# Structured Sparse Signal Recovery for mmWave Channel Estimation: Intra-vector Correlation and Modulo Compressed Sensing

A Thesis

Submitted for the Degree of

**Master of Technology (Research)**

in the Faculty of Engineering

by

**Dheeraj Prasanna**

Under the guidance of

**Prof. Chandra R. Murthy**


Electrical Communication Engineering
Indian Institute of Science, Bangalore
Bangalore – 560 012 (INDIA)

April 2021

# Acknowledgements

My sincere gratitude goes to my research supervisor, Prof. Chandra R. Murthy, for all his support and guidance throughout my term as a graduate student. I thank the Ministry of Human Resource Development (MHRD) for providing a scholarship for me to pursue my Masters' degree. I would also like to thank the management of IISc for providing the opportunity to pursue masters in a wonderful environment.

I would like to express my gratitude to Prof. Y. Gopala Rao from RVCE, whose encouragement to join IISc was a driving force for me to take up this challenge. I would also like to thank Shashank Vatedka and Karthik P.N. for motivating me to join the masters program.

I am grateful to the professors of the EECS division, for all the wonderful lectures and seminars. My special thanks goes to the members of the ECE office, especially Mr. Srinivasa Murthy, for all the administrative help. My sincere thanks goes to all my friends at IISc, with whom I have had many fun and memorable experiences.

Finally, I owe my deepest gratitude to my family. I would like to thank my parents and sister for all the love and support. I would also like to take this opportunity to thank my relatives in Bengaluru, especially my aunt, uncle, and their families, whose love and support made me feel at home while staying far away from home.

# Abstract

This thesis contributes new theoretical results and recovery algorithms for the area of sparse signal recovery motivated by applications to the problem of channel estimation in mmWave communication systems.

The thesis is written in two parts. The first part focuses on the recovery of sparse vectors with correlated nonzero entries from their noisy low dimensional projections. Such structured sparse signals can be recovered using the technique of covariance matching. Here, we first estimate the covariance of the signal from the compressed measurements, and then use the obtained covariance matrix estimate as a plug-in to the linear minimum mean squared estimator to obtain an estimate of the sparse vector. We present a novel parametric Gaussian prior model, inspired by sparse Bayesian learning (SBL), which captures the underlying correlation in addition to sparsity. Based on this prior, we develop a novel Bayesian learning algorithm called `Corr-SBL`, using the expectation-maximization procedure. This algorithm learns the parameters of the prior and updates the posterior estimates in an iterative fashion, thereby yielding a sparse vector estimate upon convergence. We present a closed form solution for the hyperparameter update based on fixed-point iterations. In case of imperfect correlation information, we present a pragmatic approach to learn the parameters of the correlation matrix in a data-driven fashion.

Next, we apply `Corr-SBL` to the channel estimation problem in mmWave multiple-input

multiple-output systems employing a hybrid analog-digital architecture. We use noisy low dimensional projections of the channel obtained in the pilot transmission phase to estimate the channel across multiple coherence blocks. We show the efficacy of the Corr-SBL prior by analyzing the error in the channel estimates. Our results show that, compared to genie-aided estimators and other existing sparse recovery algorithms, exploiting both sparsity and correlation results in significant performance gains, even under imperfect covariance estimates obtained using a limited number of samples.

In the second part of the thesis, we consider the sparse signal recovery problem when the measurement acquisition process has a limited dynamic range, for example, when low-resolution analog-to-digital converters are used. To counter the effect of signal clipping in these systems, we use modulo arithmetic to fold the measurements crossing the range back into the dynamic range of the system. For this setup, termed as *modulo-CS*, we answer the fundamental question of signal identifiability, by deriving conditions on the measurement matrix and the minimal number of measurements required for unique recovery of sparse vectors. We also show that recovery using the minimum required number of measurements is possible when the entries of the measurement matrix are drawn independently from any continuous distribution. Finally, we present an algorithm based on convex relaxation, and formulate a mixed integer linear program (MILP) for recovery of sparse vectors under modulo-CS. Our empirical results show that the minimum number of measurements required for the MILP is close to the theoretical result, for signals with low variance.

# Notation

**Fields**

| | |
|---|---|
| $\mathbb{R}$ | Field of real numbers |
| $\mathbb{R}_+$ | Field of non-negative real numbers |
| $\mathbb{C}$ | Field of complex numbers |
| $\mathbb{Z}$ | Field of integers |
| $\mathbb{F}^N$ | Set of vectors of length $N$ with entries from the field $\mathbb{F}$ |

**Other Sets:** Denoted by calligraphy alphabets

| | |
|---|---|
| $\lvert\mathcal{S}\rvert$ | Cardinality of set $\mathcal{S}$ |
| $\mathcal{S}^{\mathrm{C}}$ | Complement of $\mathcal{S}$ |
| $[N]$ | Set of integers from 1 to $N$ |
| $a : b$ | Set of integers from $a$ to $b$ |
| $\bigcup$ | Union of sets |

**Vectors:** Denoted by bold lowercase letters

| | |
|---|---|
| $x_i$ | $i^{\text{th}}$ element of vector $\mathbf{x}$ |
| $\mathbf{x}_{\mathcal{S}}$ | Vector supported on index set $\mathcal{S}$ |
| $\lVert\mathbf{x}\rVert_p$ | $\ell_p$-norm of vector $\mathbf{x}$ |
| $\lVert\mathbf{x}\rVert_0$ | $\ell_0$-norm of vector $\mathbf{x}$: Number of nonzero entries |
| $\langle\mathbf{a}, \mathbf{x}\rangle$ | Inner product between $\mathbf{a}$ and $\mathbf{x}$ |
| $\frac{1}{\mathbf{x}}$ | Element wise inverse of vector $\mathbf{x}$ |
| $\mathrm{diag}(\mathbf{x})$ | Diagonal matrix with entries of the vector $\mathbf{x}$ on the diagonal |
| $\mathrm{supp}(\mathbf{x})$ | Support of vector $\mathbf{x}$: Index set of nonzero elements in $\mathbf{x}$ |
| $\mathbf{0}$ | All zero vector |
| $\mathbf{1}$ | All ones vector |

**Matrices:** Denoted by bold uppercase letters

| | |
|---|---|
| $[\mathbf{A}]_{(i,j)}$ or $(\mathbf{A})_{ij}$ | $\{i,j\}^{\text{th}}$ element of matrix $\mathbf{A}$ |
| $[\mathbf{A}]_{(\mathcal{S}_r,\mathcal{S}_c)}$ | Submatrix consisting of intersection of rows and columns of $\mathbf{A}$ with respect to index sets $\mathcal{S}_r$ and $\mathcal{S}_c$ respectively |
| $\mathbf{A}_{\mathcal{S}}$ | Submatrix with columns of $\mathbf{A}$ corresponding on index set $\mathcal{S}$ |
| $\mathbf{A}^T$ | Transpose of matrix $\mathbf{A}$ |
| $\mathbf{A}^H$ | Conjugate transpose of matrix $\mathbf{A}$ |
| $\det(\mathbf{A})$ | Determinant of matrix $\mathbf{A}$ |
| $\operatorname{logdet}(\mathbf{A})$ | Logarithm of the determinant of matrix $\mathbf{A}$ |
| $\mathbf{A} \odot \mathbf{B}$ | Hadamard product of $\mathbf{A}$ and $\mathbf{B}$ |
| $\operatorname{Tr}[\mathbf{A}]$ | Trace of a square matrix $\mathbf{A}$ |
| $\operatorname{diag}(\mathbf{A})$ | Vector with diagonal entries of $\mathbf{A}$ |
| $\mathbf{I}_N$ | Identity matrix of size $N$ |
| $\mathbf{J}_{i,i}$ | Square matrix with 1 in $i^{\text{th}}$ diagonal entry and zeros elsewhere |

**Probability theory**

| | |
|---|---|
| $p(\mathbf{x};\theta)$ | Probability distribution on $\mathbf{x}$ with parameters $\theta$ |
| $\mathbb{E}$ | Expectation operator |
| $\mathbb{E}_x$ | Expectation with respect to probability distribution $p(x)$ |
| $\mathcal{CN}(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma})$ | Complex normal distribution (Complex Gaussian) on $\mathbf{x}$ with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ |
| i.i.d. | Independent and identically distributed |

**Miscellaneous**

| | |
|---|---|
| $\lvert x \rvert$ | Absolute value of a scalar $x$ |
| $\operatorname{Re}\{x\}$ | Real part of $x$ where $x$ can be scalar, vector or a matrix |
| $\lfloor t \rfloor$ | Integer part of scalar $t$: Floor function |
| $[\![t]\!]$ | Fractional part of scalar $t$ |

# Contents

# List of Figures

# List of Tables

# Publications from this thesis

## Conference

1. Dheeraj Prasanna and Chandra R. Murthy, "On the Role of Sparsity and Intra vector Correlation in mmWave Channel Estimation," 2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Atlanta, GA, USA, 2020, pp. 1-5.

## Journal

1. Dheeraj Prasanna, Chandrasekhar Sriram, and Chandra R. Murthy, "On the identifiability of sparse vectors from modulo compressed sensing measurements", in IEEE Signal Processing Letters, vol. 28, pp. 131-134, 2021.

2. Dheeraj Prasanna and Chandra R. Murthy, "mmWave channel estimation via compressive covariance estimation: Role of sparsity and intra-vector correlation," IEEE Transactions on Signal Processing, vol. 69, pp. 2356-2370, 2021.

# Chapter 1

# Introduction

Cellular communication systems have come a long way since their inception in the early 1980s. A combination of market demands and technological research have driven new generations of cellular systems every decade. Riding on the success of the fourth generation (4G), research and development efforts are now concentrating on the next generation of communication systems that has been termed as 5G technology. This generation is expected to connect a massive number of devices, support extremely high data rates, and provide ultra-high reliability. To cater to these design targets of 5G, innovations in newer technologies has gained utmost importance, and one key technology is millimeter wave (mmWave) communications.

## 1.1 mmWave Communications

The existing cellular networks including the 4G technology largely operate in the sub-6 GHz frequency spectrum. The bandwidth available at these frequencies are not sufficient to meet the capacity requirement of the next-gen wireless networks. A promising approach to address the demands is to utilize the large bandwidths available in the frequency bands

above 6 GHz, which includes the mmWave frequencies ranging from about 30-300 GHz.

### 1.1.1 mmWave MIMO Systems

mmWave communications has been in use for some years now, especially with it being standardized for the IEEE 802.11ad WiGig network standard. The high frequencies allow the use of large bandwidths, which, in turn, enables transmission at very high data rates. However, signals at these high frequencies experience high attenuation, scattering and absorption, leading to high propagation and penetration losses [1]. These challenges have hindered the use of mmWave communications for cellular systems. In the recent years, research and development advances in areas like massive deployment of small cells, multiple-input multiple-output (MIMO), have significantly improved the reliability to the mmWave systems, and have resulted in mmWave communications being a potential technology to the meet the 5G requirements [2].

The large antenna gains obtained in MIMO systems from the use of large antenna arrays help overcome the severe path loss at the high frequencies. At the same time, the short wavelengths of mmWave signals are conducive to the deployment of MIMO because a larger number of antennas can be accommodated in a small area due to the lower antenna spacing requirements [2]. In the cellular communications realm, the concept of multi-user MIMO (MU-MIMO) [3] has been introduced to add the capability of simultaneously supporting multiple users to MIMO. In this setup, a base station/access point with multiple antennas communicates with multiple users employing single or multiple antennas each.

Traditionally, to handle the interface between the analog and digital parts of a communication system, each antenna is equipped with a dedicated radio frequency (RF) chain

and analog-to-digital converters (ADCs). The high cost and power consumption of these circuits hinder the use of a dedicated RF chain for each antenna, and as a result traditional MIMO baseband precoding and combining schemes cannot be used. To overcome this limitation, the beamforming operations can be divided into the analog and digital domains leading to the hybrid analog-digital architecture [4]. Similarly, low resolution ADCs are employed to reduce the hardware cost and power consumption [5]. While these systems lead to lower hardware complexity, it comes at a cost of increased software complexity in implementing the transceivers. As a result, signal processing plays a critical role in the implementation of mmWave MIMO systems. Specifically, techniques such as compressed sensing, which exploit underlying structures in the system, are beneficial in the design of channel estimation and beamforming algorithms.

### 1.1.2 Channel Estimation

The performance of mmWave MIMO systems depends critically on the availability of channel state information (CSI) at both transmitter and receiver. In general, CSI is obtained in a separate training or channel estimation phase using pilot transmissions, which leads to additional pilot overhead. In the MU-MIMO scenario, the overhead increases with the number of users. This overhead can severely effect the spectral efficiency of the system, especially when the channel coherence time is small. Thus, both accuracy and low overhead play crucial role in the design of channel estimation algorithms.

The hybrid analog-digital architecture used in mmWave communications precludes the use of traditional channel estimation techniques, as the channel is only observed through the lens of the analog beams used at the radio-frequency (RF) front-end, which leads

to an underdetermined system. In addition to the hybrid architecture, the use of low resolution ADCs can also be detrimental to the accuracy of channel estimation due to the quantization errors introduced in the ADCs.

Multiple techniques have been studied in literature for mmWave channel estimation, including analog beam sweeping and compressed sensing based approaches. Measurement campaigns have revealed structures like sparsity and correlation in the channel, which can be exploited to obtain better estimates of the channel and/or reduce the pilot overhead. The measurements in the pilot transmission phase can be represented using a linear system model. Due to the underdetermined nature of the system caused due to the hybrid architecture, algorithms based on compressed sensing have played a major role in estimation by exploiting the sparsity in the signal. Exploiting additional structures along with sparsity (e.g., correlation) can offer even better estimates [6]. To achieve this, an important step is to develop better structural models that can fit the measurement campaigns in practical scenarios.

In this thesis, we concentrate on design and analysis of two techniques motivated by the above-mentioned problem of obtaining reliable channel estimates in mmWave MIMO systems. In the first part, we discuss the importance of correlation in the mmWave channels and how it can be exploited for channel estimation. In the second part, we consider the application of modulo operation for low resolution ADCs to increase the efficiency of channel estimation.

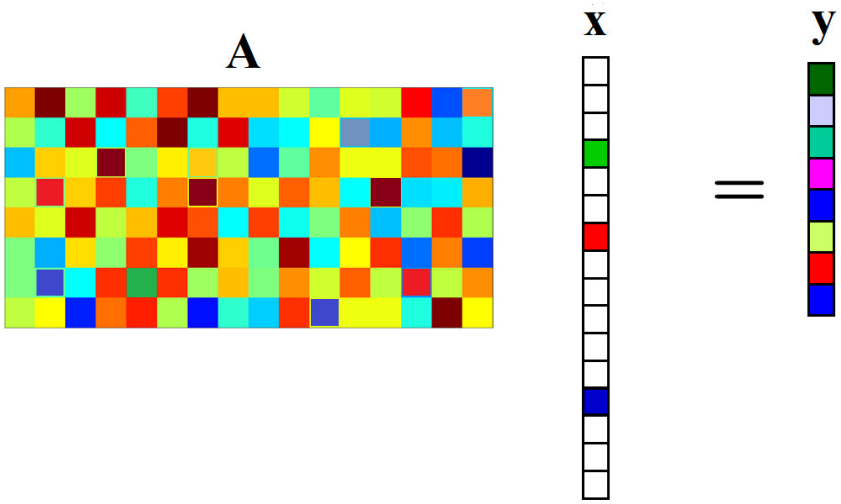



Fig. 1.1. Compressed sensing setup

## 1.2 Structured Signal Recovery

As seen in the previous section, exploiting structure in the signals is important for the performance of mmWave systems. However, the structured signal recovery problem can be useful in other applications as well. To this end, we first present the general framework of structured sparse signal recovery in Chapter 2, and present a novel solution. In Chapter 3, we formulate the mmWave channel estimation problem as one of structured sparse signal recovery and show that our algorithm yields better channel estimates compared to state-of-the-art methods. In the remainder of this section, we present a brief discussion of sparse signal recovery, concentrating on results from and algorithms in compressed sensing literature, and discuss a few extensions to structured signal recovery, summarized from [7].

### 1.2.1 Review of Compressed Sensing (CS)

The CS framework deals with reconstructing high-dimensional sparse signals from their low-dimensional noisy linear measurements. A signal $\mathbf{x} \in \mathbb{C}^N$ is said to be $s$-sparse if

$\|\mathbf{x}\|_0 \leq s \ll N$. The signal acquisition is typically assumed to be linear, which leads to a linear system of equations. The general setup for CS, as depicted in Fig. 1.1, involves a system of compressed (underdetermined) measurements of an $s$-sparse vector $\mathbf{x}$ given by the equation:

$$\mathbf{y} = \mathbf{A}\mathbf{x} \in \mathbb{C}^m \ (m < N). \tag{1.1}$$

In the classical linear algebra theory, it is impossible to recover $\mathbf{x}$ from $\mathbf{y}$ uniquely, since the condition $m < N$ leads to an underdetermined system having infinitely many solutions. However, an important facet of the CS or sparse signal recovery area is that the structural assumption of sparsity on the input vector can lead to the above system being identifiable, i.e., there can be a one-to-one correspondence between all $s$-sparse vectors and the corresponding compressed measurements, when $\mathbf{A}$ satisfies certain assumptions. Additionally, it has also been shown that efficient algorithms exist that guarantee successful recovery of sparse vectors from the compressed measurements. In this section, we present some important results pertaining to both theoretical and algorithmic aspects of CS which are useful in understanding the results in this thesis.

### Theory

A natural cost function to minimize in order to find the sparsest solution vector to (1.1) is the $\ell_0$-norm, which returns the number of nonzero entries in the vector. This leads to the optimization problem given by [7, Section 2.2]:

$$\arg\min_{\mathbf{z}} \|\mathbf{z}\|_0 \text{ subject to } \mathbf{A}\mathbf{z} = \mathbf{y}. \tag{$\mathrm{P}_{0-CS}$}$$

It can be shown that the existence of unique solution to the above optimization problem is equivalent to the identifiability of sparse vectors from compressed measurements. The important results with respect to this setup are:

1. Every $s$-sparse vector $\mathbf{x} \in \mathbb{C}^N$ is the unique solution to the ($\mathrm{P}_{0-CS}$) problem if and only if every set of $2s$ columns of the measurement matrix $\mathbf{A}$ are linearly independent [7, Theorem 2.13].

2. The minimal number of measurements required for identifiability of every $s$-sparse vector from compressed measurements is $m = 2s$ [7, Theorem 2.14].

3. For any $N \geq 2s$, there exists a measurement matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ with $m = 2s$ rows such that every $s$-sparse vector $\mathbf{x}$ can be recovered from its measurement vector $\mathbf{y} = \mathbf{A}\mathbf{x}$ as a solution to the ($\mathrm{P}_{0-CS}$) optimization problem. [7, Theorem 2.14]

Unfortunately, $\ell_0$-minimization is an NP-hard problem. Hence, the design of fast and provable reconstruction algorithms is not a straightforward task. A popular and well-understood method in CS literature is the convex relaxation of the $\ell_0$-norm by using the $\ell_1$-norm, which leads to the convex optimization problem [7, Section 3.1]:

$$\arg\min_{\mathbf{z}} \|\mathbf{z}\|_1 \text{ subject to } \mathbf{A}\mathbf{z} = \mathbf{y}. \qquad (\mathrm{P}_{1-CS})$$

Due to the convexity of the $\ell_1$-norm, the optimization problem can be solved using methods from convex optimization. In fact, it can be rewritten as a linear program. To obtain recovery guarantees for ($\mathrm{P}_{1-CS}$), the null space property on the measurement matrix was introduced, which can be shown to be necessary and sufficient for the unique recovery of every $s$-sparse signal [7, Section 4.1, Theorem 4.5].

**Definition 1** (Null space property)**.** *A matrix* $\mathbf{A} \in \mathbb{C}^{m \times N}$ *is said to satisfy the null space property of order* $s$*, if for all index sets* $\mathcal{S}$ *with* $|S| \leq s$*,*

$$\|\mathbf{u}_{\mathcal{S}}\|_1 < \|\mathbf{u}_{\mathcal{S}^{\mathrm{C}}}\|_1$$

*holds for every* $\mathbf{u} \in \mathbb{C}^N \setminus \{\mathbf{0}\}$ *with* $\mathbf{A}\mathbf{u} = \mathbf{0}$*.*

Other properties of the measurement matrix, such as mutual coherence [7, Chapter 5] and restricted isometric property (RIP) [7, Chapter 6] have also been considered to obtain recovery guarantees for the $\ell_1$ norm minimization and other sparse recovery algorithms. These recovery guarantees mainly deal with certain conditions or assumptions on the measurement matrix $\mathbf{A}$. As a result, design of matrices that satisfy these conditions also form an important part of CS literature. A breakthrough result for construction of $\mathbf{A}$ has been the use of random matrices such as Gaussian random matrices. To this end, a key result in CS states that, with high probability, all $s$-sparse vectors can be reconstructed using a variety of algorithms, when the entries of the measurement matrix are drawn independently from a sub-Gaussian random matrix, with the number of rows satisfying $m \geq Cs \log(N/s)$, where $C > 0$ is a universal constant [7, Theorem 9.12]. This important discovery has lead to the popularity of CS in a variety of applications. While sub-Gaussian random matrices have been popular choices for measurement matrices, there exist other constructions as well that have been widely used in various applications.

### Algorithms

A variety of algorithms have been designed for the sparse signal recovery problem. In this subsection, we discuss a subset of these algorithms which come under the purview of this

thesis. The algorithms can be divided into three important categories.

1. **Greedy algorithms:** These are iterative algorithms, where, in the $n^{\text{th}}$ iteration, locally optimal choices of index are added to obtain a target support $\mathcal{S}^n$ (updated in each iteration), and the target vector $\mathbf{x}^n$ is updated based on the best vector supported on $\mathcal{S}^n$ that fits the measurement. Some popular greedy methods are orthogonal matching pursuit (OMP), compressive sampling matching pursuit (CoSaMP) and subspace pursuit (SP). These algorithms generally tend to have low computational complexity and are popular choices in many applications [7, Section 3.2].

2. **Basis pursuit (BP):** This category includes the algorithms which solve the convex optimization problem given in ($\mathrm{P}_{1-CS}$). An alternative formulation in terms of linear programming is also popular. Other variants of BP such as LASSO and the Dantzig selector have also been studied in literature [7, Section 3.1].

3. **Bayesian methods:** These algorithms impose a fictitious prior on the unknown vector and use probabilistic inference based methods to obtain posterior estimates for the vector. Sparse Bayesian learning (SBL) [8] is a popular Bayesian method that has been considered for a large number of applications. These algorithms tend to have higher complexity, but usually perform better than the other methods. Low complexity versions of these algorithms have also recently become available.

### 1.2.2 Structured Sparse Signals

Early research in CS mainly dealt with finding sparse solutions to underdetermined systems of linear equations. However, in practical applications, signals tend to have additional structure, and prior knowledge of these structures can be exploited to improve the recovery

performance, especially in term of reducing the number of measurements required to obtain solutions with a given level of accuracy. Two of the important structures considered in literature are block sparsity and joint sparsity.

1. **Block sparsity:** In this model, the signal is divided into blocks or groups. A signal is called as group/block sparse if only a small number of the groups have nonzero entries [9]. By exploiting the group structure, algorithms based on greedy and Bayesian methods have been developed in literature.

2. **Joint sparsity:** Consider the setup where measurements are obtained from multiple signals given by:

$$\mathbf{y}_i = \mathbf{A}\mathbf{x}_i,\ i = 1, 2, \ldots, T. \tag{1.2}$$

   This setup is well known in the CS literature as the multiple measurement vector (MMV) paradigm. In this setup, along with each signal being sparse independently, if all the signals have a common support, the vectors are termed as jointly sparse [10]. To exploit this structure, instead of $\ell_1$-minimization, the use of a mixed $\ell_1/\ell_2$-minimization has been considered. Extensions of OMP, SBL algorithms discussed in the last section have also been proposed for the MMV setup.

The structures in addition to sparsity that has been presented here is not exhaustive. In fact, towards defining theory along similar lines to that of CS, there has been work on model-based compressed sensing [11] and distributed compressed sensing [12], which cater to more general structured sparsity setups. Generative models describing structured sparse models are presented in more detail in Chapter 2. In the following sections, we introduce the two structured signal recovery problems that are considered in this thesis.

## 1.3 Intra-vector Correlation

The structured sparse models discussed in the previous section concentrated primarily on the structure in the position of the nonzero entries in the signal. Very few works have considered correlation among the nonzero entries, and how to exploit it. The block sparsity model was extended to include correlation among the nonzero entries inside a block, which is called as intra-block correlation [9]. In the joint sparsity model, correlation across the different sparse vectors have been studied, and extensions to the MMV algorithms have been considered [13]. However, the general scenario of intra-vector correlation, where all the nonzero coefficients of the sparse vector can be correlated has not been well studied in the literature.

Revisiting the channel estimation problem in mmWave, a number of works have considered sparse models for the mmWave channels. Spatial sparsity arises in mmWave because the signal arrives at the receiver in a small number of paths [1, 14]. However, these models do not consider any correlations present between the multipaths of the channel. In sub-6 GHz systems, spatial correlation is typically introduced by the close spacing of antennas (mutual coupling) or due to co-located transmitters or receivers [15, 16]. In the mmWave MIMO scenario, new sources of correlation can exist due to the physical blockage and common scattering, which can degrade the system performance [2]. Spatial correlation can be beneficial for beamforming if the users have sufficiently different covariance matrices, but it is detrimental to the multi-antenna diversity gain because the effective rank of the channel gets reduced. However, correlation is unavoidable in general, which makes it necessary to incorporate correlation along with sparsity into statistical models to obtain accurate channel estimates.

Our goal in the first part of the thesis is to study the interplay between correlation and sparsity in the recovery of high-dimensional vectors from a small number of low dimensional projections. It is well known that, when the covariance matrix of a vector is known, an optimal estimator is the minimum mean squared error (MMSE) estimator. To reduce complexity, a simple linear minimum mean squared error (LMMSE) estimator can also be considered. The performance of this estimator increases when the entries of the vector are highly correlated, as we will see from the LMMSE curve in Fig. 3.6. in Chapter 3. Loosely speaking, when a vector is highly correlated, it can be reliably estimated from a small number of its noisy linear projections using an LMMSE estimator. On the other hand, if the vector is sparse, it can again be recovered from a small number of noisy linear projections, but more sophisticated sparse recovery algorithms are necessary. In this context, it is pertinent to ask the question of how to recover vectors from noisy low dimensional linear projections when both structures - sparsity and correlation - are simultaneously present. Moreover, how can one extend this to the case where the covariance matrix of the sparse vectors is unknown, and needs to be estimated from a set of measurements obtained using independent instantiations of the sparse vectors? And finally, what are the performance benefits that can be obtained?

We have tried to explore these questions in the thesis, with the mmWave channel estimation as a case study. To this end, we consider a statistical model incorporating both sparsity and correlation to model the mmWave MIMO channel. For the model, we derive a Bayesian algorithm on the lines of SBL algorithm, and analyze the performance in terms of the error in the channel estimates and the spectral efficiency of the system.

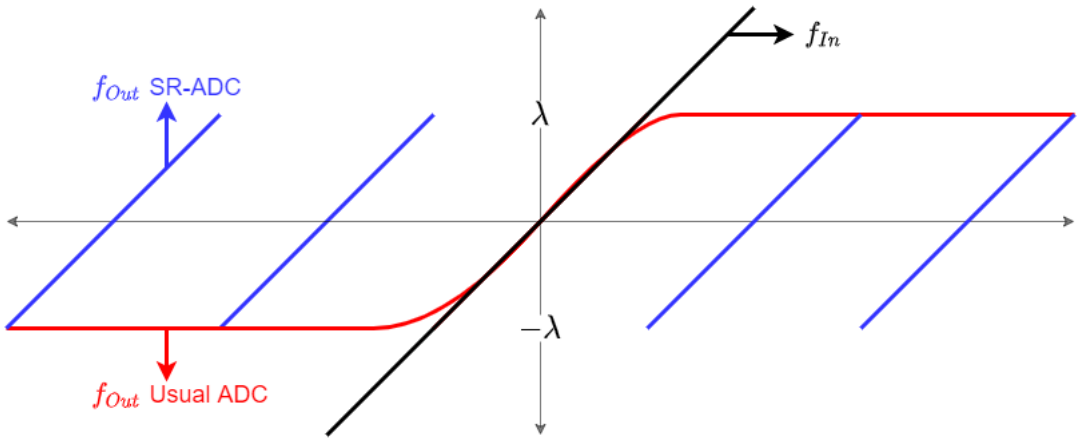



Fig. 1.2. Transfer function of conventional ADC compared with self-reset ADC

## 1.4 Dynamic Range in Signal Acquisition

The second problem is motivated by the use of analog-to-digital converters (ADCs), which form an integral part of the signal acquisition at the receiver of any digital communication system. A key design consideration for ADCs is the dynamic range of the system, which has a direct bearing on the quantization error. Dynamic range is defined as the ratio between the maximum and minimum value that can be represented by the ADC. Hence, dynamic range is dependent on both the resolution (number of bits) and the maximum allowable signal amplitude that can be represented without clipping the signal. Assuming the resolution is fixed, lower quantization error is obtained when the ADCs are designed with smaller dynamic range, i.e., smaller quantization range. However, if this quantization range is smaller than the highest amplitude of the input signal, it leads to signal loss due to clipping.

The literature has in general concentrated on algorithms to de-clip the signal (e.g., using its band-limitedness) or ignore the clipped signal. A recent and new direction of research

to avoid clipping has been the so-called self-reset ADCs (SR-ADCs) [17], which fold the amplitudes back into the dynamic range using the modulo arithmetic, thus mitigating the clipping effect. The transfer function of SR-ADC with parameter $\lambda$ is given in Fig. 1.2, where, if the input $f_{\text{In}}$ crosses threshold $|f_{\text{In}}| > \lambda$, it is folded such that output $f_{\text{Out}}$ is in the range $[-\lambda, \lambda]$. The SR-ADC encounters a different information loss, namely, the loss caused due to the modulo operation. As a result, the Nyquist-Shannon sampling theorem does not hold, as it assumes that the signal can be measured using an infinite dynamic range and with infinite precision. This led to the development of an alternative sampling theory called the unlimited sampling framework [18], which showed that if a bandlimited signal is sampled at a sampling rate slightly above Nyquist rate, the signal can be recovered from its folded samples. This framework provided a breakthrough for a principled use of SR-ADCs for signal acquisition.

In the context of the channel estimation for mmWave systems, the relation between observed data and the channel vector is usually modeled as a linear system. This model assumes infinite dynamic range in the signal acquisition. Due to the large number of antennas in mmWave MIMO systems, these ADCs are predominantly designed with low resolution, i.e., the number of bits used to represent the signal is small. As a result, ADCs with small quantization range is preferred. Hence, SR-ADCs can play a crucial role in the design of ADCs for mmWave systems. To model the SR-ADCs along with the hybrid analog-digital architecture, the low pass filtered samples in the unlimited sampling framework can be replaced with modulo compressed measurements, leading to the so-called modulo compressed sensing (modulo-CS) problem.

Modulo-CS is an emerging area with a few initial works presenting algorithms under

restricted settings. However, a general theory for the identifiability of the sparse signals has not been studied. To this end, the second part of the thesis aims at answering the questions listed below.

1. What are the conditions on the measurement matrix $\mathbf{A}$ such that, every $s$-sparse vector $\mathbf{x}$ results in an unique modulo measurement $\mathbf{z}$?

2. What is the minimum number of measurements ($m$) required for the identifiability of sparse signals under modulo measurements?

3. How to construct a measurement matrix $\mathbf{A}$ that satisfies the above conditions?

4. How to solve the problem efficiently?

## 1.5 Thesis Outline and Key Contributions

This thesis is presented in two parts. The first part, comprising of two chapters, discusses the correlated sparse signal recovery problem and a case study of the mmWave channel estimation problem. The second part of the thesis is devoted to developing theory and algorithms for the modulo-CS problem. In this section, we summarize the key contributions of each chapter of the thesis.

### 1.5.1 Correlated Sparse Recovery

In Chapter 2, we study the sparse signal recovery problem when the nonzero entries of the sparse vector are correlated. To exploit the correlation information, the sparse recovery in the MMV setup is interpreted as a covariance estimation problem. We study the effect of intra-vector correlation on two correlation-aware covariance matching algorithms, where

the correlation is assumed to be zero. To obtain better estimates by exploiting correlation, a novel covariance matching algorithm is presented. The key contributions are:

1. We formulate a hierarchical complex Gaussian prior with covariance matrix that can incorporate known correlation structure while at the same time induce sparsity.

2. Using the above prior, we develop a novel Bayesian sparse recovery algorithm based on evidence maximization called `Corr-SBL`. A closed-form solution to update the hyper-parameters is obtained as a fixed point iteration, and a guarantee to show that the update leads to convergence of the algorithm is provided.

3. We also present a heuristic approach for learning the correlation coefficient in the unknown correlation case.

4. We derive an alternative representation of the output of the algorithm as a plug-in LMMSE estimator, which in turn allows us to connect Corr-SBL to the general approach of covariance matching. In effect, we show that Corr-SBL is a covariance matching algorithm.

### 1.5.2 mmWave Channel Estimation

In Chapter 3, we address the mmWave channel estimation problem by modeling both spatial sparsity and correlation in the same model. We also present the application of Corr-SBL described in Chapter 2 for the above problem. Our key contributions include:

1. We investigate the utility of exploiting both spatial sparsity and correlation in the multi-paths of a mmWave channel, for uplink channel estimation in a multi-user MIMO setup with the hybrid analog-digital architecture.

2. We extend Corr-SBL to jointly estimate the channels over multiple coherence blocks. To reduce the overall latency in channel estimation, we develop an online version of the algorithm.

3. We derive an expression for the normalized mean squared error (NMSE) in channel estimation, and discuss the efficacy of Corr-SBL for channel estimation.

4. We present a design for hybrid combining of the received data using the channel estimates, and derive a lower bound on the spectral efficiency achieved by each user.

We discuss the significance of the choice of prior using the analytical expression for the NMSE, and present empirical comparisons against optimal genie-aided estimators in highly measurement-constrained scenarios. The Monte Carlo simulation results illustrate that it is advantageous to exploit correlation as well as sparsity even under imperfect correlation information, depending on the correlation level and number of independent channel instantiations that are available to estimate the covariance.

### 1.5.3 Modulo Compressed Sensing

In Chapter 4, we consider the sparse recovery problem from modulo compressed measurements. We present the identifiability conditions for the recovery of sparse signals under this framework without any assumptions on the input signal or the measurement matrix. We also present a simple algorithm which does not assume any prior distribution on the sparse signal, and present theoretical guarantees for unique recovery using the algorithm. The key contributions can be summarized as follows:

1. We present an optimization problem for the Modulo-CS setup, and obtain necessary

and sufficient conditions for unique recovery of sparse signals from their modulo compressed measurements.

2. We show that the minimum number of measurements $m$ to uniquely reconstruct every $s$-sparse signal from modulo measurements is $2s + 1$.

3. We show that $m = 2s + 1$ suffices, and that a measurement matrix with $2s + 1$ rows and entries derived from any continuous distribution satisfies the identifiability conditions with high probability.

4. We develop an algorithm using convex relaxation to the optimization problem for Modulo-CS setup, and present integer range space property for the measurement matrix which guarantees unique recovery of sparse signals.

### 1.5.4 Conclusions and Future Work

The thesis is concluded in Chapter 5, where we offer a brief summary of the key results of the thesis, and present some possible extensions and directions for further study.

# Chapter 2

# Correlated Sparse Signal Recovery

In the current era of digital economy, data powers the economy similar to how oil fueled industrial economy, leading to the metaphor "Data is the new oil". In fact, most of today's applications are becoming more and more reliant on the availability of data. While data collection from a multitude of devices makes virtually infinite data collection possible, factors like storage, transmission bandwidth, latency and data processing drive the data availability at other locations. The traditional Nyquist theorem based sampling approaches are fast turning out to be prohibitively expensive, and unable to keep up with the demand for data. One promising approach to alleviate this problem is to exploit signal structure to reduce the sampling rate requirements. In particular, sparse signal recovery has been a breakthrough in sub-Nyquist (it has also been termed beyond-Nyquist in the literature) sampling based approaches.

In Section 1.2, we presented an overview of the area of sparse signal recovery. Traditional sparse recovery algorithms including greedy methods and basis pursuit discussed in

A part of this chapter was presented in SPAWC conference [19] and has been published in IEEE transactions on signal processing journal [20].

the review yield accurate point estimates when the measurement matrix satisfies strong requirements such as restricted isometric property (RIP), which are rarely met in highly measurement-constrained scenarios. Further, they do not reveal the posterior distribution which can add flexibility in dealing with additional structures. With the increasing demand in data, it becomes necessary to exploit additional structures along with sparsity. To this end, Bayesian inference algorithms such as SBL have been known to be well suited to exploit these structures, especially the joint sparsity structure. Their primary goal is to infer the best-fitting distribution from a parameterized class of distributions. Point estimates can then be obtained from the posterior distribution.

In this chapter, we consider modeling the additional structure of correlation among the nonzero entries of the sparse signal. Recovery of sparse vectors under this model (with unknown covariance matrix) is termed as correlated sparse signal recovery. To this end, a pragmatic approach could be first to estimate the covariance matrix from the measurements, and then estimate the vectors using a plug-in-LMMSE estimator based on the estimated covariance matrix. The problem of estimating the covariance matrix of a set of vectors from noisy linear measurements has been an important area of research, and a number of works have presented theoretical results for the error in the estimation of the covariance matrix. The specific case when underdetermined linear measurements of the vector are available has also been analyzed [21]. However, the final goal of correlated sparse recovery is recovering the vectors themselves, and not just the covariance matrix. In this context, the question of what is the best covariance estimate and what is the best way to recover the vectors are open problems, to the best of our knowledge.

The goal of this chapter is to develop a Bayesian inference algorithm for the correlated

sparse recovery problem. The problem setup and signal model is presented in Section 2.1. In Section 2.2, an interpretation of the above problem as a covariance estimation problem is presented. Finally in Section 2.3 and 2.4, we present the Bayesian learning methodology and the `Corr-SBL` algorithm based on this methodology for the correlated sparse recovery problem. We apply the algorithm to the problem of mmWave channel estimation in Chapter 3.

## 2.1 Structured Sparse Signal Recovery

### 2.1.1 Multiple Measurement Vector Setup

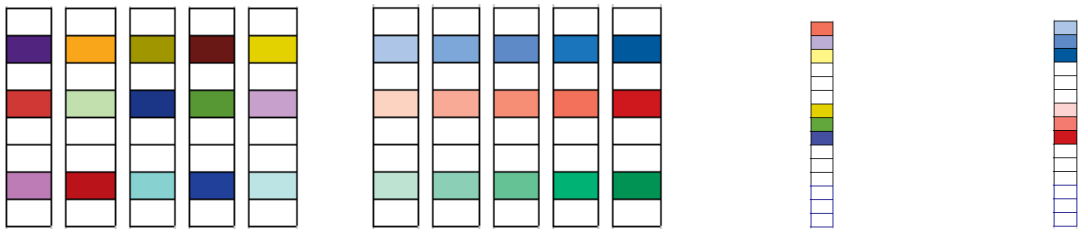
We consider the multiple measurement vector (MMV) setup, where noisy low dimensional projections of vectors $\mathbf{x}_i \in \mathbb{C}^N$ obtained according to the equation:

$$\mathbf{y}_i = \mathbf{A}\mathbf{x}_i + \mathbf{w}_i \in \mathbb{C}^m,\ i = 1, 2, \ldots, T, \tag{2.1}$$

where $\mathbf{A} \in \mathbb{C}^{m \times N} (m < N)$ is the measurement matrix and $\mathbf{w}_i \overset{i.i.d.}{\sim} \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I}_m)$ denotes the measurement noise corresponding to each measurement vector $\mathbf{y}_i$. By stacking the vectors $\mathbf{y}_i$ as columns of a matrix $\mathbf{Y}$, (2.1) can be rewritten as:

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{W} \in \mathbb{C}^{m \times T}. \tag{2.2}$$

The above model presents a relation between the input signal $\mathbf{X} \in \mathbb{C}^{N \times T}$ and the corresponding noisy measurements $\mathbf{Y}$. Since the total number of measurements ($mT$) is less than the signal size ($NT$), unique recovery of $\mathbf{X}$ from $\mathbf{Y}$ is not possible. However, the joint sparsity pattern in $\mathbf{X}$, along with other structures such as correlation, can be exploited to uniquely recover $\mathbf{X}$ from the above low-dimensional noisy measurement matrix. In the



(a) (b) (c) (d)

Fig. 2.1. Generative model for sparse signals (a) MMV model, (b) MMV with temporal correlation, (c) Block sparse, and (d) Block sparse with intra-block correlation. Similar colors in different shades in contiguous locations indicate the presence of correlation.

following, we present a generative model for the input signal $\mathbf{X}$ that can help capture the structures present in the signal.

### 2.1.2 Generative Model for Correlated Sparse Signals

Observations in the real world such as images, sound measurements, channel measurements, etc tend to be structured. One approach to model these structures has been the use of generative models. A generative model includes the distribution of the data along with the relation between the input and the output.

Compressed sensing deals with the recovery of sparse signals, i.e., most of the entries of the signal are zero. An extension to the MMV setup is the joint sparsity setup, sometimes called as the row sparse model, which forces a common sparsity support onto the multiple vectors being measured.

$$\text{supp}(\mathbf{x}_i) = \mathcal{S},\ i = 1, 2, \ldots, T,\ \text{with } |\mathcal{S}| = K < N. \tag{2.3}$$

A pictorial representation of the input signal under this model is given in Fig. 2.1a. It can be observed that there is no systematic relationship between the nonzero entries of the sparse vector, implying an absence of correlation across the sparse vectors. Generative

(a) (b)

Fig. 2.2. Covariance matrix for sparse signals with (a) Independent entries and (b) Correlated entries.

models for this input is obtained by forcing an i.i.d. prior with a common support across the vectors. To model existing knowledge of correlation in the model, non i.i.d. generative models are required. Some of the existing models are presented in Fig. 2.1, which include:

1. MMV setup with temporal correlation [13]. Here, the covariance matrix of the sparse vector is diagonal, but the corresponding entries of different jointly sparse vectors are correlated.
2. Block sparsity with and without intra-block correlation [9]. Here, the covariance matrix can either be diagonal or of block diagonal form.

In general, correlation can exist among all the nonzero entries of each sparse vector $\mathbf{x}$. The correlation among the entries can be modeled by a correlation matrix $\mathbf{U} \in \mathbb{R}^{N \times N}$, where $(\mathbf{U})_{ij}$ denotes the correlation coefficient between the $x_i$ and $x_j$. Various correlation matrix design such as uniform correlation, exponential correlation, Toeplitz correlation matrix, have been considered in literature. For the simulations in the thesis, we will concentrate mainly on a simple single parameter correlation matrix model called the Uniform correlation model [22], to elucidate the role of both sparsity and correlation in signal recovery.

$$(\mathbf{U})_{ij} = \begin{cases} 1, & i = j \\ \rho, & i \neq j \end{cases}, \tag{2.4}$$

where the parameter $\rho \in [0, 1)$ is called the correlation coefficient.

For the correlated sparse recovery problem, we start with a general statistical model for the signal, where all nonzero entries of the sparse representation are assumed to be correlated with each other. The $(i, j)^{\text{th}}$ entry of the covariance matrix $\mathbf{\Sigma_x} = \mathbb{E}\left[\mathbf{x}\mathbf{x}^H\right] \in \mathbb{C}^{N\times N}$ can be obtained from the Pearson product-moment correlation coefficient definition [23] as $[\mathbf{\Sigma_x}]_{(i,j)} = (\mathbf{U})_{ij}\sqrt{\gamma_i^*}\sqrt{\gamma_j^*}$ where $\gamma_i^*$[1] denotes the variance of $i^{\text{th}}$ entry of $\mathbf{x}$ and $(\mathbf{U})_{ij}$ is the correlation coefficient between the two entries, which is a function of the pair of indices $(i, j)$ and is governed by a known correlation model. With this notation, we have $\mathbf{\Sigma_x} = (\mathbf{\Gamma}^*)^{1/2}\mathbf{U}(\mathbf{\Gamma}^*)^{1/2}$ where $\mathbf{\Gamma}^* = \text{diag}(\gamma_1^*, \gamma_2^*, \ldots, \gamma_N^*) \in \mathbb{R}_+^N$. Note that when the index $j$ does not belong to the support, the variance of the $j^{\text{th}}$ entry, $\gamma_j^*$, is zero. Thus, $\mathbf{\Sigma_x}$ is a $N \times N$ positive semi-definite (PSD) matrix which contains a nonzero $K \times K$ positive definite (PD) principal submatrix corresponding to the index set $\mathcal{S}$, with its other entries equal to 0. The covariance matrix of this model as compared to that of i.i.d. model is shown in Fig. 2.2. We assume that in the MMV setup the columns of the signal $\mathbf{X}$ are obtained independent of each other, but the covariance matrix of all the vectors are equal to $\mathbf{\Sigma_x}$.

In the next section, we present a framework called covariance matching for the above system model and study the effect of correlation for the algorithms in the framework.

## 2.2 Covariance Matching Algorithms

Covariance matching approach is a popular methodology in the array processing literature. This approach is well suited to exploit correlation, as they estimate the covariance as a

[1]The asterisk notation (*) is used to denote the true value of the variances.

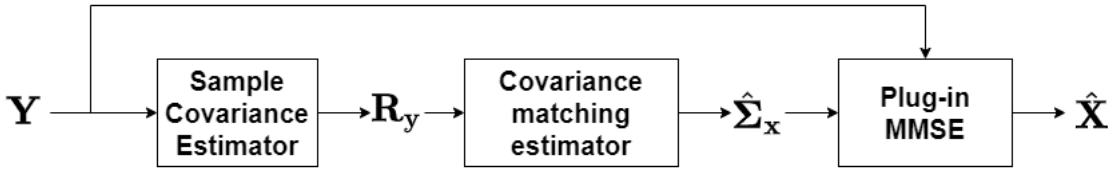



Fig. 2.3. Signal recovery using covariance matching framework

key step. Consider the covariance matrix of the input signal $\mathbf{\Sigma_x}$ and the covariance of measurements $\mathbf{\Sigma_y} = \mathbb{E}[\mathbf{y}\mathbf{y}^H] \in \mathbb{C}^{m\times m}$. These matrices are related by the covariance matching criterion given by,

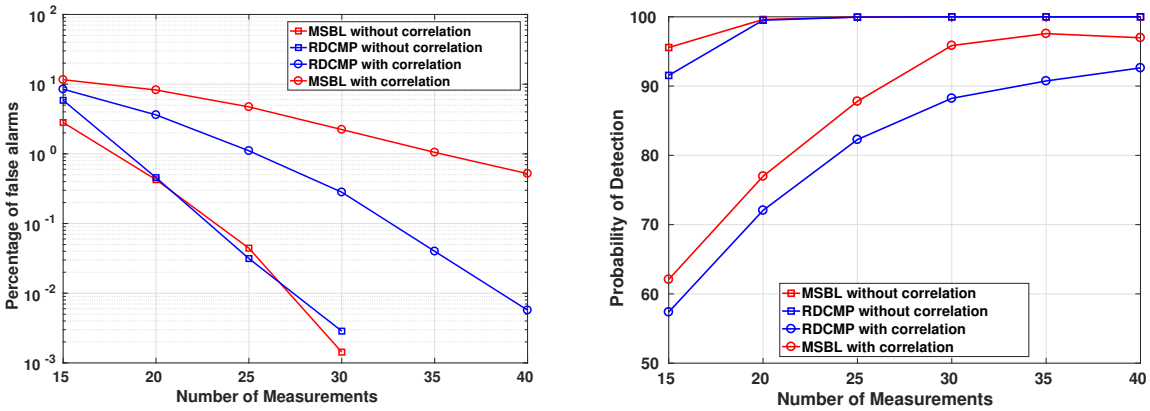
$$\mathbf{\Sigma_y} = \mathbf{A}\mathbf{\Sigma_x}\mathbf{A}^H + \sigma_n^2\mathbf{I}_m. \tag{2.5}$$

The covariance matching framework aims at recovering $\mathbf{\Sigma_x}$ by minimizing a distance measure $d(.,.)$ between the empirical covariance matrix $\hat{\mathbf{R}}_\mathbf{y} \triangleq \frac{1}{T}\mathbf{Y}\mathbf{Y}^H$ and the covariance matrix induced by the generative model (2.5):

$$\hat{\mathbf{\Sigma}}_\mathbf{x} = \arg\min_{\mathbf{\Sigma_x}} d\left(\mathbf{R_y}, \mathbf{A}\mathbf{\Sigma_x}\mathbf{A}^H + \sigma_n^2\mathbf{I}_m\right)$$

The covariance estimate $\hat{\mathbf{\Sigma}}_\mathbf{x}$ is then plugged in to a minimum mean square estimator (MMSE) to obtain an estimate for the input signal, as illustrated in Fig. 2.3.

An interpretation of MSBL as a covariance matching algorithm was presented in [24], and a general framework for this approach was presented which has the potential to spawn new MMV sparse recovery algorithms. In [25], it was shown that the extension of basis pursuit to the MMV setup can be decoupled as a covariance matching algorithm. Other algorithms based on this framework are Co-Lasso [26], RDCMP [27] among others. However, it is important to note that these algorithms assume the absence of correlation among the nonzero entries of the sparse vectors.




(a) False alarm rate (b) Detection probability

Fig. 2.4. Support recovery performance of MSBL and RDCMP

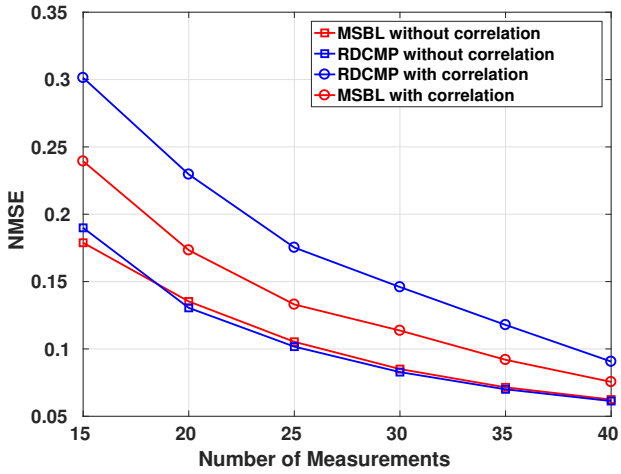



Fig. 2.5. NMSE performance of MSBL and RDCMP

### 2.2.1 Effect of Intra-vector Correlation

We compare the performance of two covariance matching based algorithms; MSBL and RDCMP, with and without the presence of correlation. We set the signal length $N = 100$, the number of MMV vectors $T = 200$ and the sparsity level $K = 30$. The number of measurements $m$ is varied from $15 - 40$. Noise variance is chosen as $\sigma_n^2 = 0.1$, which corresponds to an SNR of 10 dB when the signal power is unity. The support of the sparse vectors is obtained by drawing $K$ samples from the $N$ points uniformly at random

without replacement. The MMV sparse vectors are obtained i.i.d. from a complex normal distribution with covariance model $\mathbf{\Sigma_x}$ of the form described in Section 2.1.2.

Consider $\mathcal{S}$ to be the original support, and the recovered signal is $\mathbf{X}_{\text{out}}$ with support $\mathcal{S}_{\text{out}}$. The performance is measured under three metrics,

- False Alarm percentage= $\frac{|\mathcal{S}_{\text{out}}|-|\mathcal{S}\cap\mathcal{S}_{\text{out}}|}{N-K} \times 100$
- Probability of Detection (%)= $\frac{|\mathcal{S}\cap\mathcal{S}_{\text{out}}|}{K} \times 100$
- NMSE = $\frac{\|\mathbf{X}-\mathbf{X}_{\text{out}}\|^2}{\|\mathbf{X}\|^2}$

From Fig. 2.4, it can be observed that false alarm is higher and the detection probability is lower for correlated signals when compared to uncorrelated signals for both RDCMP and MSBL. The RDCMP algorithm which concentrates specifically on support recovery has better performance in terms of support recovery when compared to MSBL. However, it can be observed in Fig. 2.5 that the performance loss in terms of NMSE is less in MSBL when compared to RDCMP. It is also noteworthy, that the performance hit due to correlation is significant mainly in the highly measurement-constrained region ($K \geq M$).

The above results showed that the existing algorithms can perform poorly when there is high intra-vector correlation. Hence, it is important to exploit correlation to obtain better recovery performance. To this end, we concentrate on the Bayesian learning framework which can be interpreted as a covariance matching approach for the correlated sparse signal recovery problem.

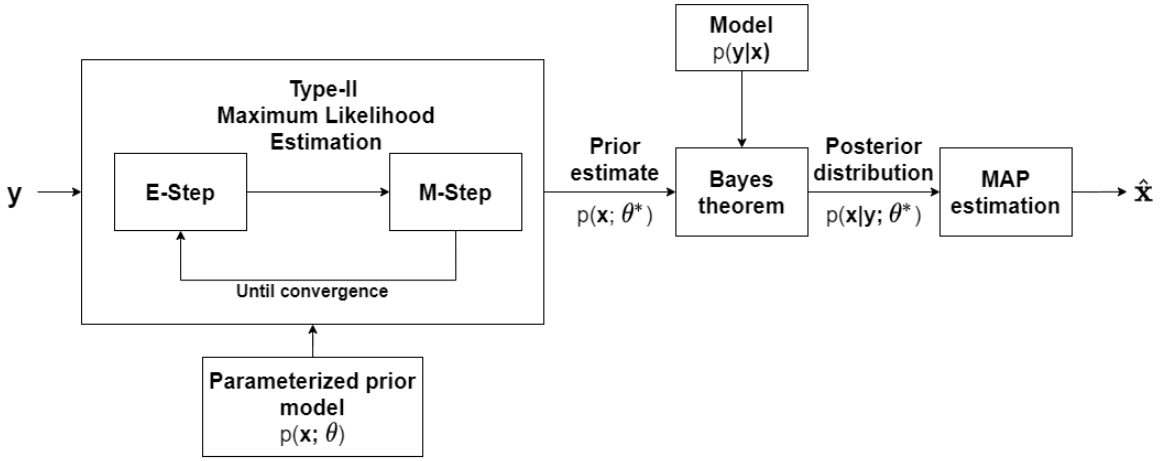



Fig. 2.6. Bayesian learning framework

## 2.3 Sparse Bayesian Learning

In this section, we present an overview of the Bayesian framework used to derive an algorithm for correlated sparse recovery.

### 2.3.1 Bayesian Learning Methodology

Bayesian learning is a statistical inference methodology that uses Bayes' theorem to determine the conditional probability of a hypothesis given some observations. A block diagram depicting the framework is presented in Fig. 2.6. In this methodology, a parameterized fictitious prior $p(\mathbf{x};\theta)$ with parameter $\theta$ is imposed on the signal of interest. The aim of the Bayesian inference procedure is to learn the prior parameters such that the resulting distribution best fits the observed data according to the underlying measurement model. Using the posterior distribution obtained using the Bayes theorem for the prior estimate, estimates for the signals or parameters of the signal can be obtained, for example using the maximum a posteriori estimate (MAP estimate). The two important design considerations in the framework are:

1. Parameterized prior model: The choice of the prior is crucial to the success of the algorithm. The prior must be designed to promote the structure in the signal and at the same time facilitate the computation of the posterior.

2. Inference procedure: In general, estimating the optimal value of the prior parameters $\theta$ is not straightforward. A popular method is the type-II maximum likelihood estimation procedure. The design of the inference procedure plays a major role in both accuracy and complexity of the algorithm.

### 2.3.2 Prior Design in SBL

Sparse Bayesian learning (SBL) is a popular algorithm for the sparse recovery problem derived based on the Bayesian learning framework. The crux of this algorithm is the use of a parameterized Gaussian distribution for the prior model. Mathematically, a parameterized Gaussian prior with the covariance matrix $\mathbf{\Sigma}_\theta \in \mathbb{C}^{N\times N}$ is modeled as $\mathbf{\Sigma}_\theta = \text{diag}(\boldsymbol{\gamma})$, where the unknown hyperparameter $\theta$ consists of the entries of the vector $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \ldots, \gamma_N]^T \in \mathbb{R}_+^N$ which denote the variances of the entries of $\mathbf{x}$. This choice of prior model is known to induce sparsity in the final estimate. The diagonal nature of the covariance matrix implicitly assumes that the entries of $\mathbf{x}$ are uncorrelated. Hence, this prior cannot accommodate any knowledge of the correlation structure in $\mathbf{x}$.

By modifying the parameters of the prior design, various extensions of SBL have been considered to exploit additional structures in the sparse signal. TMSBL [13] was developed for the MMV model with or without temporal correlation, BSBL [9] and PCSBL [28] for block sparse signals. In the following section, we present a novel choice of prior to promote intra-vector correlation and develop a SBL based algorithm on similar lines to the above

algorithms. To the best of our knowledge, there is no existing algorithm based on Bayesian inference, which considers possible correlation among the nonzero entries of a sparse vector.

## 2.4 Corr-SBL Algorithm

In this section, we present the derivation of the Corr-SBL algorithm for the single measurement vector (SMV) case, where the goal is to recover a sparse vector $\mathbf{x} \in \mathbb{C}^N$ from its noisy measurements $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w} \in \mathbb{C}^m$. The extension to the MMV setup is simple, and will be discussed in Section 3.3 while presenting an application of Corr-SBL.

### 2.4.1 Prior Model

Similar to SBL, in this chapter, we consider a parameterized Gaussian prior for $\mathbf{x}$ with hyperparameter set $\theta$ as

$$p(\mathbf{x};\theta) = \mathcal{CN}(\mathbf{x};\mathbf{0},\mathbf{\Sigma}_\theta). \tag{2.6}$$

Recall that the true covariance matrix $\mathbf{\Sigma_x}$ is a $N \times N$ positive semidefinite matrix with a $K \times K$ positive definite principal submatrix corresponding to $K$ entries in the support, and other entries equal to zero. In order to incorporate the above structure, the covariance matrix $\mathbf{\Sigma}_\theta$ induced by the prior is modeled using two hyperparameter sets: $\{\boldsymbol{\gamma} \in \mathbb{R}_+^N\}$ denoting the vector containing unknown variances of the entries of $\mathbf{x}$, and $\mathbf{U} \in \mathbb{C}^{N\times N}$ denoting a positive definite matrix with correlation coefficients between the entries of $\mathbf{x}$. For ease of exposition, we deal with the hyperparameter $\mathbf{U}$ in two cases: (i) when $\mathbf{U}$ is known; and (ii) when $\mathbf{U}$ is an unknown matrix parameterized based on the

underlying channel model.[2] Using the definition for Pearson product-moment correlation coefficient [23], $\mathbf{\Sigma}_{\boldsymbol{\gamma},\mathbf{U}} \triangleq \mathbf{\Sigma}_\theta$ is modeled as

$$\mathbf{\Sigma}_{\boldsymbol{\gamma},\mathbf{U}} = \mathbf{\Gamma}^{\frac{1}{2}}\mathbf{U}\mathbf{\Gamma}^{\frac{1}{2}}, \tag{2.7}$$

where $\mathbf{\Gamma} \triangleq \text{diag}(\boldsymbol{\gamma})$. To see how the above model incorporates the structure in $\mathbf{\Sigma}_\mathbf{x}$, consider the case where $i^{\text{th}}$ element of $\boldsymbol{\gamma}$ is zero. Since it denotes the variance of a zero mean random variable, $i^{\text{th}}$ entry of $\mathbf{x}$ is zero, and the $i^{\text{th}}$ row and column of $\mathbf{\Sigma}_\theta$ are zero vectors. If *all but $K$ entries* are zero, then $\mathbf{\Sigma}_\theta$ has a structure similar to $\mathbf{\Sigma}_\mathbf{x}$. Consequently, the MAP estimate for $\mathbf{x}$ is also an $K$-sparse signal, with the same support of $\boldsymbol{\gamma}$. Also, the prior model is a generalization to SBL, as it reduces to the SBL prior $\mathbf{\Sigma}_{\boldsymbol{\gamma}} = \mathbf{\Gamma}$ when $\mathbf{U} = \mathbf{I}_N$.

In the sequel, it will be more convenient to work with the precision matrix $\mathbf{\Omega}_{\mathbf{c},\mathbf{U}} \triangleq \mathbf{\Sigma}_{\boldsymbol{\gamma},\mathbf{U}}^{-1}$. Note that $\mathbf{\Omega}_{\mathbf{c},\mathbf{U}} = \mathbf{C}\mathbf{U}^{-1}\mathbf{C}$, where $\mathbf{C} \triangleq \text{diag}(\mathbf{c})$ and $\mathbf{c} \in \mathbb{R}_+^N$ has $1/\sqrt{\gamma_i}$ as its $i^{\text{th}}$ entry. In the following section, we present an iterative algorithm to estimate the value of $c_i$. When an index $i$ is not in the support, the value of $c_i$ goes to infinity as the iterations proceed. To counter the numerical instability, in any iteration, if the value of $c_i$ exceeds a predetermined threshold $\varepsilon$, we remove the index $i$ and the corresponding column of $\mathbf{A}$ from the support. This also speeds up the algorithm.

### 2.4.2 Algorithm Derivation

We now proceed with developing a Bayesian algorithm for the proposed choice of prior when the correlation matrix $\mathbf{U}$ is known, and drop the subscript $\mathbf{U}$ in $\mathbf{\Omega}_{\mathbf{c},\mathbf{U}}$. This involves two steps: obtaining the optimal value for the hyperparameters $\mathbf{c}$, and computing the

[2] An algorithm is derived for the first case in Sec. 2.4.2, while a pragmatic procedure is presented in Sec. 2.4.3 to estimate the parameters of $\mathbf{U}$.

posterior distribution.

We use type-II ML estimation to obtain the optimal value of $\mathbf{c}$. This is based on the evidence maximization framework, where the cost function is the marginal likelihood of $\mathbf{y}$.

$$\mathbf{c}_{\mathrm{ML}} = \arg\max_{\mathbf{c}} p(\mathbf{y}; \mathbf{c}, \sigma_n^2). \tag{2.8}$$

By marginalizing the joint density $p(\mathbf{x}, \mathbf{y}; \mathbf{c}, \sigma_n^2)$ with respect to $\mathbf{x}$, it is straightforward to show that the marginal likelihood is given by $p(\mathbf{y}; \mathbf{c}, \sigma_n^2) = \mathcal{CN}(\mathbf{y}; \mathbf{0}, \mathbf{\Omega}_{\mathbf{y}}^{-1})$, where $\mathbf{\Omega}_{\mathbf{y}} \triangleq \left[\sigma_n^2 \mathbf{I}_m + \mathbf{A}\mathbf{\Omega}_{\mathbf{c}}^{-1}\mathbf{A}^H\right]^{-1}$ denotes the precision matrix of $\mathbf{y}$. Thus, the cost function that needs to be maximized for finding $\mathbf{c}$ is obtained from the log likelihood $\log(p(\mathbf{y}; \mathbf{c}, \sigma_n^2))$ as

$$L(\mathbf{c}) \triangleq \mathrm{logdet}(\mathbf{\Omega}_{\mathbf{y}}) - \mathbf{y}^H \mathbf{\Omega}_{\mathbf{y}} \mathbf{y}. \tag{2.9}$$

The optimal $\mathbf{c}$ is then used to compute the posterior distribution and the estimate using the following Lemma [29, Theorem 10.3].

**Lemma 1.** *Let the prior distribution on* $\mathbf{x}$ *be modeled as* $p(\mathbf{x}; \mathbf{c}) = \mathcal{CN}(\mathbf{x}; \mathbf{0}, \mathbf{\Omega}_{\mathbf{c}}^{-1})$. *Then, the posterior distribution of* $\mathbf{x}$ *given the observation* $\mathbf{y}$ *and hyperparameter* $\mathbf{c}$, $p(\mathbf{x}|\mathbf{y}; \mathbf{c}) = \mathcal{CN}(\mathbf{x}; \boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}}, \mathbf{\Omega}_{\mathbf{x}|\mathbf{y}})$, *where*

$$\mathbf{\Omega}_{\mathbf{x}|\mathbf{y}} = \frac{1}{\sigma_n^2}\mathbf{A}^H\mathbf{A} + \mathbf{\Omega}_{\mathbf{c}}; \quad \boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}} = \frac{1}{\sigma_n^2}\mathbf{\Omega}_{\mathbf{x}|\mathbf{y}}^{-1}\mathbf{A}^H\mathbf{y}. \tag{2.10}$$

*Proof.* The posterior distribution of $\mathbf{x}$ given the observations $\mathbf{y}$ and hyperparameter value $\mathbf{c}$ is given by

$$p(\mathbf{x}|\mathbf{y}; \mathbf{c}, \sigma_n^2) = \frac{p(\mathbf{y}|\mathbf{x}; \sigma_n^2)p(\mathbf{x}; \mathbf{c})}{p(\mathbf{y}; \mathbf{c}, \sigma_n^2)}$$

Using the prior on $\mathbf{x}:\ \mathcal{CN}(\mathbf{x};\mathbf{0},\mathbf{\Omega_c}^{-1})$, we get

$$\begin{aligned} p(\mathbf{x}|\mathbf{y};\mathbf{c},\sigma_n^2) &= \frac{\mathcal{CN}(\mathbf{y};\mathbf{Ax},\sigma_n^2\mathbf{I}_m)\ \mathcal{CN}(\mathbf{x};\mathbf{0},\mathbf{\Omega_c}^{-1})}{\mathcal{CN}(\mathbf{y};\mathbf{0},\mathbf{\Omega}_y^{-1})} \\ &= k\exp\left(-\mathbf{x}^H(\mathbf{\Omega_c}+\frac{\mathbf{A}^H\mathbf{A}}{\sigma_n^2})\mathbf{x}+\frac{\mathbf{x}^H\mathbf{A}^H\mathbf{y}+\mathbf{y}^H\mathbf{Ax}}{\sigma_n^2}\right), \end{aligned}$$

where $k$ is a normalization constant. Using $\mathbf{\Omega_{x|y}} = \mathbf{\Omega_c} + \frac{1}{\sigma_n^2}\mathbf{A}^H\mathbf{A}$ and $\boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}} = \frac{1}{\sigma_n^2}\mathbf{\Omega}_{\mathbf{x}|\mathbf{y}}^{-1}\mathbf{A}^H\mathbf{y}$ and completing the squares, the posterior distribution can be written as

$$p(\mathbf{x}|\mathbf{y};\mathbf{c},\sigma_n^2) = k\ \exp\left(-\left(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}}\right)^H\mathbf{\Omega_{x|y}}\left(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}}\right)\right),$$

which is the complex Gaussian distribution with mean $\boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}}$ and covariance $\mathbf{\Omega}_{\mathbf{x}|\mathbf{y}}^{-1}$, as given in the statement of the Lemma. □

Since the posterior distribution of $\mathbf{x}$ is Gaussian, its mode (i.e., the MAP estimate) is the same as its mean. Hence, the posterior mean $\boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}}$ computed using the optimal value of the hyperparameters $\mathbf{c}$ is the estimate for the sparse signal. Note that (2.10) involves inverting the $N \times N$ matrix $\mathbf{\Omega_{x|y}}$. The complexity can be reduced from $\mathcal{O}(N^3)$ to $\mathcal{O}(m^3)$ by using the Woodbury matrix identity, which also speeds up the algorithm.

The problem of maximizing the cost function (2.9) is non-convex and does not admit a closed form solution. Hence, we use the expectation-maximization (EM) procedure to maximize (2.9) by treating $\mathbf{x}$ as a latent variable. The EM procedure involves iterating between an expectation step (E-step) and a maximization step (M-step) [30, Section 9.3].

### Expectation Step

This step involves computing the expected value of the complete-data log likelihood with respect to the posterior distribution for $\mathbf{x}$ computed at the hyperparameter value $\mathbf{c}_{\text{old}}$

obtained from previous iteration of the EM algorithm. The expected value is denoted by the so-called $Q$ function[3], which is defined as follows:

$$Q(\mathbf{c}, \mathbf{c}_{\text{old}}) \triangleq \mathbb{E}_{\mathbf{x}|\mathbf{y};\mathbf{c}_{\text{old}},\sigma_n^2} \left[\log(p(\mathbf{x}, \mathbf{y}; \mathbf{c}, \sigma_n^2))\right]. \tag{2.11}$$

The posterior distribution $p(\mathbf{x}|\mathbf{y}; \mathbf{c}_{\text{old}}, \sigma_n^2)$ is computed using Lemma 1. Using this distribution, the Q function is computed as given by the following theorem.

**Theorem 1.** *The expected value of complete-data log likelihood evaluated using the hyper-parameter value* $\mathbf{c}_{old}$ *corresponding to the cost function* $L(\mathbf{c}) = \text{logdet}(\mathbf{\Omega}_\mathbf{y}) - \mathbf{y}^H\mathbf{\Omega}_\mathbf{y}\mathbf{y}$ *is given by*

$$Q(\mathbf{c}, \mathbf{c}_{old}) = \ k' + (\text{logdet}(\mathbf{\Omega}_\mathbf{c})) - Tr\left[\mathbf{\Omega}_\mathbf{c}\hat{\mathbf{R}}_\mathbf{x}\right], \tag{2.12}$$

*where* $k'$ *is a constant independent of* $\mathbf{c}$*, and* $\hat{\mathbf{R}}_\mathbf{x} \triangleq \left[\mathbf{\Omega}_{\mathbf{x}|\mathbf{y}}^{-1} + \boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}}\boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}}^H\right]$.

*Proof.* The lower bound $Q$ on the cost function $L$ using the EM framework is given as

$$\begin{aligned} Q(\mathbf{c}, \mathbf{c}_{\text{old}}) &= \mathbb{E}_{\mathbf{x}|\mathbf{y};\mathbf{c}_{\text{old}},\sigma_n^2} \left[\log(p(\mathbf{x}, \mathbf{y}; \mathbf{c}, \sigma_n^2))\right] \\ &= \mathbb{E}_{\mathbf{x}|\mathbf{y},\mathbf{c}_{\text{old}}} \left[\log(p(\mathbf{y}|\mathbf{x}; \sigma_n^2))\right] + \mathbb{E}_{\mathbf{x}|\mathbf{y},\mathbf{c}_{\text{old}}} \left[\log(p(\mathbf{x}; \mathbf{c}))\right]. \end{aligned} \tag{2.13}$$

The first expectation term is a constant independent of $\mathbf{c}$, which does not affect the M-step. The second expectation is computed as

$$\begin{aligned} \mathbb{E}_{\mathbf{x}|\mathbf{y},\mathbf{c}_{\text{old}}}[\log(p(\mathbf{x}; \mathbf{c}))] &= \mathbb{E}_{\mathbf{x}|\mathbf{y},\mathbf{c}_{\text{old}}} \left[\log(\mathcal{CN}(\mathbf{x}; \mathbf{0}, \mathbf{\Omega}_\mathbf{c}^{-1}))\right] \\ &= -N\log(\pi) + \log(\det(\mathbf{\Omega}_\mathbf{c})) - \mathbb{E}_{\mathbf{x}|\mathbf{y},\mathbf{c}_{\text{old}}} \left[\mathbf{x}^H\mathbf{\Omega}_\mathbf{c}\mathbf{x}\right]. \end{aligned} \tag{2.14}$$

The scalar term $\mathbf{x}^H\mathbf{\Omega}_\mathbf{c}\mathbf{x}$ can be rewritten using the trace operator. Using the product

[3]The $Q$ function used here is not the popularly used Gaussian tail probability $Q$-function.

property of trace ($\text{Tr}[\mathbf{AB}] = \text{Tr}[\mathbf{BA}]$)), the expectation term in (2.14) can be rewritten as

$$\mathbb{E}_{\mathbf{x}|\mathbf{y},\mathbf{c}_{\text{old}}}\left[\mathbf{x}^H\boldsymbol{\Omega}_{\mathbf{c}}\mathbf{x}\right] = \text{Tr}\left[\boldsymbol{\Omega}_{\mathbf{c}}\mathbb{E}_{\mathbf{x}|\mathbf{y},\mathbf{c}_{\text{old}}}\left[\mathbf{x}\mathbf{x}^H\right]\right].$$

The expectation term is the second moment matrix with respect to the posterior probability distribution. Using Lemma 1, the second moment matrix is obtained as $\hat{\mathbf{R}}_{\mathbf{x}} \triangleq \mathbb{E}_{\mathbf{x}|\mathbf{y},\mathbf{c}_{\text{old}}}\left[\mathbf{x}\mathbf{x}^H\right] = \boldsymbol{\Omega}_{\mathbf{x}|\mathbf{y}}^{-1} + \boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}}\boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}}^H$. Substituting this into (2.13), the $Q$ function is given by

$$Q(\mathbf{c}, \mathbf{c}_{\text{old}}) = \text{constant} + \log(\det(\boldsymbol{\Omega}_{\mathbf{c}})) - \text{Tr}\left[\boldsymbol{\Omega}_{\mathbf{c}}\hat{\mathbf{R}}_{\mathbf{x}}\right].$$

□

**Maximization Step**

In this step, the hyperparameter $\mathbf{c}$ is updated by maximizing the $Q$ function. The first order optimality condition for the stationary points of $Q(\mathbf{c}, \mathbf{c}_{\text{old}})$ is given by the following theorem.

**Theorem 2.** *The first order optimality condition for the optimization problem* $\mathbf{c}^* = \arg\max_{\mathbf{c}}\ Q(\mathbf{c}, \mathbf{c}_{old})$ *is given by*

$$\frac{1}{c_i} = \text{Re}\left\{\sum_{k=1}^{N} c_k[\mathbf{U}^{-1}]_{(i,k)}[\hat{\mathbf{R}}_{\mathbf{x}}]_{(k,i)}\right\},\ i \in [N], \tag{2.15}$$

*Proof.* Using the first order optimality condition for maximizing $Q(\mathbf{c}, \mathbf{c}_{\text{old}})$,

$$\frac{\partial Q}{\partial c_i} = 0,\ i \in \{1, 2, \ldots N\}, \tag{2.16}$$

and from Theorem 1, we get

$$\frac{\partial \left( \log(\det(\mathbf{\Omega_c})) - \mathrm{Tr}\left[ \mathbf{\Omega_c} \hat{\mathbf{R}}_\mathbf{x} \right] \right)}{\partial c_i} = 0. \tag{2.17}$$

The first term is

$$\frac{\partial \log(\det(\mathbf{\Omega_c}))}{\partial c_i} = \frac{\partial \log(\det(\mathbf{C}\mathbf{U}^{-1}\mathbf{C}))}{\partial c_i} = \frac{2}{c_i}. \tag{2.18}$$

Considering the second term,

$$\begin{aligned} \frac{\partial \left( \mathrm{Tr}\left[ \mathbf{\Omega_c} \hat{\mathbf{R}}_\mathbf{x} \right] \right)}{\partial c_i} &= \frac{\partial \left( \mathrm{Tr}\left[ \mathbf{C}\mathbf{U}^{-1}\mathbf{C} \hat{\mathbf{R}}_\mathbf{x} \right] \right)}{\partial c_i} \\ &= \mathrm{Tr}\left[ \mathbf{J}_{ii}\mathbf{U}^{-1}\mathbf{C}\hat{\mathbf{R}}_\mathbf{x} + \hat{\mathbf{R}}_\mathbf{x}\mathbf{C}\mathbf{U}^{-1}\mathbf{J}_{ii} \right], \end{aligned} \tag{2.19}$$

where $\mathbf{J}_{ii}$ an $N \times N$ matrix with a single 1 in the $i^{\text{th}}$ diagonal entry and zeros elsewhere. Utilizing the fact that the two matrices inside the trace expression are Hermitian transpose of each other and using the property of single entry matrices, the term is simplified as,

$$\frac{\partial \left( \mathrm{Tr}\left[ \mathbf{\Omega_c} \hat{\mathbf{R}}_\mathbf{x} \right] \right)}{\partial c_i} = 2\,\mathrm{Re}\left\{ \left( \mathbf{U}^{-1}\mathbf{C}\hat{\mathbf{R}}_\mathbf{x} \right)_{ii} \right\}. \tag{2.20}$$

Substituting (2.18) and (2.20) in (2.17), and simplifying expressions, the optimality condition is obtained as

$$\frac{1}{c_i} = \mathrm{Re}\left\{ \sum_{k=1}^{N} c_k [\mathbf{U}^{-1}]_{(i,k)} [\hat{\mathbf{R}}_\mathbf{x}]_{(k,i)} \right\}. \tag{2.21}$$

□

In the conventional sparse Bayesian learning algorithm, the optimality condition above decouples into separate equations in each hyperparameter $c_i$, and the update for $c_i$ can be obtained independent of $c_j,\ j \neq i$. However, for the prior considered here, the optimality

condition in (2.15) is a coupled quadratic equation, which cannot be solved in closed form. Gradient based methods can be used to search for the solution, but the computational complexity involved is large. Instead, we draw from the generalized EM theory [31]: instead of achieving the maximum of the $Q$ function, any hyperparameter update rule which ensures that $Q(\mathbf{c}, \mathbf{c}_{\text{old}})$ is non-decreasing in each EM iteration will lead to convergence of the EM iterations to a local maximum or saddle point of $L(\mathbf{c})$.

To this end, we consider a vector representation for (2.15) as $\mathbf{c} = (\text{Re}\{\mathbf{K}\})^{-1} \frac{1}{\mathbf{c}}$, where $\mathbf{K} \triangleq \mathbf{U}^{-1} \odot \hat{\mathbf{R}}_{\mathbf{x}}^T \in \mathbb{C}^{N \times N}$ and $\frac{1}{\mathbf{c}}$ denotes the element-wise inverse. A single iteration of this fixed point equation results in a non-decreasing cost function value, as asserted by the following proposition.

**Proposition 1.** *Consider the update for the hyperparameter given by*

$$\mathbf{c}_{new} = (\text{Re}\{\mathbf{K}\})^{-1} \frac{1}{\mathbf{c}_{old}}. \tag{2.22}$$

*This satisfies the condition* $\sum_i \left( \frac{dQ}{dc_i} ((c_{new})_i - (c_{old})_i) \right) \geq 0$. *Consequently, by the generalized EM theory, the cost function in* (2.12) *does not decrease after the update. In turn, this guarantees the convergence of the overall algorithm.*

*Proof.* Let $s_{i,j}$ and $t_{i,j}$ denote the $\{i,j\}^{\text{th}}$ entry of the real symmetric matrix $\mathbf{S} \triangleq \text{Re}\{\mathbf{K}\}$ and its inverse $\mathbf{T}$, respectively. These follow the condition $\sum_{i=1}^N s_{i,k} t_{i,l} = \sum_{i=1}^N s_{i,k} t_{l,i} = 1$ if $k = l$ and 0 otherwise. Also, since $\mathbf{U}^{-1}$ and $\hat{\mathbf{R}}_{\mathbf{x}}^T$ are positive definite, Re$\{\mathbf{K}\}$ is a positive definite matrix.

We need to show that the update in (2.22) results in $\sum_i \left( \frac{dQ}{dc_i} ((c_{\text{new}})_i - (c_{\text{old}})_i) \right)$ being greater than or equal to zero. This would in turn imply that the cost function is nondecreasing in each iteration. Using the expression for $\frac{dQ}{dc_i}$ derived in the proof of Theorem 2

**Algorithm 1:** `Corr-SBL procedure`

**Input**: $\mathbf{A}, \mathbf{y}, \mathbf{U}, \sigma_n^2$

**Initialize:** $\mathbf{c}$

**repeat**

  **E-Step:** Compute $\mathbf{\Omega}_{\mathbf{x}|\mathbf{y}}$, $\boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}}$ using (2.10) and $\hat{\mathbf{R}}_{\mathbf{x}}$ using (2.12)

  **M-Step:** Update $\mathbf{c}$ using (2.22)

**until** *Stopping Criterion*;

and the expression for $\mathbf{c}_{\text{new}}$ from (2.22), the above quantity can be written as

$$\begin{aligned}
&\sum_{i=1}^{N}\left(\left(\frac{2}{(c_{\text{old}})_i}-2\sum_{k=1}^{N}(c_{\text{old}})_k s_{i,k}\right)\left(\sum_{k=1}^{N}\frac{t_{i,k}}{(c_{\text{old}})_k}-(c_{\text{old}})_i\right)\right)\\
&=2\sum_{i=1}^{N}\left(\sum_{k=1}^{N}\frac{t_{i,k}}{(c_{\text{old}})_i(c_{\text{old}})_k}+s_{i,k}(c_{\text{old}})_i(c_{\text{old}})_k-1-\sum_{k,l=1}^{N}\frac{(c_{\text{old}})_k}{(c_{\text{old}})_l}s_{i,k}t_{i,l}\right)\\
&=2\sum_{i=1}^{N}\left(\sum_{k=1}^{N}\frac{t_{i,k}}{(c_{\text{old}})_i(c_{\text{old}})_k}+s_{i,k}(c_{\text{old}})_i(c_{\text{old}})_k-1\right)-2\sum_{k,l=1}^{N}\frac{(c_{\text{old}})_k}{(c_{\text{old}})_l}\sum_{i=1}^{N}s_{i,k}t_{i,l}.
\end{aligned}$$

Using the properties of $\mathbf{S}$ and $\mathbf{T}$, the last term is equal to 2. Further, by considering $\mathbf{B} = \mathbf{CSC}$, where $\mathbf{C}$ is a diagonal matrix with $(c_{\text{old}})_i$ as its entries and using the eigen decomposition of $\mathbf{B} = \mathbf{U\Lambda U}^T$, the above can be simplified as

$$2\mathbf{1}^T\left(\mathbf{B}+\mathbf{B}^{-1}-2\mathbf{I}\right)\mathbf{1} = 2\mathbf{1}^T\mathbf{U}\left(\mathbf{\Lambda}+\mathbf{\Lambda}^{-1}-2\mathbf{I}\right)\mathbf{U}^T\mathbf{1} \geq 0,$$

where $\mathbf{1}$ is the all ones vector of length $N$. The last inequality is because entries of $\mathbf{\Lambda}$ are positive since $\mathbf{B}$ is positive definite, and as a consequence $\mathbf{\Lambda}+\mathbf{\Lambda}^{-1}-2\mathbf{I}$ is a diagonal matrix with non-negative entries, hence is positive semi-definite. □

Hence, by iterating over the E-step and the hyperparameter update given by (2.22), the algorithm is guaranteed to converge to a local optimum. The pseudocode for the Corr-SBL algorithm is summarized in Algorithm 1. Initialization for $\mathbf{c}$ and the stopping criterion

can be chosen depending on the application.

In the preceding derivation, it was assumed that the noise variance $\sigma_n^2$ was known. In some applications, it is desirable for the algorithm to learn the noise variance also. This can be incorporated by learning a hyperparameter $\lambda \triangleq \sigma_n^2$. The E-step of the algorithm remains unchanged, while the M-step decouples into independent updates for $\mathbf{c}$ (given by (2.22)) and $\lambda$. The update for $\lambda$ is given by

$$\lambda = \frac{1}{m}\left[\mathbf{y}^H\mathbf{y} - \mathbf{y}^H\mathbf{A}\boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}} - \boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}}^H\mathbf{A}^H\mathbf{y} + \mathrm{Tr}\left[\mathbf{A}^H\mathbf{A}\boldsymbol{\Omega}_{\mathbf{x}|\mathbf{y}}^{-1}\right]\right].$$

### 2.4.3 Learning Correlation

The Bayesian inference discussed above assumed that correlation matrix $\mathbf{U}$ is known. One approach for estimating $\mathbf{U}$ could be to consider it as an additional hidden parameter in the EM algorithm, and use the optimality condition to obtain an update. However, due to the matrix derivatives involved, deriving a closed form update for $\mathbf{U}$ is not straightforward. Instead, we present a pragmatic approach for learning the correlation in this subsection.

In the previous section, we saw that the mean $\boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}}$ of the posterior distribution is the MAP estimate for $\mathbf{x}$ upon convergence of the algorithm. Similarly, the matrix $\hat{\mathbf{R}}_{\mathbf{x}} = \boldsymbol{\Omega}_{\mathbf{x}|\mathbf{y}}^{-1} + \boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}}\boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}}^H$ can be interpreted as an estimate for the covariance matrix $\boldsymbol{\Sigma}_{\mathbf{x}}$. In fact, SBL uses the diagonal entries of this estimate to update the variance hyperparameters in each iteration. For the correlated case considered in this paper, a similar update for the inverse variance entries $\mathbf{c}$ does not satisfy the first order optimality condition (2.15), and hence is not a viable choice for an update rule for $\mathbf{c}$ in the EM algorithm. Instead, an update for $\mathbf{U}$ can be obtained by using (2.7) to project $\hat{\mathbf{R}}_{\mathbf{x}}$ onto the space of correlation matrices. To this end, the estimate for $\boldsymbol{\Gamma}$ in (2.7), denoted by $\hat{\boldsymbol{\Gamma}}$, is obtained by considering

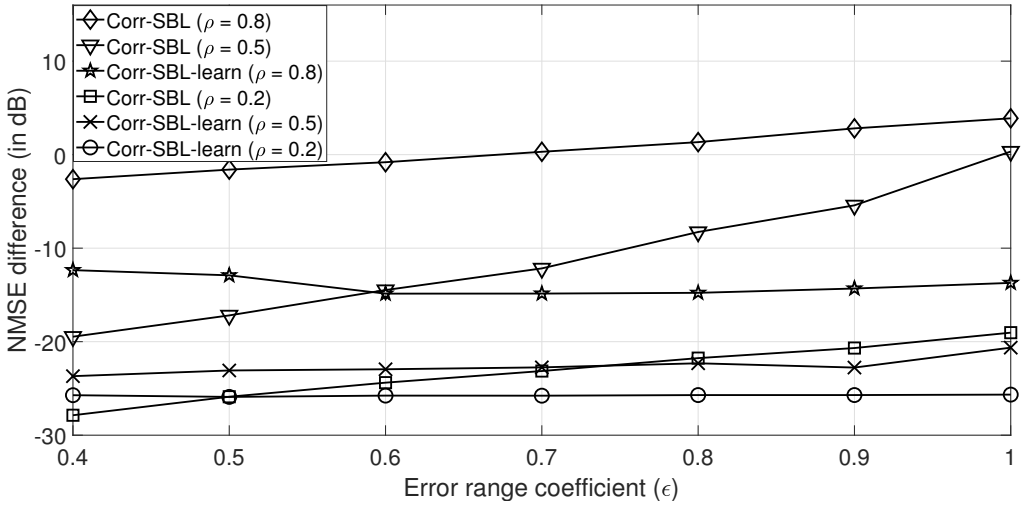



Fig. 2.7. Loss in NMSE performance with and without learning the correlation due to imperfect correlation information.

the diagonal entries of $\hat{\mathbf{R}}_{\mathbf{x}}$. An estimate for the correlation matrix, denoted by $\hat{\mathbf{U}}$, is then obtained by projecting onto space of correlation matrices with unit diagonal entries, given by

$$\hat{\mathbf{U}} = \hat{\mathbf{\Gamma}}^{-1/2}\hat{\mathbf{R}}_{\mathbf{x}}\hat{\mathbf{\Gamma}}^{-1/2}. \tag{2.23}$$

In case a parameterized model for $\mathbf{U}$ is available, e.g., if $\mathbf{U}$ is determined by a scalar parameter $\rho$, the above estimate can be used as a sufficient statistic to estimate $\rho$. In particular, for the uniform correlation model given in (2.4), $\rho \in [0, 1)$ and can be obtained by averaging the off-diagonal entries of $\hat{\mathbf{U}}$.

We term the algorithm for unknown $\mathbf{U}$ case as `Corr-SBL-learn`. To demonstrate the advantage of this algorithm for imperfect correlation information, we compare its performance with `Corr-SBL` when the exact correlation is not known. The simulation is performed for the uniform correlation model. Let $\rho$ be the correlation coefficient of the true channel. The initial value of the correlation coefficient $\hat{\rho}$ for both Corr-SBL and Corr-SBL-learn is chosen uniformly at random in the range $[(1-\epsilon)\rho, (1+\epsilon)\rho]$. The other

parameters for the setup are size of the signal $N = 256$, number of measurements $m = 16$ and sparsity level $K = 16$. The performance loss due to imperfect correlation is presented in Fig. 2.7, where we plot the difference between the NMSE of each of the algorithms initialized with the imperfect correlation $\hat{\rho}$ and the NMSE of genie-Corr-SBL which has the exact knowledge of $\rho$, as a function of the error range coefficient $\epsilon$. Since Corr-SBL continues with the initialization of the correlation coefficient, the performance deteriorates with increase in $\epsilon$. Corr-SBL-learn learns the correlation coefficient from the data, and exhibits relatively stable performance irrespective of $\epsilon$, and outperforms Corr-SBL with inaccurate knowledge of the correlation coefficient. The NMSE increases slightly with the size of the uncertainty interval, but the loss in NMSE is small even for high values of $\epsilon$.

### 2.4.4 Corr-SBL as a Covariance Matching Algorithm

The focus of first part of this chapter was the compressive covariance matching algorithms. The Corr-SBL algorithm was derived using the Bayesian methodology. The following result connects the two methodology by providing an alternative representation of the output of the Corr-SBL algorithm.

**Proposition 2.** *The output of the Corr-SBL algorithm is given by*

$$\hat{\mathbf{x}} = \mathbf{M}\mathbf{y} = \hat{\mathbf{\Sigma}}_{\mathbf{x}}\mathbf{A}^H \left(\mathbf{A}\hat{\mathbf{\Sigma}}_{\mathbf{x}}\mathbf{A}^H + \sigma_n^2\mathbf{I}_m\right)^{-1}\mathbf{y} \tag{2.24}$$

*with the covariance estimate* $\hat{\mathbf{\Sigma}}_{\mathbf{x}}$ *given by* $\hat{\mathbf{\Sigma}}_{\mathbf{x}} = \mathbf{\Gamma}_{Corr\text{-}SBL}^{\frac{1}{2}}\mathbf{U}\mathbf{\Gamma}_{Corr\text{-}SBL}^{\frac{1}{2}}$, *where* $\mathbf{\Gamma}_{Corr\text{-}SBL} = \mathbf{C}_{opt}^{-2}$ *is the hyperparameter matrix at convergence of the Corr-SBL algorithm.*

*Proof.* Let $\mathbf{C}_{\text{opt}}$ denote the value of the hyperparameters obtained upon termination of the Corr-SBL algorithm, and let $\mathbf{\Omega}_{\mathbf{c}} = \mathbf{C}_{\text{opt}}\mathbf{U}^{-1}\mathbf{C}_{\text{opt}}$. The posterior mean estimate $\hat{\mathbf{x}}$ is

obtained as

$$\hat{\mathbf{x}} = \boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}} = \frac{1}{\sigma_n^2}\boldsymbol{\Omega}_{\mathbf{x}|\mathbf{y}}^{-1}\mathbf{y} = \frac{1}{\sigma_n^2}\left(\frac{1}{\sigma_n^2}\mathbf{A}^H\mathbf{A} + \boldsymbol{\Omega}_{\mathbf{c}}\right)^{-1}\mathbf{A}^H\mathbf{y}.$$

Using the matrix identity

$$\left(\frac{1}{\sigma_n^2}\mathbf{A}^H\mathbf{A} + \boldsymbol{\Omega}_{\mathbf{c}}\right)^{-1}\mathbf{A}^H = \sigma_n^2\boldsymbol{\Omega}_{\mathbf{c}}^{-1}\mathbf{A}^H\left(\mathbf{A}\boldsymbol{\Omega}_{\mathbf{c}}^{-1}\mathbf{A} + \sigma_n^2\mathbf{I}_m\right)^{-1},$$

we obtain $\hat{\mathbf{x}} = \hat{\boldsymbol{\Sigma}}_{\mathbf{x}}\mathbf{A}^H\left(\mathbf{A}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}\mathbf{A} + \sigma_n^2\mathbf{I}_m\right)^{-1}\mathbf{y}$, where $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}} \triangleq \mathbf{C}_{opt}^{-1}\mathbf{U}\mathbf{C}_{opt}^{-1}$. From Sec. 2.4.1, we have $\mathbf{C}_{opt}^{-1} = \boldsymbol{\Gamma}_{\text{Corr-SBL}}^{\frac{1}{2}}$, which results in the expression (2.24) as given in the proposition. □

In Section 2.2, a plug-in MMSE estimator was considered in the covariance matching framework. Since Corr-SBL uses a Gaussian prior, the MMSE estimator is linear. Hence, the plug-in LMMSE estimator (2.24) can be viewed as a covariance matching algorithm.

## 2.5 Chapter Summary

In this chapter, we developed a novel covariance matching algorithm based on the Bayesian methodology called Corr-SBL, for the recovery of sparse signals in the presence of intra-vector correlation. The algorithm exploits the knowledge of presence of sparsity and correlation to facilitate recovery of the signal, especially when the covariance structure is unknown and has to be estimated from the data itself. To elucidate the advantage of the algorithm for the correlated sparse recovery problem, we consider a case study of the mmWave MIMO channel estimation problem in a hybrid analog-digital architecture in the next chapter. However, the framework developed can be potentially useful in other applications besides mmWave channel estimation, where the data admits a sparse representation with the nonzero entries being correlated with each other.

# Chapter 3

# mmWave Channel Estimation

Millimeter wave (mmWave) communication has been investigated as a promising technology for the fifth generation (5G) cellular networks [1, 32, 33]. In the introduction chapter, we discussed the challenges in the implementation of mmWave MIMO systems. An important challenge lies in the channel estimation, especially due to hybrid analog-digital architecture that is employed to keep the hardware cost low. In this chapter, our goal is to investigate the role of spatial sparsity and intra-vector correlation to obtain reliable channel estimates with low pilot overhead. We start with a review of literature.

The hybrid analog-digital architecture used in mmWave communications precludes the use of traditional channel estimation techniques. Analog beam sweeping based procedures have been proposed to sample the channel subspace and estimate the mmWave channel links [34–36], but these procedures typically incur large pilot overheads. An alternative approach is to exploit structure in the channel and estimate the channel by solving an

The contents of this chapter have been published in the IEEE transactions on signal processing journal [20].

optimization problem. Measurement campaigns for mmWave channels [1, 14, 15] have revealed structures like sparsity and correlation, which can be incorporated into a statistical model to estimate channel using far fewer pilot transmissions compared to beam scanning based approaches.

Spatial sparsity arises in mmWave channels because the signals arrive at the receiver in a small number of path clusters [1, 14]. Different sparse representations of the mmWave channel have been studied in [37–39], and sparse recovery algorithms such as orthogonal matching pursuit (OMP) have been applied to estimate the channel using a reduced number of pilots. In [40, 41], different training strategies for sparse channel estimation has been discussed. In [42], a parameter-perturbation framework combined with a low-complexity simultaneous OMP algorithm is presented mmWave channel estimation, accounting for off-grid effects. These algorithms based on traditional sparse recovery algorithms are generally not suitable for highly measurement-constrained scenarios and do not provide flexibility in dealing with additional structures.

In addition to sparsity, structures such as spatial correlation have been observed due to mutual coupling at the antennas [15, 43]. In [15], spatial fading models were provided to fit measured spatial correlation using a parameterized exponential model. In a general massive MIMO scenario, [16] developed a Toeplitz model for the covariance matrix under the assumption that the angle of arrivals (AoAs) and path gains are i.i.d. random variables. They also presented an algorithm to exploit the correlation for channel estimation.

To the best of our knowledge, very few of the existing studies capture sparsity (and the resulting spatial correlation) as well as correlation among the nonzero entries of the sparse vector in the same statistical model. Such correlations can arise, for example, due to

signal propagation and scattering in clusters (called block sparsity). Channel estimation techniques that exploit block sparsity have been presented in [44, 45]. While these studies consider a common support in the sparse representation of the channel from the user to the different antennas at the base station, correlation among the nonzero entries are not modeled or exploited. The block sparsity setup can be modified to include correlation among path gains in each block [9], but it does not generalize to correlation across blocks.

In mmWave MIMO systems, new correlation sources may be present [2]. In the multipath channel model that is typically used in the literature, the path gains are modeled as Rayleigh distributed by assuming a rich scattering environment and exploiting the law of large numbers. This does not necessarily hold true for mmWave channels, where the channel consists of a significant line-of-sight component along with a small number of multipath components, and, due to the short range of communication, rich scattering cannot be assumed to occur. In these systems, the covariance matrix of the sparse representation need not be diagonal, and can have non-negligible off-diagonal entries [46]. For example, fading statistics of the multipaths can be correlated, especially when the mmWave communications are used for short range communications. This can lead to correlation between the different links/beams in the multipath setting. This scenario has been observed in practical measurements for rain attenuation in the mmWave band [47, 48]. Also, atmospheric turbulence and correlated shadowing [49] lead to correlation among paths that are further apart, which cannot be considered under the ambit of block sparsity.

From the above discussion, correlation among the nonzero entries of the sparse representation of the channel is unavoidable in general, which makes it necessary to incorporate correlation along with sparsity into statistical models to obtain accurate channel estimates.

To this end, we start with a general statistical model for the channel, where all nonzero entries of the sparse representation are assumed to be correlated with each other, similar to the model considered by Park et al. [46]. The focus of Park et al. was on precoder design, and a modification of OMP was developed for recovering the covariance matrix of the sparse channel vector. An important aspect of correlation-aware channel estimation is the estimation of the channel covariance itself. However, due to the compressed measurements obtained in the hybrid architecture, simple techniques such as the sample covariance do not provide reliable estimates, and new approaches are needed. Bayesian inference algorithms such as sparse Bayesian learning (SBL) [8] are well suited to exploit correlation, as they estimate the covariance as a key step within the algorithm. Their primary goal is to infer the best-fitting distribution from a parameterized class of distributions. Point estimates can then be obtained from the posterior distribution. A message passing algorithm for mmWave channel estimation was developed in [50], and SBL based algorithms are presented in [45, 51] for recovery of spatially uncorrelated sparse channels.

In this chapter, we address the problem of multiple-input multiple-output mmWave channel estimation in a hybrid analog-digital architecture, and investigate the role of underlying spatial sparsity as well as the spatial correlation in the channel to obtain reliable channel estimates with low pilot overhead. In Section 3.1 we present the system model for the mmWave MIMO system. Estimation schemes considered in the work is presented in Section 3.2. For the above problem, we present the application of Corr-SBL algorithm derived in the previous chapter. The analysis of the algorithms is presented in Section 3.4. Finally, the efficacy of the algorithm is presented in Section 3.5 using simulation results.

## 3.1 System Model

In this section, we present the wireless system setup and the channel model for single cell multi-user mmWave MIMO uplink communication with hybrid analog-digital architecture.

### 3.1.1 mmWave MIMO Channel Model

The mmWave wireless system consists of a single MIMO base station (BS) equipped with a uniform linear array (ULA) consisting of $N$ equally spaced antennas, which serves $K$ spatially distributed single antenna users. The base station employs a hybrid MIMO architecture with $M$ (with $K \leq M \ll N$) RF chains. The system model is shown in Fig. 3.1.

The signal at the antenna is converted to the digital domain by a network of phase shifters in a fully connected structure, represented by an analog combining matrix $\mathbf{W}_r \in \mathbb{C}^{M\times N}$, with the phase-only control of the phase shifters satisfied by the constant modulus constraint $|\,[\mathbf{W}_r]_{(m,n)}\,| = \frac{1}{\sqrt{N}}\ \forall m \in [M], n \in [N]$. The digital baseband combiner $\mathbf{F} \in \mathbb{C}^{K\times M}$ then converts the signal into $K$ data streams.

The uplink mmWave channel from each user to the base station is assumed to have a block flat-fading structure, with coherence time $T_c$ (in s) and coherence bandwidth $B_c$ (in Hz), i.e., the channel can be modeled as constant within time-frequency coherence blocks of $\tau_c = B_c T_c$ channel uses. Further, the channel from $k^{\text{th}}$ user to the BS is comprised of $L_k \ll N$ multi-paths. The $l^{\text{th}}$ path at the $r^{\text{th}}$ coherence block is characterized using a complex baseband path gain $\bar{g}_{r,k,l}$ and the corresponding AoA $\psi_{r,k,l}$. The array response

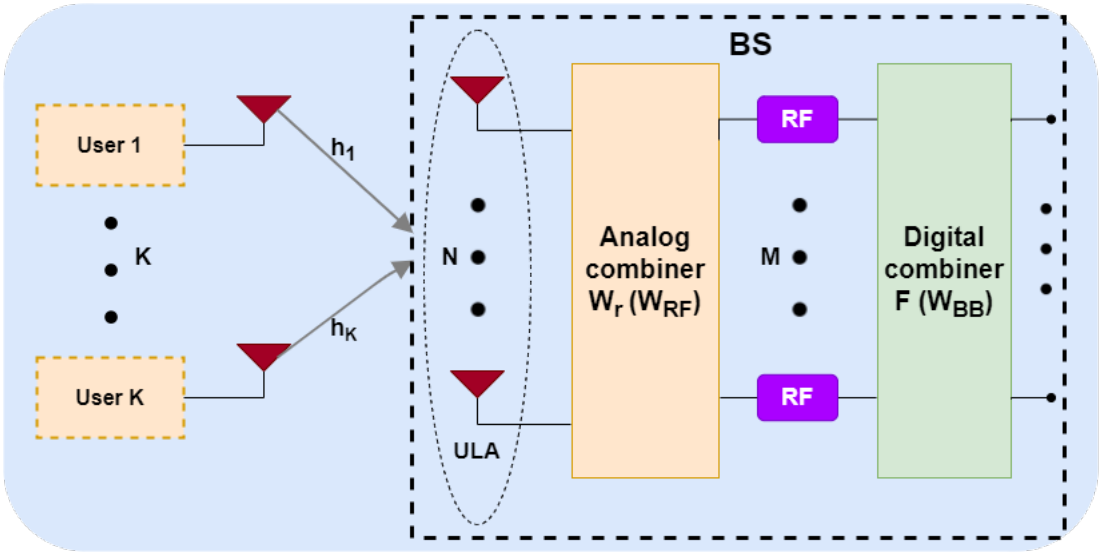



Fig. 3.1. Wireless uplink system with a single $N$ antenna base station serving $K$ single antenna users.

vector $\bar{\mathbf{a}}(\psi_{r,k,l}) \in \mathbb{C}^N$ for the ULA is

$$\bar{\mathbf{a}}(\psi_{r,k,l}) = \frac{1}{\sqrt{N}} \left[1, e^{-j2\pi v_{r,k,l}}, \ldots, e^{-j2\pi(N-1)v_{r,k,l}}\right]^T. \tag{3.1}$$

where $v_{r,k,l} = \frac{d}{\lambda}\cos(\psi_{r,k,l})$ is the directional cosine corresponding to the AoA $\psi_{r,k,l}$. The uplink mmWave channel can then be represented as [46]

$$\mathbf{h}_{r,k} = \sum_{l=1}^{L_k} \bar{g}_{r,k,l}\bar{\mathbf{a}}(\psi_{r,k,l}) = \bar{\mathbf{A}}\bar{\mathbf{g}}_{r,k} \in \mathbb{C}^N, \tag{3.2}$$

where columns of $\bar{\mathbf{A}} \in \mathbb{C}^{N \times L_k}$ are the array response vectors given by (3.1). If we grid the range of possible directional cosine values ($-1$ to 1) using $D \gg L_k$ points, (3.2) can be approximated as

$$\mathbf{h}_{r,k} = \mathbf{A}\mathbf{g}_{r,k}. \tag{3.3}$$

The matrix $\mathbf{A}$ consists of the ULA array response vectors for the AoAs corresponding to the grid points. The index of columns of $\bar{\mathbf{A}}$ in $\mathbf{A}$, denoted by the set $\mathcal{S}_k$ (with $|\mathcal{S}_k| = L_k$),

are the grid points corresponding to AoAs given in (3.2), and the corresponding nonzero entries of $\mathbf{g}_{r,k}$ are the path gains $\bar{g}_{r,k,l}$. Since $D \gg L_k$, $\mathbf{g}_{r,k}$ is a sparse vector.

We assume that the nonzero entries in the complex baseband path gain vector $\mathbf{g}_{r,k}$ have zero mean and are correlated with each other. The $(i,j)^{\text{th}}$ entry of the covariance matrix $\mathbf{R}_{\mathbf{g}_{r,k}} = \mathbb{E}\left[\mathbf{g}_{r,k}\mathbf{g}_{r,k}^H\right] \in \mathbb{C}^{D\times D}$ can be obtained from the Pearson product-moment correlation coefficient definition [23] as $\left[\mathbf{R}_{\mathbf{g}_{r,k}}\right]_{(i,j)} = \rho_{ij}\sqrt{\gamma_i^*}\sqrt{\gamma_j^*}$ where $\gamma_i^*$ denotes the variance of $i^{\text{th}}$ entry of $\mathbf{g}_{r,k}$ and $\rho_{ij}$ is the correlation coefficient between the two entries, which is a function of the pair of indices $(i,j)$ and is governed by a known correlation model (e.g., uniform correlation model, exponential model, Toeplitz model etc.). A matrix $\mathbf{U} \in \mathbb{C}^{D\times D}$ with $\rho_{ij}$ as its $(i,j)^{\text{th}}$ entry is called the correlation matrix. With this notation, we have $\mathbf{R}_{\mathbf{g}_{r,k}} = (\mathbf{\Gamma}^*)^{1/2}\mathbf{U}(\mathbf{\Gamma}^*)^{1/2}$ where $\mathbf{\Gamma}^* = \text{diag}(\gamma_1^*, \gamma_2^*, \ldots, \gamma_D^*)$.

Note that, for a grid point $j$ which does not correspond to any of the $L_k$ paths, the variance of the $j^{\text{th}}$ entry, $\gamma_j^*$, is zero. This results in the covariance structure similar to [46], where $\mathbf{R}_{\mathbf{g}_{r,k}}$ is a $D \times D$ positive semi-definite (PSD) matrix which contains a nonzero $L_k \times L_k$ positive definite (PD) principal submatrix corresponding to the index set $\mathcal{S}_k$, with its other entries equal to 0. Using (3.3), the channel covariance matrix $\mathbf{R}_{\mathbf{h}_{r,k}} = \mathbb{E}\left[\mathbf{h}_{r,k}\mathbf{h}_{r,k}^H\right] = \mathbf{A}\mathbf{R}_{\mathbf{g}_{r,k}}\mathbf{A}^H \in \mathbb{C}^{N\times N}$. The channel statistics vary slowly compared to the channel instantiations [52], i.e., the channel statistics remain constant over a time interval $T_s > T_c$. Hence, the channel covariance and AoAs remain constant over $\tau_s = \frac{T_s}{T_c}$ coherence blocks. In the sequel, since we focus on estimating the covariance matrix within $\tau_s$ coherence blocks where the covariance matrix remains constant, we drop the subscript $r$ in the covariance matrices $\mathbf{R}_{\mathbf{g}_{r,k}}$ and $\mathbf{R}_{\mathbf{h}_{r,k}}$.

### 3.1.2 Pilot Transmission

At the start of each coherence block, the $k$th user transmits a unique orthonormal pilot $\mathbf{p}_k \in \mathbb{C}^{\tau_p}$, $k = 1, 2, \ldots, K$, where $\tau_p$ is the pilot length satisfying $\tau_p = K \ll \tau_c$. This pilot signal is used by the BS to estimate the uplink channels from all the users. The pilot signal received at the BS is processed using an analog combining matrix $\mathbf{W}_r \in \mathbb{C}^{M \times N}$,

$$\mathbf{Y}_r^p = \mathbf{W}_r \left( \sum_{k=1}^{K} \mathbf{h}_{r,k} \mathbf{p}_k^H + \mathbf{N} \right) \in \mathbb{C}^{M \times \tau_p}, \tag{3.4}$$

where the entries of additive noise $\mathbf{N}$ are independent and identically distributed (i.i.d.) Gaussian with zero mean and variance $\sigma_n^2$.[1] Since the pilot signals are orthonormal, by post-multiplying $\mathbf{Y}_r^p$ with $\mathbf{p}_k$, we obtain the pilot signal for estimating the $k^{\text{th}}$ user's channel as

$$\mathbf{y}_{r,k} = \mathbf{Y}_r^p \mathbf{p}_k = \mathbf{W}_r \mathbf{h}_{r,k} + \mathbf{W}_r \mathbf{N} \mathbf{p}_k \in \mathbb{C}^M. \tag{3.5}$$

The covariance of the effective noise $\mathbf{n}_r = \mathbf{W}_r \mathbf{N} \mathbf{p}_k$ is $\sigma_n^2 \mathbf{W}_r \mathbf{W}_r^H$. In Sec. 3.2.2, we present a choice of $\mathbf{W}_r$ such that $\mathbf{W}_r \mathbf{W}_r^H$ approaches $\mathbf{I}_M$ asymptotically ($N \to \infty$) [53]. Hence, $\mathbf{n}_r$ is assumed to comprise of i.i.d. $\mathcal{CN}(0, \sigma_n^2)$ entries.

From (3.5), it can be observed that the measurement $\mathbf{y}_{r,k}$ does not involve interference from other users. Also, since the pilot sequences are orthogonal, the noise in the post-processed received training signals are independent across the users, and we can perform channel estimation independently for each user. Hence, in the sequel, we drop the subscript $k$ and consider the channel estimation for a single user. Estimating $\mathbf{h}_r$ from $\mathbf{y}_r$ in

[1] Without loss of generality, we consider the pilot signal to be of unit power, and include the effect of pilot transmission power in $\sigma_n^2$. Note that (3.4) assumes the users are time and frequency synchronized with the BS. In practice, this synchronization can be achieved using the primary and secondary synchronization signals transmitted by the BS.

(3.5) constitutes the channel estimation problem. Specifically, using (3.3), the channel estimation problem (3.5) can be posed as a sparse recovery algorithm as:

$$\mathbf{y}_r = \mathbf{\Phi}_r \mathbf{g}_r + \mathbf{n}_r, \tag{3.6}$$

where $\mathbf{\Phi}_r = \mathbf{W}_r \mathbf{A} \in \mathbb{C}^{M \times D}$. Hence, our goal boils down to exploiting sparsity and correlation among the nonzero entries of the sparse vector to obtain the channel estimate.

## 3.2 Channel Estimation: Preliminaries

### 3.2.1 Channel Estimation Schemes

In this subsection, we present an overview of two existing classes of channel estimation techniques, which form the baseline for the comparisons in this chapter.

#### Linear Estimation

The first class of estimators are linear estimators, where the estimate of the channel in each coherence block can be represented as $\hat{\mathbf{h}}_r = \mathbf{M}_r \mathbf{y}_r$. This includes linear minimum mean squared error (LMMSE) and least squared (LS) estimation, among others. When the channel covariance matrix $\mathbf{R_h}$ is known, LMMSE is the optimal linear estimator.

$$\hat{\mathbf{h}}_r^{\text{LMMSE}} = \mathbf{R_h} \mathbf{W}_r^H \left( \mathbf{W}_r \mathbf{R_h} \mathbf{W}_r^H + \sigma_n^2 \mathbf{I}_M \right)^{-1} \mathbf{y}_r. \tag{3.7}$$

However, in practice $\mathbf{R_h}$ is not known at the receiver. In this case, the simplest estimator is the least squares (LS) estimator:

$$\hat{\mathbf{h}}_r^{\text{LS}} = \mathbf{W}_r^{\dagger} \mathbf{y}_r, \tag{3.8}$$

where $\mathbf{W}_r^\dagger$ denotes the pseudo-inverse of $\mathbf{W}_r$. The LS estimator does not exploit the correlation $\mathbf{R_h}$ in the channel. To exploit correlation, LS estimates of the channel in $T$ coherence blocks are computed, and the sample covariance obtained from the estimates is used as an estimate for $\mathbf{R_h}$ as

$$\hat{\mathbf{R}}_\mathbf{h} = \mathbf{W}_r^\dagger \left( \frac{1}{T} \sum_{r=1}^{T} \mathbf{y}_r \mathbf{y}_r^H \right) \left( \mathbf{W}_r^\dagger \right)^H . \tag{3.9}$$

$\hat{\mathbf{R}}_\mathbf{h}$ is plugged into the LMMSE estimator resulting in an imperfect channel covariance information (IPCI) based estimator, which exploits correlation structure, but ignores sparsity in the channel.

$$\hat{\mathbf{h}}_r^{\text{IPCI}} = \hat{\mathbf{R}}_\mathbf{h} \mathbf{W}_r^H \left( \mathbf{W}_r \hat{\mathbf{R}}_\mathbf{h} \mathbf{W}_r^H + \sigma_n^2 \mathbf{I}_M \right)^{-1} \mathbf{y}_r. \tag{3.10}$$

**Compressed Sensing based Estimation**

The second class of estimators utilize sparse signal recovery algorithms such as OMP [37] and SBL [8] to recover sparse representation of the channel $\mathbf{g}_r$ from the measurements using (3.6). Using $\hat{\mathbf{g}}_r$, the channel is estimated as $\hat{\mathbf{h}}_r = \mathbf{A}\hat{\mathbf{g}}_r$. These algorithms exploit the sparsity structure, but neglect channel correlation. An OMP-based greedy sparse recovery algorithm, CovOMP, was proposed in [46]. This estimator exploits both sparsity and correlation, and is also used for performance comparisons. Our simulation results show that Bayesian methods significantly outperform CovOMP, especially in highly measurement-constrained scenarios.

We also consider two genie-aided estimators. When the covariance matrix $\mathbf{R_g}$ is known, the LMMSE estimate results in the same estimator as in (3.7), with $\mathbf{R_h}$ replaced with

$\mathbf{A}\mathbf{R}_{\mathbf{g}}\mathbf{A}^H$. To characterize performance of estimators that neglect the correlation information, using the support and individual variances as the diagonal matrix $\mathbf{\Gamma}_{\mathbf{g}}$, a genie-aided element-wise plug-in LMMSE estimator is

$$\hat{\mathbf{g}}_r^{\text{E-LMMSE}} = \mathbf{\Gamma}_g \mathbf{\Phi}_r^H \left( \mathbf{\Phi}_r \mathbf{\Gamma}_g \mathbf{\Phi}_r^H + \sigma_n^2 \mathbf{I}_M \right)^{-1} \mathbf{y}_r. \tag{3.11}$$

**MMV Recovery Algorithms**

Channel estimators based on covariance estimation require knowledge of the covariance matrix. Estimating the covariance entails computing the sample-averaged covariance using multiple channel instantiations. Since the channel statistics are constant over $\tau_s$ coherence blocks, pilots from multiple coherence blocks can be used to estimate the covariance matrix without additional pilot overhead. The support is also constant within a coherence block, and sparse recovery algorithms developed in the multiple measurement vector (MMV) paradigm such as SOMP [46] and MSBL [13] can be used to obtain better performance. However, estimating the channel at the end of $\tau_s$ coherence blocks leads to large latency and memory requirements. Instead, we consider averaging over $T \leq \tau_s$ coherence blocks, with $T$ being chosen to trade-off between performance and complexity.

### 3.2.2 Analog Combiner Matrix Design

The analog combining matrix $\mathbf{W}_r \in \mathbb{C}^{M \times N}$ plays an key role in the performance of channel estimation. In this subsection, we discuss two schemes used in the chapter for the choice of combining matrices across coherence blocks.

The analog combining matrix can be represented as $\mathbf{W}_r = \frac{1}{\sqrt{N}} e^{i\mathbf{\Theta}_r}$ to satisfy the constant modulus constraint, where the entries of $\mathbf{\Theta}_r$ represent the phase of each entry of $\mathbf{W}_r$. The

first of the two schemes uses the same combining matrix across coherence blocks and is referred to as "shared $\mathbf{W}_r$ scheme". In this scheme, $\mathbf{\Theta}_r = \mathbf{\Theta}\ \forall r$, with the entries of $\mathbf{\Theta}$ chosen independently from a uniform distribution in $[0, 2\pi]$.

In [21, Proposition 2], the authors proved that when a shared compression matrix is applied on every sample of a signal, all possible estimators are asymptotically biased. Instead, if independent compression matrices are applied on different samples, unbiased estimators can be designed. Using this for the channel estimation problem, better estimation performance can be obtained when a different combining matrix is chosen for each coherence block. Thus, in our second scheme, $\mathbf{\Theta}_r$ is chosen independently across $r$, with the distribution for each $\mathbf{\Theta}_r$ being the same as in the shared $\mathbf{W}_r$ scheme. This scheme is referred to as "i.i.d. $\mathbf{W}_r$ scheme". We note that the memory and computational complexity of the i.i.d. $\mathbf{W}_r$ scheme is higher than that of the shared $\mathbf{W}_r$ scheme.

In the sequel, we borrow terminology from the compressed sensing literature to refer to estimating the channel within a single coherence block as single measurement vector (SMV) channel estimation and estimation over multiple blocks by exploiting the joint sparsity structure as multiple measurement vector (MMV) channel estimation. In the next section, we will present the application of Corr-SBL algorithm for channel estimation.

## 3.3 Corr-SBL

We recall that our goal for SMV channel estimation is to recover $\mathbf{g}_r \in \mathbb{C}^D$ from the measurements $\mathbf{y}_r = \mathbf{\Phi}_r \mathbf{g}_r + \mathbf{n}_r \in \mathbb{C}^M$ with the knowledge of the structure of covariance matrix of $\mathbf{g}_r$ as given in Section 3.1.1. This setup is similar to the correlated sparse signal recovery problem discussed in Chapter 2, and Corr-SBL algorithm can be applied to the

**Algorithm 2:** `Corr-SBL for mmWave channel estimation`

**Input**: $\{\mathbf{\Phi}_r\}_{r=1}^T, \{\mathbf{y}_r\}_{r=1}^T, \mathbf{U}, \sigma_n^2$

**Initialize**: $k \leftarrow 0, \mathbf{c} \leftarrow \mathbf{1}$

**while** $k \leq k_{max}$ *and* $\| \left[\boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}\right]_k - \left[\boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}\right]_{k-1} \|_F < \epsilon$ **do**

$\hat{\mathbf{R}}_{\mathbf{g}} = \mathbf{0}_{D\times D}$

**for** *r = 1 to T* **do**

$\mathbf{\Omega}_{\mathbf{g}|\mathbf{y}} = \frac{1}{\sigma_n^2}\mathbf{\Phi}_r^H\mathbf{\Phi}_r + \mathbf{\Omega}_{\mathbf{c}}$

$\left[\boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}\right]_{([D],r)} = \frac{1}{\sigma_n^2}\mathbf{\Omega}_{\mathbf{g}|\mathbf{y}}^{-1}\mathbf{\Phi}_r^H\mathbf{y}_r$

$\hat{\mathbf{R}}_{\mathbf{g}} \leftarrow \hat{\mathbf{R}}_{\mathbf{g}} + \frac{1}{T}\left(\mathbf{\Omega}_{\mathbf{g}|\mathbf{y}}^{-1} + \left[\boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}\right]_{([D],r)}\left[\boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}\right]_{([D],r)}^H\right)$

**end**

$\mathbf{c} \leftarrow \left(\mathrm{Re}\left\{\mathbf{U}^{-1} \odot \hat{\mathbf{R}}_{\mathbf{g}}^T\right\}\right)^{-1}\frac{1}{\mathbf{c}}$

$\mathbf{\Omega}_{\mathbf{c}} = \mathrm{diag}(\mathbf{c})\mathbf{U}^{-1}\,\mathrm{diag}(\mathbf{c})$

$k \leftarrow k + 1$

**end**

**Output**: $\{\hat{\mathbf{g}_r}\}_{r=1}^T = \boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}, \mathbf{c}$.

above problem.

The performance of correlation-aware algorithms depends on accuracy of the estimates for the channel covariance. Since the channel covariance is constant for $T$ coherence blocks, multiple measurements can be used to estimate the channel statistics, and improve the performance. The extension of the algorithm derived above to the MMV case involves averaging the E-step updates over all coherence blocks under consideration. In the sequel, the name `Corr-SBL` is used for the resulting MMV algorithm for the known $\mathbf{U}$ case, and the algorithm in the i.i.d. $\mathbf{W}_r$ scheme is presented as Algorithm 2. When the heuristic approach to learn the correlation is incorporated, the algorithm is called `Corr-SBL-learn`. In the shared $\mathbf{W}_r$ scheme, the updates in the inner for-loop are independent of the loop index, and thus a computationally simpler algorithm can be obtained by vectorizing the updates.

**Algorithm 3:** `Online Corr-SBL`

**Input**: $\mathbf{\Phi}, \mathbf{Y}_{r,\Delta} = \begin{bmatrix}\mathbf{y}_r & \mathbf{y}_{r+1} & \dots & \mathbf{y}_{r+\Delta-1}\end{bmatrix}, \mathbf{U}, \sigma_n^2, r, \Delta$

**Prior**: $\mathbf{R}_{\text{old}} = \sum_{s=1}^{r-1} \left(\mathbf{\Phi}^H \mathbf{y}_s \mathbf{y}_s^H \mathbf{\Phi}\right)$

**Initialize**: $k \leftarrow 0, \mathbf{c} \leftarrow \mathbf{1}, \hat{\mathbf{U}} = \mathbf{U}$

**while** $k \leq k_{max}$ *and* $\| \left[\boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}\right]_k - \left[\boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}\right]_{k-1} \|_F < \epsilon$ **do**

$\mathbf{\Omega}_{\mathbf{g}|\mathbf{y}} = \frac{1}{\sigma_n^2} \mathbf{\Phi}^H \mathbf{\Phi} + \mathbf{\Omega}_{\mathbf{c}}$

$\boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}} = \frac{1}{\sigma_n^2} \mathbf{\Omega}_{\mathbf{g}|\mathbf{y}}^{-1} \mathbf{\Phi}^H \mathbf{Y}_{r,\Delta}$

$\hat{\mathbf{R}}_{\mathbf{g}} = \mathbf{\Omega}_{\mathbf{g}|\mathbf{y}}^{-1} + \frac{1}{r+\Delta} \left( \mathbf{\Omega}_{\mathbf{g}|\mathbf{y}}^{-1} \mathbf{R}_{\text{old}} \mathbf{\Omega}_{\mathbf{g}|\mathbf{y}}^{-1} + \boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}} \boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}^H \right)$

$\mathbf{c} \leftarrow \left( \text{Re} \left\{ \hat{\mathbf{U}}^{-1} \odot \hat{\mathbf{R}}_{\mathbf{g}}^T \right\} \right)^{-1} \frac{1}{\mathbf{c}}$

$\hat{\mathbf{\Gamma}} = \text{diag}(\text{diag}(\hat{\mathbf{R}}_{\mathbf{g}}))$

$\hat{\mathbf{U}} = \hat{\mathbf{\Gamma}}^{-\frac{1}{2}} \hat{\mathbf{R}}_{\mathbf{g}} \hat{\mathbf{\Gamma}}^{-\frac{1}{2}}$

$\mathbf{\Omega}_{\mathbf{c}} = \text{diag}(\mathbf{c}) \hat{\mathbf{U}}^{-1} \text{diag}(\mathbf{c})$

$k \leftarrow k + 1$

**end**

**Output**: $\{\hat{\mathbf{g}}_t\}_{s=r}^{r+\Delta-1} = \boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}, \mathbf{c}$,

$\mathbf{R}_{\text{old}} \leftarrow \mathbf{R}_{\text{old}} + \sum_{s=r}^{r+\Delta-1} \left(\mathbf{\Phi}^H \mathbf{Y}_{r,\Delta} \mathbf{Y}_{r,\Delta}^H \mathbf{\Phi}\right)$

### 3.3.1 Online Estimation

In the MMV estimation problem, the channel is estimated at the end of $T$ coherence blocks, which could be impractical in terms of latency when $T$ is large. In this subsection, we present an online version of Corr-SBL, which also incorporates the advantages of MMV estimation.

We first consider the shared $\mathbf{W}_r$ scheme. The only update which involves averaging over multiple coherence blocks is in computing $\boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}} \boldsymbol{\mu}_{\mathbf{g}|\mathbf{y}}^H$. A running sum $\sum_{s=1}^{r-1} \left(\mathbf{\Phi}^H \mathbf{y}_s \mathbf{y}_s^H \mathbf{\Phi}\right)$ for the $r-1$ preceding coherence blocks can be used as prior knowledge for estimating the channel in the $r^{\text{th}}$ coherence block. Building on this idea, we perform joint channel estimation over $\Delta \ll T$ blocks starting from the $r^{\text{th}}$ coherence block. The final algorithm

proposal of Online Corr-SBL with learning correlation called Online-Corr-SBL-learn is presented in Algorithm 3. For the i.i.d. $\mathbf{W}_r$ scheme, a running sum of $\hat{\mathbf{R}}_{\mathbf{g}}$ computed in Theorem 1 is required for each iteration of EM algorithm, resulting in a large memory overhead. Instead, an approximation for the MMV setup can be obtained by storing the running sum of $\hat{\mathbf{R}}_{\mathbf{g}}$ at the end of the final EM iteration of the preceding block.

### 3.3.2 Algorithm Extensions

The main goal of this chapter is to elucidate the role of spatial sparsity and correlation in mmWave channel estimation. For simplicity of exposition, we considered a frequency-flat channel model and single-antenna users in developing our solution. In this section, we briefly discuss extensions of our approach to the cases where the channels are frequency-selective and the users are equipped with multiple antennas.

#### Multiple Antennas at the Users

In this subsection, we present a measurement model for the correlated sparse recovery problem when the users are equipped with multiple antennas. Suppose the BS and the $K$ users employ ULAs with $N_B$ and $N_U$ antennas, and have $M_B < N_B$ and $M_U < N_U$ RF chains, respectively. For this setup, the channel in (3.2) can be written as,

$$\mathbf{H}_{r,k} = \sum_{l=1}^{L_k} \bar{g}_{r,k,l} \bar{\mathbf{a}}_B(\psi_{r,k,l}) \bar{\mathbf{a}}_U^H(\theta_{r,k,l}) = \bar{\mathbf{A}}_B \bar{\mathbf{g}}_{r,k} \bar{\mathbf{A}}_U^H \in \mathbb{C}^{N_B \times N_U},$$

where $\bar{\mathbf{A}}_B \in \mathbb{C}^{N_B \times L_k}$ and $\bar{\mathbf{A}}_U \in \mathbb{C}^{N_U \times L_k}$ denote the array response vectors at the BS and the $k^{\text{th}}$ user, respectively. Using grids of size $D_B$ and $D_U$, $\mathbf{H}_{r,k}$ can be approximated as

$$\mathbf{H}_{r,k} = \mathbf{A}_B \mathbf{G}_{r,k} \mathbf{A}_U^H$$

where $\mathbf{G}_{r,k} \in \mathbb{C}^{D_B \times D_U}$ is a sparse matrix with $L_k$ nonzero path gains corresponding to the respective AoAs and AoDs. If the BS uses a combiner $\mathbf{W}_r \in \mathbb{C}^{M_B \times N_B}$ and the user employs a precoder $\mathbf{F}_r \in \mathbb{C}^{N_U \times M_U}$, the received signal at the BS for estimating the $k^{\text{th}}$ user's channel (corresponding to (4)) is obtained as

$$\mathbf{Y}_{r,k} = \mathbf{W}_r \mathbf{A}_B \mathbf{G}_{r,k} \mathbf{A}_U^H \mathbf{F}_r + \mathbf{N}_{r,k}.$$

By vectorizing $\mathbf{Y}_{r,k}$, we obtain the linear system given by

$$\mathbf{y}_{r,k} = \left(\mathbf{F}^T \otimes \mathbf{W}\right) \left(\mathbf{A}_U^{\mathrm{c}} \otimes \mathbf{A}_B\right) \mathbf{g}_{r,k} + \mathbf{n}_{r,k} \in \mathbb{C}^{M_B M_U \times 1},$$

where $\mathbf{A}_U^{\mathrm{c}}$ denotes the element wise conjugate of $\mathbf{A}_U$, and $\mathbf{g}_{r,k} \in \mathbb{C}^{D_B D_U \times 1}$ is vectorized form of $\mathbf{G}_{r,k}$. This is now in the same form as the sparse recovery framework developed in this paper, albeit with a larger dimensional measurement matrix and sparse vector. At the cost of higher computational complexity, the Corr-SBL can now be directly used for learning the channel $\mathbf{g}_{r,k}$ from the measurements $\mathbf{y}_{r,k}$.

**Frequency-selective Channel Models**

In OFDM systems, the use of subcarriers for data transmission allows one to convert frequency selective channels into multiple parallel frequency flat channels. In this case, the algorithm described in the paper can be applied independently over the different sub-carriers. However, this approach cannot exploit the correlation in the channels across sub-carriers. The work in [46] considers a combination of time-domain and frequency-domain algorithms. A challenge in extending Corr-SBL to the frequency-selective case is in working out the parameters of the channel correlation across subcarriers, and studying

Table 3.1: Memory and computational complexity

| Algorithm | Mode | Memory complexity | Computational complexity |
|---|---|---|---|
| MSBL | i.i.d. $\mathbf{W}_r$ | $\mathcal{O}\left(DMT\right)$ | $\mathcal{O}\left(DM^2T\right)$ |
| | Shared $\mathbf{W}_r$ | $\mathcal{O}\left(M(D+T)\right)$ | $\mathcal{O}\left(DM(M+T)\right)$ |
| Corr-SBL | i.i.d. $\mathbf{W}_r$ | $\mathcal{O}\left(DMT+D^2\right)$ | $\mathcal{O}\left(DM^2T+D^3\right)$ |
| | Shared $\mathbf{W}_r$ | $\mathcal{O}\left(M(D+T)+D^2\right)$ | $\mathcal{O}\left(DM(M+T)+D^3\right)$ |
| Online-Corr-SBL | i.i.d. $\mathbf{W}_r$ | $\mathcal{O}\left(DM\Delta+D^2\right)$ | $\mathcal{O}\left(DM^2\Delta+D^3\right)$ |
| | Shared $\mathbf{W}_r$ | $\mathcal{O}\left(M(D+\Delta)+D^2\right)$ | $\mathcal{O}\left(DM(M+\Delta)+D^3\right)$ |

the performance of the resulting algorithm.

## 3.4 Analysis

### 3.4.1 Complexity Analysis

Table 3.1 compares the per-iteration memory and computational complexity of Corr-SBL, its online version, and MSBL. Here, the computational complexity is measured in terms of the number of floating point operations. The overall complexity of Corr-SBL and Corr-SBL-learn (similar for online versions) are the same since the additional computation of $\hat{\mathbf{U}}$ given by $\mathcal{O}(D^2)$ is included in $\mathcal{O}(D^3)$. The additional complexity of $\mathcal{O}(D^3)$ of Corr-SBL over MSBL is due to the $\mathbf{c}$-update step in Algorithm 2. The complexity is higher for the i.i.d. $\mathbf{W}_r$ scheme due to the additional inner-loop. The online version of Corr-SBL has lower computational complexity, in addition to lower latency. From our experiments, Corr-SBL converges in a similar number of iterations as MSBL, both with and without learning the correlation.

The additional complexity of $\mathcal{O}(D^3)$ can be computationally expensive, especially when large grid sizes are used. An approach to reduce complexity is to use lower grid sizes, but this can result in performance loss due to grid mismatch. A recent approach[2] has

[2]https://ece.iisc.ac.in/∼cmurthy/Learned_Chester_AI5GPHY_Challenge.pdf

considered integrating a greedy search procedure to obtain coarse estimates of the AoAs with a statistical interference model (based on MSBL), followed by dictionary refinement with smaller grid sizes. Another approach to reduce complexity is to consider the algorithm unrolling framework [54], where model based approaches can be used to develop deep learning techniques that perform similar to or better than the optimization based approaches while reducing the complexity [55].

The memory complexity is measured by the storage required for input information. Storing the correlation matrix information leads to higher memory complexity in Corr-SBL compared to MSBL. However, this can be significantly reduced for parameterized models by storing only the parameters. The memory complexity is higher for the i.i.d. $\mathbf{W}_r$ scheme compared to the shared $\mathbf{W}_r$ scheme, as expected.

### 3.4.2 Plug-in LMMSE Estimators

Corr-SBL algorithm is designed based on the maximum likelihood principle and was shown to attain a stationary point of the cost function. However, the performance of channel estimation is usually measured using other cost functions such as NMSE or spectral efficiency of the system. To analyze the performance of Corr-SBL under these measures, we present a unified framework by considering the class of plug-in LMMSE estimators, which includes the LMMSE, E-LMMSE and IPCI estimators described in 3.2. We also showed in Section 2.4.4, that the output of Corr-SBL can be represented as a plug-in LMMSE estimator, and a similar argument can be presented for MSBL also. The following result summarizes the class of plug-in-LMMSE estimators, unifying algorithms described above.

**Proposition 3.** *The class of plug-in LMMSE estimators can be represented as*

$$\hat{\mathbf{h}}_r = \mathbf{M}_r \mathbf{y}_r = \hat{\mathbf{R}}_{\mathbf{h}} \mathbf{W}_r^H \left( \mathbf{W}_r \hat{\mathbf{R}}_{\mathbf{h}} \mathbf{W}_r^H + \sigma_n^2 \mathbf{I}_M \right)^{-1} \mathbf{y}_r, \tag{3.12}$$

*where the covariance matrix* $\hat{\mathbf{R}}_{\mathbf{h}}$ *depends on the estimator:*

$$\hat{\mathbf{R}}_{\mathbf{h}} = \begin{cases} \mathbf{R}_{\mathbf{h}} = \mathbf{A}\boldsymbol{\Gamma}^{\frac{1}{2}}\mathbf{U}\boldsymbol{\Gamma}^{\frac{1}{2}}\mathbf{A}^H & \textit{LMMSE} \\ \mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^H & \textit{E-LMMSE} \\ \mathbf{A}\boldsymbol{\Gamma}_{\textit{Corr-SBL}}^{\frac{1}{2}}\mathbf{U}\boldsymbol{\Gamma}_{\textit{Corr-SBL}}^{\frac{1}{2}}\mathbf{A}^H & \textit{Corr-SBL} \\ \mathbf{A}\boldsymbol{\Gamma}_{\textit{MSBL}}\mathbf{A}^H & \textit{MSBL} \\ \hat{\mathbf{R}}_{\mathbf{h}}^{\textit{IPCI}} & \textit{IPCI} \end{cases}$$

*where* $\boldsymbol{\Gamma}$ *and* $\mathbf{U}$ *denote the true values for the variances and the correlation matrix, and* $\hat{\mathbf{R}}_{\mathbf{h}}^{\textit{IPCI}} \triangleq \mathbf{W}_r^{\dagger} \left( \frac{1}{T} \sum_{r=1}^{T} \mathbf{y}_r \mathbf{y}_r^H \right) \left( \mathbf{W}_r^{\dagger} \right)^H$ *with* $\mathbf{W}_r^{\dagger}$ *denoting the pseudoinverse of* $\mathbf{W}$*.*

### 3.4.3 Normalized Mean Squared Error (NMSE)

The NMSE in the channel estimate is defined as $\text{NMSE} = \mathbb{E}\left[ \|\hat{\mathbf{h}}_r - \mathbf{h}\|_2^2 \right] / \mathbb{E}\left[ \|\mathbf{h}\|_2^2 \right]$. The following theorem provides the NMSE for the plug-in LMMSE estimators. Its proof follows from direct computation and is omitted.

**Theorem 3.** *Consider the estimator* $\hat{\mathbf{h}}_r = \mathbf{M}_r \mathbf{y}_r$*, where the the plug-in LMMSE matrix* $\mathbf{M}_r$ *is given by* $\mathbf{M}_r = \hat{\mathbf{R}}_{\mathbf{h}} \mathbf{W}_r^H \left( \mathbf{W}_r \hat{\mathbf{R}}_{\mathbf{h}} \mathbf{W}_r^H + \sigma_n^2 \mathbf{I}_M \right)^{-1}$*. Assuming that the estimate* $\hat{\mathbf{R}}_{\mathbf{h}}$ *is independent of the measurement* $\mathbf{y}_r$*, i.e.,* $\hat{\mathbf{R}}_{\mathbf{h}}$ *is computed using measurements from other coherence blocks, the NMSE in the channel estimate is given by*

$$\textit{NMSE} = \frac{1}{\text{Tr}\left[\mathbf{R}_{\mathbf{h}}\right]} \times \mathbb{E}\left[ \text{Re}\left\{ \text{Tr}\left[ \mathbf{M}_r \left( \mathbf{W}_r \mathbf{R}_{\mathbf{h}} \mathbf{W}_r^H + \sigma_n^2 \mathbf{I}_M \right) \mathbf{M}_r^H + \mathbf{R}_{\mathbf{h}} - 2\mathbf{M}_r \mathbf{W}_r \mathbf{R}_{\mathbf{h}} \right] \right\} \right], \tag{3.13}$$

*where the expectation is over the randomness in* $\mathbf{M}_r$ *and* $\mathbf{W}_r$*.*

In the above, we assumed that $\hat{\mathbf{R}}_{\mathbf{h}}$ is independent of the measurement $\mathbf{y}_r$. If this does not hold, the expression for the MSE becomes complicated due to the coupling between the two, making the analysis more involved. In any case, one typically uses multiple previous channel instantiations to estimate $\hat{\mathbf{R}}_{\mathbf{h}}$, hence, this is not unduly restrictive.

We note that the error in estimating the covariance matrix using performance measures like Frobenius and spectral norm have been studied in literature [21]. However, very few works consider the *error in signal recovery* using the noisy covariance estimates, especially for the compressed sensing case. Theorem 3 presents the NMSE performance for the class of plug-in LMMSE estimators. For a given set of system parameters $\mathbf{W}, \mathbf{R}_{\mathbf{h}}$ and $\sigma_n^2$, the NMSE depends on $\mathbf{M}_r$, which in turn depends on the covariance estimate. From the theorem, it is straightforward to verify that the least NMSE is obtained by the LMMSE estimator, which assumes perfect knowledge of the covariance matrix $\mathbf{R}_{\mathbf{h}}$. It can be verified that the genie-aided LMMSE achieves the Cramér-Rao bound for the above problem. The E-LMMSE estimator and MSBL force a diagonal structure correlation, and therefore do not exploit the full covariance structure. The IPCI estimator uses the sample covariance, which requires large number of samples to learn the structure of $\mathbf{R}_{\mathbf{h}}$ because the underlying sparsity is not exploited. The covariance estimate of Corr-SBL has a structure similar to actual covariance, and has the potential to learn $\mathbf{R}_{\mathbf{h}}$ accurately using only a small number of samples. We illustrate this by comparing the NMSE value computed using Theorem 3 with the simulated NMSE in Fig. 3.4 in Sec. 3.5.

### 3.4.4 Spectral Efficiency (SE) Analysis

Let $\hat{\mathbf{H}} = \left[\hat{\mathbf{h}}_1 \hat{\mathbf{h}}_2 \dots \hat{\mathbf{h}}_K\right]$ denote the channel estimates for all users obtained using (3.12) in a single coherence block. If the $k^{\text{th}}$ user transmits a symbol $x_k$ $\forall k$ with zero mean and power $P$, the received combined vector at the base station is

$$\mathbf{y} = \mathbf{F}\mathbf{W}_{RF}\left(\sum_{k=1}^{K}\mathbf{h}_k x_k + \mathbf{n}\right). \tag{3.14}$$

where we use $\mathbf{W}_{RF}$ as the analog combiner matrix to distinguish it from the matrix $\mathbf{W}_r$ used in the pilot transmission phase. Various designs for the analog combiner $\mathbf{W}_{RF}$ and digital combiner $\mathbf{F}$ matrices have been studied in the literature when perfect CSI is available at the BS. One simple design aims at designing them independently [56]. In this work, we use the same methodology, with channel estimates $\hat{\mathbf{H}}$ replacing the unknown CSI $\mathbf{H}$. We let the number of users served by the BS to be equal to the number of RF chains, i.e., $K = M$. Then, the phase-only control of the analog combining matrix $\mathbf{W}_{RF}$ is satisfied by constraining the amplitude of all entries to $\frac{1}{\sqrt{N}}$, and the phase of $(m,k)^{\text{th}}$ entry of $\mathbf{W}_{RF}$, denoted by $\hat{\theta}_{m,k}$, is set using the channel estimate as

$$\hat{\theta}_{m,k} = \hat{\phi}_{m,k},\ m \in [M], k \in [K], \tag{3.15}$$

where $\hat{\phi}_{m,k}$ is the phase of $\left[\hat{\mathbf{H}}^H\right]_{m,k}$. The effective baseband channel $\hat{\mathbf{H}}_B = \mathbf{W}_{RF}\hat{\mathbf{H}}$ is used to design the digital combiner using the regularized zero forcing (RZF) combining with regularization parameter $\eta = \frac{M\sigma_n^2}{P}$:

$$\mathbf{F} = \left(\hat{\mathbf{H}}_B^H \hat{\mathbf{H}}_B + \eta \mathbf{I}_M\right)^{-1} \hat{\mathbf{H}}_B^H. \tag{3.16}$$

In order to compute the SE, we write the $k^{\text{th}}$ row of (3.14) as

$$y_k = \underbrace{\mathbb{E}\left[\mathbf{f}_k \mathbf{W}_{RF} \mathbf{h}_k\right] x_k}_{\text{signal}} + \underbrace{\left(\mathbf{f}_k \mathbf{W}_{RF} \mathbf{h}_k x_k - \mathbb{E}\left[\mathbf{f}_k \mathbf{W}_{RF} \mathbf{h}_k\right] x_k\right)}_{\text{self-interference}} + \underbrace{\sum_{q=1, q \neq k}^{K} \mathbf{f}_k \mathbf{W}_{RF} \mathbf{h}_q x_q}_{\text{inter-user interference}} + \underbrace{\mathbf{f}_k \mathbf{W}_{RF} \mathbf{n}}_{\text{AWGN}}, \tag{3.17}$$

where $\mathbf{f}_k$ is the $k^{\text{th}}$ row of $\mathbf{F}$. The term $\mathbb{E}\left[\mathbf{f}_k \mathbf{W}_{RF} \mathbf{h}_k\right]$ is treated as a known deterministic channel. Noting that it is uncorrelated with other terms in the summation, we obtain the following result. The proof follows from applying the use-and-then forget bound [57, Section 3.2] and is omitted.

**Theorem 4.** *The spectral efficiency (SE) of user $k$ is lower-bounded by*

$$SE_k \geq \left(1 - \frac{K}{\tau_c}\right) \log_2\left(1 + \tilde{\gamma}_k\right) \ \ \textit{bits/s/Hz}, \tag{3.18}$$

*where the pre-log factor accounts for pilot overhead, and $\tilde{\gamma}$ is a lower bound on SINR given by*

$$\tilde{\gamma}_k = \frac{P \left|\mathbb{E}\left[\mathbf{f}_k \mathbf{W}_{RF} \mathbf{h}_k\right]\right|^2}{P \sum_{q=1}^{K} \mathbb{E}\left[\left|\mathbf{f}_k \mathbf{W}_{RF} \mathbf{h}_q\right|^2\right] - P \left|\mathbb{E}\left[\mathbf{f}_k \mathbf{W}_{RF} \mathbf{h}_k\right]\right|^2 + \alpha_k} \tag{3.19}$$

*where $\alpha_k = \sigma_n^2 \mathbb{E}\left[\mathbf{f}_k \mathbf{W}_{RF} \mathbf{W}_{RF}^H \mathbf{f}_k^H\right]$ is the effective noise power in the combined signal.*

The expectations in Theorem 4 cannot be obtained in closed form. They are computed using simulations in the next section.

Table 3.2: Simulation parameters

| **Parameters** | **Values** |
|---|---|
| Number of antenna $N$ | 256 |
| Grid size $D$ | 256 |
| Number of RF chains $M$ | 16 |
| Number of users $K$ | 16 |
| Number of multipaths $L_k$ | 16 |
| Number of snapshots $T$ | 512 |
| SNR | 10 dB |
| Correlation coefficient $\rho$ | 0.5 |
| Coherence interval $\tau_c$ | 2000 |
| Hyperparameter pruning threshold $\varepsilon$ | $10^5$ |

## 3.5 Results

### 3.5.1 Simulation Setup

In this section, we present simulation results to elucidate the role of correlation and sparsity in mmWave channel estimation. We benchmark the NMSE performance of Corr-SBL against the genie-aided LMMSE, LS, IPCI, E-LMMSE estimators given by equations (3.7), (3.8), (3.10), (3.11) in Appendix 3.2, respectively, and sparse recovery algorithms SOMP [46], CovOMP [46] and MSBL [10]. We also compare the average sum rate achieved by the hybrid MIMO architecture with channel estimates obtained using Corr-SBL against that of a system equipped with a fully digital architecture (where the number of RF chains equals the number of antennas and hence there is no need for an analog beamforming stage) and genie-aided LMMSE channel estimates. We consider the uniform correlation model, where $\mathbf{U}_{ij}$ equals unity when $i = j$ and equals $\rho \in [0, 1)$ otherwise [22]. The support of the sparse vector is obtained by drawing $L_k$ samples from the $D$ grid points uniformly at random without replacement. The path gain vector is obtained from a complex normal distribution with zero mean and covariance matrix of the form described in Section. 3.1.1.

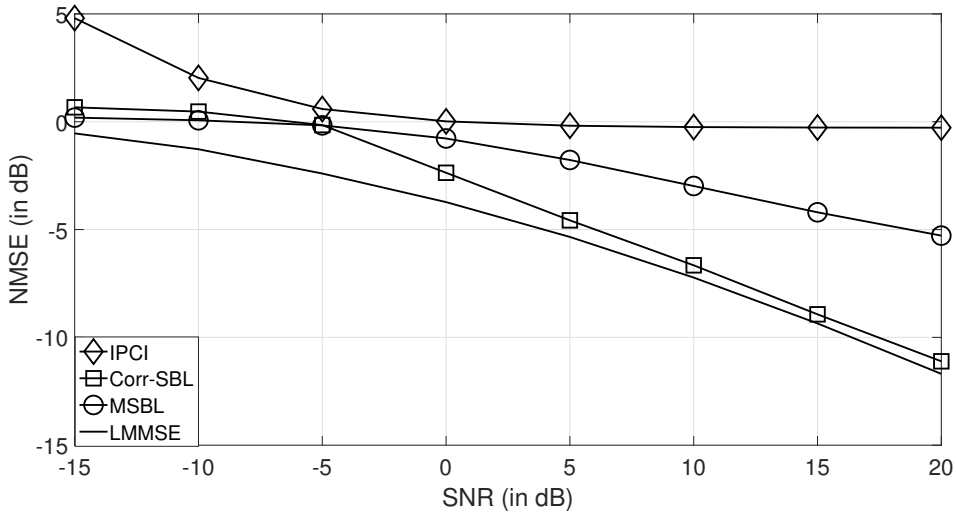



Fig. 3.2. Comparison of NMSE against SNR for the class of plug-in LMMSE estimators.

Orthogonal Zadoff-Chu sequences of length equal to the number of users are used for pilot signals. Specifically, for $K = 16$ users, we use the command `zadoffChuSeq(19,17)` (in MATLAB) to generate a base sequence of length 17. Then, we consider 15 successive cyclically permuted sequences of the base sequence, along with the base sequence, as pilot sequences for the 16 users. The other parameter values are as in Table 3.2, unless mentioned otherwise.

### 3.5.2 Effect of SNR

In the first experiment, we study the NMSE performance of the class of plug-in LMMSE estimators as a function of the SNR. From Fig. 3.2, we see that for SNR less than 0 dB, the advantage of exploiting correlation is not significant since the noise overwhelms the signal. However, as the SNR increases beyond 0 dB, performance of all algorithms exploiting sparsity increases, with the gain being higher for Corr-SBL and genie aided LMMSE estimator compared to MSBL. In the further simulations, we fix the SNR to 10 dB, and study the effect of other parameters of the system to elucidate the role of

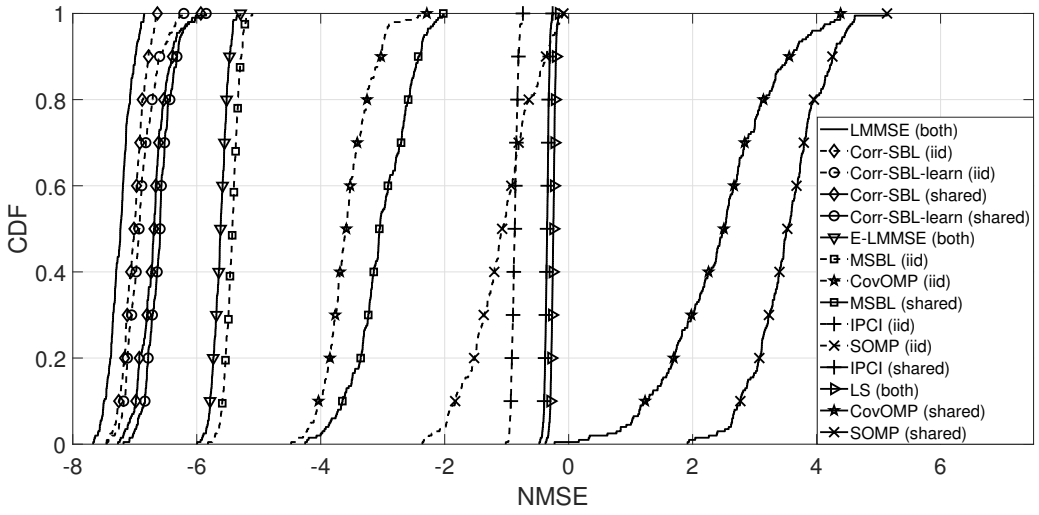



Fig. 3.3. Comparison of NMSE for different algorithms for both shared $\mathbf{W}_r$ and i.i.d. $\mathbf{W}_r$ schemes.

correlation in sparse signal recovery.

### 3.5.3 mmWave Channel Estimation

In the second set of experiments, we present the cumulative distribution function (CDF) of the NMSE of different algorithms, run for 1000 independent realizations of the channel and $T$ coherence blocks. The curves for both i.i.d. $\mathbf{W}_r$ and shared $\mathbf{W}_r$ schemes are presented in Fig. 3.3. The genie based LMMSE estimator (curve labeled `LMMSE`), sets the best achievable benchmark for all algorithms and the performance of Corr-SBL (curve labeled `Corr-SBL`) is only marginally worse. For the case where the correlation coefficient $\rho$ is not known, we estimate it by averaging the off-diagonal entries of $\hat{\mathbf{U}}$ as explained in Sec. 2.4.3. This curve, labeled `Corr-SBL-learn`, matches the performance of Corr-SBL, showing that the algorithm can learn $\rho$ without appreciable loss in performance. Exploiting only correlation (`IPCI`) or only sparsity (MSBL) performs worse than Corr-SBL.

The NMSE with the i.i.d. $\mathbf{W}_r$ scheme is lower than that of the shared $\mathbf{W}_r$ scheme,

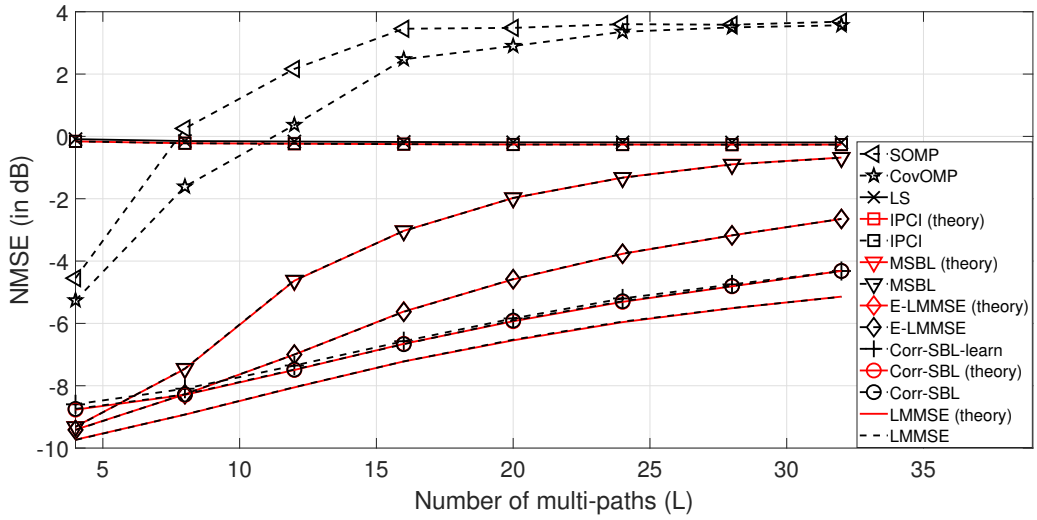
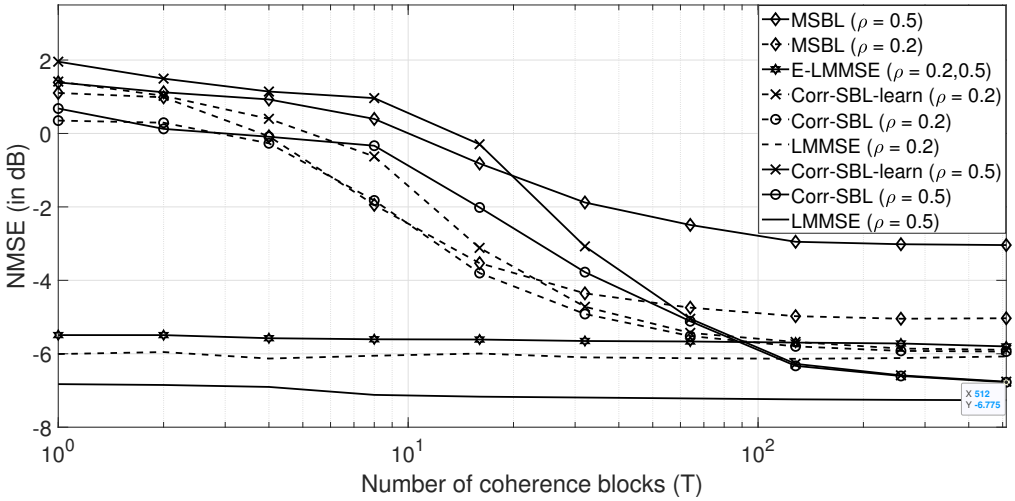



Fig. 3.4. NMSE performance against sparsity level.

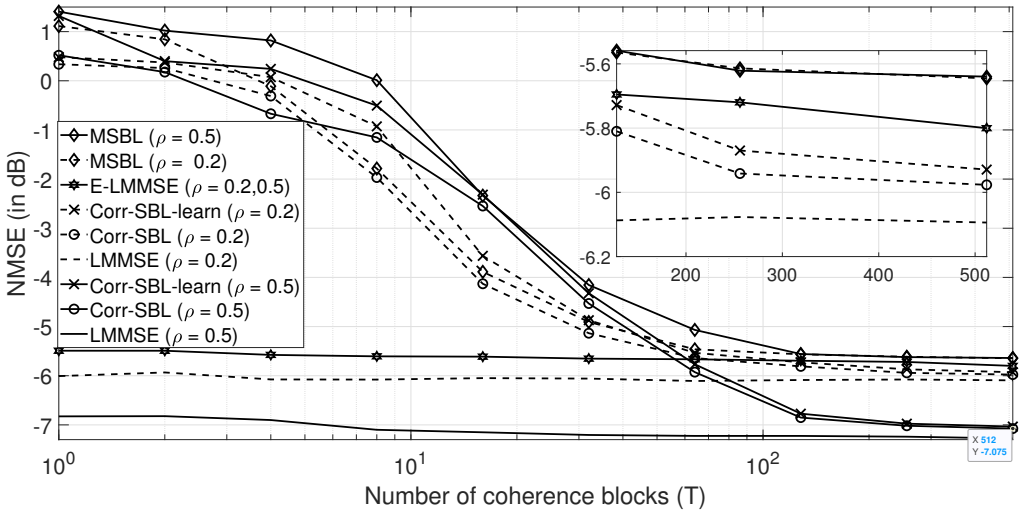
showing that it is better to use independent measurement matrices across coherence blocks. The genie-based estimator `E-LMMSE` only exploits sparsity and is a lower bound on the performance of sparse recovery algorithms that do not exploit correlation. It performs worse than Corr-SBL, which illustrates the importance of exploiting correlation in addition to sparsity. However, although it does not exploit correlation, the hierarchical Bayesian prior used in MSBL results in better performance than CovOMP (which does exploit both correlation and sparsity). In the shared $\mathbf{W}_r$ scheme, the algorithms CovOMP and SOMP in fact perform worse than a trivial estimator which outputs all zeros (and yields an NMSE of 0 dB).

### 3.5.4 Performance for Varying Sparsity Levels

The NMSE performance of different algorithms is compared as a function of the number of multi-path components, i.e., the sparsity level of the channel, in Fig. 3.4. The performance of all sparse recovery algorithms are similar at low sparsity levels, hence exploiting the correlation is not crucial. The performance of SOMP, CovOMP and MSBL deteriorate




(a) Shared $\mathbf{W}_r$ scheme.




(b) i.i.d. $\mathbf{W}_r$ scheme.

Fig. 3.5. NMSE performance with averaging over multiple coherence blocks.

quickly with increasing sparsity. In contrast, both Corr-SBL and Corr-SBL-learn continue to perform close to the optimal genie-aided estimator even at high sparsity levels, and significantly outperform the other methods. Thus, Bayesian methods can offer significant advantages especially in highly measurement-constrained scenarios, when the prior model is chosen to best-fit the underlying model. Also, the NMSE values computed using Theorem 3 overlap perfectly with the simulated NMSE values, illustrating the accuracy of the theoretical expressions.

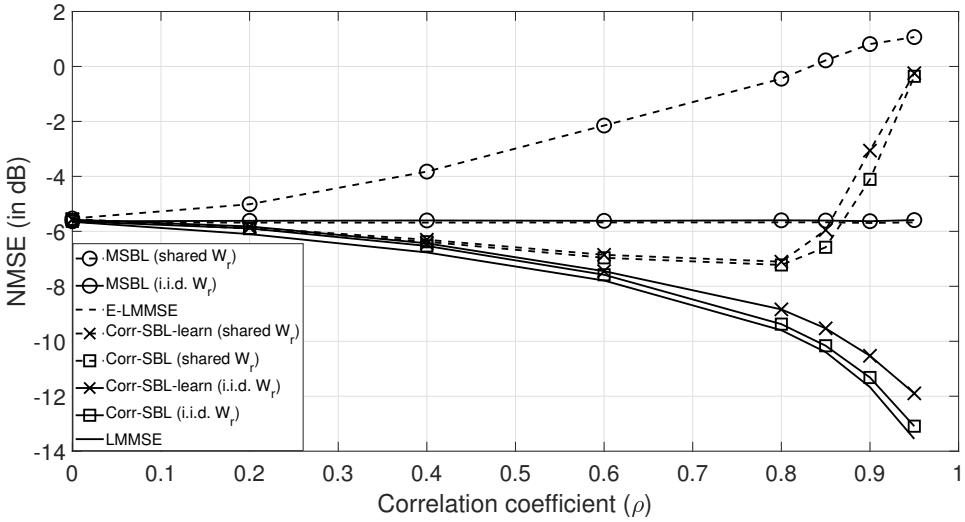



Fig. 3.6. NMSE performance against the correlation level.

### 3.5.5 Shared $\mathbf{W}_r$ vs i.i.d. $\mathbf{W}_r$ Schemes

Fig. 3.5 compares the channel estimation performance of the i.i.d. $\mathbf{W}_r$ and shared $\mathbf{W}_r$ schemes. The estimators which perform poorly in the highly measurement-constrained scenarios are not plotted to avoid clutter. As the number of coherence blocks used to estimate the channel covariance ($T$) increases, the performance of Corr-SBL and Corr-SBL-learn converge to the LMMSE estimator and that of MSBL converges to the E-LMMSE estimator, for both schemes. The NMSE value of the Corr-SBL algorithm for $T = 512$ is lower for the i.i.d. $\mathbf{W}_r$ scheme compared the shared $\mathbf{W}_r$ scheme. Fig. 3.6 shows that the performance of MSBL and Corr-SBL deteriorate with increase in correlation only for the shared $\mathbf{W}_r$ scheme. The i.i.d. $\mathbf{W}_r$ scheme offers better performance compared to the shared $\mathbf{W}_r$ scheme at high correlation, as shown in Fig. 3.6. In the shared $\mathbf{W}_r$ scheme, the performance of MSBL deteriorates with increasing correlation, and Corr-SBL follows the performance of the genie aided LMMSE estimator until a correlation threshold, after which the high correlation overwhelms the covariance estimation procedure, possibly due

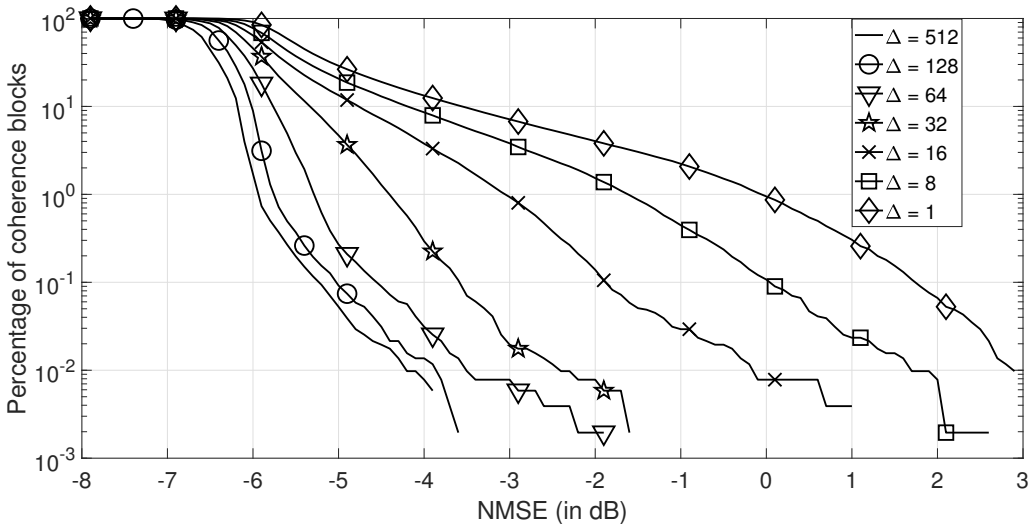



Fig. 3.7. Expected percentage of coherence blocks with NMSE value more than $x$, where $x$ is the $x$-axis coordinate value.

to the larger condition number of $\mathbf{U}$. However, as seen for both MSBL and Corr-SBL, in the i.i.d. $\mathbf{W}_r$ scheme, better covariance estimation leads to better performance even at high correlation levels.

### 3.5.6 Online Estimation

Here, we compare the performance of the online version of Corr-SBL for different block lengths $\Delta$. We consider the scenario of estimating the channel over $T = 512$ channel coherence blocks. In Fig. 3.7, we plot the percentage of the 512 coherence blocks where the NMSE is greater than a given value, say $x$, as a function of $x$. With $\Delta = 64$, which offers a significant reduction in latency compared to using all 512 coherence blocks, less than 0.2% of the blocks have an NMSE higher than $-5$ dB, while the MMV solution that uses all the 512 coherence blocks achieves an NMSE slightly lower than $-6$ dB. Even with $\Delta = 1$, less than 10% of the blocks have an NMSE higher than $-3$ dB. The higher NMSE at smaller $\Delta$ is due to poor estimates in the initial few blocks; the NMSE in fact

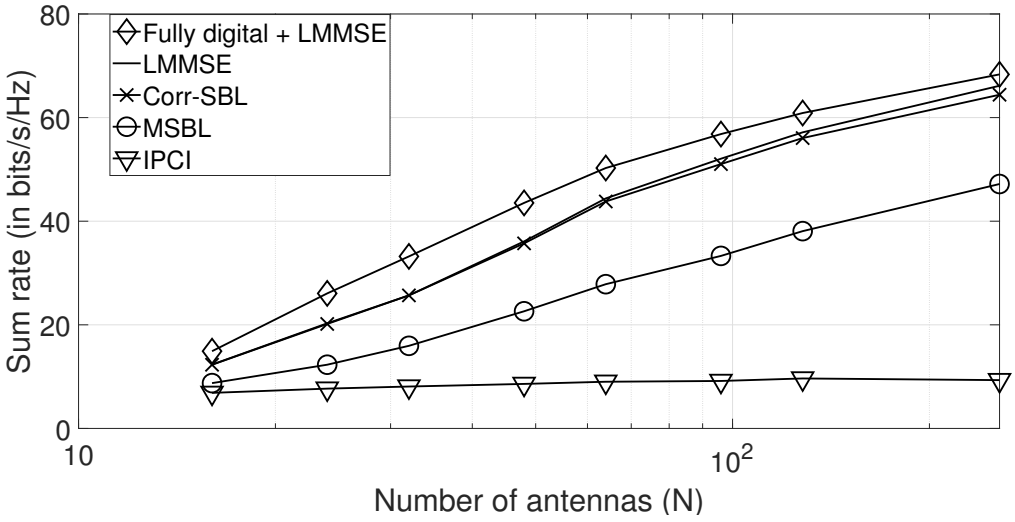



Fig. 3.8. Sum-rate performance of the system with channel estimates obtained by different algorithms.

rapidly decreases in the later blocks. This small performance loss can be insignificant when compared to the latency reduction advantage in practical applications.

### 3.5.7 Spectral Efficiency

Lastly, in Fig. 3.8, we compare the sum-rate performance of the system when the channel estimates are obtained using the different algorithms, calculated using (3.18). The hybrid architecture in Section. 3.4.4 with genie-aided LMMSE channel estimates performs close to the fully digital architecture with the optimal genie aided LMMSE channel estimates. The sum-rate performance using Corr-SBL channel estimates is close to the genie-aided estimator. It offers nearly 50% higher sum rate compared to MSBL across the range of number of antennas. The performance of IPCI is limited by the accuracy of estimating the covariance itself, and as a consequence, it does not improve with the number of antennas.

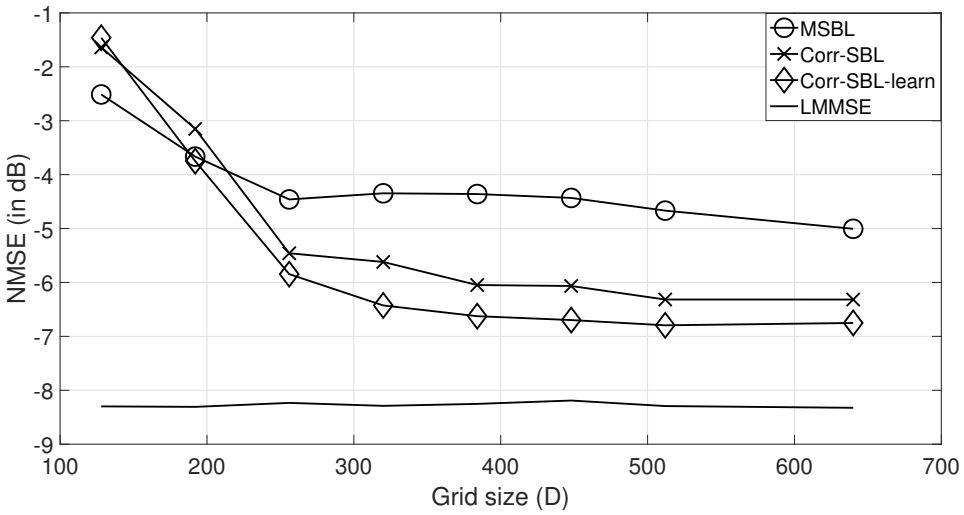



Fig. 3.9. NMSE performance under a practical clustered channel model (NYUSIM) with grid mismatch.

### 3.5.8 NYUSIM Channel Model

Lastly, we compare the performance of Corr-SBL and MSBL with the genie-aided LMMSE estimator for the channel model obtained from the NYUSIM mmWave channel simulator [58, 59], which considers a practical clustering model. For the NYUSIM model, we considered a carrier frequency of 28 GHz and a bandwidth of 800 MHz. The large and small scale fading parameters for the urban microcell scenario with LoS environment include: Path loss exponent = 2, transmit-receive separation of $100 - 500$ m, free space reference distance of 1 m, shadow fading standard deviation of 4 dB, mean angle of arrival of $1.9^\circ$ and mean excess delay of 123 ns. Since the NYUSIM model does not consider correlation among the path gains, we premultiply the obtained path gain vector with a matrix that results in the covariance matrix of the final vector to be similar to the model considered in this paper. The parameters of the simulation are same as given in Table 3.2, but with the number of multipaths obtained from the NYUSIM varying between 12 and 30 across different Monte Carlo simulations. In Fig. 3.9, we plot the $95^{\text{th}}$ percentile NMSE

(NMSE of top 95% runs, to remove outliers) as a function of the grid size. We can observe that as the grid size increases, the approximation of the channel using (3.2) becomes more accurate, and, as a result, the performance of the algorithm improves before it saturates beyond a grid size of about 350. Also, unlike the previous simulations, Corr-SBL-learn performs slightly better than Corr-SBL. This is because, with the practical channel model, the covariance matrix may not exactly match the $\mathbf{U}$ assumed by Corr-SBL. This shows that learning the correlation from data, rather than fixing it based on a model, can result in better performance in practice.

## 3.6 Chapter Summary

In this chapter, we explored the role of sparsity and intravector sparsity in the mmWave channel estimation problem. We presented the application of Corr-SBL algorithm described in the previous chapter for estimating the mmWave channel over multiple coherence blocks. Experimental results showed that Corr-SBL outperforms the existing approaches and achieves close to genie-aided optimal performance over a wide range of scenarios. The algorithm is also robust to imperfect correlation information. In practical implementations, the online version reduces the latency at a slight loss in NMSE performance.

# Chapter 4

# Modulo CS

The effect of dynamic range in data acquisition systems has been an important research topic in various domains of signal processing [61–64]. Systems with low dynamic range lead to signal loss due to clipping, and high dynamic range systems with finite resolution sampling are affected by high quantization noise. A direction of research in recent years to counter this problem has been the so-called self-reset analog to digital converters (SR-ADCs) [65, 66], which fold the amplitudes back into the dynamic range of the ADCs using the modulo arithmetic, thus mitigating the clipping effect. However, these systems encounter information loss due to the modulo operation. The transfer function of the SR-ADC with parameter $\lambda$ is given by

$$\mathcal{M}_\lambda(t) = 2\lambda \left( \left[\!\!\left[ \frac{t}{2\lambda} + \frac{1}{2} \right]\!\!\right] - \frac{1}{2} \right), \tag{4.1}$$

where $[\![t]\!] \triangleq t - \lfloor t \rfloor$ is the fractional part of $t$ [17].

The Nyquist-Shannon sampling theorem assumes infinite dynamic range and cannot be considered for the SR-ADCs, since it does not cater to the information loss due to the

The contents of this chapter have been published in IEEE signal processing letters [60].

modulo operation. In the context of SR-ADCs, an alternative sampling theory called the *unlimited sampling framework* was developed in [17, 18], which provides sufficient conditions on the sampling rate for guaranteeing the recovery of band-limited signals from its folded samples, provided a norm constraint independent of the dynamic range of ADC is satisfied. Extending these results, the work in [67] considered the inverse problem of recovering $K$ low pass filtered spikes in a continuous-time sparse signal, and developed a new sampling theorem and a signal recovery algorithm. In [68], the authors studied the quantization of oversampled signals in the SR-ADC architecture with the goal of reducing the overload distortion error.

A novel HDR imaging system that employs SR-ADCs to overcome limitations due to limited dynamic range was studied in [62, 69]. Mathematically, this involves applying an SR-ADC individually to multiple linear measurements of the images, and is termed as modulo compressed sensing (modulo-CS) [69]. Modulo-CS is an up-and-coming area with good promise in handling the large dynamic range inherent with many signal acquisition systems. While specific applications are yet to be developed, modulo-CS can be considered as an alternative sensing framework to avoid signal clipping where where compressed measurements of sparse signals are available, such as communication systems [64, 70], audio processing [71], biomedical and physiological signals [72]. Another application area is the phase unwrapping problem studied in applications like optical metrology [73], magnetic resonance imaging (MRI) [74] and synthetic aperture radar [75]. In the somewhat restrictive setting where the modulo-CS measurements are assumed to span at most two periods (the raw measurements lie in the interval $[-1, 1)$, while the modulo measurements are in $[0, 1)$), [69] proposes an algorithm and analyzes the sample complexity under Gaussian

measurement matrices. In [70], a generalized approximate message passing algorithm tailored to modulo-CS was proposed by assuming a Bernoulli-Gaussian distribution on the sparse signal. The results in these papers suggest that sparsity is very useful in the recovery of signals from modulo-CS measurements. However, a theoretical understanding of the minimal number of measurements for the identifiability of sparse signals from modulo-CS observations has not been developed to date.

In this chapter, we address the identifiability issue in the modulo-CS problem. The problem setup is presented in Section 4.1. In Section 4.2, we derive necessary and sufficient conditions on the measurement matrix under which sparse signals are identifiable under modulo-CS measurements. A novel algorithm for modulo-CS recovery is presented in Section 4.3 and theoretical guarantees for unique recovery are derived.

## 4.1 Modulo Compressed Sensing

In this section, we describe the modulo-CS problem and present an optimization problem for this setup.

Let $\mathbf{x} \in \mathbb{R}^N$ denote an $s$-sparse vector, i.e., $\|\mathbf{x}\|_0 \leq s$, with $s < \frac{N}{2}$. For ease of exposition, instead of SR-ADC transfer function given in (4.1), we consider an equivalent modular arithmetic which returns the fractional part of a real number, i.e., it returns $[\![t]\!] \triangleq t - \lfloor t \rfloor$. We obtain $m$ projections of $\mathbf{x}$ as follows:

$$z_i = [\![\langle \mathbf{a}_i, \mathbf{x} \rangle]\!], \ i = 1, 2, \ldots, m. \tag{4.2}$$

Usually, $m \leq N$ in the compressed sensing paradigm, but we will also present extensions to dense vectors ($s \geq \frac{N}{2}$) in the overdetermined system setup ($m > N$).

Stacking the projections $\langle \mathbf{a}_i, \mathbf{x} \rangle$ as a vector $\mathbf{y}$, we can rewrite (4.2) in a form similar to the CS framework as

$$\mathbf{z} = [\![\mathbf{y}]\!] = [\![\mathbf{A}\mathbf{x}]\!], \tag{4.3}$$

where $\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_m \end{bmatrix}^T \in \mathbb{R}^{m \times N}$ is the measurement matrix and $[\![\cdot]\!]$ represents the element wise modulo-1 operation on a vector, as before.

The non-linearity introduced by the modulo operation along with the underdetermined compressive measurements could lead to an indeterminate system, i.e., it may not have a unique solution. In this chapter, we explore the role of sparsity in uniquely recovering an $s$-sparse input signal $\mathbf{x}$ from the modulo-CS measurements $\mathbf{z}$ obtained using (4.3).

### $\mathrm{P}_0$ formulation

Any real valued vector $\mathbf{y} \in \mathbb{R}^m$ can be uniquely decomposed as $\mathbf{y} = \mathbf{z} + \mathbf{v}$, where $\mathbf{z} \in [0, 1)^m$ and $\mathbf{v} \in \mathbb{Z}^m$ denote the fractional part and integer part (the floor function) of $\mathbf{y}$, respectively. Using this decomposition, the non-linearity in (4.3) can be represented using a linear equation $\mathbf{A}\mathbf{x} = \mathbf{z} + \mathbf{v}$. The following proposition uses this decomposition to state an equivalent optimization problem ($\mathrm{P}_0$) for the unique recovery of $\mathbf{x}$ from $\mathbf{z}$.

**Proposition 4** ($\mathrm{P}_0$ optimization)**.** *Given $\mathbf{A} \in \mathbb{R}^{m \times N}$, an $s$-sparse $\mathbf{x} \in \mathbb{R}^N$, and modulo-CS measurements $\mathbf{z} \triangleq [\![\mathbf{A}\mathbf{x}]\!]$, the following statements are equivalent:*

1. *The vector $\mathbf{x}$ is a unique $s$-sparse vector $\mathbf{w}$ that satisfies $\mathbf{A}\mathbf{w} = \mathbf{y}$ with $\mathbf{y} = \mathbf{z} + \mathbf{v}$ and $\mathbf{v} \in \mathbb{Z}^m$.*

2. *The vector $\mathbf{x}$ is the unique solution of*

$$\arg\min_{\mathbf{w}, \mathbf{v}} \|\mathbf{w}\|_0 \text{ subject to } \mathbf{A}\mathbf{w} = \mathbf{z} + \mathbf{v}; \mathbf{v} \in \mathbb{Z}^m. \tag{$\mathrm{P}_0$}$$

*Proof.* (1) $\Rightarrow$ (2): Let $\mathbf{x}$ be the unique $s$-sparse solution of $\mathbf{A}\mathbf{w} = \mathbf{y}$ with $\mathbf{y} = \mathbf{z} + \mathbf{v}$ where $\mathbf{z} = [\![\mathbf{A}\mathbf{x}]\!]$ and $\mathbf{v} \in \mathbb{Z}^m$. Then a solution $\mathbf{x}^\#$ of ($\mathrm{P}_0$) is $s$-sparse and satisfies $\mathbf{A}\mathbf{x}^\# = \mathbf{z} + \mathbf{v}$, so that $\mathbf{x} = \mathbf{x}^\#$. The implication (2) $\Rightarrow$ (1) is direct. □

Thus, unique identifiability of an $s$-sparse $\mathbf{x}$ from modulo-CS measurements is equivalent to the existence of a unique solution to ($\mathrm{P}_0$), which we discuss next.

## 4.2 Identifiability

In this section, we derive conditions under which ($\mathrm{P}_0$) admits a unique $s$-sparse solution, which is presented in the following Lemma.

**Lemma 2** (Necessary and sufficient conditions)**.** *Any vector* $\mathbf{x}$ *satisfying* $\|\mathbf{x}\|_0 \leq s < \frac{N}{2}$ *is an unique solution to the optimization problem* ($\mathrm{P}_0$) *if and only if any* $2s$ *columns of matrix* $\mathbf{A}$ *are linearly independent of all* $\mathbf{v} \in \mathbb{Z}^m$*.*

*Proof.* We first prove sufficiency by contradiction. Let $\mathbf{z} = [\![\mathbf{A}\mathbf{x}]\!]$, $\mathbf{x}$ is an $s$-sparse vector, and $\mathbf{A} \in \mathbb{R}^{m \times N}$. Suppose the optimization problem ($\mathrm{P}_0$) returned another $s$-sparse vector $\mathbf{x}^\#$ (so that $\|\mathbf{x}^\#\|_0 \leq s$), then

$$\mathbf{A}(\mathbf{x} - \mathbf{x}^\#) = \mathbf{v} \Rightarrow \mathbf{A}_\mathcal{S}(\mathbf{x}_\mathcal{S} - \mathbf{x}^\#_\mathcal{S}) = \mathbf{v} \in \mathbb{Z}^m,$$

where the set $\mathcal{S}$ is the union of the supports of $\mathbf{x}$ and $\mathbf{x}^\#$. Since $|\mathcal{S}| \leq 2s$, a set of $2s$ columns of $\mathbf{A}$ span an integer vector $\mathbf{v}$, which violates the condition in the Lemma.

To prove the necessary part, suppose $\exists\, \mathcal{S}$ such that $|\mathcal{S}| = 2s$ and $\mathbf{0} \neq \mathbf{u} \in \mathbb{R}^{2s}$ such that $\mathbf{A}_\mathcal{S}\mathbf{u} = \mathbf{v}$, where $\mathbf{v} \in \mathbb{Z}^m$. We construct two $s$-sparse vectors $\mathbf{x}^0, \mathbf{x}^\# \in \mathbb{R}^N$ from $\mathbf{u}$, where the first $s$ indices of $\mathcal{S}$ constitute the support of $\mathbf{x}^0$ with the values equal to first $s$

entries of $\mathbf{u}$, and the remaining $s$ entries of $\mathcal{S}$ constitute the support of $\mathbf{x}^{\#}$ with the values equal to the corresponding $s$ entries of $\mathbf{u}$. Also, define $\mathbf{z} = [\![\mathbf{A}\mathbf{x}^0]\!]$ and $\mathbf{y}^0 = \mathbf{A}\mathbf{x}^0$, so that $\mathbf{y}^0 = \mathbf{z} + \mathbf{v}^0$ for some $\mathbf{v}^0 \in \mathbb{Z}^m$. Then, using $\mathbf{A}_S\mathbf{u} = \mathbf{v}$, we have $\mathbf{A}\mathbf{x}^{\#} + \mathbf{A}\mathbf{x}^0 = \mathbf{v}$ which implies $\mathbf{y}^0 - \mathbf{v} = -\mathbf{A}\mathbf{x}^{\#}$. Thus, $-\mathbf{x}^{\#}$ is also a solution to the optimization problem since $\|\mathbf{x}^{\#}\|_0 \leq s$ and $[\![-\mathbf{A}\mathbf{x}^{\#}]\!] = [\![\mathbf{y}^0 - \mathbf{v}]\!] = \mathbf{z}$, which is a contradiction. □

The following corollary presents a similar result for the recovery of dense vectors, which requires $m > N$ (overdetermined system).

**Corollary 1.** *Any vector* $\mathbf{x}$ *satisfying* $\|\mathbf{x}\|_0 \geq \frac{N}{2}$ *is a unique solution to* $\mathbf{A}\mathbf{w} = \mathbf{y}$ *with* $\mathbf{y} = [\![\mathbf{A}\mathbf{x}]\!] + \mathbf{v}$ *and* $\mathbf{v} \in \mathbb{Z}^m$ *if only if the columns of matrix* $\mathbf{A}$ *are linearly independent of all* $\mathbf{v} \in \mathbb{Z}^m$*. Consequently, the minimum number of measurements required for unique recovery is* $m = N + 1$*.*

*Proof.* The sufficiency part of the proof is similar to that of Lemma 2. The difference stems from the fact that the vector $\mathbf{x} - \mathbf{x}^{\#}$ can be any $N$ length vector in the dense case and hence, $\mathbf{A}\left(\mathbf{x} - \mathbf{x}^{\#}\right) = \mathbf{v}$ violates the condition that the columns of $\mathbf{A}$ are linearly independent of all $\mathbf{v} \in \mathbb{Z}^m$. For the necessary part, consider $\mathbf{u} \in \mathbb{R}^N$ such that $\mathbf{A}\mathbf{u} = \mathbf{v} \in \mathbb{Z}^m$ and a decomposition $\mathbf{u} = \mathbf{x} + \mathbf{x}^{\#}$. Similar to the proof of Lemma 2, it is evident that both $\mathbf{x}$ and $\mathbf{x}^{\#}$ result in the same modulo measurements. Hence, we have a contradiction, and the condition in the corollary is necessary. Finally, since the columns of $\mathbf{A}$ are linearly independent of all $\mathbf{v} \in \mathbb{Z}^m$, the minimum number of rows of $\mathbf{A}$ should be $N + 1$, thus proving the the corollary. □

To compare the modulo-CS problem to the standard CS problem, we state two necessary conditions for modulo-CS recovery in the corollary below.

**Corollary 2.** *The following two conditions are necessary for recovering any vector* $\mathbf{x}$ *satisfying* $\|\mathbf{x}\|_0 \leq s$ *as a unique solution of the optimization problem* $(\mathrm{P}_0)$*:*

1. *Any* $2s$ *columns of* $\mathbf{A}$ *are linearly independent., and*

2. $m \geq 2s + 1$.

*Proof.* From Lemma 2, for unique recovery of $s$-sparse signals from modulo-CS measurements, it is necessary that any $2s$ columns of $\mathbf{A}$ are linearly independent of all $\mathbf{v} \in \mathbb{Z}^m$. The first condition is necessary since if any $2s$ columns of $\mathbf{A}$ are not linearly independent, they cannot be linearly independent of an integer vector. Also, since any $2s+1$ vectors of dimension less than $2s+1$ are linearly dependent, $m \geq 2s+1$ is required for the necessary condition to hold true, and thus we arrive at the second condition. □

Comparing the above result to the necessary and sufficient conditions for the $(\mathrm{P}_{0-CS})$ optimization problem in standard compressed sensing problem discussed in Section 1.2.1, we can observe two important differences. First, the first condition in the corollary is both sufficient and necessary for the standard-CS, but in modulo-CS a stronger condition given in Lemma 2 is required for sufficiency. Second, the number of measurements $m = 2s$ is necessary and sufficient for unique sparse signal recovery in the standard CS setup [7]. Thus, we see that the penalty for unique sparse signal recovery due to the modulo operation is just one additional measurement. In fact, although the above corollary shows that the minimum number of measurements needed to reconstruct all $s$-sparse vector from its modulo measurements is $2s + 1$, in the following theorem we will show that $m = 2s + 1$ measurements are also sufficient.

**Theorem 5** (Sufficiency)**.** *For any $N \geq 2s+1$, there exists a matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ with $m = 2s+1$ rows such that every $s$-sparse $\mathbf{x} \in \mathbb{R}^N$ can be uniquely recovered from its modulo measurements $\mathbf{z} = [\![\mathbf{A}\mathbf{x}]\!]$ as a solution to* $(\mathrm{P}_0)$.

*Proof.* Let $\mathbf{A} \in \mathbb{R}^{(2s+1)\times N}$ be a matrix for which at least one $s$-sparse vector $\mathbf{x}$ cannot be recovered from its modulo measurements $\mathbf{z} = [\![\mathbf{A}\mathbf{x}]\!]$ via $(\mathrm{P}_0)$. Hence, $\mathbf{A}$ does not satisfy the condition in Lemma 2. We will show that the set of all such matrices is of Lebesgue measure 0. To this end, we define two sets:

1. Let $\mathcal{V} = \{\mathbf{v} | \mathbf{v} \in \mathbb{Z}^m\}$ denote the countably infinite set of all integer vectors.

2. Let $\mathfrak{S} = \{T | T \subset [N], |T| = 2s\}$ denote the set of all index sets on $[N]$ whose cardinality is $2s$. Note that the cardinality of $\mathfrak{S}$ is $\binom{N}{2s}$.

For a given $\mathbf{u} \in \mathcal{V}$ and $\mathcal{S} \in \mathfrak{S}$, construct $\mathbf{B}(\mathbf{u}, \mathcal{S}) = \begin{bmatrix} \mathbf{u} & \mathbf{A}_{\mathcal{S}} \end{bmatrix}$. Hence, the condition in Lemma 2 fails if $\det(\mathbf{B}(\mathbf{u}, \mathcal{S})) = 0$. This is a nonzero polynomial function of the entries of $\mathbf{A}_{\mathcal{S}}$, and therefore the set of matrices which satisfy this condition have Lebesgue measure 0. Now, consider $\cup_{\mathcal{S}\in\mathfrak{S}} \cup_{\mathbf{u}\in\mathcal{V}} \{\mathbf{A} | \det(\mathbf{B}(\mathbf{u}, \mathcal{S})) = 0\}$. This is a finite union of countable unions of Lebesgue measure 0 sets and hence is also of Lebesgue measure 0. Hence, a matrix $\mathbf{A}$ chosen outside of this set will ensure that any $s$-sparse vector $\mathbf{x}$ can be recovered from its modulo measurements $\mathbf{y} = [\![\mathbf{A}\mathbf{x}]\!]$. □

*Remark 1:* If the entries of $\mathbf{A}$ are drawn independently from any continuous distribution, $\mathbf{A}$ lies outside the set of Lebesgue measure 0 described in Theorem 5 and hence is a valid candidate for modulo-CS recovery.

*Remark 2:* From [76, Proposition 1], for any integer vector $\mathbf{a} \in \mathbb{Z}^K$ and $\mathbf{x} \in \mathbb{R}^K$ it holds that $[\![\mathbf{a}^T [\![\mathbf{x}]\!]]\!] = [\![\mathbf{a}^T \mathbf{x}]\!]$. As consequence, if the entries of $\mathbf{A}$ are integers, then the two

vectors $\mathbf{x}$ and $[\![\mathbf{x}]\!]$ result in the same modulo measurements, and hence a unique recovery is not possible. Hence, integer matrices cannot be used as candidate measurement matrices for modulo-CS.

*Remark 3:* Extending Theorem 5 to dense vectors similar to Corollary 1, it can be shown that $m = N + 1$ suffices for unique recovery of all $\mathbf{x} \in \mathbb{R}^N$.

We next study the recoverability of sparse vectors when the $\ell_0$-norm in ($\mathrm{P}_0$) is replaced with the $\ell_1$-norm, thus making the objective function convex.

## 4.3 Convex Relaxation Based Algorithm

In the previous section, we derived conditions for unique sparse vector recovery from modulo-CS measurements via ($\mathrm{P}_0$). However, both the objective function and the constraint set of ($\mathrm{P}_0$) are non-convex, and solving it requires an exhaustive search over all possible index sets and integer vectors of length $m$. Instead, in this section, we consider an alternative optimization problem using a convex relaxation to the objective function, and study its performance.

Replacing the $\ell_0$-norm in ($\mathrm{P}_0$) with the $\ell_1$-norm, we obtain the combinatorial optimization problem:

$$\arg\min_{\mathbf{x},\mathbf{v}} \|\mathbf{x}\|_1 \text{ subject to } \mathbf{A}\mathbf{x} = \mathbf{z} + \mathbf{v};\ \mathbf{v} \in \mathbb{Z}^m. \tag{$\mathrm{P}_1$}$$

### 4.3.1 Integer Range Space Property

In order to develop conditions on $\mathbf{A}$ for unique recoverability of the original sparse vector via ($\mathrm{P}_1$), we introduce the following property and show that it is both necessary and sufficient for the recovery of all $s$-sparse vectors using the ($\mathrm{P}_1$) problem.

**Definition 2** (Integer range space property (IRSP))**.** *A matrix* $\mathbf{A}$ *is said to satisfy the IRSP of order* $s$ *if, for all sets* $\mathcal{S} \subset [N]$ *with* $|\mathcal{S}| \leq s$*,*

$$\|\mathbf{u}_{\mathcal{S}}\|_1 < \|\mathbf{u}_{\mathcal{S}^{\mathrm{C}}}\|_1,$$

*holds for every* $\mathbf{u} \in \mathbb{R}^N$ *with* $\mathbf{A}\mathbf{u} = \mathbf{v} \in \mathbb{Z}^m$*.*

*Remark 4:* In the above, if the integer vector $\mathbf{v}$ is replaced with the all zero vector, the IRSP boils down to the null space property, which is necessary and sufficient for the $\ell_1$ norm based relaxation of the standard CS problem.

**Theorem 6** ($\ell_1$ recovery from modulo-CS)**.** *Every* $s$*-sparse* $\mathbf{x}$ *is the unique solution of* ($\mathrm{P}_1$) *if and only if the matrix* $\mathbf{A}$ *satisfies the IRSP of order* $s$*.*

*Proof.* Consider a fixed index set $\mathcal{S}$ with $|\mathcal{S}| \leq s$, and suppose that every $\mathbf{x}$ supported on $\mathcal{S}$ is a unique minimizer of ($\mathrm{P}_1$). Then, for any $\mathbf{u}$ such that $\mathbf{A}\mathbf{u} = \mathbf{v} \in \mathbb{Z}^m$, the vector $\mathbf{u}_{\mathcal{S}}$ is the unique minimizer of ($\mathrm{P}_1$). But, $\mathbf{A}(\mathbf{u}_{\mathcal{S}} + \mathbf{u}_{\mathcal{S}^{\mathrm{C}}}) = \mathbf{v}$. Thus, $\|\mathbf{u}_{\mathcal{S}}\|_1 < \|\mathbf{u}_{\mathcal{S}^{\mathrm{C}}}\|_1$, which proves the necessary condition. Conversely, suppose that the IRSP holds with respect to the set $\mathcal{S}$. Consider $\mathbf{x}$ supported on $\mathcal{S}$ and another vector $\mathbf{x}^{\#}$ that result in the same modulo measurements, i.e., $\mathbf{A}\mathbf{x} + \mathbf{v}_1 = \mathbf{A}\mathbf{x}^{\#} + \mathbf{v}_2$, where $\mathbf{v}_1$ and $\mathbf{v}_2$ are integer vectors. Letting $\mathbf{u} = \mathbf{x} - \mathbf{x}^{\#}$, the vector $\mathbf{A}\mathbf{u} = \mathbf{v}_2 - \mathbf{v}_1 = \mathbf{v} \in \mathbb{Z}^m$. Hence, by virtue of the IRSP, $\|\mathbf{u}_{\mathcal{S}}\|_1 < \|\mathbf{u}_{\mathcal{S}^{\mathrm{C}}}\|_1$. Then,

$$\begin{aligned}
\|\mathbf{x}\|_1 &\leq \|\mathbf{x} - \mathbf{x}^{\#}_{\mathcal{S}}\|_1 + \|\mathbf{x}^{\#}_{\mathcal{S}}\|_1 \\
&= \|\mathbf{u}_{\mathcal{S}}\|_1 + \|\mathbf{x}^{\#}_{\mathcal{S}}\|_1 \\
&< \|\mathbf{u}_{\mathcal{S}^{\mathrm{C}}}\|_1 + \|\mathbf{x}^{\#}_{\mathcal{S}}\|_1 = \| - \mathbf{x}^{\#}_{\mathcal{S}^{\mathrm{C}}}\|_1 + \|\mathbf{x}^{\#}_{\mathcal{S}}\|_1 = \|\mathbf{x}^{\#}\|_1.
\end{aligned}$$

Thus, $\mathbf{x}$ is the unique minimizer of ($\mathrm{P}_1$), and thus the IRSP relative to $\mathcal{S}$ is sufficient. Finally, letting $\mathcal{S}$ vary, we see that $\mathbf{A}$ satisfying the IRSP of order $s$ is necessary and sufficient. □

The above theorem shows that the IRSP is a key property for guaranteeing sparse vector recovery from modulo-CS measurements via ($\mathrm{P}_1$). In the following subsection, we restrict the possible set of $s$-sparse vectors catering to certain practical applications and provide conditions for recovery of the restricted set of vectors.

### 4.3.2 $\mathcal{L}$-restricted Integer Range Space Property

In practical applications, one can usually assume that there exists an integer $l$ for which the measurements $\mathbf{y} = \mathbf{A}\mathbf{x}$ satisfy $\|\mathbf{y}\|_\infty < l$, which results in the measurements spanning a maximum of $2l$ modulo periods, thus reducing the search space for the vector $\mathbf{v}$ to $(2l)^m$ possible integer vectors. For this setup, the ($\mathrm{P}_1$) problem can be modified (denoted as $\mathrm{P}_{1,l}$ problem) to include the additional constraint $\|\mathbf{A}\mathbf{x}\|_\infty < l$ to restrict the set of feasible solutions, resulting in:

$$\begin{aligned} &\arg\min_{\mathbf{x},\mathbf{v}} \|\mathbf{x}\|_1 \\ &\text{subject to } \mathbf{A}\mathbf{x} = \mathbf{z} + \mathbf{v};\ \mathbf{v} \in \mathbb{Z}^m;\ \|\mathbf{A}\mathbf{x}\|_\infty < l \end{aligned} \tag{$\mathrm{P}_{1,l}$}$$

We define a restricted integer range space property to study the guarantees for this setup.

**Definition 3** ($\mathcal{L}$-restricted integer range space property ($\mathcal{L}$-restricted IRSP))**.** *A matrix* $\mathbf{A}$ *is said to satisfy the* $\mathcal{L}$*-restricted integer range space property of order* $s$ *if for all sets* $\mathcal{S} \subset [N]$ *with* $|\mathcal{S}| \leq s$*,*

$$\|\mathbf{u}_{\mathcal{S}}\|_1 < \|\mathbf{u}_{\mathcal{S}^{\mathrm{C}}}\|_1,$$

*holds for every* $\mathbf{u} \in \mathcal{L} \subseteq \{\mathbf{u}|\mathbf{A}\mathbf{u} = \mathbf{v} \in \mathbb{Z}^m\}$.

We introduce three sets and derive necessary and sufficient conditions for the $\mathrm{P}_{1,l}$ problem using the restricted integer range space property with respect to these sets.

**Proposition 5.** *Consider two sets* $\mathcal{L}_k$ *and* $\mathcal{K}_{k,\mathcal{S}}$ *defined for* $k \in \mathbb{Z}$ *and* $\mathcal{S} \subseteq [N]$*:*

1. $\mathcal{L}_k = \{\mathbf{u}|\mathbf{A}\mathbf{u} = \mathbf{v} \in \mathbb{Z}^m, \|\mathbf{v}\|_\infty \leq k\}$

2. $\mathcal{K}_{k,\mathcal{S}} = \{\mathbf{u}|\mathbf{A}\mathbf{u} = \mathbf{v} \in \mathbb{Z}^m, \|\mathbf{A}\mathbf{u}_{\mathcal{S}}\|_\infty < k, \|\mathbf{A}\mathbf{u}_{\mathcal{S}^{\mathrm{C}}}\|_\infty < k\}$

*Then* $\mathcal{K}_{k,\mathcal{S}} \subset \mathcal{L}_{2k-1} \forall \mathcal{S}$ *and* $\mathcal{K}_k = \bigcup_{\mathcal{S}:|\mathcal{S}|\leq s}(\mathcal{K}_{k,\mathcal{S}}) \subseteq \mathcal{L}_{2k-1}$.

*Proof.* Let $\mathbf{u} \in \mathcal{K}_{k,\mathcal{S}}$. Then,

$$\begin{aligned}\|\mathbf{v}\|_\infty &= \|\mathbf{A}\mathbf{u}\|_\infty \\ &\leq \|\mathbf{A}\mathbf{u}_{\mathcal{S}}\|_\infty + \|\mathbf{A}\mathbf{u}_{\mathcal{S}^{\mathrm{C}}}\|_\infty < 2k\end{aligned}$$

Since $\mathbf{v} \in \mathbb{Z}^m$, the above strict inequality can be rewritten as $\|\mathbf{v}\|_\infty \leq 2k - 1$. Hence $\mathbf{u} \in \mathcal{L}_{2k-1}$. This holds true for all $\mathcal{S}$ and hence if $\mathbf{u}$ belongs to union over all $\mathcal{S}$ with $2s$ entries, it belongs to $\mathcal{L}_{2k-1}$. Hence both $\mathcal{K}_{k,\mathcal{S}}$ and $\mathcal{K}_k$ are subsets of $\mathcal{L}_{2k-1}$. □

The following theorem presents the conditions on $\mathbf{A}$ for unique recovery of $s$-sparse vectors from $(\mathrm{P}_{1,l})$ optimization problem.

**Theorem 7.** *Given a matrix* $\mathbf{A} \in \mathbb{R}^{m\times N}$*, the guarantees for unique recovery of every* $s$*-sparse vector* $\mathbf{x}$ *as a solution to* $(\mathrm{P}_1)$ *with the additional constraint* $\|\mathbf{A}\mathbf{x}\|_\infty < l$ *are:*

- *Necessary condition:* $\mathbf{A}$ *satisfies* $\mathcal{K}_{l,\mathcal{S}}$*-restricted IRSP for all sets* $\mathcal{S}$ *with* $|\mathcal{S}| \leq s$*. Equivalently,* $\mathbf{A}$ *satisfies* $\mathcal{K}_l$*-restricted IRSP.*

- *Sufficient condition:* $\mathbf{A}$ *satisfies* $\mathcal{L}_{2l-1}$*-restricted IRSP.*

*Proof.* We first prove the necessary condition. Given a fixed index set $\mathcal{S}$, let us assume that every $s$-sparse vector $\mathbf{x}$ supported on $\mathcal{S}$ satisfying the condition $\|\mathbf{Ax}\|_\infty < l$ is a unique minimizer of the $(\mathrm{P}_{1,l})$ optimization problem. Then for any $\mathbf{u} \in \mathcal{K}_{k,\mathcal{S}}$, $\mathbf{u}_\mathcal{S}$ is the unique minimizer of the $(\mathrm{P}_{1,l})$ problem. But, $\mathbf{A}(\mathbf{u}_\mathcal{S} + \mathbf{u}_{\mathcal{S}^\mathrm{C}}) = \mathbf{v}$. Thus, $\|\mathbf{u}_\mathcal{S}\|_1 < \|\mathbf{u}_{\mathcal{S}^\mathrm{C}}\|_1$. This holds for all $\mathcal{S}$ with $|\mathcal{S}| \leq s$, thus establishing the necessary condition.

To prove sufficient condition, let us assume that the $\mathcal{L}_{2l-1}$-restricted IRSP holds. Then for a given index set $\mathcal{S}$ of order $s$, consider two vectors $\mathbf{x}$ supported on $\mathcal{S}$ and $\mathbf{x}^\#$ which satisfy the condition $\|\mathbf{Ax}\|_\infty < l$ and $\|\mathbf{Ax}^\#\|_\infty < l$ that result in the same modulo measurements, i.e. $\mathbf{Ax}+\mathbf{v}_1 = \mathbf{Ax}^\#+\mathbf{v}_2$ (where $\mathbf{v}_1$ and $\mathbf{v}_2$ are integer vectors, which take values in $[-l+1, l]$ due to the norm condition). Considering $\mathbf{u} = \mathbf{x}-\mathbf{x}^\#$, the integer vector $\mathbf{v} = \mathbf{Au} = \mathbf{v}_2-\mathbf{v}_1$, can take values in $[-(2l-1), (2l-1)]$. Then, since $\mathbf{u} \in \mathcal{L}_{2l-1}$,

$$\begin{aligned}
\|\mathbf{x}\|_1 &\leq \|\mathbf{x} - \mathbf{x}^\#_\mathcal{S}\|_1 + \|\mathbf{x}^\#_\mathcal{S}\|_1 \\
&= \|\mathbf{u}_\mathcal{S}\|_1 + \|\mathbf{x}^\#_\mathcal{S}\|_1 \\
&< \|\mathbf{u}_{\mathcal{S}^\mathrm{C}}\|_1 + \|\mathbf{x}^\#_\mathcal{S}\|_1 = \| - \mathbf{x}^\#_{\mathcal{S}^\mathrm{C}}\|_1 + \|\mathbf{x}^\#_\mathcal{S}\|_1 = \|\mathbf{x}^\#\|_1
\end{aligned}$$

Thus, $\mathbf{x}$ is the unique minimizer of $(\mathrm{P}_1)$. This holds for all possible index sets $\mathcal{S}$ of order $s$, proving the sufficient condition. □

The necessary and sufficient conditions do not coincide when $k$ is finite. However, using the relation between $\mathcal{K}_l$ and $\mathcal{L}_{2l-1}$ as given in Proposition 5, the gap between two conditions in Theorem 7 can be characterized by whether $\mathbf{A}$ satisfies the $\{\mathcal{L}_{2l-1} \setminus \mathcal{K}_l\}$-restricted IRSP. At $l = \infty$, i.e., when the restriction $\|\mathbf{Ax}\|_\infty < l$ is removed, the set characterizing the gap

is a null set and hence both conditions boil down to the integer range space property. It can also be observed that for the limiting condition $l \to 0$, both the necessary and sufficient conditions boil down to the null space property used in the standard CS problem.

Both the IRSP and $\mathcal{L}$-restricted IRSP were shown to be a key-property for obtaining guarantees for the sparse recovery from modulo measurements via the $\ell_1$ relaxation problem. Next, we present a practical algorithm for solving ($\mathrm{P}_{1,l}$) based on the mixed-integer linear programming problem and empirically evaluate its performance.

### 4.3.3 Mixed Integer Linear Program (MILP)

In this subsection, we develop an algorithm to solve the ($\mathrm{P}_{1,l}$) problem. First, note that since $\mathbf{z} \in [0,1)^m$, the constraint $\|\mathbf{A}\mathbf{x}\|_\infty < l$ can be written as a bound on the extreme values of entries of the integer vector $\mathbf{v}$. Also, the $\ell_1$ norm can be rewritten as a linear function using two positive vectors $\mathbf{x}^+$ and $\mathbf{x}^-$ such that $\mathbf{x} = \mathbf{x}^+ - \mathbf{x}^-$. This leads to the MILP formulation:

$$\begin{aligned} &\min_{\mathbf{x}^+,\mathbf{x}^-,\mathbf{v}} \ \mathbf{1}^T\left(\mathbf{x}^+ + \mathbf{x}^-\right) \\ &\text{subject to } \begin{bmatrix}\mathbf{A} & -\mathbf{A} & -\mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{x}^+\\ \mathbf{x}^-\\ \mathbf{v}\end{bmatrix} = \mathbf{z}; \\ &\mathbf{v} \in \mathbb{Z}^m; \quad \mathbf{x}^+, \mathbf{x}^- \geq 0. \end{aligned} \tag{4.4}$$

The MILP can be solved efficiently using the branch-and-bound algorithm [77]. Once $\mathbf{x}^+$ and $\mathbf{x}^-$ are obtained, we can solve for $\mathbf{x}$ as $\mathbf{x} = \mathbf{x}^+ - \mathbf{x}^-$.

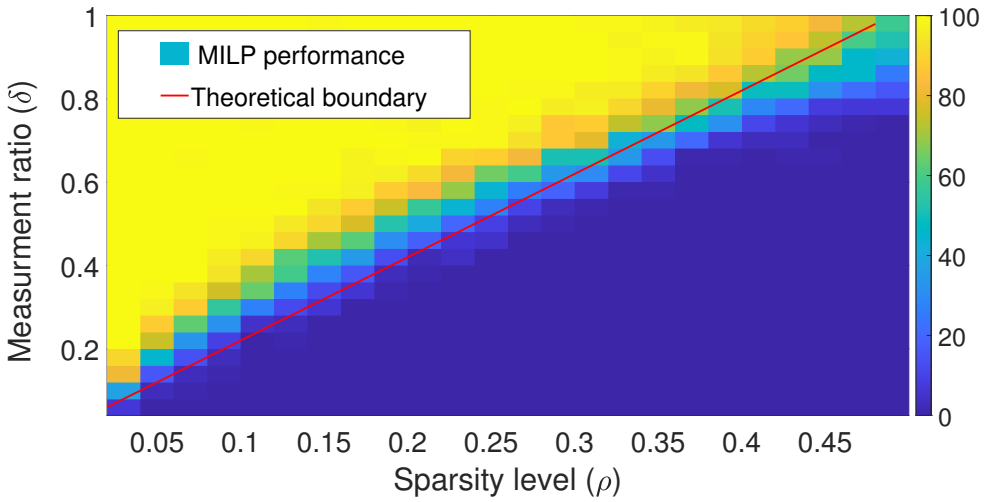



Fig. 4.1. Percentage of success recovery for MILP.

## 4.4 Simulation Results

We now empirically evaluate performance of the MILP for the modulo-CS problem. We solve the MILP using the *intlinprog* function in the MATLAB optimization toolbox. We set $N = 50$, and for a given sparsity level $\rho = \frac{s}{N}$, we randomly select $\rho N$ indices of the input signal to be nonzero, and set the others to zero. The nonzero entries are drawn from either the uniform or Gaussian distributions with zero-mean and different variances as specified in the figures. For a given measurement ratio $\delta = \frac{m}{N}$, the entries of the measurement matrix with $\delta N$ rows are drawn from an i.i.d. Gaussian distribution with mean zero and variance $\frac{1}{m}$.

We first present the phase transition curve of the MILP problem by plotting the success rate over 1000 Monte Carlo simulations when the nonzero entries of the sparse signal are obtained from a uniform distribution on $[-1, 1]$, denoted by Unif $[-1, 1]$. From Fig. 4.1, we see that the transition region between success and failure roughly follows the theoretical result (solid line in red) in Theorem 5. In particular, the performance of the MILP based

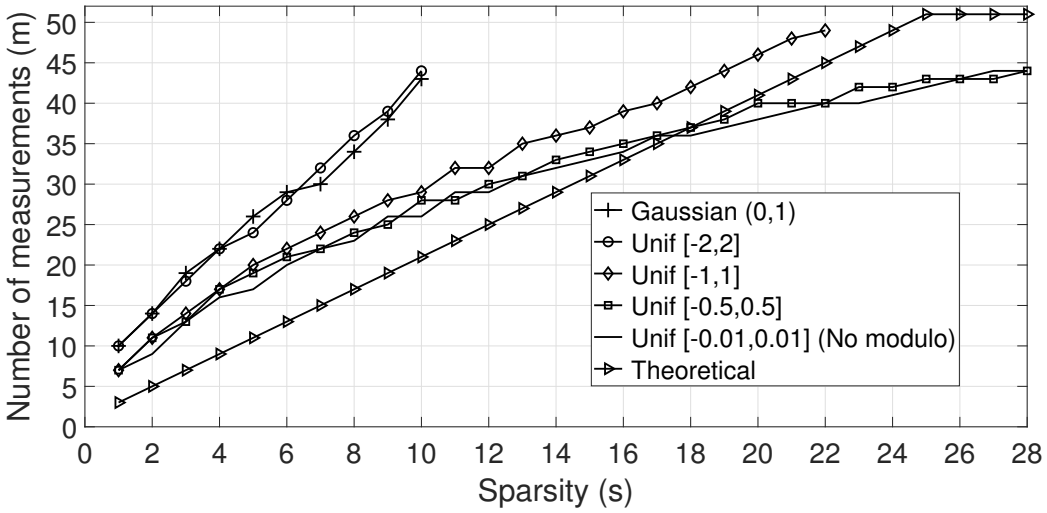



Fig. 4.2. Phase transition curves for 80% recovery accuracy.

algorithm, which is based on convex relaxation, is close to the theoretical bound for the ($P_0$) problem, i.e., it is near-optimal in the settings considered here.

In the next experiment, we compare the performance of the MILP algorithm when the measurements span different number of modulo periods. To this end, we evaluate the performance of the MILP for different distributions on the nonzero entries of the sparse signal. We plot the minimum value of $m$ required for 80% recovery accuracy (i.e., exact recovery of the sparse signals in 80% of the random experiments) for sparsity levels varying from 1 to $\frac{N}{2}$ in Fig. 4.2. For the Unif $[-0.01, 0.01]$ curve, the measurements are always in the range $[-0.5, 0.5]$, and by shifting by 0.5, we obtain all measurements within a single modulo period, hence the modulo operation does not introduce any nonlinearity. As seen in the figure, the curve for Unif $[-0.5, 0.5]$ which spans at least 2 modulo periods is close to the Unif $[-0.01, 0.01]$ case without the modulo operation. When the variance of the signal is low, MILP performs close to the theoretical limit for the minimal number of measurements required. However, with increase in variance of the input signal, the measurements span a larger number of modulo periods, and the performance starts to

deteriorate. We also notice that the simulated curves for the Unif $[-0.01, 0.01]$ and Unif $[-0.5, 0.5]$ cases cross the theoretical bound for sparsity levels beyond $s = 18$. There are two reasons for this. First, the simulated curves correspond to 80% recovery success rate, while the theoretical results were for perfect recovery of all sparse signals. Second, the simulated curves are for specific source distributions, while the theoretical result is for arbitrary (even adversarially chosen) sparse vectors. Nonetheless, the theoretical curve forms a useful benchmark for the performance of modulo-CS recovery algorithms.

## 4.5 Chapter Summary

In this chapter, we considered the problem of recovering sparse signals from modulo compressed sensing measurements. We presented an equivalent optimization problem for the modulo-CS setup using the $\ell_0$ norm. For this optimization problem, we showed that the $2s + 1$ measurements are necessary and sufficient for recovering $s$-sparse signals. Finally, we considered a convex relaxation for the $\ell_0$-norm and presented an algorithm based on mixed-integer linear programming. For this algorithm, we obtained theoretical guarantees as a property of the measurement matrix.

# Chapter 5

# Conclusion

In this final chapter, we present the key contributions presented so far, and discuss the significance of these results for the channel estimation problem. We then conclude by discussing some extensions to these results and open questions that are promising for future work.

## 5.1 Summary of Contributions

The problems addressed in this thesis can be structured into two setups; sparse signal recovery in the presence of intra-vector correlation and sparse signal recovery from modulo measurements. We list the specific contributions associated with each setup in the following subsections.

### Sparse signal recovery in the presence of intra-vector correlation

In this setup, we explored the sparse recovery problem when the nonzero entries of the sparse vector are correlated with each other, termed as correlated sparse recovery. To this end, we first explored the covariance matching framework and presented the effect

of intra-vector correlation on the algorithms which neglect the presence of correlation. To obtain better recovery performance by exploiting the correlation, we presented a novel prior model that has the potential to promote both sparsity and correlation. We developed a Bayesian inference algorithm named `Corr-SBL` for recovering correlated sparse vectors in an MMV setting, and showed that Corr-SBL comes under the purview of covariance matching algorithms. For the case with imperfect correlation information, we presented an approach for learning the correlation. The framework developed in this setup has the potential to be useful in applications such as imaging [78], face recognition [79], biomedical and physiological signals [80,81], where the intra-vector correlation can be useful to better fit the structures in the signal. To study the performance of the proposed algorithm, we considered a case study of mmWave channel estimation problem under scenarios such as rain attenuation, where the channel estimation can be represented as a correlated sparse recovery problem. Using the plug-in LMMSE structure of covariance matching algorithms, we provided analytical expressions for the normalized mean square error and spectral efficiency of the system and discussed the efficacy of the Corr-SBL prior. Experimental results for both simulated data and a practical channel model suggested that Corr-SBL outperforms the existing approaches and achieves close to genie-aided optimal performance in a wide range of scenario. For practical implementation, an online version of the algorithm was considered which reduces latency at a slight loss in NMSE.

## Sparse signal recovery from modulo measurements

This setup was motivated by the use of low resolution ADCs, where the quantization range is small to reduce quantization errors. To counter the effect of clipping that can occur

if input crosses the range, we considered the measurement model where the compressed measurements of the signal are observed through modulo operation, termed as modulo-CS. We presented an optimization problem for the recovery of sparse vectors under this setup using the $\ell_0$-norm. We then derived conditions on the measurement matrix to guarantee unique recovery from the optimization problem and showed that $2s+1$ measurements are necessary and sufficient for recovery of all $s$-sparse signals. For algorithmic tractability, we presented convex relaxation of the $\ell_0$-norm using $\ell_1$-norm and formulated a mixed integer linear program (MILP) to recover the sparse signals. We also presented a property of the measurement matrix called integer range space property and derived recovery guarantees for the MILP algorithm.

## 5.2 Future Work

Some possible avenues for future work related to the results presented in the thesis are:

- **Correlated sparse recovery**

  The Corr-SBL algorithm uses a pragmatic approach to learn the correlation. A principled method to learn the correlation can possibly lead to faster convergence of the algorithm. Furthermore, the Corr-SBL framework can be used to derive model-based deep learning approaches that can lead to faster recovery algorithms that perform similar to Corr-SBL.

- **Extensions in mmWave channel estimation application**

  Exploiting the inter-vector correlation among channels across blocks in addition to intra-vector correlation can be beneficial. Similarly, exploiting correlation among sub-carriers when Corr-SBL is applied to sub-carriers in an OFDM system can be

useful in frequency-selective channels. Off-grid based Corr-SBL algorithm to solve grid mismatch issues is also an important direction for future work.

- **Modulo-CS recovery**

  The thesis considered modulo-CS in a noiseless setting. For practical purposes, it is important to extend the setup for noisy scenarios and design algorithms for the same. In an algorithmic standpoint, applications of frameworks such as sparse Bayesian learning, variational Bayes can result in better recovery performance. Further, deriving sample complexity of these algorithms including the MILP algorithm is an interesting direction for future work.

- **Modulo quantizers** The main application of modulo-CS is for the design of low-resolution ADCs. Towards this, integrating finite resolution quantizers with the modulo-CS framework will be important. The tradeoff between the number of folds and the quantization levels will be a key question to be studied. Design of optimal quantizers, algorithms to recover from quantized measurements in the modulo-CS scenario are few of the other directions for future work in this emerging area.

To conclude, the results in the thesis have potential in a number of applications, with a common application area and the main motivation of this thesis being the mmWave channel estimation problem. While the thesis presented a framework for exploiting correlation, measurement campaigns to model the correlation structure seen in practice is required to fine-tune the algorithms for practical implementation. Further, integrating both correlated sparse recovery and modulo-CS for the channel estimation problem will pose additional challenges in the design of the communication systems, and is an important direction for further work due to the role of mmWave systems in the future of cellular communication.

# Bibliography


[1] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, 2014.

[2] S. Mumtaz, J. Rodriguez, and L. Dai, *mmWave Massive MIMO: A Paradigm for 5G*. Cambridge, Massachusetts: Academic Press, 2016.

[3] D. Gesbert, M. Kountouris, R. W. Heath, C. Chae, and T. Salzer, "Shifting the MIMO paradigm," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 36–46, 2007.

[4] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, 2014.

[5] J. Zhang, L. Dai, X. Li, Y. Liu, and L. Hanzo, "On low-resolution ADCs in practical 5G millimeter-wave massive mimo systems," *IEEE Commun. Mag.*, vol. 56, pp. 205–211, 2018.

[6] B. D. Rao, Z. Zhang, and Y. Jin, "Sparse signal recovery in the presence of intra-vector and inter-vector correlation," in *Proc. SPCOM*, 2012, pp. 1–5.

[7] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Basel: Birkhäuser, 2013.

[8] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.

[9] Z. Zhang and B. D. Rao, "Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 2009–2015, 2013.

[10] D. P. Wipf and B. D. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3704–3716, Jul. 2007.

[11] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.

[12] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, "Distributed compressed sensing of jointly sparse signals," in *Proc. Asilomar Conf. on Signals, Syst., and Comput.*, 2005, pp. 1537–1541.

[13] Z. Zhang and B. D. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 912–926, 2011.

[14] H. Zhang, S. Venkateswaran, and U. Madhow, "Channel modeling and MIMO capacity for outdoor millimeter wave links," in *Proc. WCNC*, 2010, pp. 1–6.

[15] M. K. Samimi, G. R. MacCartney, S. Sun, and T. S. Rappaport, "28 GHz millimeter-wave ultrawideband small-scale fading models in wireless channels," in *IEEE Veh. Technol. Conf.*, May 2016, pp. 1–6.

[16] L. Sanguinetti, E. Björnson, and J. Hoydis, "Toward massive MIMO 2.0: Understanding spatial correlation, interference suppression, and pilot contamination," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 232–257, 2020.

[17] A. Bhandari, F. Krahmer, and R. Raskar, "On unlimited sampling," in *Proc. SampTA*, 2017, pp. 31–35.

[18] A. Bhandari and F. Krahmer, "On identifiability in unlimited sampling," in *Proc. SampTA*, 2019, pp. 1–4.

[19] D. Prasanna and C. R. Murthy, "On the role of sparsity and intra-vector correlation in mmwave channel estimation," in *Proc. SPAWC*, May 2020, pp. 1–5.

[20] D. Prasanna and C. R. Murthy, "mmWave channel estimation via compressive covariance estimation: Role of sparsity and intra-vector correlation," *IEEE Trans. Signal Process.*, vol. 69, pp. 2356–2370, 2021.

[21] M. Azizyan, A. Krishnamurthy, and A. Singh, "Extreme compressive sampling for covariance estimation," *IEEE Trans. Inf. Theory*, vol. 64, no. 12, pp. 7613–7635, Dec. 2018.

[22] S. L. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 369–371, Sep. 2001.

[23] J. Lee Rodgers and W. A. Nicewander, "Thirteen ways to look at the correlation coefficient," *The Amer. Statistician*, vol. 42, no. 1, pp. 59–66, 1988.

[24] S. Khanna and C. R. Murthy, "On the support recovery of jointly sparse gaussian sources using sparse bayesian learning," *arXiv preprint arXiv:1703.04930*, 2020.

[25] S. Haghighatshoar and G. Caire, "Multiple measurement vectors problem: A decoupling property and its applications," *CoRR*, vol. abs/1810.13421, 2018.

[26] P. Pal and P. P. Vaidyanathan, "Pushing the limits of sparse support recovery using correlation information," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 711–726, 2015.

[27] S. Khanna and C. R. Murthy, "Rényi divergence based covariance matching pursuit of joint sparse support," in *Proc. SPAWC*, 2017, pp. 1–5.

[28] J. Fang, Y. Shen, H. Li, and P. Wang, "Pattern-coupled sparse Bayesian learning for recovery of block-sparse signals," *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 360–372, 2015.

[29] S. Kay, *Fundamentals of Statistical Signal Processing: Detection theory*, ser. Fundamentals of Statistical Si. Prentice-Hall PTR, 1998. [Online]. Available: https://books.google.co.in/books?id=vA9LAQAAIAAJ

[30] C. M. Bishop, *Pattern recognition and machine learning*. New York, NY: Springer, 2006.

[31] R. M. Neal and G. E. Hinton, *A View of the EM Algorithm that Justifies Incremental, Sparse, and other Variants*. Dordrecht: Springer Netherlands, 1998, pp. 355–368.

[32] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[33] M. Cudak, T. Kovarik, T. A. Thomas, A. Ghosh, Y. Kishiyama, and T. Nakamura, "Experimental mmWave 5G cellular system," in *Proc. Globecom*, 2014, pp. 377–381.

[34] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.

[35] J. Palacios, D. De Donno, and J. Widmer, "Tracking mm-Wave channel dynamics: Fast beam training strategies under mobility," in *IEEE Conf. on Comput. Commun. (INFOCOM)*, 2017, pp. 1–9.

[36] J. Singh and S. Ramakrishna, "On the feasibility of codebook-based beamforming in millimeter wave systems with multiple antenna arrays," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2670–2683, 2015.

[37] J. Lee, G. Gil, and Y. H. Lee, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2370–2386, June 2016.

[38] K. Venugopal, A. Alkhateeb, N. González Prelcic, and R. W. Heath, "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1996–2009, 2017.

[39] M. S. Dastgahian and H. Khoshbin, "Rank-defective millimeter-wave channel estimation based on subspace-compressive sensing," *Digit. Commun. Networks*, vol. 2, no. 4, pp. 206 – 217, 2016.

[40] A. Manoj and A. P. Kannu, "Multi-user millimeter wave channel estimation using generalized block OMP algorithm," in *Proc. SPAWC*, 2017, pp. 1–5.

[41] J. P. González-Coma et al., "FDD channel estimation via covariance estimation in wideband massive MIMO systems," *Sensors (Basel)*, vol. 20, no. 930, pp. 1–20, Feb. 2020.

[42] C. K. Anjinappa, A. C. Gürbüz, Y. Yapıcı, and I. Güvenç, "Off-grid aware channel and covariance estimation in mmwave networks," *IEEE Trans. Commun.*, vol. 68, no. 6, pp. 3908–3921, Jun. 2020.

[43] C. T. Neil, A. Garcia-Rodriguez, P. J. Smith, P. A. Dmochowski, C. Masouros, and M. Shafi, "On the performance of spatially correlated large antenna arrays for millimeter-wave frequencies," *IEEE Trans. Antennas Propag.*, vol. 66, no. 1, pp. 132–148, 2018.

[44] X. Li, J. Fang, H. Li, and P. Wang, "Millimeter wave channel estimation via exploiting joint sparse and low-rank structures," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 1123–1133, 2018.

[45] S. Srivastava, A. Mishra, A. Rajoriya, A. K. Jagannatham, and G. Ascheid, "Quasi-static and time-selective channel estimation for block-sparse millimeter wave hybrid MIMO systems: Sparse Bayesian learning (SBL) based approaches," *IEEE Trans. Signal Process.*, vol. 67, no. 5, pp. 1251–1266, Mar. 2019.

[46] S. Park and R. W. Heath, "Spatial channel covariance estimation for the hybrid MIMO architecture: A compressive sensing-based approach," *IEEE Trans. Wireless Commun.*, vol. 17, pp. 8047–8062, Dec. 2018.

[47] V. K. Sakarellos, D. Skraparlis, A. D. Panagopoulos, and J. D. Kanellopoulos, "Cooperative diversity performance in millimeter wave radio systems," *IEEE Trans. Commun.*, vol. 60, pp. 3641–3649, Dec. 2012.

[48] M. Cheffena, L. E. Braten, and T. Ekman, "On the space-time variations of rain attenuation," *IEEE Trans. Antennas Propag.*, vol. 57, no. 6, pp. 1771–1782, Jun. 2009.

[49] Y. W. E. Chan, Cam Chung La, and B. Soong, "Comparative study of correlated shadowing loss model for wireless sensor networks," in *Proc. Int. Conf. Inf. Commun., and Signal Process.*, Dec. 2013, pp. 1–5.

[50] C. Huang, L. Liu, C. Yuen, and S. Sun, "Iterative channel estimation using lse and sparse message passing for mmwave MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 1, pp. 245–259, 2019.

[51] H. Tang, J. Wang, and L. He, "Off-grid sparse Bayesian learning-based channel estimation for mmWave massive MIMO uplink," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 45–48, 2019.

[52] I. Viering, H. Hofstetter, and W. Utschick, "Spatial long-term variations in urban, rural and indoor environments," in *Proc. COST*, vol. 273. Citeseer, 2002.

[53] D. Kudathanthirige and G. Amarasuriya, "Sum rate analysis of massive MIMO downlink with hybrid beamforming," in *Proc. Globecom*, 2017, pp. 1–7.

[54] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *arXiv preprint arXiv:1912.10557*, 2019.

[55] R. J. Peter and C. R. Murthy, "Learned-sbl: A deep learning architecture for sparse signal recovery," *arXiv preprint arXiv:1909.08185*, 2019.

[56] A. A. Nasir, H. D. Tuan, T. Q. Duong, H. V. Poor, and L. Hanzo, "Hybrid beamforming for multi-user millimeter-wave networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 2943–2956, 2020.

[57] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge: Cambridge University Press, 2016.

[58] M. K. Samimi and T. S. Rappaport, "3-D millimeter-wave statistical channel model for 5g wireless system design," *IEEE Trans. Microw. Theory Tech.*, vol. 64, no. 7, pp. 2207–2225, 2016.

[59] S. Ju, O. Kanhere, Y. Xing, and T. S. Rappaport, "A millimeter-wave channel simulator NYUSIM with spatial consistency and human blockage," in *Proc. Globecom*, 2019, pp. 1–6.

[60] D. Prasanna, C. Sriram, and C. R. Murthy, "On the identifiability of sparse vectors from modulo compressed sensing measurements," *IEEE Signal Processing Letters*, vol. 28, pp. 131–134, 2021.

[61] F. Esqueda, S. Bilbao, and V. Välimäki, "Aliasing reduction in clipped signals," *IEEE Trans. Signal Process.*, vol. 64, no. 20, pp. 5255–5267, 2016.

[62] H. Zhao, B. Shi, C. Fernandez-Cull, S. Yeung, and R. Raskar, "Unbounded high dynamic range photography using a modulo camera," in *Proc. Int. Conf. on Comput. Photography*, 2015, pp. 1–10.

[63] W. Chou and R. M. Gray, "Modulo sigma-delta modulation," *IEEE Trans. Commun.*, vol. 40, no. 8, pp. 1388–1395, 1992.

[64] N. Kolomvakis, T. Eriksson, M. Coldrey, and M. Viberg, "Reconstruction of clipped signals in quantized uplink massive MIMO systems," *IEEE Trans. Commun.*, vol. 68, no. 5, pp. 2891–2905, 2020.

[65] W. Kester, "ADC architectures VI: Folding ADCs (MT-025 tutorial)," *Analog Devices, Technical Report*, vol. 27, pp. 623–656, 2009.

[66] Jehyuk Rhee and Youngjoong Joo, "Wide dynamic range CMOS image sensor with pixel level ADC," *Electron. Lett.*, vol. 39, no. 4, pp. 360–361, 2003.

[67] A. Bhandari, F. Krahmer, and R. Raskar, "Unlimited sampling of sparse signals," in *Proc. ICASSP*, 2018, pp. 4569–4573.

[68] O. Ordentlich, G. Tabak, P. K. Hanumolu, A. C. Singer, and G. W. Wornell, "A modulo-based architecture for analog-to-digital conversion," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 5, pp. 825–840, 2018.

[69] V. Shah and C. Hegde, "Signal reconstruction from modulo observations," in *Proc. IEEE GlobalSIP*, 2019, pp. 1–5.

[70] O. Musa, P. Jung, and N. Goertz, "Generalized approximate message passing for unlimited sampling of sparse signals," in *Proc. IEEE GlobalSIP*, 2018, pp. 336–340.

[71] T. Olofsson, "Deconvolution and model-based restoration of clipped ultrasonic signals," *IEEE Trans. Instrum. Meas.*, vol. 54, no. 3, pp. 1235–1240, 2005.

[72] A. Adler, V. Emiya, M. G. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley, "A constrained matching pursuit approach to audio declipping," in *Proc. ICASSP*, 2011, pp. 329–332.

[73] J. C. Estrada, M. Servin, and J. Vargas, "2D simultaneous phase unwrapping and filtering: A review and comparison," *Opt. and lasers in Eng.*, vol. 50, no. 8, pp. 1026–1029, 2012, fringe Analysis Methods and Applications. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0143816612000097

[74] S. Chavez, Q.-S. Xiang, and L. An, "Understanding phase maps in MRI: a new cutline phase unwrapping method," *IEEE Trans. Med. Imag.*, vol. 21, no. 8, pp. 966–977, 2002.

[75] R. Gens, "Two-dimensional phase unwrapping for radar interferometry: Developments and new challenges," *Int. J. of Remote Sens.*, vol. 24, no. 4, pp. 703–710, 2003.

[76] E. Romanov and O. Ordentlich, "Blind unwrapping of modulo reduced Gaussian vectors: Recovering MSBs from LSBs," in *Proc. IEEE Int. Symp. Inf. Theory*, 2019, pp. 2329–2333.

[77] C. A. Floudas, *Nonlinear and mixed-integer optimization: fundamentals and applications*. Oxford, UK: Oxford University Press, 1995.

[78] L. Zhang, W. Wei, Y. Zhang, F. Li, C. Shen, and Q. Shi, "Hyperspectral compressive sensing using manifold-structured sparsity prior," in *Proc. IEEE Int. Conf. Comp. Vision*, 2015, pp. 3550–3558.

[79] T. Li and Z. Zhang, "Robust face recognition via block sparse bayesian learning," *Mathematical Problems in Eng.*, vol. 2013, 2013.

[80] Z. Zhang, T.-P. Jung, S. Makeig, and B. D. Rao, "Compressed sensing for energy-efficient wireless telemonitoring of noninvasive fetal ecg via block sparse bayesian learning," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 2, p. 300–309, Feb 2013.

[81] Z. Zhang, T.-P. Jung, S. Makeig, Z. Pi, and B. D. Rao, "Spatiotemporal sparse bayesian learning with applications to compressed sensing of multichannel physiological signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 6, pp. 1186–1197, 2014.