UNIVERSITY OF CALIFORNIA SAN DIEGO

Efficient Techniques for Millimeter Wave Sensing and Beam Alignment, Sparse Recovery, and DoA Estimation

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

in

Electrical Engineering (Communication Theory and Systems)

by

Rohan Ramchandra Pote

Committee in charge:

Professor Bhaskar D. Rao, Chair
Professor Philip E. Gill
Professor Piya Pal
Professor Rayan Saab
Professor Xinyu Zhang

2023

The Dissertation of Rohan Ramchandra Pote is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2023

DEDICATION

To my family
whose constant love and support made this momentous achievement possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my gratitude towards Prof. Bhaskar D. Rao, who trusted in me, and nurtured my research progress. His unparalleled enthusiasm, support, and kindness guided me in gaining skills for research, and beyond. The opportunity he provided for lecturing a graduate level class on DSP will always stay memorable.

I would also like to express my gratitude towards the other committee members - Prof. Piya Pal, Prof. Xinyu Zhang, Prof. Philip E. Gill and Prof. Rayan Saab for agreeing to be part of my doctoral committee. Each member contributed a unique field of expertise, and I am deeply humbled for the opportunity to have my work be presented, and be shaped by their enquiries. Special thanks to Prof. Pal for sharing her profound insights in research, and for her support.

At UCSD, the several courses I took, have helped me to think fundamentally about research. These courses not only taught me about the subject matter, but also cultivated curiosity and the ability to ask research questions in me. I have grown here academically, and UCSD will remain a home. I am also grateful for the professional support provided by the ECE staff, ISPO advisors and the Graduate Division. Geisel and libraries in San Diego (and Naperville) hold a special place for providing much needed change of environment for study and an escape from numerous fire alarm testings at EBU1.

Research would have been half the fun without the awesome labmates (and friends). I was blessed to be sharing the lab with Aditya, Govind, Kuan-lin, Hitesh, Yonghee, Jing, Maher and Markus. Afternoon coffee sessions with Aditya, Govind and Markus were invariably fun and enlightening, regardless of the variance in the quality of coffee at Sunshine. Friends in San Diego stood by me, supported me when I needed them most - Anwesan, Aditya, Nadim and Shahar - its always fun to have you around, even while watching 'Capernaum'. I am forever grateful for that. I would also like to thank Phani Kumar and Pulak for their collaboration during Ph.D.. The course project with Phani marked the beginning of my research with Prof. Rao. Pulak, we perhaps took the just-

in-time philosophy quite literally when submitting our Qualcomm (QIF) proposal merely minutes or even seconds before the deadline. I wish to convey special thanks to Rohit Parasnis, for being there to discuss difficult times. I am also grateful to Heng, Yacong, Sung-En and Jing who shared their wisdom at multiple occasions.

I am grateful for the wonderful internship experiences at Samsung Semiconductor, NOKIA Bell Labs and Sony. My mentors included Dr. Pranav Dayal and Dr. Mahmoud Abdelgelil (Samsung), Dr. Amitava Ghosh (NOKIA) and Dr. Daniel Schneider (Sony). They believed in my ability and ensured I have freedom to innovate and provided me with ample opportunities to present and network.

I also wish to acknowledge all the teachers and friends who knowingly or unknowingly have helped shape this path for me. The list includes Prof. Arun Kumar Singh at IITJ, Prof. Bixio Rimoldi at EPFL, Ashutosh sir, Avinash sir and Suresh Apte sir. They helped me see a future which I would never have believed otherwise.

A constant source of love and support during this journey, and always, has been my family. My parents magically always got things right for me to succeed, for else, my presence here and achieving what I could is unlikely. I am forever grateful and this dissertation is dedicated to them. I miss you Papa.

Another person who supported me tremendously in this journey is my wife. Sukanya has endured my mood swings (which correlated well with the results I got), and helped critique my work, only to improve it by huge margins. She made sacrifices to join me in the US, supported me in defeats, and endured countless hours of boredom while I refined my drafts. This dissertation is dedicated to you and our future together.

Chapter 1, in part is currently submitted for publication of the material, and in part, is a reprint of the material as it appears in R. R. Pote and B. D. Rao, "Novel Sensing Methodology for Initial Alignment Using MmWave Phased Arrays," 2023 57th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, 2023. The dissertation author was the primary investigator and author of this material.

Chapter 2, in part is currently being prepared for submission for publication of the material, and in part, is a reprint of the material as it appears in R. R. Pote and B. D. Rao, "Light-Weight Sequential SBL Algorithm: An Alternative to OMP," ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 2023, pp. 1-5. The dissertation author was the primary investigator and author of this material.

Chapter 3, in part, is a reprint of the material as it appears in R. R. Pote and B. D. Rao, "Maximum Likelihood-Based Gridless DoA Estimation Using Structured Covariance Matrix Recovery and SBL With Grid Refinement," in IEEE Transactions on Signal Processing, vol. 71, pp. 802-815, 2023, R. R. Pote and B. D. Rao, "Maximum Likelihood Structured Covariance Matrix Estimation and connections to SBL: A Path to Gridless DoA Estimation," 2022 56th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, 2022, pp. 970-974. The dissertation author was the primary investigator and author of this paper.

VITA

2014	Bachelor of Technology, Indian Institute of Technology Jodhpur
2017	Master of Science, École Polytechnique Fédérale de Lausanne (EPFL)
2023	Doctor of Philosophy, University of California San Diego

PUBLICATIONS

R. R. Pote and B. D. Rao, "Novel Active Sensing and Inference for MmWave Beam Alignment Using Single RF Chain Systems," Submitted.

R. R. Pote and B. D. Rao, "Novel Sensing Methodology for Initial Alignment Using MmWave Phased Arrays," 2023 57th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, 2023.

R. R. Pote and B. D. Rao, "Light-Weight Sequential SBL Algorithm: An Alternative to OMP," ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 2023, pp. 1-5.

R. R. Pote and B. D. Rao, "Maximum Likelihood-Based Gridless DoA Estimation Using Structured Covariance Matrix Recovery and SBL With Grid Refinement," in IEEE Transactions on Signal Processing, vol. 71, pp. 802-815, 2023.

R. R. Pote and B. D. Rao, "Maximum Likelihood Structured Covariance Matrix Estimation and connections to SBL: A Path to Gridless DoA Estimation," 2022 56th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, 2022, pp. 970-974.

R. R. Pote and B. D. Rao, "A Novel Bayesian Approach for the Two-Dimensional Harmonic Retrieval Problem," ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 2021, pp. 4525-4529.

R. R. Pote and B. D. Rao, "Robustness of Sparse Bayesian Learning in Correlated Environments," ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 2020, pp. 9100-9104.

R. R. Pote and B. D. Rao, "Reduced Dimension Beamspace Design Incorporating Nested Array For Mmwave Channel Estimation," 2019 53rd Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, 2019, pp. 1212-1216.

ABSTRACT OF THE DISSERTATION

Efficient Techniques for Millimeter Wave Sensing and Beam Alignment, Sparse Recovery, and DoA Estimation

by

Rohan Ramchandra Pote

Doctor of Philosophy in Electrical Engineering (Communication Theory and Systems)

University of California San Diego, 2023

Professor Bhaskar D. Rao, Chair

Parametric and non-parametric measurement models provide means to more insights when the amount of available data is limited. We focus on measurements from multi-sensor systems in three applications of practical interest and provide inference techniques to estimate underlying unknowns with higher accuracy than previous approaches under limited resources. In this pursuit we adopt the maximum likelihood estimation (MLE) framework to estimate the parameters from the measurements. The framework is widely used with varying degrees of success. In certain applications, framing problems under MLE leads to difficult non-convex optimization problems, and thus model selection

becomes crucial. Other applications may even require one to design measurements upon which the MLE framework may be applied.

In the first problem, we consider the initial beam alignment in millimeter wave systems using phased arrays. For a single RF chain system, we propose a novel sensing methodology inspired from synthetic aperture radar, that enables more informative measurements in a structured manner. We also provide an inference technique that utilizes the measurements under proposed sensing to efficiently compute a posterior density on the unknown angle. The inference is carried without the knowledge of complex path gain, and demonstrates significant improvement over competing techniques.

In the second problem, we study sparse signal recovery with the aim to bridge the computational gap between the widely used Orthogonal Matching Pursuit algorithm and methods derived from the MLE objective. We propose a novel Light-Weight Sequential Sparse Bayesian Learning (LWS-SBL) algorithm and provide efficient recursive procedures to update the internal variables of the algorithm. We demonstrate superior support recovery performance using LWS-SBL over OMP and further elucidate the subtle differences in the underlying mechanisms in the two algorithms.

Lastly, we delve into the Direction-of-Arrival (DoA) estimation problem for narrowband signals and propose a novel two-step algorithm. In the first step, we recover a structured covariance matrix estimate for the received signal in the MLE sense. The second step involves estimating DoAs using root-MUSIC from the recovered structured covariance matrix. The first step draws inspiration from the SBL formulation, as it provides the basic model which we fit to the measurements. The proposed approach improves resolution, bias, identifiability, and can identify two or more sources with a single snapshot, unlike the traditional subspace-based algorithms.

Introduction

Many applications involve taking measurements using more than one sensor with the hope to gain better insights about the environment or an object under examination. The inherent geometry of the sensor placement and the characteristics of the sources under study dictates the efficacy of combining measurements from these multiple sensors. Such systems are ubiquitous such as biomagnetic imaging, underwater acoustics, echo cancellations, and wireless systems. In this dissertation, we focus on three such applications and develop methods that exploit available prior information about the sources.

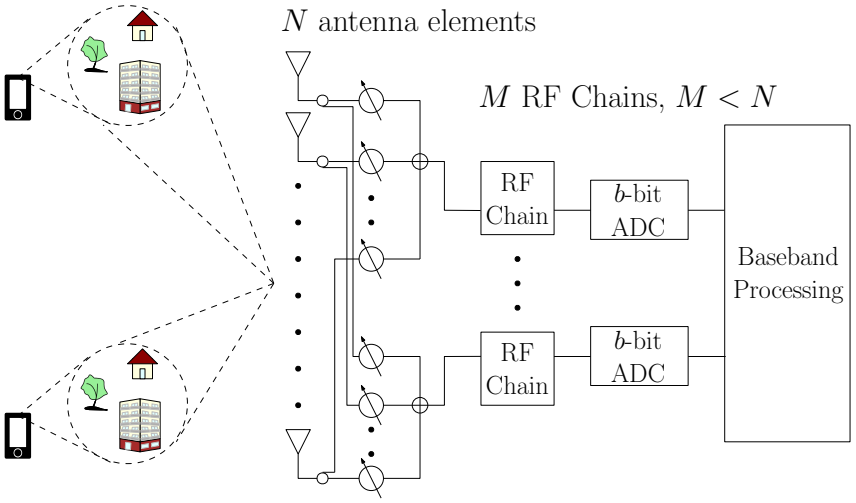


Figure 1. Millimeter wave receiver with fewer RF chains than antenna elements.

Millimeter wave (mmWave) technology is key to satisyfing the ever increasing wireless data demand, wherein numerous multimedia services require very high throughput. The large spectrum available in the millimeter wave band can solve many of these issues [1, 2] and has been a subject of active research and standardization efforts (3GPP).

Channel modelling efforts have revealed some inherent attributes of the mmWave channel that makes it unique [3]. Unlike the traditional cellular or Wi-Fi bands, mmWave channel suffers from large propagation losses and incurs poor scattering. The latter issue lends the mmWave channel to be sparse in nature which can be characterized in terms of angle of departure, angle of arrival, and complex path gain of the multiple paths. As a result, array signal processing techniques for Direction of Arrival (DoA) estimation become relevant for the mmWave channel estimation problem. Yet, there are unique challenges while estimating the multipath DoAs. Unlike, the traditional setting in the array signal processing problem where the signal from the multiple antenna elements is available for inference in a digital form, since the mmWave systems typically have large number of antenna elements, the signal at antennas is combined in analog domain first (see Fig. 1). Depending on the number of radio-frequency (RF) chains available for A/D conversion, only a low-dimensional projection of the high-dimensional antenna signal is available for inference. Thus, a simple plug-and-play approach is not applicable. In Chapter 1, we consider the beam alignment problem in millimeter wave systems using phased arrays (i.e., single RF chain systems). These systems are widely considered due to their low hardware cost and battery power savings.

We focus on two fundamental tools in the remaining chapters, that are applicable to mmWave systems, but can be applied more broadly. In Chapter 2, we consider the Sparse Signal Recovery (SSR) problem. SSR is an important problem that has witnessed numerous applications (e.g., biomagnetic imaging [4], functional approximation [5], and echo cancellation [6]). Consequently, many algorithms have been proposed that offer favourable tradeoffs between recovery performance, speed, and storage [7–14]. On the computationally favourable paradigm, greedy algorithms (e.g. pursuit algorithms like matching pursuit (MP) [7], orthogonal MP (OMP) [15–18], compressive sampling MP (CoSaMP) [19] etc.) are well-studied algorithms that offer faster recovery at the cost of slight degradation in performance. OMP is a widely employed [20, 21] iterative technique

in this category that models the unknown sparse vector as a deterministic variable, and improves over MP by ensuring that the residual error is orthogonal to already selected columns of the measurement matrix. In our work, we offer a stochastic alternative to the OMP algorithm with a similar complexity while improving support recovery performance. The proposed approach is based on maximum likelihood estimation (MLE) framework, unlike OMP which is based on matching-criteria[1].

In Chapter 3, we revisit the classical DoA estimation problem with an aim to overcome the model complexities faced by MLE methods of the past. It is known that parametric methods like MLE allow one to introduce meaningful parameters as a means to incorporate information about antenna array geometry and prior. These parameters can be estimated even with a single snapshot, unlike subspace-based methods. Other recent techniques include [8, 10, 11, 22–25], which utilize sparsity as an explicit regularizer. Such methods are thus sensitive to setting the regularization parameter appropriately. SBL [12,26,27] formulates the sparse recovery problem on-grid, under the MLE framework, and recovers sparse solutions via implicit regularization [28]. We combine the classical MLE framework and the modern SBL formulation to estimate the DoAs in a gridless manner.

We defer further discussion to the respective chapters, and briefly explore the common theme of the dissertation in the remainder of this Chapter.

0.1 Measurement Model

We consider parametric as well as non-parametric measurement models in this dissertation. Such models exploit information about the underlying physics of the experiment, and thus can do away with fewer measurements than typically required for

[1]Note that OMP is MLE when the support size to be recovered is one.

completely data-driven methods. One drawback of driving inference based on models is that if the model assumptions do not hold true for the measurements, then the inference and insights carried out under the model may be less useful. In this dissertation, we assume that the models carry relevant information about the problem at-hand. We also briefly discuss a case when the model assumptions do not hold perfectly, and study the impact of this mismatch on the inference.

0.1.1 Parametric Measurement Model

Consider the following measurement model

$$\mathbf{y}_l = \mathbf{\Phi}_{\boldsymbol{\theta}} \mathbf{x}_l + \mathbf{n}_l, \qquad 0 \leq l < L, \tag{1}$$

where $\mathbf{y}_l \in \mathbb{C}^M$ denotes the measurements, and L denotes the total number of snapshots available. The kth column of $\mathbf{\Phi}_{\boldsymbol{\theta}} \in \mathbb{C}^{M \times K}$ is a vector function of the parameter θ_k i.e., $[\mathbf{\Phi}_{\boldsymbol{\theta}}]_k = \boldsymbol{\phi}(\theta_k)$ for some known $\boldsymbol{\phi}(.), k \in \{1, \ldots, K\}$. $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_K]^T$ and θ_k's lie in some known continuous domain. K denotes the number of sources. The sources' signal $\mathbf{x}_l \in \mathbb{C}^K$ and the zero mean noise $\mathbf{n}_l \in \mathbb{C}^M$ are independent of each other, and i.i.d. over time. The noise, $\mathbf{n}_l$, is distributed as $\mathscr{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I})$. In (1), the parameters $(\boldsymbol{\theta}, \mathbf{x}_l, \sigma_n^2)$ are the unknowns. The model parameters affect the measurements in a non-linear manner, which makes the inverse problem extremely difficult to solve, even in the absence of noise.

0.1.2 Non-Parametric Measurement Model

Given a measurement matrix $\mathbf{\Phi} \in \mathbb{C}^{M \times n} (M < n)$ and measurement vector $\mathbf{y} \in \mathbb{C}^M$ such that

$$\mathbf{y} = \mathbf{\Phi} \mathbf{x} + \mathbf{n}, \tag{2}$$

may be corrupted by noise vector $\mathbf{n} \in \mathbb{C}^M$, the goal is to recover the vector $\mathbf{x} \in \mathbb{C}^n$ which is known to be sparse i.e., $\|\mathbf{x}\|_0 \ll n$. Similar assumptions on pdf of noise, as made in

the parametric model case are considered here. Note that without sparsity the problem is underdetermined and can have infinite number of solutions. The field of study of the problem in (2) is known as Sparse Signal Recovery (SSR). The parametric problem in (1) can be re-modeled as a SSR problem under suitable conditions (more on this is described in later chapters), and thus SSR is an attractive modeling approach that is applicable quite generally.

The problem in (2) can be extended to account for multiple measurements over time as

$$\mathbf{y}_l = \mathbf{\Phi}\mathbf{x}_l + \mathbf{n}_l, \qquad 0 \leq l < L, \tag{3}$$

where it is known that $\mathbf{X} = [\mathbf{x}_0, \ldots, \mathbf{x}_{L-1}]$ is row-sparse i.e., most of the rows are zero. The problem in (3) is known as the multiple measurement vector (MMV) problem when $L > 1$ [29], compared to the single measurement vector or SMV problem when just a single snapshot is available i.e., $L = 1$. The non-zero rows correspond to active sources, and one of the key problems in SSR is to identify these non-zero rows.

Next we describe the Hierarchical Bayesian Framework and the Type-II estimation procedure that is widely employed in this dissertation. We focus on the non-parametric model for the remainder of this Chapter.

0.2 Hierarchical Bayesian Framework and Type-II Estimation

We model the unknown $\mathbf{x}$ as a stochastic variable, and invoke a parameterized prior on $\mathbf{x}$. We learn the parameters of the prior (also called as hyperparameters) along with an estimate for $\mathbf{x}$. One implication of this approach is that we estimate a posterior density on the unknown, $\mathbf{x}$, instead of just a point-estimate. We now describe the choice of the parameterized prior and introduce the hyperparameters.

We impose a Gaussian paramaterized prior on $\mathbf{x}$ with mean zero and uncorrelated

components. In other words,

$$\mathbf{x} \sim \mathscr{CN}(\mathbf{0}, \mathbf{\Gamma}), \tag{4}$$

where $\mathbf{\Gamma}$ is a diagonal matrix, $\mathbf{\Gamma} \in \mathbb{R}^{n \times n}$ and $\mathbf{\Gamma} \succeq \mathbf{0}$; let $\mathbf{diag}(\mathbf{\Gamma}) = \boldsymbol{\gamma}$. Since the noise is Gaussian distributed with pdf as discussed in previous subsections, we have that, $\mathbf{y} \sim \mathscr{CN}(\mathbf{0}, \mathbf{\Phi\Gamma\Phi}^H + \lambda\mathbf{I})$, where λ denotes the estimate for the noise variance. In Fig. 2 we plot

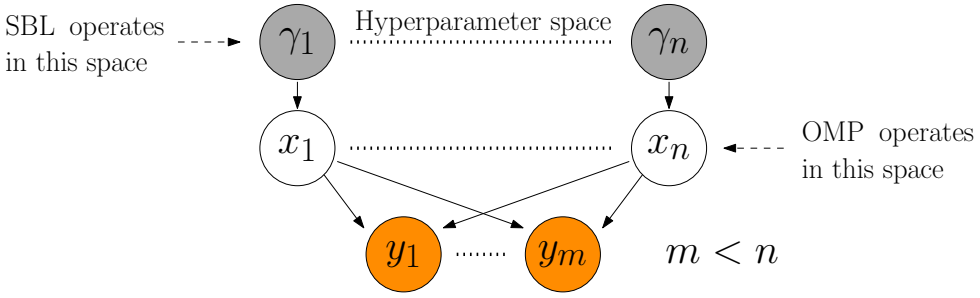


Figure 2. Hierarchical Bayesian Framework (Credits: Adapted from ECE 285 by Prof. B. D. Rao)

the probabilistic graphical model representing the relationship between the measurements and the different unknown variables.

0.2.1 Type-II Estimation & Sparse Bayesian Learning (SBL)

The approach adopted in this work is to solve (2) in the hyperparameter $\boldsymbol{\gamma}$-space, post marginalization with respect to $\mathbf{x}$. This approach is known as Type-II estimation framework, compared to Type-I estimation where the problem is solved in $\mathbf{x}$ space after marginalization with respect to $\boldsymbol{\gamma}$ [30] (e.g., Lasso algorithm [31]). The choice of the parameterized Gaussian prior, together with the Type-II estimation procedure for the problem in (2) falls under the rubric-Sparse Bayesian Learning (SBL). The hyperparameter $\boldsymbol{\gamma}$ is estimated by minimizing the negative marginal log-likelihood function [12, 26]

$$\min_{\boldsymbol{\gamma} \geq \mathbf{0}, \lambda \geq 0} \mathscr{L}(\boldsymbol{\gamma}, \lambda) := \log\det\mathbf{\Sigma}_\mathbf{y} + \mathbf{y}^H \mathbf{\Sigma}_\mathbf{y}^{-1} \mathbf{y}. \tag{5}$$

where, $\mathbf{\Sigma_y} = \mathbf{\Phi \Gamma \Phi}^H + \lambda \mathbf{I}$. (5) is a non-convex problem in $(\boldsymbol{\gamma}, \lambda)$. The cost function consists of two terms: the first term, $\log\det \mathbf{\Sigma_y}$, promotes sparsity [28], whereas the second term, $\mathbf{y}^H \mathbf{\Sigma_y}^{-1} \mathbf{y}$ fits the model, $\mathbf{\Sigma_y}$, to the measurements, $\mathbf{y}$. Sparsity in $\mathbf{x}$ is achieved as a consequence of sparsity in recovered $\boldsymbol{\gamma}$. More details about the SBL formulation will be discussed in the later chapters.

0.3 Enforcing Model Order When Known

The formulation in (5) utilized the array geometry, (uncorrelated) source correlation prior, and the noise distribution. In some applications, the number of sources to be identified, K, may be known a priori. Such information can be incorporated in (5) as

$$\min_{\|\boldsymbol{\gamma}\|_0 = K, \boldsymbol{\gamma} \geq \mathbf{0}, \lambda \geq 0} \mathscr{L}(\boldsymbol{\gamma}, \lambda) := \log\det \mathbf{\Sigma_y} + \mathbf{y}^H \mathbf{\Sigma_y}^{-1} \mathbf{y}. \tag{6}$$

The above formulation is combinatorial in nature, and thus not practical, in general. The case when $K = 1$ is special, as the above optimization can be further simplified. This will be elaborated in Chapter 1, where we discuss the millimeter wave beam alignment problem, and assume line-of-sight (LoS). The general problem of $K \geq 1$ is dealt with in Chapter 2 and Chapter 3. In Chapter 2, (6) is solved in a greedy manner, and the solution is compared with the widely used SSR technique known as Orthogonal Matching Pursuit (OMP). Whereas in Chapter 3, the model order information is incorporated in the second step of the proposed two-step algorithm for the DoA estimation problem.

0.4 Outline and Contributions of the Dissertation

We briefly outline and summarize the contributions discussed in this dissertation.

The first contribution is in sensing and beam alignment problem in millimeter wave systems using phased arrays. Conventionally, sensing using a single Radio Frequency (RF) chain involves taking multiple power-based measurements, each measurement focusing in

different spatial regions. We propose a novel sensing methodology, named as Synthesis of Virtual Array Manifold (SVAM), inspired from synthetic aperture radar systems, that realizes a virtual array geometry over temporal measurements. Thus, with just two snapshots, it is possible to estimate the dominant path angle-of-arrival (AoA), along with the complex path gain. We demonstrate the benefits of SVAM using Cramér-Rao bound (CRB) analysis over schemes that repeat beam pattern to boost signal-to-noise (SNR) ratio. We also showcase versatile applicability of the proposed SVAM sensing by incorporating it within beam alignment procedures that assume perfect channel state information (CSI). We further consider the practical scenario wherein we estimate the CSI and propose a novel beam alignment procedure based on efficient computation of posterior density on dominant path AoA. The performance of the proposed sensing and beam alignment is empirically observed to approach the perfect CSI performance, even at low SNR. We also provide numerical experiments to study the impact of parameters involved in estimating the posterior. Finally, we briefly highlight other systems and scenarios in which proposed SVAM sensing can be employed. This is fully discussed in Chapter 1.

The second contribution is the development of a low complexity algorithm for the Sparse Signal Recovery (SSR) problem based on the sparse Bayesian learning (SBL) formulation. The proposed algorithm, called as Light-Weight Sequential SBL (LWS-SBL), is compared to the widely used iterative and greedy algorithm known as Orthogonal Matching Pursuit (OMP). In contrast to OMP, which models the unknown sparse vector as a deterministic variable, the same is modeled as a stochastic variable within LWS-SBL. Specifically, the proposed algorithm is derived from the stochastic maximum likelihood estimation framework, and it iteratively selects columns that maximally increase the likelihood. We derive efficient recursive procedure to update the internal parameters of the algorithm, and maintain a similar asymptotic computational complexity as OMP. Additional perspectives to understand the underlying differences in the mechanisms of the two algorithms are provided, that reveal avenues where LWS-SBL improves over OMP. These

are verified thoroughly in the numerical section in terms of improved support recovery performance. For SSR problems involving parametric dictionaries, the flexibility of the proposed approach is demonstrated by extending LWS-SBL to recover multi-dimensional parameters, and in a gridless manner. This is fully discussed in Chapter 2.

The third contribution is the development of a novel maximum likelihood-based gridless DoA estimation algorithm. We focus on the stochastic maximum likelihood estimation (MLE) framework and overcome the model complexities of the past by reparameterization of the objective and exploiting the sparse Bayesian learning (SBL) approach. SBL is shown to be a correlation-aware method and, for the underlying problem, a grid-based technique for recovering a structured covariance matrix of the measurements. For the case when measurements are spatial (or temporal) samples at regular intervals, the structured matrix is expressible as a sampled Toeplitz matrix. In this case, additional constraints and reparameterization of the SBL objective leads to the proposed structured matrix recovery technique based on MLE. The optimization problem is non-convex and a majorization-minimization based iterative procedure is proposed to estimate the structured matrix; each iteration solves a semidefinite program. We recover the parameter of interest in a gridless manner by appealing to the Carathéodory-Féjer result on decomposition of positive semidefinite Toeplitz matrices. For the general case of irregularly spaced samples, we propose an iterative SBL procedure that refines grid points to increase resolution near potential source locations, while maintaining a low per iteration complexity. We provide numerical results to compare the performance of the proposed techniques with other gridless techniques, and the Cramér-Rao bound. The proposed correlation-aware approach is more robust to issues such as fewer snapshots, correlated or closely separated sources, and improves sources identifiability. This is fully discussed in Chapter 3.

Chapter 1

Active Sensing & Beam Alignment Using MmWave Single RF Chain Systems

1.1 Introduction

Millimeter wave (mmWave) technology is essential for expanding the existing capabilities of cellular networks [1, 2]. The availability of the large spectrum in the 30-300 GHz spectrum range, and the ability to place many more antennas in the same form factor on a device are very promising avenues for the next generation wireless systems. This has propelled the interest, both in the industry and academia. The envisioned benefits include much higher throughput and low latency. The impact of this technology can be gauged from the numerous use cases enabled by the mmWave technology, which includes industrial-IoT, virtual/augmented reality, biomedical applications, and non-terrestrial networks [32, 33].

MmWave technology also faces many challenges. The mmWave channel incurs large propagation losses thereby restricting coverage per base station (BS), and requiring additional infrastructure compared to legacy cellular networks. A second challenge is the specular nature of mmWave channel rendering it to be sparse and requiring accurate beam alignment. This challenge is only further exacerbated by the narrow beamwidths and consequent large codebook size due to the large antenna array dimensions. Several options are being considered for enhancing coverage at low-cost such as integrated access backhaul

and intelligent reflective surfaces. On the other hand, reducing the beam alignment phase duration is a critical and active area of research. Hardware cost also impacts the ability of the transceivers to sense the mmWave channel. The large number of antenna elements are typically supported by only a few Radio Frequency (RF) chains [34], and thus necessitates for a low-dimensional projection of the received signal at the antennas. Beam alignment using such a low-dimensional signal is a challenging problem, and it is the main focus of this work.

There is a need to build a better sensing approach coupled with efficient inference mechanisms that exploit the array geometry and channel characteristics under the hardware constraints. An interesting direction adopted in [35] formulates the beam alignment problem under the posterior matching framework, and actively learns the single path Direction of Arrival (DoA). The work is shown to improve over the detection-based algorithm in [36]. However, the authors in [35] assume that the small-scale fading coefficient is perfectly known. Subsequent effort build upon this work, and estimate both the DoA as well as the fading coefficient [37]. The Kalman filter-based posterior matching algorithm in [37] still requires good prior density on the small-scale fading coefficient. Similar assumptions on availability of good prior density was made in the variational hierarchical posterior matching algorithm proposed in [38].

In this work, we reflect on the beam alignment problem from the perspective of active sensing for improved estimation performance. We do so without relying on additional information such as good prior knowledge about the small-scale fading coefficient. The contributions of this work are as follows:

- A novel sensing methodology, inspired from Synthetic Aperture Radar (SAR), is proposed for the single RF chain mmWave systems. Under the proposed sensing approach, a virtual Uniform Linear Array (ULA) manifold is synthesized over temporal measurements. Extension to construct a virtual arbitrary array geometry such as Sparse Linear

Array (SLA) is also discussed.

- Benefits of the proposed sensing are described, when the channel small-scale fading coefficient is known, in terms of i. its impact on the Cramér-Rao lower Bound (CRB) on the variance of estimation of the unknown dominant path angle, when compared to a benchmark scheme, and ii. its ability to be incorporated within existing active beam alignment procedures. The improvement in terms of lower training overhead is demonstrated using numerical experiments.

- A novel beam alignment procedure is proposed which adapts the beamformer based on the current estimate of the posterior on the unknown angle. The proposed algorithm estimates a posterior on the small-scale fading along with the angular posterior. Both, flexible and hierarchical codebook-based beam alignment procedures are presented.

- Finally, the proposed sensing and beam alignment procedures are empirically studied, and compared with the performance using perfect knowledge of the channel state information. Impact of the parameters involved in the estimation procedure is also studied, which also reveal the ability of the adaptive beam alignment procedure to self-correct in case of premature misalignment during the early phase of the training period.

1.1.1 Relevant Prior Work

A virtual array synthesis from spatial measurements under the reduced number of RF chains constraint for mmWave systems was proposed in [39]. A similar sensing scheme as in [39], was proposed in [40] for mmWave multipath angle estimation and in [41] for the DoA estimation problem. In [40], the authors proposed using random precoders and combiners that are submatrices of banded Toeplitz matrices. The work focuses on mmWave systems with multiple RF chains, and the ideas are extended to the single RF chain case. In [41], the authors investigated the applicability of root-MUSIC and ESPRIT algorithms, as a consequence of preserving the Vandermonde structure and the

shift-invariance under the virtual array synthesis procedure. These sensing methodologies can be traced back to Silverstein [42] and Tkacenko [43]. The sensing scheme proposed in this work synthesizes a virtual array over temporal measurements, and considers the practical single phased array system. To the best of our knowledge, the presented adaptive sensing methodology is the first of its kind.

Many non-adaptive and adaptive beamforming approaches have been proposed for the mmWave beam alignment problem in the past. Random beamforming was proposed in [44–46], wherein the inference was carried out using compressed sensing algorithms. This approach does not exploit beamforming gain needed to combat large path loss in the mmWave channels. An exhaustive beam steering approach proposed in IEEE 802.11ad and 5G standards improves the beamforming gain, but can be slow in selecting the appropriate beam. The acquisition time is improved in literature by replacing the non-adaptive linear search with an adaptive binary (or in general n-ary, $n \geq 2$) search within a hierarchical codebook [36, 47–49]. A comparison in terms of asymptotic misalignment probability between the exhaustive search and the hierarchical search was studied in [50]. For adaptive schemes using hierarchical codebook, the inference at each hierarchical level is typically carried out by comparing power at the output of different beams within a hierarchical node. The inference mechanism was improved in [35] by computing posterior on the dominant path angle and selecting next beam based on the posterior within the hierarchical codebook of [36]. The work assumes that the small-scale fading coefficient is perfectly known. However, such assumption is difficult to satisfy in practice. The problem of estimating small-scale fading coefficient (along with the unknown path angle) was considered in later works [37, 38]. A grid-approach was also proposed in [37], but the relevant issue on how to choose appropriate grid on the small-scale fading coefficient was not addressed. An adaptive beam search algorithm scheme based on posterior computation to compare beams was proposed in [51]. Many learning-based approaches have been proposed in literature as well [52–54]. These include approaches that frame the beam alignment prob-

lem as multi-armed bandit problem, or train an end-to-end neural network to design a model-free or codebook-free architecture. In this work, we emphasize the model and propose novel sensing (for improved acquisition) and inference procedures to estimate the unknown parameters. The paper builds on our previous work in [55].

1.1.2 Organization of the Chapter and Notations

We describe the problem tackled in this work, and introduce the new sensing approach in Section 1.2. In Section 1.3, we investigate the impact of the proposed sensing for the case when the fading coefficient is perfectly known, and only the unknown DoA is to be recovered. The more practical case, when the fading coefficient needs to be estimated along with the DoA is discussed in Section 1.4. We provide empirical results in Section 1.5, and present our conclusions in Section 1.6.

Notations: We represent scalars, vectors, and matrices by lowercase, boldface-lowercase, and boldface-uppercase letters, respectively. Sets are represented using blackboard bold letters. $(.)^T, (.)^H, (.)^c$ denotes transpose, Hermitian, and complex conjugate operation respectively. $\otimes$ denotes matrix Kronecker product, and $\odot$ denotes Hadamard product of two conformable matrices. $*$ denotes convolution operation. $[M] = \{0, 1, \ldots, M-1\}, M \in \mathbb{Z}^+$.

1.2 Problem Statement & Proposed Novel Sensing

We consider a receiver (base station or user equipment) equipped with a Uniform Linear antenna Array (ULA) of size N and a single RF chain. We assume a flat fading channel, with a single dominant path between the transmitter and receiver. We further assume that the channel remains coherent within the training duration due to low receiver mobility.

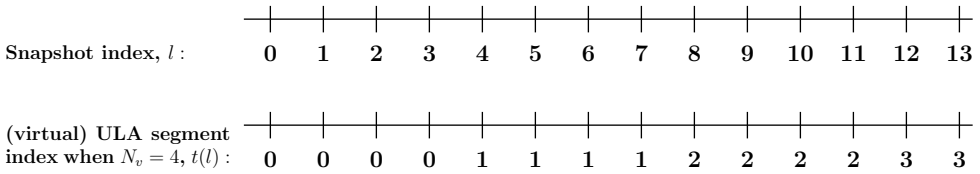


Figure 1.1. A (virtual) ULA segment of size $N_v = 4$ is created using 4 snapshots. The beamformer, $\mathbf{f}_{t(l)}$, is adapted once in every segment duration.

1.2.1 Problem Statement

The received signal at the antennas at instant l, $\mathbf{x}_l \in \mathbb{C}^N$, is given by

$$\mathbf{x}_l = \sqrt{P_s}\alpha\boldsymbol{\phi}_N(u) + \bar{\mathbf{n}}_l, \quad l \in [L], \tag{1.1}$$

where L denotes the total training duration. $P_s \geq 0$ denotes the combined contribution of transmitted power and the large-scale fading (path loss and shadowing), $\alpha \in \mathbb{C}$ is the unknown small-scale fading coefficient. Since the transmitted symbol is known, it can be easily absorbed within the signal term of $\mathbf{x}_l$ in (1.1). Thus, we assume the transmitted symbol value to be 1 without loss of generality. $\boldsymbol{\phi}_N(u)$ is the array manifold or response vector for an incoming narrowband signal along the angle u; $u = \sin\theta, u \in [-1, 1)$, where $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2})$ denotes physical DoA. Noise, $\bar{\mathbf{n}}_l \in \mathbb{C}^N$, is distributed as $\mathscr{CN}(\mathbf{0}, \sigma_n^2\mathbf{I})$ and i.i.d. over time. Since there is a single RF chain, the received signal is processed using an analog combiner, $\mathbf{w}_l \in \mathbb{C}^N$. The output, $y_l \in \mathbb{C}$, available for inference is given by

$$y_l = \mathbf{w}_l^H\mathbf{x}_l = \sqrt{P_s}\alpha\mathbf{w}_l^H\boldsymbol{\phi}_N(u) + \mathbf{w}_l^H\bar{\mathbf{n}}_l = \sqrt{P_s}\alpha\mathbf{w}_l^H\boldsymbol{\phi}_N(u) + n_l. \tag{1.2}$$

The goal is to design $\mathbf{w}_l$ and infer u; $\mathbf{w}_l$ can be adapted over time to improve the inference. For a ULA with $\lambda/2$ inter-element spacing[1], where λ denotes the wavelength of the

[1] $\lambda/2$ inter-element spacing prevents ambiguity in angular estimation.

received signal, we have

$$\boldsymbol{\phi}_N(u) = [1 \exp(j\pi u) \cdots \exp(j\pi(N-1)u)]^T. \tag{1.3}$$

In this paper, we consider the ULA geometry for easier exposition of ideas. However, these ideas can be extended to planar geometries as well, such as uniform rectangular arrays [56]. Next, we describe a high-level structure that we impose when designing the beamformer, $\mathbf{w}_l$.

1.2.2 Synthesis of Virtual Array Manifold (SVAM) Sensing

The sensing methodology is inspired from SAR systems used in remote sensing and automotive radar [57]. In typical SAR systems, the sensor motion allows synthesis of larger aperture than physical antenna, which helps to improve resolution. In this work, we mimic the sensor motion by designing $\mathbf{w}_l$ appropriately. We exploit the coherence interval to synthesize virtual apertures over time. An important consequence is that, such measurements preserve phase information from the physical antenna, which captures rich information about the DoA of the incoming signal. The proposed sensing can be applied more broadly to multi-path angles. Moreover, leveraging the complex exponential structure present at the combiner output $y_l, l \in [L]$, it is possible to apply an unlimited number of digital filters on these measurements.

Constructing a Virtual ULA with $\lambda/2$ Inter-element Spacing: Let N_v denote the aperture size of the virtual ULA we wish to create. We assume that the total training duration, L, is divisible by N_v for simplicity. Let $t(l) = \mathbf{floor}(l/N_v)$ denote[2] the (ULA) segment index (see Fig. 1.1). We design a beamformer of size $M = N - N_v + 1$, $\mathbf{f}_{t(l)} \in \mathbb{C}^M$, such that $\|\mathbf{f}_{t(l)}\|_2 = 1$, initially spanning the region of interest (RoI). As the system gathers information about the unknown DoA, u, the design of the beamformer, $\mathbf{f}_{t(l)}$, is adapted.

[2] $\mathbf{floor}(\cdot)$ denotes the floor function.

The analog combiner at instant l is given by

$$\mathbf{w}_l = \begin{bmatrix} \mathbf{0}^T_{\text{mod}(l,N_v)} & \mathbf{f}^T_{t(l)} & \mathbf{0}^T_{N_v-\text{mod}(l,N_v)-1} \end{bmatrix}^T. \tag{1.4}$$

Thus, within a segment duration, the beamformer slides along the antenna aperture. In contrast to the work in [39], here only a single RF chain is available and thus a virtual ULA segment is synthesized over time. Let

$$\beta_{t(l)}(u) = \mathbf{f}^H_{t(l)} \boldsymbol{\phi}_M(u), \tag{1.5}$$

denote the complex gain of the beamformer along the angle u. The signal y_l post-combining can be expressed as

$$y_l = \mathbf{w}_l^H \mathbf{x}_l = \sqrt{P_s}\alpha\beta_{t(l)}(u) \cdot \exp\left(j\pi u \text{mod}(l,N_v)\right) + n_l. \tag{1.6}$$

Note that the gain $\beta_{t(l)}(u)$ does not change within a segment, but '$\exp\left(j\pi u \text{mod}(l,N_v)\right)$' varies within the segment.

Remark 1. The beamforming gain, measured in terms of $|\beta_{t(l)}(u)|^2$, in the passband depends primarily on the beamwidth of the beamformer, $\mathbf{f}_{t(l)}$. For an ideal beamformer design, the gain in the beamformer passband corresponding to a beamwidth of $\frac{2}{R}, R \geq 1$, in u-space is given by $|\beta_{t(l)}(u)|^2 = R$. As the beamformer size, M, increases the beamformer response approaches the ideal response.

We drop the notation for dependence of t on l for simplicity. We stack the measurements within a segment duration to form $\mathbf{y}_t = \begin{bmatrix} y_{tN_v} & y_{tN_v+1} & \cdots & y_{(t+1)N_v-1} \end{bmatrix}^T \in \mathbb{C}^{N_v}, t \in [L/N_v]$,

$$\begin{aligned} \mathbf{y}_t &= \sqrt{P_s}\alpha\beta_t(u)\left[1 \ \exp\left(j\pi u\right) \cdots \exp\left(j\pi(N_v-1)u\right)\right]^T + \mathbf{n}_t \\ &= \sqrt{P_s}\alpha\beta_t(u)\boldsymbol{\phi}_{N_v}(u) + \mathbf{n}_t. \end{aligned} \tag{1.7}$$

We identify the following design parameters: a) $N_v \in \{1,2,\ldots,N\}$, the virtual ULA size, and b) beamformer, $\mathbf{f}_t$, design, which includes the beam direction and beamwidth. $N_v = 1$ reduces to the conventional beam design. Thus, the proposed sensing strategy includes the methodology adopted for sensing in [35,36] as a special case. The hierarchical codebook in [36] designed using a least squared error criterion imposes a constant amplitude and phase in the passband. The inference is improved by relaxing the constant phase requirement in the passband. Thus, in this work we design the beamformers as linear-phase Finite Impulse Response (FIR) filter using the Parks-McClellan algorithm [58].

Let $\tilde{\mathbf{y}}_t = [\mathbf{y}_0^T,\ldots,\mathbf{y}_t^T]^T \in \mathbb{C}^{(t+1)N_v}$ denote the measurements until snapshot index, $l = (t+1)N_v - 1$. Then

$$\begin{aligned}\tilde{\mathbf{y}}_t = \begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_t \end{bmatrix} &= \sqrt{P_s}\alpha \begin{bmatrix} \beta_0(u)\boldsymbol{\phi}_{N_v}(u) \\ \beta_1(u)\boldsymbol{\phi}_{N_v}(u) \\ \vdots \\ \beta_t(u)\boldsymbol{\phi}_{N_v}(u) \end{bmatrix} + \begin{bmatrix} \mathbf{n}_0 \\ \mathbf{n}_1 \\ \vdots \\ \mathbf{n}_t \end{bmatrix} \\ &= \sqrt{P_s}\alpha \left(\tilde{\boldsymbol{\beta}}_t(u) \otimes \boldsymbol{\phi}_{N_v}(u)\right) + \tilde{\mathbf{n}}_t, \end{aligned} \tag{1.8}$$

where $\tilde{\boldsymbol{\beta}}_t(u) = [\beta_0(u),\beta_1(u),\ldots,\beta_t(u)]^T$ and $\tilde{\mathbf{n}}_t = [\mathbf{n}_0^T,\mathbf{n}_1^T,\ldots,\mathbf{n}_t^T]^T$. In Section 1.4 we discuss how to estimate u along with α using $\tilde{\mathbf{y}}_t$, and how to adaptively design the beamformer i.e., $\mathbf{f}_{t+1}$ for the next segment.

Remark 2. It is important to highlight the significance of the proposed sensing methodology. Given just two measurements, it is possible to construct a virtual ULA under the proposed sensing with $N_v = 2$. This is equivalent to a contrived single snapshot measurement from a physical array of size 2. Owing to the rich (array) geometrical information preserved in the measurements, it is thus possible to estimate the dominant path DoA in a gridless manner using existing techniques [59]. In contrast, the beam scan operation using $N_v = 1$ requires as many measurements as the codebook size to detect the DoA.

Constructing a Virtual Sparse Linear Array: The construction presented in the previous subsection can be extended to form virtual ULAs with more than $\lambda/2$ spacing[3]. More generally, a Sparse Linear Array (SLA) can also be realized as the virtual array geometry, for example, minimum redundancy arrays [60], nested arrays [61] or co-prime arrays [62]. These can help to increase the virtual aperture and improve resolution for the same segment duration. Let N_v denote the number of antenna elements in the virtual SLA we wish to construct over time. Let $\mathbb{P} = \{P_i : 0 \leq P_i < N, P_i \in \mathbb{Z}, i \in [N_v]\}$ denote the set of sensor positions in the SLA ordered in an increasing manner; $P_0 = 0$, without loss of generality. We design a beamformer, $\mathbf{f}_t$, of length $M = N - P_{N_v-1}$. The analog combiner at time l, in the case of SLA, is given by

$$\mathbf{w}_l = \begin{bmatrix} \mathbf{0}^T_{P_{\text{mod}(l,N_v)}} & \mathbf{f}_t^T & \mathbf{0}^T_{N-M-P_{\text{mod}(l,N_v)}} \end{bmatrix}^T. \tag{1.9}$$

Using identical notation to describe the complex gain, $\beta_t(u)$, as in (1.5), the signal y_l post-combining can be expressed as

$$y_l = \mathbf{w}_l^H \mathbf{x}_l = \sqrt{P_s}\alpha\beta_t(u) \cdot \exp\left(j\pi u P_{\text{mod}(l,N_v)}\right) + n_l. \tag{1.10}$$

Note that (1.10) generalizes (1.6) for the SLA case. Finally, the measurements within the t-th SLA segment can be stacked as

$$\begin{aligned} \mathbf{y}_t &= \sqrt{P_s}\alpha\beta_t(u)[1 \exp(j\pi P_1 u) \cdots \exp(j\pi P_{N_v-1} u)]^T + \mathbf{n}_t \\ &= \sqrt{P_s}\alpha\beta_t(u)\mathbf{S}_{\mathbb{P}}\boldsymbol{\phi}_N(u) + \mathbf{n}_t, \end{aligned} \tag{1.11}$$

[3]Any ambiguity in angular estimation can be resolved if the RoI is an appropriate fraction of the spatial region.

where $\mathbf{S}_{\mathbb{P}} \in \mathbb{R}^{N_v \times N}$ is a binary sampling matrix given by

$$[\mathbf{S}_{\mathbb{P}}]_{m,n} = \begin{cases} 1 & \text{if } n = P_m \\ 0 & \text{otherwise} \end{cases}, m \in [N_v], n \in [N]. \tag{1.12}$$

In the remainder of the work, we focus on the virtual ULA with $\lambda/2$ spacing-based sensing for ease of exposition, but the ideas presented can be easily extended to the virtual SLA case.

We refer to the sensing methodology described in this section as Synthesis of Virtual Array Manifold (SVAM). Furthermore, we describe the SVAM sensing in conjunction with the virtual ULA size (with $\lambda/2$ spacing) by SVAM-N_v. For example, SVAM-2 indicates the sensing methodology employed is as described in subsection 1.2.2 with $N_v = 2$.

1.3 Benefits of SVAM For Beam Alignment

We discuss some of the benefits of the proposed SVAM sensing for estimating the DoA, u. We demonstrate the benefits in two settings: i. agnostic to the adaptive scheme used, ii. when hierarchical posterior matching (hiePM) [35] scheme is used. In both settings we assume that α is known. We also briefly discuss the role of virtual aperture size N_v for improving the estimation performance. The more practical scenario, where α is unknown, is discussed in the next section.

1.3.1 Adaptive Scheme-Agnostic Analysis

We begin by first deriving the CRB after L snapshots, assuming α is known. Let $\mathbf{W} = [\mathbf{w}_0, \ldots, \mathbf{w}_{L-1}] \in \mathbb{C}^{N \times L}, \|\mathbf{w}_l\|_2 = 1$ denote the matrix of beamformers used to generate measurements $\mathbf{y} = [y_0, \ldots, y_{L-1}]^T \in \mathbb{C}^L$ as in (1.2). Note that $\mathbf{w}_l$'s may be designed generally, and not necessarily under SVAM for the following result to hold.

Theorem 1. The $\mathbf{CRB}(u)$ on the variance for estimating u using the beamformer matrix

$\mathbf{W} = [\mathbf{w}_0, \ldots, \mathbf{w}_{L-1}], \|\mathbf{w}_l\|_2 = 1$ as in (1.2) over L snapshots when $\boldsymbol{\alpha}$ is known is given by

$$\mathrm{CRB}(u) = \frac{\sigma_n^2}{2P_s|\boldsymbol{\alpha}|^2} \left\{ \left(\frac{\partial}{\partial u} \boldsymbol{\phi}_N(u) \right)^H \mathbf{W}\mathbf{W}^H \frac{\partial}{\partial u} \boldsymbol{\phi}_N(u) \right\}^{-1}. \tag{1.13}$$

Proof. The proof follows standard steps for deriving CRB and provided in Section 1.7.1 for completion. □

In this subsection, we compare two sensing strategies - both deterministically modify the beamformer design after every N_v snapshots. The proposed sensing strategy involves a shift in space as described in (1.4), and requires the beamformer to be fixed for N_v snapshots by design. The alternative (benchmark) strategy designs the beamformer $\mathbf{w}_l$ of size N without inserting 0's - used in SVAM to effect a linear time invariant operation. The beamformer for the two strategies considered here are to be designed with identical specifications except the length. The SVAM beamformer $\mathbf{f}_t$ has a size of $M(=N-N_v+1)$, whereas the alternative strategy utilizes the total antenna aperture of size N. We expect that for a large antenna size N, which is typical in mmWave systems, the slightly different length of the beamformers will have negligible impact. The rest of the beamformer specifications, which can change every N_v snapshots, may be chosen arbitrarily for the discussion in this subsection. We defer the discussion that involves using the specific adaptive scheme - HiePM [35] to the subsection 1.3.3. We specialize the CRB expression in Theorem 1 for the benchmark strategy and the proposed sensing strategy. This exercise helps to understand: how informative are the measurements available post analog combining about the unknown DoA using either of the two techniques for designing the analog combiners? For both the cases, we treat as if the same ordered set of L received signal snapshots $\mathbf{x}_l \in \mathbb{C}^N, l \in [L]$ were available at the antenna.

CRB for the Benchmark Strategy: Let $\mathbf{W}^{\mathrm{B}} = [\mathbf{w}_0^{\mathrm{B}}, \mathbf{w}_1^{\mathrm{B}}, \ldots, \mathbf{w}_{L-1}^{\mathrm{B}}] \in \mathbb{C}^{N \times L}$ be the L beam-

formers used to generate the measurements as in (1.2); superscript B highlights the benchmark sensing strategy. Note that

$$\mathbf{w}_l^{\mathrm{B}} = \mathbf{w}^{\mathrm{B}}_{N_v \times \mathrm{floor}(l/N_v)}, \tag{1.14}$$

under the benchmark scheme. Let $\mathbf{F}^{\mathrm{B}} = [\mathbf{w}_0^{\mathrm{B}}, \mathbf{w}_{N_v}^{\mathrm{B}}, \ldots, \mathbf{w}_{L-N_v}^{\mathrm{B}}] \in \mathbb{C}^{N \times \frac{L}{N_v}}$ denote the matrix of unique beamformers from $\mathbf{W}^{\mathrm{B}}$. We have the following result.

Corollary 1.1. The $\mathrm{CRB}(u)$ using measurements in (1.2) from the beamformer matrix $\mathbf{W}^{\mathrm{B}}$ under the constraint in (1.14) over L snapshots when α is known is given by

$$\mathrm{CRB}(u) = \frac{1}{N_v} \frac{\sigma_n^2}{2P_s|\alpha|^2} \left\{ \left(\frac{\partial}{\partial u} \boldsymbol{\phi}_N(u) \right)^H \mathbf{F}^{\mathrm{B}} (\mathbf{F}^{\mathrm{B}})^H \frac{\partial}{\partial u} \boldsymbol{\phi}_N(u) \right\}^{-1}. \tag{1.15}$$

Proof. The proof follows from simplifying (1.13) using (1.14), and is provided in Section 1.1. □

CRB for the Proposed Sensing Strategy: Let $\mathbf{W}^{\mathrm{P}} = [\mathbf{w}_0^{\mathrm{P}}, \mathbf{w}_1^{\mathrm{P}}, \ldots, \mathbf{w}_{L-1}^{\mathrm{P}}] \in \mathbb{C}^{N \times L}$ be the L beamformers designed under SVAM as described in (1.4); superscript P highlights the proposed SVAM sensing. Let $\mathbf{F}^{\mathrm{P}} = [\mathbf{f}_0, \mathbf{f}_1, \ldots, \mathbf{f}_{L/N_v-1}] \in \mathbb{C}^{M \times \frac{L}{N_v}}$ denote the matrix of SVAM beamformers. We first present a simple lemma which will be useful in simplifying the CRB expression. Let $[\mathbf{v}]_{m:n}$ denote the vector formed from a vector $\mathbf{v} \in \mathbb{C}^N$, using the components indexed from m to $n \geq m$ and $m, n \in [N]$.

Lemma 2. For $\boldsymbol{\phi}_N(u)$ denoting ULA manifold vector with N elements and half-wavelength inter-element spacing:

$$\left[\frac{\partial}{\partial u} \boldsymbol{\phi}_N(u) \right]_{m:n} = \exp(jm\pi u) \left(\frac{\partial}{\partial u} \boldsymbol{\phi}_{n-m+1}(u) + jm\pi \boldsymbol{\phi}_{n-m+1}(u) \right), \tag{1.16}$$

where $n \geq m$ and $m, n \in [N]$.

Proof. The proof follows from expanding the LHS, and identifying the two terms from

the RHS. □

We have the following result.

Corollary 2.1. The $\mathrm{CRB}(u)$ using measurements in (1.2) from the beamformer matrix $\mathbf{W}^{\mathrm{P}}$ under the construction in (1.4) over L snapshots when α is known is given by

$$\mathrm{CRB}(u) = \frac{1}{N_v}\frac{\sigma_n^2}{2P_s|\alpha|^2}\left\{\left(\frac{\partial}{\partial u}\boldsymbol{\phi}_M(u)\right)^H \mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H \frac{\partial}{\partial u}\boldsymbol{\phi}_M(u) + G\right\}^{-1}, \text{ where} \quad (1.17)$$

$$\begin{aligned} G = {} & \frac{\pi^2(N_v-1)(2N_v-1)}{6}\boldsymbol{\phi}_M(u)^H \mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H \boldsymbol{\phi}_M(u) \\ & -\pi(N_v-1)\mathrm{Im}\left\{\left(\frac{\partial}{\partial u}\boldsymbol{\phi}_M(u)\right)^H \mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H \boldsymbol{\phi}_M(u)\right\} \end{aligned} \quad (1.18)$$

Proof. The proof follows from simplifying (1.13) using (1.4), and is provided in Section 2.1. □

The CRB expression in (1.17) differs from the expression in (1.15) in two ways. The first term in (1.17) involves an array of dimension $M = N - N_v + 1$ instead of N in (1.15). For large array sizes, N, which are typical in mmWave systems, we expect this term to be similar to that in (1.15). Secondly, the denominator in (1.17) has an additional second term '$+G$'. We show that the conditions under which $G \geq 0$ is not difficult to satisfy, by deriving a sufficient condition to ensure the same.

Theorem 3. G in (1.18) is non-negative if

$$\frac{\boldsymbol{\phi}_M^H(u)\mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H\boldsymbol{\phi}_M(u)}{\|\boldsymbol{\phi}_M(u)\|^2} \geq \frac{\lambda_{\max}\left((\mathbf{F}^{\mathrm{P}})^H\mathbf{P}_{u,\perp}\mathbf{F}^{\mathrm{P}}\right)}{4}, \quad (1.19)$$

where $\lambda_{\max}(\mathbf{X})$ denotes the largest eigenvalue of the matrix $\mathbf{X}$. $\mathbf{P}_{u,\perp} = [\boldsymbol{\phi}_M(u)\ \boldsymbol{\phi}_M^{\perp}(u)] \times \begin{bmatrix} \|\boldsymbol{\phi}_M(u)\|^2 & 0 \\ 0 & \|\boldsymbol{\phi}_M^{\perp}(u)\|^2 \end{bmatrix}^{-1} [\boldsymbol{\phi}_M(u)\ \boldsymbol{\phi}_M^{\perp}(u)]^H$ denotes a projection onto the subspace of

orthogonal vectors $\boldsymbol{\phi}_M(u)$ and $\boldsymbol{\phi}_M^{\perp}(u) = \begin{bmatrix} -\frac{(M-1)}{2} & (1-\frac{(M-1)}{2}) & \dots & \frac{(M-1)}{2} \end{bmatrix}^T \odot \boldsymbol{\phi}_M(u)$.

Proof. The proof is provided in Appendix section 1.7.4. □

The following remark discusses the implication of Theorem 3.

Remark 3. If the left singular vectors of $\mathbf{F}^{\mathrm{P}}$ includes $\boldsymbol{\phi}_M(u)$ and $\boldsymbol{\phi}_M^{\perp}(u)$ (post-normalization), we can further simplify (1.19). If $\boldsymbol{\phi}_M(u)$ leads $\boldsymbol{\phi}_M^{\perp}(u)$, then (1.19) is trivially satisfied. If the opposite is true, in that, $\boldsymbol{\phi}_M^{\perp}(u)$ leads $\boldsymbol{\phi}_M(u)$, then (1.19) describes the required gap in the two corresponding singular values within $\mathbf{F}^{\mathrm{P}}$. In practice, we expect an adaptive scheme to choose beamformers close to the direction of the DoA. Also, it was found out that $G \geq 0$ very often even when $\mathbf{F}^{\mathrm{P}}$ contained i.i.d. complex Gaussian random entries.

The above analysis evaluates the impact of the SVAM sensing when compared to a benchmark scheme. The benchmark analysis can be extended to include more general scenario, wherein it need not repeat the beamformer in N_v snapshots. Since the contribution from each snapshot appears independent of other snapshots, as a linear sum in the denominator in the general CRB expression in (1.13), any time a beamformer is repeated in a benchmark scheme, it can be replaced with the proposed scheme. A similar impact and analysis can be carried out to reveal ensuing benefits.

1.3.2 Case Study: Combining SVAM with HiePM Framework

The hiePM algorithm [35] processes each new snapshot and updates the beamformer based on the current estimate of the posterior density on the unknown DoA. To incorporate the proposed sensing, one approach is to update the beamformer after every N_v snapshots, where N_v denotes the size of the virtual ULA. Within each N_v interval, a SVAM beamformer, $\mathbf{f}_t$, is designed; $\mathbf{f}_t$ is simply a codeword from the hierarchical codebook that satisfies the selection criteria within hiePM framework. The beamfomer (of physical antenna aperture size, N) for N_v snapshots within this interval is constructed as

in (1.4). The remainder steps in Algorithm 1 in [35] are compatible with the proposed sensing. This modification also becomes crucial when α is unknown; this is discussed in Section 1.4.

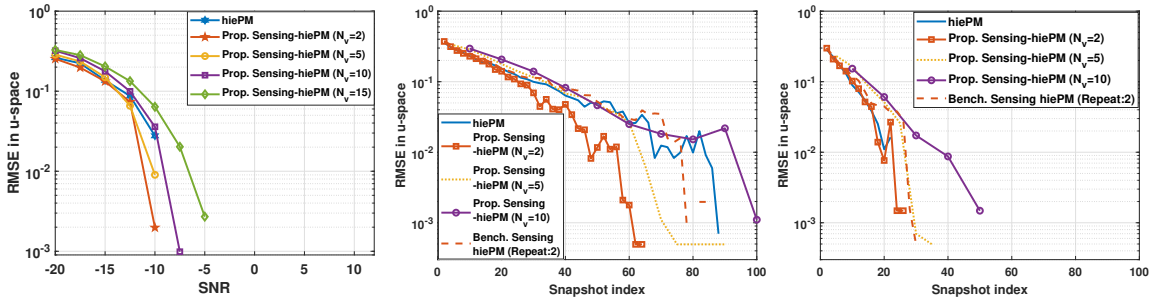
Remark 4. The impact of the modification can be understood in the following manner. The hiePM strategy in [35] may repeat the same beamformer multiple times until the posterior condition triggers a new beamformer. While it repeats the beamformer, it gains only in terms of the Signal-to-Noise Ratio (SNR), which is also evident from the CRB analysis for the benchmark scheme in (1.15)[4]. Instead of a repeated beamformer, the proposed sensing aims to achieve both, a SNR boost and at the same time an (virtual) aperture gain ('$+G$' term in the CRB analysis). Since the adaptive beamformer sequence is unknown apriori, analysis of the proposed modification to hiePM in this subsection is more involved. Instead, we provide empirical studies in the next subsection.

Remark 5. The proposed sensing can be incorporated in other beam alignment procedures as well. The inference procedure and adaptive strategy may require suitable modifications, or be left unaltered as demonstrated using the hiePM case study in this subsection.

1.3.3 Numerical Results

We numerically analyze the benefits of the proposed sensing when used with the hiePM framework in [35], and when α is assumed to be known. We compare the hiePM [35] approach with i. the proposed modified algorithm in Section 1.3.2, and ii. the benchmark scheme where the beamformer is simply repeated, as described in Section 1.3.1. We assume that the DoA, u, lies in between $[0,1)$ i.e., RoI is $\frac{1}{2}$ of the entire space. The information about the RoI is incorporated into the posterior calculation. We set the following parameters as: Physical antenna size: $N = 64$, number of grid points uniformly spaced in RoI: $G = 64$, number of random realizations for averaging: $Q = 1000$. The main metric employed for comparison is the root mean squared error (RMSE) computed in

[4] N_v multiplied with P_s in (1.15) indicates a N_v-fold SNR boost.



(a) Snapshots, $L = 60$ (b) SNR $= -10$ dB (c) SNR $= -5$ dB

Figure 1.2. (a) RMSE vs SNR: Proposed SVAM sensing with modified-hiePM improves over hiePM [35] when $N_v \in \{2,5\}$. (b) RMSE vs number of snapshots at SNR$=-10$ dB: Proposed sensing reduces the duration needed for beam alignment when $N_v \in \{2,5\}$. (c) RMSE vs number of snapshots at SNR$=-5$ dB: At high SNR, the proposed scheme still improves over the benchmark scheme, but takes slightly more time to converge.

u-space as RMSE $= \sqrt{\frac{1}{Q}\sum_{q=1}^{Q}\left(\hat{u}_q - u_q\right)^2}$.

Performance as a function of SNR: In Fig. 1.2 (a) we plot the RMSE as a function of SNR for the modified-hiePM under the proposed SVAM sensing using different virtual ULA sizes. As observed in this plot, given $L = 60$ snapshots, the curves using $N_v = 2$ and $N_v = 5$ improve over the hiePM in [35]. As the virtual aperture increases, the number of beamformer updates, given by 'L/N_v', reduces. Beyond $N_v = 5$, the reduced number of updates or longer acquisition time before a beamformer update, is seen to negatively impact the performance. Studying the tradeoff between the virtual aperture size and the frequency of update as they impact the beam alignment performance is an important future direction.

Performance as a function of number of snapshots: In Fig. 1.2 (b) and (c), we plot the RMSE as a function of number of snapshots at SNR$=-10$ dB and -5 dB, respectively. As observed from the Fig. 1.2 (b), the performance of hiePM improves under the proposed sensing when N_v is set to either 2 or 5. Setting it to a higher value degrades the performance, for reasons similar to those discussed for Fig. 1.2 (a). In red dashed curve we plot the performance for hiePM with benchmark sensing scheme (in Section 1.3.1). Although

the benchmark scheme improves the beam alignment duration over hiePM (solid blue curve), it still is much higher compared to setting $N_v = 2$ in the proposed sensing. This indicates that the improvement under SVAM is not merely due to more reliable beamformer adaptations caused by basing adaptations on more measurements. It is inherent to the phase response characteristics of the beamformer under SVAM, as described in (1.6). Note that as we increase SNR to -5 dB, the performance gains can be limited as seen from Fig. 1.2(b). This highlights that SVAM is beneficial in reducing the training time at low SNR. Also it is again observed that SVAM improves over the benchmark scheme. Thus, a dynamic method to adapt the virtual ULA size is an interesting future direction.

1.4 Beam Alignment Algorithm with Unknown α

We begin with a wide RoI, $u \in [u_l, u_r], u_l < u_r, u_l, u_r \in [-1, 1)$, and presume that the DoA lies within the RoI. Therefore, we initially set the SVAM beamformer to span this region and suppress any interference coming from outside the RoI. This ensures that we incorporate any prior information available about the DoA. The approach is also practical, as base stations are typically deployed with dedicated antennas to serve a specific RoI. We collect measurements over time and process them to compute an approximate posterior on the unknown angle. We adapt SVAM beamformer once we accrue enough posterior mass around the mode of the posterior. We now describe the methodology adopted in this work to estimate the posterior and to adapt the SVAM beamformer over time.

1.4.1 Algorithm Preliminaries: Initializing Angular Grid and Stochastic Modeling of both α and u

Since α is unknown but assumed to be fixed during the training phase, it can be estimated along with the DoA given two or more measurements. We observe this from the conditional CRB on variance for angle estimation when α is unknown. Let $\mathbf{W} = [\mathbf{w}_0, \ldots, \mathbf{w}_{L-1}] \in \mathbb{C}^{N \times L}$ denote the beamformer matrix used to generate measurements

$\mathbf{y} = [y_0, \ldots, y_{L-1}]^T \in \mathbb{C}^L$ as in (1.2). The beamformers may be designed generally, and not necessarily under SVAM. The $\mathbf{CRB}(u)$ is given by

$$\begin{aligned}\mathrm{CRB}(u) = \frac{\sigma_n^2}{2P_s|\alpha|^2} \Bigg\{ \left(\frac{\partial}{\partial u} \boldsymbol{\phi}_N(u) \right)^H \mathbf{W} \\ \times \left(\mathbf{I} - \frac{\mathbf{W}^H \boldsymbol{\phi}_N(u) \boldsymbol{\phi}_N(u)^H \mathbf{W}}{\boldsymbol{\phi}_N(u)^H \mathbf{W} \mathbf{W}^H \boldsymbol{\phi}_N(u)} \right) \mathbf{W}^H \frac{\partial}{\partial u} \boldsymbol{\phi}_N(u) \Bigg\}^{-1}, \end{aligned} \tag{1.20}$$

and can be derived from the conditional CRB expression (Theorem 4.1) in [63] by stacking all L snapshots along a column and replacing $\boldsymbol{\phi}_N(u)$ and $\frac{\partial}{\partial u}\boldsymbol{\phi}_N(u)$ by $\mathbf{W}^H\boldsymbol{\phi}_N(u)$ and $\mathbf{W}^H \frac{\partial}{\partial u}\boldsymbol{\phi}_N(u)$, respectively. For $L = 1$, we can see that the CRB is infinite. Furthermore, if the beamformer is kept fixed over time i.e., if $\mathbf{w}_l = \mathbf{w}_0, l \in [L]$, or in general if $\mathbf{W}$ is rank-one, even then the CRB is infinite. This exercise highlights a key requirement for being able to estimate both DoA and α simultaneously. The magnitude or phase response of the beamformer $\mathbf{w}_l$ should have variation over time, and it must vary differently for different angles. The requirement is not satisfied in the case of hiePM algorithm, if initial measurements are all taken using a fixed codeword, which is possible, early on, within the hiePM framework[5]. In the proposed sensing scheme this condition is naturally prevented, as the phase response (over time) leads to the virtual array manifold, which is sufficient to estimate both α and u when $N_v > 1$ and given at least two measurements.

The two basic pre-requisities for adapting the SVAM beamformer are the next beam direction to steer towards, and the beamwidth. Since the problem-at-hand considers the single path scenario, the beam direction can be estimated by matching-based criteria which results in a maximum likelihood estimate (MLE), and it also yields a deterministic estimate for α. However, for a good beamwidth selection we need to account for the uncertainty in the estimation procedure. The bounds for the CRB analysis can be used. In this work, we achieve this by modeling both α and the unknown DoA, u, as stochastic

[5]The CRB analysis is applicable when the unknowns (α, u) are estimated in the maximum likelihood sense. The issues can be circumvented by imposing a grid or strong prior information on α [37].

variables. For the latter, we impose a uniform prior distribution in the wide RoI. We model α with a complex circular Gaussian distribution with a parameterized prior.

We introduce a uniform grid, $\mathbf{u}_{\text{grid}}$, of size G within the RoI, $[u_l, u_r]$. Let $u_i, i \in [G]$, denote the i-th grid point. The initial grid can be refined as we successively reduce uncertainty of the estimate for u over snapshots. The grid refinement aspect is not focused in this paper, and left as future work. For each candidate DoA, u_i, we estimate a corresponding value for the complex path gain; let us denote the same as α_i. We impose the following parameterized prior on α_i

$$\alpha_i \sim \mathscr{CN}(0, \gamma_i). \tag{1.21}$$

Remark 6. Assigning a different complex gain, α_i, per candidate angle, u_i, is a non-trivial choice. By doing so, we allow each candidate u_i to explain the measurements at its best. Going forward, we explicitly utilize the presence of a single path by identifying suitable value for hyperparameters, γ_i, separately, instead of jointly. A joint optimization is widely adopted in the absence of knowledge of the model order [12].

A consequence of the prior imposed in (1.21) is that the marginalized pdf of the measurements $\tilde{\mathbf{y}}_t$ in (1.8) after $(t+1)N_v$ snapshots, conditioned on angle $u = u_i$ is given by

$$\begin{aligned} f(\tilde{\mathbf{y}}_t \mid u = u_i; \gamma_i) &= \int f(\tilde{\mathbf{y}}_t \mid u = u_i, \alpha_i) f(\alpha_i; \gamma_i)\, d\alpha_i \\ &= \frac{1}{\pi^{(t+1)N_v} \det(\boldsymbol{\Sigma}_{i,t})} \exp\left(-\tilde{\mathbf{y}}_t^H \boldsymbol{\Sigma}_{i,t}^{-1} \tilde{\mathbf{y}}_t\right), \text{ where} \end{aligned} \tag{1.22}$$

$$\begin{aligned} \boldsymbol{\Sigma}_{i,t} &= P_s \gamma_i \left(\tilde{\boldsymbol{\beta}}_t(u_i) \otimes \boldsymbol{\phi}_{N_v}(u_i)\right) \left(\tilde{\boldsymbol{\beta}}_t(u_i) \otimes \boldsymbol{\phi}_{N_v}(u_i)\right)^H + \sigma_n^2 \mathbf{I} \\ &= P_s \gamma_i \left(\tilde{\boldsymbol{\beta}}_t(u_i) \tilde{\boldsymbol{\beta}}_t(u_i)^H\right) \otimes \left(\boldsymbol{\phi}_{N_v}(u_i) \boldsymbol{\phi}_{N_v}(u_i)^H\right) + \sigma_n^2 \mathbf{I}, \end{aligned} \tag{1.23}$$

where in the last line, we use the mixed product property of Kronecker product, namely

if matrix products **AC** and **BD** are defined, then $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = \mathbf{AC}\otimes\mathbf{BD}$. Note that σ_n^2 is assumed to be known[6]. In the absence of such knowledge, it can be included in the estimation procedure (see [12]). We focus on the estimation of α in this work, and study the impact of imperfect knowledge of σ_n^2 in the simulation section.

1.4.2 Estimation of Hyperparameters under the MLE Framework

We find the hyperparameters of the imposed prior, namely $\gamma_i, i \in [G]$, in the MLE sense. The derivation is simple, and similar to the work in [13,64]. Since $\mathbf{\Sigma}_{i,t}$ in (1.23) can be expressed as a rank-one perturbation to the noise covariance matrix, the determinant can be expressed in closed-form as

$$\det(\mathbf{\Sigma}_{i,t}) = (P_s\gamma_i g_t(u_i)\|\boldsymbol{\phi}_{N_v}(u_i)\|_2^2 + \sigma_n^2)\cdot(\sigma_n^2)^{(t+1)N_v-1}, \tag{1.24}$$

where $g_t(u_i) = \sum_{t'=0}^{t}|\beta_{t'}(u_i)|^2$. Also, using matrix-inversion lemma, we can simplify $\tilde{\mathbf{y}}_t^H\mathbf{\Sigma}_{i,t}^{-1}$ $\times\tilde{\mathbf{y}}_t$ as

$$\tilde{\mathbf{y}}_t^H\mathbf{\Sigma}_{i,t}^{-1}\tilde{\mathbf{y}}_t = \sigma_n^{-2}\|\tilde{\mathbf{y}}_t\|_2^2 - \frac{P_s\gamma_i}{\sigma_n^2}\times\frac{|\boldsymbol{\phi}_{N_v}^H(u_i)\sum_{t'=0}^{t}\beta_{t'}^c(u_i)\mathbf{y}_{t'}|^2}{(P_s\gamma_i g_t(u_i)\|\boldsymbol{\phi}_{N_v}(u_i)\|_2^2+\sigma_n^2)}. \tag{1.25}$$

Using (1.24) and (1.25), we maximize $f(\tilde{\mathbf{y}}_t \mid u = u_i;\gamma_i)$ over γ_i by setting the derivative w.r.t. the same as zero. We get

$$\hat{\gamma}_{i,t} = \max\left\{0, \frac{1}{P_s g_t(u_i)\|\boldsymbol{\phi}_{N_v}(u_i)\|_2^2}\times\left(\left|\frac{\left(\tilde{\boldsymbol{\beta}}_t(u_i)\otimes\boldsymbol{\phi}_{N_v}(u_i)\right)^H}{\|\tilde{\boldsymbol{\beta}}_t(u_i)\otimes\boldsymbol{\phi}_{N_v}(u_i)\|_2}\tilde{\mathbf{y}}_t\right|^2 - \sigma_n^2\right)\right\}. \tag{1.26}$$

The above can be interpreted as normalized beamforming output power that is compensated for the noise power. The beamformer in this case is the conventional beamformer [56] and takes into account the complex gain, $\tilde{\boldsymbol{\beta}}_t(u_i)$, from the SVAM beamformer $\mathbf{f}_{t'}, t' \in [t+1]$.

[6]This assumption is also made in other works [35,37].

The posterior density of α_i is also complex circular Gaussian, and can be computed as

$$f(\alpha_i \mid \tilde{\mathbf{y}}_t, u_i; \hat{\gamma}_{i,t}) = \frac{1}{\pi \hat{\sigma}^2_{\alpha_i,t}} \exp\left(-\frac{|\alpha_i - \hat{\mu}_{\alpha_i,t}|^2}{\hat{\sigma}^2_{\alpha_i,t}} \right), \tag{1.27}$$

$$\begin{aligned} \text{where } \hat{\mu}_{\alpha_i,t} &= \sqrt{P_s}\hat{\gamma}_{i,t} \frac{\left(\tilde{\boldsymbol{\beta}}_t(u_i) \otimes \boldsymbol{\phi}_{N_v}(u_i)\right)^H \tilde{\mathbf{y}}_t}{P_s \hat{\gamma}_{i,t} g_t(u_i) \|\boldsymbol{\phi}_{N_v}(u_i)\|_2^2 + \sigma_n^2} \\ \hat{\sigma}^2_{\alpha_i,t} &= \hat{\gamma}_{i,t} \frac{\sigma_n^2}{P_s \hat{\gamma}_{i,t} g_t(u_i) \|\boldsymbol{\phi}_{N_v}(u_i)\|_2^2 + \sigma_n^2}. \end{aligned} \tag{1.28}$$

1.4.3 Computing Posterior on u

We begin with describing the posterior probability, $p(u = u_i \mid \tilde{\mathbf{y}}_t)$, and identifying the missing pieces for computing this posterior. Using Bayes' rule, we can write

$$p(u = u_i \mid \tilde{\mathbf{y}}_t) = \frac{p(u = u_i) f(\tilde{\mathbf{y}}_t \mid u_i)}{f(\tilde{\mathbf{y}}_t)} \overset{(a)}{\propto} f(\tilde{\mathbf{y}}_t \mid u_i), \tag{1.29}$$

where (a) follows from two facts: i. we rely on discrete uniform (prior) distribution on u, ii. $f(\tilde{\mathbf{y}}_t)$ does not depend on u. Consequently, the goal is to be able to compute $f(\tilde{\mathbf{y}}_t \mid u_i)$ as accurately possible. We explore this direction next.

Using marginalization and Bayes' rule we get

$$\begin{aligned} f(\tilde{\mathbf{y}}_t \mid u_i) &= \int f(\tilde{\mathbf{y}}_t, \alpha \mid u_i) d\alpha \\ &= \int f(\alpha \mid u_i) f(\tilde{\mathbf{y}}_t \mid \alpha, u_i) d\alpha. \end{aligned} \tag{1.30}$$

The integrand consists of two factors, of which, the second factor, $f(\tilde{\mathbf{y}}_t \mid \alpha, u_i)$, is well-defined, provided the distribution of noise. However, the first factor, $f(\alpha \mid u_i)$, is unavailable.

In the absence of this knowledge, we approximate the unknown distribution with

the best proxy available at-hand that uses all the available measurements i.e., we replace $f(\alpha \mid u_i)$ with $f(\alpha_i \mid \tilde{\mathbf{y}}_t, u_i; \hat{\gamma}_{i,t})$. The required quantity, $f(\alpha_i \mid \tilde{\mathbf{y}}_t, u_i; \hat{\gamma}_{i,t})$, was computed in the previous subsection, and we update this proxy as we collect more measurements. In other words, we compute the following approximate likelihood

$$\begin{aligned}\hat{f}_t(\tilde{\mathbf{y}}_t \mid u_i) &= \int f(\alpha_i \mid \tilde{\mathbf{y}}_t, u_i; \hat{\gamma}_{i,t}) f(\tilde{\mathbf{y}}_t \mid \alpha_i, u_i) d\alpha_i \\ &= \frac{1}{\pi^{(t+1)N_v} \det(\tilde{\mathbf{\Sigma}}_{i,t})} \exp\left(-\tilde{\mathbf{e}}_{i,t}^H \tilde{\mathbf{\Sigma}}_{i,t}^{-1} \tilde{\mathbf{e}}_{i,t}\right), \end{aligned} \tag{1.31}$$

where $\tilde{\mathbf{e}}_{i,t} = \tilde{\mathbf{y}}_t - \tilde{\boldsymbol{\mu}}_{i,t}$, $\tilde{\boldsymbol{\mu}}_{i,t} = \sqrt{P_s}\hat{\mu}_{\alpha_i,t}\left(\tilde{\boldsymbol{\beta}}_t(u_i) \otimes \boldsymbol{\phi}_{N_v}(u_i)\right)$ and $\tilde{\mathbf{\Sigma}}_{i,t} = P_s \hat{\sigma}^2_{\alpha_i,t}\left(\tilde{\boldsymbol{\beta}}_t(u_i)\tilde{\boldsymbol{\beta}}_t(u_i)^H\right) \otimes \left(\boldsymbol{\phi}_{N_v}(u_i)\boldsymbol{\phi}_{N_v}(u_i)^H\right) + \sigma_n^2 \mathbf{I}$. $(\hat{\mu}_{\alpha_i,t}, \hat{\sigma}^2_{\alpha_i,t})$ were estimated in the previous subsection.

Next, we demonstrate the procedure to compute the required quantities in (1.31) efficiently. We provide the final result below

$$\det(\tilde{\mathbf{\Sigma}}_{i,t}) = (P_s \sigma^2_{\alpha_i,t} g_t(u_i) \|\boldsymbol{\phi}_{N_v}(u_i)\|_2^2 + \sigma_n^2) \times (\sigma_n^2)^{(t+1)N_v - 1} \tag{1.32}$$

$$\tilde{\mathbf{e}}_{i,t}^H \tilde{\mathbf{\Sigma}}_{i,t}^{-1} \tilde{\mathbf{e}}_{i,t} = \sigma_n^{-2} \|\tilde{\mathbf{e}}_{i,t}\|_2^2 - \frac{P_s \sigma^2_{\alpha_i,t}}{\sigma_n^2} \times \frac{|\boldsymbol{\phi}_{N_v}^H(u_i) \sum_{t'=0}^{t} \beta_{t'}^c(u_i) \tilde{\mathbf{e}}_{i,t'}|^2}{(P_s \sigma^2_{\alpha_i,t} g_t(u_i) \|\boldsymbol{\phi}_{N_v}(u_i)\|_2^2 + \sigma_n^2)}. \tag{1.33}$$

Once the required quantities in (1.31) are computed, we compute the likelihood estimate, $\hat{f}_t(\tilde{\mathbf{y}}_t \mid u_i)$, for all grid points $u_i, i \in [G]$, in the RoI. We get the estimate for the posterior as

$$\hat{p}_t(u = u_i \mid \tilde{\mathbf{y}}_t) = \frac{\hat{f}_t(\tilde{\mathbf{y}}_t \mid u_i)}{\sum_{i'} \hat{f}_t(\tilde{\mathbf{y}}_t \mid u_{i'})}. \tag{1.34}$$

Let $\hat{\mathbf{p}}_t = [\hat{p}_t(u_0 \mid \tilde{\mathbf{y}}_t), \hat{p}_t(u_1 \mid \tilde{\mathbf{y}}_t), \ldots, \hat{p}_t(u_{G-1} \mid \tilde{\mathbf{y}}_t)]^T$. We adapt the SVAM beamformer, $\mathbf{f}_{t+1}$, for the next snapshot based on the computed posterior in (1.34). We discuss this step next.

1.4.4 Proposed Adaptive SVAM Beamforming Algorithm

The aim is to ensure that the DoA lies within the passband of the SVAM beamformer, $\mathbf{f}_t$, so that effective received SNR for inference is high. A high effective SNR helps

Algorithm 1: Proposed Adaptive SVAM Beamforming

Result: Beam direction: u^*

Input: $L, N_v, P_s, \sigma_n^2, \text{beam\_dir}, \text{BW}_{\mathbf{f}} := \text{BW}_{\text{initial}}, p_{\text{thresh}}$

Initialize: $\text{beam\_spec\_initial} := \{\text{beam\_dir}, \text{BW}_{\mathbf{f}}\}, \mathbf{f}_0 :=$ beamformer_design(beam_spec_initial)

for $l := 0$ to $L-1$ do

  $\mathbf{y}_l = \mathbf{w}_l^H \mathbf{x}_l$ (New measurement, $\mathbf{w}_l$ as in (1.4))

  if $\text{mod}(l+1, N_v) = 0$ then

    $t := (l+1)/N_v - 1$

    Form $\mathbf{y}_t$, then $\tilde{\mathbf{y}}_t$ as in (1.7) and (1.8) respectively

    (α Prior): Compute $\hat{\gamma}_{i,t}$ as in (1.26), $i \in [G]$

    (α Posterior): Compute $\hat{\mu}_{\alpha_i,t}, \hat{\sigma}^2_{\alpha_i,t}$ as in (1.28)

    (Likelihood $\mid u_i$): Compute likelihood in (1.31)

    (u_i Posterior): $\hat{\mathbf{p}}_t :=$Compute pmf in (1.34)

    $\text{BW}_{\text{check}} := 0.5 \times \text{BW}_{\mathbf{f}}$

    $[\text{peak\_prob}, \text{beam\_spec}] := \text{cumul\_peak}(\hat{\mathbf{p}}_t, \text{BW}_{\text{check}})$ (Algorithm 2)

    while $\text{peak\_prob} < p_{\text{thresh}}$ do

      $\text{BW}_{\text{check}} := 2 \times \text{BW}_{\text{check}}$

      $[\text{peak\_prob}, \text{beam\_spec}] := \text{cumul\_peak}(\hat{\mathbf{p}}_t, \text{BW}_{\text{check}})$ (Algorithm 2)

    end

    $\text{BW}_{\mathbf{f}} := \text{BW}_{\text{check}}$

    $\mathbf{f}_{t+1}$:=beamformer_design(beam_spec)

  end

end

$i^* := \arg\max_i \hat{p}_{L/N_v-1}(u_i \mid \tilde{\mathbf{y}}_{L/N_v-1}), u^* := u_{i^*}$

further to ensure a successful beam alignment. The presented approach still may not ensure that the DoA always stays in the passband. We circumvent this issue by evaluating the posterior $\hat{p}_t(u_i \mid \tilde{\mathbf{y}}_t)$ over the entire RoI. This allows the posterior mass to move freely and helps to recover when the SVAM beamformer adapted prematurely in the incorrect region. In other words, even if we adapt the SVAM beamformer, the RoI is kept fixed.

Proposed Algorithm: We initially set the beam direction of the SVAM beamformer, $\mathbf{f}_0$, to the centre of the RoI, and the beamwidth to $\text{BW}_{\mathbf{f}} = \text{BW}_{\text{initial}} < 2$ in u-space; $\text{BW}_{\text{initial}}$ chosen to cover the RoI. We adapt the SVAM beamformer only if there is a sufficient posterior mass concentrated around the posterior mode. More specifically, we adapt the beamformer if the peak of the posterior mass in any contiguous span of a specific

Algorithm 2: Cumulative Posterior Peak

Result: $\text{peak\_prob}, \text{beam\_spec} := \{\text{BD\_sel}, \text{BW}'\}$
Input: $\mathbf{p}, \text{BW}'$
Initialize: $\mathbf{1}_{\text{BW}'}$ (vector of 1's of size $\equiv \text{BW}'$)
$\text{mode} := \arg\max_j \mathbf{p}(j)$
$\text{cumul\_prob} := \mathbf{p} * \mathbf{1}_{\text{BW}'}$ ('$*$': convolution operation)
$[\text{peak\_prob}, k] := \max(\text{cumul\_prob}(\text{mode} : \text{mode} + \text{size}(\mathbf{1}_{\text{BW}'}) - 1)$
$\text{BD\_sel} := \mathbf{u}_{\text{grid}}(\text{mode} + k) - \text{BW}'/2$

beamwidth, that includes the posterior mode, exceeds a fixed threshold p_{thresh}. We begin with setting the beamwidth for this search to $\text{BW}_{\text{check}} = 0.5 \times \text{BW}_{\mathbf{f}}$ i.e., half of the current beamwidth for the SVAM sensing. If the threshold condition is not satisfied, BW_{check} is doubled. This continues until the posterior threshold condition is met. Note that in the default scenario, the next beam resorts to the initial beam specifications. When the posterior threshold condition is met, the corresponding contiguous span of angular grid points is selected, and the next SVAM beamformer, $\mathbf{f}_{t+1}$, is designed to cover the selected region. The threshold parameter p_{thresh} may be set to a fixed value or chosen dynamically. A lower value for p_{thresh} allows the SVAM beamformer to adapt often, whereas a higher value makes the adaptations more cautious. We discuss the role of this crucial parameter in detail in Section 1.5. Algorithm 1 summarizes the overall approach.

Remark 7. The proposed adaptive beam search mechanism has the ability to both refine, as well as correct erroneous adaptations. In the case when the beam is adapted in the wrong portion of the spatial region, the posterior mass along with the mode is expected to either shift or widen as more measurements are collected. The beam gets rectified within the proposed scheme as the search is carried out around the updated posterior mode, and it has the ability to widen the beamwidth beyond the $\text{BW}_{\mathbf{f}}$ presently in use for the sensing.

The presented approach to adapt the beamformer has an equivalent representation within the compact hierarchical codebook, and is discussed next.

Using Hierarchical Codebook: The proposed scheme in Algorithm 1 adopts a flexible

Algorithm 3: Beam Search In Hierarchical Codebook

Result: $l^h_{\text{final}} \in [\log_2 G], k^h_{\text{final}} \in [2^{l^h_{\text{final}}} - 1]$
Input: l^h_{init} (current level in hierarchy), $\mathbf{p}$ (posterior pmf), G, p_{thresh}
1 Initialize: $l^h := l^h_{\text{init}} + 1, G_{l^h} := 2^{l^h}$
2 $\text{mode} := \arg\max_j \mathbf{p}(j)$
3 (node at level l^h): $k^h := \text{floor}\left(\text{mode} \times \frac{G_{l^h}}{G}\right)$
4 for $\bar{l}^h := l^h \text{ to } 0$ (descending) do
5 $\quad \text{node\_k\_ind} := \frac{G}{G_{l^h}} k^h : \frac{G}{G_{l^h}}(k^h + 1) - 1$
6 $\quad$ if $\text{sum}(\mathbf{p}(\text{node\_k\_ind})) \geq p_{\text{thresh}}$ then
7 $\quad\quad$ break
8 $\quad$ else
9 $\quad\quad k^h := \text{floor}(k^h/2)$
10 $\quad\quad G_{l^h} := G_{l^h}/2$
11 $\quad$ end
12 end
13 $l^h_{\text{final}} := l^h, k^h_{\text{final}} := k^h$

beam design which requires beam direction and beamwidth to design the SVAM beamformer. In some cases, such flexible beam designs may not be possible due to complexity, and a smaller codebook may be desired. We take an example of the (binary) hierarchical codebook [36] and demonstrate the procedure to select a codeword based on the posterior computed in (1.34). The procedure closely mimics the approach taken in Algorithm 1, while constraining the beam to belong to the hierarchical codebook.

The basic idea is to begin at one level below in hierarchy compared to the current level used for SVAM sensing, and traverse up the hierarchy to satisfy the posterior condition. At one-level below current, we begin with that node which contains the posterior mode. The principle in doing so, is to try and ensure that the DoA is included in the next beam. We summarize this beam search in Algorithm 3 which can replace line $11-18$ in Algorithm 1. We also study the impact of using a hierarchical codebook on the performance of the proposed adaptive scheme in simulation Section 1.5.6.

Remark 8. A variation on the beam search Algorithm 3 is to initialize the search at a

deeper hierarchical level, which can be favourable to converge early at high SNRs, but may lead to premature beam focusing at low SNRs. Similar idea can be incorporated with Algorithm 1 as well. This is not explored further in this paper, but left as future work.

On HiePM-based Beam Alignment: The posterior computed in (1.31) involves approximating the posterior on $\boldsymbol{\alpha}$, $f(\boldsymbol{\alpha} \mid u_i)$, using a Gaussian density, $f(\boldsymbol{\alpha}_i \mid \tilde{\mathbf{y}}_t, u_i; \hat{\boldsymbol{\gamma}}_{i,t})$. This allows to bypass a grid approach on $\boldsymbol{\alpha}$, which was explored in [37] under Algorithm 1. Although the grid approach can be more accurate, it is computationally expensive. Also initializing an appropriate grid on $\boldsymbol{\alpha}$ is a non-trivial problem.

Although the adopted prior on $\boldsymbol{\alpha}$ in this work and within Alg. 2 in [37], both are Gaussian distributed, there is an important distinction. The presented approach utilizes a parameterized Gaussian prior unlike the fixed Gaussian prior in [37]. This has a couple of implications: a) We allow the framework to learn the relevant parameters directly from the measurements without relying on any additional information such as prior mean or variance, b) The choice of prior here allows the framework to fit the best Gaussian posterior on $\boldsymbol{\alpha}$ using the measurements, in contrast to the fixed prior case where only the posterior mean is learned from measurements and the variance in the passband depends only on the prior and the noise variance level. It is known that the inference is less sensitive to the values of higher-level hyperparameters than the values for the $\boldsymbol{\alpha}$-prior distribution (see [30] and references therein). This leads to more robust learning.

Most importantly, the proposed approach processes measurements jointly for computing the posterior on $\boldsymbol{u}$ (see (1.29)). In contrast, in [37] the $\boldsymbol{u}$-posterior is updated sequentially. Since $\boldsymbol{\alpha}$ is considered static in time, both in this work and within Algorithm 2 in [37], and because the posteriors are approximate, it was observed that processing measurements jointly instead of sequentially performs better. In other words, the $\boldsymbol{u}$-posterior at previous snapshot is not a sufficient statistic when computed approximately. The

impact is studied in Section 1.5.1.

1.5 Numerical Results

In this section we perform numerical experiments to study the performance of the proposed sensing and beam alignment Algorithm 1 under different scenarios, and the impact of the parameters involved. The parameters that are central to the performance include i. Size of the virtual ULA, N_v, ii. noise variance parameter σ_n^2, iii. posterior threshold, p_{thresh}. In Section 1.3.3, we studied the role of N_v; we focus on the parameters σ_n^2 and p_{thresh} in this section. We assume that the DoA lies in between $[0,1)$ i.e., RoI is $\frac{1}{2}$ of the entire space. The information about the RoI is incorporated into the posterior calculation. Note that the RoI can be arbitrary, and one may choose a wider or narrower RoI depending on the use case specifications. The virtual ULA constructed here has the same inter-element spacing as the physical antenna. Depending on the RoI, a wider virtual inter-element spacing can be realized[7], but in the following simulation we do not exploit it further. We set the following parameters, unless otherwise specified as: Physical antenna size: $N = 64$, number of grid points uniformly spaced in RoI: $G = 64$, virtual ULA size: $N_v = 4$, total number of snapshots: $L = 120$, number of random realizations for averaging: $Q = 100$. The SVAM beamformer is designed using the Parks-McClellan algorithm [58] with the passband edge set to realize the desired beamwidth, and the transition width set as a fraction of the desired beamwidth.

1.5.1 Performance of Algorithm 1 as a function of SNR

In Fig. 1.3, we study the RMSE in u-space as a function of SNR in dB. The red curve (with pentagram markers) plots the performance of hiePM with the SVAM sensing scheme (Section 1.3.2). It is seen to improve over the hiePM algorithm in blue curve

[7]For e.g., here a λ-spacing instead of $\lambda/2$ may be realized without causing ambiguity in angular estimation, since the DoA lies in $\frac{1}{2}$ of the spatial region.

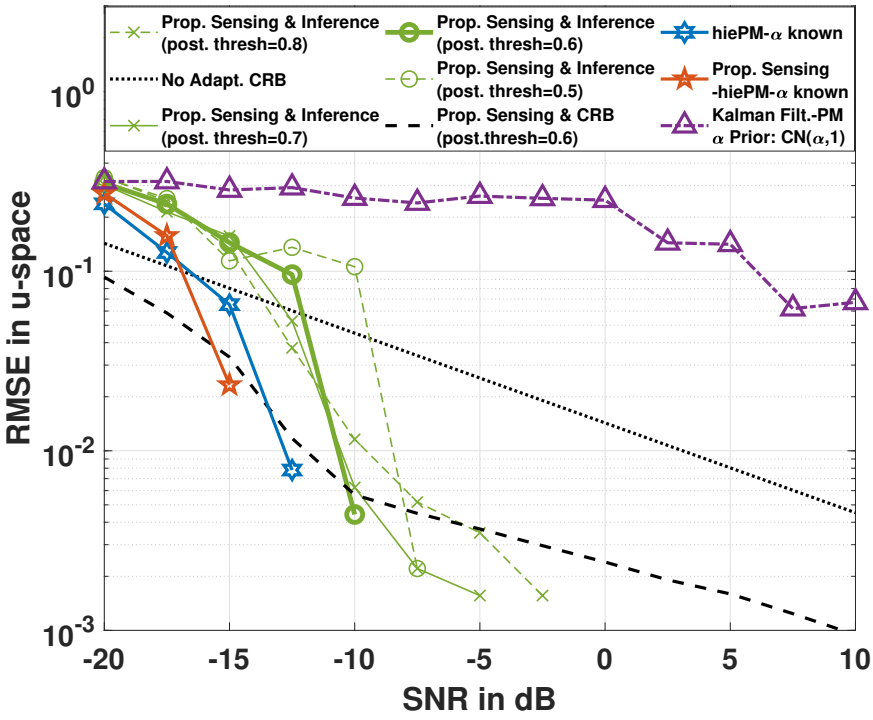
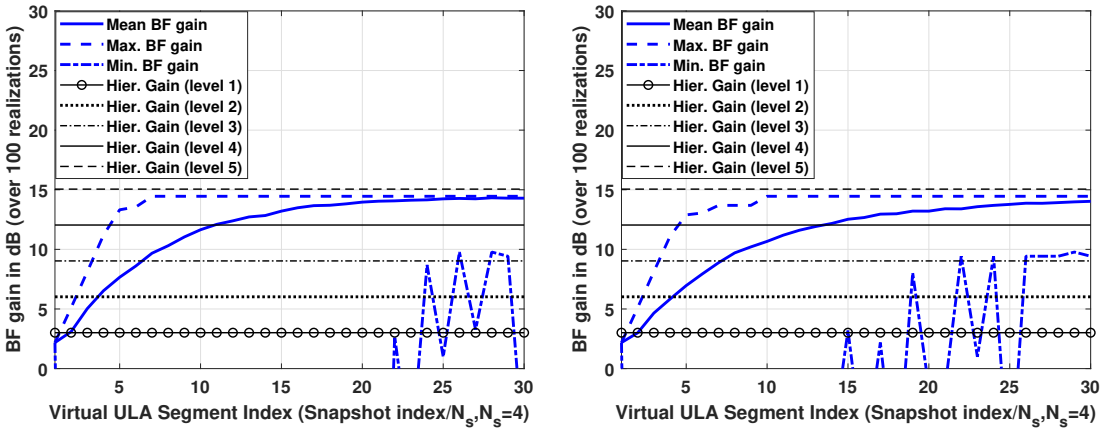


Figure 1.3. RMSE as a function of SNR. Performance of Algorithm 1 is observed to be close to the performance of algorithms using true α.

with hexagram markers, at SNR=$\{-15, -12.5\}$ dB. These two curves assume that α is known. At high SNR, the schemes incur zero error in this experiment as the DoA is on-grid. The green curves plots the performance of the proposed approach when α is unknown. The different curves correspond to different posterior threshold, p_{thresh}, levels for adapting the beamformer. At lower thresholds, the algorithm has the ability to adapt often. This may lead to slightly unstable performance, especially at low SNR, which is observed in the green-dashed curve with $\circ$ markers where $p_{\text{thresh}} = 0.5$ is used. Setting a high p_{thresh} requires a higher posterior mass to be accumulated for the beamformer to adapt. This adds more stability to the algorithm, but may incur slow beamformer adaptations, especially at high SNR. This is seen in green-dashed curve with x markers where $p_{\text{thresh}} = 0.8$ is set. Note that all the green curves are already close to the blue and red curves, without assuming any knowledge of α. The curve for the Algorithm 2 in [37] is plotted in purple with Δ markers. The algorithm is provided with the true value of α as mean along with variance set to 1 for the Gaussian α-prior. It is competitive at low SNR as it hinges on the already good prior information, whereas at high SNR it disregards the provided prior. Another important reason behind the large gap between this curve, and the blue curve with hexagram markers is the fact that the static α-case considered in [37] is not exploited[8]. The proposed beam alignment procedure exploits the same, and thus demonstrates a drastic improvement as the resulting posterior on angle is more accurate. We plot the (conditional) CRB on variance of angular estimation for two schemes i. using a non-adaptive SVAM beamformer with beam direction as 0.5 and beamwidth as 1 ii. using adaptive SVAM beamformers chosen during runtime of Algorithm 1 with $p_{\text{thresh}} = 0.6$. As seen in Fig. 1.3, the curves using proposed approach attains (& surpasses) the CRBs as SNR increases. Note that since the DoA is on grid which is also employed by the algorithms, they are biased. Thus, CRB is not a valid lower

[8]Posterior on α at current snapshot is used to compute likelihood of current measurement, but not utilized to update likelihoods of previous measurements.



(a) $p_{\text{thresh}} = 0.6$ (b) $p_{\text{thresh}} = 0.7$

Figure 1.4. Beamforming gain over time (SNR$=-10$ dB).

bound. Still it provides useful insight about the performance of the proposed inference procedure.

1.5.2 Beamforming gain over time using Algorithm 1

In Fig. 1.4, we study the beamforming gain experienced by the DoA using Algorithm 1 at SNR of -10 dB. We plot the mean, maximum and the minimum beamforming gain over time (in terms of the virtual ULA segment index t) over 100 random realizations. In Fig. 1.4 (a) and (b) we plot the results when p_{thresh} is set to 0.6 and 0.7, respectively. As expected, at a lower posterior threshold the algorithm is more flexible, and thus the maximum beamforming gain is achieved earlier than compared to the case when a higher posterior threshold is set. In contrast, the minimum beamforming gain may be less stable when the beamformer is adapted frequently at such low SNR. This is seen in plot (a) where the minimum gain fluctuates even after 25 virtual ULA segments i.e., $25 \times 4 = 100$ snapshots. The solid curve representing the mean beamforming gain indicates how the gain improves over time as the beamwidth adaptively narrows. We also plot the beamforming gain achieved by an ideal beamformer. A hierarchical level $l^h \in \{1, 2, \ldots, 5\}$ indicates a beam focused in $2/2^{l^h+1}$ of the spatial region and consequently achieves a beamforming

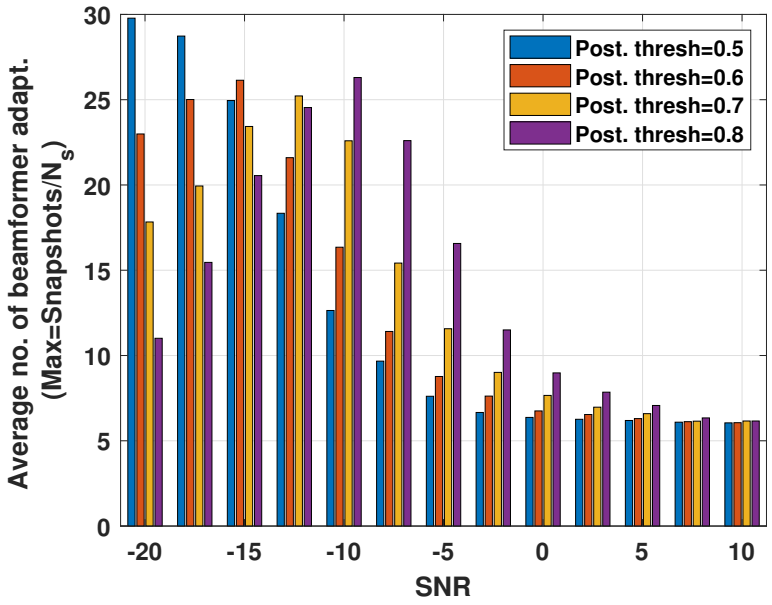


Figure 1.5. Number of Beamformer Adaptations as a function of SNR. Different bar plot curves for different Posterior Threshold indicates at lower thresholds and low SNR, the the beamformer adapts often. At high SNR, the BF adapts to the DoA quickly, indicated by low number of adaptions.

gain of $2^{l^h+1}/2$, in absolute scale. As observed in both the plots, beginning with a beamformer focusing on $\frac{1}{2}$ of the space (equivalently $10\log_{10}(2) \approx 3$ dB gain), the beamformer is able to adapt to $1/32$-fraction, garnering a beamforming gain of around $10\log_{10}(32) \approx 15$ dB. Thus, the proposed algorithm adds around 12 dB gain at -10 dB SNR.

1.5.3 Impact of Posterior Thresholds on the amount of Adaptation

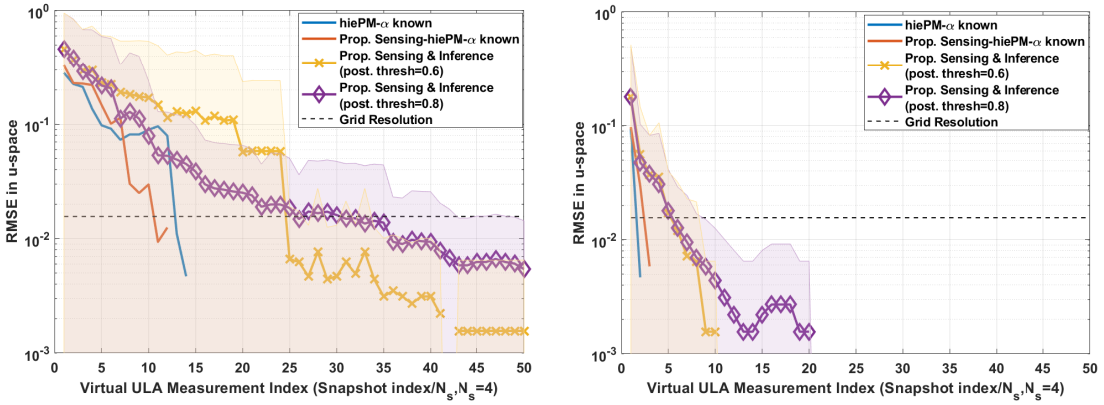
In Fig. 1.5, we study the average number of BF adaptations made as a function of SNR. We study using bar plots for the proposed algorithm under different settings of the posterior thresholds $= \{0.5, 0.6, 0.7, 0.8\}$. At high SNR, we expect less impact of the different settings as the posterior mass is expected to concentrate quickly around the true angle. This is seen in the Fig. 1.5, as all settings result in $6-7$ beam adaptations. At low SNR, setting a high threshold results in the algorithm using the widest beam for longer time. This is observed in the purple bars at low SNR, where the number of adaptations

is lower compared to other posterior threshold settings. It can be seen that the number of adaptations in this case (i.e., posterior threshold set to 0.8) initially increases as SNR increases, then peaks at SNR=-7.5 dB, and then drops as SNR increases further. This trend shifts to the right as the posterior threshold is reduced. This is expected, as the BF adapts more aggressively. With the posterior threshold set to 0.5, the BF is observed to adapt almost for every virtual ULA measurement i.e., $L/N_v = 120/4 = 30$ times at SNR= -20 dB.

1.5.4 Performance as a function of number of snapshots

In Fig. 1.6 we plot the RMSE over time (in terms of the number of virtual ULA measurements t). We plot the hiePM algorithms with α perfectly known for comparison. In Fig. 1.6 (a) and (b), we plot curves corresponding to SNR= -10 dB and 0 dB, respectively. We study two curves corresponding to the proposed scheme i. $p_{\text{thresh}} = 0.6$ (yellow curve with x markers) ii. $p_{\text{thresh}} = 0.8$ (purple curve with $\diamond$ markers). The former setting allows the beamformer to aggressively adapt, which can be rewarding at high SNR, but can lead to poor beamforming gain in the low SNR regime as it adapts to incorrect regions. This is seen in the two plots, at high SNR the curve corresponding to $p_{\text{thresh}} = 0.6$ converges quickly, whereas at low SNR it leads to larger error and variance initially, compared to setting $p_{\text{thresh}} = 0.8$. Note that even at low SNR and low threshold setting (Fig. 1.6 (a)), the yellow curve manages to adapt back to the correct spatial region, which is indicated by a drop in RMSE around $t = 25$ and is able to retain the improvement over time. Overall, it is observed that a high p_{thresh} leads to stable yet slow convergence, which is useful when the SNR is low and if the training duration is short. At high SNR or if the training duration is longer, a low p_{thresh} pays off, as the algorithm is robust to recover from its mistakes!

One may dynamically select p_{thresh} such that it is set to higher value initially during the training phase, but it may be reduced over time as the beamformer narrows further



(a) SNR $= -10$ dB (a) SNR $= 0$ dB

Figure 1.6. RMSE over time. Shaded region covers the RMSE $\pm$ root standard deviation of squared error over time at every time index along x-axis.

and enjoys a high beamforming gain. This is left for future work.

1.5.5 Studying Impact of Noise Variance Parameter

We plot the RMSE as a function of different settings for the noise variance parameter in Fig. 1.7. The optimal value is observed to be around the true noise variance value. We also plot curves corresponding to different posterior thresholds. For a low posterior threshold, the optimal value of the noise variance parameter is observed to be slightly higher ($\times 2$) than or equal to the true value. This is due to the aggressive nature of the algorithm to adapt, which can be compensated by setting a slightly higher noise variance parameter. On the other hand, a high posterior threshold already imparts more stable adaptations, and consequently does not need setting a higher noise variance parameter. In fact, at such large thresholds, the algorithm benefits from lower noise variance parameter setting, which intuitively offsets the conservative posterior threshold setting. This is observed by the curve corresponding to $p_{\text{thresh}} = 0.9$, where setting noise variance to $0.5 \times \sigma_n^2$ leads to much better performance than setting it to the true value.

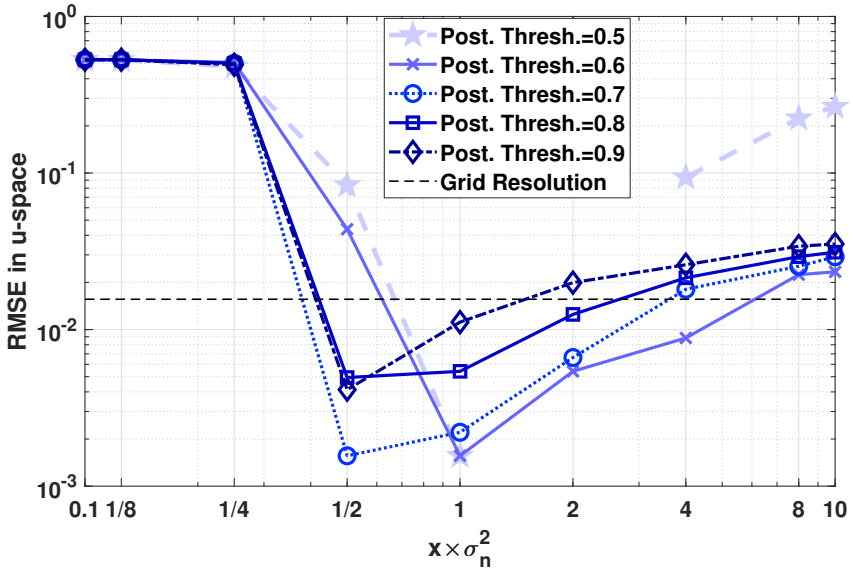


Figure 1.7. RMSE vs. $x \times \sigma_n^2, x \in \{0.1, 1/8, 1/4, 1/2, 1, 2, 4, 8, 10\}$. SNR$= -10$ dB, $L = 200$. As the posterior threshold increases, the optimal value for the noise variance parameter is observed to be lower than the true value.

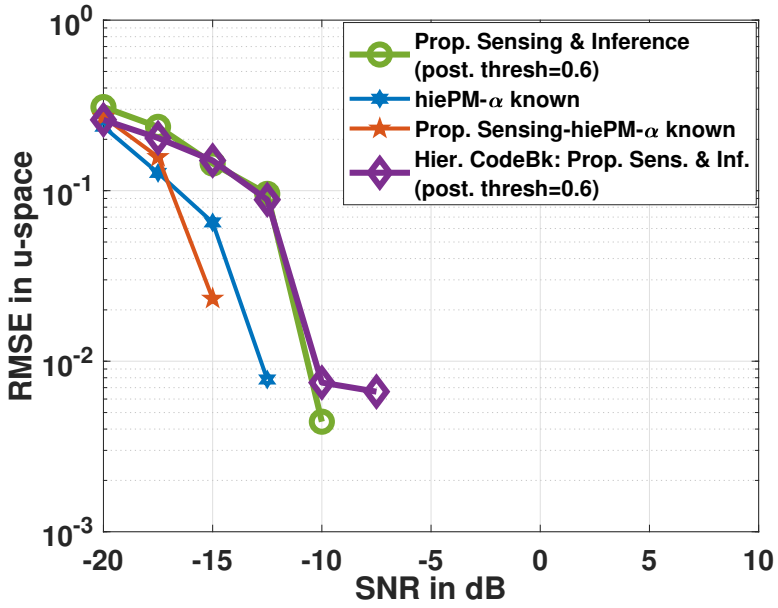


Figure 1.8. RMSE as a function of SNR. A hierarchical codebook is used, which results in little to no performance loss, indicating that the overall performance gain is unaffected by the compact hierarchical codebook.

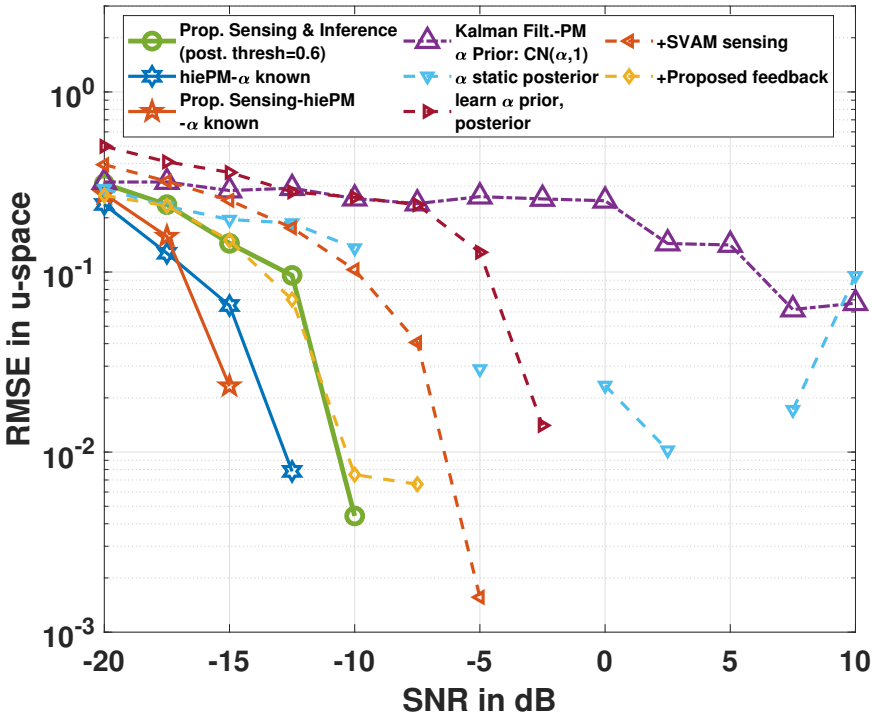


Figure 1.9. Studying the impact of each component in Proposed Initial Alignment algorithm

1.5.6 Impact of compact hierarchical codebook (Algorithm 3)

In Fig. 1.8, we plot the curve corresponding to using a hierarchical codebook (see purple curve with $\diamond$ markers). The approach is described in Algorithm 3. We set $p_{\text{thresh}} = 0.6$ and compare the performance with using the flexible codebook (see green curve with $\circ$ markers). As observed in Fig. 1.8, the two implementations have very similar performance. This suggests that the superior performance is due to other ingredients, namely i. improved sensing (and filter design), ii. good posterior estimate, and iii. good adaptive strategy. A large codebook size helps, but is not the key to the performance improvement demonstrated in this paper.

1.5.7 From Kalman Filtering-based HiePM algorithm to Proposed SVAM sensing and Inference: Step-By-Step Analysis

In Fig. 1.9, we study the impact of each component within proposed algorithm beginning from the purple curve (with triangle pointed up) corresponding to Kalman-Filtering-based HiePM algorithm, through a step-by-step modification. First, a static alpha over time is imposed in the reference algorithm, and it uses the common statistics to compute angular posterior. This leads to the light-blue curve (with triangle pointed down). This marks significant improvement overall. Then, with the aim to relax the known alpha prior mean assumption, the prior and posterior on alpha is learned from measurements. This leads to slight degradation of performance at low SNR but improvement at high SNR (as expected), as shown in the brown curve (with triangle pointed right). SVAM sensing is then included, which again leads to significant performance improvement, as seen in red curve (with triangle pointed left). Finally, hiePM's feedback is replaced with proposed feedback including filter design with Parks-McClellan algorithm. This leads to the best performance presented in this work, as shown in the yellow curve (with diamond markers). It was observed that if the filter design is replaced with LS-based filter design in [36], the performance is bad. This highlights the importance of having a linear phase filter as used throughout in this work.

1.6 Conclusion

We proposed a novel Synthesis of Virtual Array Manifold (SVAM) sensing approach for the mmWave single RF chain systems and discussed the ensuing benefits. More specifically, the proposed sensing is demonstrated to lead to faster and more robust beam alignment. We believe this contribution will have significant impact on the traditional paradigm for sensing in mmWave systems. We also proposed a novel inference scheme that estimates a posterior density on the small-scale fading coefficient and the unknown

dominant path angle. Based on the proposed inference procedure, an adaptive beamforming scheme is provided that aims to collect high SNR measurements. Finally, the performance of the proposed active sensing scheme is evaluated under different scenarios, and a significant improvement over various benchmarks is demonstrated. The empirical study also reveals the impact of the different design parameters on the beam alignment performance.

Chapter 1, in part is currently submitted for publication of the material, and in part, is a reprint of the material as it appears in R. R. Pote and B. D. Rao, "Novel Sensing Methodology for Initial Alignment Using MmWave Phased Arrays," 2023 57th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, 2023. The dissertation author was the primary investigator and author of this material.

1.7 Appendix

1.7.1 Proof of Theorem 1

Proof. Given that $\boldsymbol{\alpha}$ is known, the likelihood of the measurements $\mathbf{y} = [y_0, \ldots, y_{L-1}]^T$ as a function of u is given by

$$f(\mathbf{y}; u) = \frac{1}{\pi^L \det(\sigma_n^2 \mathbf{I})} \exp - \frac{(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}}(u))^H (\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}}(u))}{\sigma_n^2}, \tag{1.35}$$

where $\boldsymbol{\mu}_{\mathbf{y}}(u) = \sqrt{P_s}\alpha\mathbf{W}^H\boldsymbol{\phi}_N(u)$. We can compute the required second derivative w.r.t. u of the negative log-likelihood and the expectation as

$$\begin{aligned}
-\frac{\partial^2}{\partial u^2}\ln f(\mathbf{y};u) &= \frac{P_s|\alpha|^2}{\sigma_n^2}2\mathrm{Re}\left\{\left(\frac{\partial^2}{\partial u^2}\boldsymbol{\phi}_N(u)\right)^H\mathbf{W}\mathbf{W}^H\boldsymbol{\phi}_N(u)\right\} \\
&\quad + \frac{P_s|\alpha|^2}{\sigma_n^2}2\left(\frac{\partial}{\partial u}\boldsymbol{\phi}_N(u)\right)^H\mathbf{W}\mathbf{W}^H\frac{\partial}{\partial u}\boldsymbol{\phi}_N(u) \\
&\quad - \frac{\sqrt{P_s}\alpha}{\sigma_n^2}2\mathrm{Re}\left\{\mathbf{y}^H\mathbf{W}^H\frac{\partial^2}{\partial u^2}\boldsymbol{\phi}_N(u)\right\} \qquad (1.36) \\
E\left[-\frac{\partial^2}{\partial u^2}\ln f(\mathbf{y};u)\right] &= \frac{P_s|\alpha|^2}{\sigma_n^2}2\left(\frac{\partial}{\partial u}\boldsymbol{\phi}_N(u)\right)^H\mathbf{W}\mathbf{W}^H\frac{\partial}{\partial u}\boldsymbol{\phi}_N(u).
\end{aligned}$$

The Cramér-Rao lower bound is given by, $\mathrm{CRB}(u) = E\left[-\frac{\partial^2}{\partial u^2}\ln f(\mathbf{y};u)\right]^{-1}$. This completes the proof. □

1.7.2 Proof of Corollary 1.1

Proof. We need to simplify the quadratic term in the $\mathrm{CRB}(u)$ expression from Theorem 1 using the property of the beamformer in (1.14).

$$\begin{aligned}
\left(\mathbf{W}^{\mathrm{B}}\right)^H\frac{\partial}{\partial u}\boldsymbol{\phi}_N(u) &= \left(\left(\mathbf{F}^{\mathrm{B}}\right)^H\otimes\mathbf{1}_{N_v}\right)\frac{\partial}{\partial u}\boldsymbol{\phi}_N(u) \\
&= \left(\left(\mathbf{F}^{\mathrm{B}}\right)^H\frac{\partial}{\partial u}\boldsymbol{\phi}_N(u)\right)\otimes\mathbf{1}_{N_v}. \qquad (1.37)
\end{aligned}$$

And,

$$\begin{aligned}
\left(\frac{\partial}{\partial u}\boldsymbol{\phi}_N(u)\right)^H\mathbf{W}^{\mathrm{B}}\left(\mathbf{W}^{\mathrm{B}}\right)^H\frac{\partial}{\partial u}\boldsymbol{\phi}_N(u) &= \left(\left(\left(\mathbf{F}^{\mathrm{B}}\right)^H\frac{\partial}{\partial u}\boldsymbol{\phi}_N(u)\right)\otimes\mathbf{1}_{N_v}\right)^H \\
&\quad\times\left(\left(\left(\mathbf{F}^{\mathrm{B}}\right)^H\frac{\partial}{\partial u}\boldsymbol{\phi}_N(u)\right)\otimes\mathbf{1}_{N_v}\right) \\
&= N_v\left(\frac{\partial}{\partial u}\boldsymbol{\phi}_N(u)\right)^H\mathbf{F}^{\mathrm{B}}(\mathbf{F}^{\mathrm{B}})^H\frac{\partial}{\partial u}\boldsymbol{\phi}_N(u). \qquad (1.38)
\end{aligned}$$

This completes the proof. □

1.7.3 Proof of Corollary 2.1

Proof. Using Lemma 2, we can express

$$\begin{aligned}\left(\mathbf{W}^{\mathrm{P}}\right)^{H}\frac{\partial}{\partial u}\boldsymbol{\phi}_{N}(u) &= \left(\left(\mathbf{F}^{\mathrm{P}}\right)^{H}\otimes\boldsymbol{\phi}_{N_v}(u)\right)\frac{\partial}{\partial u}\boldsymbol{\phi}_{M}(u)+\left(\left(\mathbf{F}^{\mathrm{P}}\right)^{H}\otimes\frac{\partial}{\partial u}\boldsymbol{\phi}_{N_v}(u)\right)\boldsymbol{\phi}_{M}(u)\\ &\overset{(a)}{=}\left(\left(\mathbf{F}^{\mathrm{P}}\right)^{H}\frac{\partial}{\partial u}\boldsymbol{\phi}_{M}(u)\right)\otimes\boldsymbol{\phi}_{N_v}(u)+\left(\left(\mathbf{F}^{\mathrm{P}}\right)^{H}\boldsymbol{\phi}_{M}(u)\right)\otimes\frac{\partial}{\partial u}\boldsymbol{\phi}_{N_v}(u). \quad (1.39)\end{aligned}$$

where in (a) we use the following result on Kronecker product: $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D})=\mathbf{AC}\otimes\mathbf{BD}$, provided the matrix products $\mathbf{AC}$ and $\mathbf{BD}$ are conformable. Then, the required quadratic term can be simplified as

$$\begin{aligned}&\left(\frac{\partial}{\partial u}\boldsymbol{\phi}_{N}(u)\right)^{H}\mathbf{W}^{\mathrm{P}}\left(\mathbf{W}^{\mathrm{P}}\right)^{H}\frac{\partial}{\partial u}\boldsymbol{\phi}_{N}(u)\\ &\quad\overset{(a)}{=}N_v\left(\frac{\partial}{\partial u}\boldsymbol{\phi}_{M}(u)\right)^{H}\mathbf{F}^{\mathrm{P}}\left(\mathbf{F}^{\mathrm{P}}\right)^{H}\frac{\partial}{\partial u}\boldsymbol{\phi}_{M}(u)\\ &\qquad+\frac{\pi^2(N_v-1)N_v(2N_v-1)}{6}\boldsymbol{\phi}_{M}^{H}(u)\mathbf{F}^{\mathrm{P}}\left(\mathbf{F}^{\mathrm{P}}\right)^{H}\boldsymbol{\phi}_{M}(u)\\ &\qquad-\pi(N_v-1)N_v\mathrm{Im}\left\{\left(\frac{\partial}{\partial u}\boldsymbol{\phi}_{M}(u)\right)^{H}\mathbf{F}^{\mathrm{P}}\left(\mathbf{F}^{\mathrm{P}}\right)^{H}\boldsymbol{\phi}_{M}(u)\right\} \quad (1.40)\end{aligned}$$

□

1.7.4 Proof of Theorem 3

Proof. We first note that the following holds

$$\frac{\partial}{\partial u}\boldsymbol{\phi}_{M}(u)=j\pi\left\{\frac{(M-1)}{2}\boldsymbol{\phi}_{m}(u)+\boldsymbol{\phi}_{M}^{\perp}(u)\right\}, \quad (1.41)$$

where $\boldsymbol{\phi}_M^{\perp}(u) = \begin{bmatrix} -\frac{(M-1)}{2} & (1-\frac{(M-1)}{2}) & \dots & \frac{(M-1)}{2} \end{bmatrix}^T \odot \boldsymbol{\phi}_M(u)$, '$\odot$' denotes Hadamard product. Using the result in (1.41) we simplify the second term in G as

$$\begin{aligned}
&-\pi(N_v-1)\mathrm{Im}\left\{\left(\frac{\partial}{\partial u}\boldsymbol{\phi}_M(u)\right)^H \mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H \boldsymbol{\phi}_M(u)\right\} \\
&= \pi^2(N_v-1)\mathrm{Re}\left\{\left(\frac{(M-1)}{2}\boldsymbol{\phi}_M^H(u) + \left(\boldsymbol{\phi}_M^{\perp}(u)\right)^H\right)\mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H\boldsymbol{\phi}_M(u)\right\} \\
&= \pi^2(N_v-1)\left\{\frac{(M-1)}{2}\boldsymbol{\phi}_M^H(u)\mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H\boldsymbol{\phi}_M(u) + \mathrm{Re}\left\{\left(\left(\boldsymbol{\phi}_M^{\perp}(u)\right)^H\right)\mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H\boldsymbol{\phi}_M(u)\right\}\right\} \\
&\geq \pi^2(N_v-1)\left\{\frac{(M-1)}{2}\boldsymbol{\phi}_M^H(u)\mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H\boldsymbol{\phi}_M(u) - \left|\boldsymbol{\phi}_M^H(u)\mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H\boldsymbol{\phi}_M^{\perp}(u)\right|\right\} \\
&= \pi^2(N_v-1)\boldsymbol{\phi}_M^H(u)\mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H\boldsymbol{\phi}_M(u)\left\{\frac{(M-1)}{2} - \frac{\left|\boldsymbol{\phi}_M^H(u)\mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H\boldsymbol{\phi}_M^{\perp}(u)\right|}{\boldsymbol{\phi}_M^H(u)\mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H\boldsymbol{\phi}_M(u)}\right\}. \qquad (1.42)
\end{aligned}$$

Thus to ensure

$$\begin{aligned}
& G \geq 0 \\
\Longrightarrow\ & c\left\{\frac{(2N_v-1)}{6} + \frac{(M-1)}{2} - \frac{\left|\boldsymbol{\phi}_M^H(u)\mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H\boldsymbol{\phi}_M^{\perp}(u)\right|}{\boldsymbol{\phi}_M^H(u)\mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H\boldsymbol{\phi}_M(u)}\right\} \geq 0 \qquad (1.43)
\end{aligned}$$

where $c = \pi^2(N_v-1)\boldsymbol{\phi}_M^H(u)\mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H\boldsymbol{\phi}_M(u)$. The above implies

$$\begin{aligned}
\frac{\left|\boldsymbol{\phi}_M^H(u)\mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H\boldsymbol{\phi}_M^{\perp}(u)\right|}{\boldsymbol{\phi}_M^H(u)\mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H\boldsymbol{\phi}_M(u)} &\leq \left\{\frac{(3M+2N_v-4)}{6}\right\} \\
\frac{\left|\boldsymbol{\phi}_M^H(u)\mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H\boldsymbol{\phi}_M^{\perp}(u)\frac{\|\boldsymbol{\phi}_M(u)\|}{\|\boldsymbol{\phi}_M^{\perp}(u)\|}\right|}{\boldsymbol{\phi}_M^H(u)\mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H\boldsymbol{\phi}_M(u)} &\leq C(N,N_v), \qquad (1.44)
\end{aligned}$$

where $C(N,N_v) = \frac{\|\boldsymbol{\phi}_M(u)\|}{\|\boldsymbol{\phi}_M^{\perp}(u)\|}\left\{\frac{(3M+2N_v-4)}{6}\right\}$, $\|\boldsymbol{\phi}_M(u)\| = \sqrt{M}$, $\|\boldsymbol{\phi}_M^{\perp}(u)\| = \frac{\sqrt{M(M^2-1)}}{2\sqrt{3}}$. Note that $C(N,N_v) \geq \sqrt{3}$ when $N_v > 1$. Let $\boldsymbol{\phi}_{M,\mathrm{n}}^{\perp}(u) = \boldsymbol{\phi}_M^{\perp}(u)\frac{\|\boldsymbol{\phi}_M(u)\|}{\|\boldsymbol{\phi}_M^{\perp}(u)\|}$. The LHS can be further sim-

plified to get

$$\frac{\left|\boldsymbol{\phi}_M^H(u)\mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H\boldsymbol{\phi}_M^{\perp}(u)\frac{\|\boldsymbol{\phi}_M(u)\|}{\|\boldsymbol{\phi}_M^{\perp}(u)\|}\right|}{\boldsymbol{\phi}_M^H(u)\mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H\boldsymbol{\phi}_M(u)} = \frac{\sqrt{\boldsymbol{\phi}_M^H(u)\mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H\boldsymbol{\phi}_{M,\mathrm{n}}^{\perp}(u)(\boldsymbol{\phi}_{M,\mathrm{n}}^{\perp}(u))^H\mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H\boldsymbol{\phi}_M(u)}}{\boldsymbol{\phi}_M^H(u)\mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H\boldsymbol{\phi}_M(u)}$$

$$\overset{(a)}{=} \sqrt{\frac{\|\boldsymbol{\phi}_M(u)\|^2\boldsymbol{\phi}_M^H(u)\mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H\mathbf{P}_{u,\perp}\mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H\boldsymbol{\phi}_M(u)}{(\boldsymbol{\phi}_M^H(u)\mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H\boldsymbol{\phi}_M(u))^2} - 1}$$

$$\overset{(b)}{\leq} \sqrt{\frac{\|\boldsymbol{\phi}_M(u)\|^2\lambda_{\max}\left((\mathbf{F}^{\mathrm{P}})^H\mathbf{P}_{u,\perp}\mathbf{F}^{\mathrm{P}}\right)}{\boldsymbol{\phi}_M^H(u)\mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H\boldsymbol{\phi}_M(u)} - 1}. \tag{1.45}$$

where in (a) we express $\frac{\boldsymbol{\phi}_{M,\mathrm{n}}^{\perp}(u)(\boldsymbol{\phi}_{M,\mathrm{n}}^{\perp}(u))^H}{\|\boldsymbol{\phi}_{M,\mathrm{n}}^{\perp}(u)\|^2} = \mathbf{P}_{u,\perp} - \frac{\boldsymbol{\phi}_M(u)\boldsymbol{\phi}_M(u)^H}{\|\boldsymbol{\phi}_M(u)\|^2}$. $\mathbf{P}_{u,\perp} = [\boldsymbol{\phi}_M(u)\ \boldsymbol{\phi}_M^{\perp}(u)]$ $\times \begin{bmatrix} \|\boldsymbol{\phi}_M(u)\|^2 & 0 \\ 0 & \|\boldsymbol{\phi}_M^{\perp}(u)\|^2 \end{bmatrix}^{-1} [\boldsymbol{\phi}_M(u)\ \boldsymbol{\phi}_M^{\perp}(u)]$. In (b) we upper bound the quadratic term using the largest eigenvalue, $\lambda_{\max}\left((\mathbf{F}^{\mathrm{P}})^H\mathbf{P}_{u,\perp}\mathbf{F}^{\mathrm{P}}\right)$, of the Hermitian-symmetric matrix $(\mathbf{F}^{\mathrm{P}})^H\mathbf{P}_{u,\perp}\mathbf{F}^{\mathrm{P}}$. Thus a more stricter condition that satisfies the inequality in (1.44) is given by

$$\sqrt{\frac{\|\boldsymbol{\phi}_M(u)\|^2\lambda_{\max}\left((\mathbf{F}^{\mathrm{P}})^H\mathbf{P}_{u,\perp}\mathbf{F}^{\mathrm{P}}\right)}{\boldsymbol{\phi}_M^H(u)\mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H\boldsymbol{\phi}_M(u)} - 1} \leq C(N,N_s) \tag{1.46}$$

which implies

$$\frac{\boldsymbol{\phi}_M^H(u)\mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H\boldsymbol{\phi}_M(u)}{\|\boldsymbol{\phi}_M(u)\|^2} \geq \frac{\lambda_{\max}\left((\mathbf{F}^{\mathrm{P}})^H\mathbf{P}_{u,\perp}\mathbf{F}^{\mathrm{P}}\right)}{C^2(N,N_s)+1} \quad . \tag{1.47}$$

Since $C^2(N,N_s) \geq 3$, an even stricter but simplified condition than (1.47) is

$$\frac{\boldsymbol{\phi}_M^H(u)\mathbf{F}^{\mathrm{P}}(\mathbf{F}^{\mathrm{P}})^H\boldsymbol{\phi}_M(u)}{\|\boldsymbol{\phi}_M(u)\|^2} \geq \frac{\lambda_{\max}\left((\mathbf{F}^{\mathrm{P}})^H\mathbf{P}_{u,\perp}\mathbf{F}^{\mathrm{P}}\right)}{4}. \tag{1.48}$$

This concludes the proof.

□

Chapter 2

Light Weight Sequential SBL: An Alternative to OMP

2.1 Introduction

Sparse signal recovery (SSR) has been recognized as an important signal processing problem with numerous applications. Applications include biomedical imaging [4, 65, 66], functional approximation [5], echo cancellation [6] and wireless communications [67–69]. The underlying problem in SSR can be stated as follows: Given a measurement matrix $\mathbf{\Phi} \in \mathbb{C}^{m \times n} (m < n)$ and measurement vector $\mathbf{y} \in \mathbb{C}^m$ such that

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{n}, \tag{2.1}$$

may be corrupted by noise vector $\mathbf{n} \in \mathbb{C}^m$, the goal is to recover the vector $\mathbf{x} \in \mathbb{C}^n$ which is known to be sparse i.e., $\|\mathbf{x}\|_0 \ll n$. The noise, $\mathbf{n}$, is distributed as $\mathscr{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I})$. $\mathbf{x}$ and $\mathbf{n}$ are independent of each other.

Realizing the vibrant need for approaches, many algorithms have been proposed that offer favourable tradeoffs between recovery performance, speed, and storage [7–14]. If the support size of $\mathbf{x}$ is known i.e., $|\mathbf{x}\|_0 = k$, where $\|\cdot\|_0$ indicates the ℓ_0 quasi-norm,

then the problem in (2.1) may be written as

$$\arg\min_{\mathbf{x}\in\mathbb{C}^n} \|\mathbf{y}-\mathbf{\Phi}\mathbf{x}\|_2 \quad \text{s.t. } \|\mathbf{x}\|_0 = k. \tag{2.2}$$

The problem in (2.2) is non-convex and involves a combinatorial search over the columns of $\mathbf{\Phi}$. Furthermore, in many problems of interest the support size is not explicitly provided, and it must be estimated along with the non-zero entries in $\mathbf{x}$. The approach in (2.2) can be relaxed to the more amenable convex formulations such as LASSO, Basis Pursuit, etc. Bayesian methods have also been investigated in the past, and they provide an alternative perspective wherein the presumed deterministic nature of $\mathbf{x}$ is replaced with a stochastic modeling. Such methods may lead to non-convex formulation. The use of Laplacian prior is a well-known Bayesian method for enforcing sparsity. Another sub-class within Bayesian methods is the Hierarchical Bayes framework. This will be subject of enquiry in this work. In brief, the framework involves introducing hyperparameters that shape the prior and a search for the right (empirical) prior in the hyperparameter space. The hierarchical Bayes framework will be discussed more in the next section.

On the computationally favourable paradigm, greedy algorithms (e.g. pursuit algorithms like matching pursuit (MP) [7], orthogonal MP (OMP) [15–18], compressive sampling MP (CoSaMP) [19] etc.) are well-studied algorithms that offer faster recovery at the cost of slight degradation in performance. OMP is a widely employed [20, 21] iterative technique in this category that recovers $\mathbf{x}$ deterministically. Each iteration freezes the past recovered support and optimizes (2.2) with respect to one additional column. The non-zero entries in $\mathbf{x}$ of the past recovered support is also fixed during this step. OMP improves over MP, in that in the second step, it ensures that the residual error is orthogonal to all the selected columns of $\mathbf{\Phi}$.

In this work, an alternative to OMP is provided wherein $\mathbf{x}$ is modeled as a stochastic variable. Like OMP, the proposed algorithm is a greedy iterative algorithm. The

contributions made in this work are identified below:

- Light weight sequential sparse Bayesian learning (LWS-SBL) algorithm is proposed as an alternative to OMP. In contrast to OMP, it is derived from the Type-II estimation framework and the stochastic maximum likelihood objective
- Computationally efficient steps based on recursive implementations are developed to maintain a complexity similar to that of OMP
- Two perspectives are provided to comprehend LWS-SBL in comparison to OMP. These perspectives highlight key differences between the two algorithms. The analysis also highlights avenues where LWS-SBL improves over the OMP algorithm; more theoretical justifications in this regard are provided
- For parametric dictionaries, the underlying goal of estimating parameter in a gridless manner is identified and a gridless extension of proposed LWS-SBL algorithm is provided. In comparison to the work in [64], the computational step uses Newton updates and thus is more economical
- Flexibility of the proposed approach to the multi-dimensional harmonic retrieval problem is demonstrated by describing the corresponding LWS-SBL extension and its gridless counterpart

The ideas presented in this work are empirically analyzed in the Section 2.7.

2.1.1 Organization of the Chapter and Notations

In Section 2.2 the proposed low complexity sparse signal recovery algorithm is developed and its computational complexity is compared with OMP. Two novel perspectives to understand the operating mechanisms for the proposed algorithm is discussed under Section 2.3 and Section 2.4 by focusing on global and local characteristics, respectively, of the two algorithms. Extension of proposed algorithm to single dimension parametric

problems for gridless sparse signal recovery is presented in Section 2.5. More extensions to two dimensional harmonic retrieval problem is discussed in Section 2.6. Numerical results for the algorithms proposed in this work are presented in Section 2.7. Section 2.8 concludes this work.

Notations: We represent scalars, vectors, and matrices by lowercase, boldface-lowercase, and boldface-uppercase letters, respectively. Sets are represented using blackboard bold letters. $(.)^T, (.)^H, (.)^c$ denotes transpose, Hermitian, and complex conjugate operation respectively. $\otimes$ denotes matrix Kronecker product, and $\odot$ denotes Hadamard product of two conformable matrices. $*$ denotes convolution operation. $[M] = \{0, 1, \ldots, M-1\}, M \in \mathbb{Z}^+$.

2.2 Proposed Light-Weight Sequential SBL Algorithm

In this work, we focus on iterative approach to find sparse solutions to (2.1) while maintaining a low computational complexity. Within an iterative scheme the goal is to identify the support over iterations. Greed-based iterative approaches optimize a certain objective function, $\mathscr{L}(.)$ at every iteration with respect to candidate support elements. We focus on the greedy-iterative paradigm and develop algorithms that add one column of $\mathbf{\Phi}$ at every iteration. As a benchmark we consider the widely used OMP algorithm which sequentially solves a least squares objective. Next, we discuss the SBL formulation which motivates the objective function adopted in the proposed algorithm. Without loss of generality, we assume that the columns of $\mathbf{\Phi}$ are normalized throughout in this paper i.e., $\|\boldsymbol{\phi}_j\|_2 = 1$, where $\boldsymbol{\phi}_j$ denotes the j-th column of $\mathbf{\Phi}$.

2.2.1 On Sparse Bayesian Learning

SBL is a Bayesian technique to compute a sparse decomposition of measurements $\mathbf{y}$. SBL models $\mathbf{x}$ as a random variable, in contrast to the OMP algorithm [14, 15] where $\mathbf{x}$ is modeled as a deterministic unknown. More specifically, a parameterized Gaussian prior is

imposed on $\mathbf{x}$ with mean zero and uncorrelated components. In other words, $\mathbf{x} \sim \mathscr{CN}(\mathbf{0}, \mathbf{\Gamma})$ where $\mathbf{\Gamma}$ is a diagonal matrix; let $\mathbf{diag}(\mathbf{\Gamma}) = \boldsymbol{\gamma}$. Under the zero mean Gaussian noise assumption, $\mathbf{y} \sim \mathscr{CN}(\mathbf{0}, \mathbf{\Phi\Gamma\Phi}^H + \lambda\mathbf{I})$, where λ denotes the noise variance estimate. The SBL approach is to solve (2.1) in the hyperparameter $\boldsymbol{\gamma}$-space, post marginalization with respect to $\mathbf{x}$. This approach is known as Type-II estimation framework, compared to Type-I where the problem is solved in $\mathbf{x}$ space after marginalization with respect to $\boldsymbol{\gamma}$ [30] (e.g., Lasso algorithm [31]). Note that the Gaussian density imposition on $\mathbf{x}$ is not a limitation, and the method generalizes well to the case when $\mathbf{x}$ was in fact drawn from a non-Gaussian density (e.g., see Section VI.A. in [12]). The hyperparameter $\boldsymbol{\gamma}$ is estimated by minimizing the negative marginal log-likelihood function [12, 26]

$$\min_{\boldsymbol{\gamma} \geq \mathbf{0}, \lambda \geq 0} \mathscr{L}(\boldsymbol{\gamma}, \lambda) := \log\det \mathbf{\Sigma}_\mathbf{y} + \mathbf{y}^H \mathbf{\Sigma}_\mathbf{y}^{-1} \mathbf{y}. \tag{2.3}$$

where, $\mathbf{\Sigma}_\mathbf{y} = \mathbf{\Phi\Gamma\Phi}^H + \lambda\mathbf{I}$. (2.3) is a non-convex problem in $(\boldsymbol{\gamma}, \lambda)$. In this work, we initialize with $\gamma_j = 0, \forall j \in \{1, \ldots, n\}$, and add a column l that minimizes the negative log-likelihood, the most. Thus, similar to OMP, the proposed algorithm [13] selects one column per iteration. However, in contrast to OMP, each iteration in the proposed scheme optimizes the maximum likelihood based cost function in (2.3). The following remark is in order.

Remark 9. The sequential SBL approach in [13] is a framework and does not provide a specific algorithm, but rather suggestions (see Section 4 in [13]) for developing variants with different options, e.g. adding/deleting selected columns, modifying the variances of the columns already selected, etc. To ensure a computational complexity comparable to OMP, we develop a specific algorithm, which like OMP, runs the sequential steps for K iterations only; K denotes the desired support size.

Next we derive steps to iterative optimize (2.3).

2.2.2 Proposed Light Weight Sequential-SBL Algorithm

We focus on estimating $\boldsymbol{\gamma}$ for a fixed $\boldsymbol{\lambda}$. The latter may be estimated [13] but is not discussed in this paper. We begin by separating out the contribution of the j-th column to the cost function, $\mathscr{L}(\boldsymbol{\gamma})$, in (2.3). Let $\mathbb{T} \subset \{1,\ldots,n\}$ denote the set of column indices of $\boldsymbol{\Phi}$ already selected and $\mathbf{C} = \boldsymbol{\Phi}_{\mathbb{T}}\boldsymbol{\Gamma}_{\mathbb{T}}\boldsymbol{\Phi}_{\mathbb{T}}^H + \lambda\mathbf{I}$, where $\boldsymbol{\Gamma}_{\mathbb{T}}$ denotes the diagonal matrix with rows and columns in $\mathbb{T}$. For $j \notin \mathbb{T}$, we can write

$$\mathscr{L}(\boldsymbol{\gamma}_{\mathbb{T}\cup\{j\}}) = \mathscr{L}(\boldsymbol{\gamma}_{\mathbb{T}}) + L(\gamma_j, \mathbf{C}), \tag{2.4}$$

where $L(\gamma_j, \mathbf{C}) = \log(1+\gamma_j\boldsymbol{\Phi}_j^H\mathbf{C}^{-1}\boldsymbol{\Phi}_j) - |\boldsymbol{\Phi}_j^H\mathbf{C}^{-1}\mathbf{y}|^2/(\gamma_j^{-1}+\boldsymbol{\Phi}_j^H\mathbf{C}^{-1}\boldsymbol{\Phi}_j)$ (see eq. (18) in [13] for detailed derivation of (2.4)). Let us introduce the following quantities for ease of presentation:

$$s_j = \boldsymbol{\Phi}_j^H\mathbf{C}^{-1}\boldsymbol{\Phi}_j \text{ and } q_j = \boldsymbol{\Phi}_j^H\mathbf{C}^{-1}\mathbf{y}. \tag{2.5}$$

Since γ_j is initialized with zero i.e., $\gamma_{\text{prev},j} = 0$, we have $\forall j \notin \mathbb{T}$

$$\Delta L_{\mathbf{C}}(\gamma_j, \gamma_{\text{prev},j}) = L(\gamma_j, \mathbf{C}) - L(\gamma_{\text{prev},j}, \mathbf{C}) = L(\gamma_j, \mathbf{C}). \tag{2.6}$$

In other words, the change in negative log-likelihood due to updating γ_j is simply its contribution to the cost function. This helps to simplify the objective further. The column to be added is given by

$$l = \underset{j\notin\mathbb{T}}{\arg\min} \min_{\gamma_j \geq 0} L(\gamma_j, \mathbf{C}) \left(= \Delta L_{\mathbf{C}}(\gamma_j, \gamma_{\text{prev},j})\right). \tag{2.7}$$

Minimization with respect to γ_j can be obtained in closed-form as

$$\gamma_j^{\text{opt}} = \max\left\{\frac{|q_j|^2 - s_j}{s_j^2}, 0\right\} \tag{2.8}$$

At the optimal value γ_j^{opt} we have

$$L(\gamma_j^{\text{opt}}, \mathbf{C}) = \begin{cases} \log \frac{|q_j|^2}{s_j} - \frac{|q_j|^2}{s_j} + 1 & \text{if } |q_j|^2 > s_j \\ 0 & \text{otherwise.} \end{cases} \tag{2.9}$$

Note that $\log \frac{|q_j|^2}{s_j} - \frac{|q_j|^2}{s_j} + 1 \leq 0$ with equality if and only if $q_j = s_j$. Since $L(\gamma_j^{\text{opt}}, \mathbf{C})$ is a monotonic non-increasing function of $\frac{|q_j|^2}{s_j}$ with $L(\gamma_j^{\text{opt}}, \mathbf{C}) = 0$ if $|q_j|^2 \leq s_j$ we can simplify the underlying problem in (2.7) as

$$l = \underset{j \notin \mathbb{T}}{\arg\max} \quad R_{\mathbf{C}}(j) := \max \left\{ \frac{|q_j|^2}{s_j}, 1 \right\}. \tag{2.10}$$

Algorithm 4 summarizes the proposed steps.

Remark 10. An important characteristic of the sequential SBL [13] and the proposed LWS-SBL Algorithm is that they may be run for more than m iterations if needed, unlike OMP [70]. This flexibility is achieved by avoiding the orthogonal residue computation step as in OMP. Such a flexibility may be useful to correct erroneous support that is possible in initial iterations.

Efficient updating of q_j and s_j: To lower the complexity of the steps in Algorithm 4, we exploit the fact that one column of $\mathbf{\Phi}$ is added per iteration and update $\mathbf{C}^{-1}$ using matrix inversion lemma. We highlight the value at iteration i with superscript $(\cdot)^{[i]}$. More specifically, let $\left(\mathbf{C}^{[i]}\right)^{-1}$ denote the value used to compute (q_j, s_j) as per (2.5), and $l^{[i]}$ denote the column index to be added at iteration i (Step 3 & 4 of Algorithm 4). Then

$$(\mathbf{C}^{[i+1]})^{-1} = (\mathbf{C}^{[i]} + \hat{\gamma}_{l^{[i]}} \mathbf{\Phi}_{l^{[i]}} \mathbf{\Phi}_{l^{[i]}}^H)^{-1} = (\mathbf{C}^{[i]})^{-1} - \frac{\mathbf{w}^{[i]} (\mathbf{w}^{[i]})^H}{1/\hat{\gamma}_{l^{[i]}} + s_{l^{[i]}}^{[i]}}$$

where $\mathbf{w}^{[i]} = \left(\mathbf{C}^{[i]}\right)^{-1} \mathbf{\Phi}_{l^{[i]}}$ and $s_{l^{[i]}}^{[i]} = \mathbf{\Phi}_{l^{[i]}}^H \left(\mathbf{C}^{[i]}\right)^{-1} \mathbf{\Phi}_{l^{[i]}}$. Then we can update $\left(q_j^{[i]}, s_j^{[i]}\right)$ to

Algorithm 4: Light-Weight Sequential SBL Algorithm

Result: $\hat{\boldsymbol{\gamma}}$, Posterior mean $\hat{\boldsymbol{\mu}}_{\mathbf{x}} = \hat{\mathbf{x}}$, covariance $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}$

Input: $\mathbf{y}, \boldsymbol{\Phi}, K$

Initialize: $\hat{\boldsymbol{\gamma}} = \mathbf{0}, \lambda =$ some sensible value (e.g., $0.1\text{var}(\mathbf{y})$), $\mathbf{C}^{-1} = \lambda^{-1}\mathbf{I}, \mathbb{T} = \emptyset$

for $i := 1$ to K do

  Compute q_j and $s_j, \forall j \notin \mathbb{T}$ as in (2.5) (efficient recursive implementation in (2.11))

  $l = \arg\max_{j \notin \mathbb{T}} \quad R_{\mathbf{C}}(j) := \max\left\{\frac{|q_j|^2}{s_j}, 1\right\}$

  $\hat{\gamma}_l = \max\left\{\frac{|q_l|^2 - s_l}{s_l^2}, 0\right\}$; rank-one update of $\mathbf{C}^{-1}$

  $\mathbb{T} = \mathbb{T} \cup l$ (unless $\hat{\gamma}_l = 0$, rare)

end

Compute posterior mean, $\hat{\boldsymbol{\mu}}_{\mathbf{x}}$, and covariance, $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}$

get $\left(q_j^{[i+1]}, s_j^{[i+1]}\right)$ as

$$\begin{aligned} q_j^{[i+1]} &= \boldsymbol{\Phi}_j^H \left(\mathbf{C}^{[i+1]}\right)^{-1} \mathbf{y} = q_j^{[i]} - \frac{\boldsymbol{\Phi}_j^H \mathbf{w}^{[i]}}{1/\hat{\gamma}_{l^{[i]}} + s_{l^{[i]}}^{[i]}} q_{l^{[i]}}^{[i]}, \\ s_j^{[i+1]} &= \boldsymbol{\Phi}_j^H \left(\mathbf{C}^{[i+1]}\right)^{-1} \boldsymbol{\Phi}_j = s_j^{[i]} - \frac{|\boldsymbol{\Phi}_j^H \mathbf{w}^{[i]}|^2}{1/\hat{\gamma}_{l^{[i]}} + s_{l^{[i]}}^{[i]}}. \end{aligned} \tag{2.11}$$

Computing posterior mean, $\hat{\boldsymbol{\mu}}_{\mathbf{x}}$, and covariance, $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}$: The posterior on $\mathbf{x}$ is a complex Gaussian distribution [12, 26]. We report $\hat{\mathbf{x}} = \hat{\boldsymbol{\mu}}_{\mathbf{x}}$ as a point estimate after K iterations using

$$\hat{\boldsymbol{\mu}}_{\mathbf{x}} = \hat{\boldsymbol{\Gamma}}\boldsymbol{\Phi}^H \left(\boldsymbol{\Phi}\hat{\boldsymbol{\Gamma}}\boldsymbol{\Phi}^H + \lambda \mathbf{I}\right)^{-1} \mathbf{y} = \hat{\boldsymbol{\Gamma}}\mathbf{q}^{[K+1]}, \tag{2.12}$$

where $\mathbf{q}^{[K+1]} = \left[q_1^{[K+1]}, \ldots, q_n^{[K+1]}\right]^T$. Note that $\hat{\gamma}_j \neq 0$ only for $j \in \mathbb{T}$, and thus the above equation provides a sparse solution. Also, the diagonal entries of $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}$ can be easily obtained as following

$$[\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}]_{j,j} = \hat{\boldsymbol{\gamma}}_j - \hat{\boldsymbol{\gamma}}_j^2 \boldsymbol{\Phi}_j^H \left(\boldsymbol{\Phi}\hat{\boldsymbol{\Gamma}}\boldsymbol{\Phi}^H + \lambda \mathbf{I}\right)^{-1} \boldsymbol{\Phi}_j = \hat{\boldsymbol{\gamma}}_j - \hat{\boldsymbol{\gamma}}_j^2 s_j^{[K+1]}, \tag{2.13}$$

which is similarly non-zero only for $j \in \mathbb{T}$.

2.2.3 Computational Complexity of Algorithm 4

We analyze the computational complexity for the proposed LWS-SBL algorithm and compare with OMP.

Efficient computation of $\mathbf{w}^{[i]}$: The update equations in (2.11) require the vector $\mathbf{w}^{[i]} = \left(\mathbf{C}^{[i]}\right)^{-1}\mathbf{\Phi}_{l^{[i]}}$ which can be computed by first computing $\left(\mathbf{C}^{[i]}\right)^{-1}$ from $\left(\mathbf{C}^{[i-1]}\right)^{-1}$ using a rank-one update, followed by multiplying $\mathbf{\Phi}_{l^{[i]}}$. These steps need $O(m^2)$ computations. Instead we propose the following steps

$$\begin{aligned}\mathbf{w}^{[i]} &= (\mathbf{C}^{[i]})^{-1}\mathbf{\Phi}_{l^{[i]}} = (\mathbf{C}^{[i-1]})^{-1}\mathbf{\Phi}_{l^{[i]}} - \frac{\mathbf{w}^{[i-1]}(\mathbf{w}^{[i-1]})^H\mathbf{\Phi}_{l^{[i]}}}{1/\gamma_{l^{[i-1]}} + s_{l^{[i-1]}}^{[i-1]}} \\ &= (\mathbf{C}^{[1]})^{-1}\mathbf{\Phi}_{l^{[i]}} - \sum_{\tilde{i}=1}^{i-1}\frac{\mathbf{w}^{[\tilde{i}]}(\mathbf{w}^{[\tilde{i}]})^H\mathbf{\Phi}_{l^{[i]}}}{1/\gamma_{l^{[\tilde{i}]}} + s_{l^{[\tilde{i}]}}^{[\tilde{i}]}},\end{aligned} \tag{2.14}$$

where $\left(\mathbf{C}^{[1]}\right)^{-1} = \lambda^{-1}\mathbf{I}$. $\left(\mathbf{C}^{[1]}\right)^{-1}$ may be pre-computed; the rest requires $O(m(i-1))$ computations, an improvement over $O(m^2)$. In this manner, we avoid computing $\mathbf{C}^{-1}$ in Step 5 of Algorithm 4.

Computational Complexity: Analyzing and comparing two algorithms based on computational complexity is challenging because it depends on the metric employed, i.e. number and type of operations, parallel[1] versus sequential computations, storage, etc. For this work, we track the number of arithmetic operations, in particular, multiplications and divisions as they are more computationally demanding than additions or subtractions. We further assume that multiplication and division have similar complexity and do not distinguish them in this aspect, but highlight them by mentioning 'div' for divisions explicitly (e.g., n divisions represented as n 'div'). We summarize the computational complexity for LWS-SBL and compare with OMP in Table 2.1.

Remark 11. The extra $2(n-i)+(n-i)$ 'div' for LWS-SBL (than OMP) corresponds to

[1]Note that updating q_j's and s_j's may be parallelized.

Table 2.1. LWS-SBL vs. OMP: Computational Complexity ($p = i-1$)

	LWS-SBL	OMP (QR) [14, 71]
Iteration $i+1$	$mn+2mp+n+m$ $+2(n-i)+(n-i)$ 'div'	$mn+2mp+n+m$
Solution $\hat{\mathbf{x}}$	$m(3K-1)+2K$	$(K-1)K/2+K$ 'div'

updating s_j and dividing $|q_j|^2$ and $s_j, \forall j \notin \mathbb{T}$.

During the first iteration s_j does not depend on j as $\left(\mathbf{C}^{[1]}\right)^{-1} = \lambda^{-1}\mathbf{I}$ and thus the divisions may be avoided. Note that both the algorithms have $O(mn)$ per iteration computational complexity.

2.3 Global Analysis of OMP and LWS-SBL

In this section we compare the two algorithms from an operational perspective. The aim is to explore a standard framework that will help to benchmark different algorithms for sparse signal recovery beyond just the two focused in this work. Thus, it is envisioned that the framework may be applicable more generally.

2.3.1 On OMP Algorithm

The first iteration in OMP (and LWS-SBL) begins with comparing the absolute value of inner product of the dictionary columns with the received measurement vector $\mathbf{y}$. This leads to

$$\mathscr{L}_{\text{OMP}} := |\boldsymbol{\phi}_j^H \mathbf{y}|^2 \tag{2.15}$$

during the first iteration for both OMP and LWS-SBL. The column corresponding to the largest absolute value gets selected. The following iterations proceed in a similar manner, simply replacing the measurement vector by a residual vector. Thus, columns get selected in a greedy manner. The residual, $\mathbf{y}_{\text{res}}$, is computed as orthogonal to columns already selected (let's say $\mathbf{C} = [\boldsymbol{\phi}_{l_1}, \ldots, \boldsymbol{\phi}_{l_k}]$) and is given by $\mathbf{y}_{\text{res}} = \mathbf{P}_{\mathbf{C}}^{\perp}\mathbf{y}$, where $\mathbf{P}_{\mathbf{C}}^{\perp}$ denotes

the projection matrix on to the subspace orthogonal to the range space of $\mathbf{C}$. Note that

$$\boldsymbol{\phi}_j^H \mathbf{y}_{\text{res}} = (\boldsymbol{\phi}_j^H \mathbf{P}_{\mathbf{C}}^{\perp})\mathbf{y}, \tag{2.16}$$

can be thought of as the output of a beamformer $\mathbf{w}_{\text{OMP},j} = \mathbf{P}_{\mathbf{C}}^{\perp}\boldsymbol{\phi}_j$ in the direction $\boldsymbol{\phi}_j$ while simultaneously placing nulls at columns within $\mathbf{C}$. The above beamformer can be realized as the solution to the following optimization problem

$$\mathbf{w}_{\text{OMP},j} = \arg\min_{\mathbf{w}} \|\mathbf{w} - \boldsymbol{\phi}_j\|_2^2 \text{ subject to } \mathbf{C}^H\mathbf{w} = \mathbf{0}, \tag{2.17}$$

which implies that the OMP beamformer at column j minimizes the ℓ_2 distance to $\boldsymbol{\phi}_j$ while placing an absolute null at columns in $\mathbf{C}$. The metric for selecting a column in OMP is simply the output power of the beamformer.

Remark 12. The OMP beamformer has couple of issues associated with it. Firstly, the absolute null constraint, much like zero-forcing equalizer, is prone to amplify noise. Secondly, the OMP metric only accounts for the output power without accounting for the resulting interference and noise power.

Note, that the second issue raised here is handled better in Order-Recursive MP (ORMP) [71] as it enforces that the beamformer is appropriately normalized at every direction. Next we describe the two aspect of interest i. beamformer design ii. metric for selecting a column, for the LWS-SBL algorithm.

2.3.2 On LWS-SBL

At each iteration, LWS-SBL computes $\frac{|q_j|^2}{s_j}$ for each column $j \notin \mathbb{T}$. The metric can be equivalently expressed as

$$\frac{|q_j|^2}{s_j} = \frac{|q_j|^2/s_j^2}{1/s_j} = \frac{|\boldsymbol{\Phi}_j^H \mathbf{S}^{-1}\mathbf{y}|^2/(\boldsymbol{\Phi}_j^H \mathbf{S}^{-1}\boldsymbol{\Phi}_j)^2}{1/\boldsymbol{\Phi}_j^H \mathbf{S}^{-1}\boldsymbol{\Phi}_j}, \tag{2.18}$$

which is the ratio of the well-known minimum variance distortionless response (MVDR) total beamforming output power [56] and the expected interference and noise beamforming output power. The MVDR beamformer above is given by $\mathbf{w}_{\text{LWS-SBL},j} = \frac{\mathbf{S}^{-1}\boldsymbol{\phi}_j}{\boldsymbol{\Phi}_j^H \mathbf{S}^{-1}\boldsymbol{\Phi}_j}$ and is obtained as the solution to the following optimization problem

$$\mathbf{w}_{\text{LWS-SBL},j} = \arg\min_{\mathbf{w}} \mathbf{w}^H \mathbf{S}\mathbf{w} \text{ subject to } \mathbf{w}^H \boldsymbol{\phi}_j = 1. \tag{2.19}$$

Using $\mathbf{S} = \boldsymbol{\Phi}_{\mathbb{T}} \boldsymbol{\Gamma}_{\mathbb{T}} \boldsymbol{\Phi}_{\mathbb{T}}^H + \lambda \mathbf{I}$ and the constraint $\mathbf{w}^H \boldsymbol{\Phi}_j = \boldsymbol{\Phi}_j^H \mathbf{w} = 1$, we simplify the cost function above as

$$\begin{aligned} \mathbf{w}^H \mathbf{S}\mathbf{w} &= \mathbf{w}^H (\lambda \mathbf{I} + \boldsymbol{\Phi}_{\mathbb{T}} \boldsymbol{\Gamma}_{\mathbb{T}} \boldsymbol{\Phi}_{\mathbb{T}}^H) \mathbf{w} \\ &= \lambda \mathbf{w}^H \mathbf{w} + (\mathbf{w}^H \boldsymbol{\Phi}_{\mathbb{T}}) \boldsymbol{\Gamma}_{\mathbb{T}} (\boldsymbol{\Phi}_{\mathbb{T}}^H \mathbf{w}) \\ &= \lambda \|\mathbf{w} - \boldsymbol{\Phi}_j\|_2^2 + (\mathbf{w}^H \boldsymbol{\Phi}_{\mathbb{T}}) \boldsymbol{\Gamma}_{\mathbb{T}} (\boldsymbol{\Phi}_{\mathbb{T}}^H \mathbf{w}) - c, \end{aligned} \tag{2.20}$$

where $c = (2\lambda + \lambda \|\boldsymbol{\Phi}_j\|_2^2)$. The above exploration helps to highlight the soft null constraint imposed within the BF design in LWS-SBL by virtue of the regularizing second term above. It further helps to contrast with the absolute null constraint in the BF design in OMP.

Thus, the LWS-SBL algorithm departs from the OMP algorithm in terms of both i. the underlying beamformer design ii. the criterion adopted for selecting the next column to be added to the model, in that it accounts for the beamforming output interference and noise power and chooses the Signal-to-Interference and Noise Ratio (SINR) as the metric. This framework provides unique perspective into the operational mechanisms of the two algorithms. This framework can be extended to other SSR algorithms as well. We explore one connection of OMP with CLEAN algorithm in the following remark.

Remark 13. The OMP algorithm can be considered as a special case of the CLEAN algorithm [72–74] wherein a 'clean' spectrum is obtained from a 'dirty' spectrum over multiple iterations. In CLEAN algorithm, the direction corresponding to the peak of a (residual)

power spectral density estimate is found, and a new residual is computed by subtracting a (scaled) contribution along the recovered direction. A similar analysis presented in this section can be carried out to identify the beamformer design and corresponding metric that is operational within CLEAN algorithm. This is not discussed ahead but forms an interesting future research direction.

Next, we extend the perspective presented in this section to analyze the case when sources can be closely separated. We highlight the usefulness of the beamformer design and metric employed under LWS-SBL in such cases.

2.4 Local Analysis for the Closely Separated Sources' case

In this section we consider a noiseless scenario and study the case when two sources are closely separated. The impact of the beamformer and the metric employed for selecting columns in OMP and LWS-SBL is highlighted. In particular, we show that owing to the metric within OMP, the algorithm fails deterministically when two sources are sufficiently close and the grid size is sufficiently large. We begin by first describing the exact conditions.

Let $\mathbb{S}$ denote the true support set. We provide the algorithms with the cardinality of the support set and let them run for the same number of iterations as support set size. An algorithm fails to recover the true support exactly if at any iteration it recovers an element not in the true support set. Thus, let $\mathbb{S}_{\text{rec}}$ denote the recovered support set, and let us assume that $\mathbb{S}_{\text{rec}} \subseteq \mathbb{S}$. Let $\mathbf{P}^{\perp}_{\mathbb{S}_{\text{rec}}}$ denote the projection matrix into the orthogonal subspace of the recovered support. Then the residual vector is given by

$$\mathbf{P}^{\perp}_{\mathbb{S}_{\text{rec}}}\mathbf{y} = \sum_{j\in\mathbb{S}} \mathbf{P}^{\perp}_{\mathbb{S}_{\text{rec}}}\mathbf{\Phi}_j x_j = \sum_{j\in\mathbb{S}\setminus\mathbb{S}_{\text{rec}}} \mathbf{P}^{\perp}_{\mathbb{S}_{\text{rec}}}\mathbf{\Phi}_j x_j. \tag{2.21}$$

The next iteration within OMP algorithm is unsuccessful if

$$\max_{l\in\mathbb{S}} |\mathbf{\Phi}_l^H \mathbf{P}^{\perp}_{\mathbb{S}_{\text{rec}}} \mathbf{y}| = \max_{l\in\mathbb{S}\setminus\mathbb{S}_{\text{rec}}} |\mathbf{\Phi}_l^H \mathbf{P}^{\perp}_{\mathbb{S}_{\text{rec}}} \mathbf{y}| < \max_{k\in\mathbb{S}^c} |\mathbf{\Phi}_k^H \mathbf{P}^{\perp}_{\mathbb{S}_{\text{rec}}} \mathbf{y}|. \tag{2.22}$$

The approach we follow next is to find stricter yet simplified conditions that satisfy (2.22). Based on triangle inequality and operating the projection matrix on $\mathbf{y}$, we get

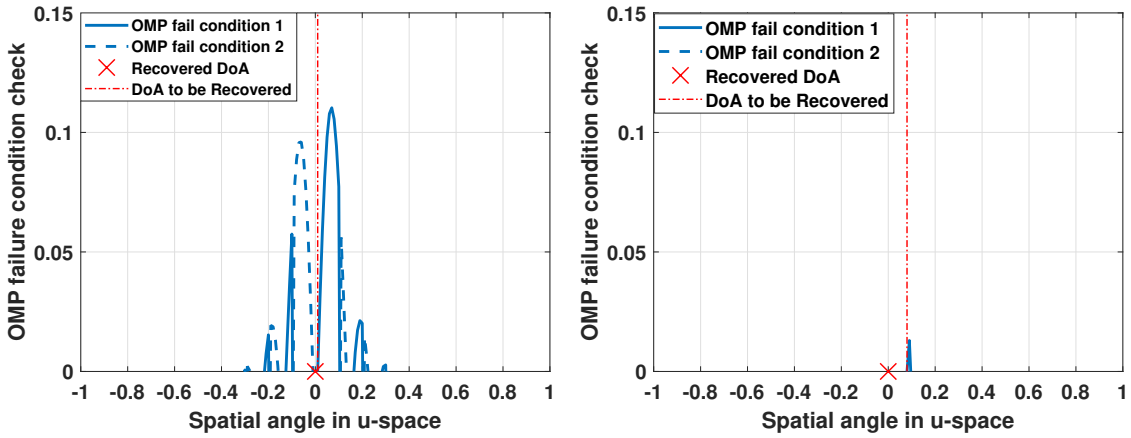
$$\begin{aligned}
|\mathbf{\Phi}_l^H \mathbf{P}^{\perp}_{\mathbb{S}_{\text{rec}}} \mathbf{y}| &\leq |\mathbf{\Phi}_l^H \mathbf{P}^{\perp}_{\mathbb{S}_{\text{rec}}} \mathbf{\Phi}_l||x_l| + |\mathbf{\Phi}_l^H \mathbf{P}^{\perp}_{\mathbb{S}_{\text{rec}}} \mathbf{\Phi}_{\mathbb{S}\setminus\mathbb{S}_{\text{rec}}\setminus l} \mathbf{x}_{\mathbb{S}\setminus\mathbb{S}_{\text{rec}}\setminus l}| \\
&= \left(1 - \mathbf{\Phi}_l^H \mathbf{P}_{\mathbb{S}_{\text{rec}}} \mathbf{\Phi}_l\right)|x_l| + |\mathbf{\Phi}_l^H \mathbf{P}^{\perp}_{\mathbb{S}_{\text{rec}}} \mathbf{\Phi}_{\mathbb{S}\setminus\mathbb{S}_{\text{rec}}\setminus l} \mathbf{x}_{\mathbb{S}\setminus\mathbb{S}_{\text{rec}}\setminus l}| \qquad (2.23) \\
|\mathbf{\Phi}_k^H \mathbf{P}^{\perp}_{\mathbb{S}_{\text{rec}}} \mathbf{y}| &\geq \left| |\mathbf{\Phi}_k^H \mathbf{P}^{\perp}_{\mathbb{S}_{\text{rec}}} \mathbf{\Phi}_l||x_l| - |\mathbf{\Phi}_k^H \mathbf{P}^{\perp}_{\mathbb{S}_{\text{rec}}} \mathbf{\Phi}_{\mathbb{S}\setminus\mathbb{S}_{\text{rec}}\setminus l} \mathbf{x}_{\mathbb{S}\setminus\mathbb{S}_{\text{rec}}\setminus l}| \right| \\
&= \Big| |\mathbf{\Phi}_k^H \mathbf{\Phi}_l - \mathbf{\Phi}_k^H \mathbf{P}_{\mathbb{S}_{\text{rec}}} \mathbf{\Phi}_l||x_l| \\
&\quad - |\mathbf{\Phi}_k^H \mathbf{P}^{\perp}_{\mathbb{S}_{\text{rec}}} \mathbf{\Phi}_{\mathbb{S}\setminus\mathbb{S}_{\text{rec}}\setminus l} \mathbf{x}_{\mathbb{S}\setminus\mathbb{S}_{\text{rec}}\setminus l}| \Big| \qquad (2.24)
\end{aligned}$$

We consider the last iteration and assume all the previous iterations successfully recovered elements from true support. In this case, $\mathbb{S} = \mathbb{S}_{\text{rec}} \cup \{l\}$, and both the bounds are tight. Then, the condition for failure in OMP can be written as

$$\left(1 - \mathbf{\Phi}_l^H \mathbf{P}_{\mathbb{S}_{\text{rec}}} \mathbf{\Phi}_l\right) \leq |\mathbf{\Phi}_k^H \mathbf{\Phi}_l - \mathbf{\Phi}_k^H \mathbf{P}_{\mathbb{S}_{\text{rec}}} \mathbf{\Phi}_l| \tag{2.25}$$

Qualitatively, the condition implies that the last support element needs to have high correlation with the recovered support set, compared to the confusing column in k. The latter should have high correlation with the unrecovered support element. This intuitive condition stems from OMP's metric for selecting next column wherein it ignores the impact of underlying beamformer on interference and noise.

We continue our treatment to get stricter yet intuitive bounds for the case when



(a) Second source at $u = 2/n$ (b) Second source at $u = 8/n$

Figure 2.1. OMP analysis to identify scenarios with deterministic misdetection

OMP fails. Let $p = \max_{i \in \mathbb{S}_{\text{rec}}} |\mathbf{\Phi}_l^H \mathbf{\Phi}_i| \leq 1$. Then we have

$$\mathbf{\Phi}_l^H \mathbf{P}_{\mathbb{S}_{\text{rec}}} \mathbf{\Phi}_l \geq p^2. \tag{2.26}$$

This implies

$$1 - \mathbf{\Phi}_l^H \mathbf{P}_{\mathbb{S}_{\text{rec}}} \mathbf{\Phi}_l \leq 1 - p^2. \tag{2.27}$$

Thus we can pose a stricter condition as

$$1 - p^2 \leq |\mathbf{\Phi}_k^H \mathbf{\Phi}_l - \mathbf{\Phi}_k^H \mathbf{P}_{\mathbb{S}_{\text{rec}}} \mathbf{\Phi}_l|. \tag{2.28}$$

We further use triangle inequality to lower bound RHS as

$$|\mathbf{\Phi}_k^H \mathbf{\Phi}_l - \mathbf{\Phi}_k^H \mathbf{P}_{\mathbb{S}_{\text{rec}}} \mathbf{\Phi}_l| \geq \left| |\mathbf{\Phi}_k^H \mathbf{\Phi}_l| - |\mathbf{\Phi}_k^H \mathbf{P}_{\mathbb{S}_{\text{rec}}} \mathbf{\Phi}_l| \right|. \tag{2.29}$$

Thus a stricter condition than (2.22) for unsuccessful recovery at the last OMP iteration, provided all the previous iterations have been successful, can be written as

$$1 - p^2 \leq \left| |\mathbf{\Phi}_k^H \mathbf{\Phi}_l| - |\mathbf{\Phi}_k^H \mathbf{P}_{\mathbb{S}_{\text{rec}}} \mathbf{\Phi}_l| \right|. \tag{2.30}$$

$$|\mathbf{\Phi}_k^H\mathbf{\Phi}_l - \mathbf{\Phi}_k^H\mathbf{P}_{\mathbb{S}_{\text{rec}}}\mathbf{\Phi}_l| \geq \max\{|\mathbf{\Phi}_k^H\mathbf{\Phi}_l| - |\mathbf{\Phi}_k^H\mathbf{P}_{\mathbb{S}_{\text{rec}}}\mathbf{\Phi}_l|, \\ |\mathbf{\Phi}_k^H\mathbf{P}_{\mathbb{S}_{\text{rec}}}\mathbf{\Phi}_l| - |\mathbf{\Phi}_k^H\mathbf{\Phi}_l|\}. \tag{2.31}$$

Next we simplify one of the lower bound from above. Similar ideas can be applied for simplifying the second argument within max. We write

$$\begin{aligned} |\mathbf{\Phi}_k^H\mathbf{\Phi}_l - \mathbf{\Phi}_k^H\mathbf{P}_{\mathbb{S}_{\text{rec}}}\mathbf{\Phi}_l| &\geq |\mathbf{\Phi}_k^H\mathbf{\Phi}_l| - |\mathbf{\Phi}_k^H\mathbf{P}_{\mathbb{S}_{\text{rec}}}\mathbf{\Phi}_l| \\ &\geq |\mathbf{\Phi}_k^H\mathbf{\Phi}_l| - \|\mathbf{P}_{\mathbb{S}_{\text{rec}}}\mathbf{\Phi}_k\|\|\mathbf{P}_{\mathbb{S}_{\text{rec}}}\mathbf{\Phi}_l\| \\ &\geq |\mathbf{\Phi}_k^H\mathbf{\Phi}_l| - \|\mathbf{P}_{\mathbb{S}_{\text{rec}}}\mathbf{\Phi}_k\| \\ &\overset{(a)}{\geq} |\mathbf{\Phi}_k^H\mathbf{\Phi}_l| - \lambda_{\min}^{-0.5}\|\mathbf{\Phi}_{\mathbb{S}_{\text{rec}}}^H\mathbf{\Phi}_k\| \\ &\overset{(b)}{\geq} |\mathbf{\Phi}_k^H\mathbf{\Phi}_l| - \lambda_{\min}^{-0.5}\sqrt{s-1}q(k). \end{aligned} \tag{2.32}$$

In (a), $\lambda_{\min}$ denotes the minimum eigenvalue of the matrix $\mathbf{\Phi}_{\mathbb{S}_{\text{rec}}}^H\mathbf{\Phi}_{\mathbb{S}_{\text{rec}}}$. In (b), s denotes the support size including the column vector l, and $q(k) = \max_i |\mathbf{\Phi}_k^H\mathbf{\Phi}_{\mathbb{S}_{\text{rec},i}}|$. The strict condition can be written as

$$1 - p^2 \leq |\mathbf{\Phi}_k^H\mathbf{\Phi}_l| - \lambda_{\min}^{-0.5}\sqrt{s-1}q(k). \tag{2.33}$$

We now consider the special case when there are two sources to be recovered. For this case, when one source has already been recovered successfully, (2.30) simplifies to

$$1 - |\mathbf{\Phi}_l^H\mathbf{\Phi}_{\mathbb{S}_{\text{rec}}}|^2 \leq \left||\mathbf{\Phi}_k^H\mathbf{\Phi}_l| - |\mathbf{\Phi}_k^H\mathbf{\Phi}_{\mathbb{S}_{\text{rec}}}||\mathbf{\Phi}_l^H\mathbf{\Phi}_{\mathbb{S}_{\text{rec}}}|\right|. \tag{2.34}$$

We numerically show that this strict condition is non-trivial in Fig. 2.1. Figure plots RHS-LHS from (2.34) with condition 1 referring to using the operand of modulus in RHS as is, and condition 2 referring to using the negative of the operand of modulus in RHS. As evident from the plot, when the second source is just one grid point away from first

source with $n = 200$, there are other grid points which are better under OMP's objective, and thus it will fail. This failure condition is satisfied as long as source 2 is sufficiently close and the grid is sufficiently fine. Such resolution for OMP is not explored ahead, instead this aspect is delved into further in the numerical section.

2.5 Gridless LWS-SBL Algorithm

We now discuss the case when $\mathbf{\Phi}$ has a parametric representation, and extend LWS-SBL to perform gridless parameter estimation. Consider the following parametric data model

$$\mathbf{y} = \mathbf{\Phi}_{\boldsymbol{\theta}}\mathbf{x} + \mathbf{n}, \tag{2.35}$$

where the kth column of $\mathbf{\Phi}_{\boldsymbol{\theta}} \in \mathbb{C}^{m\times K}$ is a vector function of the parameter θ_k i.e., $[\mathbf{\Phi}_{\boldsymbol{\theta}}]_k = \boldsymbol{\phi}(\theta_k)$ for some known $\boldsymbol{\phi}(.), k \in \{1,\ldots,K\}$. $\boldsymbol{\theta} = [\theta_1,\ldots,\theta_K]^T$ and θ_k's lie in some known continuous domain. K denotes the number of active sources. Model assumptions made in (2.1) are also applicable in (2.35). The above problem is ubiquitous, with applications such as line spectral estimation [73] and direction-of-arrival (DoA) estimation [75] for narrowband signals; we emphasize the latter as means for exposition.

2.5.1 Grid-based Remodeling of (2.35)

The proposed Algorithm 4 in its present form is not amenable to the parametric problem in (2.35). Mathematically, the approach can be readily developed by modifying equation (2.7) to reflect optimization over the continuous parameter space rather than the index set. However, this step when implemented will require a grid search though the grid is not predetermined. As a practical implementation of this concept, we adopt a two-step approach where in step-1 we recover an on-grid support set using Algorithm 4. This is followed by step-2 where we locally optimize the recovered grid points over a smaller parameter space. This step is inspired by a previous work by the authors [76] on batch

EM-SBL where all components of $\boldsymbol{\gamma}$ are simultaneously updated. In contrast to [76], this paper proposes a fixed iterations strategy to identify a sparse solution, with an emphasis on computational complexity reduction.

Assuming a grid size n, we introduce a dictionary $\boldsymbol{\Phi}$ with array manifold vectors as columns

Algorithm 5: Gridless LWS-SBL Algorithm

Result: $\tilde{\boldsymbol{\theta}} = [\tilde{\theta}_1, \ldots, \tilde{\theta}_K]^T, \tilde{\boldsymbol{\gamma}}, \tilde{\boldsymbol{\mu}}_{\mathbf{x}} = \tilde{\mathbf{x}}, \tilde{\boldsymbol{\Sigma}}_{\mathbf{x}}$
Input: $\mathbf{y}, \boldsymbol{\Phi}, K$; $(\hat{\boldsymbol{\gamma}}, \mathbf{C}^{-1}, \lambda)$ from Algorithm 4; $\tilde{n}$, ITER

```
Initialize: γ̃ = γ̂
for iter := 1 to ITER do
    for i := 1 to K do
        Rank-one update of C^{-1} to get C_{-l_i}^{-1}
        u^opt := Newton steps (with backtracking)
        Compute γ^opt as in (2.40); update (γ̃_{l_i}, u_{l_i}) := (γ^opt, u^opt)
        Rank-one update of C_{-l_i}^{-1} to get C^{-1}
    end
end
```

Compute $\tilde{\boldsymbol{\mu}}_{\mathbf{x}}, \tilde{\boldsymbol{\Sigma}}_{\mathbf{x}}$; $\tilde{\theta}_i = \arcsin u_{l_i}, i \in \{1, \ldots, K\}$

i.e., $[\boldsymbol{\Phi}]_j = \boldsymbol{\phi}(\theta_j), \theta_j \in [\frac{-\pi}{2}, \frac{\pi}{2}), j \in \{1, \ldots, n\}$. For example, consider a uniform linear array (ULA) with m sensors and $d = \bar{\lambda}/2$ distance between adjacent sensors to prevent ambiguity in DoA estimation; $\bar{\lambda}$ denotes the wavelength of the incoming narrowband source signals. Then $[\boldsymbol{\Phi}]_j = \boldsymbol{\phi}(\theta_j) = [1, \exp(i\pi u_j), \ldots, \exp(i(m-1)\pi u_j)]^T, u_j = \sin\theta_j$.

2.5.2 Gridless LWS-SBL Algorithm Development

The goal is to further maximize the likelihood after the initial K iterations by locally optimizing the selected grid points. This enables to go beyond the limitations of the initial grid. We begin by separating out the l_i-th dictionary component, $i \in \{1, \ldots, K\}$, selected previously as part of Algorithm 4 and optimize its contribution to the likelihood not only with respect to (w.r.t.) the corresponding γ_{l_i}, but also w.r.t. the grid point

$u_{l_i} = \sin\theta_{l_i}$. Thus, we write

$$\mathscr{L}(\boldsymbol{\gamma}_{\mathbb{T}}) = \mathscr{L}(\boldsymbol{\gamma}_{\mathbb{T}\setminus\{l_i\}}) + \tilde{L}(\gamma_{l_i}, u_{l_i}, \mathbf{C}_{-l_i}), \tag{2.36}$$

where $\mathbf{C}_{-l_i} = \boldsymbol{\Phi}_{\mathbb{T}\setminus\{l_i\}}\boldsymbol{\Gamma}_{\mathbb{T}\setminus\{l_i\}}\boldsymbol{\Phi}^H_{\mathbb{T}\setminus\{l_i\}} + \lambda\mathbf{I}$, $\tilde{L}(\gamma, u, \mathbf{C}_{-l_i}) = \log(1+\gamma s(u)) - |q(u)|^2/(\gamma^{-1}+s(u))$, and

$$s(u) = \boldsymbol{\phi}(u)^H\mathbf{C}^{-1}_{-l_i}\boldsymbol{\phi}(u), \quad q(u) = \boldsymbol{\phi}(u)^H\mathbf{C}^{-1}_{-l_i}\mathbf{y}. \tag{2.37}$$

For ease of exposition, we represent a grid point with the notation $\Psi = (\gamma, u)$. We locally maximize the likelihood in the neighbourhood of $\Psi_{\text{prev}} = (\hat{\gamma}_{l_i}, u_{l_i})$ by equivalently minimizing

$$\Delta\tilde{L}_{\mathbf{C}_{-l_i}}(\Psi, \Psi_{\text{prev}}) = \tilde{L}(\gamma, u, \mathbf{C}_{-l_i}) - \tilde{L}(\hat{\gamma}_{l_i}, u_{l_i}, \mathbf{C}_{-l_i}). \tag{2.38}$$

Note that unlike in the previous section, the second term above is non-zero. However, it does not depend on Ψ and is therefore fixed. Another perspective to understand this is that we are replacing the same grid point Ψ_{prev} with any other point in its neighbourhood, and thus in order to maximize likelihood we only need to minimize the first term on RHS in (2.38). The underlying problem reduces to

$$\Psi^{\text{opt}} = (\gamma^{\text{opt}}, u^{\text{opt}}) = \underset{u\in[u_{l_i}-\delta, u_{l_i}+\delta)}{\arg\min} \min_{\gamma\geq 0} \tilde{L}(\gamma, u, \mathbf{C}_{-l_i}), \tag{2.39}$$

for some δ[2]. Similar steps as in previous section can be followed to simplify the objective and we get

$$\begin{aligned} u^{\text{opt}} &= \underset{u\in[u_{l_i}-\delta, u_{l_i}+\delta)}{\arg\max} \tilde{R}_{\mathbf{C}_{-l_i}}(u) \coloneqq \max\left\{\frac{|q(u)|^2}{s(u)}, 1\right\}, \\ \gamma^{\text{opt}} &= \max\left\{\frac{|q(u^{\text{opt}})|^2 - s(u^{\text{opt}})}{s(u^{\text{opt}})^2}, 0\right\}. \end{aligned} \tag{2.40}$$

[2]We set $\delta = 1/n$ during the first iteration to avoid any grid point overlap. Future iterations may involve non-uniform grid, and a similar asymmetric δ in either directions is used to avoid the same issue.

The optimization for u in the neighbourhood of Ψ_{prev} can be implemented by a fine grid of size $\tilde{n}$. This approach was presented in [64]. In this work, we propose an alternative strategy that is more computationally favourable, and we discuss this in the next subsection.

2.5.3 Newtonized Gridless LWS-SBL

The optimization in some neighbourhood of u_{l_i} in (2.40) can be carried out in a completely gridless manner by using a quadratic approximation of the objective function in (2.40). We implement this using Newton steps and the update equation is given by

$$u^{\text{opt}}[t] = u^{\text{opt}}[t-1] - \frac{\tilde{R}'_{\mathbf{C}_{-l_i}}(u = u^{\text{opt}}[t-1])}{\tilde{R}''_{\mathbf{C}_{-l_i}}(u = u^{\text{opt}}[t-1])}, \tag{2.41}$$

where $\tilde{R}'_{\mathbf{C}_{-l_i}}(u) = \frac{1}{s(u)}\left\{2\text{Re}\left(q'(u)q(u)^H\right) - \frac{|q(u)|^2}{s(u)}s'(u)\right\}$
and $\tilde{R}''_{\mathbf{C}_{-l_i}}(u) = \frac{1}{s(u)}\left\{2\text{Re}\left(q''(u)q(u)^H\right) + 2q'(u)q'(u)^H - \frac{|q(u)|^2}{s(u)}s''(u)\right\} - 2\frac{s'(u)}{s(u)}\tilde{R}'_{\mathbf{C}_{-l_i}}(u)$. The additional required quantities can be computed as follows:

$$q'(u) = \left(\frac{\partial \boldsymbol{\phi}(u)}{\partial u}\right)^H \mathbf{C}_{-l_i}^{-1}\mathbf{y} \tag{2.42}$$

$$s'(u) = 2\text{Re}\left(\left(\frac{\partial \boldsymbol{\phi}(u)}{\partial u}\right)^H \mathbf{C}_{-l_i}^{-1}\boldsymbol{\phi}(u)\right) \tag{2.43}$$

$$q''(u) = \left(\frac{\partial^2 \boldsymbol{\phi}(u)}{\partial u^2}\right)^H \mathbf{C}_{-l_i}^{-1}\mathbf{y} \tag{2.44}$$

$$s''(u) = 2\text{Re}\left(\left(\frac{\partial^2 \boldsymbol{\phi}(u)}{\partial u^2}\right)^H \mathbf{C}_{-l_i}^{-1}\boldsymbol{\phi}(u)\right) + 2\left(\frac{\partial \boldsymbol{\phi}(u)}{\partial u}\right)^H \mathbf{C}_{-l_i}^{-1}\left(\frac{\partial \boldsymbol{\phi}(u)}{\partial u}\right). \tag{2.45}$$

The Newton update is carried out only if $\tilde{R}''_{\mathbf{C}_{-l_i}}(u^{\text{opt}}[t-1]) < 0$ as the underlying optimization is a maximization problem. Furthermore, the updated point is accepted only if it increases the objective function i.e., $\tilde{R}_{\mathbf{C}_{-l_i}}(u^{\text{opt}}[t]) > \tilde{R}_{\mathbf{C}_{-l_i}}(u^{\text{opt}}[t-1])$.

Newton Algorithm with backtracking line search: The Newton steps provided in the previous subsection are effective if the recovered grid point is in the close vicinity of the ground truth solution. The proximity is governed by the measurement size, m, as for larger values of m the recovered point must be closer to the ground truth source location, for Newton updates to perform reliably. This is not always the case, and the updates may qualitatively be too slow or too fast. In the former case, running more iterations of Newton update suffices. This case can be identified by checking if the relative change in $u^{\text{opt}}[t]$ is above certain threshold to continue, and setting the threshold appropriately. If the Newton step is too large, one may need to employ a backtracking strategy to ensure effectiveness of the Newton steps. This case occurs when the updated point decreases the objective function i.e., $\tilde{R}_{\mathbf{C}_{-l_i}}(u^{\text{opt}}[t]) < \tilde{R}_{\mathbf{C}_{-l_i}}(u^{\text{opt}}[t-1])$. We employ the following simple backtracking strategy to overcome this problem.

2.6 Extension to Two Dimensional Harmonic Retrieval Problem

The proposed algorithms can be readily extended to the multi-dimensional Harmonic Retrieval (HR) problem. In this section, we discuss the two dimensional HR case, and demonstrate the application of proposed ideas. We provide empirical analysis in Section 2.7.3. In the 2D HR problem, the goal is to recover frequency tuples (u_k, v_k) of 2D complex exponential signals from linear measurements given by

$$\mathbf{Y}[p_x, p_y] = \sum_{k=1}^{K} x_k \exp\left(i\pi\left(p_x u_k + p_y v_k\right)\right) + \mathbf{N}[n_x, n_y], \tag{2.46}$$

where $u_k, v_k \in [-1, 1)$ and $u_k^2 + v_k^2 \leq 1$, x_k denotes the complex amplitude for the k-th source. m_x and m_y denote the size of the measurements along x and y dimension, respectively; $0 \leq p_x < m_x$ and $0 \leq p_y < m_y, p_x, p_y \in \mathbb{Z}$. K denotes the number of active sources. The

equation in (2.46) can be expressed in a compact form as

$$\mathbf{Y} = \sum_{k=1}^{K} x_k \mathbf{\Phi}_{\mathbf{u}} \mathbf{\Phi}_{\mathbf{v}}^T + \mathbf{N}, \quad \mathbf{Y} \in \mathbb{C}^{m_x \times m_y}, \tag{2.47}$$

where $\mathbf{u} = [u_1, u_2, \ldots, u_K]^T$ and $\mathbf{v} = [v_1, v_2, \ldots, v_K]^T$, and the k-th column of $\mathbf{\Phi}_{\mathbf{u}} \in \mathbb{C}^{m_x \times K}$ and $\mathbf{\Phi}_{\mathbf{v}} \in \mathbb{C}^{m_y \times K}$ are given by

$$\begin{aligned} [\mathbf{\Phi}_{\mathbf{u}}]_k &= [1, \exp(i\pi u_k), \ldots, \exp(i\pi(m_x - 1)u_k)]^T \\ [\mathbf{\Phi}_{\mathbf{v}}]_k &= [1, \exp(i\pi v_k), \ldots, \exp(i\pi(m_y - 1)v_k)]^T. \end{aligned} \tag{2.48}$$

The 2D HR problem has many applications including MIMO radar [77], MIMO wireless channel sounding [78], and nuclear magnetic resonance spectroscopy [79]. One issue that may arise in 2D HR, and multi-dimensional HR in general, is that even thought the parameter pairs are distinct for different sources, the harmonics may have the same frequencies in one dimension, for example (u, v_1) and (u, v_2).

2.6.1 Grid-based remodeling of (2.47)

The problem in (2.47) can be modeled as a sparse signal recovery problem on grid, as discussed in the previous section for one-dimensional parametric problems. We introduce a grid of size n_u and n_v evaluated in u- and v-space, respectively. Let the corresponding measurement matrices be represented by $\mathbf{\Phi}_u \in \mathbb{C}^{m_x \times n_u}$ and $\mathbf{\Phi}_v \in \mathbb{C}^{m_y \times n_v}$. Then, (2.47) can be reformulated as

$$\mathbf{Y} = \mathbf{\Phi}_u \mathbf{X} \mathbf{\Phi}_v^T + \mathbf{N}. \tag{2.49}$$

$\mathbf{X} \in \mathbb{C}^{n_u \times n_v}$ denotes the matrix of complex amplitudes; each matrix entry in $\mathbf{X}$ corresponds to the amplitude associated with the corresponding harmonic. Since the number of active sources is limited to K, the number of non-zero entries in $\mathbf{X}$ is K as well. Next, we

transform (2.49) to a form that is readily amenable to algorithms presented in this work. We begin by vectorizing the measurements in $\mathbf{Y}$.

$$\begin{aligned}\mathbf{y} = \text{vec}(\mathbf{Y}^T) &= (\mathbf{\Phi}_u \otimes \mathbf{\Phi}_v)\,\text{vec}(\mathbf{X}^T) + \text{vec}(\mathbf{N}^T) \\ &= (\mathbf{\Phi}_u \otimes \mathbf{\Phi}_v)\,\mathbf{x} + \mathbf{n}, \end{aligned} \tag{2.50}$$

where $\otimes$ denotes Kronecker product, $\mathbf{x} = \text{vec}(\mathbf{X}^T) \in \mathbb{C}^{n_u n_v}$ is the sparse unknown vector, and $\mathbf{n} = \text{vec}(\mathbf{N}^T) \in \mathbb{C}^{m_x m_y}$ denotes the noise corrupting the measurement. The proposed LWS-SBL algorithm can be applied to the problem in (2.50) with the combined measurement matrix $\mathbf{\Phi}_u \otimes \mathbf{\Phi}_v$. The issue where two harmonics (or distinct parameter pairs) share the same frequency along a dimension is automatically handled here, as the Kronecker product of two 1-D measurement matrices ensures that such pairs are included in the search space. The corresponding grid points are independently evaluated and selected based on the LWS-SBL cost function.

Next, we discuss steps to update the grid points recovered by applying LWS-SBL to (2.50), and recover the frequency tuples (u, v) in a gridless manner. The algorithm development is similar to the 1D case, and thus we discuss the algorithm in brief.

2.6.2 Gridless LWS-SBL Algorithm for 2D HR problem

The aim here is to further maximize the likelihood after the K iterations of LWS-SBL algorithm, K denotes the recovered support size. This is executed by locally optimizing the selected grid points. We begin by separating out the l_i-th dictionary component, $i \in \{1, \ldots, K\}$ and optimize its contribution w.r.t. the parameters γ_{l_i} and (u_{l_i}, v_{l_i}). Similar

steps as in the 1D case can be followed which leads to the following optimization problem

$$\begin{aligned}
&(u^{\text{opt}}, v^{\text{opt}}) \\
&= \underset{\substack{u\in[u_{l_i}-\delta,u_{l_i}+\delta)\\ v\in[v_{l_i}-\delta,v_{l_i}+\delta)}}{\arg\max} \tilde{R}_{\mathbf{C}_{-l_i}}(u,v) := \max\left\{\frac{|q(u,v)|^2}{s(u,v)}, 1\right\}, \\
&\gamma^{\text{opt}} = \max\left\{\frac{|q(u^{\text{opt}}, v^{\text{opt}})|^2 - s(u^{\text{opt}}, v^{\text{opt}})}{s(u^{\text{opt}}, v^{\text{opt}})^2}, 0\right\}.
\end{aligned} \tag{2.51}$$

A fine grid in 2D may be implemented locally to (u_{l_i}, v_{l_i}) in order to perform the optimization in (2.51). Alternatively, we propose the following Newton update steps, and they are implemented along with backtracking steps similar to that in Section 2.5.3

$$\begin{aligned}
\begin{bmatrix} u^{\text{opt}}[t] \\ v^{\text{opt}}[t] \end{bmatrix} = \begin{bmatrix} u^{\text{opt}}[t-1] \\ v^{\text{opt}}[t-1] \end{bmatrix} &- \left(\nabla^2_{u,v}\tilde{R}_{\mathbf{C}_{-l_i}}(u^{\text{opt}}[t-1], v^{\text{opt}}[t-1])\right)^{-1} \\
&\cdot \nabla_{u,v}\tilde{R}_{\mathbf{C}_{-l_i}}(u^{\text{opt}}[t-1], v^{\text{opt}}[t-1]),
\end{aligned} \tag{2.52}$$

where

$$\begin{aligned}
\nabla_{u,v}\tilde{R}_{\mathbf{C}_{-l_i}}(u,v) &= \begin{bmatrix} \frac{\partial \tilde{R}_{\mathbf{C}_{-l_i}}(u,v)}{\partial u} \\ \frac{\partial \tilde{R}_{\mathbf{C}_{-l_i}}(u,v)}{\partial v} \end{bmatrix} \\
&= \frac{1}{s(u,v)} \begin{bmatrix} \frac{\partial |q(u,v)|^2}{\partial u} \\ \frac{\partial |q(u,v)|^2}{\partial v} \end{bmatrix} - \frac{|q(u,v)|^2}{s(u,v)^2} \begin{bmatrix} \frac{\partial s(u,v)}{\partial u} \\ \frac{\partial s(u,v)}{\partial v} \end{bmatrix}, \text{ and}
\end{aligned} \tag{2.53}$$

$$\begin{aligned}
\nabla^2_{u,v}\tilde{R}_{\mathbf{C}_{-l_i}}(u,v) &= \begin{bmatrix} \frac{\partial^2 \tilde{R}_{\mathbf{C}_{-l_i}}(u,v)}{\partial u^2} & \frac{\partial^2 \tilde{R}_{\mathbf{C}_{-l_i}}(u,v)}{\partial u \partial v} \\ \frac{\partial^2 \tilde{R}_{\mathbf{C}_{-l_i}}(u,v)}{\partial v \partial u} & \frac{\partial^2 \tilde{R}_{\mathbf{C}_{-l_i}}(u,v)}{\partial v^2} \end{bmatrix} \\
&= \frac{1}{s(u,v)} \begin{bmatrix} \frac{\partial^2 |q(u,v)|^2}{\partial u^2} & \frac{\partial^2 |q(u,v)|^2}{\partial u \partial v} \\ \frac{\partial^2 |q(u,v)|^2}{\partial v \partial u} & \frac{\partial^2 |q(u,v)|^2}{\partial v^2} \end{bmatrix} - \frac{|q(u,v)|^2}{s(u,v)^2} \begin{bmatrix} \frac{\partial^2 s(u,v)}{\partial u^2} & \frac{\partial^2 s(u,v)}{\partial u \partial v} \\ \frac{\partial^2 s(u,v)}{\partial v \partial u} & \frac{\partial^2 s(u,v)}{\partial v^2} \end{bmatrix} \\
&\quad - \frac{1}{s(u,v)} \left(\nabla_{u,v}\tilde{R}_{\mathbf{C}_{-l_i}}(u,v) \begin{bmatrix} \frac{\partial s(u,v)}{\partial u} \\ \frac{\partial s(u,v)}{\partial v} \end{bmatrix}^T + \begin{bmatrix} \frac{\partial s(u,v)}{\partial u} \\ \frac{\partial s(u,v)}{\partial v} \end{bmatrix} \nabla_{u,v}\tilde{R}_{\mathbf{C}_{-l_i}}(u,v)^T \right).
\end{aligned} \tag{2.54}$$

The required single and double (partial) derivatives of $|q(u,v)|^2$ and $s(u,v)$ can be computed in a manner similar to the 1D case; we skip the details for brevity.

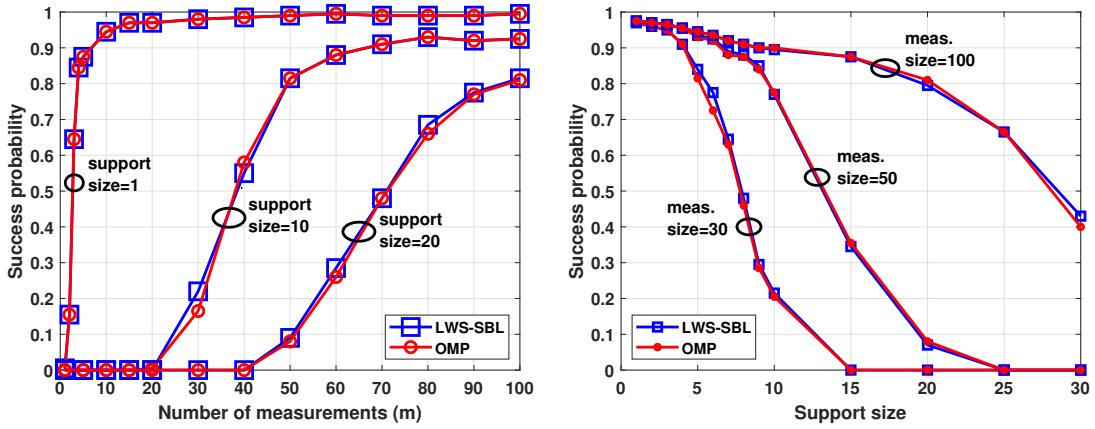
2.7 Numerical Results

In this section we study the empirical performance of the algorithms presented in this work, and compare them with OMP. For the 1D parametric measurement models we compare gridless LWS-SBL with Newtonized OMP algorithm [80]. We present results for the grid-based, 1D gridless, and 2D HR problems in subsections 2.7.1,2.7.1 ,and 2.7.3, respectively. For the grid-based scenario, we consider two types of measurement matrices, namely i. matrices with Gaussian random entries ii. parametric matrices. Parametric matrices allows to study the impact of coherence among columns of support in the measurement matrix on the sparse signal recovery performance. For grid-based algorithms, the metric employed for performance comparison is the probability of successful recovery of the true support. Let $\mathbb{S}$ and $\mathbb{S}^{\text{rec}}$ denote the true and recovered support sets, respectively. A successful event is characterized by $\mathbb{S} = \mathbb{S}^{\text{rec}}$ i.e., the algorithm recovers all the support elements present in ground truth and nothing more. For figures involving gridless algorithms, root mean squared error is computed in u-space as $\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{u}_t - u_t)^2}$, where T denotes the total number of random trials.

2.7.1 Grid-Based Sparse Signal Recovery

The following parameters are used for the experiments in this subsection, unless stated otherwise. The grid size is $n = 256$, the results are summarized over $T = 200$ random realizations. The support set is generated uniformly at random. The entries of the measurement matrix and the non-zero entries in $\mathbf{x}$ are generated from the standard normal distribution. The noise follows real normal distribution, and the per measurement and per source SNR is 30 dB.

Probability of success as function of measurement size: In Fig. 2.2 (a), we study the prob-



(a) (b)

Figure 2.2. (a) Probability of successful support recovery as a function of measurement size, for different support sizes. (b) Probability of successful support recovery as a function of support size, for different measurement sizes.

ability of successful support recovery as a function of measurement size. As expected and seen in the plot, the probability of success increases as the measurement size increases. For cases with high support size to be recovered, a larger number of measurements are required to ensure the same probability of success. For the case with a single support element, both the proposed LWS-SBL and OMP have identical performance. This is expected as OMP exactly solves the maximum likelihood problem, and both OMP and LWS-SBL have identical underlying cost function. The proposed LWS-SBL algorithm improves over OMP when there are multiple sources to be handled. This is evident from the plot, as the blue curves (for LWS-SBL) are to the left of red curves (for OMP), which indicates that fewer measurements are required for LWS-SBL to ensure the same level of success probability as OMP.

Probability of success as function of support size: In Fig. 2.2 (b), we study the recovery performance in terms of probability of successful recovery as a function of support size. We plot three different measurement sizes $m = \{30, 50, 100\}$, and plot one curve corresponding to each case. As expected, the probability of success decreases as the support

size increases. Also, the recovery performance improves as the measurement size increases. The proposed algorithm can be seen to improve over OMP, in these different settings.

It is observed that the gap between the two algorithms is small when support set is selected uniformly at random. The gap is observed to widen when the support set is selected so as to gradually increase coherence among selected columns. This is studied next.

Parametric measurement matrices: Success probability as a function of separation between two sources: We consider 1D-parametric measurement models. In contrast to the random measurement matrix setting, a more controlled study is possible when parametric dictionaries are involved. One perspective to see this is consider the (on-grid) DoA estimation problem. There exists additional structure to the coherence between columns of the measurement matrix. It is high locally and decreases as the (wrap around) distance between grid points increases. The study is however more general, in that it also helps to reflect on random measurement matrix settings, as such high coherent columns are also likely there. The only difference is that the coherence is more localized in parametric model.

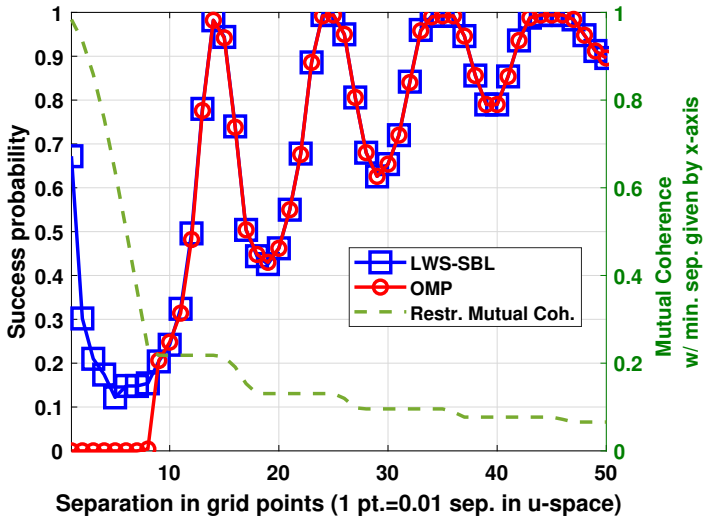


Figure 2.3. Probability of success as function of separation between sources

In this experiment we consider a measurement size of $m = 20$ and select a grid of size $n = 200$ and fix one source at $u = 0$. We place the second source at a spacing given by the x-axis. In Fig. 2.3 we study the recovery performance of LWS-SBL and OMP as a function of separation between the two sources. The SNR is 30 dB and we summarize the results over $T = 500$ random realizations. As observed in the plot, the two algorithms have similar performance except when the two sources are close. In contrast to OMP, the proposed LWS-SBL algorithm has a non-zero success probability in this regime (separation < 8). This is also a regime where the mutual coherence of the measurement matrix computed over those columns that ensure a higher (wrap around) distance than mentioned along x-axis is high (green curve marked as restricted mutual coherence). As discussed in Section 2.3, the proposed LWS-SBL algorithm employs SINR as the metric, whereas OMP ignores the impact of the beamformer on the interference and noise. This issue within OMP impacts the second and higher iterations, and biases the support away from the true support (see (2.25) and the following discussion therein). We study this behaviour more thoroughly next. The characteristic variation of performance over the separation between sources is due to the sidelobes of the underlying beamformer within OMP and LWS-SBL. Note that the separtion between nulls for coherent beamformer is $2/m = 0.1$ which is equal to 10 grid points. The separation between performance peaks in Fig. 2.3 is also around 10 grid points which helps explain the nature of the curves.

We consider the separation of one grid point between the two sources in Fig. 2.4. The true source location is the grid point number 101 and 102 (marked by black-dashed curve). We had observed in Fig. 2.3, that for this case the LWS-SBL is able to recover the true support with probability≈ 0.65. On the other hand, the OMP algorithm failed completely. In Fig. 2.4, we plot the histogram of the recovered support. As can be observed from the first row corresponding to LWS-SBL, the histogram peaks around the true source location. The histogram for OMP is plotted along second row and reveals the multi-modal distribution of the support recovered. Such a behaviour comes from the fact

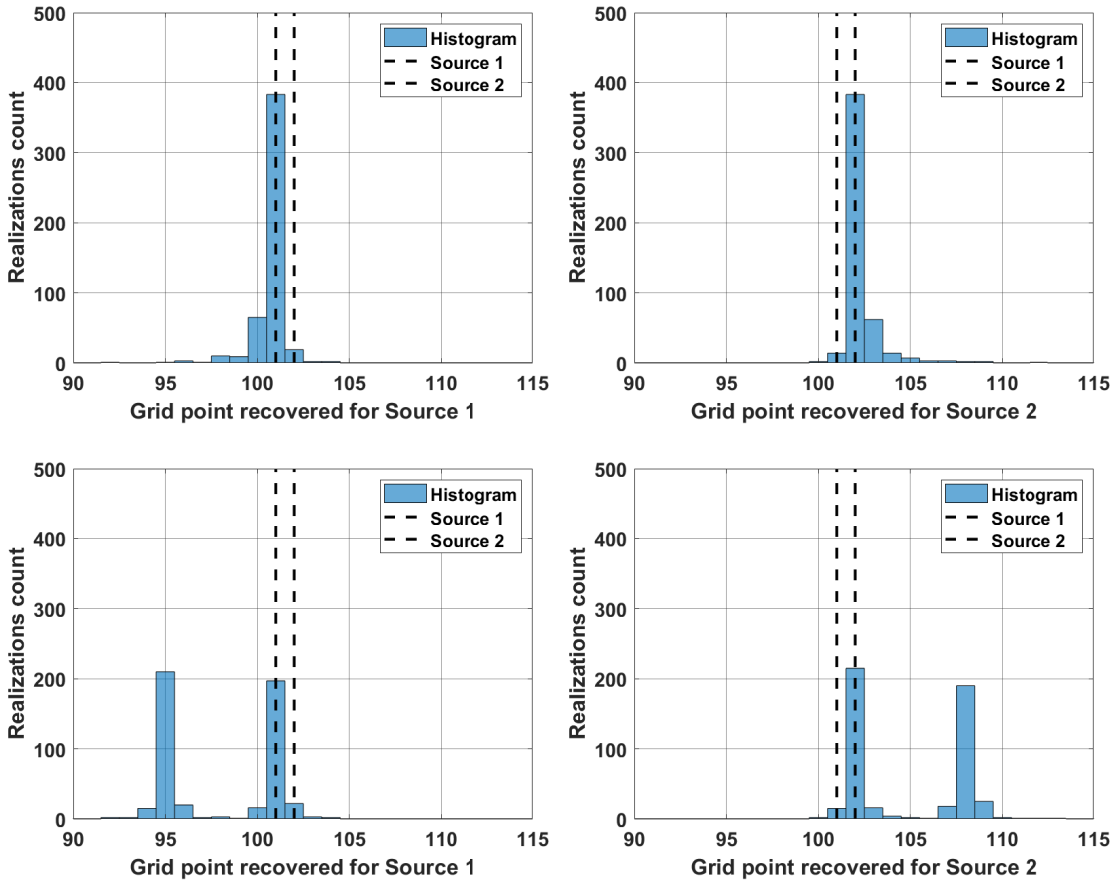
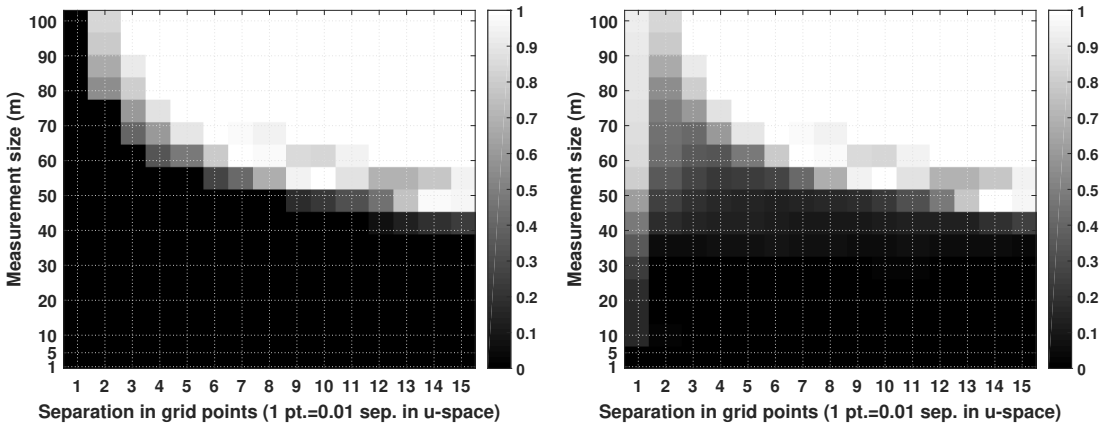


Figure 2.4. Histogram of recovered support elements for LWS-SBL (top row) and OMP (bottom row)

that even though a true grid point may be recovered in the first iteration of OMP, the second iteration drives away the peak of the OMP cost function. This was observed by tracking the separation between the grid points recovered by the two algorithms.

How many measurements are needed? An Empirical Study: In this experiment, we empirically analyze the probability of success as a function of both measurement size and the separation between two sources. The grid size is set to $n = 200$, SNR is 30 db. We summarize the results over $T = 500$ realizations in Fig. 2.5. As evident from the plot, the success probability increases for both algorithms as the measurement size or separation between sources increases. For separation greater than ≈ 7 grid points i.e., 0.07 in u-space, both OMP and LWS-SBL have similar performance. However, in the lower separation



OMP LWS-SBL

Figure 2.5. Probability of success as function of measurement size and separation between sources

regime, the proposed LWS-SBL algorithm can be seen to outperform OMP. It requires fewer measurements to achieve similar performance as OMP. The difference is even more clear when the separation between sources is one grid point, as the OMP is observed to fail completely, but LWS-SBL is able to recover the sparse support with high probability of success and uses few measurements.

2.7.2 Gridless Sparse Signal Recovery

Performance as a function of SNR: In this experiment we compare the proposed gridless LWS-SBL algorithm to the Newtonized OMP (NOMP) algorithm and the Cramér-Rao bound. We also plot the curve for the grid refinement based gridless LWS-SBL proposed in [64]. In Fig. 2.6 we plot the root mean squared error in u-space as a function of SNR. The measurement size $m = 50$ and the grid size for recovering the initial support is $n = 200$. A total of $K = 10$ sources are considered. Their locations are simulated by first considering grid points that are 10 points apart. These locations are then randomly perturbed (maximum perturbation $= 2/n$). The algorithms are provided with $L = 500$ snapshots, and they are adapted so as to handle more than one snapshot. As evident from the Fig. 2.6, all the algorithms have similar performance in this setting, and they

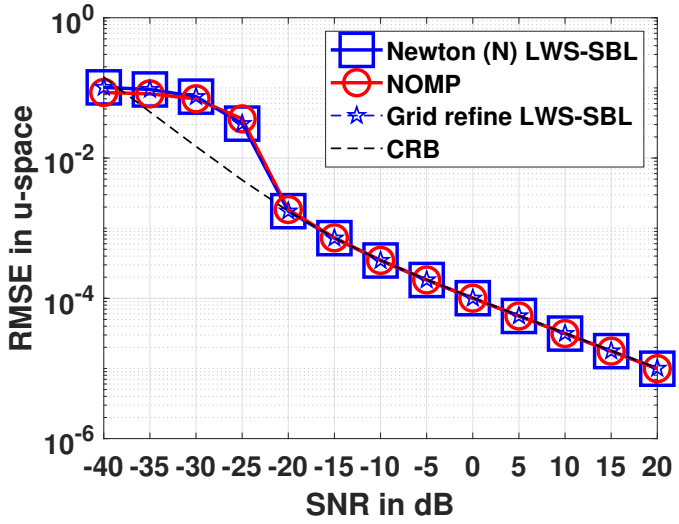
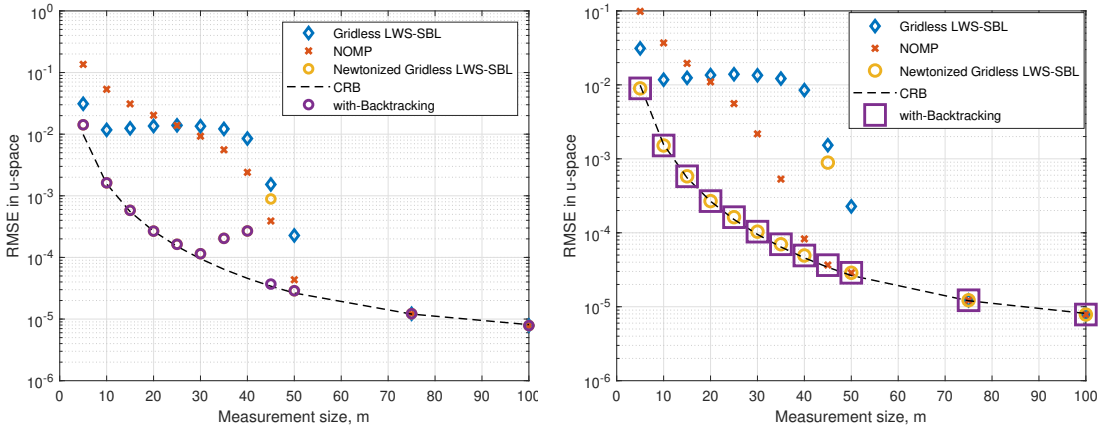


Figure 2.6. Performance of gridless LWS-SBL: RMSE as a function of SNR

approach CRB as SNR increases.

Performance as a function of measurement size: In Fig. 2.7 (a), we consider $K = 2$ sources placed roughly 10 grid points apart with $n = 500$. The Newton steps are allowed to run for three iterations for the proposed algorithm. We also set $R_s = 1$ and $R_c = 3$ for NOMP algorithm that ensures that all recovered support elements at least run $R_s + R_c = 4$ steps. As is evident from the figure, the proposed algorithm performs better than NOMP except at $m = 45$. The performance degradation as m increases can be explained using the fact that the underlying beamwidths become narrower as m increases. For $m \approx 45$, the beamwidths are such that the grid-based algorithm recovers a support set between the two sources, adding to the difficulty of Newton steps. The difficulty manifests itself in two ways - a) Newton steps are too fast, which is addressed by backtracking line search, and b) Newton steps are too slow, which is addressed by running more steps. The latter is demonstrated in Fig. 2.7b. For completion we plot the separation between two sources and the half power beamwidth as a function of measurement size m for a Rectangular window in Fig. 2.8. As evident from the plot, the difficulty of the problem increases when the separation between the two sources is close to the half power beamwidth, which



(a) Three Newton iterations (b) 10 Newton iterations

Figure 2.7. RMSE as a function of measurement size, m. Other parameters: $n = 500$, $L = 500, T = 100$, SNR$= 30$ dB

explains the performance degradation around $m = 40$ in Fig.2.7a.

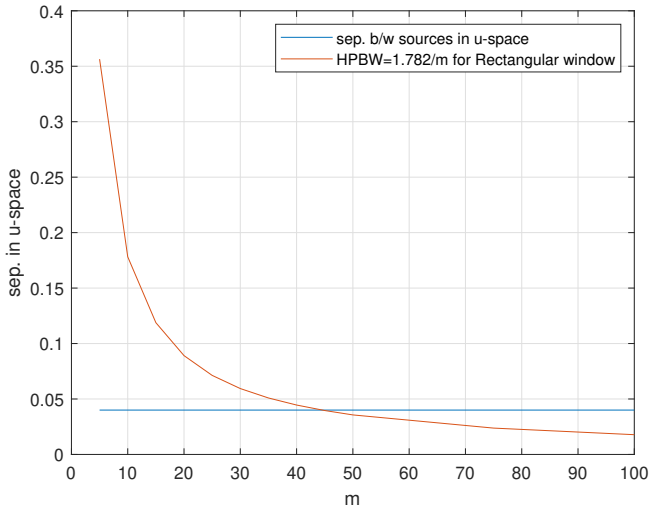


Figure 2.8. Half power beamwidth as a function of measurement size, m, for a Rectangular window.

2.7.3 Two-dimensional Harmonic Retrieval Problem

In this experiment, we evaluate the performance of the proposed algorithm to 2DHR problem. In Fig. 2.9, we study the performance in terms of RMSE as a function

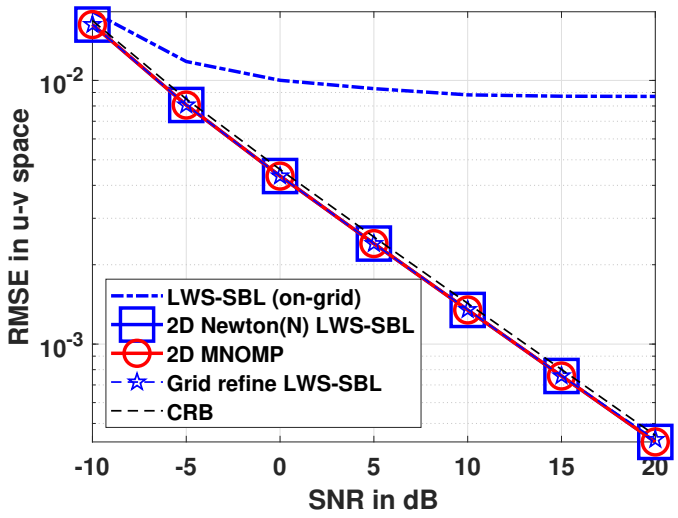


Figure 2.9. Root mean squared error in (u, v)-space as a function of SNR

of SNR. We compare the performance with the Cramér-Rao lower bound and the 2D-MNOMP algorithm in [81]. As evident from the plot, the algorithms are able to improve as SNR increases, unlike the grid-based algorithm which saturates to the grid resolution. The Newton steps proposed in this paper converges to the grid-refinement based gridless algorithm in [64], and thus is able to offer a similar performance with lower complexity than the algorithm in [64]. Note that the CRB is observed to be slightly worse than the performance achieved by the proposed algorithm. This may be explained by the bias introduced by the grid selected. As the source locations are fixed over the random realizations, the selected positions seem to be favourable for the grid adopted. This benign bias is thus a result of the experimental setting, and it can be eradicated by randomizing the true source locations as well. We skip this as the additional settings provides only a limited insight.

2.8 Conclusion

A novel low complexity algorithm for the sparse signal recovery problem is proposed based on the Sparse Bayesian Learning (SBL) formulation, named as Light-Weight

Sequential SBL algorithm. The proposed algorithm recovers support over iterations, similar to OMP, but unlike OMP it is based on the stochastic maximum likelihood estimation framework. Support vectors are added to the recovery set that gradually increases the likelihood over iterations. Efficient recursive iterations are developed with the computational complexity similar to OMP. A beamformer and metric-based perspective is developed that can provide a fundamental framework to classify and understand different sparse signal recovery algorithms. Under this framework, LWS-SBL and OMP are analyzed to understand their subtle differences. Numerical results are provided that both verify the theory and demonstrate the improvement achieved using LWS-SBL. Gridless extensions of the proposed algorithm are provided with applications to both 1D parametric model and multi-dimensional harmonic retrieval problem.

Chapter 2, in part is currently being prepared for submission for publication of the material, and in part, is a reprint of the material as it appears in R. R. Pote and B. D. Rao, "Light-Weight Sequential SBL Algorithm: An Alternative to OMP," ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 2023, pp. 1-5. The dissertation author was the primary investigator and author of this material.

Chapter 3

Maximum Likelihood-based Gridless DoA Estimation

3.1 Introduction

Consider the following parametric data model

$$\mathbf{y}_l = \mathbf{\Phi}_{\boldsymbol{\theta}} \mathbf{x}_l + \mathbf{n}_l, \qquad 0 \leq l < L, \tag{3.1}$$

where $\mathbf{y}_l \in \mathbb{C}^M$ denotes the measurements, and L denotes the total number of snapshots available. The kth column of $\mathbf{\Phi}_{\boldsymbol{\theta}} \in \mathbb{C}^{M \times K}$ is a vector function of the parameter θ_k i.e., $[\mathbf{\Phi}_{\boldsymbol{\theta}}]_k = \boldsymbol{\phi}(\theta_k)$ for some known $\boldsymbol{\phi}(.), k \in \{1, \ldots, K\}$. $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_K]^T$ and θ_k's lie in some known continuous domain. K denotes the number of sources. The sources' signal $\mathbf{x}_l \in \mathbb{C}^K$ and noise $\mathbf{n}_l \in \mathbb{C}^M$ are independent of each other, and i.i.d. over time. The noise, $\mathbf{n}_l$, is distributed as $\mathscr{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I})$. In (3.1), the parameters $(\boldsymbol{\theta}, \mathbf{x}_l, \sigma_n^2)$ are the unknowns. The model parameters affect the measurements in a non-linear manner, which makes the inverse problem extremely difficult to solve, even in the absence of noise. The above problem is ubiquitous, with applications including biomagnetic imaging [4], functional approximations [5], and echo cancellation [6]. In this work we are concerned with problems such as in line spectral estimation [73] and direction-of-arrival (DoA) estimation [75] for narrowband signals; we emphasize the latter as means for exposition. Approaches to solve (3.1) have

Table 3.1. Summary of traditional algorithms for DoA estimation: (a) Spectral based methods (b) Parametric methods

Methods	Primary Bottleneck [75]
(a) i. Spatial filtering (beamforming)	Aperture/ degrees of freedom
(a) ii. Subspace based methods	Number of snapshots
(b) Deterministic/ Stochastic MLE	Model & computational complexity

a rich history and can broadly be classified as traditional vs. modern, both significant in insights and contributions.

On traditional approaches: They can be further classified into spectral based [82–84] and parametric methods [75]. The typical ingredients to solve (3.1) include geometrical properties (e.g., subspace orthogonality in MUltiple SIgnal Classification (MUSIC) [83] or Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT) [84]) and statistical properties of the model in (3.1). A common thread that unites these methods is the usage of the second order statistics of the data. A second order statistic offers benefits such as a) compact representation of the data when[1] $L \geq M$ (also, sample covariance matrix serves as a sufficient statistic when data is Gaussian distributed) b) model based interpretation of data with much fewer parameters. Parametric methods are particularly attractive as they do not suffer from the bottlenecks faced by beamforming and subspace based methods (summary in Table 3.1). Parametric methods like maximum likelihood estimation (MLE) allow one to introduce meaningful parameters as a means to incorporate information about geometry and prior, which may be inferred even with a single snapshot. The main issues with MLE methods are the model complexity, as the resulting cost function may be highly non-linear in the parameters to solve, and often the model order is unknown.

On modern approaches: These techniques, under the rubric of sparse signal recovery (SSR), involve a) reparameterization of the original problem in (3.1) b) explicit or implicit sparsity regularization and corresponding optimization problem. They recover the

[1]This condition was rightfully pointed out by a reviewer.

parameter of interest in either grid-based or grid-less manner, and most often explicitly impose sparsity. Under the grid-based reparameterization [7–12], the methods first discretize the possible values of $\boldsymbol{\theta}$ and introduce the measurement matrix $\boldsymbol{\Phi} \in \mathbb{C}^{M \times G}$, G denotes the grid size. The i-th column $[\boldsymbol{\Phi}]_i = \boldsymbol{\phi}(\theta_i), i = 1, \ldots, G$, and $M \ll G$. The original problem in (3.1) can be re-written as

$$\mathbf{y}_l = \boldsymbol{\Phi}\bar{\mathbf{x}}_l + \bar{\mathbf{n}}_l, \qquad 0 \leq l < L, \tag{3.2}$$

where it is known that $\bar{\mathbf{X}} = [\bar{\mathbf{x}}_0, \ldots, \bar{\mathbf{x}}_{L-1}]$ is row-sparse i.e., most of the rows are zero. The problem in (3.2) is known as the multiple measurement vector or MMV problem when $L > 1$ [29], compared to the single measurement vector or SMV problem when just a single snapshot is available i.e., $L = 1$. The non-zero rows correspond to active sources, and one of the key problems in SSR is to identify these non-zero rows. For the gridless approach [22–25] the reparameterization involves Toeplitz matrix fitting of appropriate size. Note that modern techniques are applicable more generally even when there is no underlying parametric model, for example Gaussian random entries in $\boldsymbol{\Phi}$. Sparsity can be explicitly enforced by adding suitable p-pseudo-mixed norm[2] ($p \in (0, 1]$), $\|\bar{\mathbf{X}}\|_{2,p}$, regularizer for the grid case or atomic norm for the gridless formulations. The core emphasis in these approaches is on optimizing an appropriate fit to the measurements with an additional (sparsity) regularizer [8, 10, 11, 22–25]. Such methods are therefore sensitive to setting the regularization parameter properly. An exception to the explicit regularization based methods includes sparse Bayesian learning (SBL) [12, 26, 27] which recovers sparse solutions for (3.2) via implicit regularization [28]. SBL formulates the recovery problem under the MLE framework and therefore demonstrates superior performance.

The question we seek to answer is: how can we enhance the SBL formulation to overcome the model complexities faced by MLE methods of the past, and solve (3.1) i.e., perform

[2]Note that for $p = 1$ we get a norm, as it satisfies all the required axioms.

gridless estimation of $\boldsymbol{\theta}$? We identify the following contributions:

- It was shown in [85] that correlation-aware techniques effectively utilize available geometry and prior information and thus, can recover support as high as $O(M^2)$. In [86], it was shown that SBL can indeed identify $O(M^2)$ sources in the noiseless case under certain sufficient conditions on the dictionary and sources, and was shown empirically in the noisy case. In this work we reexamine the SBL formulation and show that it places a similar emphasis on available structure i.e., geometry and prior information, and thus is a correlation-aware technique!

- We reformulate the SBL problem as a novel structured matrix recovery (SMR) problem under the MLE framework. We will also show that the cost function employed by the proposed method can be derived using the Kullback-Leibler (KL) divergence between the true (data) distribution and the one assumed in this work. This insight provides a new perspective for understanding the underlying strategy to handle the case when sources may be arbitrarily correlated, extending the benefits of correlation-aware methods.

- A majorization-minimization (MM) procedure [87] to minimize the negative log-likelihood function is provided. One of the advantages of such an approach over other algorithms like sequential quadratic programming (SQP) is that more information is retained as we only majorize the concave terms in the cost. Thus, all information about third order and higher, of the convex terms is retained, unlike in SQP. Also, unlike SQPs where trust regions are required which limit progress per iteration, such conservative measures are prevented using convex-concave procedure (CCP) [88]. Thus, the linear MM procedure allows for more progress per iteration. We further discuss how array geometry can play an important role in identifying more sources than sensors. We also provide perspectives to understand the proposed approach and connect with the traditional MLE framework and the modern SBL formulation.

Table 3.2. Summary of Gridless Sparse Signal Recovery Algorithms

Methods	Applicable Array Geometries	Sparsity Regularizer	Iterative	Optimization Tool*	Noise variance (σ_n^2)
ANM [22]	ULA†	Atomic Norm (AN)	No	SDP	known
RAM [24]	ULA†	Reweighted AN	Yes	SDP	known
Gridless SPARROW [23, 90]	ULA†	$\ell_{2,1}$ mixed norm	No	SDP	known
Gridless SPICE [25]	ULA†	Implicit (trace norm)	No	SDP	unknown
Proposed	Non-uniform linear array	Implicit (logdet)	Yes	SDP	known**

ULA† includes ULA with missing sensors' case. Non-uniform linear array includes ULA† as a special case.
*First-order methods have been proposed for some of the above algorithms, although they were primarily derived as SDPs.
**σ_n^2 can be assumed unknown and estimated as part of the procedure.

- Finally, we consider arbitrary geometries where it is difficult to identify simplifying structures, that are otherwise possible for array geometries such as uniform linear arrays (ULA) with potential missing sensors. For this case, we propose adaptive grid-based strategies to extend SBL to alleviate the initial grid limitation.

The proposed techniques set us apart from other family of approaches in the literature that albeit put together a cost function with a similar essence (i.e. $\textbf{Simple Model} + \textbf{Data Fitting}$), but lack a (MLE) principled approach and hence the associated insights, performance guarantees and rich options. We provide numerical results to further elucidate the impact of the proposed techniques and compare them with other gridless approaches and the Cramér-Rao bound (CRB). Some of the work presented here was also discussed in [89] by the authors. We will now review some relevant prior work in this field.

3.1.1 Relevant Prior Work

Early works, primarily in the field of DoA estimation using the MLE based cost function include [91–96]. In [91], the authors proposed an iterative algorithm to solve the necessary gradient equations for moderate sized problems. An expectation-maximization (EM) based approach was proposed in [93] wherein the incomplete observed data is assumed to have a Toeplitz structured covariance, and where it is shown that it is possible to embed the incomplete data into a larger size periodic data series. A separable solution, consisting of an optimization problem for recovering support and a closed form expression for estimating the source covariance matrix was proposed in [94], which was further extended to

the case when noise variance is unknown in [97]. The problem was later considered in the presence of spatially correlated noise fields in [98]. A closed-form formula for estimating Hermitian Toeplitz covariance matrices using the extended invariance principle was suggested in [96]. A covariance matching based estimation to bypass the model complexity associated with the MLE based cost function was proposed in [99]. The approach developed in this paper can be viewed as a natural progression of this line of work, benefiting from the developments in the field and in optimization tools.

In [22], authors proposed a gridless scheme for estimating the frequency components of a mixture of complex sinusoids based on the concept of atomic norm [100]. They formulated a semidefinite program (SDP) which recovered a low rank Toeplitz matrix. Such a Toeplitz matrix can be further decomposed to identify the DoAs. In our work we similarly break the task into two steps. First, we recover a structured covariance matrix approximation for the sample covariance matrix (SCM). This recovery is based on the MLE cost function, unlike the work in [22]. The second step is similar to that in [22]. At each step we process the SCM, and do not process the received samples directly. As a result, the problem dimension is bounded, and results into a compact formulation. A similar compact reformulation, called SPARse ROW-norm reconstruction (SPARROW), for the atomic norm minimization problem was proposed in [23]. The atomic-norm minimization (ANM) technique in [22] builds on the mathematical theory of super-resolution developed by Candés et al. [101], in that it extends to the cases of partial/compressive samples and/or multiple measurement vectors. ANM, however, requires sources to be adequately separated, prohibiting true super-resolution. A re-weighted ANM (RAM) strategy that potentially overcomes the shortfalls of ANM was proposed in [24]. SParse Iterative Covariance-based Estimation (SPICE) was proposed in [102] as a grid-based sparse parameter estimation technique based on covariance matching, as opposed to the MLE formulation, and was later extended to the gridless case in [25]. It was shown in [25] that gridless SPICE and atomic norm-based techniques are equivalent, under varied as-

sumptions of noise. LIKelihood-based Estimation of Sparse parameters (LIKES) [103] was proposed as a grid-based method following the MLE principle, with the same application as SPICE. Table 3.2 summarizes recent gridless SSR approaches.

3.1.2 Organization of the Chapter and Notations

In Section 3.2 we begin with a simple insight into SBL formulation, and demonstrate that SBL is a correlation-aware technique. We further compare SBL with another line of correlation-aware algorithms based on minimizing diversity measures. We take this insight further and present the structured matrix recovery (SMR) reformulation and highlight benefits of the proposed approach when sources may be arbitrarily correlated. In Section 3.3, we propose an iterative algorithm to solve the SMR problem. We consider both ULA without missing sensors and ULA wherein some sensors may be missing, in this section. We also connect the proposed SMR approach with the traditional MLE framework and the modern SBL formulation. In Section 3.4, we discuss the general case where sensors may be placed arbitrarily, and may not lie on a uniform grid. We present numerical results in Section 3.5 and conclude the work in Section 3.6.

We represent scalars, vectors, and matrices by lowercase, boldface-lowercase, and boldface-uppercase letters, respectively. Sets are represented using blackboard bold letters. $(.)^T$ denotes transpose and $(.)^H$ denotes Hermitian of the operand matrix, and $(.)^c$ denotes element-wise complex conjugate. $\odot$ denotes Khatri-Rao product between two matrices of appropriate sizes.

3.2 SBL Revisited: Correlation Aware Interpretation, Robustness, and Structured Matrix Reformulation

A correlation-aware technique [85, 104] satisfies the following three general requirements[3]: a) it depends on the measurements only through its second order statistics b) it assumes a source correlation prior, usually that sources are uncorrelated, and fits a resulting structured received signal covariance matrix to the second order statistics of the measurements c) any further inference is carried using the recovered parameters characterizing the estimated structured covariance matrix. In this work we assume that the sources are uncorrelated. This assumption may not always hold, and some sources may in fact be correlated. The impact of this mismatch between assumed model and true model is discussed at the end of this section. The discussion highlights another aspect of the MLE framework, as it provides interpretable and superior results even in the mismatched model case.

For the purpose of simplicity, we focus on the ULA geometry in this section, and postpone the general case of ULAs with missing sensors until next section. However, the insights presented here are applicable to the general case as well.

3.2.1 On the SBL Algorithm

SBL is a Bayesian technique to find a row-sparse decomposition of the received measurements, $\mathbf{Y} = [\mathbf{y}_0, \ldots, \mathbf{y}_{L-1}]$, (i.e., to solve the MMV problem in (3.2)) using an overcomplete dictionary $\mathbf{\Phi} \in \mathbb{C}^{M \times G}$ consisting of G suitably chosen vectors (may be non-parametric in general). In the DoA estimation problem, these vectors are array manifold vectors evaluated on a grid of angular space representing potential DoAs i.e., $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2})$ or $u \in [-1, 1)$ in u-space. Note that there is a bijective mapping $u = \sin\theta$ in the domains of interest [56] and thus we use the two notations interchangeably. Consider a ULA

[3]To our knowledge, a formal set of requirements to be a 'correlation-aware' technique is missing in literature. Thus, we propose these requirements based on the conditions for superior source identifiability reported in [85, 86, 104].

with M sensors and $d = \bar{\lambda}/2$ distance between adjacent sensors to prevent ambiguity in DoA estimation; $\bar{\lambda}$ denotes the wavelength of the incoming narrowband source signals. The array manifold vector for a source signal incoming at angle $u \in [-1,1)$, is given by $\boldsymbol{\phi}(u) = [1, \exp(-j\pi u), \ldots, \exp(-j(M-1)\pi u)]^T$.

SBL imposes a parameterized Gaussian prior on the source signal $\bar{\mathbf{x}}_l \in \mathbb{C}^G$ as $\bar{\mathbf{x}}_l \sim \mathscr{CN}(\mathbf{0}, \boldsymbol{\Gamma})$. Note that SBL explicitly imposes an uncorrelated sources prior, and thus $\boldsymbol{\Gamma}$ is a diagonal matrix; let $\mathbf{diag}(\boldsymbol{\Gamma}) = \boldsymbol{\gamma}$. Thus we have $\mathbf{y}_l \sim \mathscr{CN}(\mathbf{0}, \boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^H + \lambda \mathbf{I})$, λ denotes the estimate for noise variance. In the case with uninformative prior for $\boldsymbol{\gamma}$, the hyperparameter $\boldsymbol{\Gamma}$ and λ can be estimated under the MLE framework [27] as

$$\min_{\boldsymbol{\Gamma} \succeq \mathbf{0}, \lambda \geq 0} \log\det\left(\boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^H + \lambda \mathbf{I}\right) + \mathrm{tr}\left(\left(\boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^H + \lambda \mathbf{I}\right)^{-1} \hat{\mathbf{R}}_{\mathbf{y}}\right), \tag{3.3}$$

where $\hat{\mathbf{R}}_{\mathbf{y}} = \frac{1}{L}\sum_{l=0}^{L-1} \mathbf{y}_l \mathbf{y}_l^H$ denotes the SCM. Choices for solving the problem in (3.3) include the Tipping iterations [26], EM iterations [12], sequential SBL [13], and generalized approximate message passing (GAMP) implementations [105, 106]. A MM approach for solving (3.3) was introduced in [107].

Remark 14. Note that if the number of sources K is known exactly in (3.1), such model order information is not used in the SBL formulation. Instead, the $\log\det$ penalty in (3.3) helps to promote sparsity and to deal with small but unknown number of sources. If there is prior knowledge on K, then $\|\boldsymbol{\gamma}\|_0 = K$ would have to be imposed on the objective function.

We now present the following useful insight.

Proposition 1. $\forall \boldsymbol{\gamma} \geq \mathbf{0}$ such that $(\boldsymbol{\Phi} \odot \boldsymbol{\Phi}^c)\boldsymbol{\gamma} = \mathbf{w}$, for some fixed $\mathbf{w} \in \mathbb{C}^{M^2}$, the SBL cost is a constant i.e.,

$$\log\det\left(\boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^H + \lambda \mathbf{I}\right) + \mathrm{tr}\left(\left(\boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^H + \lambda \mathbf{I}\right)^{-1} \hat{\mathbf{R}}_{\mathbf{y}}\right) = C(\lambda),$$

where $C(\lambda)$ is some constant.

Proof. The proof follows simply by observing that $(\mathbf{\Phi} \odot \mathbf{\Phi}^c)\boldsymbol{\gamma} = \mathbf{w}$ implies $\mathbf{\Phi\Gamma\Phi}^H$ is a fixed structured matrix with entries dictated by components of $\mathbf{w}$. □

The above result demonstrates that, the hyperparameter $\boldsymbol{\gamma}$ affects the SBL cost function only through the entries of the structured covariance matrix of the measurements. The sources are localized by peaks in the output $\boldsymbol{\gamma}$ pseudospectrum. This procedure satisfies the general requirements for correlation-aware algorithms. Thus, we conclude that SBL is indeed a correlation-aware technique. The procedure also marks some key requirements for superior sources' identifiability (see Theorem 1 and following remarks in [86]).

3.2.2 Connecting to Correlation-Aware SSR Techniques based on Minimizing Diversity Measures

Consider the class of problems given by

$$\begin{aligned} \min_{\mathbf{z} \geq \mathbf{0}} \quad & f(\mathbf{z}) \\ \text{subject to} \quad & \|\hat{\mathbf{r}}_{\mathbf{y}} - \mathbf{\Phi}_{KR}\mathbf{z}\|_2 \leq \varepsilon, \end{aligned} \tag{3.4}$$

where, $\hat{\mathbf{r}}_{\mathbf{y}} = \text{vec}(\hat{\mathbf{R}}_{\mathbf{y}})$ and $\mathbf{\Phi}_{KR} = \mathbf{\Phi}^c \odot \mathbf{\Phi}$ denotes the Khatri-Rao product of $\mathbf{\Phi}$ with its conjugate. $f(\mathbf{z})$ is a sparsity promoting objective function and choices include ℓ_1 norm, ℓ_0 or $\ell_{1/2}$ as considered in [108]. The above problem satisfies the requirements for being correlation-aware, namely a) it matches the model to the second order statistics of the data b) uses uncorrelated sources' correlation prior to fit a structured matrix to the measurements c) further performs inference using the parameters of this estimated structured matrix. Next, we reformulate SBL as a constrained optimization problem to highlight the data-fitting term and to compare with (3.4).

The MLE optimization problem in (3.3) can be reformulated as a constrained optimization

problem as follows:

$$\begin{aligned} \min_{\boldsymbol{\Gamma}\succeq\mathbf{0},\lambda\geq 0} \quad & \log\det\left(\boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^H+\lambda\mathbf{I}\right) \\ \text{subject to} \quad & \operatorname{tr}\left(\left(\boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^H+\lambda\mathbf{I}\right)^{-1}\hat{\mathbf{R}}_{\mathbf{y}}\right)\leq\varepsilon. \end{aligned} \tag{3.5}$$

Note that the constraint imposes a Mahalanobis distance-based bound on the optimization variables. Another perspective to understand the data-fitting term above based on regularized least-squares fit to measurements can be found in [107].

Proposition 2. Let $(\boldsymbol{\Gamma}^*,\lambda^*)$ be a global minimizer of the optimization problem in (3.3) such that $\lambda^*>0$. $(\boldsymbol{\Gamma}^*,\lambda^*)$ globally minimizes problem in (3.5) as well, if and only if $\varepsilon=\operatorname{tr}\left(\left(\boldsymbol{\Phi}\boldsymbol{\Gamma}^*\boldsymbol{\Phi}^H+\lambda^*\mathbf{I}\right)^{-1}\hat{\mathbf{R}}_{\mathbf{y}}\right)$.

Proof. Proof in Appendix section 3.7.1. □

The constraint in the formulation of (3.5) allows to match the model to the observation (through the sample covariance matrix) and the objective function promotes a simpler model to be picked. Note that the constrained optimization problem in (3.5) is exactly MLE only when ε is set appropriately. The proposition indicates the difficulty in transforming the MLE to a constrained problem, although the latter can be explored as a viable option with ε set heuristically. This is not discussed further and left as future work. The above outlook only tries to highlight the two components of the SBL objective and allows one to compare the constrained formulation in (3.5) to the other correlation-aware technique in (3.4). The data fitting term in (3.4) lacks the MLE framework for data fitting used in (3.5). This insight highlights one of the key difference between our approach and that used in many other works in the literature.

An alternative treatment of the SBL cost function that also reveals connections to reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions to (3.2) can be found in [28,107, 109].

3.2.3 Proposed SMR Approach: ULA with No Missing Sensors

The structure for $\mathbf{\Phi\Gamma\Phi}^H$ in the case of ULA is a Toeplitz matrix, and is informed by the array geometry and the uncorrelated sources prior. In other words, SBL attempts to find the 'best' positive semidefinite (PSD) Toeplitz matrix approximation to the SCM $\hat{\mathbf{R}}_{\mathbf{y}}$. The grid-based formulation restricts the solution to lie in the union of PSD cones. We use this insight and reparameterize the SBL cost function to directly estimate the entries of the Toeplitz covariance matrix. Let $\mathbf{v}$ denote the first row of such a Toeplitz matrix, denoted by $\mathbf{Toep}(\mathbf{v})$. We reformulate the SBL optimization problem as

$$\min_{\substack{\mathbf{v}\in\mathbb{C}^M \text{ s.t.} \\ \mathbf{Toep}(\mathbf{v})\succeq\mathbf{0},\lambda\geq 0}} \log\det\left(\mathbf{Toep}(\mathbf{v})+\lambda\mathbf{I}\right)+\mathrm{tr}\left(\left(\mathbf{Toep}(\mathbf{v})+\lambda\mathbf{I}\right)^{-1}\hat{\mathbf{R}}_{\mathbf{y}}\right). \tag{3.6}$$

Once the solution $\mathbf{v}^*$ is obtained, we estimate the DoAs by decomposing the Toeplitz matrix, $\mathbf{Toep}(\mathbf{v}^*)$. In our simulations we use root-MUSIC to estimate the DoAs [110].

Remark 15. It is known that a low rank ($D<M$) PSD Toeplitz matrix such as $\mathbf{Toep}(\mathbf{v}^*)$ can be uniquely decomposed as $\mathbf{Toep}(\mathbf{v}^*)=\sum_{i=1}^{D} p_i\boldsymbol{\phi}(\theta_i)\boldsymbol{\phi}(\theta_i)^H, p_i>0$, and θ_i's are distinct [111]. In (3.6), a low-rank solution is encouraged by the $\log\det$ term [112], while its effect is being moderated by the additional noise variance term, '$+\lambda\mathbf{I}$'.

The SBL formulation in (3.3) not only finds a structured matrix fit to the measurements, it also factorizes it. The same is true with the classical MLE approach, and is briefly discussed in Section 3.3.3. The structured matrix factorization is a crucial step. In the proposed approach, we find a structured matrix in the MLE sense. We therefore refer to the proposed approach as 'StructCovMLE'. The problem in (3.6) is non-convex and we discuss an iterative algorithm to solve it, along with an extension to allow ULAs with missing sensors, in Section 3.3.

Next, we briefly discuss an important aspect of the chosen approach in (3.6) to solve the original problem in (3.1).

3.2.4 Performance under a Correlation Prior Mismatch

We discuss the case when there is a prior misfit, between the assumed model and the actual (data) model. This insight is another feature resulting from the MLE formulation used by SBL as opposed to a regularization framework. In particular, we discuss the case when the sources may be arbitrarily correlated. As briefly mentioned before, in the case of a ULA, the structure SBL imposes by virtue of the array geometry and the (uncorrelated) source correlation prior is a Toeplitz matrix. If some of the sources are correlated, the approach fits Toeplitz structured covariance to a non-Toeplitz structure obeyed by the data. Our aim is not to correct but to quantify the model misfit. In particular, we show that the recovered Toeplitz fit to the SCM minimizes the KL divergence between the assumed and the true distribution.

Let $p_{\mathbf{y}}$ and $f_{\mathbf{y}|\boldsymbol{\Psi}}$ denote the true probability density function (pdf) and the pdf for the mismatched model, respectively, where $\boldsymbol{\Psi} = (\mathbf{v}, \lambda)$ s.t. $\text{Toep}(\mathbf{v}) \succeq \mathbf{0}, \lambda \geq 0$. Since the source and noise vectors are uncorrelated with each other, $p_{\mathbf{y}}$ is a zero mean Gaussian pdf with covariance matrix $\mathbf{R}_{\mathbf{y}} = \boldsymbol{\Phi}_{\boldsymbol{\theta}} \mathbf{R}_{\mathbf{x}} \boldsymbol{\Phi}_{\boldsymbol{\theta}}^H + \sigma_n^2 \mathbf{I}$, where $\mathbf{R}_{\mathbf{x}}$ denotes the source covariance matrix. Similarly $f_{\mathbf{y}|\boldsymbol{\Psi}}$ is zero mean Gaussian pdf with covariance $\boldsymbol{\Sigma}_{\boldsymbol{y}} = \text{Toep}(\mathbf{v}) + \lambda \mathbf{I}$. The KL divergence between these two normal distributions is well known and is given by

$$D(p_{\mathbf{y}} \| f_{\mathbf{y}|\boldsymbol{\Psi}}) = \log\det \boldsymbol{\Sigma}_{\boldsymbol{y}} - \log\det \mathbf{R}_{\mathbf{y}} - M + \text{tr}(\boldsymbol{\Sigma}_{\boldsymbol{y}}^{-1} \mathbf{R}_{\mathbf{y}}). \tag{3.7}$$

The effective optimization problem to minimize the KL divergence between the two distributions is given by

$$\boldsymbol{\Psi}^* = \underset{\boldsymbol{\Psi} \text{ s.t. } \text{Toep}(\mathbf{v}) \succeq \mathbf{0}, \lambda \geq 0}{\text{argmin}} \log\det\big(\boldsymbol{\Sigma}_{\boldsymbol{y}}\big) + \text{tr}\big(\boldsymbol{\Sigma}_{\boldsymbol{y}}^{-1} \mathbf{R}_{\mathbf{y}}\big). \tag{3.8}$$

Note that this optimization problem is similar to (3.6) for the proposed approach (or (3.3) used within SBL), where instead of the actual received signal covariance matrix, $\mathbf{R}_{\mathbf{y}}$, we

used the SCM. Note that the SCM is the unconstrained/unstructured MLE estimate of the received signal covariance matrix. In [113] it was shown for SBL using the two sources example and when the DoAs were known that, when sources are far apart, the estimate for the source powers under the uncorrelated model matches the true source power using the problem in (3.8). Such a mismatched model was also used in [114] to propose more robust beamformers that can resist source correlation.

3.3 Maximum Likelihood Structured Covariance Matrix Recovery

We focus on ULA, first on the case with no missing sensors, and then on the case of ULA with missing sensors. We assume that the noise variance is known and set $\lambda = \sigma_n^2$ in (3.6), but it can be estimated as well, similar to $\mathbf{v}$ in this section.

3.3.1 Uniform Linear Array Geometry

Based on the concavity of the $\log\det$ term, we majorize the $\log\det$ term in (3.6) and replace it with a linear term using its Taylor expansion [87]

$$\begin{aligned}\log\det\left(\text{Toep}(\mathbf{v})+\lambda\mathbf{I}\right) \leq & \log\det\left(\text{Toep}(\mathbf{v}^{(k)})+\lambda\mathbf{I}\right) \\ & +\text{tr}\left((\text{Toep}(\mathbf{v}^{(k)})+\lambda\mathbf{I})^{-1}\text{Toep}(\mathbf{v}-\mathbf{v}^{(k)})\right), \end{aligned} \tag{3.9}$$

where $\mathbf{v}^{(k)}$ denotes the iterate value at the kth iteration. Note that the linear term from Taylor expansion provides a supporting hyperplane to the hypograph $\{(\mathbf{v},t) : t <= \log\det(\text{Toep}(\mathbf{v})+\lambda\mathbf{I}))\}$ [115]. We ignore the constant terms above and get the following majorized objective function

$$\text{tr}\left((\text{Toep}(\mathbf{v}^{(k)})+\lambda\mathbf{I})^{-1}\text{Toep}(\mathbf{v})\right)+\text{tr}\left((\text{Toep}(\mathbf{v})+\lambda\mathbf{I})^{-1}\hat{\mathbf{R}}_{\mathbf{y}}\right).$$

Rewriting second term above using Schur complement lemma:

$$\begin{aligned}\operatorname{tr}\left((\text{Toep}(\mathbf{v})+\lambda\mathbf{I})^{-1}\hat{\mathbf{R}}_{\mathbf{y}}\right) = \min_{\mathbf{U}\in\mathbb{C}^{M\times M}} & \operatorname{tr}\left(\mathbf{U}\hat{\mathbf{R}}_{\mathbf{y}}\right) \\ \text{s.t.} & \begin{bmatrix} \mathbf{U} & \mathbf{I}_M \\ \mathbf{I}_M & \text{Toep}(\mathbf{v})+\lambda\mathbf{I} \end{bmatrix} \succeq \mathbf{0}, \end{aligned} \tag{3.10}$$

which is a SDP. The overall optimization problem is convex and can be formulated as a SDP as follows

$$\begin{aligned}&\min_{\mathbf{v}\in\mathbb{C}^M,\mathbf{U}\in\mathbb{C}^{M\times M}} \operatorname{tr}\left((\text{Toep}(\mathbf{v}^{(k)})+\lambda\mathbf{I})^{-1}\text{Toep}(\mathbf{v})\right)+\operatorname{tr}\left(\mathbf{U}\hat{\mathbf{R}}_{\mathbf{y}}\right) \\ &\text{subject to} \quad \begin{bmatrix} \mathbf{U} & \mathbf{I}_M \\ \mathbf{I}_M & \text{Toep}(\mathbf{v})+\lambda\mathbf{I} \end{bmatrix} \succeq \mathbf{0}, \text{Toep}(\mathbf{v}) \succeq \mathbf{0}, \end{aligned} \tag{3.11}$$

and can be solved using any standard solvers (e.g. CVX solvers such as SDPT3, SeDuMi [116]). It can be solved iteratively and we summarize the proposed steps in Algorithm 6. The following remark briefly discusses the choice of initialization.

Algorithm 6: Proposed 'StructCovMLE' Algorithm

Result: $\mathbf{v}^*$
Input: Measurements: $\mathbf{Y}\in\mathbb{C}^{M\times L}, \lambda=\sigma_n^2, \text{ITER}$
Initialize: $\hat{\mathbf{R}}_{\mathbf{y}}=\mathbf{Y}\mathbf{Y}^H/L, \mathbf{v}^*=\mathbf{e}_1=[1,0,\ldots,0]^T$
for $k:=1$ to ITER do
    $\mathbf{v}^{(k)}\leftarrow\mathbf{v}^*$
    $\mathbf{v}^*\leftarrow$ Solve the problem in (3.11)
end

Remark 16. We initialize the proposed algorithm with the unit vector $\mathbf{v}_0=\mathbf{e}_1$, following the suggestion in [112] for effective rank minimization. This initialization reduces the majorized term to a trace function in the first iteration. It is known that trace function is a convex envelope for the rank function for matrices with spectral norm less than one [117]. Furthermore, the iterative weighted trace minimization in the following iterations helps

to preserve relevant signal components.

A criterion such as stopping when the relative change ($\|\mathbf{v}^{(k)} - \mathbf{v}^{(k-1)}\|_2 / \|\mathbf{v}^{(k-1)}\|_2$) is small may be considered in Algorithm 6, instead of a fixed number of iterations. The fixed number of iterations approach is used to compare different iterative algorithms in Section 3.5.

3.3.2 ULA with Missing Sensors

We begin by identifying the relevant structure for the general case of ULAs with missing sensors. Consider a linear array with M sensors on a grid with minimum inter-element spacing $d = \bar{\lambda}/2$. Let $\mathbb{P} = \{p_i \mid p_i \in \mathbb{Z}, 0 \leq i < M\}$ denote the set of normalized (w.r.t. d) sensor positions. We assume $p_0 = 0$ without loss of generality. The array manifold vector is given by $\boldsymbol{\phi}(u) = [1, \exp(-jp_1\pi u), \ldots, \exp(-jp_{M-1}\pi u)]^T, u = \sin\theta$. The difference coarray is given by $\mathbb{D} = \{z \mid z = r - s,\, r,s \in \mathbb{P}\}$. The concept of difference coarray influences the structure we seek to identify, and also arises naturally when computing the received signal covariance matrix. It represents the set of unique lags experienced by the physical array. The received signal covariance matrix under the SBL formulation is given by $\boldsymbol{\Phi\Gamma\Phi}^H + \lambda\mathbf{I}$, as discussed previously. The (m,n) entry in $\boldsymbol{\Phi\Gamma\Phi}^H$ is given by $[\boldsymbol{\Phi\Gamma\Phi}^H]_{m,n} = \sum_{i=1}^{G} \gamma_i \exp(-j(p_m - p_n)\pi u_i)$, and $[\boldsymbol{\Phi\Gamma\Phi}^H]_{m,n} = [\boldsymbol{\Phi\Gamma\Phi}^H]^c_{n,m}$. Thus, $[\boldsymbol{\Phi\Gamma\Phi}^H]_{m,n} = [\boldsymbol{\Phi\Gamma\Phi}^H]_{m',n'}, \forall$ tuples (m,n) and (m',n') such that $p_m - p_n = p_{m'} - p_{n'}$. In other words, the entries in $\boldsymbol{\Phi\Gamma\Phi}^H$ can be distinct only corresponding to distinct elements in $\mathbb{D}$. $\boldsymbol{\Phi\Gamma\Phi}^H$ is Hermitian symmetric, which further restricts the number of distinct entries. This reveals the underlying structure that the model $\boldsymbol{\Phi\Gamma\Phi}^H$ satisfies, and we formalize it below.

Let M_{apt} denote the aperture of the array, $M_{\text{apt}} = \max_{d\in\mathbb{D}} d + 1$. We define a linear mapping $\mathbf{T}(\mathbf{v}) : \mathbb{C}^{M_{\text{apt}}} \to \mathbb{C}^{M\times M}$ as

$$[\mathbf{T}(\mathbf{v})]_{i,j} = \begin{cases} v_{|p_i - p_j|} & j \geq i \\ v^c_{|p_i - p_j|} & \text{otherwise} \end{cases}, 0 \leq i, j < M. \tag{3.12}$$

$$\mathbf{T}(\mathbf{v}) = \begin{bmatrix} v_0 & v_1 & v_3 \\ v_1^c & v_0 & v_2 \\ v_3^c & v_2^c & v_0 \end{bmatrix} \quad \mathbf{T}(\mathbf{v}) = \begin{bmatrix} v_0 & v_1 & v_4 \\ v_1^c & v_0 & v_3 \\ v_4^c & v_3^c & v_0 \end{bmatrix}$$

(a) (b)

Figure 3.1. Structured Covariance Matrix $\mathbf{T}(\mathbf{v})$

The mapping $\mathbf{T}(\mathbf{v})$ in general is many-to-one. It is only when the difference coarray has no holes, the mapping is one-to-one. For such cases we define $\mathbf{T}^{-1}(\mathbf{R}) : \mathbb{C}^{M\times M} \to \mathbb{C}^{M_{\text{apt}}}$ as a function that extracts the entries of a given structured matrix $\mathbf{R}$, formed using (3.12), to form a column vector. For the ULA with no missing sensors' case, we have $\mathbf{T}(\mathbf{v}) = \text{Toep}(\mathbf{v})$.

Example 1: Consider $\mathbb{P} = \{0,1,3\}$. This leads to $\mathbb{D} = \{-3,-2,-1,0,1,2,3\}$ and $M_{\text{apt}} = 4$. We therefore define $\mathbf{v} \in \mathbb{C}^4$, and the structured coavariance matrix as in Fig. 3.1(a). The mapping $\mathbf{T}(\mathbf{v})$ is one-to-one here and consequently $\mathbf{T}^{-1}$ is defined and we have $\mathbf{T}^{-1}(\mathbf{T}(\mathbf{v})) = [v_0, v_1, v_2, v_3]^T$.

Example 2: Consider $\mathbb{P} = \{0,1,4\}$. This leads to $\mathbb{D} = \{-4,-3,-1,0,1,3,4\}$ and $M_{\text{apt}} = 5$. We therefore define $\mathbf{v} \in \mathbb{C}^5$, and the structured coavariance matrix as in Fig. 3.1(b). The mapping $\mathbf{T}(\mathbf{v})$ is many-to-one here, as the component v_2 is missing in $\mathbf{T}(\mathbf{v})$. Consequently $\mathbf{T}^{-1}$ is not defined.

Thus, for the general case, (3.3) can be reformulated as:

$$\min_{\substack{\mathbf{v}\in\mathbb{C}^{M_{\text{apt}}} \text{ s.t.} \\ \text{Toep}(\mathbf{v})\succeq\mathbf{0}, \lambda\geq 0}} \log\det\left(\mathbf{T}(\mathbf{v}) + \lambda\mathbf{I}\right) + \text{tr}\left((\mathbf{T}(\mathbf{v}) + \lambda\mathbf{I})^{-1}\hat{\mathbf{R}}_{\mathbf{y}}\right). \tag{3.13}$$

Remark 17. We would like to highlight a non-trivial choice made above of imposing $\text{Toep}(\mathbf{v}) \succeq \mathbf{0}$, instead of only requiring $\mathbf{T}(\mathbf{v}) \succeq \mathbf{0}$. Note that the former constraint ensures that the latter is satisfied. The choice imposes a relevant constraint and is an important aspect of the model we wish to fit to the data in MLE sense. It also helps to connect the proposed reformulation to the traditional and modern MLE approaches, and is discussed in Section 3.3.3.

Remark 18. As in the case for SBL, if the number of sources, K, is known, a rank constraint $\mathbf{rank}(\mathbf{Toep}(\mathbf{v})) = K$ should be imposed. Since imposing a rank constraint is difficult, surrogate measures like in compressed sensing may be used, such as '$+\beta \log\det(\mathbf{Toep}(\mathbf{v}) + \varepsilon\mathbf{I})$' as a regularizer in (3.13) to further promote sparse solutions. In this work, we do not exploit knowledge of K to solve (3.13).

Like in the previous case of ULA with no missing sensors, we majorize the cost function in (3.13) to get a convex function and rewrite it as a SDP, assuming knowledge of noise variance and setting $\lambda = \sigma_n^2$. The majorized objective is given by

$$\mathrm{tr}\left((\mathbf{T}(\mathbf{v}^{(k)}) + \lambda\mathbf{I})^{-1}\mathbf{T}(\mathbf{v})\right) + \mathrm{tr}\left((\mathbf{T}(\mathbf{v}) + \lambda\mathbf{I})^{-1}\hat{\mathbf{R}}_{\mathbf{y}}\right). \tag{3.14}$$

The resulting SDP is given below

$$\begin{aligned} &\min_{\mathbf{v}\in\mathbb{C}^{M_{\text{apt}}}, \mathbf{U}\in\mathbb{C}^{M\times M}} \mathrm{tr}\left((\mathbf{T}(\mathbf{v}^{(k)}) + \lambda\mathbf{I})^{-1}\mathbf{T}(\mathbf{v})\right) + \mathrm{tr}\left(\mathbf{U}\hat{\mathbf{R}}_{\mathbf{y}}\right) \\ &\text{subject to} \quad \begin{bmatrix} \mathbf{U} & \mathbf{I}_M \\ \mathbf{I}_M & \mathbf{T}(\mathbf{v}) + \lambda\mathbf{I} \end{bmatrix} \succeq \mathbf{0}, \mathbf{Toep}(\mathbf{v}) \succeq \mathbf{0}, \end{aligned} \tag{3.15}$$

where $\mathbf{v}^{(k)}$ denotes the value at the kth iteration. Steps similar to Algorithm 6 can be followed to find the optimal point $\mathbf{v}^*$. To estimate the DoAs we perform root-MUSIC on $\mathbf{T}(\mathbf{v}^*)$.

Remark 19. It was shown in [118] that sparse arrays with a larger number of consecutive lags than the number of sensors, M, can identify more sources than M. Under the proposed approach, a similar higher identifiability can be achieved by instead performing root-MUSIC on $\mathbf{Toep}(\mathbf{v}^*)$, and we numerically verify this in Section 3.5.2.

3.3.3 On Proposed Method: From MLE to SBL

We connect the proposed technique with the classical MLE framework and the grid SBL formulation. We hope to answer the following question: how has the reparameterization affected the original problem in (3.1) of solving for $\boldsymbol{\theta}$?

1) Connection with the classical MLE formulation: We begin by first stating the traditional MLE formulation. In this approach, we impose a parametrized Gaussian prior on $\mathbf{x}_l$ i.e., $\mathbf{x}_l \sim \mathscr{CN}(\mathbf{0}, \mathbf{P})$. Note that an explicit knowledge of model order information is a requisite here. We further assume that the sources are uncorrelated, and thus $\mathbf{P}$ is a diagonal matrix. The resulting optimization problem is given by

$$\min_{\boldsymbol{\theta} \in [-\frac{\pi}{2}, \frac{\pi}{2})^K, \mathbf{P} \succ \mathbf{0}, \lambda \geq 0} \log\det\left(\boldsymbol{\Phi}_{\boldsymbol{\theta}} \mathbf{P} \boldsymbol{\Phi}_{\boldsymbol{\theta}}^H + \lambda \mathbf{I}\right) + \mathrm{tr}\left((\boldsymbol{\Phi}_{\boldsymbol{\theta}} \mathbf{P} \boldsymbol{\Phi}_{\boldsymbol{\theta}}^H + \lambda \mathbf{I})^{-1} \hat{\mathbf{R}}_{\mathbf{y}}\right). \tag{3.16}$$

The model is also referred to as the unconditional model in the DoA literature [119], compared to the conditional model where $\mathbf{x}_l$ is assumed deterministic. Consider the following updated MLE optimization problem:

$$\min_{\substack{K \in \mathbb{Z}^+ \\ 0 < K < M_{\text{apt}}}} \min_{\boldsymbol{\theta} \in [-\frac{\pi}{2}, \frac{\pi}{2})^K, \mathbf{P} \succ \mathbf{0}, \lambda \geq 0} \log\det\left(\boldsymbol{\Phi}_{\boldsymbol{\theta}} \mathbf{P} \boldsymbol{\Phi}_{\boldsymbol{\theta}}^H + \lambda \mathbf{I}\right) + \mathrm{tr}\left((\boldsymbol{\Phi}_{\boldsymbol{\theta}} \mathbf{P} \boldsymbol{\Phi}_{\boldsymbol{\theta}}^H + \lambda \mathbf{I})^{-1} \hat{\mathbf{R}}_{\mathbf{y}}\right). \tag{3.17}$$

The difference with the traditional MLE formulation is that, in the above we consider all model orders, $0 < K < M_{\text{apt}}$, to optimize the cost function. We then have the following result.

Theorem 4. The problem in (3.13) and in (3.17) are equivalent, in that they achieve the same globally minimum cost.

Proof. Proof is provided in the Appendix section 3.7.2. □

2) Connection with SBL in (3.3): Consider the following updated SBL optimization prob-

lem:

$$\min_{\boldsymbol{\Phi}} \min_{\boldsymbol{\Gamma}\succeq\mathbf{0},\lambda\geq 0} \log\det\left(\boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^H + \lambda\mathbf{I}\right) + \mathrm{tr}\left(\left(\boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^H + \lambda\mathbf{I}\right)^{-1}\hat{\mathbf{R}}_{\mathbf{y}}\right), \tag{3.18}$$

where we also allow all possible dictionaries $\boldsymbol{\Phi}$ with array manifold vectors as columns, to optimize the cost function. The following result follows similarly.

Theorem 5. The problem in (3.13) and in (3.18) are equivalent, in that they achieve the same globally minimum cost.

Proof. The proof follows similarly as for Theorem 4, and we present it it Appendix section 3.7.3 for completion. □

The above results help to understand the proposed approach in (3.13): (3.13) estimates a structured covariance matrix fit to the measurements in the MLE sense over all model orders for classical MLE or all appropriate dictionaries for SBL.

The entries of a structured matrix and noise variance may be combined as presented in [120]. However, the choice of explicitly involving λ parameter has two important consequences: a) If σ_n^2 is known, the proposed approach allows a mechanism to feed this information, which is absent in [120] b) If σ_n^2 is unknown, a better learning strategy to estimate the noise variance and then feeding it as part of the model may result in better DoA estimates than jointly estimating $\boldsymbol{\theta}$ and σ_n^2. Finally, although the optimization problem in [120] and the proposed are similar, an algorithm for solving it is missing in [120]. During the preparation of this manuscript we came across another recent work in [121] which derives from the classical MLE formulation. Consequently, it involves the non-linear rank constraint which is implemented by a truncated eigen-decomposition step. In contrast, the presented algorithm builds on the success of SBL algorithm and relaxes the rank constraint, similar to SBL. The presented approach also guarantees that the likelihood increases over the iterations. K is utilized for root-MUSIC. [121] focuses on sparse linear arrays i.e., sensors on grid. We, however, consider the general non-uniform

linear array case as well, and is discussed next.

3.4 Gridless SBL with Likelihood-based Grid Refinement

In this section, we consider the case when sensors may be placed arbitrarily on a linear aperture. The presented ideas can be extended to other shapes or higher dimensional (2D, 3D, etc.) geometries. Note in both the sections we assume that the sensor positions are known. Issues concerning calibration errors is not the focus here, and we request interested readers to check the relevant literature for tackling such issues [56].

Consider an array with sensor positions $\mathbb{P} = \{0, 1, 2.1, 3.5\ , 4.7, 10\}$. The difference coarray for this geometry is given by $\mathbb{D} = \{0, 1, 1.1, 1.2, 1.4, 2.1, 2.5, 2.6, 3.5, 3.7, 4.7, 5.3\ , 6.5, 7.9\ , 9, 10\}$. The structured received signal covariance matrix is neither Toeplitz, nor is sampled from a higher order Toeplitz matrix. This implies that (3.13), where we enforced a Toeplitz PSD constraint, is not applicable. Similarly, the second step wherein we estimate DoAs in a gridless manner using root-MUSIC is not applicable. Finally, the number of distinct lags is $|\mathbb{D}| = M(M+1)/2 - (M-1) = 16$ which essentially enforces structure only on the diagonal entries, in that, they be equal. This indicates poor availability of structural constraints on the received signal covariance, compared to the geometries where sensors are present on a uniform grid. Note that the SBL formulation in (3.3) is devoid of such limitations. Using the recovered $\mathbf{\Gamma}^*$ one can construct a Toeplitz matrix of order $\lfloor M_{\text{apt}} \rfloor$, where $\lfloor \cdot \rfloor$ indicates the floor function, and beyond although the accuracy may not be reliable as the measurements lack information about larger lags. This highlights the versatility with which SBL can handle arbitrary array geometries. However, as we already know that SBL does not quite solve (3.1) that we ventured out to solve in the first place, because the DoAs may not lie on the chosen grid. One can employ a very fine grid, but the per iteration computational complexity increases linearly with the grid size. We extend the SBL procedure to progressively refine the initial uniform coarse grid

by adding more points near potential source locations. We achieve this in two steps: (a) Grid point adjustment around peaks, in the solution $\boldsymbol{\gamma}^*$ of (3.3) using sequential SBL [13] to simultaneously update both grid point and power estimate (b) Multi-resolution grid refinement. Note that the latter builds on the former step by re-running SBL after the local step (a), pruning, and increasing grid resolution near top peaks in the $\boldsymbol{\gamma}$ pseudospectrum. As will be shown next, the grid point adjustment around peaks in $\boldsymbol{\gamma}$ pseudospectrum is a computationally simpler procedure to further increase the likelihood after SBL iterations on a coarse grid.

3.4.1 Grid Point Adjustment around Peaks in Solution

We begin by rewriting the SBL objective function to separate out the i-th grid component characterized by the tuple (γ_i, u_i); $u = \sin\theta$ is used here. Let $\mathbf{C} = \boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^H + \lambda\mathbf{I}$ and $\mathbf{C}_{-i} = \boldsymbol{\Phi}_{-i}\boldsymbol{\Gamma}_{-i}\boldsymbol{\Phi}_{-i}^H + \lambda\mathbf{I}$, where $\boldsymbol{\Phi}_{-i}$ denotes the dictionary without the i-th column in $\boldsymbol{\Phi}$, and $\boldsymbol{\Gamma}_{-i}$ denotes the matrix without the the i-th row and the i-th column in $\boldsymbol{\Gamma}$. Then

$$\mathscr{L}(\boldsymbol{\gamma}) = \log\det\mathbf{C} + \mathrm{tr}\left(\mathbf{C}^{-1}\hat{\mathbf{R}}_{\mathbf{y}}\right) = \mathscr{L}(\boldsymbol{\gamma}_{-i}) + L(\gamma_i, u_i), \tag{3.19}$$

where $\mathscr{L}(\boldsymbol{\gamma}_{-i}) = \log\det\mathbf{C}_{-i} + \mathrm{tr}\left(\mathbf{C}_{-i}^{-1}\hat{\mathbf{R}}_{\mathbf{y}}\right)$ is devoid of (γ_i, u_i),
and $L(\gamma_i, u_i) = \log(1 + \gamma_i\boldsymbol{\Phi}_i^H\mathbf{C}_{-i}^{-1}\boldsymbol{\Phi}_i) - \frac{\boldsymbol{\Phi}_i^H\mathbf{C}_{-i}^{-1}\hat{\mathbf{R}}_{\mathbf{y}}\mathbf{C}_{-i}^{-1}\boldsymbol{\Phi}_i}{\gamma_i^{-1} + \boldsymbol{\Phi}_i^H\mathbf{C}_{-i}^{-1}\boldsymbol{\Phi}_i}$ (see eq. (18) in [13] for detailed derivation of (3.19)). Let $i_k^{(0)}, k \in \{1, \ldots, K\}$ denote the indices for the K top peaks in the $\boldsymbol{\gamma}$ pseudospectrum. The index superscript (.) in $i_k^{(0)}$ indicates the iteration number of the overall two step procedure, and will be discussed more in the next subsection. The idea is to fix the first term in RHS of (3.19) and minimize the objective $L(\boldsymbol{\gamma}, u)$ with respect to

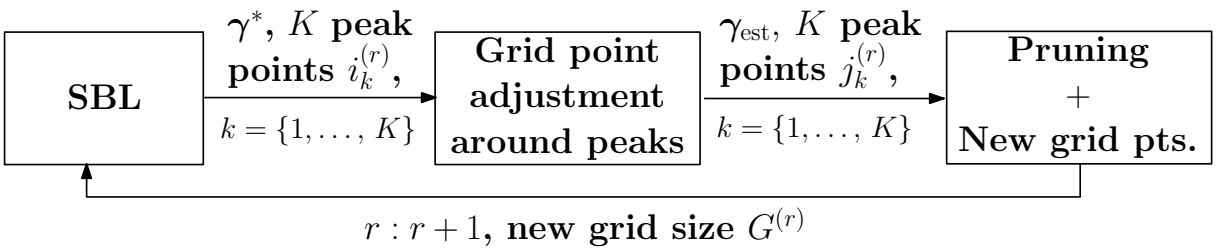


Figure 3.2. Proposed SBL with likelihood-based grid refinement procedure. At $r = 0$, SBL is run with a uniform grid.

$(\gamma, u), u \in [u_{i_k^{(0)}} - \delta, u_{i_k^{(0)}} + \delta], k = \{1, \ldots, K\}$, one peak at a time[4]; the bound[5] $\delta < 1/G$ is to avoid grid point overlap. In other words, the aim is to solve

$$\min_{u \in [u_{i_k^{(0)}} - \delta, u_{i_k^{(0)}} + \delta]} \min_{\gamma \geq 0} L(\gamma, u) = \log(1 + \gamma s(u)) - \frac{q(u)}{\gamma^{-1} + s(u)} \tag{3.20}$$

where $q(u) = \boldsymbol{\phi}(u)^H \mathbf{C}_{-i}^{-1} \hat{\mathbf{R}}_{\mathbf{y}} \mathbf{C}_{-i}^{-1} \boldsymbol{\phi}(u)$ and $s(u) = \boldsymbol{\phi}(u)^H \mathbf{C}_{-i}^{-1} \boldsymbol{\phi}(u)$. The minimization with respect to γ for a fixed u can be obtained in closed-form as

$$\gamma_{\text{opt}}(u) = \begin{cases} \frac{q(u) - s(u)}{s(u)^2} & q(u) > s(u) \\ 0 & q(u) \leq s(u) \end{cases}. \tag{3.21}$$

And thus we have

$$L(\gamma_{\text{opt}}(u), u) = \begin{cases} \log\left(\frac{q(u)}{s(u)}\right) - \frac{q(u)}{s(u)} + 1 & q(u) > s(u) \\ 0 & q(u) \leq s(u) \end{cases}. \tag{3.22}$$

Note that $\log\left(\frac{q(u)}{s(u)}\right) - \frac{q(u)}{s(u)} + 1 \leq 0, \forall u$, and is equal to zero only when $q(u) = s(u)$ for some u. Consequently, we are interested in $u \in [u_{i_k^{(0)}} - \delta, u_{i_k^{(0)}} + \delta]$ such that $q(u) > s(u)$. For such

[4] In general, solving for one grid point at a time (as done here) may lead to a different solution compared to solving for all grid points simultaneously. This is because the underlying cost function is non-convex. Since the objective is to perform local grid refinement where new grid points in the neighborhood of a support element are considered, a sequential SBL strategy of optimizing one grid point at a time is more suited, as it leads to simple yet efficient updates.

[5] Note that future iterations may involve non-uniform grid, and a similar bound on either directions is used to avoid grid point overlap.

points $L(\gamma_{\text{opt}}(u), u)$ is a monotonic non-increasing function of '$\frac{q(u)}{s(u)}$', and thus the problem in (3.20) reduces to the following problem

$$u^* = \underset{u \in [u_{i_k^{(0)}} - \delta, u_{i_k^{(0)}} + \delta] \text{ s.t. } q(u) > s(u)}{\arg\max} R(u) = \frac{q(u)}{s(u)}. \tag{3.23}$$

We provide the following perspective to understand the objective we wish to locally maximize.

$$R(u) = \frac{q(u)}{s(u)} = \frac{\boldsymbol{\phi}(u)^H \mathbf{C}_{-i}^{-1} \hat{\mathbf{R}}_{\mathbf{y}} \mathbf{C}_{-i}^{-1} \boldsymbol{\phi}(u)}{\boldsymbol{\phi}(u)^H \mathbf{C}_{-i}^{-1} \boldsymbol{\phi}(u)} = \frac{\boldsymbol{\phi}(u)^H \mathbf{C}_{-i}^{-1} \hat{\mathbf{R}}_{\mathbf{y}} \mathbf{C}_{-i}^{-1} \boldsymbol{\phi}(u) / \left(\boldsymbol{\phi}(u)^H \mathbf{C}_{-i}^{-1} \boldsymbol{\phi}(u)\right)^2}{1 / \left(\boldsymbol{\phi}(u)^H \mathbf{C}_{-i}^{-1} \boldsymbol{\phi}(u)\right)}, \tag{3.24}$$

which is the ratio of (numerator) actual beamforming total output power and (denominator) expected beamforming interference plus noise output power, where the model interference plus noise signal covariance is given by $\mathbf{C}_{-i}$. The expression utilizes the beamformer $\mathbf{w} = \frac{\mathbf{C}_{-i}^{-1} \boldsymbol{\phi}(u)}{\left(\boldsymbol{\phi}(u)^H \mathbf{C}_{-i}^{-1} \boldsymbol{\phi}(u)\right)}$, which represents a minimum variance distortionless response (MVDR) beamformer with $\mathbf{C}_{-i}$ as the model interference plus noise signal covariance matrix [56, 122]. Thus the criterion $R(u)$ in (3.23) picks a u in the neighbourhood of $u_{i_k^{(0)}}$ that most exceeds the expected beamforming interference plus noise output power, guided by $\mathbf{C}_{-i}$. At the true location, which is likely to be in the search region, the model $\mathbf{C}_{-i}$ expects low power but hopefully the measurements indicate higher power than expected.

We solve (3.23) by implementing a fine grid of size G' around the peak and evaluating the criterion $R(u)$. Once we find the maximum point we replace the grid point $(\gamma_{i_k^{(0)}}, u_{i_k^{(0)}})$ with $(\gamma_{\text{opt}}(u^*), u^*)$. Note that by including the previous grid point in the search region we ensure that the likelihood is steadily increasing. We then repeat this procedure for the next peak, corresponding to another source, and so on. This procedure around a peak assumes that other peaks were reliably estimated, which may hold only approximately. Therefore, we iterate over the K peaks until convergence. In practice, the procedure converges quickly over $20 - 30$ iterations.

The above procedure is quite different from that in [11], in that it offers a means to improve the DoA estimate without requiring to re-run the primary procedure (here: SBL, in [11]: ℓ_1-SVD) on all grid points. As will be shown numerically in Section 3.5.6, this step alone improves the solution significantly.

3.4.2 Multi-resolution Grid Refinement

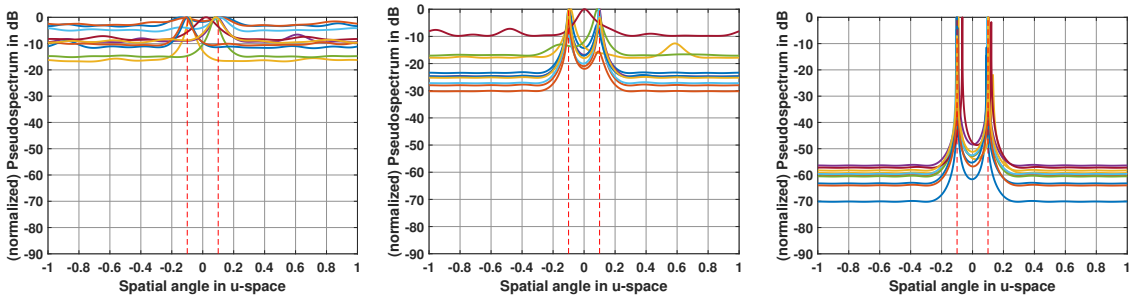
In this subsection, we take the local grid point adjustment step further by introducing a finer grid around the new peak locations and by re-running the SBL procedure. Let $r = 0$ and $j_k^{(r)}, k \in \{1, \ldots, K\}$, denote the indices for the K top peaks in the solution $\boldsymbol{\gamma}_{\text{est}}$ after grid point adjustment around the peaks, and $G^{(0)} = G$ denotes the current grid size. We run the following procedure to further refine the solution:

1. Prune the grid points i for $\{i : \boldsymbol{\gamma}_{\text{est}}(i) < \gamma_{\text{thresh}}\}$ for some γ_{thresh}. In the simulations we set $\gamma_{\text{thresh}} = 10^{-3}$.

2. Introduce new grid points in the region $[u_{j_k^{(0)}} - \frac{4}{G^{(0)}}, u_{j_k^{(0)}} + \frac{4}{G^{(0)}}]$ with finer resolution $\frac{1}{g}\frac{2}{G^{(0)}}, g > 1$. The region includes two neighbouring grid points on each side. 'g' is chosen such that the total number of grid points does not exceed $G^{(0)}$. This choice ensures that per iteration complexity is contained. In the simulations, we choose $g = 3$; in general the procedure adds $(4g+1)$ new points per peak of interest.

3. Increment $r : r+1$, and update the grid size $G^{(r)}$. Run SBL from scratch, get new set of indices for K top peaks $i_k^{(r)}, k \in \{1, \ldots, K\}$; perform grid point adjustment at these peaks to get $j_k^{(r)}, k \in \{1, \ldots, K\}$, as updated peak locations. Go to step 1).

This procedure is similar to that in [11]. We run these steps a few times and report the peak points as DoA estimates in the simulation section. A natural question that arises is: why is the local grid point adjustment not enough for improved resolution and separating two closely spaced sources? In other words, is a SBL re-run necessary?



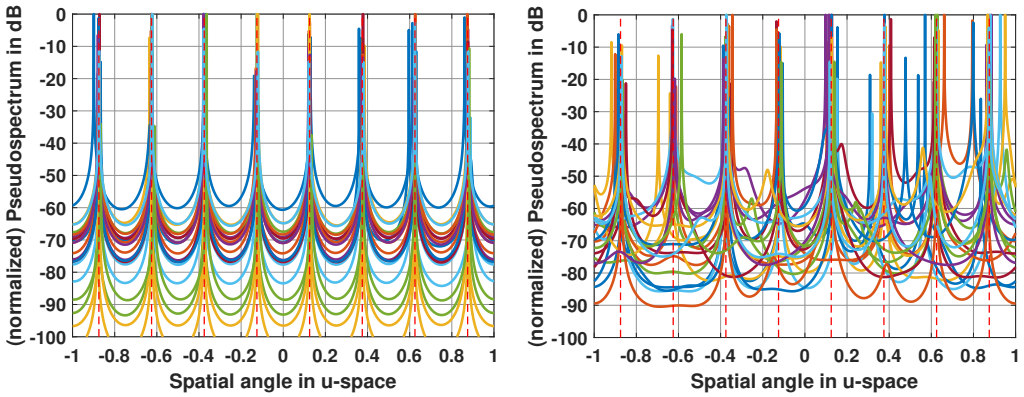
(a) SCM (b) FB averaged SCM (c) StructCovMLE (Proposed)

Figure 3.3. Single snapshot scenario

A re-run improves both, the SBL with coarse grid and the grid point adjustment around peaks. For the SBL procedure, a re-run in this manner provides a much more informed sampling of the spatial coordinates with closely-spaced grid points around locations of interest. For the grid point adjustment step where the MVDR beamformer is employed, a finer grid helps to further ensure that only a single source is present in the search region of (3.23). This is important because the beamformer $\mathbf{w} = \frac{\mathbf{C}_{-i}^{-1}\boldsymbol{\phi}(u)}{\left(\boldsymbol{\phi}(u)^H\mathbf{C}_{-i}^{-1}\boldsymbol{\phi}(u)\right)}$ engages its degrees of freedom and attempts to null interference outside of this search region, and the criterion $R(u)$ works best only if a single source is present in the search region. A block diagram summarizing the high level steps suggested in this section for the general case of sensors being placed arbitrarily is shown in Fig. 3.2.

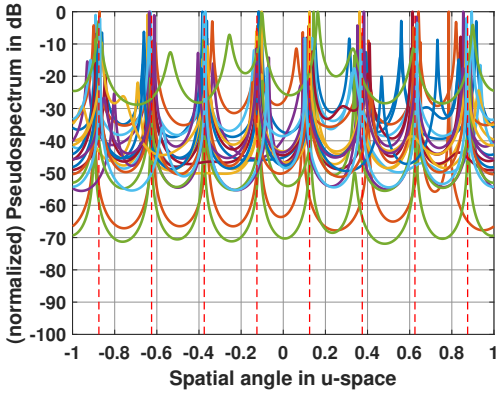
3.5 Simulation Results

We present numerical results to evaluate the performance of the proposed algorithms in Section 3.3 and 3.4 in different scenarios. We also compare the proposed 'Struct-CovMLE' algorithm with MUSIC using SCM, MUSIC using forward-backward (FB) averaged SCM [56], reweighted ANM (RAM), GridLess (GL)-SPICE, GL-SPARROW and Cramér-Rao bound (CRB) [123]. We initialize all the iterative techniques (i.e. the proposed Algorithm 6 and RAM) with the unit vector $\mathbf{v}_0 = \mathbf{e}_1$ for reasons stated in remark 16, and run 20 iterations unless otherwise specified. We provide the number of sources, K, to identify to all the algorithms. We set $\lambda = \sigma_n^2$ for the proposed algorithm. The RAM

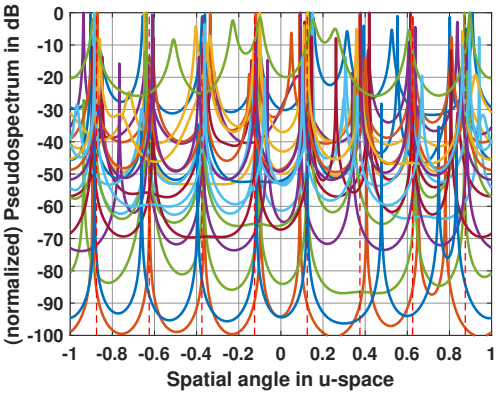


(a) StructCovMLE (RMSE = 0.005)

(b) RAM (RMSE = 0.106)



(c) GL-SPARROW (RMSE = 0.050)



(d) GL-SPICE (RMSE = 0.072)

Figure 3.4. More sources than sensors' case

implementation follows its description in [24], and we also adopt the dimension reduction mechanism suggested in the paper. We set $\eta = \sigma_n\sqrt{ML+2\sqrt{ML}}$ as suggested for DoA estimation in [24]. For GL-SPARROW, we set $\lambda = \sigma_n\sqrt{M\log M}$ as suggested in [23, 90]. GL-SPICE does not require or utilize the knowledge of the noise variance, σ_n^2. We compute root mean squared error (RMSE) in u-space ($u = \sin\theta$) as

$$\text{RMSE} = \sqrt{\frac{1}{T}\frac{1}{K}\sum_{t=1}^{T}\sum_{k=1}^{K}\left(\hat{u}_{k,t}-u_k\right)^2}, \tag{3.25}$$

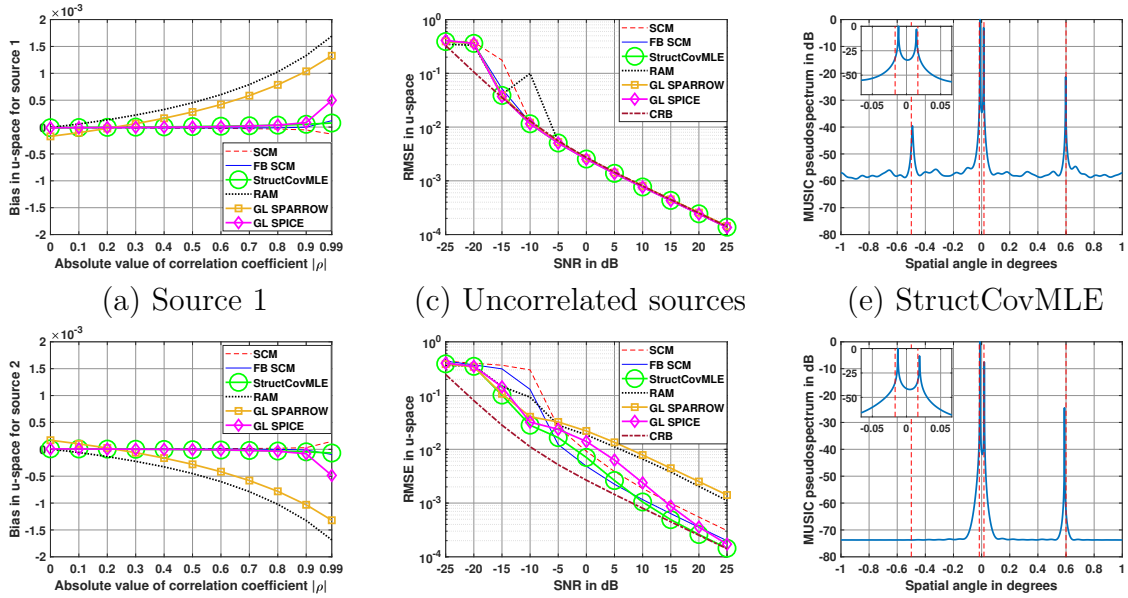
where T denotes the total number of random trials.

3.5.1 Performance in Single Snapshot Case i.e., $L = 1$

We consider a ULA with $M = 10$ sensors, and two sources at angles $\{-1/M_{\text{apt}}, 1/M_{\text{apt}}\}$,$M_{\text{apt}} = 10$, in u-space. The sources have signal-to-noise ratio (SNR) $= 20$ dB. In Fig. 3.3 we plot the normalized (with respect to maximum value) MUSIC pseudospectrum for different estimates of the structured covariance matrix, for 10 random realizations. The true DoAs are marked in vertical red dashed curve. As expected, the SCM provides the worst performance as it does not satisfy the rank requirement to identify two sources. It is observed that the forward-backward averaging helps to improve the performance by ensuring the rank requirement is satisfied. Still the overall performance is poor, as evident from the high noise floor compared to that for 'StructCovMLE' in Fig. 3.3 (c), and from its inability to resolve the two sources for some realizations. The proposed method is able to resolve both the sources. This is also true for RAM, GL-SPARROW, and GL-SPICE methods although we skip the plots in the interest of space.

3.5.2 More Sources than Sensors' Case

We consider a nested array [61] with $M = 6$ sensors at locations $\{0,1,2,3,7,11\}$, and $K = 8$ sources at angles uniformly in u-space. Their locations in MATLAB notation are $\{-1+1/K : 2/K : 1-1/K\}$. The SNR for each source is 20 dB and $L = 4$. In Fig. 3.4, we plot 20 random realizations of the normalized MUSIC pseudospectrum for the different estimates of the structured covariance matrix. As seen in Fig. 3.4 (a), the proposed algorithm is able to localize all the 8 sources, whereas the rest of the algorithms suffer from poor identifiability for some realizations. The superior performance is also evident from the lower RMSE value (in u-space) for the proposed algorithm, as compared to the other techniques.



(a) Source 1 (c) Uncorrelated sources (e) StructCovMLE

(b) Source 2 (d) Correlated sources, $|\rho| = 0.9$ (f) RAM

Figure 3.5. (a) & (b): Effect of correlation ($\rho/|\rho| = 0.5010 + j0.8654$) on empirical bias. (c) & (d): RMSE as a function of SNR. (e) & (f): Nested array with sensor locations, $\mathbb{P} = \{0,1,2,3,4,5,11,17,23,29\}$.

3.5.3 Effect of Correlation: An Empirical Bias Study

We consider a ULA with $M = 6$ sensors and two sources incoming at angles $\{-1/4, 1/4\}$. The SNR is 20 dB and $L = 500$. In Fig. 3.5 (a) and (b) we plot the empirical bias for the two sources, respectively i.e., $\frac{1}{T}\sum_{t=1}^{T}\left(\hat{u}_{k,t} - u_k^*\right), k = \{1,2\}$, as a function of the absolute value of correlation coefficient, $|\rho|$. For computing the bias, we average over $T = 50$ realizations. As observed in the plots, there is an increasing empirical bias in the angular estimates for the RAM and the GL-SPARROW techniques. This is evident as the curves drift away from the x-axis as $|\rho|$ increases. The proposed approach (shown in green curve with circular markers) has low empirical bias even when $|\rho|$ is as high as 0.99. This demonstrates the superiority of the MLE based proposed approach over the other algorithms when there may be sources that are arbitrarily correlated.

Next, we provide RMSE vs. SNR curves for uncorrelated and correlated sources' case, and compare the performance with CRB. Note that in certain scenarios the algorithms may be biased, for example in extremely low SNR regime, or as evident in Fig. 3.5 (a)

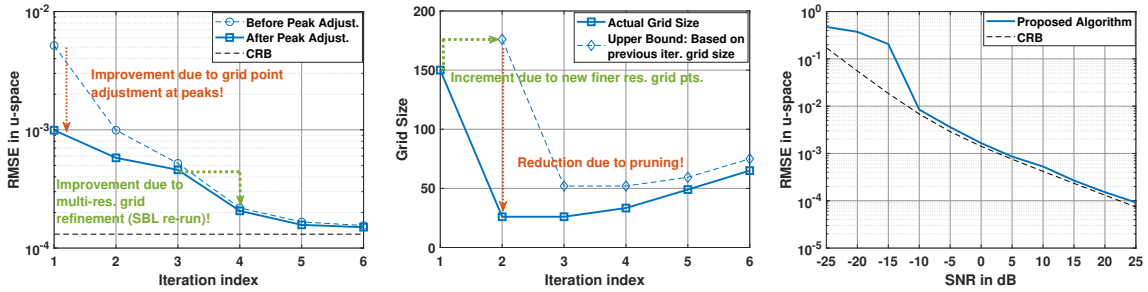
and (b) for RAM and GL SPARROW even in the high SNR scenario when the sources are correlated. We provide the curves for completion, but note that the CRB may not be a valid bound for such extreme cases.

3.5.4 Performance as a Function of SNR

We consider ULA with $M = 6$ sensors and two sources at angles $\{-1/2, 1/2\}$. $L = 500$ and we run 30 iterations for RAM and the proposed 'StructCovMLE' algorithm. In Fig. 3.5 (c) and (d) we plot the RMSE (averaged over $T = 50$ realizations) for the uncorrelated sources and correlated sources' cases, respectively, as a function of SNR. As observed in Fig. 3.5 (c), when the sources are uncorrelated, all the algorithms approach CRB as the SNR increases. When the sources are highly correlated and the SNR is high, $|\rho| = 0.9$ in Fig. 3.5 (d), we observed that the performance curve for the proposed algorithm is the closest to CRB, followed by GL-SPICE. The performance curves are worse for RAM and GL-SPARROW indicating the effect of empirical bias present in the estimates.

3.5.5 Resolution Study and Regularization-free Proposed Approach vs RAM Study

We evaluate the performance of the proposed technique for resolution and compare it with RAM. We also compare the two algorithms for the case when sources have different SNRs. We consider a nested array with $M = 10$ sensors, and allow $K = 4$ sources incoming at angles $\{-0.5, -1/2M_{\text{apt}}, 1/2M_{\text{apt}}, 0.6\}$ in u-space, where $M_{\text{apt}} = 30$. The corresponding SNR for sources is $\{5, 20, 20, 10\}$ dB and only a single snapshot ($L = 1$) is available. The two sources near broadside are $1/M_{\text{apt}}$ apart, or equivalently $0.5/M_{\text{apt}}$ apart in normalized frequencies, which is a challenging scenario. As seen in Fig. 3.5 (e) and (f) for the proposed algorithm and RAM, respectively, both are able to resolve the two sources. The proposed algorithm is able to identify all 4 sources, but RAM misses the weakest source. This behavior for RAM comes from the fact that the model is matched to an estimate of



(a) RMSE over iterations $(r = 0, 1, \ldots)$

(b) Grid size over iterations $(r = 0, 1, \ldots)$ CRB req. grid size $= \frac{1}{\text{CRB}} \approx 7632$

(c) RMSE vs. SNR

Figure 3.6. Non-uniform linear array: Performance of the proposed SBL with likelihood-based grid refinement procedure

noiseless data. In an attempt to construct such a noiseless estimate of measurement, the algorithm effectively suppressed the weakest source. It was observed that setting $\eta = 0$ helped to identify all sources for RAM here. This indicates that RAM is sensitive to setting the parameter η appropriately. Note also that the noise floor for the proposed algorithm is higher than that for RAM. This is expected because for the case of $K = 4$ sources and $M = 10$ sensors i.e., fewer sources than sensors, MUSIC is applied on the smaller $\mathbf{T}(\mathbf{v}^*) \in \mathbb{C}^{10\times 10}$ in the proposed algorithm, compared to MUSIC on $\mathbf{Toep}(\mathbf{v}^*) \in \mathbb{C}^{30\times 30}$ in RAM.

3.5.6 Performance of the Proposed SBL with Likelihood-based Grid Refinement Procedure

We consider the array geometry mentioned in Section 3.4 with $M = 6$ sensors at positions $\mathbb{P} = \{0, 1, 2.1, 3.5, 4.7, 10\}$, two sources at approximately $\{-0.5400, 0.4802\}$ in u-space, and $L = 500$. In Fig. 3.6 (a) and (b), we analyse the performance of the proposed algorithm at SNR=20 dB, while in (c) we consider a range of SNRs. We run 5000 SBL iterations each time SBL is called, and the initial grid size is $G = 150$. We plot the average results of $T = 50$ random realizations in Fig. 3.6 (a) and (b) and $T = 25$ random realizations in Fig. 3.6 (c).

In Fig. 3.6 (a), we plot the RMSE over the iterations ($r = 0, 1, \ldots$) described in Fig. 3.2 for the proposed algorithm. As observed here the RMSE decreases over iterations. Within each iteration (compare dashed vs. solid curve), it can be seen that the grid point adjustment step at peaks helps to reduce the error further, and thus establishes a simple way to further increase the likelihood. After 5 iterations, it can be seen the error is very close to the CRB. In Fig. 3.6 (b), we plot the grid size (in solid blue curve) for running SBL at every iteration of the proposed procedure. In dashed blue curve, we plot a simple upper bound on the grid size by only counting the number of new grid points added around top peaks. This helps to compute the number of points pruned at every iteration. For reaching CRB at SNR=20 dB with a single SBL run and a uniform grid, we need a grid size of $G = 7632$. In comparison, the proposed procedure only requires a maximum grid size of $G = 150$. Note that the grid resolution around the top peaks, after 5 iterations, is comparable to an initial grid size of $G \times g^(5-1) = 150 \times 3^4 = 12150$, which is more than enough to achieve CRB. In Fig. 3.6 (c), we plot the RMSE as a function of SNR. We set the maximum iterations (< 7) of the proposed procedure so as to allow for sufficiently small grid spacing at high SNR as required to reach CRB. As seen in the plot, the RMSE approaches CRB as the SNR increases. Note that for a fixed grid size i.e., for a standard SBL procedure the RMSE is expected to saturate beyond a certain SNR.

3.6 Conclusion

In this work we revisited the problem of gridless sparse signal recovery using MLE framework. We showed that SBL performs a structured covariance matrix estimation, where the structure is governed by the geometry of the measurement collection system (e.g. antenna array) and the (uncorrelated) source correlation prior. We further established that SBL is a correlation-aware technique and compared it with another class of correlation-aware techniques. Both are able to identify $O(M^2)$ sources given sparse linear array with

M sensors, like minimum-redundancy linear array [56, 60, 118] and nested array. The noteworthy aspect about SBL is the underlying objective it uses, which is MLE. In the event that some of the sources are correlated, the model misfit is characterized in terms of the KL divergence between the distribution SBL assumes and the true data distribution. We reparametrized the SBL cost function to enable gridless support recovery when the sensors are placed on uniform grid and some sensors may be switched off. We provided an iterative algorithm based on linear MM to minimize the cost function and to estimate the structured covariance matrix of measurements. The DoAs can be recovered by using any off-the-shelf root-finding technique such as root-MUSIC. In this work, we also consider geometries when the sensors may be placed off the grid, and extend the SBL procedure to include a peak adjustment and grid refinement steps. Finally we compared the proposed algorithms numerically with other state-of-the-art algorithms from the literature and demonstrated the superior performance showcased by the cost function motivated by first principles, that is maximum likelihood estimation.

Several directions are open for future work. This includes, for all sensors on grid case, developing faster methods to solve the proposed 'StructCovMLE' optimization problem. For the arbitrarily placed sensors' case, we feel the grid refinement based iterative SBL procedure is an important first step and opens up many interesting avenues of inquiry. For dictionaries parameterized by a few parameters, there is hope that discretization (grid) may not be necessary upfront except as a practical computational method as in Equation (3.23).

Chapter 3, in part, is a reprint of the material as it appears in R. R. Pote and B. D. Rao, "Maximum Likelihood-Based Gridless DoA Estimation Using Structured Covariance Matrix Recovery and SBL With Grid Refinement," in IEEE Transactions on Signal Processing, vol. 71, pp. 802-815, 2023, R. R. Pote and B. D. Rao, "Maximum Likelihood Structured Covariance Matrix Estimation and connections to SBL: A Path to Gridless DoA Estimation," 2022 56th Asilomar Conference on Signals, Systems, and Computers,

Pacific Grove, CA, USA, 2022, pp. 970-974. The dissertation author was the primary investigator and author of this paper.

3.7 Appendix

3.7.1 Proof of Proposition 2

Proof. The 'if' part can be proved simply using contradiction and follows by noting that if $(\boldsymbol{\Gamma}^*, \lambda^*)$ is not the global minimizer of (3.5), then the solution for (3.3) can be further improved. We now prove the 'only if' part. If $\varepsilon < \mathrm{tr}\left(\left(\boldsymbol{\Phi}\boldsymbol{\Gamma}^*\boldsymbol{\Phi}^H + \lambda^*\mathbf{I}\right)^{-1}\hat{\mathbf{R}}_{\mathbf{y}}\right)$, then $(\boldsymbol{\Gamma}^*, \lambda^*)$ is infeasible, and the assertion holds trivially. If $\varepsilon > \mathrm{tr}\left(\left(\boldsymbol{\Phi}\boldsymbol{\Gamma}^*\boldsymbol{\Phi}^H + \lambda^*\mathbf{I}\right)^{-1}\hat{\mathbf{R}}_{\mathbf{y}}\right)$, then $(\boldsymbol{\Gamma}^*, \lambda^*)$ lies in the feasible region. We prove that the point $(\boldsymbol{\Gamma}^*, \lambda^*)$ can be further improved. For any two matrices $\mathbf{B}, \mathbf{C} \succ \mathbf{0}$ such that $\mathbf{B} \succ \mathbf{C}$, the following holds

$$\mathrm{tr}\left((\mathbf{B}-\mathbf{C})^{-1}\hat{\mathbf{R}}_{\mathbf{y}}\right) \geq \mathrm{tr}\left(\mathbf{B}^{-1}\hat{\mathbf{R}}_{\mathbf{y}}\right) \tag{3.26}$$

$$\log\det\mathbf{B} > \log\det(\mathbf{B}-\mathbf{C}). \tag{3.27}$$

Inserting $\mathbf{B} = \boldsymbol{\Phi}\boldsymbol{\Gamma}^*\boldsymbol{\Phi}^H + \lambda^*\mathbf{I}$ and $\mathbf{C} = \alpha\mathbf{I}$, for some $\alpha \in (0, \lambda^*)$ in the above ensures that the conditions $\mathbf{B}, \mathbf{C} \succ \mathbf{0}$ such that $\mathbf{B} \succ \mathbf{C}$ are satisfied. We choose α sufficiently small to ensure that the constraint $\mathrm{tr}\left((\mathbf{B}-\mathbf{C})^{-1}\hat{\mathbf{R}}_{\mathbf{y}}\right) \leq \varepsilon$ is satisfied and consequently $(\boldsymbol{\Gamma}^*, \lambda^* + \alpha)$ is feasible. Such an α exists because $\mathrm{tr}\left((\mathbf{B}-\mathbf{C})^{-1}\hat{\mathbf{R}}_{\mathbf{y}}\right)$ is a) continuous w.r.t. α in $(0, \lambda^*)$ and b) right continuous at $\alpha = 0$ with $\mathrm{tr}\left(\mathbf{B}^{-1}\hat{\mathbf{R}}_{\mathbf{y}}\right) < \varepsilon$ as assumed. For such an α, as evident from (3.27), $(\boldsymbol{\Gamma}^*, \lambda^* + \alpha)$ further improves the solution, and thus $(\boldsymbol{\Gamma}^*, \lambda^*)$ does not globally minimize (3.5) if $\varepsilon > \mathrm{tr}\left(\left(\boldsymbol{\Phi}\boldsymbol{\Gamma}^*\boldsymbol{\Phi}^H + \lambda^*\mathbf{I}\right)^{-1}\hat{\mathbf{R}}_{\mathbf{y}}\right)$. This concludes the proof. □

3.7.2 Proof of Theorem 4

Proof. The cost functions in (3.13) and (3.17) are identical, except for the received signal covariance matrix model. The optimization variables affect their cost only through the covariance matrix. Thus, the two problems are equivalent if the effective matrix

search domains, up to an additional '$+\tilde{\lambda}\mathbf{I}$' ($\tilde{\lambda} \geq 0$) term, are same. Let $\mathscr{D}_1$ denote the matrix search region spanned by $\mathbf{T}(K, \boldsymbol{\theta}, \mathbf{P}) = \boldsymbol{\Phi}_{\boldsymbol{\theta}} \mathbf{P} \boldsymbol{\Phi}_{\boldsymbol{\theta}}^H$ in (3.17), and $\mathscr{D}_2$ for $\mathbf{T}(\mathbf{v})$ in (3.13), where the domain for the parameters are indicated in the respective problems. To prove $\mathscr{D}_1 \subseteq \mathscr{D}_2$: Let $\mathbf{T}(K', \boldsymbol{\theta}', \mathbf{P}') \in \mathscr{D}_1$ for some $(K', \boldsymbol{\theta}', \mathbf{P}')$, then the construction $\mathbf{v}' = \mathbf{T}^{-1}(\boldsymbol{\Phi}_{\boldsymbol{\theta}',\text{ULA}} \mathbf{P}' \boldsymbol{\Phi}_{\boldsymbol{\theta}',\text{ULA}}^H)$[6] ensures that $\text{Toep}(\mathbf{v}') \succeq \mathbf{0}$ and $\mathbf{T}(\mathbf{v}') = \mathbf{T}(K', \boldsymbol{\theta}', \mathbf{P}')$, i.e., $\mathbf{T}(K', \boldsymbol{\theta}', \mathbf{P}') \in \mathscr{D}_2$. This concludes $\mathscr{D}_1 \subseteq \mathscr{D}_2$. To prove $\mathscr{D}_2 \subseteq \mathscr{D}_1$: Let $\mathbf{T}(\mathbf{v}'') \in \mathscr{D}_2$ for some $\mathbf{v}''$, then we have $\text{Toep}(\mathbf{v}'') \succeq \mathbf{0}$. We skip the case when $\text{Toep}(\mathbf{v}'')$ is low rank as it follows simply from unique Vandermonde decomposition. If $\text{Toep}(\mathbf{v}'')$ is full rank, then it uniquely decomposes as $\boldsymbol{\Phi}_{\boldsymbol{\theta}'',\text{ULA}} \mathbf{P}'' \boldsymbol{\Phi}_{\boldsymbol{\theta}'',\text{ULA}}^H + \lambda'' \mathbf{I}$, for some $(\boldsymbol{\theta}'', \mathbf{P}'', \lambda'' > 0)$, where the corresponding $K'' < M_{\text{apt}}$ [73]. This ensures that $\boldsymbol{\Phi}_{\boldsymbol{\theta}''} \mathbf{P}'' \boldsymbol{\Phi}_{\boldsymbol{\theta}''}^H + \lambda'' \mathbf{I} = \mathbf{T}(\mathbf{v}'')$, which are equal up to the additional '$+\lambda'' \mathbf{I}$' term. This concludes that $\mathscr{D}_2 \subseteq \mathscr{D}_1$. □

3.7.3 Proof of Theorem 5

Proof. Similar to the proof for Theorem 4, we conclude that the two problems in (3.13) and (3.18) are equivalent if the effective matrix search domains, up to an additional '$+\tilde{\lambda}\mathbf{I}$' ($\tilde{\lambda} \geq 0$) term, are same. Let $\mathscr{D}_1$ denote the matrix search region spanned by $\mathbf{T}(\boldsymbol{\Phi}, \boldsymbol{\Gamma}) = \boldsymbol{\Phi} \boldsymbol{\Gamma} \boldsymbol{\Phi}^H$ in (3.18), and $\mathscr{D}_2$ for $\mathbf{T}(\mathbf{v})$ in (3.13), where the domain for the parameters are indicated in the respective problems. To prove $\mathscr{D}_1 \subseteq \mathscr{D}_2$: Let $\mathbf{T}(\boldsymbol{\Phi}', \boldsymbol{\Gamma}') \in \mathscr{D}_1$ for some $(\boldsymbol{\Phi}', \boldsymbol{\Gamma}')$, then the construction $\mathbf{v}' = \mathbf{T}^{-1}(\boldsymbol{\Phi}'_{\text{ULA}} \boldsymbol{\Gamma}' \boldsymbol{\Phi}'^H_{\text{ULA}})$[7] ensures that $\text{Toep}(\mathbf{v}') \succeq \mathbf{0}$ and $\mathbf{T}(\mathbf{v}') = \mathbf{T}(\boldsymbol{\Phi}', \boldsymbol{\Gamma}')$, i.e., $\mathbf{T}(\boldsymbol{\Phi}', \boldsymbol{\Gamma}') \in \mathscr{D}_2$. This concludes $\mathscr{D}_1 \subseteq \mathscr{D}_2$. To prove $\mathscr{D}_2 \subseteq \mathscr{D}_1$: Let $\mathbf{T}(\mathbf{v}'') \in \mathscr{D}_2$ for some $\mathbf{v}''$, then we have $\text{Toep}(\mathbf{v}'') \succeq \mathbf{0}$. Using Vandermonde decomposition we get $\text{Toep}(\mathbf{v}'') = \boldsymbol{\Phi}_{\boldsymbol{\theta}'',\text{ULA}} \mathbf{P}'' \boldsymbol{\Phi}_{\boldsymbol{\theta}'',\text{ULA}}^H$ for some $(\boldsymbol{\theta}'', \mathbf{P}'' \succ \mathbf{0})$ which may not be unique. This decomposition leads to a valid dictionary $\boldsymbol{\Phi}'' = \boldsymbol{\Phi}_{\boldsymbol{\theta}''}$ and diagonal source covariance matrix $\boldsymbol{\Gamma}'' = \mathbf{P}''$. This concludes that $\mathscr{D}_2 \subseteq \mathscr{D}_1$. □

[6] $\boldsymbol{\Phi}_{\boldsymbol{\theta}',\text{ULA}}$ denotes the array manifold matrix for a ULA of size M_{apt}.

[7] $\boldsymbol{\Phi}_{\text{ULA}}$ denotes the overcomplete dictionary for a ULA of size M_{apt}, evaluated at same grid points as $\boldsymbol{\Phi}$.

Chapter 4

Conclusions

In this dissertation we consider three problems of practical relevance in today's times. These problems seem different yet are inherently connected through their respective measurement models and parameters of interest.

The first problem focused on the millimeter wave initial alignment problem where high training overhead is of prime concern. Large propagation losses and specular characteristics of the channel are some of the challenges tackled in this work. A novel sensing scheme is proposed which synthesizes a virtual array manifold over time while achieving beamforming gain to overcome the propagation losses. Benefits of replacing conventional beamformers with proposed Synthesis of Virtual Array Manifold (SVAM) sensing is discussed in detail, and wide applicability of proposed sensing within existing beam alignment algorithms is elaborated. A novel inference algorithm is developed that exploits coherence within channel and outperforms benchmark algorithms. Future work includes methods to adapt the virtual array geometry and aperture size over time as more measurements are collected, dynamically selecting posterior threshold, and extensions to time-varying channels. The proposed SVAM sensing (and initial alignment procedure) is especially relevant for future technologies such as Integrated Sensing and Communication (ISAC) wherein sensing and communication services coexist to enhance overall performance. Consequently, such applications of the proposed SVAM sensing is an exciting future research

direction.

The second problem considered the Sparse Signal Recovery (SSR) problem, and a novel low complexity algorithm, named as the Light-Weight Sequential Sparse Bayesian Learning (LWS-SBL), is proposed. The proposed algorithm is compared with the widely used Orthogonal Matching Pursuit (OMP) algorithm in terms of support recovery performance and computational complexity. Both algorithms have identical asymptotic computational complexity but have different foundational principles. OMP solves a sequential least squares problem whereas LWS-SBL optimizes the stochastic maximum likelihood estimation formulation. In this regard, the latter increases (or maintains) the likelihood over iteration as more support is added to the model. A fundamental framework is proposed to compare OMP and LWS-SBL (which can be applied more generally to other SSR algorithms) in terms of the underlying beamformer and metric employed to select columns from the measurement matrix. The proposed framework helps to understand the superior performance of LWS-SBL compared to OMP when sources are closely located as evident from the empirical results. Extensions of the proposed algorithm to single and two-dimensional gridless parameter estimation problem is discussed and empirically studied. This contribution transforms a Bayesian algorithm for SSR into an efficient greedy approach and provides associated and additional insights. An interesting future research direction is to extend the framework proposed in this work, to compare OMP and LWS-SBL i.e., the underlying beamformer and metric to select columns, to classify other SSR algorithms. Another research direction is the inherent connection between LWS-SBL and Order Recursive MP, wherein the former derives from a stochastic formulation and the latter is derived from a deterministic one. However, the two seem more related and exploring this connection is an interesting research problem.

Finally the third problem considers the maximum likelihood estimation framework for the Direction of Arrival (DoA) estimation problem. The proposed approach helps to overcome the model complexities associated with the traditional MLE formulation by

reformulating the SBL formulation as a novel structured matrix recovery problem. It recovers a structured covariance matrix estimate using the maximum likelihood estimation framework, and is named as StructCovMLE. The resulting approach is correlation-aware and thus enjoys the rich properties of such procedures, for instance, the proposed algorithm can identify more sources than sensors in suitable scenarios. Moreover by virtue of the MLE formulation, the proposed algorithm is demonstrated to outperform existing modern and traditional DoA estimation algorithms in low number of snapshots, correlated sources, and low SNR scenarios. The proposed algorithm requires to solve a semi-definite program at every iteration. Developing first-order methods and application on field experiments where practical requirements render classical algorithms less useful such as low snapshots or SNR are interesting directions. For the arbitrarily placed sensors' case, the grid refinement based iterative SBL procedure is an important first step and opens up many interesting avenues of inquiry. For dictionaries parameterized by a few parameters, a discretization (grid) may not be necessary upfront except as a practical computational method as in Equation (3.23).

In this dissertation, accurate modeling and exploitation of array geometry and correlation prior information is emphasized. In practice, such accurate prior knowledge or high-quality measurements may not always be available. For instance, there may be perturbation in the sensor locations, the sources may be arbitrarily correlated or only highly quantized measurements such as few bits may be available. Even when the available information may be reliable, the complexity of the proposed algorithms may be an impediment for certain real-time applications. In such scenarios data-driven techniques may be suited as refinement to the solutions from the model-based algorithms. Such research enquiries form an exciting research direction, as well.

Bibliography

[1] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," IEEE Communications Magazine, vol. 49, no. 6, pp. 101–107, June 2011.

[2] 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.

[3] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," IEEE Journal on Selected Areas in Communications, vol. 32, no. 6, pp. 1164–1179, June 2014.

[4] I.F. Gorodnitsky, J.S. George, and B.D. Rao, "Neuromagnetic source imaging with FOCUSS: a recursive weighted minimum norm algorithm.," J. Electroencephalog. Clinical Neurophysiol., vol. 95, no.4, pp. 231–251, 1995.

[5] B. K. Natarajan, "Sparse Approximate Solutions to Linear Systems," SIAM J. Comput., vol. 24, no. 2, pp. 227–234, 1995.

[6] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," IEEE Transactions on Speech and Audio Processing, vol. 8, no. 5, pp. 508–518, 2000.

[7] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," IEEE Transactions on Signal Processing, vol. 41, no. 12, pp. 3397–3415, 1993.

[8] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," SIAM Journal on Scientific Computing, vol. 20, no. 1, pp. 33–61, 1998.

[9] B. D. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," IEEE Transactions on Signal Processing, vol. 47, no. 1, pp. 187–200, 1999.

[10] J.A. Tropp, "Just relax: convex programming methods for identifying sparse signals in noise," IEEE Transactions on Information Theory, vol. 52, no. 3, pp. 1030–1051, 2006.

[11] D. Malioutov, M. Cetin, and A.S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," IEEE Transactions on Signal Processing, vol. 53, no. 8, pp. 3010–3022, 2005.

[12] D. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," IEEE Transactions on Signal Processing, vol. 52, no. 8, pp. 2153–2164, Aug 2004.

[13] M. E. Tipping and A. C. Faul, "Fast marginal likelihood maximisation for sparse Bayesian models," in Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics. 03–06 Jan 2003, vol. R4, pp. 276–283, PMLR.

[14] Y. C. Eldar and G. Kutyniok, Compressed Sensing: Theory and Applications, Cambridge University Press, Cambridge ; New York :, 1 edition, 2012.

[15] Y.C. Pati, R. Rezaiifar, and P.S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in Proceedings of 27th Asilomar Conference on Signals, Systems and Computers, 1993, pp. 40–44 vol.1.

[16] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," Constructive Approximation, vol. 13, pp. 57–98, 1997.

[17] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," IEEE Transactions on Information Theory, vol. 53, no. 12, pp. 4655–4666, 2007.

[18] B. L. Sturm and M. G. Christensen, "Comparison of orthogonal matching pursuit implementations," in 2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO), 2012, pp. 220–224.

[19] D. Needell and J.A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," Applied and Computational Harmonic Analysis, vol. 26, no. 3, pp. 301–321, 2009.

[20] G.Z. Karabulut and A. Yongacoglu, "Sparse channel estimation using orthogonal matching pursuit algorithm," in IEEE 60th Vehicular Technology Conference, 2004. VTC2004-Fall. 2004, 2004, vol. 6, pp. 3880–3884 Vol. 6.

[21] J. A. Becerra, M. J. Madero-Ayora, J. Reina-Tosina, C. Crespo-Cadenas, J. García-Frías, and G. Arce, "A doubly orthogonal matching pursuit algorithm for sparse predistortion of power amplifiers," IEEE Microwave and Wireless Components Letters, vol. 28, no. 8, pp. 726–728, 2018.

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

[23] C. Steffens, M. Pesavento, and M. E. Pfetsch, "A compact formulation for the $\ell_{2,1}$ mixed-norm minimization problem," IEEE Transactions on Signal Processing, vol. 66, no. 6, pp. 1483–1497, 2018.

[24] Z. Yang and L. Xie, "Enhancing sparsity and resolution via reweighted atomic norm minimization," IEEE Transactions on Signal Processing, vol. 64, no. 4, pp. 995–1006, 2016.

[25] Z. Yang and L. Xie, "On gridless sparse methods for line spectral estimation from complete and incomplete data," IEEE Transactions on Signal Processing, vol. 63, no. 12, pp. 3139–3153, 2015.

[26] M. Tipping, "Sparse Bayesian learning and the relevance vector machine," Machine Learning Research, vol. 1, pp. 211–244, 2001.

[27] D. P. Wipf and B. D. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," IEEE Transactions on Signal Processing, vol. 55, no. 7, pp. 3704–3716, 2007.

[28] D. Wipf and S. Nagarajan, "Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions," IEEE Journal of Selected Topics in Signal Processing, vol. 4, no. 2, pp. 317–329, 2010.

[29] S.F. Cotter, B.D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," IEEE Trans. on Signal Processing, vol. 53, no. 7, pp. 2477–2488, 2005.

[30] R. Giri and B. D. Rao, "Type I and Type II bayesian methods for sparse signal recovery using scale mixtures," IEEE Transactions on Signal Processing, vol. 64, no. 13, pp. 3418–3428, 2016.

[31] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," Journal of the Royal Statistical Society. Series B (Methodological), vol. 58, no. 1, pp. 267–288, 1996.

[32] E. Bastug, M. Bennis, M. Medard, and M. Debbah, "Toward interconnected virtual reality: Opportunities, challenges, and enablers," IEEE Communications Magazine, vol. 55, no. 6, pp. 110–117, 2017.

[33] A. Ghosh, A. Maeder, M. Baker, and D. Chandramouli, "5g evolution: A view on 5g cellular technology beyond 3gpp release 15," IEEE Access, vol. 7, pp. 127639–127651, 2019.

[34] C.H. Doan, S. Emami, D.A. Sobel, A.M. Niknejad, and R.W. Brodersen, "Design considerations for 60 ghz cmos radios," IEEE Communications Magazine, vol. 42, no. 12, pp. 132–140, 2004.

[35] Sung-En Chiu, Nancy Ronquillo, and Tara Javidi, "Active learning and csi acquisition for mmwave initial alignment," IEEE Journal on Selected Areas in Communications, vol. 37, no. 11, pp. 2474–2489, 2019.

[36] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," IEEE Journal of Selected Topics in Signal Processing, vol. 8, no. 5, pp. 831–846, Oct 2014.

[37] Nancy Ronquillo, Sung-En Chiu, and Tara Javidi, "Sequential learning of csi for mmwave initial alignment," in 2019 53rd Asilomar Conference on Signals, Systems, and Computers, 2019, pp. 1278–1283.

[38] N. Akdim, C. N. Manchón, M. Benjillali, and P. Duhamel, "Variational hierarchical posterior matching for mmwave wireless channels online learning," in 2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), 2020, pp. 1–5.

[39] Rohan R. Pote and Bhaskar D. Rao, "Reduced dimension beamspace design incorporating nested array for mmwave channel estimation," in 2019 53rd Asilomar Conference on Signals, Systems, and Computers, 2019, pp. 1212–1216.

[40] Y. Lin and T. Yang, "Random sbt precoding for angle estimation of mmwave massive mimo systems using sparse arrays spacing," IEEE Access, vol. 8, pp. 163380–163393, 2020.

[41] Zhilin Chen, Foad Sohrabi, Ya-Feng Liu, and Wei Yu, "Phase transition analysis for covariance based massive random access with massive mimo," 2020.

[42] S.D. Silverstein, W.E. Engeler, and J.A. Tardif, "Parallel architectures for multirate superresolution spectrum analyzers," IEEE Transactions on Circuits and Systems, vol. 38, no. 4, pp. 449–453, 1991.

[43] A. Tkacenko and P.P. Vaidyanathan, "The role of filter banks in sinusoidal frequency estimation," Journal of the Franklin Institute, vol. 338, no. 5, pp. 517–547, 2001.

[44] D. Ramasamy, S. Venkateswaran, and U. Madhow, "Compressive adaptation of large steerable arrays," in 2012 Information Theory and Applications Workshop, 2012, pp. 234–239.

[45] D. E. Berraki, S. M. D. Armour, and A. R. Nix, "Application of compressive sensing in sparse spatial channel recovery for beamforming in mmwave outdoor systems," in 2014 IEEE Wireless Communications and Networking Conference (WCNC), 2014, pp. 887–892.

[46] A. Alkhateeb, G. Leus, and R. W. Heath, "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?," in 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015, pp. 2909–2913.

[47] 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 Transactions on Communications, vol. 61, no. 10, pp. 4391–4403, 2013.

[48] Z. Xiao, T. He, P. Xia, and X. Xia, "Hierarchical codebook design for beamforming training in millimeter-wave communication," IEEE Trans. on Wireless Communications, vol. 15, no. 5, pp. 3380–3392, 2016.

[49] J. Zhang, Y. Huang, Q. Shi, J. Wang, and L. Yang, "Codebook design for beam alignment in millimeter wave communication systems," IEEE Transactions on Communications, vol. 65, no. 11, pp. 4980–4995, 2017.

[50] C. Liu, M. Li, S. V. Hanly, I. B. Collings, and P. Whiting, "Millimeter wave beam alignment: Large deviations analysis and design insights," IEEE Journal on Selected Areas in Communications, vol. 35, no. 7, pp. 1619–1631, 2017.

[51] C. Liu, L. Zhao, M. Li, and L. Yang, "Adaptive beam search for initial beam alignment in millimetre-wave communications," IEEE Trans. on Vehicular Technology, vol. 71, no. 6, pp. 6801–6806, 2022.

[52] F. Sohrabi, T. Jiang, W. Cui, and W. Yu, "Active sensing for communications by learning," IEEE Journal on Selected Areas in Communications, vol. 40, no. 6, pp. 1780–1794, 2022.

[53] Y. Wei, Z. Zhong, and V. Y. F. Tan, "Fast beam alignment via pure exploration in multi-armed bandits," IEEE Transactions on Wireless Communications, vol. 22, no. 5, pp. 3264–3279, 2023.

[54] M. Hussain and N. Michelusi, "Second-best beam-alignment via bayesian multi-armed bandits," in 2019 IEEE Global Communications Conference (GLOBECOM), 2019, pp. 1–6.

[55] R. R. Pote and B. D. Rao, "Novel sensing methodology for initial alignment using mmwave phased arrays," in Asilomar Conference on Signals, Systems, and Computers, 2023.

[56] H. L. Van Trees, "Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory," John Wiley & Sons, Ltd, p. 178, 2002.

[57] Christian Waldschmidt, Juergen Hasch, and Wolfgang Menzel, "Automotive radar — from first efforts to future systems," IEEE Journal of Microwaves, vol. 1, no. 1, pp. 135–148, 2021.

[58] A. V. Oppenheim and R. W. Schafer, Discrete-Time Signal Processing, Prentice Hall Press, USA, 3rd edition, 2009.

[59] R. R. Pote and B. D. Rao, "Maximum likelihood-based gridless doa estimation using structured covariance matrix recovery and sbl with grid refinement," IEEE Trans. on Signal Proc., vol. 71, pp. 802–815, 2023.

[60] A. Moffet, "Minimum-redundancy linear arrays," IEEE Transactions on Antennas and Propagation, vol. 16, no. 2, pp. 172–175, 1968.

[61] P. Pal and P. P. Vaidyanathan, "Nested arrays: A novel approach to array processing with enhanced degrees of freedom," IEEE Transactions on Signal Processing, vol. 58, no. 8, pp. 4167–4181, 2010.

[62] Palghat P. Vaidyanathan and Piya Pal, "Sparse sensing with co-prime samplers and arrays," IEEE Transactions on Signal Processing, vol. 59, no. 2, pp. 573–586, 2011.

[63] P. Stoica and A. Nehorai, "Music, maximum likelihood, and cramer-rao bound," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 37, no. 5, pp. 720–741, 1989.

[64] Rohan R. Pote and Bhaskar D. Rao, "Light-weight sequential sbl algorithm: An alternative to omp," in ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023, pp. 1–5.

[65] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," Magnetic Resonance in Medicine, vol. 58, no. 6, pp. 1182–1195, 2007.

[66] J. Trzasko and A. Manduca, "Highly undersampled magnetic resonance image reconstruction via homotopic ℓ_0 -minimization," IEEE Transactions on Medical Imaging, vol. 28, no. 1, pp. 106–121, 2009.

[67] Q. Ling and Z. Tian, "Decentralized sparse signal recovery for compressive sleeping wireless sensor networks," IEEE Transactions on Signal Processing, vol. 58, no. 7, pp. 3816–3827, 2010.

[68] Z. Qin, J. Fan, Y. Liu, Y. Gao, and G. Y. Li, "Sparse representation for wireless communications: A compressive sensing approach," IEEE Signal Processing Magazine, vol. 35, no. 3, pp. 40–58, 2018.

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

[70] S. Foucart and H. Rauhut, A Mathematical Introduction to Compressive Sensing, Birkhäuser New York, NY, 2013.

[71] S.F. Cotter, J. Adler, B.D. Rao, and K. Kreutz-Delgado, "Forward sequential algorithm for best basis selection," IEE Proc Vision, Image Signal Proc, pp. 235–244, 1999.

[72] J. A. Högbom and Tim J. Cornwell, "Aperture synthesis with a non-regular distribution of interferometer baselines. commentary," Astronomy and Astrophysics, vol. 500, pp. 55–66, 1974.

[73] P. Stoica and R. Moses, Spectral Analysis of Signals, Prentice-Hall, Upper Saddle River, NJ, USA, 2005.

[74] M. Trinh-Hoang, W. K. Ma, and M. Pesavento, "A partial relaxation doa estimator based on orthogonal matching pursuit," in ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020, pp. 4806–4810.

[75] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," IEEE Signal Processing Magazine, vol. 13, no. 4, pp. 67–94, July 1996.

[76] R. R. Pote and B. D. Rao, "Maximum likelihood-based gridless doa estimation using structured covariance matrix recovery and sbl with grid refinement," arXiv, 2022.

[77] D. Nion and N. D. Sidiropoulos, "Tensor Algebra and Multidimensional Harmonic Retrieval in Signal processing for MIMO Radar," IEEE Transactions on Signal Processing, vol. 58, no. 11, pp. 5693–5705, 2010.

[78] Xiangqian Liu, Nikos D. Sidiropoulos, and Tao Jiang, Multidimensional Harmonic Retrieval with Applications in MIMO Wireless Channel Sounding, chapter 2, pp. 41–75, John Wiley & Sons, Ltd, 2005.

[79] Y. Li, J. Razavilar, and K. J. R. Liu, "A high-resolution technique for multidimensional NMR spectroscopy," IEEE Transactions on Biomedical Engineering, vol. 45, no. 1, pp. 78–86, 1998.

[80] B. Mamandipoor, D. Ramasamy, and U. Madhow, "Newtonized orthogonal matching pursuit: Frequency estimation over the continuum," IEEE Transactions on Signal Processing, vol. 64, no. 19, pp. 5066–5081, 2016.

[81] Lin Han, Xingchuan Liu, Ning Zhang, Sheng Wu, Jiang Zhu, and Zhiwei Xu, "Two-dimensional multi-snapshot newtonized orthogonal matching pursuit for doa estimation," Digital Signal Processing, vol. 121, pp. 103313, 2022.

[82] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," Proceedings of the IEEE, vol. 57, no. 8, pp. 1408–1418, 1969.

[83] R. Schmidt, "Multiple emitter location and signal parameter estimation," IEEE Transactions on Antennas and Propagation, vol. 34, no. 3, pp. 276–280, 1986.

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

[85] P. Pal and P. P. Vaidyanathan, "Pushing the limits of sparse support recovery using correlation information," IEEE Transactions on Signal Processing, vol. 63, no. 3, pp. 711–726, 2015.

[86] O. Balkan, K. Kreutz-Delgado, and S. Makeig, "Localization of more sources than sensors via jointly-sparse Bayesian learning," IEEE Signal Processing Letters, vol. 21, no. 2, pp. 131–134, 2014.

[87] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," IEEE Trans. on Signal Processing, vol. 65, no. 3, pp. 794–816, 2017.

[88] T. Lipp and S. Boyd, "Variations and extension of the convex–concave procedure," Optimization and Engineering, vol. 17, no. 9, pp. 1573–2924, 2016.

[89] R. R. Pote and B. D. Rao, "Maximum likelihood structured covariance matrix estimation and connections to SBL: A path to gridless DoA estimation," in 2022 56th Asilomar Conference on Signals, Systems, and Computers, 2022.

[90] W. Suleiman, C. Steffens, A. Sorg, and M. Pesavento, "Gridless compressed sensing for fully augmentable arrays," in 2017 25th European Signal Processing Conference (EUSIPCO), 2017, pp. 1986–1990.

[91] J. P. Burg, D. G. Luenberger, and D. L. Wenger, "Estimation of structured covariance matrices," Proceedings of the IEEE, vol. 70, no. 9, pp. 963–974, 1982.

[92] J. Bohme, "Separated estimation of wave parameters and spectral parameters by maximum likelihood," in ICASSP '86. IEEE International Conference on Acoustics, Speech, and Signal Processing, 1986, vol. 11, pp. 2819–2822.

[93] M. I. Miller and D. L. Snyder, "The role of likelihood and entropy in incomplete-data problems: Applications to estimating point-process intensities and toeplitz constrained covariances," Proceedings of the IEEE, vol. 75, no. 7, pp. 892–907, 1987.

[94] A. G. Jaffer, "Maximum likelihood direction finding of stochastic sources: a separable solution," in ICASSP-88., International Conference on Acoustics, Speech, and Signal Processing, 1988, pp. 2893–2896 vol.5.

[95] D. R. Fuhrmann, "Progress in structured covariance estimation," in Fourth Annual ASSP Workshop on Spectrum Estimation and Modeling, 1988, pp. 158–161.

[96] H. Li, P. Stoica, and J. Li, "Computationally efficient maximum likelihood estimation of structured covariance matrices," IEEE Transactions on Signal Processing, vol. 47, no. 5, pp. 1314–1323, 1999.

[97] P. Stoica and A. Nehorai, "On the concentrated stochastic likelihood function in array signal processing," Circuits, Systems and Signal Process., vol. 14, pp. 669–674, 1995.

[98] M. Viberg, P. Stoica, and B. Ottersten, "Maximum likelihood array processing in spatially correlated noise fields using parameterized signals," IEEE Trans. on Signal Processing, vol. 45, no. 4, pp. 996–1004, 1997.

[99] B Ottersten, P Stoica, and R Roy, "Covariance matching estimation techniques for array signal processing applications," Digital Signal Processing, vol. 8, no. 3, pp. 185 – 210, 1998.

[100] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," Foundations of Computational Mathematics, vol. 12, pp. 805–849, 2012.

[101] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," Communications on Pure and Applied Mathematics, vol. 67, no. 6, pp. 906–956, 2014.

[102] P. Stoica, P. Babu, and J. Li, "SPICE: A sparse covariance-based estimation method for array processing," IEEE Transactions on Signal Processing, vol. 59, no. 2, pp. 629–638, 2011.

[103] P. Stoica and P. Babu, "SPICE and LIKES: Two hyperparameter-free methods for sparse-parameter estimation," Signal Processing, vol. 92, no. 7, pp. 1580 – 1590, 2012.

[104] P. Pal and P. P. Vaidyanathan, "Correlation-aware techniques for sparse support recovery," in 2012 IEEE Statistical Signal Processing Workshop (SSP), 2012, pp. 53–56.

[105] M. Al-Shoukairi, P. Schniter, and B. D. Rao, "A GAMP-based low complexity sparse Bayesian learning algorithm," IEEE Transactions on Signal Processing, vol. 66, no. 2, pp. 294–308, 2018.

[106] S. Liu, H. Wu, Y. Huang, Y. Yang, and J. Jia, "Accelerated structure-aware sparse Bayesian learning for three-dimensional electrical impedance tomography," IEEE Transactions on Industrial Informatics, vol. 15, no. 9, pp. 5033–5041, 2019.

[107] D. Wipf and S. Nagarajan, "A new view of automatic relevance determination," in Advances in Neural Information Processing Systems. 2008, vol. 20, Curran Associates, Inc.

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

[109] D. P. Wipf, B. D. Rao, and S. Nagarajan, "Latent variable Bayesian models for promoting sparsity," IEEE Transactions on Information Theory, vol. 57, no. 9, pp. 6236–6255, 2011.

[110] A. Barabell, "Improving the resolution performance of eigenstructure-based direction-finding algorithms," in IEEE International Conference on Acoustics, Speech, and Signal Processing, 1983, vol. 8, pp. 336–339.

[111] C. Carathéodory and L. Fejér, "Uber den Zusammenhang der Extremen von harmonischen Funktionen mit ihren Koeffizienten und uber den Picard-Landau'schen Satz," Rendiconti del Circolo Matematico di Palermo (1884-1940), vol. 32, no. 1, pp. 218–239, 1911.

[112] M. Fazel, H. Hindi, and S. P. Boyd, "Log-det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices," in Proceedings of the 2003 American Control Conference, 2003., 2003, vol. 3, pp. 2156–2162 vol.3.

[113] R. R. Pote and B. D. Rao, "Robustness of sparse Bayesian learning in correlated environments," in IEEE International Conference on Acoustics, Speech and Signal Processing, 2020, pp. 9100–9104.

[114] D. Wipf and S. Nagarajan, "Beamforming using the relevance vector machine," in Proceedings of the 24th International Conference on Machine Learning, New York, NY, USA, 2007, ICML '07, p. 1023–1030, Association for Computing Machinery.

[115] Stephen Boyd and Lieven Vandenberghe, Convex Optimization, Cambridge University Press, 2004.

[116] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[117] M. Fazel, H. Hindi, and S. P. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in Proceedings of the 2001 American Control Conference. (Cat. No.01CH37148), 2001, vol. 6, pp. 4734–4739 vol.6.

[118] S. U. Pillai, Y. Bar-Ness, and F. Haber, "A new approach to array geometry for improved spatial spectrum estimation," Proceedings of the IEEE, vol. 73, no. 10, pp. 1522–1524, 1985.

[119] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation," IEEE Trans. on Acoustics, Speech, and Signal Processing, vol. 38, no. 10, pp. 1783–1795, 1990.

[120] H. Qiao and P. Pal, "On maximum-likelihood methods for localizing more sources than sensors," IEEE Signal Processing Letters, vol. 24, no. 5, pp. 703–706, 2017.

[121] Z. Yang, X. Chen, and X. Wu, "A robust and statistically efficient maximum-likelihood method for doa estimation using sparse linear arrays," arXiv, 2022.

[122] M. AlShoukairi, "Message passing algorithms and extensions of sparse Bayesian learning," UC San Diego. ProQuest ID: AlShoukairi_ucsd_0033D_20301. Merritt ID: ark:/13030/m52g3pxb. Retrieved from https://escholarship.org/uc/item/88812150, 2021.

[123] P. Stoica, E.G. Larsson, and A.B. Gershman, "The stochastic CRB for array processing: a textbook derivation," IEEE Signal Processing Letters, vol. 8, no. 5, pp. 148–150, 2001.