, the structure of the resulting matrix $\mathbf{B} \in \mathbb{C}^{J \times E}$ is similar to that of $\mathbf{B}^e$, *i.e.* only one element per row and column is different from zero.

5.1.4 Non-uniform array

Physical constraints may prevent the deployment of a uniform array [18], [35], [66], [67]: for example, this may occur when the antenna is embedded in the car body to save space and costs (see Fig. 5.2). To handle this situation, we consider here a non-uniform array with J_a elements, obtained by sampling a larger virtual uniform array containing J_x, J_y, and J_z elements,*i.e.* $J_a < \mathsf{J}_x\mathsf{J}_y\mathsf{J}_z$. We define an index set $\mathcal{I}_a \subset [\mathsf{J}_x\mathsf{J}_y\mathsf{J}_z]$, with $|\mathcal{I}_a| = J_a$, which will identify the active antennas from the uniform 3-D deployment.

Thus, in this non-uniform scenario, the available $S = J_a\mathsf{J}_\nu\mathsf{J}_\tau$ received samples $y_{n,(l-1),(m-(F+1))}$ with $n \in \mathcal{I}_a \subset [\mathsf{J}_x\mathsf{J}_y\mathsf{J}_z]$, $l \in [\mathsf{J}_\nu]$ and $m \in [\mathsf{J}_\tau]$ can be read accordingly, as in previous section, to generate the measurement vector $\mathbf{y} \in \mathbb{C}^S$ in the atom set $\mathcal{A}_{\mathsf{J}}$. We recall here that $\mathsf{J} = [\mathsf{J}_1, \mathsf{J}_2, \mathsf{J}_3, \mathsf{J}_4, \mathsf{J}_5]^\top = \mathbf{\Pi}[\mathsf{J}_x, \mathsf{J}_y, \mathsf{J}_z, \mathsf{J}_\nu, \mathsf{J}_\tau]^\top$, where $\mathbf{\Pi}$ is a $d \times d$ permutation matrix which orders the dimensions in an ascending way, *i.e.* $\mathsf{J}_1 \leq \mathsf{J}_2 \leq \cdots \leq \mathsf{J}_5$. To remove the elements in the atoms that represent missing (not active) antennas from the virtual uniform array we use a binary sampling matrix $\mathbf{B}_a \in \{0,1\}^{S \times \mathsf{J}_x\mathsf{J}_y\mathsf{J}_z\mathsf{J}_\nu\mathsf{J}_\tau}$. Thence, the measurement vector $\mathbf{y} \in \mathbb{C}^S$ is formulated as

$$\begin{aligned} \mathbf{y} =& \mathbf{B}_a\mathbf{D}\sum_{k=1}^{K} u_k \mathbf{a}_{\mathsf{J}}(\mathsf{f}_k) + \mathbf{w} \\ =& \mathbf{B}_a\mathbf{D}\sum_{k=1}^{K} u_k \mathbf{B}^e \mathbf{a}_{\mathsf{E}}(\mathsf{f}_k) + \mathbf{w} \\ =& \mathbf{B}\sum_{k=1}^{K} u_k \mathbf{a}_{\mathsf{E}}(\mathbf{f}_k) + \mathbf{w} \end{aligned} \tag{5.20}$$

where $\mathbf{B} = \mathbf{B}_a\mathbf{D}\mathbf{B}^e \in \mathbb{C}^{S\times E}$. Note that the structure of $\mathbf{B} \in \mathbb{C}^{S\times E}$ in this scenario is similar to the one in (5.19) where there is only one non-zero element per row and column.

Fig. 5.2 shows an example of a non-uniform planar array with $J_a = 14$ antennas which is deployed on the side of the car. The non-uniform array is obtained by sampling a virtual uniform array with $\mathsf{J}_x = 1$, $\mathsf{J}_y = 3$, and $\mathsf{J}_z = 6$; the blue dots are the missing elements of the virtual array,

5.1.5 Conditions to meet injectivity properties defined in Definition 5

In Definition 5, the mathematical conditions for a sampling matrix to meet injectivity properties are shown. Particularly, in this radar application the resulting sampling matrix obtained in both scenarios (5.19) (enlarged) and (5.20) (sampled), named both as an abuse of notation $\mathbf{B}$ to keep the notation uncluttered, do not belong to the binary sampling matrices set $\mathcal{B}$. However, thanks to the diagonal structure of $\mathbf{D}$, the conditions in Definition 5 can be easily checked. When multiplying the diagonal matrix $\mathbf{D}$, its rank is preserved which implies that the condition for injectivity is also preserved.

Thus, for the uniform antenna deployment, we have a resulting sampling matrix

$\mathbf{B}$ such that:

$$\begin{aligned}\mathbf{B}\mathbf{a}_{\mathsf{E}}(\mathsf{f}) =& \mathbf{D}\mathbf{B}^{e}\mathbf{a}_{\mathsf{E}} \\ =& \mathbf{D}\mathbf{a}_{\mathsf{J}}(\mathsf{f})\end{aligned} \tag{5.21}$$

The largest non trivial underlying atom structure we can find in the array, *i.e.* in $\mathcal{A}_{\mathsf{J}}$, has the relevant parameters $L_S = \sum_{i=1}^{d} \mathsf{J}_i = \mathsf{J}_x + \mathsf{J}_y + \mathsf{J}_z + \mathsf{J}_\nu + \mathsf{J}_\tau$ and $N_S = \prod_{i=1}^{d} \mathsf{J}_i = J$, given that the largest atom structure is the uniform array itself.

For the non-uniform antenna deployment, the resulting sampling matrix $\mathbf{B}$ is such that:

$$\begin{aligned}\mathbf{B}\mathbf{a}_{\mathsf{E}}(\mathsf{f}) =& \mathbf{B}_a\mathbf{D}\mathbf{B}^{e}\mathbf{a}_{\mathsf{E}} \\ =& \mathbf{B}_a\mathbf{D}\mathbf{a}_{\mathsf{J}}(\mathsf{f})\end{aligned} \tag{5.22}$$

where, in this scenario matrix $\mathbf{B}$ preserves all the samples from Doppler and delay domains, *i.e.* J_ν and J_τ, and thus, the atom structure for these two dimensions is preserved. However, the sampling is done in the 3-D spatial domain, *i.e.* dimensions J_x, J_y and J_z. Following (2.8), the largest underlying structure has dimensions $\mathsf{S}_x \geq 1$, $\mathsf{S}_y \geq 1$ and $\mathsf{S}_z \geq 1$. In this case, the relevant parameters are $L_S = \mathsf{S}_x + \mathsf{S}_y + \mathsf{S}_z + \mathsf{J}_\nu + \mathsf{J}_\tau$ and $N_S = \mathsf{S}_x\mathsf{S}_y\mathsf{S}_z\mathsf{J}_\nu\mathsf{J}_\tau \leq J_a$. In Fig. 5.2, we have an example of a non-uniform array in the side of the car with $J_a = 14$ antennas. Note that the largest underlying structure we can find is highlighted with a green dotted line and its parameters are $\mathsf{S}_x = 1$, $\mathsf{S}_y = 2$ and $\mathsf{S}_z = 5$.

5.2 Gridless signal and echo parameter recovery

The measurement models (5.19) and (5.20), obtained in the previous section from processing and properly piling the received samples, are noisy measurements of a linear combination of K atoms taken from the atomic set $\mathcal{A}_{\mathsf{E}} = \left\{\mathbf{a}_{\mathsf{E}}(\mathsf{f}) | \mathsf{f} \in \mathbb{T}^d\right\}$, with $d = 5$. Here $\mathbf{B}$, which is either in the set $\mathbb{C}^{J\times E}$ or in $\mathbb{C}^{S\times E}$ is a known observation matrix, while $\mathbf{x} \in \mathbb{C}^E$ is the unknown signal of interest containing the information about the prospective echoes.

Therefore, we can apply either (P.5) or its convex relaxation or (P.6) in the enlarged dimension E to estimate the number of received echoes K and their multi-dimensional frequency vectors $\{\mathsf{f}_1, \mathsf{f}_2 \ldots, \mathsf{f}_K\}$. The corresponding numerical results are given in section 5.3.

5.2.1 Complexity analysis

Recalling section 2.3, and given the enlargement in the radar application given by the structure $\mathsf{E} = [\mathsf{E}_1, \mathsf{E}_2, \mathsf{E}_3, \mathsf{E}_4, \mathsf{E}_5]^\top$, we have that the complexity of (P.6) is given by $O(E^{3.5})$. Furthermore, Theorems 1 and 2 provide some useful guidelines to set the relevant system parameters determining E. For example, we can set $\mathsf{E}_i = \mathsf{J}_i$ for all $i = 1, \ldots, d-1$ and $\mathsf{E}_d = (L_S - (d-1))/2$ to ensure the largest resolvabilty region in the noiseless case, while limiting the computational complexity implied by the enlargement of the composite steering vector in (5.19) and (5.20). Also, if a bound on the number on the maximum number of targets K^* would be known, the composite steering vector can be shortened (rather than enlarged) if $K^* < \min\{(L_S - (d-1))/2, \mathsf{J}_d\}$; this idea will be developed in next section, where we propose to reduce the complexity, by accordingly reducing the dimensionality of the problem through splitting the available measurements in multiple sub-vectors.

5.2.2 Reduced-complexity signal and echo parameter recovery

In some applications, the number of data samples along one or more dimensions can be much larger than the number of prospective targets, which gives abundant degrees of freedom. For example, this may happen when the receive array has a massive number of antennas along any of the axes and/or when a large range interval is inspected in wide-band systems and/or when a large number of pulses are processed in low-mobility scenarios. This gap would be even more apparent if motivated by improved performance, we use virtual enlarged scenarios. Recall that in [24] it is shown that by enlarging the last dimension of the structure $\mathsf{E} = [\mathsf{E}_1, \mathsf{E}_2, \mathsf{E}_3, \mathsf{E}_4, \mathsf{E}_5]^\top$, we have better performance in the recovery of the frequencies, both in the noiseless and noisy cases and when we solve for either (P.5) or its convex relaxation or (P.6). Being this the case, we still need to take into account the complexity associated to solve (P.6) and find the proper balance between performance and complexity, specially in those cases where the computational complexity required due to the system parameters may be prohibitive; to overcome this drawback, we can reduce the size of the involved vectors and matrices, at eventually the price of some performance loss in the presence of noise. To this end, we discuss next a low-complexity formulation of (P.6), wherein the data samples acquired along one dimension (typically the one containing the largest number of observations) are split into multiple reduced-sized groups (corresponding to as many sub-arrays or sub-bands or pulse subtrains in the aforementioned examples).

We assume we are in the non-uniform array scenario and that no virtual enlargement is provided, *i.e.* $\mathsf{J} = \mathsf{E}$. Without loss of generality, we focus on the delay dimension, but similar arguments can be replicated for any other dimension. For each pulse and antenna, we take $L_\tau \mathsf{J}_\tau$ Fourier coefficients; these coefficients are then split into $L_\tau \geq 1$ groups of size $\mathsf{J}_\tau \geq 2$, corresponding to as many contiguous and non-overlapping sub-bands, as illustrated in Fig. 5.3. L_τ and J_τ are design parameters that allow to control the computational complexity and the resolvability region, as it will be explained later on. Now, after piling up the samples paralleling

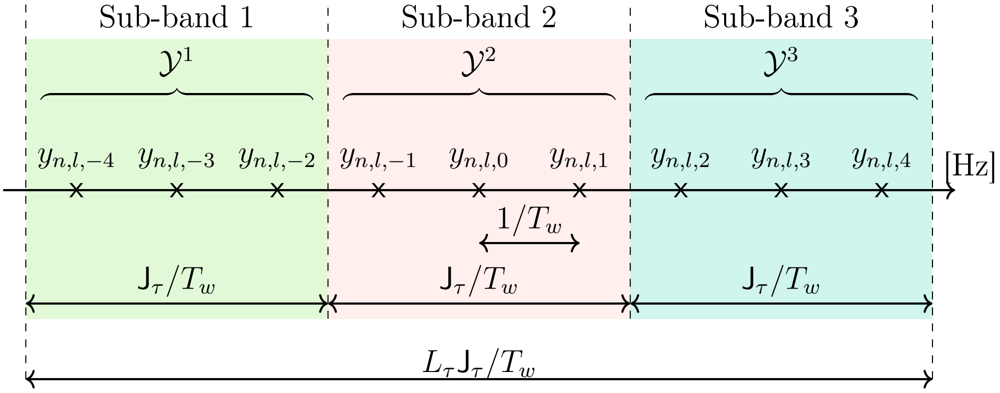


Fig. 5.3. Example of processed sub-bands. It is assumed here that there are 9 samples in the frequency domain within the processed bandwidth (corresponding to the cross-marked frequencies), which are split into $L_\tau = 3$ groups of $\mathsf{J}_\tau = 3$ elements.

the derivations in Sections 5.1.3 and 5.1.4 we obtain L_τ vectors, one per sub-band. Thus, for the ℓ-th sub-band, vector $\mathbf{y}^\ell \in \mathbb{C}^S$ is defined as

$$\mathbf{y}^\ell = \mathbf{B}^\ell \mathbf{x}^\ell + \mathbf{w}^\ell \tag{5.23}$$

with $\ell \in [L_\tau]$, where the sampling matrix $\mathbf{B}^\ell \in \mathbb{C}^{S \times E}$ and the noise vector $\mathbf{w}^\ell \in \mathbb{C}^S$ are defined similarly to $\mathbf{B}$ and $\mathbf{w}$ in (5.20), and

$$\mathbf{x}^\ell = \sum_{k=1}^{K} u_k e^{j2\pi \mathsf{f}_k^\tau (\ell-1)\mathsf{J}_\tau} \mathbf{a}_{\mathbf{J}}(\boldsymbol{\mathsf{f}}_k) = \mathbf{A}_{\mathbf{J}}\left(\boldsymbol{\mathsf{f}}_{1:K}\right) \mathbf{u}^\ell \in \mathbb{C}^E \tag{5.24}$$

is a linear combination of K atoms in $\mathcal{A}_{\mathbf{J}}$, with

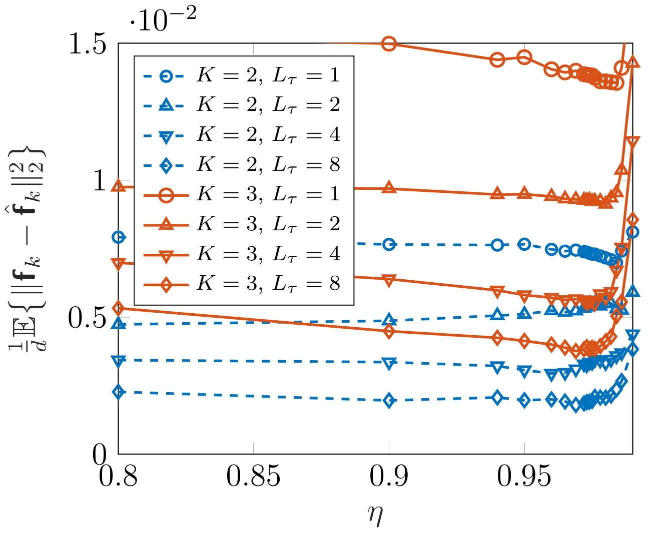
$$\mathbf{u}^\ell = [u_1 e^{j2\pi \mathsf{f}_1^\tau (\ell-1)\mathsf{J}_\tau}, \dots, u_K e^{j2\pi \mathsf{f}_K^\tau (\ell-1)\mathsf{J}_\tau}]^\top \in \mathbb{C}^K.$$

We highlight that the target parameters $\{\boldsymbol{\mathsf{f}}_k, u_k\}$ with $k \in [K]$ specifying $\mathbf{x}^\ell$ are the same on all sub-bands; hence, $\mathbf{x}^\ell$ with $\ell \in [L_\tau]$ not only have the same ℓ_0-AN, but also a similar atomic decomposition, which only differs in the phase of the weighting coefficients, leading to a Multiple Measurement Vector (MMV) scenario similar to the one in (3.12).

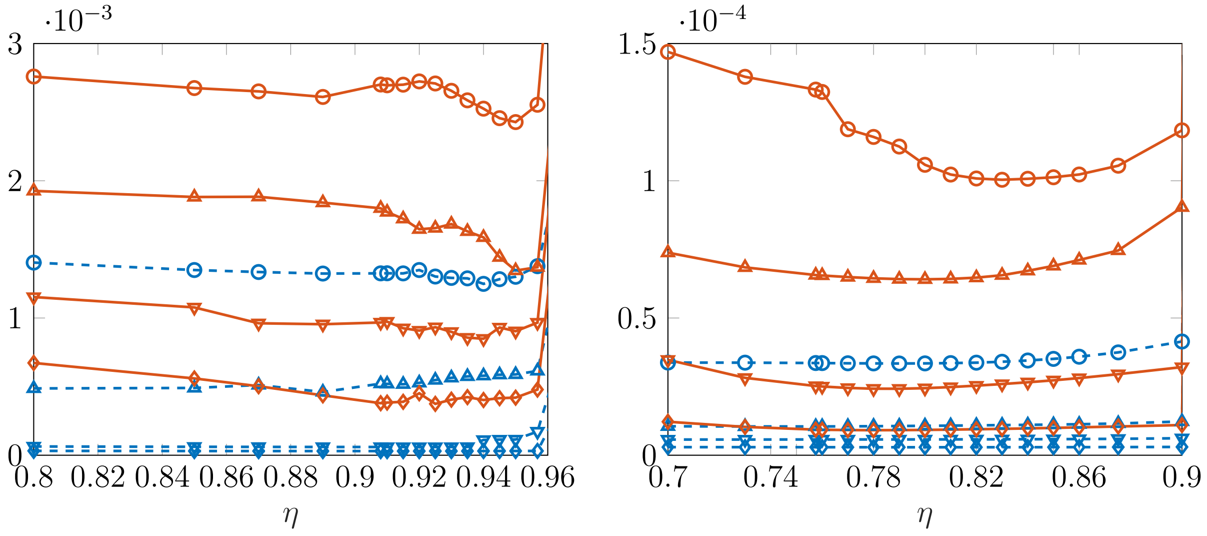
Given the measurements $\mathbf{y}^\ell$, with $\ell \in [L_\tau]$, we propose to recover $\mathbf{x}^\ell$ with $\ell \in [L_\tau]$ as the solution to

$$\begin{aligned} \min_{t,\, \mathbf{x}^\ell,\, \mathbf{T}_{\mathbf{J}} \in \mathcal{T}} \; & \frac{1-\eta}{L_\tau} \sum_{\ell \in [L_\tau]} \left\| \mathbf{y}^\ell - \mathbf{B}^\ell \mathbf{x}^\ell \right\|_2^2 + \frac{\eta}{2}\big(t + \mathrm{Tr}\left\{\mathbf{T}_{\mathbf{J}}\right\}\big) \\ \text{s.t.} \; & \begin{bmatrix} \mathbf{T}_{\mathbf{J}} & \mathbf{x}^\ell \\ \left(\mathbf{x}^\ell\right)^\dagger & t \end{bmatrix} \succeq 0, \quad \ell \in [L_\tau]. \end{aligned} \tag{P.7}$$

For both uniform and non-uniform antenna deployments, and recalling how L_S is obtained from section 5.1.5, we now have the following recovery result.



(a) $\mathsf{SNR} = 0$dB



(b) $\mathsf{SNR} = 10$dB

(c) $\mathsf{SNR} = 20$dB

Fig. 5.4. MSE in the estimation of entire frequency vector vs η when $[\mathsf{J}_x, \mathsf{J}_y, \mathsf{J}_\nu, \mathsf{J}_z, \mathsf{J}_\tau]^\top = [1, 2, 2, 4, 6]^\top$, $K = \{2, 3\}$ and $L_\tau = \{1, 2, 4, 8\}$.

Theorem 3. *If* $K < \min\{(L_S-(d-1))/2, \mathsf{J}_d\}$, (P.7) *allows to uniquely identify the target parameters* $\{\mathsf{f}_k, u_k\}_{k=1}^{K}$ *specifying* $\mathbf{x}^\ell$ *in the noiseless case if the optimal canonical d-LT matrix, say* $\mathbf{T}^\circ_{\mathsf{J}}$*, is such that* $\operatorname{rank}\{\mathbf{T}^\circ_{\mathsf{J}}\} < \mathsf{J}_d$.

The complexity of (P.7) is obtained upon setting $n = J$ and $m = L_\tau$ in (2.9) and therefore given by $O\left(L_\tau J^{3.5} + L_\tau^2 J^{2.5} + L_\tau^3\right)$. Clearly, increasing L_τ and/or J_τ may be helpful in the presence of noise (as also verified in Sec. 5.3) at the price of a increased complexity; specifically, increasing the number of sub-bands results into additional constraints, while increasing the size of each sub-band results into a larger dimension of the vector $\mathbf{x}^\ell$ and of the matrix $\mathbf{T}_{\mathsf{J}}$. For $L_\tau = 1$, we obtain the case discussed in the previous subsections.

5.3 Numerical analysis

We consider an automotive radar operating with a carrier frequency of $f_o = 79$ GHz, a two-side bandwidth of $B = 4$ GHz, and a PRT of $T = 0.5$ μs. The chip waveform $\psi(t)$ is a raised cosine with roll-off factor 0.22 and support in $[0, 6T_c]$, while $\mathbf{c}_l$ is a Barker code of length $C = 10$, whereby $B \approx 1/T_c$. The car is equipped with a wide-beam illuminator and a receive uniform array with $\mathsf{J}_x = 1$, $\mathsf{J}_y = 2$ and $\mathsf{J}_z = 4$, as the one in the front of the vehicle in Fig. 5.2. We assume that the car can accelerate from 0 to 60 Km/h in 4 s and consider a minimum and maximum radial velocity of $v_{\min} = -54$ Km/h and $v_{\max} = 54$ Km/h, respectively. We assume that $\mathsf{J}_\nu = 2$ pulses are jointly processed and that the minimum and maximum inspected range are $d_{\min} = 58$ cm (tied to the pulse length) and $d_{\max} = 7$ m, respectively, whereby $\tau_{\max} - \tau_{\min} \approx 187T_c$. To limit the computational complexity, we resort to the formulation in (P.7), with $\mathsf{J}_\tau = 6$, and a variable number of sub-bands L_τ, namely, $L_\tau = \{1, 2, 4, 8\}$.Also, no enlargement of the composite steering vector is enforced. With the above parameters, we have $[\mathsf{J}_x, \mathsf{J}_y, \mathsf{J}_\nu, \mathsf{J}_z, \mathsf{J}_\tau]^\top = [1, 2, 2, 4, 6]^\top$ and $J = E = 96$.

The noise process $w_n(t)$ is independent across the antennas and white, with a power spectral density σ_w^2. Unless otherwise stated, the targets are randomly positioned within the inspected region. As to the target response γ_k, we assume that its phase ϕ_k is uniformly distributed in $[0, 2\pi)$; also, its amplitude a_k is set equal to $\bar{a}_k/(c\tau_k/2)^2$, where $\bar{a}_k$ is a unit-power Rayleigh random variable accounting for the target fluctuation and $(c\tau_k/2)^2$ models the path-loss, for $k = 1, \dots, K$; finally, we define the average (over both $\bar{a}_k$ and τ_k) SNR of the k-th target across the exploited sub-bands as

$$\mathsf{SNR}_k = \frac{\mathrm{E}[|a_k|^2]}{\sigma_w^2 L_\tau} \mathbf{a}_{\mathsf{J}}(\mathsf{f}_k)^\dagger \left(\sum_{\ell=1}^{L_\tau} \mathbf{B}^{\ell\dagger} \mathbf{B}^\ell \right) \mathbf{a}_{\mathsf{J}}(\mathsf{f}_k) \tag{5.25}$$

which is assumed to be equal for all targets, *i.e.* $\mathsf{SNR}_1 = \cdots = \mathsf{SNR}_K = \mathsf{SNR}$.

5.3.1 Benchmarks

We evaluate the theoretical results presented in chapter 2 [24] applied to our radar scenario, and compare it with the methods in [22] and [63]. Prior to [24] results, the technique in [22] was the state-of-the-art solution for gridless multidimensional frequency recovery; specifically, we present here the performance of the solution in [22, Eq. 53], which we reformulate using a convex combination of [22, Sec. IV.C, Eq. 45] and $\|\mathbf{y}-\mathbf{B}\mathbf{x}\|_2^2$ to handle the noise presence. The IIC-AMF in [63] extracts the prospective targets one-by-one from the noisy measurement in (5.17). While the algorithm was originally presented with reference to a 2D measurement space involving the delay and Doppler dimensions, it has been extended here to handle the additional spatial dimensions and the multiple sub-bands. The implementation complexity of IIC-AMF is tied to the adopted dictionary length named as L_D and given by $O\left(K(L_G J^2 + J^3 + 5J^2 + 2L_G)\right)$. In Sec. 5.3.3 a uniform grid with $L_D = 940$ elements is used to maintain the same implementation complexity required to solve either (P.6) or (P.7), while in Sec. 5.3.4 a uniform grid with $L_D = 10.000$ elements is used, leading to a implementation complexity larger than that required to solve (P.6) or (P.7).

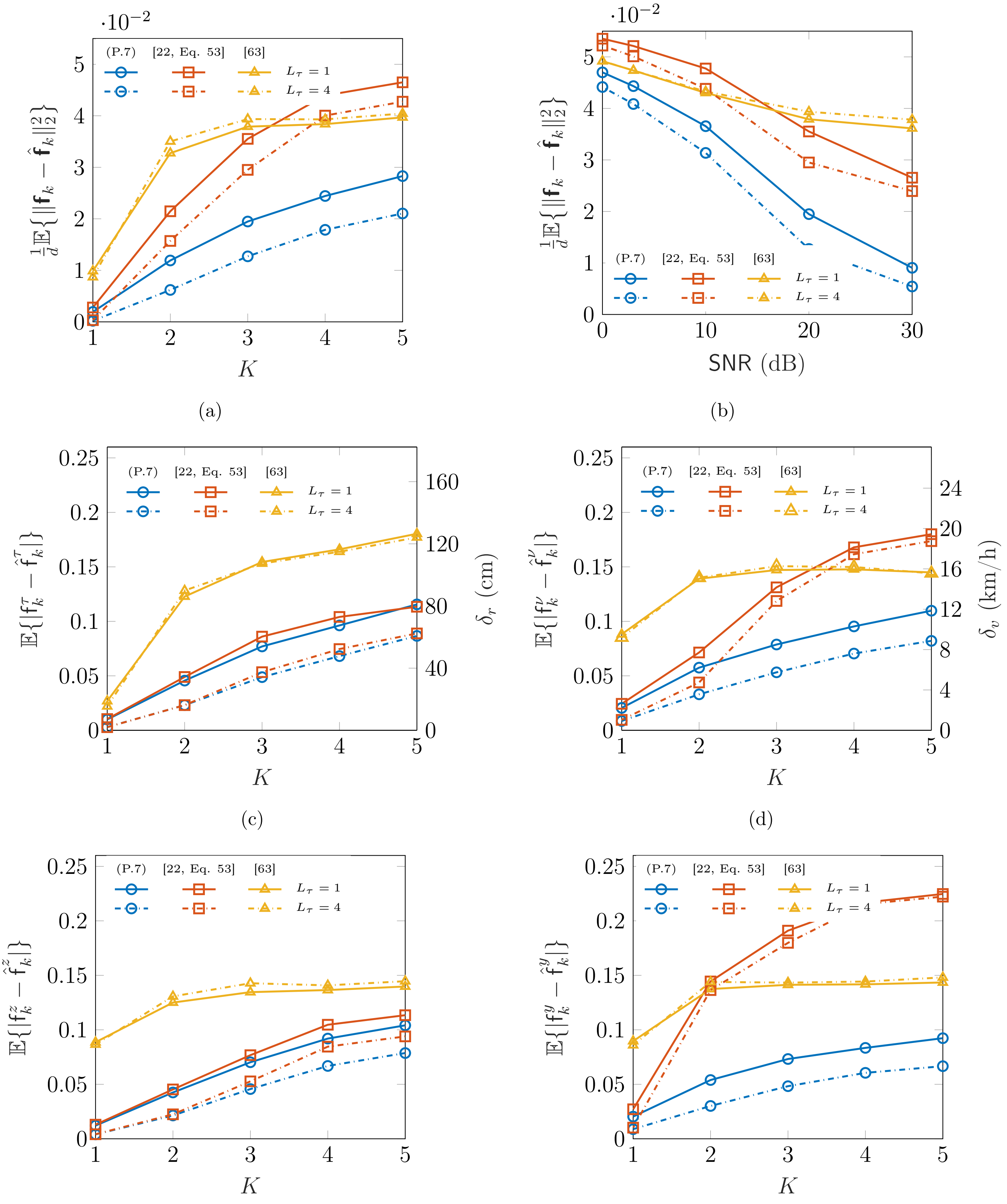
In the implementation of all algorithms, the number K of prospective targets is assumed known, by using the methodology proposed in Chapter 4.

5.3.2 Optimization of η

In the proposed optimization problems, the parameter η can be chosen in order to minimize the recovery error. Our numerical simulations indicate that the optimal η depends upon the SNR, the number of sub-bands L_τ, and on the structure of the composite steering vector specified by the vector J. Since L_τ and J are known a priory, the value of the η can be optimized off-line, at least for a nominal design SNR. Fig. 5.4 reports the MSE in the estimation of entire frequency vector, defined as $\frac{1}{d}\mathbb{E}\{\|\mathsf{f}_k - \hat{\mathsf{f}}_k\|_2^2\}$, versus η for $[\mathsf{J}_x, \mathsf{J}_y, \mathsf{J}_\nu, \mathsf{J}_z, \mathsf{J}_\tau]^\top = [1,2,2,4,6]^\top$, $K = \{2,3\}$, $L_\tau = \{1,2,4,8\}$, and $\mathsf{SNR} = \{0,10,20\}$ dB. The optimal η decreases as SNR or L_τ is increased; also, it is seen by inspection that, for a given SNR and L_τ, the optimal η does not depend much on the number of targets K to be resolved.

5.3.3 Estimation performance vs K, SNR, and L_τ

Upon choosing the optimal η, we assess here the effects of the parameters K, SNR, and L_τ on the estimation performance, when $[\mathsf{J}_x, \mathsf{J}_y, \mathsf{J}_\nu, \mathsf{J}_z, \mathsf{J}_\tau]^\top = [1,2,2,4,6]^\top$ and $L_\tau = \{1,4\}$. Figs. 5.5(a) and 5.5(b) show the MSE in the estimation of entire frequency vector versus K when $\mathsf{SNR} = 20$ dB and versus SNR when $K = 3$, respectively. Next, Figs. 5.5(c)- 5.5(f) show the MAE in the estimation of the



(a) (b)

(c) (d)

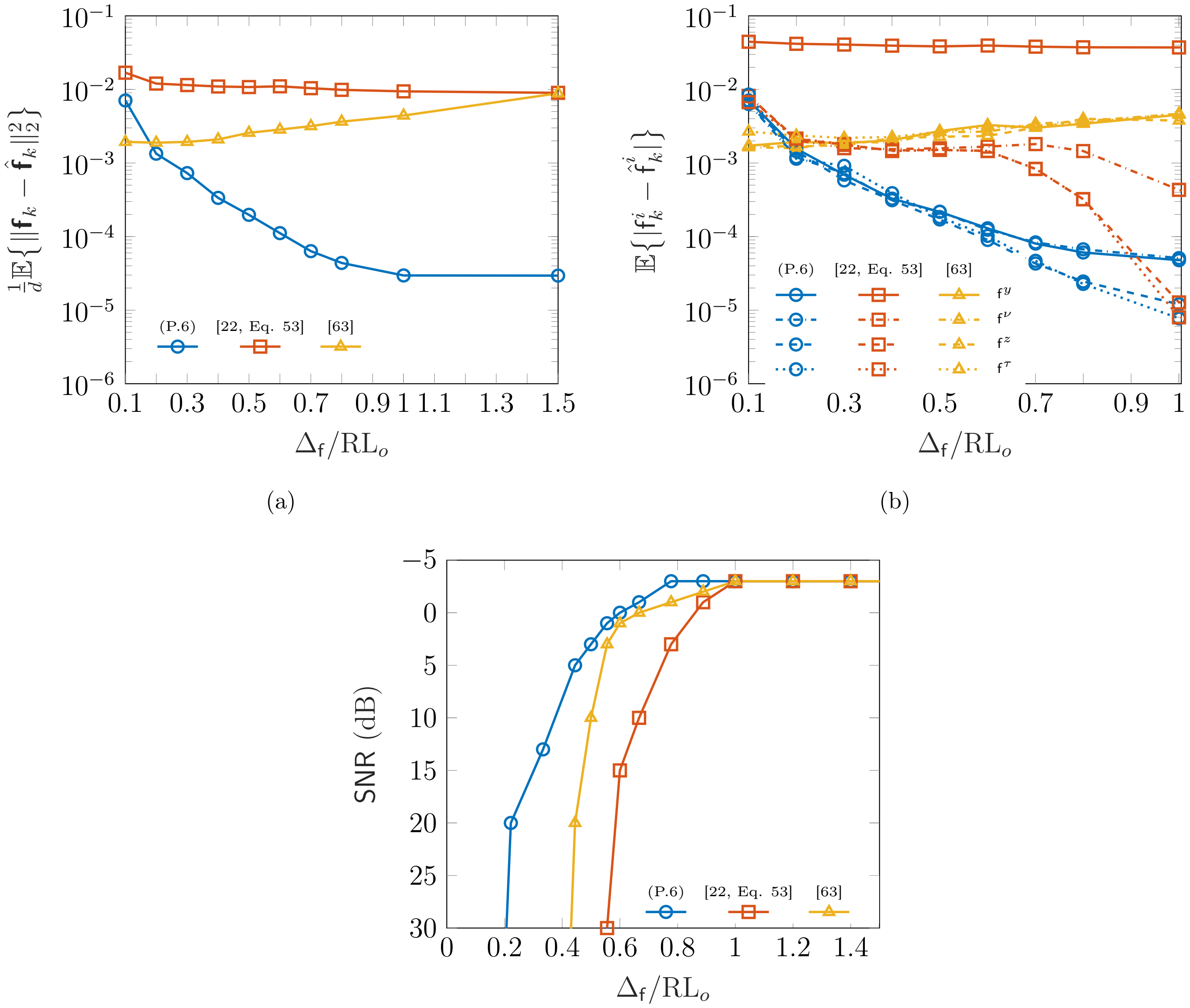
(e) (f)

Fig. 5.5. Analysis of the estimation error performance when $[\mathsf{J}_x, \mathsf{J}_y, \mathsf{J}_\nu, \mathsf{J}_z, \mathsf{J}_\tau]^\top = [1, 2, 2, 4, 6]^\top$, $L_\tau = \{1, 4\}$, and η is optimally designed. (a) MSE in the estimation of entire frequency vector vs K for $\mathsf{SNR} = 20$ dB. (b) MSE in the estimation of entire frequency vector vs SNR for $K = 3$. (c-f) MAE in the estimation of the scalar frequency along each non-singleton dimension $i \in \{\tau, \nu, z, y\}$, respectively, vs K for $\mathsf{SNR} = 20$ dB.

scalar frequency along each non-singleton dimension $i \in \{\tau, \nu, z, y\}$, respectively, defined as $\mathbb{E}\{|\mathsf{f}_k^i - \hat{\mathsf{f}}_k^i|\}$. For the reader's sake, Fig. 5.5(c) also reports on the right vertical axis the MAE in the estimation of the target range, *i.e.* $\delta_r = \frac{c_o T_w}{2}\mathbb{E}\{|\mathsf{f}_k^\tau - \hat{\mathsf{f}^\tau}_k|\}$, while Fig. 5.5(d) the MAE in the estimation of the target radial velocity, *i.e.* $\delta_v = \frac{c_o}{2 f_o T}\mathbb{E}\{|\mathsf{f}_k^\nu - \hat{\mathsf{f}^\nu}_k|\}$. As expected, the estimation error increases with K, while decreases with SNR and L_τ. Different performance/complexity trade-offs can be achieved by changing L_τ. Finally, it is verified by inspection that our solution outperforms both the gridless approach in [22] and the IIC-AMF [63].

5.3.4 Estimation performance vs the target separation

Upon choosing the optimal η, we simulate here $K = 2$ targets with $\mathbf{f}_1 \in \mathbb{T}^d$ and $\mathbf{f}_2 = \mathbf{f}_1 + [\pm\Delta_{\mathsf{f}}, \dots, \pm\Delta_{\mathsf{f}}]^\top$ and evaluate the estimation performance as a function of Δ_{f}; we assume here $[\mathsf{J}_x, \mathsf{J}_y, \mathsf{J}_\nu, \mathsf{J}_z, \mathsf{J}_\tau]^\top = [1, 2, 2, 4, 6]^\top$ and $L_\tau = 1$. A key parameter in classical resolution theory is the RL, considered to be the minimum resolvable separation of two objects with equal intensities [68], [69] along a given dimension. Since we have here a different number of samples along the considered dimensions, we take as a reference unit, say RL_o, the RL along the τ-axis (which is the smallest RL among the involved non-singleton dimensions), namely, $\mathrm{RL}_o = 1/\mathsf{J}_\tau = 1/6$, and vary Δ_{f} from $0.1/\mathrm{RL}_o$ to $1.5/\mathrm{RL}_o$. The corresponding results are shown in Fig. 5.6. Specifically, Fig. 5.6(a) shows the MSE in the estimation of entire frequency vector for $\mathsf{SNR} = 20$dB, while Fig. 5.6(b) shows the MAE in the estimation of the scalar frequency along each non-singleton dimension $i \in \{y, z, \nu, \tau\}$. Finally, Fig. 5.6(c) shows the minimum SNR ensuring a successful recovery of the frequency vectors (also referred to as the phase transition curve in [69]); following [69], we declare here a successful recovery if $\frac{1}{d}\mathbb{E}\left\{\|\mathbf{f}_k - \hat{\mathbf{f}}_k\|_1\right\} < \frac{\Delta_{\mathsf{f}}}{2}$. Again, the proposed solution always outperforms the considered benchmarks.



(a)

(b)

(c)

Fig. 5.6. MSE in the estimation of entire frequency vector (a), MAE in the estimation of each scalar frequency (b), and the minimum SNR ensuring a successful recovery of the frequency vectors (c) versus Δ_f/RL_o when $K = 2$, $[\mathsf{J}_x, \mathsf{J}_y, \mathsf{J}_\nu, \mathsf{J}_z, \mathsf{J}_\tau]^\top = [1, 2, 2, 4, 6]^\top$, and η is optimally designed. In (a-b), it is assumed $\mathsf{SNR} = 20$ dB.

Chapter 6

Low complexity hybrid precoding strategies for MIMO in sparse mmWave channels

mmWave band usage for 5G-6G presents many challenges, as discussed in chapter 3, some of which are recalled next. First, this band experiences significant free–space path loss due to the ten-fold increase in carrier frequency. Fortunately, this increase in carrier frequency implies a decrease in wavelength, which facilitates the deployment of large antenna arrays in small areas that can take advantage of the well-known multiplexing and diversity gain of MIMO to mitigate path loss. A second important challenge, very relevant for this chapter, relates to mmWave hardware design, a subject of significant research since the 1970s. While the first mmWave system implementations were based on gallium arsenide (GaAs), the improvement that CMOS technology brought to traditional microwave systems has recently led researchers and industry practitioners to develop CMOS subsystems for mmWave systems. Nonetheless, mmWave system design based on CMOS technology is still under ongoing investigation, as mmWave CMOS is not easily extrapolated from microwave CMOS [70]. A last challenge to take into consideration is the cost of mmWave hardware (low noise amplifiers, power amplifiers, antennas, etc.), mainly due to its reduced size and low power consumption requirements [38], which creates the need for the design of low complexity MIMO communication solutions in these bands. With MIMO becoming a key technology for 5G due to its improved spectral efficiency and diversity gains, it is important to address the specific challenges of implementing MIMO, or even massive MIMO, in the mmWave band.

In this chapter, we focus on the design of a low complexity precoder in a multiuser MIMO (MU-MIMO) downlink (DL) scenario. Under conventional microwave propagation, MIMO precoding can be easily implemented digitally at baseband (BB), and requires dedicated Radio Frequency (RF) hardware, *i.e.* an RF chain for each antenna element. This approach is named fully-digital (FD) implementation. In mmWave bands we have two circumstances that refrain from using one dedicated RF chain for each antenna element, which are *i)* typically antenna deployments in mmWave are very large, precisely to overcome the challenging propagation scenario and *ii)* RF hardware for mmWave wavelengths is more expensive and power consuming than for conventional microwave wavelengths. Hybrid (HB) precoding divides the precoding implementation into digital BB processing, bandpass modulation and analog processing, allowing a potential reduction on the number of RF chains and thus enabling a viable precoding architecture for mmWave systems.

In this chapter, by reformulating the hybrid precoder design as a matrix factorization problem, and adopting the atomic norm minimization approach from chapter 2, we propose a new hybrid precoding algorithm that takes advantage of the sparse nature of the mmWave channel and that it is able to closely approach the performance of the optimal fully-digital precoder.

The results presented in this chapter are an extension of the results presented in my Master Thesis [71] by addressing the following novel challenges:

- Gridless recovery of parameters: compared to the results in [71] where an on-grid approach was used for parameter retrieval, here we propose to gridlessly resolve the recovery of the relevant channel parameters that allow for the design of hybrid precoders.

- Multidimensional parameter recovery compared to the one-dimensional recovery proposed in [71]. This scenario indeed addresses transmission strategies for future wireless communication systems where the signal processing in the base station needs to be aware of the 3D positioning of the user equipment.

6.1 System description

6.1.1 Propagation in MU-MIMO mmWave sparse channels

Consider a MU-MIMO propagation scenario in the downlink (DL) with a d-D antenna deployment at the base station (BS) with M antennas, *i.e.* $\mathbf{M} = [\mathsf{M}_1, \ldots, \mathsf{M}_d]^\top$, being M_i the number of antennas in dimension $i \in [d]$, and K single-antenna users at the receiver side, *i.e.* $\mathbf{N} = [1]^\top$ for every user. Due to the nature of mmWave propagation, this can be modeled as a sparse channel matrix $\mathbf{H} \in \mathbb{C}^{K \times M}$ as formulated in (3.9). In the MU-MIMO scenario we assume line of sight (LOS) and thus, only K propagation paths are present, one per user. Also, the steering vectors matrix at the receiver, $\mathbf{R}_{\mathbf{N}}(\mathbf{f}_{1:K})$ is considered an identity matrix $\mathbf{I}_{K \times K}$ since each user only has an antenna and thus only sees its frequency given that the single-antenna users are not co-located. Thus, channel matrix in (3.9) can be further reformulated as

$$\begin{aligned}
\mathbf{H} &= \mathbf{R}_{\mathsf{N}}(\mathbf{f}_{1:K})\mathbf{G}\mathbf{T}^{\dagger}_{\mathsf{M}}(\mathbf{g}_{1:K})\mathbf{B}^{\dagger}_{t} \\
&= \mathbf{G}\mathbf{T}^{\dagger}_{\mathsf{M}}(\mathbf{g}_{1:K})\mathbf{B}^{\dagger}_{t} \\
&= \begin{bmatrix} \gamma_1 \mathbf{t}^{\dagger}_{\mathsf{M}}(\mathbf{g}_1) \\ \vdots \\ \gamma_K \mathbf{t}^{\dagger}_{\mathsf{M}}(\mathbf{g}_K) \end{bmatrix} \mathbf{B}^{\dagger}_{t} \\
&= [\mathbf{h}_1, \dots, \mathbf{h}_K]^{\dagger}
\end{aligned} \tag{6.1}$$

where, if a uniform array is considered at the BS, sampling matrix $\mathbf{B}_t$ is an identity matrix, $\mathbf{I}_{M\times M}$.

6.1.2 Transmission strategies

In this scenario, the definition of the transmission strategy is the real and unique challenge, given the assumption that users do not cooperate. Indeed, the best solution to overcome the co-channel interference, and eventually provide each user with reasonable performance is to preprocess, with a proper transmission strategy, the signal intended for each user at the BS.

Many transmission strategies have been studied in the literature. The DL MU-MIMO sum-rate, defined as the maximum aggregation of all the users' data rates is achieved via dirty paper coding (DPC), a highly complex transmission strategy for which finding practical dirty paper codes that approach the capacity limit is still unknown. However, linear precoding can achieve the same multiplexing gain as DPC, with a certain offset with respect to the sum-rate performance [72]. Thus, a linear precoder would be able to transmit as many data threads as the DPC, while requiring a much less complex transmission implementation [73]. All transmission strategies rely on a certain level of channel knowledge at the BS. The best scenario is characterized by full channel knowledge at the transmitter, also referred to as Channel State Information (CSI) at the transmitter (CSIT). Other more realistic assumptions provide the BS with partial information of the channel (PCSI). Examples include quantized information, delayed information, or even long-term variation (statistics). These scenarios differ in performance, with CSI outperforming PCSI, and also in complexity, with CSI being more complex to acquire than PCSI [74]. For the rest of this chapter, we will focus on linear precoders with CSIT.

Linear transmission strategies

Linear precoding strategies provide implementation advantages compared to non-linear techniques, allowing a trade-off between complexity and performance. The linear precoder implements a transformation of the K threads of information into

Linear precoder	Design criteria	Formulation
Conjugate beamforming	For large values of M is equivalent to a zero–forcing approach	$\mathbf{V}^{\mathrm{CB}} = \mathbf{H}^{\dagger}$
Zero–forcing	Eliminates spatial interference	$\mathbf{V}^{\mathrm{ZF}} = \mathbf{H}^{\dagger}\left(\mathbf{H}\mathbf{H}^{\dagger}\right)^{-1}$
MMSE	Minimizes the mean squared error of each user	$\mathbf{V}^{\mathrm{MMSE}} = \mathbf{H}^{\dagger}\left(\mathbf{C}\mathbf{H}\mathbf{H}^{\dagger} + \sigma_w^2\mathbf{I}\right)^{-1}\mathbf{C}\mathbf{B}^{-1}$

TABLE 6.1. EXAMPLE OF LINEAR PRECODERS WITH THEIR CORRESPONDING DESIGN CRITERIA.

the M transmitted symbols, which is linearly modeled in its low-pass equivalent by a precoding matrix named $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_K] \in \mathbb{C}^{M\times K}$, being $\mathbf{v}_k$ the transformation applied to user k's signal and to be transmitted through the M antennas.

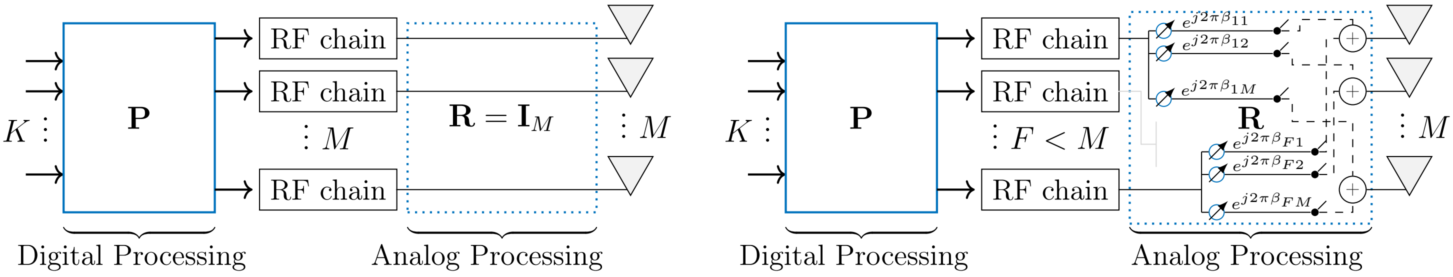
When designing a linear precoder, an optimization criteria should be determined and applied according to the level of channel knowledge at the transmitter. Focusing on full CSI, where the BS has full access to $\mathbf{H}$, the linear precoder could be chosen to optimize *i)* the sum-rate, *ii)* the signal to interference and noise ratio (SINR), *iii)* the mean squared error (MSE), or *iv)* it could even be proposed to eliminate the inter-user interference (block-diagonalization), or the spatial interference (zero-forcing, or conjugate beamforming).[3] Some examples of linear precoders, together with the optimized performance criteria, are shown in Table 6.1. Among the several linear precoders, in the following we will focus on the MSE precoder as the reference linear precoder for the proposed hybrid implementation and its performance.

Building linear precoders

Any linear precoding strategy has two particular elements. The first one is digital processing, where transformations are done in baseband (BB) to the discrete symbols. The second one is analog processing, where the discrete symbols are transformed and modulated in band-pass, see Fig. 6.1. These two stages, digital and analog processing, can be modeled within the low-pass equivalent linear precoding matrix as $\mathbf{V} = \mathbf{R}\mathbf{P} \in \mathbb{C}^{M\times K}$, where matrix $\mathbf{P} \in \mathbb{C}^{F\times K}$ represents the BB processing and has no predefined structure and $\mathbf{R} \in \mathbb{C}^{M\times F}$ implements the analog processing; $\mathbf{P}$ will be referred to as BB matrix and $\mathbf{R}$ as RF matrix. F is the number of RF chains implemented in the band-pass modulation. Thus, RF matrix $\mathbf{R} = [\mathbf{r}_1, \ldots, \mathbf{r}_F]$ consists of a set of phase shifters, switches and mixers where the vector $\mathbf{r}_l$ represents the input phase shift to each of the M transmit antennas from the l-th RF chain. We represent this phase shift as $r_{ml} = \frac{1}{\sqrt{M}}e^{j2\pi\beta_{ml}}$, where the phase $\beta_{ml} \in \mathbb{T}$ may take

---

[3] It should be noted that when the users have one single antenna, both block-diagonalization and zero-forcing beamforming provide the same performance.



(a) (b)

Fig. 6.1. Block diagram of the processing to be implemented in the BS: (a) Digital and analog processing for a FD implementation of the transmission strategy; (b) Digital and analog processing for a HB implementation of the transmission strategy.

different values depending on the precoding scenario that will be discussed next:

- If $F = M$ typically $r_{ml} = \frac{1}{\sqrt{M}}$ if $m = l$ and zero otherwise.
- If $F < M$, then $r_{ml} = \frac{1}{\sqrt{M}} e^{j2\pi \mathbf{m}_m^\top \mathbf{g}_l}$, with $\mathbf{m}_m = \left[\mathsf{m}_m^1, \ldots, \mathsf{m}_m^d\right]^\top$ being the normalized position of the antenna elements of the transmit array, $\beta_{ml} \in \mathbb{T}$, $\mathbf{g}_l \in \mathbb{T}^d$ for $m \in [M]$ as in (3.5) and $l \in [F]$.

Finally, the BB matrix $\mathbf{P}$ is normalized so that the columns of $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_K]$, being $\mathbf{v}_k$ the beamforming vector for the information of the k-th user, are $||\mathbf{v}_k|| = 1$.

The hardware/software design of the system will depend on the choice of F and on how the split in the precoding processing is done, *i.e.* digital and analog, given that digital processing is implemented in software and analog processing is performed at hardware. The scenario known as fully-digital (FD) has $F = M$, *i.e.* one RF chain for each transmit antenna yielding a square RF matrix $\mathbf{R}$. In this case, matrix $\mathbf{R}$ is typically forced to be an identity matrix with some type of normalization, given that any phase shift or mixing performed in a square matrix can be done digitally at baseband by matrix $\mathbf{P}$, being it the reason why this scenario is called fully-digital (FD) (depicted in Fig. 6.1(a)). Conversely, we could also decide to reduce the number of RF chains making $F < M$, and in this scenario the tall $\mathbf{R}$ matrix would typically have a non-diagonal structure, allowing a certain number of mixers and phase shifters to be active in order to implement the analog processing (see Fig. 6.1(b)). In this case, the transmitter implements digital and analog processing together with the band-pass modulation. This hardware/software implementation is said to be hybrid.

In the following, we discuss the most adequate hardware/software implementation of the transmission strategy: fully-digital or hybrid. The answer is obviously not universal and it will depend on several factors, such as performance, cost of hardware for analog processing and band-pass modulators, and space issues for the deployment of a certain number of RF chains, among others. Most of the precoders in the literature for DL MU-MIMO provide a structure for the precoding matrix $\mathbf{V}$ that, apparently, and without digging into any suitable potential decomposition, would only match the FD implementation (see examples in Table 6.1). In this case, a hybrid implementation with $F < M$ would not be feasible without compromising performance unless we are able to leverage a suitable internal decomposition of the precoder matrix that matches the hybrid implementation. Assuming performance is not an issue, either because there is an internal decomposition of $\mathbf{V}$ that allows a hybrid implementation, or because the performance loss can be compensated with other benefits of the hardware implementation (*e.g.* reduced cost, reduced complexity), there is still scope for discussion on when it is suitable to reduce the number of RF chains compared to the number of antennas at the BS. In any system, reducing the number of RF chains would force analog processing, with the increase in cost that this hardware implementation would incur. However, since the number of RF chains are being reduced, there is a trade-off between the additional cost due to analog processing and the cost reduction due to the smaller number of RF chains. In this scenario, there is typically a net reduction of the hardware cost (RF chains are more expensive than the analog processing hardware required).

A remaining issue to address when deciding between a FD or HB transmitter is the role of the baseband processing. Typical system parametrization in MU-MIMO, and specially in mmWave channels, assume $M \gg K$, *i.e.* the number of antennas at the BS should be larger than the number of information threads K to be transmitted. In this case, we could still face two scenarios, $F \geq K$ or $F < K$. In the first scenario, the baseband processing is performed by a tall or square matrix, ensuring that we are providing a level of redundancy that helps overcoming the channel impairments and allows for an accurate estimate of the transmitted thread. However, the second scenario is not as straightforward since the baseband processing matrix would be fat and therefore the thread information is being compressed before the analog processing. In this scenario, we should ensure that the analog processing overcomes this compressing loss, which in many cases may not be possible due to the spatial multiplexing gain reduction [74].

Next, we depict the downlink (DL) and uplink (UL) scenario in a MU-MIMO system. From that, we unveil the hybrid structure inherent in the MMSE precoder.

6.1.3 DL scenario

The full DL system is depcited in Fig. 6.2 where a transmit information vector with unit-energy, i.i.d. elements $\mathbf{s} = [s_1, \ldots, s_K]^\top \in \mathbb{C}^K$ is scaled, prior to transmission, by means of a matrix $\mathsf{\mathbf{Q}}^{1/2} = \operatorname{diag}(\mathsf{\mathbf{q}}^{1/2}) \in \mathbb{R}^{K\times K}$, where $\mathsf{\mathbf{q}} = [\mathsf{q}_1, \ldots, \mathsf{q}_K]^\top \in \mathbb{R}^K$ contains the energy of each information symbol. Then the information vector is preprocessed by the precoder matrix $\mathbf{V} \in \mathbb{C}^{M\times K}$, obtaining the signal vector $\mathbf{x} =$

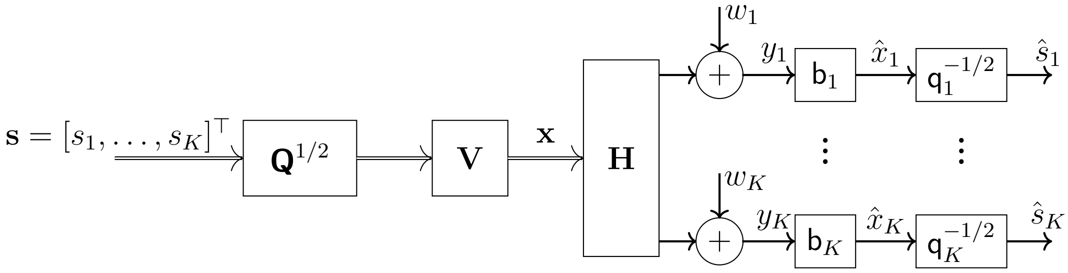


Fig. 6.2. Downlink MU-MIMO system model.

$\mathbf{V}\mathsf{Q}^{1/2}\mathbf{s} \in \mathbb{C}^M$. Finally, $\mathbf{x}$ is transmitted through a flat fading DL MIMO channel $\mathbf{H} \in \mathbb{C}^{K\times M}$ as in (6.1). The received signal is then given by

$$\begin{aligned} \hat{\mathbf{s}} =& \mathsf{Q}^{-1/2}\mathsf{B}\left(\mathbf{H}\mathbf{x} + \mathbf{w}\right) \\ =& \mathsf{Q}^{-1/2}\mathsf{B}\left(\mathbf{H}\mathbf{V}\mathsf{Q}^{1/2}\mathbf{s} + \mathbf{w}\right), \end{aligned} \tag{6.2}$$

where $\mathsf{B} = \operatorname{diag}(\mathsf{b}) \in \mathbb{R}^{K\times K}$ is a diagonal matrix that scales each user received signal. Both vectors $\mathsf{q} \in \mathbb{R}^K$ and $\mathsf{b} \in \mathbb{R}^K$ are used to achieve downlink/uplink (DL/UL) duality [75], [76], that would be needed to obtain the optimal Mean Square Error (MSE) precoder and also precoder normalization. Finally, $\mathbf{w} \in \mathbb{C}^K$ is an additive white circular-symmetric complex Gaussian noise vector with independent zero-mean components and σ_w^2-variance. Thus, the signal $\hat{s}_k$ received by the k-th user, with $k \in [K]$, is then formulated as follows:

$$\begin{aligned} \hat{s}_k &= \mathsf{q}_k^{-1/2}\mathsf{b}_k\mathbf{h}_k^\dagger\mathbf{V}\mathsf{Q}^{1/2}\mathbf{s} + \mathsf{q}_k^{-1/2}\mathsf{b}_k w_k \\ &= \mathsf{q}_k^{-1/2}\mathsf{b}_k\sum_{i=1}^{K}\mathsf{q}_i^{1/2}\mathbf{h}_k^\dagger\mathbf{v}_i s_i + \mathsf{q}_k^{-1/2}\mathsf{b}_k w_k \\ &= \underbrace{\mathsf{b}_k\mathbf{h}_k^\dagger\mathbf{v}_k s_k}_{\text{desired symbol}} + \underbrace{\mathsf{q}_k^{-1/2}\mathsf{b}_k\sum_{\substack{i=1\\ i\neq k}}^{K}\mathsf{q}_i^{1/2}\mathbf{h}_k^\dagger\mathbf{v}_i s_i}_{\text{multiuser interference}} + \underbrace{\mathsf{q}_k^{-1/2}\mathsf{b}_k w_k}_{\text{noise}}. \end{aligned} \tag{6.3}$$

Therefore assuming uncorrelated symbols, the Signal to Interference and Noise Ratio (SINR) of the k-th user in the DL, $\mathsf{SINR}_k^{\mathrm{DL}}$, is given by:

$$\begin{aligned} \mathsf{SINR}_k^{\mathrm{DL}} &= \frac{\mathsf{q}_k\mathbf{h}_k^\dagger\mathbf{v}_k\mathbf{v}_k^\dagger\mathbf{h}_k}{\sum_{\substack{i=1\\ i\neq k}}^{K}\mathsf{q}_i\mathbf{h}_k^\dagger\mathbf{v}_i\mathbf{v}_i^\dagger\mathbf{h}_k + \sigma_w^2} \\ &= \frac{\mathsf{q}_k||\mathbf{v}_k^\dagger\mathbf{h}_k||^2}{\sum_{\substack{i=1\\ i\neq k}}^{K}\mathsf{q}_i||\mathbf{v}_i^\dagger\mathbf{h}_k||^2 + \sigma_w^2}. \end{aligned} \tag{6.4}$$

And the k-th user's MSE is given by:

$$\begin{aligned}\epsilon_k^{\mathrm{DL}} &= \mathbb{E}\left\{\|\hat{s}_k - s_k\|_2^2\right\} \\ &= 1 - \mathsf{b}_k \mathbf{v}_k^\dagger \mathbf{h}_k - \mathsf{b}_k \mathbf{h}_k^\dagger \mathbf{v}_k + \frac{\mathsf{b}_k^2}{\mathsf{q}_k}\mathbf{h}_k^\dagger \mathbf{V}\mathbf{Q}\mathbf{V}^\dagger \mathbf{h}_k + \frac{\mathsf{b}_k^2}{\mathsf{q}_k}\sigma_w^2.\end{aligned} \tag{6.5}$$

It should be noted that the design of the precoder $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_K]$ under a MSE criteria minimization, $\min_{\mathbf{v}_k} \epsilon_k^{\mathrm{DL}}$, leads to a coupled solution for each of the user beamformers. For this reason, as it is shown in the next sections 6.1.4 and 6.1.5, we propose to apply DL/UL duality [75], [76] to obtain the optimal receiver in the UL and further get the equivalent precoder in the DL that achieves the same MSE region.

6.1.4 UL scenario

As stated earlier, we consider the precoder that minimizes MSE (see (6.5)) with a hybrid structure as $\mathbf{V} = \mathbf{RP}$ as the reference linear precoder. For that we apply DL/UL duality and assume Time Division Duplexing (TDD). Thus, the equivalent UL scenario is depicted in Fig. 6.3 where the k-th user transmits with power c_k and both $\mathbf{B}$ and $\mathbf{C} = \mathrm{diag}([\mathsf{c}_1, \ldots, \mathsf{c}_K]) = \mathrm{diag}(\mathbf{c})$ are scaling matrices only affecting the MSE and not the SINR when applied at the receiver.

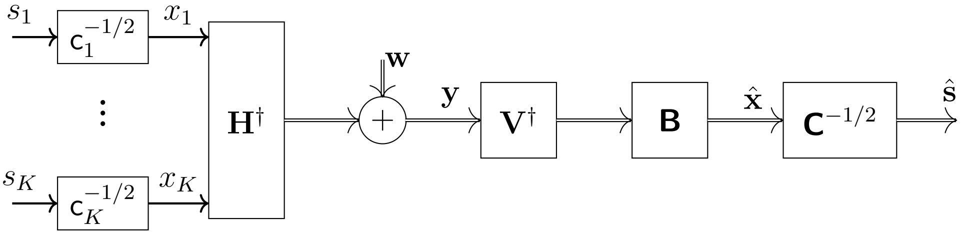


Fig. 6.3. Uplink MU-MIMO system model.

Thus, in the UL, we obtain the SINR of the k-th user as

$$\begin{aligned}\mathsf{SINR}_k^{\mathrm{UL}} &= \frac{\mathsf{c}_k \mathbf{v}_k^\dagger \mathbf{h}_k \mathbf{h}_k^\dagger \mathbf{v}_k}{\sum_{\substack{i=1\\ i\neq k}}^{K} \mathsf{c}_i \mathbf{v}_k^\dagger \mathbf{h}_i \mathbf{h}_i^\dagger \mathbf{v}_k + \sigma_w^2 \mathbf{v}_k^\dagger \mathbf{v}_k} \\ &= \frac{\mathsf{c}_k ||\mathbf{v}_k^\dagger \mathbf{h}_k||^2}{\sum_{\substack{i=1\\ i\neq k}}^{K} \mathsf{c}_i ||\mathbf{v}_k^\dagger \mathbf{h}_i||^2 + \sigma_w^2 ||\mathbf{v}_k||^2},\end{aligned} \tag{6.6}$$

and the k-th user's MSE is given by

$$\epsilon_k^{\mathrm{UL}} = 1 - \mathsf{b}_k \mathbf{h}_k^\dagger \mathbf{v}_k - \mathsf{b}_k \mathbf{v}_k^\dagger \mathbf{h}_k + \frac{\mathsf{b}_k^2}{\mathsf{c}_k}\mathbf{v}_k^\dagger \mathbf{H}^\dagger \mathbf{C}\mathbf{H}\mathbf{v}_k + \frac{\mathsf{b}_k^2}{\mathsf{c}_k}\sigma_w^2 \mathbf{v}_k^\dagger \mathbf{v}_k. \tag{6.7}$$

6.1.5 Precoder Design

The receiver for a Minimum Mean Square Error (MMSE) criteria in the UL, for fixed $\mathsf{\mathbf{C}}$ and $\mathsf{\mathbf{B}}$ matrices, is obtained by solving arg $\min_{\mathbf{v}_k} \epsilon_k^{\mathrm{UL}}$, *i.e.*

$$\frac{\partial \epsilon_k^{\mathrm{UL}}}{\partial \mathbf{v}_k^\dagger} = -\mathsf{b}_k \mathbf{h}_k + \frac{\mathsf{b}_k^2}{\mathsf{c}_k}\mathbf{H}^\dagger \mathsf{\mathbf{C}}\mathbf{H}\mathbf{v}_k + \frac{\mathsf{b}_k^2}{\mathsf{c}_k}\sigma_w^2 \mathbf{v}_k = 0, \tag{6.8}$$

which yields

$$\mathbf{v}_k^{\mathrm{MMSE}} = \frac{\mathsf{c}_k}{\mathsf{b}_k}\left(\mathbf{H}^\dagger \mathsf{\mathbf{C}}\mathbf{H} + \sigma_w^2 \mathbf{I}\right)^{-1}\mathbf{h}_k. \tag{6.9}$$

The optimal MMSE receiver is then derived as

$$\begin{aligned}\mathbf{V}^{\mathrm{MMSE}} &= \left[\mathbf{v}_1^{\mathrm{MMSE}}, \dots, \mathbf{v}_K^{\mathrm{MMSE}}\right] \\ &= \left(\mathbf{H}^\dagger \mathsf{\mathbf{C}}\mathbf{H} + \sigma_w^2 \mathbf{I}\right)^{-1}\mathbf{H}^\dagger \mathsf{\mathbf{C}}\mathsf{\mathbf{B}}^{-1}\end{aligned} \tag{6.10}$$

and since $\mathsf{\mathbf{B}}$ can be any diagonal matrix we will choose it to achieve $||\mathbf{v}_k^{\mathrm{MMSE}}|| = 1$.

Given matrices $\mathsf{\mathbf{C}} \in \mathbb{R}^{K\times K}$ and $\mathsf{\mathbf{B}} \in \mathbb{R}^{K\times K}$, the SINR region, $\left\{\mathsf{SINR}_1^{\mathrm{UL}}, \dots, \mathsf{SINR}_K^{\mathrm{UL}}\right\}$, and MSE region, $\left\{\epsilon_1^{\mathrm{UL}}, \dots, \epsilon_K^{\mathrm{UL}}\right\}$, achieved in the UL with $\mathbf{V}^{\mathrm{MMSE}\dagger}$ as the precoder exactly match those in the DL, with $\mathbf{V}^{\mathrm{MMSE}}$ as the precoder and $\mathsf{\mathbf{q}}$ being the power allocation in the DL as [75], [76]:

$$\mathsf{\mathbf{q}} = \sigma_w^2 \mathsf{\mathbf{Q}}_{\mathrm{pw}}^{-1}\mathbf{1}_K, \tag{6.11}$$

with

$$\mathsf{\mathbf{Q}}_{\mathrm{pw}} = \begin{bmatrix} \frac{1}{\mathsf{SINR}_1^{\mathrm{UL}}}||\mathbf{v}_1^{\mathrm{MMSE}\dagger}\mathbf{h}_1||^2 & -||\mathbf{v}_2^{\mathrm{MMSE}\dagger}\mathbf{h}_1||^2 & \dots & -||\mathbf{v}_K^{\mathrm{MMSE}\dagger}\mathbf{h}_1||^2 \\ -||\mathbf{v}_1^{\mathrm{MMSE}\dagger}\mathbf{h}_2||^2 & \frac{1}{\mathsf{SINR}_2^{\mathrm{UL}}}||\mathbf{v}_2^{\mathrm{MMSE}\dagger}\mathbf{h}_2||^2 & \dots & -||\mathbf{v}_K^{\mathrm{MMSE}\dagger}\mathbf{h}_2||^2 \\ \vdots & & \ddots & \vdots \\ -||\mathbf{v}_1^{\mathrm{MMSE}\dagger}\mathbf{h}_K||^2 & -||\mathbf{v}_2^{\mathrm{MMSE}\dagger}\mathbf{h}_K||^2 & \dots & \frac{1}{\mathsf{SINR}_K^{\mathrm{UL}}}||\mathbf{v}_K^{\mathrm{MMSE}\dagger}\mathbf{h}_K||^2 \end{bmatrix}. \tag{6.12}$$

Furthermore if $\mathbf{V}$ and the SINR region for the UL/DL are known, the power allocation in for the UL is given by

$$\begin{aligned}\mathsf{\mathbf{c}} =& \sigma_w^2 \mathsf{\mathbf{Q}}_{\mathrm{pw}}^{-\top}\left[||\mathbf{v}_1^{\mathrm{MMSE}}||^2, \dots, ||\mathbf{v}_K^{\mathrm{MMSE}}||^2\right]^\top \\ =& \sigma_w^2 \mathsf{\mathbf{Q}}_{\mathrm{pw}}^{-\top}\mathbf{1}_K,\end{aligned} \tag{6.13}$$

where the last equality comes from the fact that the columns of matrix $\mathbf{V}$ are normalized.

Unveiling a hybrid structure in the full-digital MMSE precoder

Further developing (6.10) using the channel definition in (6.1) and applying matrix inversion identity we have that

$$\begin{aligned}\mathbf{V}^{\text{MMSE}} &= \left(\mathbf{H}^{\dagger}\mathbf{C}\mathbf{H} + \sigma_w^2\mathbf{I}\right)^{-1}\mathbf{H}^{\dagger}\mathbf{C}\mathbf{B}^{-1}\\ &= \sigma_w^{-2}\mathbf{H}^{\dagger}\left(\sigma_w^{-2}\mathbf{H}\mathbf{H}^{\dagger} + \mathbf{C}^{-1}\right)^{-1}\mathbf{B}^{-1}\\ &= \mathbf{H}^{\dagger}\left(\mathbf{C}\mathbf{H}\mathbf{H}^{\dagger} + \sigma_w^2\mathbf{I}\right)^{-1}\mathbf{C}\mathbf{B}^{-1}\\ &= \mathbf{B}_t\mathbf{T}_{\mathbf{M}}(\mathbf{g}_{1:K})\mathbf{G}^{\dagger}\mathbf{R}_{\mathbf{N}}^{\dagger}(\mathbf{f}_{1:K})\left(\mathbf{C}\mathbf{H}\mathbf{H}^{\dagger} + \sigma_w^2\mathbf{I}\right)^{-1}\mathbf{C}\mathbf{B}^{-1}\\ &= \mathbf{B}_t\mathbf{T}_{\mathbf{M}}(\mathbf{g}_{1:K})\mathbf{P}^{\text{MMSE}}\end{aligned} \tag{6.14}$$

The factorization form used for the MMSE precoder in (6.14) immediately reveals how to obtain a hybrid structure from a fully-digital implementation. The key aspect is to have knowledge of the matrix $\mathbf{B}_t\mathbf{T}_{\mathbf{M}}(\mathbf{g}_{1:K})$, or in other words, having knowledge of the frequency parameters at the transmitter $\{\mathbf{g}_1, \ldots, \mathbf{g}_K\}$, given that the structure of the atoms in $\mathbf{T}_{\mathbf{M}}(\mathbf{g}_{1:K})$ and the sampling matrix $\mathbf{B}_t$ are known. The consideration that identifying a proper precoder factorization allows a hybrid implementation of such precoder, can be in principle extended to other CSI based precoders, like, for example, all those precoders where $\mathbf{V} = \mathbf{H}^{\dagger}\mathbf{O}$, with $\mathbf{O}$ being any arbitrary matrix (*e.g.* zero–forcing and conjugate beamforming).

Mapping the factorization given in (6.14) with the decomposition in terms of $\mathbf{R}$ and $\mathbf{P}$ described in Section 6.1.2, it follows immediately that $\mathbf{R} = \mathbf{B}_t\mathbf{T}_{\mathbf{M}}(\mathbf{g}_{1:K})$. From this consideration, it follows that in a sparse MU-MIMO channel as defined in (6.1) with a small number of users ($K \ll M$), the MMSE precoder would potentially have a desirable hybrid structure with a number of RF chains much lower than the number of antennas, and consequently at a lower cost than a classical FD implementation while maintaining the same performance.

In fact, it should be noted that hybrid implementations that arise from this type of decomposition and from a potential knowledge of the transmission parameters $\{\mathbf{g}_1, \ldots, \mathbf{g}_K\}$ has the following characteristics. First, the system performance does not change, in the sense that whatever performance is achieved by the FD implementation (minimum MSE, interference cancellation, maximum SINR, etc.) will be achieved by the hybrid implementation. Second, the number of RF chains match the number of users seen at the transmitter, which in a sparse channel is low and would allow for a corresponding cost/complexity reduction compared to the FD implementation. Finally, note that, in general, no claim can be made regarding the fact that this is the minimum number of RF chains that guarantee the same performance achieved in the FD implementation. This will in general also depend on the rank of the matrix $\mathbf{H}$.

6.2 Solving the hybrid precoding using Atomic Norm

As stated earlier, the hybrid structure is feasible when setting $\mathbf{R} = \mathbf{B}_t \mathbf{T}_{\mathsf{M}}(\mathbf{g}_{1:K})$. Thus, assuming CSI at the transmitter, we have that for the k-th column of the known precoder matrix[4] $\mathbf{V}$, $\mathbf{v}_k = \mathbf{R}\mathbf{p}_k = \mathbf{B}_t \mathbf{T}_{\mathsf{M}}(\mathbf{g}_{1:K})\mathbf{p}_k = \mathbf{B}_t \sum_{i=1}^{K} \mathbf{t}_{\mathsf{M}}(\mathbf{g}_k) p_{ki}$ with $k \in [K]$. This representation unveils that the precoder is a binary measurement of a linear combination of atoms in $\mathcal{A}_{\mathsf{M}}$ as in (2.1). If we set $\mathbf{x}_k = \mathbf{R}\mathbf{p}_k$ we can pose the problem as $\mathbf{v}_k = \mathbf{B}_t \mathbf{x}_k$, where we have the same structure as in (2.4) by setting $\mathbf{w} = \mathbf{0}$. Note that all precoder vectors $\mathbf{v}_k$ with $k \in [K]$, are generated by means of the same steering matrix $\mathbf{T}_{\mathsf{M}}(\mathbf{g}_{1:K})$, and it is the weighting applied to the combination of the steering vectors, what provides the different $\mathbf{v}_k$. This is clearly a Multiple Measurement Vector (MMV) scenario equivalent to the one in (3.12) with $\mathbf{W} = \mathbf{0}$, where we have a set of measurements $[\mathbf{v}_1, \ldots, \mathbf{v}_K] = \mathbf{B}_t \mathbf{T}_{\mathsf{M}}(\mathbf{g}_{1:K})[\mathbf{p}_1, \ldots, \mathbf{p}_K]$.

Thence, we have two options to unveil the hybrid structure. One is looking at the problem from a Single Measurement Vector (SMV) perspective. In this case, the set of frequencies $\{\mathbf{g}_1, \ldots, \mathbf{g}_K\}$ can be found by solving the optimization problem (P.6) for any of the $\mathbf{v}_k$ with $k \in [K]$ observations. The second option is to look at the problem from a MMV perspective and recall here optimization (P.7), where a low complexity alternative was proposed splitting the data samples acquired in one dimension into multiple groups which altogether solved the optimization. Here, the problem can be formulated likewise, *i.e.* a set of sample vectors ($\mathbf{v}_k$, $k \in [K]$) are obtained and they all are binary measurements of linear combinations of atoms in $\mathcal{A}_{\mathsf{M}}$ built with the same frequency parameter set ($\mathbf{g}_{1:K}$). Only one last consideration, here we are in a noiseless scenario given that measurements come directly from this linear combination without any additive disturbance. Thus, we adapt (P.7) to this application, by setting the minimization cost function in ℓ_2 to zero due to the absence of noise and by adding a constraint that ensures that the solution matches the known precoding vectors:

$$\begin{aligned} \min_{t,\, \mathbf{x}_k,\, \mathbf{T}_{\mathsf{M}} \in \mathcal{T}} \quad & \frac{1}{2}\big(t + \mathrm{Tr}\left\{\mathbf{T}_{\mathsf{M}}\right\}\big) \\ \text{s.t.} \quad & \begin{bmatrix} \mathbf{T}_{\mathsf{M}} & \mathbf{x}_k \\ (\mathbf{x}_k)^{\dagger} & t \end{bmatrix} \succeq 0, \quad \mathbf{v}_k = \mathbf{B}_t \mathbf{x}_k \quad k \in [K] \end{aligned} \tag{P.8}$$

For this chapter, we will use the MMV perspective and unveil the hybrid precoding structure by using (P.8).

The complexity of (P.8), upon setting $n = M$ and $m = K$ in (2.9), is given by $O\left(KM^{3.5} + K^2 M^{2.5} + K^3\right)$.

Thence, the frequency set $\mathbf{g}_{1:K}$ is found by solving the Vandermonde decomposition of $\mathbf{T}_{\mathsf{M}}$, solution of (P.8). As stated earlier, once the frequency set is recovered,

---

[4] Note that we can either address the proper $\mathbf{V}^{\mathrm{MMSE}}$ as in (6.14), or any other precoding structure such that $\mathbf{V} = \mathbf{H}^{\dagger}\mathbf{O}$

RF matrix can be computed easily as $\mathbf{B}_t\mathbf{T}_{\mathsf{M}}(\mathbf{g}_{1:K})$. And finally, BB matrix, is calculated using a Least Squares solution, assuming CSIT and thus, having access to $\mathbf{V}^{\mathrm{MMSE}}$ as

$$\mathbf{P} = \left(\mathbf{R}^{\dagger}\mathbf{R}\right)^{-1}\mathbf{R}^{\dagger}\mathbf{V}^{\mathrm{MMSE}}. \tag{6.15}$$

With both matrices $\mathbf{R}$ and $\mathbf{P}$, the HB precoder matrix estimation is finally computed as $\mathbf{V} = \mathbf{RP}$.

6.3 Numerical analysis

We consider a DL MU-MIMO scenario with a BS transmitting with a uniform 18-antenna deployment distributed as $\mathsf{M} = [\mathsf{M}_1, \mathsf{M}_2]^{\top} = [3, 6]^{\top}$ with K single-antenna users. The number of users varies in the set $\{1, 2, 3, 4, 5\}$. Numerical results are showed next for the proposed method in (P.8) as well as two other methods, OMP and MUSIC introduced below, used for comparison.

In Fig. 6.4(a) we plot the error (MSE) in the recovery of the frequency set $\mathbf{g}_{1:K}$ versus K, where $\hat{\mathbf{g}}_k$ denotes the recovered frequency vector. In Fig. 6.4(b) we depict the error (MSE) of the HB precoder with respect to the MMSE optimal precoder, where we denote the $\hat{\mathbf{v}}_k$ as the HB precoder based on the recovery of the frequency parameters.

The rate achieved for every $K \in [5]$ for the proposed method and the two benchmarks is depicted in Fig. 6.5(a) with respect to the SNR. In Fig. 6.5(b) we depict the MSE of the received signal, $\mathbb{E}\{\|\mathbf{s} - \hat{\mathbf{s}}\|_2^2\}$ with respect to the SNR, compared to both benchmarks.

Computational complexity is adjusted in both benchmark methods so that it matches the one in (P.8) so that comparisons are fair. We can clearly see that the proposed method outperforms the benchmarks in all the results.

6.3.1 Benchmarks

Next, for comparison purposes, two alternative state-of-the-art methods for parameter recovery are proposed: MUSIC and OMP. All the results presented in this section will be compared with these two methods. The proposed alternative algorithms, unlike (P.8), rely on a predefined grid or dictionary of frequencies which is used to find the true frequencies. The length of the dictionary for each of the two methods is chosen so that computational complexity matches that of (P.8).

Orthogonal Matching Pursuit (OMP)

One of the first approaches to solve this factorization problem is given in [42] for a single–user MIMO scenario. In this case, the reference precoder maximizes the mutual information and it is given by the right-hand side eigenvectors of the channel matrix. To make feasible the search of the RF phases, they assume a predefined dictionary of length L_{OMP} with a grid on all the frequency space that characterize the possible position of the scatterers, and they choose the "best" F vectors within that dictionary by means of an orthogonal matching pursuit (OMP) algorithm. This work is further extended to the multi-user scenario [74] by optimizing, in two stages, the RF beamformers to maximize each user desired power, and the baseband beams to overcome the remaining multiuser interference. In this case, again, the search on the phases is not over all the phase space, but it is still brute force search over a discrete predefined phase dictionary whose cardinality depends on the discrete set of phases that the phase shifters can implement.

This algorithm uses K iterations to find K frequency vectors $\mathbf{g}_k$. In each iteration, one location in the dictionary is chosen so that it best matches the residual from the observation $\mathbf{y}$. Then this dictionary entry is added to the support estimate and removed from the dictionary, and the residual is updated. The procedure is repeated until the K entries are found.

Computational complexity of OMP is given by

$$O\Bigg(K\left(L_{\text{OMP}}^2 K + MK + L_{\text{OMP}}\right) + \sum_{i=1}^{K} i^3 + 2Mi^2 + 2MKi\Bigg) \tag{6.16}$$

where L_{OMP} is the length of the dictionary and d is the number of dimensions.

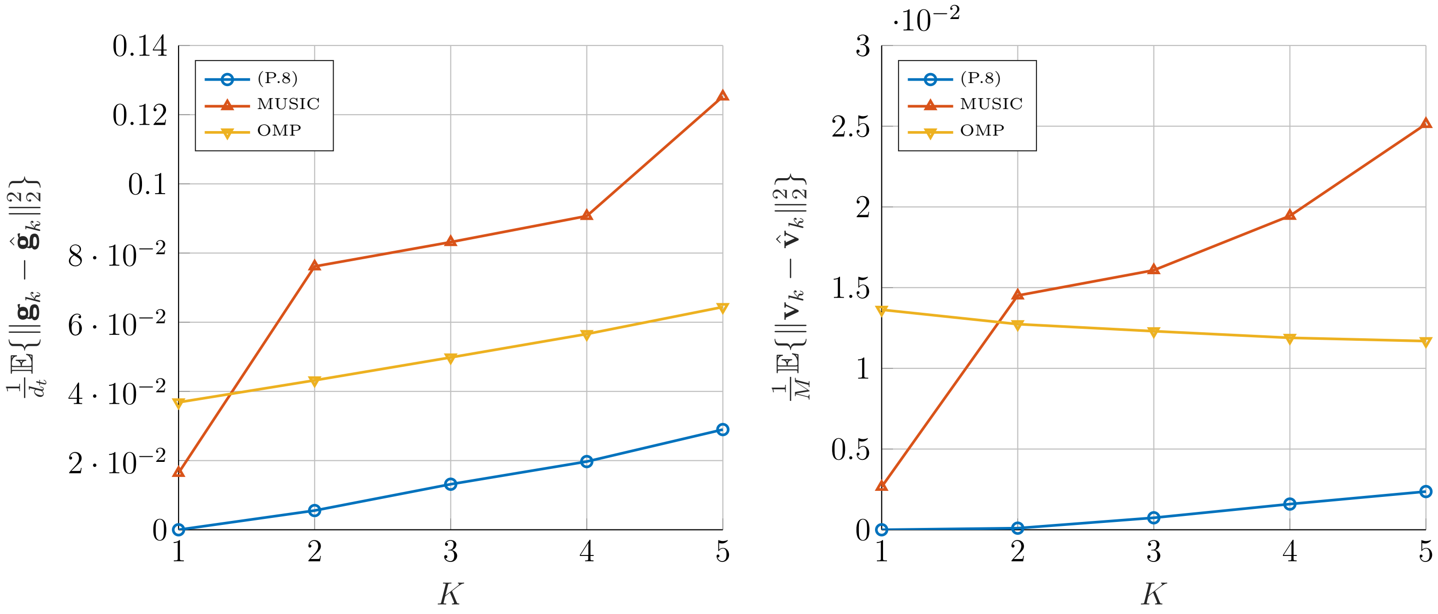
MUltiple SIgnal Classification (MUSIC)

MUSIC was first proposed in 1979 [60] and has been widely used and improved for frequency and direction of arrival estimation [69].

Given an observation system as $\mathbf{y} = \mathbf{B}\mathbf{x} + \mathbf{w}$, the covariance matrix of signal $\mathbf{x}$ is estimated based on the observation $\mathbf{y}$. Then, the matrix is split between signal (first K eigenvalues/eigenvectors) and noise and with this information a noise spectrum function is constructed over a predefined frequency dictionary of arbitrary length L_{MUSIC}. Then, the frequencies are estimated by locating the peaks. Here we use an improved version of this algorithm introduced in [69] where the observation vector is piled following a d-D Hankel matrix with dimension $H < M$.

Computational complexity of MUSIC algorithm is given by

$$O\Bigg(L_{\text{MUSIC}}\left(H^2 + Hd - HK + 2H - K + 1\right) + H(H-K)^2\Bigg) \tag{6.17}$$



(a) (b)

Fig. 6.4. MSE of frequency recovery in terms of K (a) and MSE of precoder estimation in terms of K (b) for a MU-MIMO DL scenario with $\mathbf{M} = [3, 6]^\top$ and $K \in [5]$ users.

where L_{MUSIC} is the length of the dictionary, d is the number of dimensions and H is the dimension of the d-D Hankel matrix.

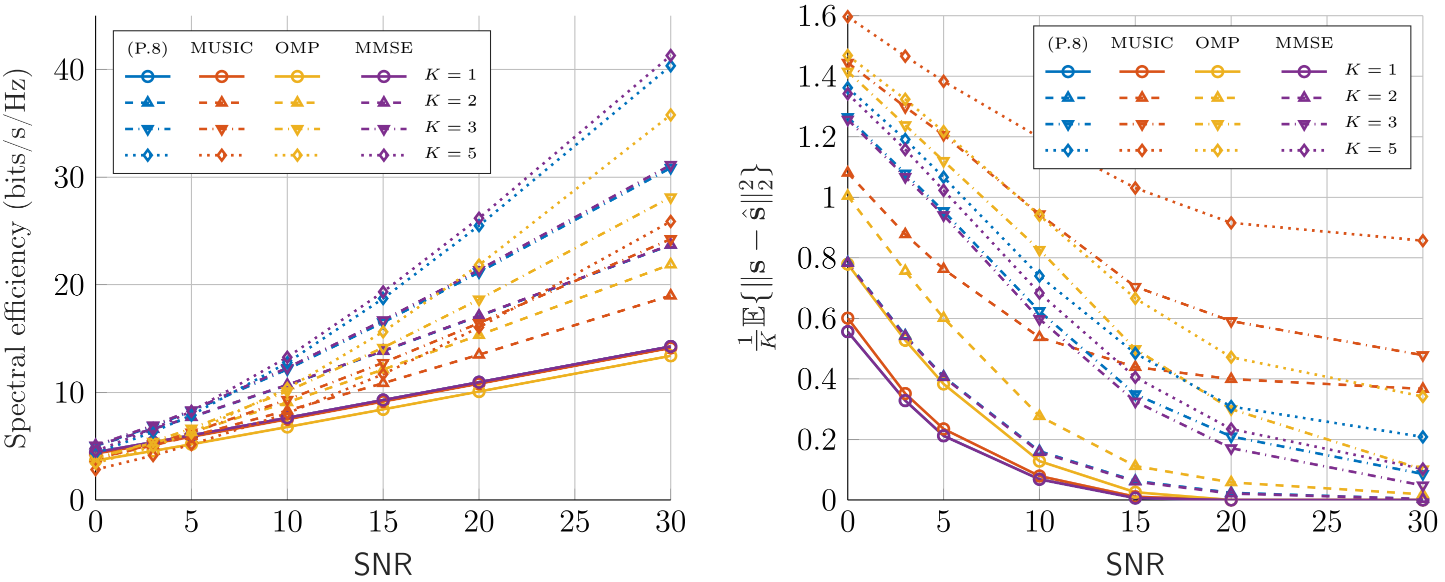


Fig. 6.5. Spectral efficiency in terms of SNR (a) and MSE of the received signal in terms of SNR (b) for a MU-MIMO DL scenario with $\mathbf{M} = [3, 6]^\top$ and $K \in \{1, 2, 3, 5\}$ users.

Chapter 7

Full channel estimation using atomic norm

As stated in previous chapters, increasing demands for higher capacity in present and upcoming generations of mobile communications have led the industry and researchers to explore the mmWave frequency range. The available bandwidth in this range is promising, being a very good candidate to meet the specified requirements in wireless communications speed. However, propagation in mmWave channels presents serious drawbacks due to the high path loss, atmospheric absorption and other conditions such as foliage loss, rain attenuation, diffraction, obstacle blockage, etc [77]. To counter this, MIMO technologies have been widely adopted in 5G deployments, being an enabling technology for ultra-high speed wireless communications due to the gains of spatial diversity [78]. Channel estimation in this type of scenarios is more challenging than in single-antenna transmissions given that the overhead grows with the number of antennas in the system [79], [80]. Thus, leveraging the sparse nature of propagation in the mmWave range to reduce the overhead in the transmission is of high interest. Several studies in the literature have explored compressed sensing techniques to recover the relevant information from the received signal which allows to reduce the number of necessary pilots and therefore, reducing the overhead in the communication [81], [82]. Other works, inspired by recent studies on deep neural networks for sparse signal recovery [83] have applied convolutional neural networks (CNN) for sparse channel estimation in mmWave MIMO channels [84].

A sparse model as in (3.9) is fully characterized by the structures of both transmitter and receiver, the angles of departure (AoD) and arrival (AoD) and the channel fading coefficients for every propagation path. In this chapter, we aim to apply the gridless methodology presented in chapter 2 to extract the angular information of the communication, AoD and AoA simultaneously, and with that, reconstruct the channel to provide a full channel estimation. We also compare the proposed gridless method with state-of-the-art techniques introduced in the previous chapter, *i.e.* OMP and MUSIC.

7.1 System model

This application follows a multiple measurement vector (MMV) as explained in section 3.5.2. We consider a MIMO propagation scenario with a d_t-dimensional antenna array deployment at the transmitter with M antennas and a d_r-dimensional antenna array deployment with N antennas at the receiver. For the transmitter, the m-th antenna is placed at $\mathbf{q}_m$ and the n-th receive element is placed at $\mathbf{v}_n$ as in (3.1). We considered both deployments as a sampled version of a uniform deployment, with sampling matrices $\mathbf{B}_t \in \{0,1\}^{M\times M_u}$ and $\mathbf{B}_r \in \{0,1\}^{N\times N_u}$ respectively for transmitter and receiver, where M_u and N_u are the number of antennas in the virtual uniform deployments at transmitter and receiver. Thus, channel matrix $\mathbf{H} \in \mathbb{C}^{N\times M}$ is modeled as in (3.9).

We focus in this chapter on full channel estimation and in gridlessly extracting all relevant propagation parameter such as AoA, AoD and channel gains. For that a sequence of known pilots $\mathbf{P} = [\mathbf{p}_1, \dots, \mathbf{p}_P] \in \mathbb{C}^{M\times P}$ of length P is transmitted over the channel generating a set of measurements $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_P] \in \mathbb{C}^{N\times P}$. So, plugging the pilot sequence into the MMV (3.12) we have that

$$\begin{aligned}\mathbf{Y} =&\mathbf{HP} + \mathbf{W}\\ =&\mathbf{B}_r\mathbf{R}_{\mathsf{N}}(\mathsf{f}_{1:K})\mathbf{G}\mathbf{T}^{\dagger}_{\mathsf{M}}(\mathsf{g}_{1:K})\mathbf{B}_t^{\dagger}\mathbf{P} + \mathbf{W}\end{aligned} \tag{7.1}$$

and recalling (3.13), we can vectorize the observation as

$$\begin{aligned}\mathbf{y} &= \mathrm{vec}(\mathbf{Y})\\ &= \left(\mathbf{P}^{\top} \otimes \mathbf{I}_N\right)\left(\mathbf{B}_t \otimes \mathbf{B}_r\right)\mathbf{A}_{\mathsf{L}}(\mathsf{l}_{1:K})\boldsymbol{\gamma} + \mathbf{w}\\ &= \mathbf{Q}\mathbf{A}_{\mathsf{L}}(\mathsf{l}_{1:K})\boldsymbol{\gamma} + \mathbf{w}\\ &= \mathbf{Q}\mathbf{h}_u + \mathbf{w}\end{aligned} \tag{7.2}$$

where matrix $\mathbf{Q} = \left(\mathbf{P}^{\top} \otimes \mathbf{I}_N\right)\left(\mathbf{B}_t \otimes \mathbf{B}_r\right) \in \mathbb{C}^{NP\times M_uN_u}$ is known and contains the pilots and the information about the non-uniform or sampled structures. Note that matrix $\mathbf{Q}$ is not a binary measuring matrix but it has a very specific structure based on the antenna structures at the transmitter and at the receiver. Frequency parameters regarding AoD, $\{\mathsf{g}_1, \dots, \mathsf{g}_K\}$, and AoA, $\{\mathsf{f}_1, \dots, \mathsf{f}_K\}$, are contained in $\mathsf{l}_k = [-\mathsf{g}_k^{\top}, \mathsf{f}_k^{\top}]^{\top} \in \mathbb{T}^{d_l}$ for $k \in [K]$, being $d_l = d_t + d_r$. The vector $\mathsf{L} = [\mathsf{M}^{\top}, \mathsf{N}^{\top}]^{\top}$ contains the number of antennas in each dimension at the transmitter and at the receiver. Finally, $\mathbf{A}_{\mathsf{L}}(\mathsf{l}_{1:K}) = [\mathbf{a}_{\mathsf{L}}(\mathsf{l}_1), \dots, \mathbf{a}_{\mathsf{L}}(\mathsf{l}_K)] \in \mathbb{C}^{M_uN_u\times K}$ is the matrix containing the K atoms (or steering vectors) defined as in (3.14). Thus, the unknown elements are the frequency parameters $\{\mathsf{l}_1, \dots, \mathsf{l}_K\}$ and the fading coefficients $\boldsymbol{\gamma} \in \mathbb{C}^K$.

Therefore, the problem is formulated so that the observation vector $\mathbf{y} \in \mathbb{C}^N$ is a binary measurement, done by means of matrix $\mathbf{Q}$, of a linear combination of atoms in $\mathcal{A}_{\mathsf{L}}$. With these formulation we can use (P.5) and its convex relaxation (P.6) to recover the frequency set $\mathsf{l}_{1:K}$. Once the frequencies are recovered, the channel fading coefficients $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_K]^\top$ can be estimated. Finally, the channel matrix $\mathbf{H}$ can be reconstructed.

7.2 Numerical analysis

We consider a transmitter with a uniform linear array deployment with $M = 4$ antenna elements, with $\mathbf{M} = [\mathsf{M}_1]^\top = [4]^\top$. At the receiver we have a uniform planar array with $N = 24$ and $\mathbf{N} = [\mathsf{N}_1, \mathsf{N}_2]^\top = [4, 6]^\top$. The propagation scenario will be characterized by $K \in \{1, 2, 3, 4, 5\}$ scatterers located randomly that will particularize the angles of departure and arrival following (3.2) and, consequently, the channel parameters $\mathbf{l}_k = [-\mathbf{g}_k^\top, \mathbf{f}_k^\top]^\top$. The sequence of pilots $\mathbf{P}$ that is transmitted over the channel would have length $P \in \{2, 3, 4, 5, 6\}$. We next show numerical results on frequency recovery error ($\mathbf{l}_{1:K}$), channel fading coefficients ($\boldsymbol{\gamma}$) and full channel estimation ($\mathbf{H}$) for different values of SNR.

7.2.1 Optimization of η

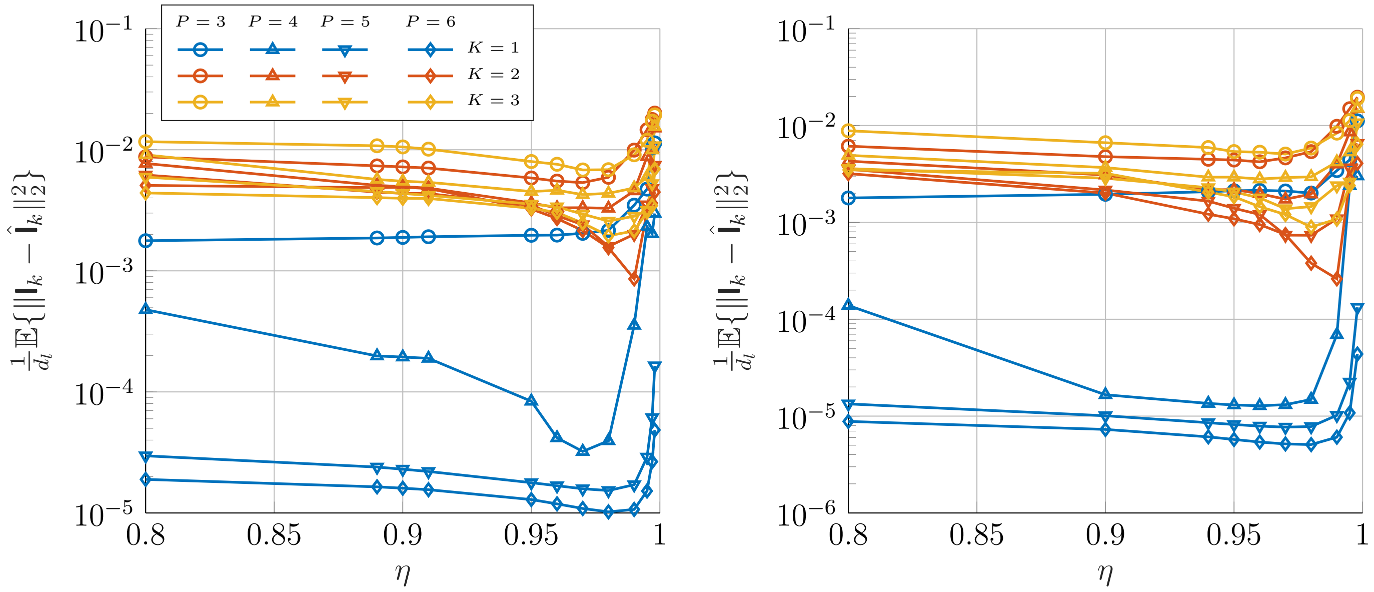
The optimization problem proposed in (P.6) has a parameter η that can be chosen so that it minimizes recovery error. The numerical simulations showed next indicate that there is an optimal value for η which depends on the SNR and the dimensional structure of the transmitter and receiver systems $\mathbf{L}$ which characterize the structure of the steering vectors $\mathbf{A}_{\mathbf{L}}(\mathbf{l}_{1:K})$. Both $\mathbf{L}$ and P are known a priori which allows to optimize η off-line. In Fig. 7.1 we report the MSE in the entire frequency vector recovery, defined $\mathbb{E}\{\|\mathbf{l}_k - \hat{\mathbf{l}}_k\|_2^2\}$, versus η. We see that the η starts making a difference from a certain value for P, being useless for short pilot sequences. We can also see by inspection that for a given value of SNR and P, the optimal η does not depend much on the number of scatterers K to be resolved.

7.2.2 Benchmarks

For comparison purposes we use the same two alternative methods used in chapter 6, *i.e.* OMP and MUSIC. Both alternative methods rely on a predefined dictionary of atoms over which the best K entries are selected. The length of the dictionary is chosen so that the computational complexity of both algorithms matches that of our proposed solution (P.6).

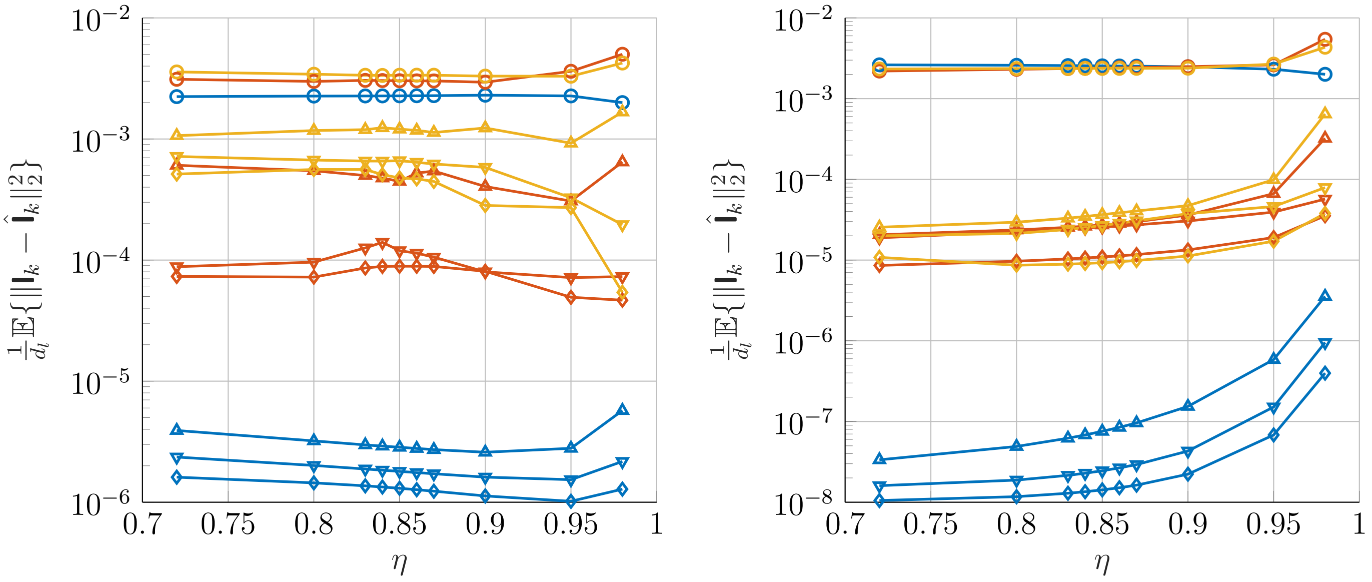
OMP

We recall that this is an K-iterative method that selects in an entry of the dictionary in each iteration. The computational complexity of this method is given in (6.16).



(a) SNR $= 0$dB

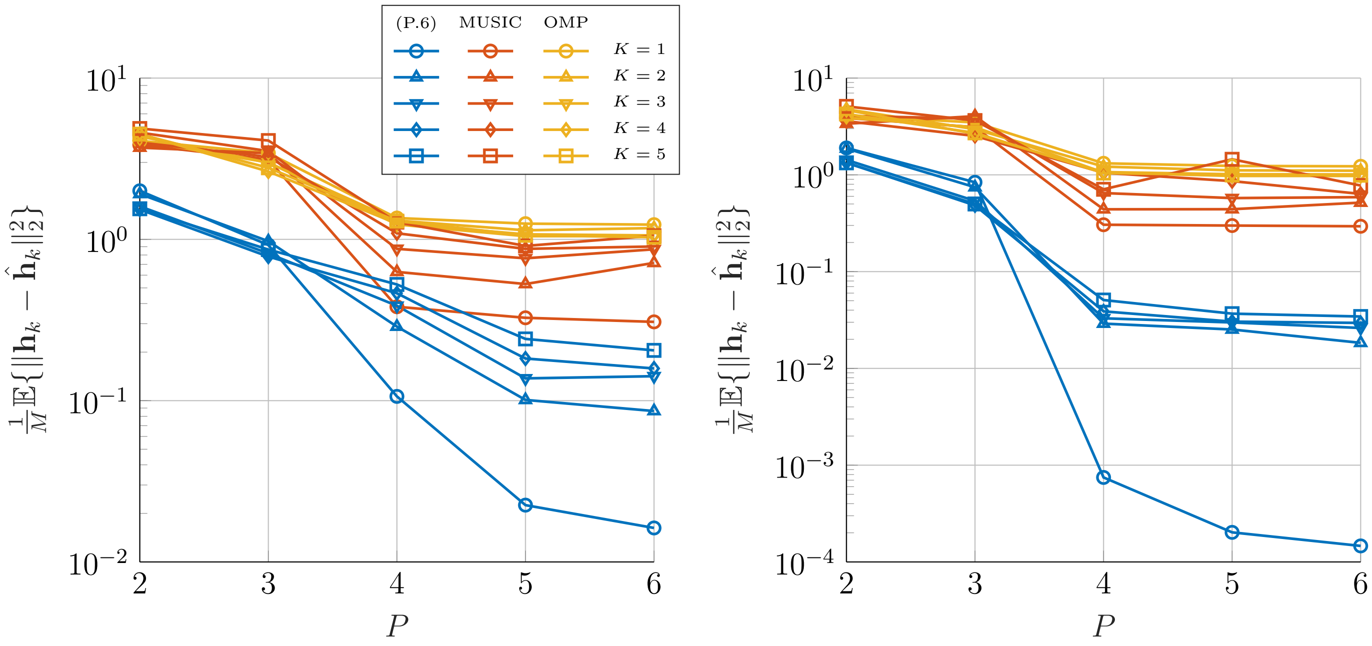
(b) SNR $= 3$dB



(c) SNR $= 10$dB

(d) SNR $= 30$dB

Fig. 7.1. Optimization of η for $\mathbf{M} = [4]^\top$, $\mathbf{N} = [4, 6]^\top$ with (a) SNR $= 0$dB, (b) SNR $= 5$dB, (c) SNR $= 10$dB and (d) SNR $= 20$dB.



(a) $\mathsf{SNR} = 0\text{dB}$ (b) $\mathsf{SNR} = 20\text{dB}$

Fig. 7.2. MSE of channel estimation versus number of pilots P for $\mathbf{M} = [4]^\top$ and $\mathbf{N} = [4, 6]^\top$ for (a) $\mathsf{SNR} = 0\text{dB}$, (b) $\mathsf{SNR} = 3\text{dB}$, etc

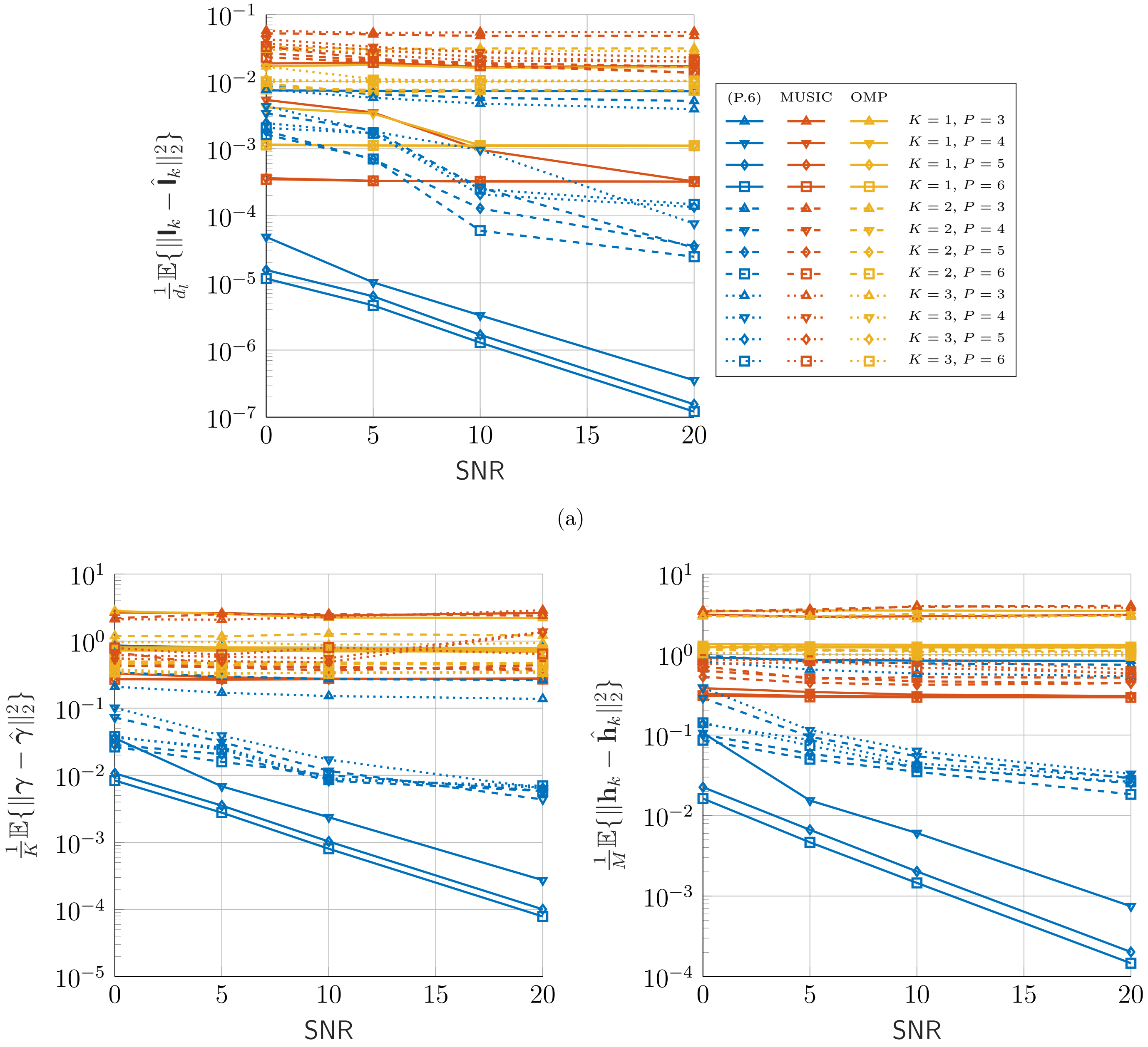
MUSIC

We recall that this method splits the estimated covariance using an eigenvalue decomposition and from that it searches through the dictionary finding the best K atoms that might have composed the signal. The computational complexity of this algorithm is given in (6.17).

7.2.3 Error in channel estimation

Once we have the optimal value of η has been found for every value of SNR, obtained by repeating the process shown in Fig. 7.1, we can run simulations comparing the performance on frequency recovery, and channel fading coefficients and channel estimation. In Fig. 7.3 we depict the MSE in the recovery of the frequency set in Fig. 7.3(a), the MSE in the channel fading coefficient estimation in Fig. 7.3(b) and finally, the MSE in full channel estimation in Fig. 7.3(c).

We can see that our proposed method (P.6) outperforms the two alternative methods for any given SNR and any length of the pilot sequence. This yields that a shorter pilot sequence is enough for the proposed solution to achieve the same performance than the state-of-the-art solutions for a given SNR. Therefore, it makes communication more efficient given that less space is required for the pilot sequence in the transmitted signal.



(a)

(b)

(c)

Fig. 7.3. Results for $\mathbf{M} = [4]^\top$, $\mathbf{N} = [4, 6]^\top$ for MSE of frequency recovery (a), channel coefficients $\boldsymbol{\gamma} = [\gamma_1, \dots, \gamma_K]^\top$ estimation MSE (b), and channel estimation MSE (c).

Chapter 8

Conclusion and future work

This thesis studies different applications of sparse signal and gridless parameter recovery via atomic norm decomposition. Sparse signals are formulated here as a linear combination of a number of atoms in the continuous space belonging to a so called atom set $\mathcal{A}_{\mathsf{J}}$ with infinite cardinality. Recovery relies on the rank minimization of a multi-level Toeplitz (MLT) matrix with a proper structure based on a d-dimensional Kronecker ordering. Also, a convex relaxation which guarantees parameter recovery in polynomial time, as opposed to the rank minimization, is explored. We propose a gridless methodology to recover the so called atoms, which are fully characterized by a d-dimensional parameter $\mathbf{f}$, and test it in different application scenarios. Also, we explore a machine learning approach to estimate the number of composite atoms that characterize the model in the scenarios where that number is unknown.

The proposed methods have been applied to several scenarios in the field of wireless communications. By numerical analysis we show that our proposed method improve significantly the performance when compared to state-of-the-art techniques. If the structure of the atoms is designed properly (via Kronecker ordering) without losing generality in the model, the complexity needed to recover the composite atoms of the model is linear in the sparsity degree, whereas in former solutions was exponential.

Next, we recapitulate the different applications studied in the thesis and finally outline some points for further research in future work.

8.1 Conclusions

Several applications are studied in this thesis that consist in parametric model recovery. We delivered numerical analysis that show how the proposed solution outperforms some of the state-of-the-art techniques for frequency and parameter extraction.

8.1.1 Automotive radar in mmWave frequency range

Millimeter-wave radars are key components of advanced driver assistant systems. In this thesis we have considered a mmWave pulse radar and, using the signal measurements at the antenna elements of a 3D receive array, we have studied the super-resolution of multiple target echoes. Upon exploiting the atomic norm to harness sparsity in the continuous parameter domain, the identification of the target parameters is re-cast in terms of the rank–minimization of a properly-structured multi-level Toeplitz matrix. As mentioned earlier, we have also developed and studied a convex relaxation that ensures the extraction of the relevant parameters in polynomial time. Finally, we have developed and discussed a low-complexity implementation of the recovery strategy that offers advantageous complexity/performance trade-offs.

8.1.2 Hybrid precoder design

For this application, we address one of the main challenges presented by mmWave systems via the design of a hybrid precoder in a MU-MIMO scenario. This precoder takes advantage of the sparse nature of mmWave channels to reduce the number or RF chains, therefore scaling down the total hardware cost, and yielding a feasible transmission strategy in this millimeter bands. We described some of the state-of-the-art algorithms for hybrid precoding, such as MUSIC and OMP, and present a novel algorithm based on AN by reformulating the hybrid precoder design as a matrix factorization problem and adopting the rank minimization of a MLT matrix, and the convex relaxation that ensures polynomial time. The AN–based precoder allows a hardware complexity (number of RF chains) reduction given that for a given error in terms of distance to the optimal fully-digital precoder, AN-based precoder needs less RF chains than the state-of-the-art precoders. Thus, mmWave spectrum can be exploited in a feasible way by the use of MIMO and hybrid precoding. Further, AN is revealed as a promising algorithm to leverage the sparse nature of the mmWave channel in order to reduce the number or RF chains.

8.1.3 Full channel estimation

In this application we leverage the sparse nature of mmWave propagation to estimate the channel. This is done by extracting the K frequency parameters of the channel which contain the information on the angles of departure and arrival and with which the channel matrix can be reconstructed. This is done, as in the other applications, with the rank minimization of a MLT matrix with a specific structure. We also studied the convex relaxation that guarantees recovery in polynomial time. We compare the proposed solution with some of the state-of-the-art methods such as OMP and MUSIC. We see via numerical analysis that our method outperforms both alternative algorithms.

8.2 Future work

Next we outline some points for further research

- Explore the theoretical conditions under which robust parameter recovery is achieved for noisy measurements.
- Study the feasibility of a different data gathering and training approach for the Machine Learning model to estimate the number of composite atoms that allow us to do the predictions without having to obtain matrix $\mathbf{T}_{\mathbf{J}}$.
- Extend the applications to other science and engineering fields where the studied sparse models fit, such as sensor networks, magnetic resonance imaging, computer vision, etc.

References

[1] A. V. Delgado, M. Sánchez-Fernández, L. Venturino, and A. Tulino, "Super-resolution in automotive pulse radars," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 4, pp. 913–926, Jun. 2021.

[2] A. V. Delgado, M. Sanchez-Fernandez, J. Llorca, and A. Tulino, "Feasible transmission strategies for downlink mimo in sparse millimeter-wave channels," *IEEE Communications Magazine*, vol. 56, no. 7, pp. 49–55, Jul. 2018.

[3] E. Crespo-Bardera, A. Vega Delgado, A. Garrido Martín, A. Fernández-Durán, and M. Sánchez-Fernández, "Textile multiantenna technology and relaying architectures for emergency networks," *Wireless Communications and Mobile Computing*, vol. 2019, Feb. 2019.

[4] D. Donoho, M. Vetterli, R. DeVore, and I. Daubechies, "Data compression and harmonic analysis," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2435–2476, 1998.

[5] E. J. Candès, 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] M. A. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2275–2284, 2009.

[7] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[8] S. Aja-Fernández, R. de Luis García, D. Tao, and X. Li, Eds., *Tensors in Image Processing and Computer Vision*, ser. Advances in Pattern Recognition. Springer, 2009.

[9] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Optimization with sparsity-inducing penalties," *Found. Trends Mach. Learn.*, vol. 4, no. 1, pp. 1–106, Jan. 2012.

[10] C. Beckmann and S. Smith, "Tensorial extensions of independent component analysis for multisubject fmri analysis," *NeuroImage*, vol. 25, no. 1, pp. 294–311, 2005.

[11] M. F. Duarte *et al.*, “Single-pixel imaging via compressive sampling,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, 2008.

[12] C. Shannon, “Communication in the presence of noise,” *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, Jan. 1949.

[13] P. Stoica, P. Babu, and J. Li, “Spice: A sparse covariance-based estimation method for array processing,” *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 629–638, Feb. 2011.

[14] R. Heckel and M. Soltanolkotabi, “Generalized line spectral estimation via convex optimization,” *IEEE Transactions on Information Theory*, vol. 64, no. 6, pp. 4001–4023, Jun. 2018.

[15] B. N. Bhaskar, G. Tang, and B. Recht, “Atomic norm denoising with applications to line spectral estimation,” *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5987–5999, Dec. 2013.

[16] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, “Compressed sensing off the grid,” *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7465–7490, Nov. 2013.

[17] S. Li, M. B. Wakin, and G. Tang, “Atomic norm denoising for complex exponentials with unknown waveform modulations,” *IEEE Transactions on Information Theory*, vol. 66, no. 6, pp. 3893–3913, 2020.

[18] Z. Tan, Y. C. Eldar, and A. Nehorai, “Direction of arrival estimation using co-prime arrays: A super resolution viewpoint,” *IEEE Transactions on Signal Processing*, vol. 62, no. 21, pp. 5565–5576, Nov. 2014.

[19] S. Semper, F. Roemer, T. Hotz, and G. D. Galdo, “Grid-free direction-of-arrival estimation with compressed sensing and arbitrary antenna arrays,” in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2018, pp. 3251–3255.

[20] Y. Wang, Y. Zhang, Z. Tian, G. Leus, and G. Zhang, “Super-resolution channel estimation for arbitrary arrays in hybrid millimeter-wave massive MIMO systems,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 5, pp. 947–960, 2019.

[21] Y. Chi and Y. Chen, “Compressive two-dimensional harmonic retrieval via atomic norm minimization,” *IEEE Transactions on Signal Processing*, vol. 63, no. 4, pp. 1030–1042, Feb. 2015.

[22] Z. Yang, L. Xie, and P. Stoica, “Vandermonde decomposition of multilevel toeplitz matrices with application to multidimensional super-resolution,” *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3685–3701, Jun. 2016.

[23] H. Chu, L. Zheng, and X. Wang, “Super-resolution mmWave channel estimation for generalized spatial modulation systems,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 6, pp. 1336–1347, Oct. 2019.

[24] Matilde Sánchez-Fernández, V. Jamali, J. Llorca, and A. Tulino, “Gridless multidimensional direction of arrival estimation for arbitrary 3D antenna arrays,” *IEEE Transactions on Wireless Communications*, 2021.

[25] J. Zheng, X. Gao, Z. Zhang, and Z. Feng, “A compact eighteen-port antenna cube for MIMO systems,” *IEEE Transactions on Antennas and Propagation*, vol. 60, no. 2, pp. 445–455, Feb. 2012.

[26] M. Wang and A. Nehorai, “Coarrays, MUSIC, and the Cramér–Rao bound,” *IEEE Transactions on Signal Processing*, vol. 65, no. 4, pp. 933–946, 2017.

[27] H. Qiao and P. Pal, “Guaranteed localization of more sources than sensors with finite snapshots in multiple measurement vector models using difference co-arrays,” *IEEE Transactions on Signal Processing*, vol. 67, no. 22, pp. 5715–5729, 2019.

[28] Z. Yang, J. Li, P. Stoica, and L. Xie, “Sparse methods for direction-of-arrival estimation,” in. Academic Press Library in Signal Processing: Array, Radar and Communications Engineering, 2018, vol. 7, ch. 11, pp. 509–581.

[29] Y. Chi and M. Ferreira Da Costa, “Harnessing sparsity over the continuum: Atomic norm minimization for superresolution,” *IEEE Signal Processing Magazine*, vol. 37, no. 2, pp. 39–57, Mar. 2020.

[30] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, “The convex geometry of linear inverse problems,” *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, 2012.

[31] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*. Springer, 2013.

[32] C. Carathéodory and L. Fejér, *Über den Zusammenhang der Extremen von harmonischen Funktionen mit ihren Koeffizienten un über den Picard-Landauschen Satz*. Tipografia matematica, 1911.

[33] K. Krishnan and T. Terlaky, “Interior point and semidefinite approaches in combinatorial optimization,” in *Graph Theory and Combinatorial Optimization*, D. Avis, A. Hertz, and O. Marcotte, Eds. Boston, MA: Springer US, 2005, pp. 101–157.

[34] I. F. Akyildiz, A. Kak, and S. Nie, “6G and beyond: The future of wireless communications systems,” *IEEE Access*, vol. 8, pp. 133 995–134 030, 2020.

[35] N. Wu, F. Zhu, and Q. Liang, “Evaluating spatial resolution and channel capacity of sparse cylindrical arrays for massive MIMO,” *IEEE Access*, vol. 5, pp. 23 994–24 003, 2017.

[36] S. Hu, F. Rusek, and O. Edfors, “Beyond massive MIMO: The potential of positioning with large intelligent surfaces,” *IEEE Transactions on Signal Processing*, vol. 66, no. 7, pp. 1761–1774, Apr. 2018.

[37] "5G Spectrum," GSM Asociation, Public Policy Position, 2016.

[38] Z. Pi and F. Khan, "An introduction to Millimeter-Wave mobile broadband systems," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[39] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[40] G. R. Maccartney, T. S. Rappaport, S. Sun, and S. Deng, "Indoor office wideband millimeter-wave propagation measurements and channel models at 28 and 73 GHz for ultra-dense 5G wireless networks," *IEEE Access*, vol. 3, pp. 2388–2424, 2015.

[41] B. Ai *et al.*, "On indoor millimeter wave massive mimo channels: Measurement and simulation," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 7, pp. 1678–1690, Jul. 2017.

[42] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[43] J. D. Jobson, "Multiple linear regression," in *Applied Multivariate Data Analysis: Regression and Experimental Design.* New York, NY: Springer New York, 1991, pp. 219–398.

[44] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16, San Francisco, California, USA: ACM, 2016, pp. 785–794.

[45] D. Berrar, "Bayes' theorem and naive bayes classifier," *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics; Elsevier Science Publisher: Amsterdam, The Netherlands*, pp. 403–412, 2018.

[46] C. Kwak and A. Clayton-Matthews, "Multinomial logistic regression," *Nursing research*, vol. 51, no. 6, pp. 404–410, 2002.

[47] J. Nocedal, "Updating quasi-newton matrices with limited storage," *Mathematics of Computation*, vol. 35, no. 151, pp. 773–782, 1980.

[48] F. Melo, "Area under the roc curve," in *Encyclopedia of Systems Biology*, W. Dubitzky, O. Wolkenhauer, K.-H. Cho, and H. Yokota, Eds. New York, NY: Springer New York, 2013, pp. 38–39.

[49] D. M. Grimes and T. O. Jones, "Automotive radar: A brief review," *Proceedings of the IEEE*, vol. 62, no. 6, pp. 804–822, Jun. 1974.

[50] J. Wenger, "Automotive radar - status and perspectives," in *Proc. of the IEEE Compound Semiconductor Integrated Circuit Symposium*, Oct. 2005, pp. 21–24.

[51] F. Fölster and H. Rohling, “Signal processing structure for automotive radar,” *Frequenz*, vol. 60, pp. 20–24, Feb. 2006.

[52] H. H. Meinel, “Evolving automotive radar — from the very beginnings into the future,” in *The 8th European Conference on Antennas and Propagation (EuCAP 2014)*, Apr. 2014, pp. 3107–3114.

[53] S. M. Patole, M. Torlak, D. Wang, and M. Ali, “Automotive radars: A review of signal processing techniques,” *IEEE Signal Processing Magazine*, vol. 34, no. 2, pp. 22–35, Mar. 2017.

[54] D. Kissinger, *Millimeter-Wave Receiver Concepts for 77 GHz Automotive Radar in Silicon-Germanium Technology.* New York: Springer Science+Business Media, 2012.

[55] V. Jain and P. Heydari, *Automotive Radar Sensors in Silicon Technologies.* New York: Springer Science+Business Media, 2013.

[56] H. Griffiths *et al.*, “Radar spectrum engineering and management: Technical and regulatory issues,” *Proceedings of the IEEE*, vol. 103, no. 1, pp. 85–102, Jan. 2015.

[57] Recommendation ITU-R M.2057-1, *Systems characteristics of automotive radars operating in the frequency band 76-81 GHz for intelligent transport systems applications*, 2018.

[58] J. Gamba, *Radar Signal Processing for Autonomous Driving.* Singapore: Springer, 2020.

[59] J. Capon, “High-resolution frequency-wavenumber spectrum analysis,” *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.

[60] R. Schmidt, “Multiple emitter location and signal parameter estimation,” *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[61] R. Roy and T. Kailath, “ESPRIT-estimation of signal parameters via rotational invariance techniques,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, Jul. 1989.

[62] W. Li, W. Liao, and A. Fannjiang, “Super-resolution limit of the ESPRIT algorithm,” *IEEE Transactions on Information Theory*, vol. 66, no. 7, pp. 4593–4608, 2020.

[63] E. Grossi, M. Lops, and L. Venturino, “Adaptive detection and localization exploiting the IEEE 802.11ad standard,” *IEEE Transactions on Wireless Communications*, vol. 19, no. 7, pp. 4394–4407, 2020.

[64] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part III: Radar-Sonar Signal Processing and Gaussian Signals in Noise.* Wiley, 2001.

[65] M. I. Skolnik, *Introduction to Radar Systems.* McGraw-Hill, 2001.

[66] X. Lan, L. Wang, Y. Wang, C. Choi, and D. Choi, “Tensor 2-D DOA estimation for a cylindrical conformal antenna array in a massive MIMO system under unknown mutual coupling,” *IEEE Access*, vol. 6, pp. 7864–7871, 2018.

[67] M. Rubsamen and A. B. Gershman, “Direction-of-arrival estimation for nonuniform sensor arrays: From manifold separation to Fourier domain MUSIC methods,” *IEEE Transactions on Signal Processing*, vol. 57, no. 2, pp. 588–599, Feb. 2009.

[68] A. J. den Dekker and A. van den Bos, “Resolution: A survey,” *Journal of the Optical Society of America A*, vol. 14, no. 3, pp. 547–557, Mar. 1997.

[69] W. Liao, “MUSIC for multidimensional spectral estimation: Stability and super-resolution,” *IEEE Transactions on Signal Processing*, vol. 63, no. 23, pp. 6395–6406, 2015.

[70] A. A. Abidi, “CMOS microwave and millimeter-wave ICs: The historical background,” *IEEE International Symposium on Radio-Frequency Integration Technology (RFIT)*, Aug. 2014.

[71] A. V. Delgado, “Low complexity precoding for MU-MIMO in millimeter wave systems,” *Master’s Thesis, Universidad Carlos III de Madrid*, Sep. 2017.

[72] J. Lee and N. Jindal, “High snr analysis for mimo broadcast channels: Dirty paper coding versus linear precoding,” *IEEE Transactions on Information Theory*, vol. 53, pp. 4787–4792, 2007.

[73] C. Peel, B. Hochwald, and A. Swindlehurst, “A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization,” *IEEE Transactions on Communications*, vol. 53, no. 1, pp. 195–202, Jan. 2005.

[74] A. Alkhateeb, G. Leus, and R. W. H. Jr., “Limited Feedback Hybrid Precoding for Multi-User Millimeter Wave Systems,” *IEEE Transactions on Wireless Communications*, vol. 14, no. 11, Nov. 2015.

[75] S. Shi and M. Schubert, “MMSE transmit optimization for multi-user multi-antenna systems,” in *Proceedings. (ICASSP ’05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 3, Mar. 2005, iii/409–iii/412 Vol. 3.

[76] S. Shi, M. Schubert, and H. Boche, “Downlink MMSE transceiver optimization for multiuser MIMO systems: MMSE balancing,” *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3702–3712, Aug. 2008.

[77] T. Bogale, X. Wang, and L. Le, “Chapter 9 - mmWave communication enabling techniques for 5G wireless systems: A link level perspective,” in *mmWave Massive MIMO*, S. Mumtaz, J. Rodriguez, and L. Dai, Eds., Academic Press, 2017, pp. 195–225.

[78] S. Chen, S. Sun, G. Xu, X. Su, and Y. Cai, "Beam-space multiplexing: Practice, theory, and trends, from 4G TD-LTE, 5G, to 6G and beyond," *IEEE Wireless Communications*, vol. 27, no. 2, pp. 162–172, 2020.

[79] Y. Li, M. Zhang, W. Zhu, M. Cheng, C. Zhou, and Y. Wu, "Performance evaluation for medium voltage MIMO-OFDM power line communication system," *China Communications*, vol. 17, no. 1, pp. 151–162, 2020.

[80] Y. Zhang, H. Cao, M. Zhou, and L. Yang, "Non-orthogonal multiple access in cell-free massive MIMO networks," *China Communications*, vol. 17, no. 8, pp. 81–94, 2020.

[81] Z. Wan, Z. Gao, B. Shim, K. Yang, G. Mao, and M.-S. Alouini, "Compressive sensing based channel estimation for millimeter-wave full-dimensional MIMO with lens-array," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 2, pp. 2337–2342, 2020.

[82] A. Liao, Z. Gao, H. Wang, S. Chen, M.-S. Alouini, and H. Yin, "Closed-loop sparse channel estimation for wideband millimeter-wave full-dimensional MIMO systems," *IEEE Transactions on Communications*, vol. 67, no. 12, pp. 8329–8345, 2019.

[83] J. Zhang and B. Ghanem, "ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1828–1837.

[84] S. Liu and X. Huang, "Sparsity-aware channel estimation for mmWave massive MIMO: A deep CNN-based approach," *China Communications*, vol. 18, no. 6, pp. 162–171, 2021.