



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

CAS-MS-2015-08

M.Sc. Thesis

Sparse Arrays: Vector Sensors and Design Algorithms

Shilpa Rao

Abstract

Direction-of-arrival (DOA) estimation of acoustic sources is of great interest in a number of applications. Acoustic vector sensors (AVSs) provide an edge over traditional scalar sensors since they measure the acoustic velocity field in addition to the acoustic pressure. It is known that a uniform linear array (ULA) of M conventional scalar sensors can identify up to $M - 1$ DOAs. However, using second-order statistics, the class of sparse scalar sensor arrays have been shown to identify more source DOAs than the number of sensors. In this thesis, we extend these results using sparse AVS arrays. We first assume that the sources are quasi-stationary and use the Khatri-Rao subspace approach to estimate the source DOAs. In addition, a spatial-velocity smoothing technique is proposed to estimate the DOAs of stationary sources. For both scenarios, we show that the number of source DOAs that can be identified is significantly greater than the number of physical vector sensors.

The second problem considered in this thesis is sensor selection for non-linear models. It is often necessary to guarantee a certain estimation accuracy by choosing the best subset of the available set of sensors. A non-linear measurement model in additive Gaussian noise is considered. To solve the sensor selection problem, which is inherently combinatorial, a greedy algorithm based on submodular cost functions is developed. The proposed low-complexity greedy algorithm is computationally attractive as compared to existing sensor selection solvers for non-linear models. The submodular cost ensures optimality of the greedy algorithm.



Sparse Arrays: Vector Sensors and Design Algorithms

THESIS

submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

in

ELECTRICAL ENGINEERING

by

Shilpa Rao
born in Hyderabad, India

This work was performed in:

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



Delft University of Technology

Copyright © 2015 Circuits and Systems Group
All rights reserved.

DELFT UNIVERSITY OF TECHNOLOGY
DEPARTMENT OF
MICROELECTRONICS & COMPUTER ENGINEERING

The undersigned hereby certify that they have read and recommend to the Faculty of Electrical Engineering, Mathematics and Computer Science for acceptance a thesis entitled “**Sparse Arrays: Vector Sensors and Design Algorithms**” by **Shilpa Rao** in partial fulfillment of the requirements for the degree of **Master of Science**.

Dated: 16 July 2015

Chairman:

prof.dr.ir. Geert Leus

Advisor:

ir. Sundeep Prabhakar Chepuri

Committee Members:

dr.ir. Hans Driessen

dr.Emiel Tijds

Abstract

Direction-of-arrival (DOA) estimation of acoustic sources is of great interest in a number of applications. Acoustic vector sensors (AVSs) provide an edge over traditional scalar sensors since they measure the acoustic velocity field in addition to the acoustic pressure. It is known that a uniform linear array (ULA) of M conventional scalar sensors can identify up to $M - 1$ DOAs. However, using second-order statistics, the class of sparse scalar sensor arrays have been shown to identify more source DOAs than the number of sensors. In this thesis, we extend these results using sparse AVS arrays. We first assume that the sources are quasi-stationary and use the Khatri-Rao subspace approach to estimate the source DOAs. In addition, a spatial-velocity smoothing technique is proposed to estimate the DOAs of stationary sources. For both scenarios, we show that the number of source DOAs that can be identified is significantly greater than the number of physical vector sensors.

The second problem considered in this thesis is sensor selection for non-linear models. It is often necessary to guarantee a certain estimation accuracy by choosing the best subset of the available set of sensors. A non-linear measurement model in additive Gaussian noise is considered. To solve the sensor selection problem, which is inherently combinatorial, a greedy algorithm based on submodular cost functions is developed. The proposed low-complexity greedy algorithm is computationally attractive as compared to existing sensor selection solvers for non-linear models. The submodular cost ensures optimality of the greedy algorithm.

Acknowledgments

I would like to thank all the people who have helped me and contributed to this thesis indirectly.

First of all, I wish to express my sincere gratitude to Sundeeep for his daily supervision and guidance. From him, I have learnt to be patient and to be critical - two qualities that every researcher must possess. His invaluable comments and suggestions have helped me improve my work greatly. His pleasant attitude has made it enjoyable to collaborate with him.

I am also immensely grateful to my professor, Dr. Geert Leus, for giving me the opportunity to work with him. I am thankful for his timely advice, his guidance and constructive feedback that have gone a long way in making this work more worthwhile. Thank you!

I would also like to thank the members of the Circuits and Systems group for being very helpful, and for inviting me to the group activities which I greatly enjoyed. I am also especially thankful to Dony for his valuable suggestions and for his interest in this work.

Finally, I want to thank my parents, my brother, and my friends for their kind words of encouragement and support.

Shilpa Rao
Delft, The Netherlands
16 July 2015

Contents

| | |
|---|-------------|
| Abstract | v |
| Acknowledgments | vii |
| Notations | xiii |
| 1 Introduction | 1 |
| 1.1 Applications of Vector Sensors | 3 |
| 1.2 Motivation | 3 |
| 1.3 Problem Statement | 4 |
| 1.4 Thesis Outline | 5 |
| 1.5 Publications | 6 |
| I DOA Estimation using Sparse AVS Arrays | 7 |
| 2 Data Model - AVS Setup | 9 |
| 2.1 Data Model | 9 |
| 2.2 Correlation Domain Data Model | 10 |
| 3 Quasi-Stationary Sources | 12 |
| 3.1 Khatri-Rao Subspace Approach | 12 |
| 3.2 Full Rank Condition of KR Product | 14 |
| 3.3 MUSIC Identifiability | 16 |
| 3.4 Simulation Results - Quasi-stationary sources | 16 |
| 3.5 Conclusions | 23 |
| Appendices | 25 |
| 3.A Appendix A: Proof of Lemma 1 | 25 |
| 3.B Appendix B: Proof of Theorem 1 | 25 |

| | | |
|-----------|---|-----------|
| 4 | Stationary Sources | 27 |
| 4.1 | Extraction of Rows | 27 |
| 4.2 | Formation of Subarrays: Example | 28 |
| 4.3 | Formation of Subarrays: General Procedure | 29 |
| 4.4 | Effective Model | 30 |
| 4.5 | Simulation Results - Stationary Sources | 32 |
| 4.6 | Conclusions | 33 |
| | | |
| II | Sensor Selection for Non-linear Measurement Models | 38 |
| | | |
| 5 | Greedy Sensor Selection for Non-Linear Models | 40 |
| 5.1 | Problem Statement | 41 |
| 5.2 | Performance Measures | 41 |
| 5.2.1 | Taylor Approximation | 41 |
| 5.2.2 | Cost Function | 42 |
| 5.3 | Proposed Algorithm | 43 |
| 5.4 | Greedy vs. Convex Optimization | 44 |
| 5.5 | Numerical Example | 45 |
| 5.6 | Conclusions | 47 |
| | | |
| | Appendices | 49 |
| 5.A | Submodularity of the cost function - weighted FP | 49 |
| 5.B | Submodularity of the cost function - weighted log-det | 50 |
| | | |
| 6 | Conclusions and Future Work | 52 |
| 6.1 | Conclusion | 52 |
| 6.2 | Suggestions for Future Research | 53 |
| | | |
| A | Preliminaries of Covariance Sensing | 55 |
| A.1 | Sparse Arrays | 55 |
| A.2 | Spatial Smoothing | 57 |
| A.3 | Related Work | 58 |

| | |
|--|-----------|
| B Preliminaries of Sensor Selection | 59 |
| B.1 Existing Art | 60 |
| B.2 Performance Measures and Convex Optimization | 61 |
| B.3 Submodular Cost Functions | 63 |

List of Figures

| | | |
|-----|--|----|
| 1.1 | Acoustic Vector Sensor | 2 |
| 1.2 | Computational complexity and performance for non-linear data models | 5 |
| 3.1 | A speech signal | 12 |
| 3.2 | KR-MUSIC: MUSIC Spectrum and RMSE - Two-level AVS nested array. | 19 |
| 3.3 | KR-MUSIC: MUSIC Spectrum and RMSE - AVS-MRA. | 20 |
| 3.4 | KR-Subspace MUSIC: Probability of resolution- two level nested array and MRA. | 21 |
| 3.5 | Probability of false alarm: Two level nested array $M = 6$ | 23 |
| 3.6 | KR-Subspace MUSIC: Probability of detection - Two level nested array and MRA. | 24 |
| 4.1 | SV-Smoothing: MUSIC Spectrum and RMSE versus SNR, two-level nested array, $M = 6$ | 34 |
| 4.2 | SV- Smoothing: Probability of resolution and probability of detection- two-level nested array. | 35 |
| 4.3 | SV Smoothing: MUSIC Spectrum and the RMSE versus SNR - MRA, $M = 6$ | 36 |
| 4.4 | SV Smoothing: Probability of resolution and probability of detection - MRA. | 37 |
| 5.1 | Sensor placement for localization. (a.) Placement based on the weighted FP cost. (b.) Placement based on the weighted log-det cost. | 46 |
| 5.2 | (a.) Sensor selection for the different costs for $K = 4$. (b.) Computational time versus K from an available set of $M = 80$ sensors. (c.) RMSE versus the number of selected sensors. | 47 |
| A.1 | Sparse Arrays. | 57 |

List of Tables

| | | |
|-----|--|----|
| 3.1 | Generation of Quasi-Stationary Signals. | 17 |
| 4.1 | Summary of subarray selection (Example) | 29 |
| A.1 | Some MRA configurations. | 56 |
| A.2 | Optimal two-level nested array configurations for even and odd M . . . | 56 |

Notations

| | |
|------------------------------|---|
| a | Single element |
| \mathbf{a} | Vector |
| \mathbf{A} | Matrix |
| \mathbf{a}^T | Transpose of \mathbf{a} |
| \mathbf{a}^H | Hermitian of \mathbf{a} |
| \mathbf{a}^* | Conjugate of \mathbf{a} , without transpose |
| $\mathbb{E}(\cdot)$ | Expectation operator |
| \mathbf{I}_M | $M \times M$ identity matrix |
| $\mathbf{1}_M(\mathbf{0}_M)$ | $M \times 1$ vector of all ones(zeros) |
| $\ \mathbf{a}\ _p$ | p -norm of \mathbf{a} |
| \otimes | Kronecker product |
| \circ | Khatri-Rao product |
| $\text{vec}(\cdot)$ | Vectorization operator |
| $\text{diag}(\cdot)$ | Vectorization of diagonal elements |
| $\text{Tr}(\cdot)$ | Trace operator |
| \mathbf{A}^{-1} | Inverse of matrix \mathbf{A} |

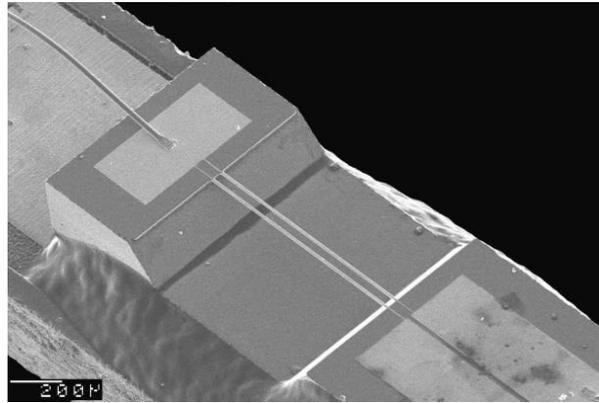
Spatial sensor arrays have found applications in several areas like seismology, optical imaging, acoustic imaging, radio astronomy and radar systems, to list a few. In all of these applications, the existing wavefield is assumed to be generated by radiation from some sources, and that it contains information about the parameters characterizing these sources. The main goal of array signal processing is *parameter estimation*, achieved by combining both the temporal and spatial outputs of the array. In this context, direction of arrival (DOA), or angle of arrival (AOA), estimation is a problem where the source bearings have to be determined. Traditionally, the problem of DOA estimation has been addressed using uniform linear arrays (ULAs) of scalar sensors, where the time-difference-of-arrival (TDOA) between scalar sensors is used to extract directional information of sources.

Pressure sensor arrays have been widely used in array processing and source localization because it is straightforward to process their outputs. With recent advances in sensor technology, it has become possible to manufacture more sophisticated sensors—those which measure quantities such as velocity and acceleration in addition to pressure. One such device is the *acoustic vector sensor* (AVS) which has made the acoustic particle velocity a measurable quantity; see Figure 1.1a. The sensor consists of three orthogonally placed velocity sensors. Each velocity sensor consists of two parallel placed platinum wires and the temperature difference across these wires generates a voltage difference which is proportional to the acoustic particle velocity. One such velocity sensor is shown in Figure 1.1b. Since the AVS observes more information in general, it has been shown to have a better estimation performance than conventional pressure sensors. An omnidirectional pressure sensor cannot identify source directions since it has equal response in all directions. A single vector sensor, on the other hand, contains directional information of sources in its measurements and can, hence, identify directions unambiguously. The inherent directivity of a vector sensor also allows construction of smaller vector sensor arrays in comparison to microphone arrays, in addition to providing better range and angle accuracy.

A commonly used method for DOA estimation is a subspace based technique known as MULTiple SIGNAL Classification (MUSIC) [3]. In particular, MUSIC can identify up to $M - 1$ DOAs using a uniform linear array of M scalar sensors. However, non-uniform arrays have found considerable interest since they are not constrained by the resolution limit of uniform arrays. Recently, it has been shown that the DOAs of more sources than the number of available sensors can be estimated. These aim at constructing the array more smartly. The processing of data for DOA estimation using such sparse arrays usually takes place in the covariance domain, and, hence, it is referred to as *covariance sensing*. In this work, we address the problem of covariance sensing with



(a) Acoustic vector sensor [1]



(b) Velocity sensor [2]

Figure 1.1: Acoustic Vector Sensor

vector sensor arrays, and this is a subject of the first part of this work.

In addition to estimating the source DOAs, it might be necessary to guarantee a certain parameter estimation accuracy. This is a subject of the second part of the thesis which addresses the problem of *sensor selection*. Sensor selection has become largely popular in several applications like localization, tracking, and remote-sensing. Modern approaches use spatially deployed sensors to measure complex spatio-temporal phenomena. Typically, a large number of sensors are involved in making observations in the applications and therefore the sensing process can be expensive. Hence, it is desirable to select the best subset of sensors from the available set so that a certain estimation accuracy is ensured. The proposed framework can be extended to any data model with additive noise when the measurements are non-linear in the unknown parameter. Although we do not explicitly apply the proposed method to an example of sensor selection for the AVS data model, it could be extended to the problem of DOA estimation with AVSs. Such a sensor selection can be applied to design sparse AVS arrays that also ensure a certain quality of DOA estimates next to their identifiability.

1.1 Applications of Vector Sensors

The proposed framework using sparse AVS arrays could be used in several applications. Vector sensors have been used to provide solutions in a diverse range of fields such as:

1. *Sound localization and tracking-* Arrays of vector sensors have found applications in acoustic as well as electromagnetic source localization and tracking. In particular, vector sensors have found widespread usage in military applications like gunshot localization, Doppler processing of passing aircrafts and missiles as well as mapping of point of origin, trajectory and point of impact of artillery. Furthermore, since the AVS has an operating range of 0.1Hz – 10kHz, it has been used to locate infrasound (or low frequency sound) sources. AVSs have also been used in near-field sound source localization in reverberant rooms and reverberant space structures. Recent applications of AVSs include wideband source localization as well as underwater source localization.
2. *Noise and Vibration Analysis-* The property of vector sensor arrays to provide high spatial resolution has been used in noise and vibration analysis of components. Noise contribution analysis and appliance monitoring are pertinent to the automotive and space industry where the machinery has to be tested for quality purposes. The acoustic vector sensor in particular is capable of providing the acoustic particle velocity map in the entire frequency range.
3. *Room Acoustics-* The ability of AVSs to measure particle velocity has made it suitable to applications which require the analysis of inhomogenous surfaces and their characteristics. This can, in turn, provide information on the acoustic reflection, impedance and absorption characteristics of materials. Moreover, AVSs can be used to map points of specular reflection which affect the acoustic properties of a room. Recent applications of AVSs allow the measurement and mapping of energy densities and energy flow in the room, which provide information about its acoustic properties.

1.2 Motivation

The aim of this thesis is motivated by the following two aspects:

- *DOA estimation using acoustic vector sensor arrays:* Sparse arrays are formed by reducing the number of redundant spacings between sensors in the array. Then, the number of redundant correlations between array elements, or the redundant differences in the co-domain, are also reduced. Some examples of such sparse arrays which provide all the phase-related differences in the co-domain are Minimum Redundancy Array (MRAs) [4], Nested Arrays [5] and the Co-prime Arrays [6]. Similarly, sparse AVS array topologies can also be constructed. Since AVSs also measure the acoustic particle velocity vector, it is expected that the sparse vector sensor array is able to estimate more DOAs than that possible with a uniform

linear scalar sensor array. For applications like source localization and tracking as in Section 1.1, a natural consequence is that more sources can be localized than that possible with a uniform linear AVS array. Hence, it is interesting to investigate the maximum number of sources that can be identified by such a sparse AVS array. Another reason to study the problem of covariance sensing with vector sensor arrays is that a sparse array constructed with the same number of AVSs as an AVS-ULA has a larger aperture. In addition to identifying more source DOAs, the sparse AVS array can be expected to have a better resolution performance than the AVS-ULA.

- *Sensor selection for non-linear measurement models:* While the first part of the thesis deals with the identifiability analysis, or the conditions for unique DOA estimation, it is also important to guarantee a pre-defined performance measure, like DOA estimation accuracy, in certain critical applications. For instance, in military applications like gunshot localization and target tracking, it may be necessary to have a high estimated angle accuracy like 1° . In such cases, sensor selection finds the optimal set of sensors so that this accuracy is achieved. Usually, we want to select K out of M existing sensors where $K \ll M$, and a straightforward method of sensor selection is to perform a search over all $\binom{M}{K}$ combinations and select the combination that best meets the our performance criterion. Clearly, such an approach is intractable for large values of M . Hence, it is solved with several approximations - convex optimization and greedy heuristics.

Convex optimization generally have a better performance with respect to the chosen cost function than greedy approaches if the cost to be minimized is convex. On the other hand, greedy heuristics aim at finding computationally attractive solutions to optimality. For a chosen cost function, convex optimization based approaches generally have a better performance, but are computationally more intensive. This has been illustrated in Figure 1.2. The sensor selection problem is even more aggravated for non-linear models because the performance measure depends on the unknown parameter itself. It is interesting to combine the benefits of the two approaches to solve the sensor selection for non-linear models. That is, the aim is to develop costs that can be optimized with a low complexity algorithm, as well as guarantee the performance of the such a greedy algorithm.

1.3 Problem Statement

The problem that this thesis addresses is two-fold:

- How many sources DOAs can be identified using a sparse, linear AVS array using covariance processing?
- How does one formulate performance measures for the non-linear model that are independent of the unknown parameter? How does the error performance of the greedy algorithm compare with that of convex optimization based approaches?

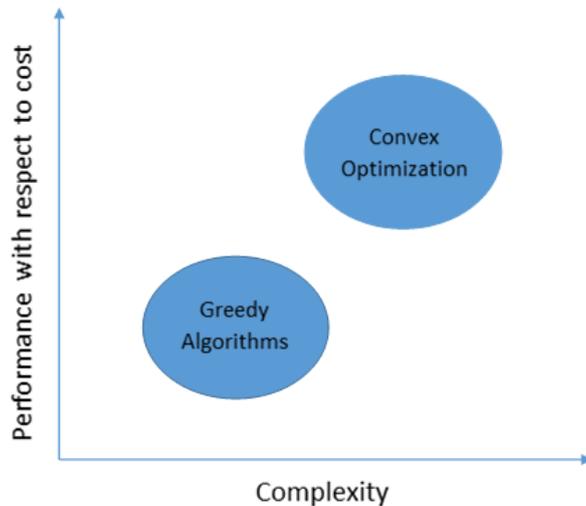


Figure 1.2: Computational complexity and performance for non-linear data models

1.4 Thesis Outline

In this section, we describe the content of this thesis and highlight the main contributions. This thesis is a two-part work. The first part of the thesis addresses the problem of compressive covariance sensing using vector sensor arrays. In particular, we discuss the concepts of sparse arrays and covariance based processing for existing scalar sensor arrays. These will be extended to AVS arrays. The first contribution of the thesis addresses the DOA estimation of quasi-stationary sources, which have a time-variant covariance matrix. Further, we analyze the number of source DOAs that can be estimated. The second contribution addresses the DOA estimation of stationary sources, where the source covariance matrix is time-invariant. A new variant of the spatial smoothing technique of [5], called the spatial-velocity smoothing due to its nature of using both the spatial- and velocity-domain invariance, is proposed. This allows the estimation of DOAs of even stationary sources.

The second part of the thesis addresses the problem of sensor selection for non-linear measurement models. We propose a *submodular* cost function which is independent of the unknown parameter. The concept of submodularity is similar to concept of diminishing returns. That is, adding a sensor index that maximizes a certain submodular function to a set is less beneficial than adding the same index to a subset of that set. Moreover, we leverage the concept of submodularity to assure the performance of the proposed greedy algorithm by using a classic result of [7]. We use a greedy algorithm to optimize the cost function which is computationally less intensive compared to convex optimization based approaches. The reader is referred to Appendix A and Appendix B for preliminaries and literature study on covariance sensing and sensor selection, respectively.

Part I

This part of the thesis discusses the covariance processing using sparse vector sensor arrays. We use sparse AVS arrays, such as the nested array or the MRA, which result in a co-array that is uniform and linear. We focus on 1D-DOA estimation of strictly uncorrelated sources. In particular, we consider *quasi-stationary* and *stationary* sources.

Chapter 2: Data model: In this chapter, we introduce the time-domain data model for an AVS array as well as the correlation-domain data model on which DOA estimation is investigated.

Chapter 3: Quasi-stationary sources: For quasi-stationary sources, we use a Khatri-Rao subspace approach from [8] and extend it to a sparse AVS array. Specifically, we can identify 5α sources, where α represents the number of sources that can be identified with a sparse scalar array. We also discuss the full-rank conditions of the array manifold in the co-domain and identifiability issues with subspace based techniques like MUSIC.

Chapter 4: Stationary Sources: For stationary sources, we propose spatial-velocity smoothing to handle the time-invariance of the source covariance matrix and to identify up to $1.5(\alpha + 1)$ sources. Here, we discuss the full-rank conditions of the array manifold in the co-domain and identifiability issues using MUSIC.

Part II

Chapter 5: Greedy Sensor Selection for Non-Linear Measurement Models:

In this chapter, the non-linear function is linearized using a first-order Taylor approximation around points where the parameter is likely to reside. A submodular cost is proposed which does not depend on the unknown parameter but uses the knowledge of the domain where the parameter resides. The cost function is based on *frame potential*, a measure of the orthogonality of the rows of the measuring matrix. Similar to the frame potential, we use the submodularity of the D-optimal criterion to propose a second cost function for non-linear measurement models. Further, a greedy algorithm to minimize the submodular cost function is proposed. Using the fundamental result of Nemhauser et al. [7], the performance of the algorithm is guaranteed with respect to the proposed cost functions.

1.5 Publications

1. Shilpa Rao, Sundeep Prabhakar Chepuri, and Geert Leus. “DOA Estimation Using Sparse Vector Sensor Arrays”. Submitted to CAMSAP 2015.
2. Shilpa Rao, Sundeep Prabhakar Chepuri, and Geert Leus. “Greedy Sensor Selection for Non-Linear Models”. Submitted to CAMSAP 2015.
3. Shilpa Rao, Sundeep Prabhakar Chepuri, and Geert Leus. “Compressive Covariance Sensing using Acoustic Vector Sensors”. Journal in preparation.

Part I

**DOA Estimation using Sparse AVS
Arrays**

Data Model - AVS Setup

In this chapter, we briefly present the data model for a general AVS array. Then, we proceed to the signal model in the covariance domain.

2.1 Data Model

We assume that the sound wave is travelling in a homogenous and isotropic medium and that the waves impinging on the array are planar. For the sake of simplicity, we assume that the acoustic sources are narrowband and that the number of sources are known *a priori*. We consider the problem of DOA estimation using a non-uniform linear AVS array with the array being constructed either with the MRA [4] or nested array [5]. In general, any array whose difference co-array is a filled ULA can be used. Let us consider D incoherent narrowband sources with distinct azimuth DOAs $\theta_d \in [-\pi, \pi)$, $d = 1, 2, \dots, D$. It is also assumed that each AVS measures the acoustic pressure as well as two components of the acoustic velocity, i.e., $\cos \theta_d$ and $\sin \theta_d$, $d = 1, 2, \dots, D$. The output of the m th AVS is then a three-element complex vector given by

$$\mathbf{y}_m(t) = \begin{bmatrix} y_{m,p}(t) \\ \mathbf{y}_{m,v}(t) \end{bmatrix} = \sum_{d=1}^D \begin{bmatrix} 1 \\ \mathbf{u}_d \end{bmatrix} p_d(t) + \mathbf{n}_m(t) \quad (2.1)$$

where $y_{m,p}$ and $\mathbf{y}_{m,v}$ are, respectively, the pressure and velocity sensor outputs and $p_d(t)$ is d th source signal. Here, \mathbf{u}_d is the bearing vector of the d th source given by $\mathbf{u}_d = [\cos \theta_d, \sin \theta_d]^T$, where θ_d is the azimuth angle of the d th source. Finally, $\mathbf{n}_m(t)$ represents the noise on the m th AVS. Let $\mathbf{n}_m(t) = [n_{m,p}(t), \mathbf{n}_{m,v}(t)]^T$ where $n_{m,p}(t)$ and $\mathbf{n}_{m,v}(t)$ represent the noise on the pressure and velocity channels, respectively. The noise covariance matrix for a single AVS is given by:

$$\mathbf{R}_n = \mathbb{E}[\mathbf{n}_m(t)\mathbf{n}_m^H(t)] = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \beta\sigma^2\mathbf{I}_2 \end{bmatrix} \in \mathbb{R}^{3 \times 3}$$

where σ^2 and $\beta\sigma^2$ are the noise variances on the pressure and velocity channels respectively, and the factor $\beta \in \mathbb{R}$ depends on how the noise is modeled. This has been discussed in [9]. Collecting the outputs of M such vector sensors, we get

$$\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{p}(t) + \mathbf{n}(t) \quad (2.2)$$

where $\mathbf{y}(t) = [\mathbf{y}_1^H(t), \mathbf{y}_2^H(t), \dots, \mathbf{y}_M^H(t)]^H \in \mathbb{C}^{3M}$ is the array output, $\mathbf{p}(t) = [p_1(t), p_2(t), \dots, p_D(t)]^T$ is a vector containing the D source signals, and $\mathbf{n}(t) =$

$[\mathbf{n}_1^H(t), \mathbf{n}_2^H(t), \dots, \mathbf{n}_M^H(t)]^H$. Here, $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_D]^T$ is the vector of unknown source DOAs and has to be estimated. Further, the array manifold is given by $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1) \mathbf{a}(\theta_2) \cdots \mathbf{a}(\theta_D)]$, where $\mathbf{a}(\theta_d)$ is given by

$$\begin{aligned} \mathbf{a}(\theta_d) &= \mathbf{g}(\theta_d) \otimes \mathbf{h}(\theta_d) \in \mathbb{C}^{3M}, \\ \mathbf{g}(\theta_d) &= [1, e^{j2\pi(r_2 \cos \theta_d)/\lambda}, \dots, e^{j2\pi(r_M \cos \theta_d)/\lambda}]^T \in \mathbb{C}^M, \\ \mathbf{h}(\theta_d) &= [1, \mathbf{u}_d^T]^T \in \mathbb{R}^3. \end{aligned} \quad (2.3)$$

Here, λ is the wavelength and r_m is the m th marking on the nested array or MRA as shown in Figure A.1. For the sake of convenience, we assume that the first element of the sparse AVS array is located at the origin, i.e., $r_1 = 0$.

2.2 Correlation Domain Data Model

We briefly describe the model based on the second-order statistics. In the correlation domain, the array model in (2.2) can be re-written as

$$\mathbf{R}_y = \mathbb{E}[\mathbf{y}(t)\mathbf{y}^H(t)] = \mathbf{A}\mathbf{R}_p\mathbf{A}^H + \mathbf{I}_M \otimes \mathbf{R}_n \quad (2.4)$$

where \mathbf{R}_y is the data correlation matrix and \mathbf{R}_p is the source correlation matrix. We have dropped the dependence of the array manifold \mathbf{A} on $\boldsymbol{\theta}$ for brevity. We further assume uncorrelated sources, i.e., \mathbf{R}_p is a diagonal matrix.

In practice, \mathbf{R}_y is estimated by local averaging over, say, N samples. That is,

$$\hat{\mathbf{R}}_y = \frac{1}{N} \sum_{t=1}^N \mathbf{y}(t)\mathbf{y}^H(t). \quad (2.5)$$

Using the property $\text{vec}(\mathbf{A}\text{diag}(\mathbf{b})\mathbf{C}) = (\mathbf{C}^T \circ \mathbf{A})\mathbf{b}$, where \circ denotes the Khatri-Rao product, $\text{vec}(\cdot)$ denotes vectorization, and $\text{diag}(\cdot)$ refers to a diagonal matrix with the vector in its argument on the main diagonal, we vectorize \mathbf{R}_y to get

$$\mathbf{r}_y = (\mathbf{A}^* \circ \mathbf{A})\mathbf{r}_p + \mathbf{e}. \quad (2.6)$$

Here, $\mathbf{r}_y = \text{vec}(\mathbf{R}_y) \in \mathbb{R}^{9M^2}$, $\mathbf{r}_p = \text{diag}(\mathbf{R}_p)$, where $\text{diag}(\cdot)$ in this context forms a vector from the main-diagonal elements, and $\mathbf{e} = \text{vec}(\mathbf{I}_M \otimes \mathbf{R}_n)$. The KR product is defined as

$$\mathbf{A}^* \circ \mathbf{A} = [\mathbf{a}^*(\theta_1) \otimes \mathbf{a}(\theta_1), \mathbf{a}^*(\theta_2) \otimes \mathbf{a}(\theta_2), \dots, \mathbf{a}^*(\theta_D) \otimes \mathbf{a}(\theta_D)] \in \mathbb{C}^{9M^2 \times D}. \quad (2.7)$$

Since σ^2 and β are known, the term \mathbf{e} is deterministic and can be subtracted from (2.6) to arrive at

$$\mathbf{z} = \mathbf{r}_y - \mathbf{e} = (\mathbf{A}^* \circ \mathbf{A})\mathbf{r}_p. \quad (2.8)$$

Comparing the above equation with (2.2), we see that \mathbf{z} behaves like the output of a virtual array whose manifold is given by $\mathbf{A}^* \circ \mathbf{A}$. The rows of $\mathbf{A}^* \circ \mathbf{A}$ are such that

the KR product contains AVSs whose locations are given by the difference set \mathcal{S} as discussed in Appendix A. Thus, the difference set occurs naturally in problems where the second-order statistics are taken. Instead of performing DOA estimation on the signal model in (2.2), we may use the model in (2.8). Further, it is known that $\mathbf{A}^* \circ \mathbf{A}$ contains a ULA due to the properties of the MRA and nested array. However, since velocity measurements corresponding to the cosine and sine terms are also present, we expect the DOF of the AVS co-array to be even higher than that of a scalar-sensor ULA. In the next chapters, we discuss DOA estimation of two kinds of sources, viz., quasi-stationary sources and stationary sources using the signal model in (2.8) .

Remark: If σ^2 is unknown, the term \mathbf{e} can be projected out using the orthogonal projection matrix \mathbf{P} where \mathbf{P} is given by

$$\mathbf{P} = \mathbf{I}_{3M} - \frac{\mathbf{e}\mathbf{e}^T}{\|\mathbf{e}\|_2^2}.$$

Quasi-Stationary Sources

In this chapter, we propose an approach based on the quasi-stationarity of sources to estimate the DOAs of sources using the model in (2.8). It can be noticed that only a single measurement vector is available from (2.8) which makes the DOA estimation problem difficult. However, there is a certain class of signals whose second-order statistics vary over time but the statistics remain constant over short intervals. These are called *quasi-stationary* signals. We use the Khatri-Rao subspace approach of [8] for DOA estimation of quasi-stationary sources using the covariance model of (2.8).

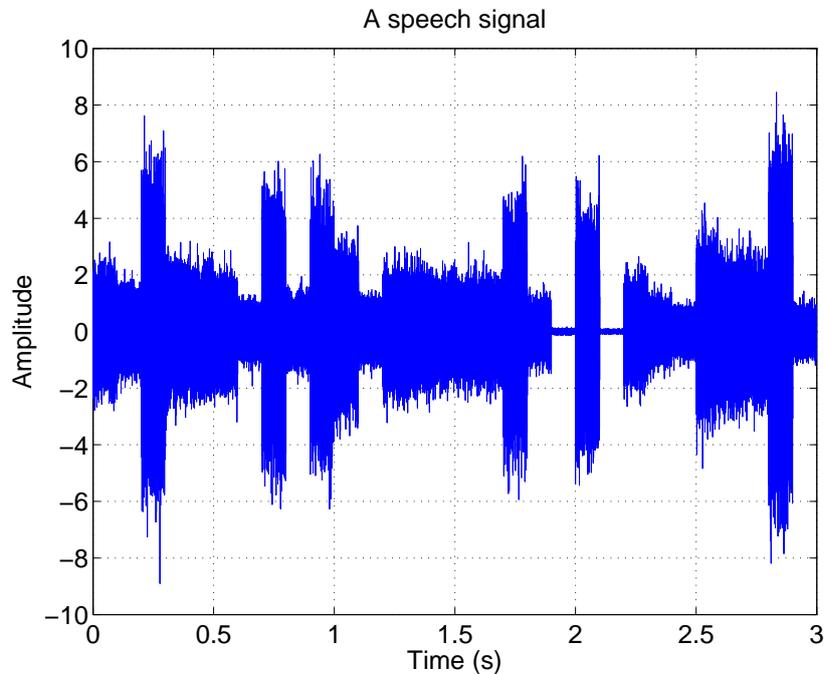


Figure 3.1: A speech signal

3.1 Khatri-Rao Subspace Approach

For quasi-stationary sources, the following assumption is made about the model. Each source signal $p_d(t)$ is assumed to be quasi-stationary with a frame length L , i.e.,

$$\mathbb{E}[|p_d(t)|^2] = r_{d,f} \quad \forall t \in [(f-1)L, fL-1] \\ f = 1, 2, \dots, F,$$

where $r_{d,f}$ is the variance of the d th source for frame index f . By allowing the source to be quasi-stationary, the second-order statistics of the source signal is time-varying, but $r_{d,f}$ is constant over duration of the frame. Moreover, for certain signals such as audio and speech, quasi-stationarity of the sources is a reasonable assumption. Figure 3.1 shows an example of one such speech signal.

Then, re-writing (2.8) for each frame f , we get

$$\mathbf{z}_f = (\mathbf{A}^* \circ \mathbf{A}) \mathbf{r}_{p,f} \quad (3.1)$$

where \mathbf{z}_f is the vector \mathbf{z} from (2.8) with the subscript f indicating the frame index. Similarly $\mathbf{r}_{p,f} = [r_{1,f}, r_{2,f}, \dots, r_{D,f}]^T$. To build a full rank correlation matrix, the source powers are allowed to vary in the different quasi-stationary intervals. Collecting the data over F different intervals and defining $\mathbf{Z} = [\mathbf{z}_1 \ \mathbf{z}_2 \ \dots \ \mathbf{z}_F]$, we get

$$\mathbf{Z} = (\mathbf{A}^* \circ \mathbf{A}) \mathbf{\Psi} \quad (3.2)$$

with $\mathbf{\Psi}$ being given by

$$\mathbf{\Psi} = [\mathbf{r}_{p,1} \ \mathbf{r}_{p,2} \ \dots \ \mathbf{r}_{p,F}] = \begin{bmatrix} r_{1,1} & r_{1,2} & \dots & r_{1,F} \\ r_{2,1} & r_{2,2} & \dots & r_{2,F} \\ \vdots & \vdots & \vdots & \vdots \\ r_{D,1} & r_{D,2} & \dots & r_{D,F} \end{bmatrix}. \quad (3.3)$$

Thus, the columns of $\mathbf{\Psi}$ contain the source powers in the frames.

The matrices \mathbf{Z} and $(\mathbf{A}^* \circ \mathbf{A})$ span the same column space when $\mathbf{\Psi}$ is full row rank. The identifiability analysis and the conditions for which the KR product $(\mathbf{A}^* \circ \mathbf{A})$ has full column rank is discussed separately in Section 3.2. The singular value decomposition (SVD) of \mathbf{Z} gives

$$\mathbf{Z} = \begin{bmatrix} \mathbf{U}_s & \mathbf{U}_n \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_s^H \\ \mathbf{V}_n^H \end{bmatrix}, \quad (3.4)$$

where $\mathbf{\Sigma}_s \in \mathbb{R}^{D \times D}$ is a diagonal matrix containing the signal powers, $\mathbf{U}_s \in \mathbb{C}^{9M^2 \times D}$ and $\mathbf{V}_s \in \mathbb{C}^{F \times D}$ are the left and right singular matrices associated with the signal subspace, respectively, and $\mathbf{U}_n \in \mathbb{C}^{9M^2 \times (9M^2 - D)}$ and $\mathbf{V}_n \in \mathbb{C}^{F \times (9M^2 - D)}$ are the left and right singular matrices associated with the noise subspace, respectively. Hence, the following subspace criterion could be used

$$\mathbf{U}_n^H [\mathbf{a}^*(\theta_d) \otimes \mathbf{a}(\theta_d)] = \mathbf{0}, \quad d = 1, 2, \dots, D. \quad (3.5)$$

Further, the KR-MUSIC spectrum is given as:

$$\mathbf{J}_{\text{MUSIC}}(\theta) = \frac{1}{\|\mathbf{U}_n^H [\mathbf{a}^*(\theta) \otimes \mathbf{a}(\theta)]\|^2}. \quad (3.6)$$

3.2 Full Rank Condition of KR Product

Since the full rank of $(\mathbf{A}^* \circ \mathbf{A})$ in (3.2) is crucial to the application of the subspace-based technique, it is essential to analyze under what conditions $(\mathbf{A}^* \circ \mathbf{A})$ is full rank. This section presents an analysis that leads to the full rank condition of the KR-product $(\mathbf{A}^* \circ \mathbf{A})$. Recalling that \mathbf{A} is given by

$$\mathbf{A} = [\mathbf{g}(\theta_1) \otimes \mathbf{h}(\theta_1), \mathbf{g}(\theta_2) \otimes \mathbf{h}(\theta_2), \dots, \mathbf{g}(\theta_D) \otimes \mathbf{h}(\theta_D)],$$

and that the d th column of \mathbf{A} is

$$\mathbf{a}(\theta_d) = \mathbf{g}(\theta_d) \otimes \mathbf{h}(\theta_d) = \begin{bmatrix} 1 \\ e^{j2\pi(r_2 \cos \theta_d)/\lambda} \\ \vdots \\ e^{j2\pi(r_M \cos \theta_d)/\lambda} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ \cos \theta_d \\ \sin \theta_d \end{bmatrix},$$

the KR-product $(\mathbf{A}^* \circ \mathbf{A})$ is then given by

$$\mathbf{A}^* \circ \mathbf{A} = [(\mathbf{g}^*(\theta_1) \otimes \mathbf{h}(\theta_1)) \otimes (\mathbf{g}(\theta_1) \otimes \mathbf{h}(\theta_1)), \dots, (\mathbf{g}^*(\theta_D) \otimes \mathbf{h}(\theta_D)) \otimes (\mathbf{g}(\theta_D) \otimes \mathbf{h}(\theta_D))].$$

Next, for the sake of exposition, we consider a single column of $(\mathbf{A}^* \circ \mathbf{A})$. The d th column of $(\mathbf{A}^* \circ \mathbf{A})$, $(\mathbf{g}^*(\theta_d) \otimes \mathbf{h}(\theta_d)) \otimes (\mathbf{g}(\theta_d) \otimes \mathbf{h}(\theta_d))$, can be re-written as

$$(\mathbf{g}^*(\theta_d) \otimes \mathbf{h}(\theta_d)) \otimes (\mathbf{g}(\theta_d) \otimes \mathbf{h}(\theta_d)) = \mathbf{E}(\mathbf{g}^*(\theta_d) \otimes \mathbf{g}(\theta_d)) \otimes (\mathbf{h}(\theta_d) \otimes \mathbf{h}(\theta_d)) \quad (3.7)$$

where $\mathbf{E} \in \mathbb{R}^{9M^2 \times 9M^2}$ is a known square exchange matrix of full rank. Further, we know from the properties of MRAs and nested arrays that the distinct rows of $(\mathbf{g}^*(\theta_d) \otimes \mathbf{g}(\theta_d))$ represent a ULA and can be collected in a vector $\tilde{\mathbf{g}}(\theta_d)$. That is, $\tilde{\mathbf{g}}(\theta_d) = [\psi_d^{\frac{-(\alpha-1)}{2}}, \dots, 1, \dots, \psi_d^{\frac{(\alpha-1)}{2}}]^T \in \mathbb{C}^\alpha$. Then, $\mathbf{g}^*(\theta_d) \otimes \mathbf{g}(\theta_d)$ can be rewritten as

$$\mathbf{g}^*(\theta_d) \otimes \mathbf{g}(\theta_d) = \mathbf{T}_G \begin{bmatrix} \psi_d^{\frac{-(\alpha-1)}{2}} \\ \vdots \\ 1 \\ \vdots \\ \psi_d^{\frac{(\alpha-1)}{2}} \end{bmatrix} = \mathbf{T}_G \tilde{\mathbf{g}}(\theta_d) \quad (3.8)$$

where $\mathbf{T}_G \in \mathbb{R}^{M^2 \times \alpha}$ is a known tall permutation matrix that depends on the kind of sparse array initially chosen and $\psi_d = e^{(j2\pi d \cos \theta_d/\lambda)}$, d being the minimum array spacing.

The value of α is obtained from Table A.1 for an MRA or calculated from Table A.2 for a nested array. Similarly, $\mathbf{h}(\theta_d) \otimes \mathbf{h}(\theta_d)$ can be written as

$$\mathbf{h}(\theta_d) \otimes \mathbf{h}(\theta_d) = \mathbf{T}_H \begin{bmatrix} 1 \\ \cos \theta_d \\ \sin \theta_d \\ \cos \theta_d \sin \theta_d \\ \cos^2 \theta_d \end{bmatrix} = \mathbf{T}_H \tilde{\mathbf{h}}(\theta_d) \quad (3.9)$$

where $\mathbf{T}_H \in \mathbb{R}^{9 \times 5}$ is also a tall permutation matrix. Similar to $\tilde{\mathbf{g}}(\theta_d)$, $\tilde{\mathbf{h}}(\theta_d)$ is defined by taking the distinct rows of $\mathbf{h}(\theta_d) \otimes \mathbf{h}(\theta_d)$, i.e., $\tilde{\mathbf{h}}(\theta_d) = [1, \cos \theta_d, \sin \theta_d, \cos \theta_d \sin \theta_d, \cos^2 \theta_d]^T \in \mathbb{R}^5$. Combining (3.8) and (3.9), we can re-write (3.7) as

$$(\mathbf{g}^*(\theta_d) \otimes \mathbf{h}(\theta_d)) \otimes (\mathbf{g}(\theta_d) \otimes \mathbf{h}(\theta_d)) = \underbrace{\mathbf{E}(\mathbf{T}_G \otimes \mathbf{T}_H)}_{\mathbf{B}' } (\tilde{\mathbf{g}}(\theta_d) \otimes \tilde{\mathbf{h}}(\theta_d)) \quad (3.10)$$

where $\mathbf{B}' = \mathbf{E}(\mathbf{T}_G \otimes \mathbf{T}_H)$ and we have used the property: $(\mathbf{WY}) \otimes (\mathbf{XV}) = (\mathbf{W} \otimes \mathbf{X})(\mathbf{Y} \otimes \mathbf{V})$. Thus, $(\mathbf{A}^* \circ \mathbf{A})$ can be represented as

$$\mathbf{A}^* \circ \mathbf{A} = \mathbf{B}' [\tilde{\mathbf{g}}(\theta_1) \otimes \tilde{\mathbf{h}}(\theta_1), \tilde{\mathbf{g}}(\theta_2) \otimes \tilde{\mathbf{h}}(\theta_2), \dots, \tilde{\mathbf{g}}(\theta_D) \otimes \tilde{\mathbf{h}}(\theta_D)] = \mathbf{B}' \tilde{\mathbf{A}} \quad (3.11)$$

where $\tilde{\mathbf{A}} = [\tilde{\mathbf{g}}(\theta_1) \otimes \tilde{\mathbf{h}}(\theta_1), \tilde{\mathbf{g}}(\theta_2) \otimes \tilde{\mathbf{h}}(\theta_2), \dots, \tilde{\mathbf{g}}(\theta_D) \otimes \tilde{\mathbf{h}}(\theta_D)] \in \mathbb{C}^{5\alpha}$.

Since \mathbf{E} , \mathbf{T}_G and \mathbf{T}_H are full column rank matrices, the rank of $(\mathbf{A}^* \circ \mathbf{A})$ depends on the rank of $\tilde{\mathbf{A}}$. It can be seen that the number of distinct rows in $\tilde{\mathbf{A}}$ is directly related to the degrees of freedom. The full rank of the effective array $(\mathbf{A}^* \circ \mathbf{A})$ is ensured by the linear independence of steering vectors. Firstly, we see that the parameter θ can take values in the interval $[0, 2\pi)$. This definition for θ is radically different from that of a scalar-sensor array which cannot differentiate between two angles θ_1 and θ_2 which are related as $\theta_2 = 2\pi - \theta_1$. Hence, the steering vectors $\{\tilde{\mathbf{g}}(\theta_1) \otimes \tilde{\mathbf{h}}(\theta_1)\}$ and $\{\tilde{\mathbf{g}}(\theta_2) \otimes \tilde{\mathbf{h}}(\theta_2)\}$ related to θ_1 and θ_2 are not parallel. To prove the full column rank of matrix $\tilde{\mathbf{A}}$, we introduce the following lemma.

Lemma 1: For $D \leq \alpha$, if the steering vectors of a scalar-sensor array $\{\tilde{\mathbf{g}}(\theta_1), \tilde{\mathbf{g}}(\theta_2), \dots, \tilde{\mathbf{g}}(\theta_D)\}$ are linearly independent, then the steering vectors of the AVS array $\{\tilde{\mathbf{g}}(\theta_1) \otimes \tilde{\mathbf{h}}(\theta_1), \tilde{\mathbf{g}}(\theta_2) \otimes \tilde{\mathbf{h}}(\theta_2), \dots, \tilde{\mathbf{g}}(\theta_D) \otimes \tilde{\mathbf{h}}(\theta_D)\}$ are also linearly independent.

Proof. Appendix A: Proof of Lemma 1. ■

Recall that the number of entries in $\tilde{\mathbf{g}}(\theta_d)$, $\forall d$ is equal to α . A natural result of lemma 1 is that for $D \leq \alpha$ and D distinct DOAs, the matrix $\tilde{\mathbf{A}}$ is full column rank. The result can be extended for $\alpha < D \leq 5\alpha$.

Theorem 1. : For $\alpha < D \leq 5\alpha$, if the D sources have distinct DOAs, then the steering vectors $\{\tilde{\mathbf{g}}(\theta_1) \otimes \tilde{\mathbf{h}}(\theta_1), \tilde{\mathbf{g}}(\theta_2) \otimes \tilde{\mathbf{h}}(\theta_2), \dots, \tilde{\mathbf{g}}(\theta_D) \otimes \tilde{\mathbf{h}}(\theta_D)\}$ are linearly independent.

Proof. Appendix B: Proof of Theorem 1. ■

The proofs for the above two theorems can be found at the end of the chapter. A direct consequence of the above theorem and lemma 1 is that with an M -element linear sparse AVS array, up to 5α steering vectors are linearly independent. Thus, $(\mathbf{A}^* \circ \mathbf{A}) \in \mathbb{C}^{9M^2 \times D}$ is full column rank when $D \leq 5\alpha$. For DOA estimation using MUSIC described in the previous sub-section the number of distinct DOAs that can be identified is $5\alpha - 1$.

3.3 MUSIC Identifiability

In this section, we discuss the necessary and sufficient condition for unique identifiability of source DOAs using MUSIC.

Theorem 2. The subspace criterion of (3.5) holds for a true DOA $\theta_d, d = 1, 2, \dots, D$, if and only if

$$D \leq 5\alpha - 1. \quad (3.12)$$

Proof. Assume that an angle $\varphi \notin \{\theta_1, \theta_2, \dots, \theta_D\}$ exists such that the subspace criterion of (3.5) is satisfied. That is,

$$\mathbf{U}_n^H [\mathbf{a}^*(\varphi) \otimes \mathbf{a}(\varphi)] = \mathbf{0}. \quad (3.13)$$

Recall that $\mathbf{U}_n \in \mathbb{C}^{9M^2 \times (9M^2 - D)}$. Since the dimension of \mathbf{U}_n^H is only $9M^2 - D$ and not $9M^2 - (D + 1)$, it means that $\mathbf{a}^*(\varphi) \otimes \mathbf{a}(\varphi)$ is linearly dependent on $\{\mathbf{a}^*(\theta_1) \otimes \mathbf{a}(\theta_1), \dots, \mathbf{a}^*(\theta_D) \otimes \mathbf{a}(\theta_D)\}$. Hence,

$$\mathbf{a}^*(\varphi) \otimes \mathbf{a}(\varphi) = \sum_{d=1}^D \mathbf{a}^*(\theta_d) \otimes \mathbf{a}(\theta_d). \quad (3.14)$$

However, we know from Theorem 1, that for $D + 1 \leq 5\alpha$, and for unique $\{\theta_1, \theta_2, \dots, \theta_D, \varphi\}$, $[\mathbf{A} \ \mathbf{a}(\varphi)]^* \circ [\mathbf{A} \ \mathbf{a}(\varphi)]$ has linearly independent columns. Thus, (3.13) is satisfied if and only if $D + 1 > 5\alpha$. Hence, the condition for unique source identifiability is $D \leq 5\alpha - 1$. ■

3.4 Simulation Results - Quasi-stationary sources

In this section we provide several simulations to illustrate the performance of the KR product based MUSIC applied on an AVS nested array as well as the AVS-MRA. In particular, the performance is evaluated in terms of the root mean squared error (RMSE),

probability of detection and the probability of resolving closely spaced sources. For the sake of illustration, the two-level nested array with $M = 6$ is considered where the sensor locations are given by $\{0, 1, 2, 3, 7, 11\}$. For comparison, we also show simulation results for an MRA with 6 AVSs whose locations are given by $\{0, 1, 2, 6, 10, 13\}$. Further, quasi-stationary sources are generated artificially as in Table 3.1 [8] where L_{low} and L_{upp} , the upper and lower limit of the frame length, are fixed to 300 and 500, respectively. The frame length for each quasi-stationary interval is drawn from a uniform random distribution $[L_{\text{low}}, L_{\text{upp}}]$ since, in practice, the local stationary periods are varying.

| | |
|--------------------|---|
| Given: | Upper limit L_{upp} , lower limit L_{low} , and a total sequence length N . |
| Initialize: | Sequence length counter $N_{\text{seq}} = 0$. |
| Step 1: | Draw frame length L from the uniform distribution $[L_{\text{low}}, L_{\text{upp}}]$. |
| Step 2: | Generate signal variance σ_d^2 randomly from a uniformly from a random distribution $[0, 1]$. |
| Step 3: | For $t = N_{\text{seq}}, N_{\text{seq}} + 1, \dots, N_{\text{seq}} + L - 1$, generate the complex signal $p_d(t)$ with real and imaginary parts drawn from a Gaussian distribution $[0, \sigma_d^2/2]$. |
| Step 4: | Update $N_{\text{seq}} = N_{\text{seq}} + L$, and repeat Step 1 if $N_{\text{seq}} < N$. |

Table 3.1: Generation of Quasi-Stationary Signals.

MUSIC Spectrum

Figure 3.2a shows the normalized MUSIC spectrum for a nested array of 6 AVSs. A total of 400 quasi-stationary intervals have been used and the SNR is assumed to be 10dB. The spectrum yields peaks at the locations of the 70 sources which are also well-resolved. In general, the number of quasi-stationary intervals needed to resolve D sources is at least $D + 1$ [5]. However, the theoretical bound on the number of sources that can be identified is $5\alpha_{\text{nested}} = 5 \times (M^2/2 + M - 1) - 1 = 5 \times 23 - 1 = 114$. Similarly, Figure 3.3a shows the normalized MUSIC spectrum for an MRA of AVSs with $M = 6$ and $D = 80$. The spectrum yields peaks at the true locations of the 80 sources and these sources are also well-resolved. The upper bound on the number of sources in this case is $5\alpha_{\text{MRA}} = 5 \times 27 - 1 = 134$.

Root Mean Squared Error (RMSE)

Next, we compare the performance in terms of the RMSE. Figure 3.2b shows the RMSE as a function of the SNR for the nested array. Since the two-level nested array of 6 AVSs has 115 unique measurements, a ULA of 38 AVSs is used for comparison which has $38 \times 3 = 114$ measurements. We also use a ULA for 6 AVSs for comparison. Further, conventional MUSIC is used on these two ULAs. The RMSE is plotted for a source at 33.4° and only 20 quasi-stationary frames are used. It is seen that the performance of the KR product based MUSIC improves with SNR and comes close to that of the conventional MUSIC on the AVS-ULA of 38 elements whereas, at low SNRs,

the AVS-ULA of 6 elements performs worse than that the nested AVS as well as the longer AVS-ULA of 38 elements.

Similarly, Figure 3.3b illustrates the RMSE versus SNR for the 6 element AVS-MRA. As before, we compare its performance to that of an AVS-ULA of 6 elements. A 45-element AVS-ULA is also used in the comparison since it has $45 \times 3 = 135 = 27 \times 5$ measurements. The performance of the KR product based MUSIC improves with SNR and comes close to that of the conventional MUSIC on the AVS-ULA of 45 elements whereas the performance of the AVS-ULA of 6 elements is even worse than that of the other two array topologies at low SNRs.

Probability of Resolution

To illustrate the performance of the sparse AVS arrays in resolving two closely spaced sources, we also plot the probability of resolution as in [10]. For this purpose, we consider two cases of source separation: (i) two sources with $\theta_1 = 30^\circ$ and $\theta_2 = 35^\circ$, $\Delta\theta = 5^\circ$, and (ii) two sources with $\theta_1 = 30^\circ$ and $\theta_2 = 32^\circ$, $\Delta\theta = 2^\circ$.

The procedure used is briefly described as follows. Define the function $b(\theta)$ as:

$$b(\theta) = 1 - \mathbf{a}^H(\theta) \left[\sum_{d=1}^D \mathbf{s}_d \mathbf{s}_d^H \right] \mathbf{a}(\theta), \quad (3.15)$$

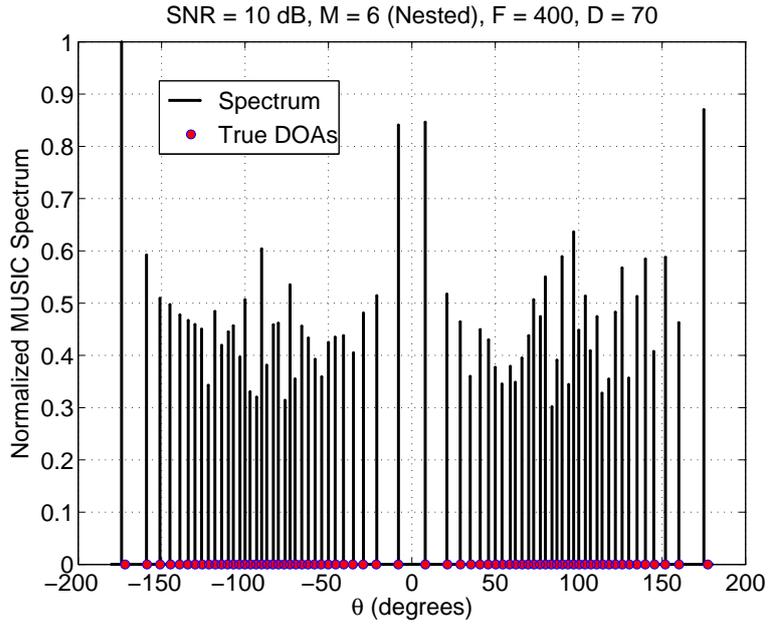
where \mathbf{s}_d is the d th left singular vector from (3.4) and $\mathbf{a}(\theta)$ is the steering direction in the direction θ . For two closely spaced equi-powered sources with DOAs θ_1 and θ_2 , [10] proposes the threshold as $b(\theta_m)$ where $\theta_m = \cos^{-1}(\frac{\cos\theta_1 + \cos\theta_2}{2})$. Since in practice $b(\theta_1)$, $b(\theta_2)$, and $b(\theta_m)$ are unknown, they are estimated from the estimated covariance matrix (2.5). Let $\hat{b}(\theta_1)$, $\hat{b}(\theta_2)$, and $\hat{b}(\theta_m)$ represent the estimates where $\hat{b}(\theta)$ is given by

$$\hat{b}(\theta) = 1 - \mathbf{a}^H(\theta) \left[\sum_{d=1}^D \hat{\mathbf{s}}_d \hat{\mathbf{s}}_d^H \right] \mathbf{a}(\theta), \quad (3.16)$$

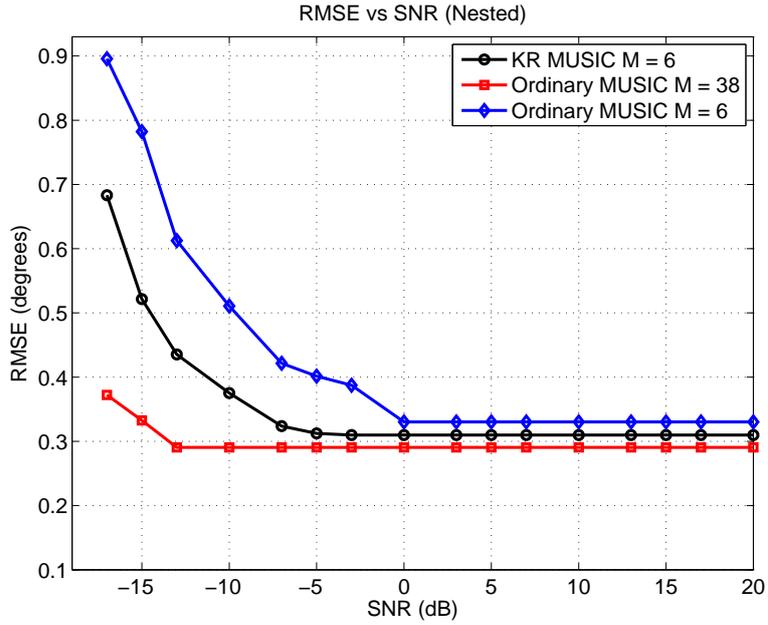
where $\hat{\mathbf{s}}_d$ is the d th estimated left singular vector. Resolution is said to be achieved when both $\hat{b}(\theta_1)$ and $\hat{b}(\theta_2)$ are less than $\hat{b}(\theta_m)$.

The number of quasi-stationary intervals is taken to be 50 and the resolution performance is averaged over 1000 Monte-Carlo simulations. The probabilities of resolution versus SNR for the KR product based MUSIC for the nested array, conventional MUSIC on the AVS-ULA of 38 elements, and conventional MUSIC on the AVS-ULA of 6 elements are plotted in Figure 3.4a. It is seen that the performance of all the three arrays improve with SNR. However, for a given SNR and a given separation of sources, the conventional MUSIC employed on the longer AVS-ULA of 38 elements clearly has a better resolution performance than the nested AVS array as well the AVS-ULA of 6 elements.

Similarly, Figure 3.4b illustrates the probability of resolution versus SNR for the AVS-MRA of 6 elements, conventional MUSIC on the AVS-ULA of 45 elements, and conventional MUSIC on the AVS-ULA of 6 elements. It is seen that both the AVS-MRA as



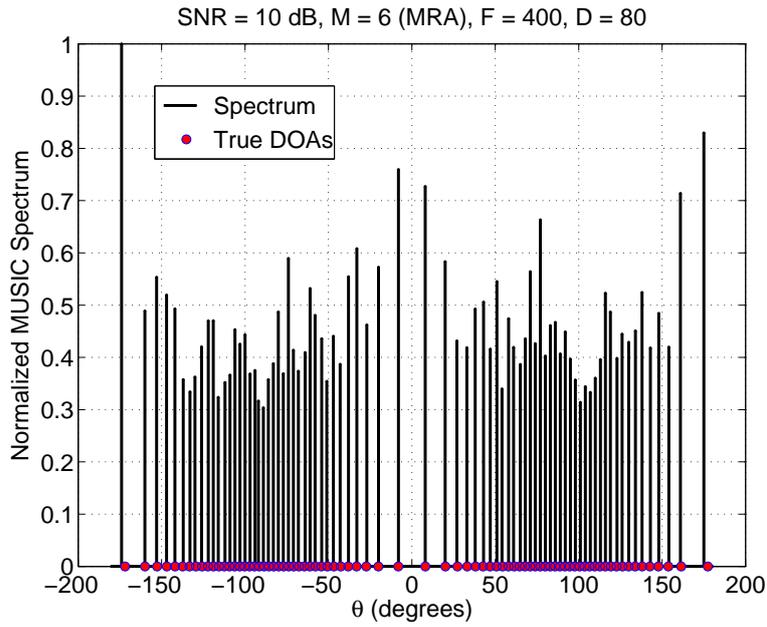
(a) KR-MUSIC: Spectrum for the two-level AVS nested array $M = 6$, $D = 70$.



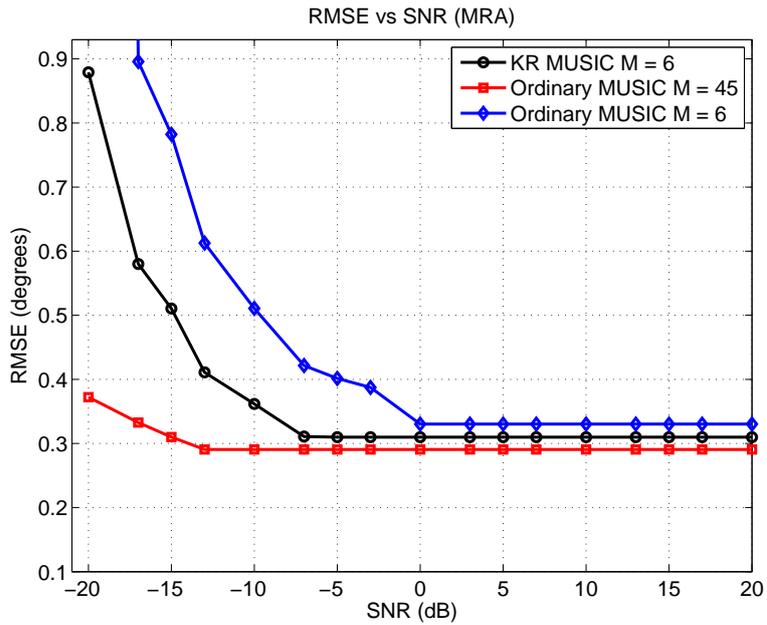
(b) KR-MUSIC: RMSE (degrees) versus SNR (dB) - two-level nested array.

Figure 3.2: KR-MUSIC: MUSIC Spectrum and RMSE - Two-level AVS nested array.

well as the AVS-ULA of 45 elements perform better than the AVS-ULA of 6 elements. Moreover, even at a smaller source separation ($\Delta\theta = 2^\circ$), the AVS-ULA of 45 elements and the AVS-MRA have a better performance than that of the AVS-ULA of 6 elements at $\Delta\theta = 5^\circ$.



(a) KR-MUSIC: Spectrum for MRA $M = 6$, $D = 80$.

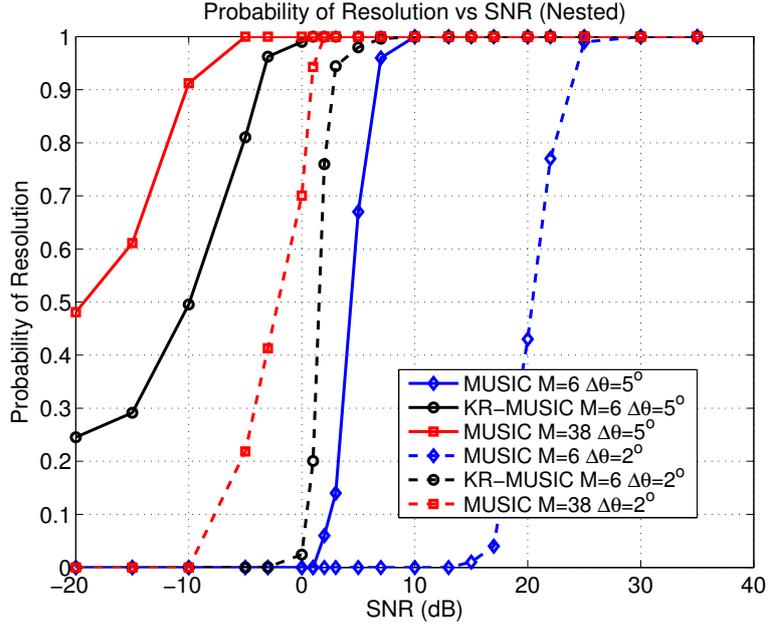


(b) KR-MUSIC: RMSE (degrees) versus SNR (dB) - AVS-MRA.

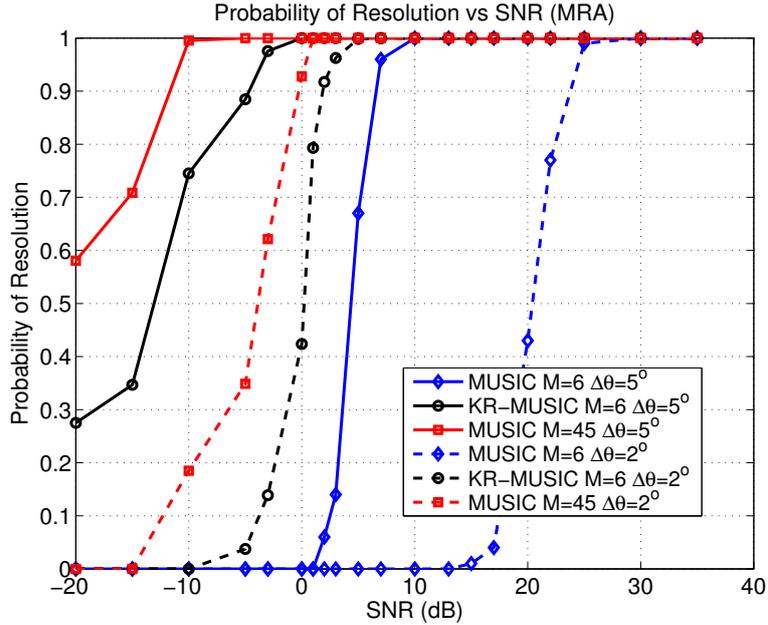
Figure 3.3: KR-MUSIC: MUSIC Spectrum and RMSE - AVS-MRA.

Probability of Detection

In the previous examples, the number of sources present was assumed to be known. However, the number of sources has to be estimated in practice. In this thesis, we use the simple singular-value threshold based technique of [11] to calculate the probability of detection empirically. The procedure is summarized as follows.



(a) KR-Subspace MUSIC: Probability of resolution - two-level nested array.



(b) KR-Subspace MUSIC: Probability of resolution - MRA.

Figure 3.4: KR-Subspace MUSIC: Probability of resolution- two level nested array and MRA.

Let $\hat{s}_j, j = 1, 2, \dots, 9M^2$, represent the j th estimated singular value from (3.4) and F be the number of quasi-stationary intervals considered. Let l_i be the average observed noise-subspace singular value given by

$$l_i = \frac{1}{9M^2 - i + 1} \left[\sum_{j=i}^{9M^2} \hat{s}_j \right].$$

The binary hypothesis for the p th iteration is given by:

$$\begin{aligned} H_0 : & D < 9M^2 - p \\ H_1 : & D = 9M^2 - p. \end{aligned} \quad (3.17)$$

Next, the either H_0 or H_1 is accepted according to:

$$\hat{s}_{9M^2-p} \underset{H_0}{\overset{H_1}{\gtrless}} \hat{s}_{9M^2-p}^u, \quad (3.18)$$

where $\hat{s}_{9M^2-p}^u$ is the upper threshold on the $(9M^2 - p)$ th singular value given by:

$$\hat{s}_{9M^2-p}^u = \left[(p+1) \frac{1 + t(F(p+1))^{-0.5}}{1 - t(Fp)^{-0.5}} - p \right] l_{9M^2-p+1}. \quad (3.19)$$

Here, t is a threshold that is fixed for a given false alarm probability P_{fa} . Consider the binary hypothesis of (3.17) and assume that H_1 is accepted at the p th step. That is, the estimated number of sources $\hat{D} = 9M^2 - p$. A false alarm occurs when $\hat{D} > D$. More specifically, P_{fa} is calculated by integrating a complex integral given by Equation (4.11) in [11]. However, for our example, we evaluate the false alarm probability numerically for various values of t as illustrated in Figure 3.5. Then, t is chosen so that the probability of false alarm is required to satisfy a certain level for a given F . In this case, for a P_{fa} of, say 1.5%, and $F = 400$, t is taken to be 2.9. Then the detection algorithm of [11] is summarized in Algorithm 1.

Algorithm 1 Detection Algorithm

Require: $9M^2$ singular values $\hat{s}_j, j = 1, 2, \dots, 9M^2$, number of quasi-stationary intervals F .

1: **Initialize:**

Iteration counter $p = 1$.

2: Check the test in (3.18).

3: If H_0 is accepted, put $p = p + 1$ and continue with Step 2.

4: If H_1 is accepted, stop testing and estimate the number of sources as $\hat{D} = 9M^2 - p$.

Figure 3.6a shows the detection performance for the AVS nested array. The probability of detection is averaged over 1000 Monte-Carlo simulations for each SNR that is considered. As expected, it is seen that the 38-element AVS-ULA shows a higher probability of detection for a given SNR. We do not consider the AVS-ULA of 6 elements since it cannot detect more than $6 \times 3 - 1 = 17$ sources.

Similarly, Figure 3.6b shows the detection performance for the AVS-MRA. The AVS-ULA of 45 elements has a slightly better performance than the AVS-MRA of 6 elements.

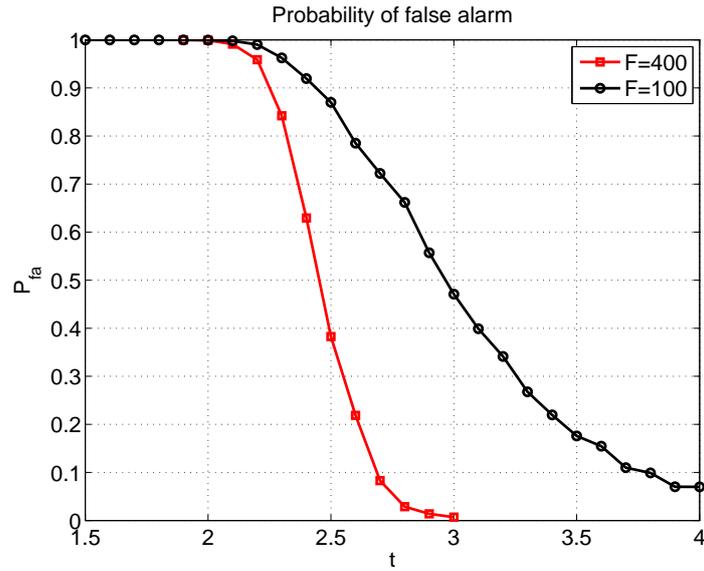
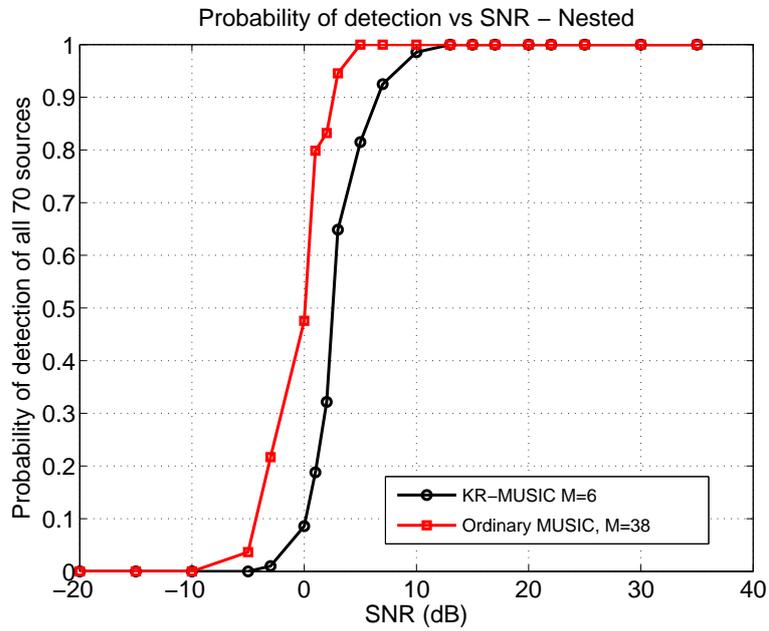


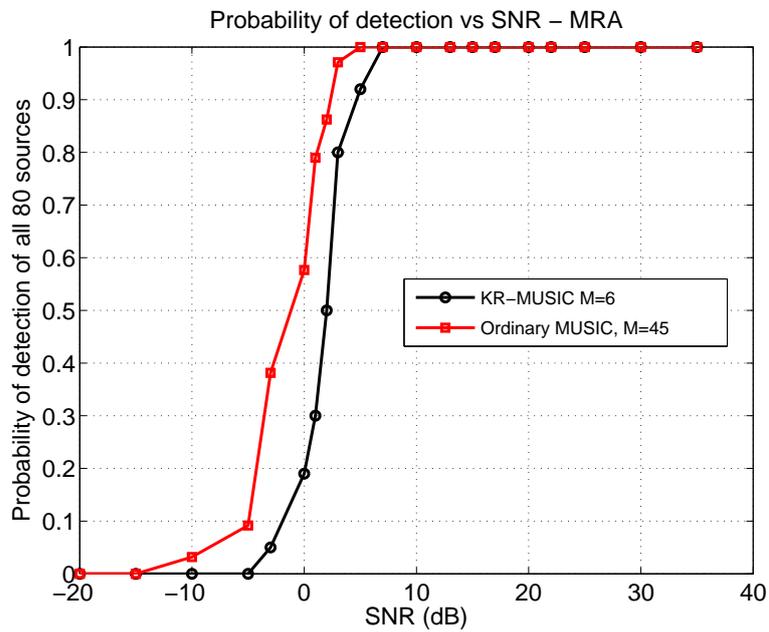
Figure 3.5: Probability of false alarm: Two level nested array $M = 6$

3.5 Conclusions

In this chapter, we investigated the upper bound on the number of sources that can be localized using a sparse AVS array by leveraging the time-varying nature of second-order statistics of quasi-stationary sources. Further, we proposed a tight upper bound on the number of DOAs that can be identified, based on the linear independence of steering vectors. More specifically, the number of source DOAs that can be uniquely identified with a sparse AVS array with MUSIC is $5\alpha - 1$, where α is the DOF of the sparse scalar sensor array. Through extensive simulations, we also verified the advantages of using a sparse AVS array over a ULA of the same number of sensors. Moreover, the sparse AVS array with only 6 elements has a resolution and detection performance that comes very close to a much longer AVS-ULA with an equivalent number of measurements.



(a) KR-Subspace MUSIC: Probability of Detection - Two level nested array.



(b) KR-Subspace MUSIC: Probability of detection - MRA

Figure 3.6: KR-Subspace MUSIC: Probability of detection - Two level nested array and MRA.

Appendix

3.A Appendix A: Proof of Lemma 1

First, assume that $\{\tilde{\mathbf{g}}(\theta_1) \otimes \tilde{\mathbf{h}}(\theta_1), \tilde{\mathbf{g}}(\theta_2) \otimes \tilde{\mathbf{h}}(\theta_2), \dots, \tilde{\mathbf{g}}(\theta_D) \otimes \tilde{\mathbf{h}}(\theta_D)\}$ are linearly dependent. Then

$$\sum_{d=1}^D c_d (\tilde{\mathbf{g}}(\theta_d) \otimes \tilde{\mathbf{h}}(\theta_d)) = \mathbf{0}_{5\alpha \times 1} = \sum_{d=1}^D \tilde{\mathbf{g}}(\theta_d) \otimes (c_d \tilde{\mathbf{h}}(\theta_d)) \quad (3.20)$$

where not all the c_d , $d = 1, 2, \dots, D$ are zeros. Since $\{\tilde{\mathbf{g}}(\theta_1), \tilde{\mathbf{g}}(\theta_2), \dots, \tilde{\mathbf{g}}(\theta_D)\}$ are linearly independent, it follows that either $c_d = 0$ or $\tilde{\mathbf{h}}(\theta_d) = 0$, $d = 1, 2, \dots, D$. Since neither is possible, our assumption is invalid. Therefore, $\{\tilde{\mathbf{g}}(\theta_1) \otimes \tilde{\mathbf{h}}(\theta_1), \tilde{\mathbf{g}}(\theta_2) \otimes \tilde{\mathbf{h}}(\theta_2), \dots, \tilde{\mathbf{g}}(\theta_D) \otimes \tilde{\mathbf{h}}(\theta_D)\}$ are linearly independent.

3.B Appendix B: Proof of Theorem 1

To show that as many as 5α steering vectors of $\tilde{\mathbf{A}}$ are linearly independent, we will use mathematical induction.

More specifically, we show that for D such that $\alpha \leq D \leq 5\alpha - 1$, if the D steering vectors $\{\tilde{\mathbf{g}}(\theta_1) \otimes \tilde{\mathbf{h}}(\theta_1), \tilde{\mathbf{g}}(\theta_2) \otimes \tilde{\mathbf{h}}(\theta_2), \dots, \tilde{\mathbf{g}}(\theta_D) \otimes \tilde{\mathbf{h}}(\theta_D)\}$ are linearly independent, then the $D+1$ steering vectors $\{\tilde{\mathbf{g}}(\theta_1) \otimes \tilde{\mathbf{h}}(\theta_1), \tilde{\mathbf{g}}(\theta_2) \otimes \tilde{\mathbf{h}}(\theta_2), \dots, \tilde{\mathbf{g}}(\theta_{D+1}) \otimes \tilde{\mathbf{h}}(\theta_{D+1})\}$ are also linearly independent. In lemma 1, it was shown that for $D = \alpha$, the D steering vectors $\{\tilde{\mathbf{g}}(\theta_1) \otimes \tilde{\mathbf{h}}(\theta_1), \tilde{\mathbf{g}}(\theta_2) \otimes \tilde{\mathbf{h}}(\theta_2), \dots, \tilde{\mathbf{g}}(\theta_D) \otimes \tilde{\mathbf{h}}(\theta_D)\}$ are linearly independent.

Assume that $\tilde{\mathbf{g}}(\theta_{D+1}) \otimes \tilde{\mathbf{h}}(\theta_{D+1})$ is linearly dependent on $\tilde{\mathbf{g}}(\theta_1) \otimes \tilde{\mathbf{h}}(\theta_1), \tilde{\mathbf{g}}(\theta_2) \otimes \tilde{\mathbf{h}}(\theta_2), \dots, \tilde{\mathbf{g}}(\theta_D) \otimes \tilde{\mathbf{h}}(\theta_D)$. We get

$$\tilde{\mathbf{g}}(\theta_{D+1}) \otimes \tilde{\mathbf{h}}(\theta_{D+1}) = \sum_{d=1}^D c_d \tilde{\mathbf{g}}(\theta_d) \otimes \tilde{\mathbf{h}}(\theta_d) \quad (3.21)$$

where none of the c_d are zeros. This is because if any $c_d = 0$, the set of D steering vectors $\{\tilde{\mathbf{g}}(\theta_1) \otimes \tilde{\mathbf{h}}(\theta_1), \dots, \tilde{\mathbf{g}}(\theta_{d-1}) \otimes \tilde{\mathbf{h}}(\theta_{d-1}), \tilde{\mathbf{g}}(\theta_{d+1}) \otimes \tilde{\mathbf{h}}(\theta_{d+1}), \dots, \tilde{\mathbf{g}}(\theta_{D+1}) \otimes \tilde{\mathbf{h}}(\theta_{D+1})\}$ will constitute a linearly independent set. Similar to the first part, $\tilde{\mathbf{g}}(\theta_{D+1})$ can be expressed as

$$\begin{aligned} \tilde{\mathbf{g}}(\theta_{D+1}) &= \sum_{d=1}^D e_d \tilde{\mathbf{g}}(\theta_d) \\ \Rightarrow \tilde{\mathbf{g}}(\theta_{D+1}) \otimes \tilde{\mathbf{h}}(\theta_{D+1}) &= \left(\sum_{d=1}^D e_d \tilde{\mathbf{g}}(\theta_d) \right) \otimes \tilde{\mathbf{h}}(\theta_{D+1}) \end{aligned} \quad (3.22)$$

where not all of the $e_d, d = 1, 2, \dots, D$ are zeros. Subtracting (3.21) from (3.22), we get

$$\sum_{d=1}^D \tilde{\mathbf{g}}(\theta_d) \otimes (e_d \tilde{\mathbf{h}}(\theta_{D+1}) - c_d \tilde{\mathbf{h}}(\theta_d)) = \mathbf{0}_{5\alpha \times 1} \quad (3.23)$$

Let

$$e_d \tilde{\mathbf{h}}(\theta_{D+1}) - c_d \tilde{\mathbf{h}}(\theta_d) = f_d \mathbf{z}_d$$

where f_d and \mathbf{z}_d are the amplitude and normalized direction of the new vector. If we show that all the f_d are non-zero and all $\mathbf{z}_d, d = 1, 2, \dots, D$ are different, we prove the linear independence of the vectors $\{\tilde{\mathbf{g}}(\theta_1) \otimes \tilde{\mathbf{h}}(\theta_1), \tilde{\mathbf{g}}(\theta_2) \otimes \tilde{\mathbf{h}}(\theta_2), \dots, \tilde{\mathbf{g}}(\theta_{D+1}) \otimes \tilde{\mathbf{h}}(\theta_{D+1})\}$.

1. If $f_d = 0, d = 1, 2, \dots, D$, then $e_d \tilde{\mathbf{h}}(\theta_{D+1}) = c_d \tilde{\mathbf{h}}(\theta_d)$. This implies that $\tilde{\mathbf{h}}(\theta_d)$ and $\tilde{\mathbf{h}}(\theta_{D+1})$ are parallel which contradicts our previous assumption. Thus, none of the $f_d, d = 1, 2, \dots, D$ are zeros.
2. To prove that all the $\mathbf{z}_d, d = 1, 2, \dots, D$ are different, assume without loss of generality, that $\mathbf{z}_1 = \mathbf{z}_2$. Then

$$\frac{e_1}{f_1} \tilde{\mathbf{h}}(\theta_{D+1}) - \frac{c_1}{f_1} \tilde{\mathbf{h}}(\theta_1) = \mathbf{z}_1 = \mathbf{z}_2 = \frac{e_2}{f_2} \tilde{\mathbf{h}}(\theta_{D+1}) - \frac{c_2}{f_2} \tilde{\mathbf{h}}(\theta_2). \quad (3.24)$$

So

$$\left(\frac{e_1}{f_1} - \frac{e_2}{f_2} \right) \tilde{\mathbf{h}}(\theta_{D+1}) - \frac{c_1}{f_1} \tilde{\mathbf{h}}(\theta_1) + \frac{c_2}{f_2} \tilde{\mathbf{h}}(\theta_2) = \mathbf{0}_{5 \times 1}. \quad (3.25)$$

Since $\tilde{\mathbf{h}}(\theta_{D+1}), \tilde{\mathbf{h}}(\theta_1)$ and $\tilde{\mathbf{h}}(\theta_2)$ are linearly independent, it is implied that $c_1 = c_2 = 0$, and $\frac{e_1}{f_1} = \frac{e_2}{f_2}$. Since neither of these possibilities are true, equation (3.25) does not hold.

This proves the linear independence of the steering vectors $\{\tilde{\mathbf{g}}(\theta_1) \otimes \tilde{\mathbf{h}}(\theta_1), \tilde{\mathbf{g}}(\theta_2) \otimes \tilde{\mathbf{h}}(\theta_2), \dots, \tilde{\mathbf{g}}(\theta_{D+1}) \otimes \tilde{\mathbf{h}}(\theta_{D+1})\}$ for any D such that $\alpha \leq D \leq 5\alpha - 1$. Thus, the steering vectors $\{\tilde{\mathbf{g}}(\theta_1) \otimes \tilde{\mathbf{h}}(\theta_1), \tilde{\mathbf{g}}(\theta_2) \otimes \tilde{\mathbf{h}}(\theta_2), \dots, \tilde{\mathbf{g}}(\theta_{5\alpha}) \otimes \tilde{\mathbf{h}}(\theta_{5\alpha})\}$ are linearly independent.

Stationary Sources

Previously, we discussed a method to exploit the degrees of freedom available in the co-array through the assumption of quasi-stationary sources. In this chapter, we propose an alternative method to DOA estimation using the co-array.

To build a rank- D covariance matrix from (2.8), the authors of [8] and [12] use the time-variant nature of \mathbf{z} for quasi-stationary sources. This was discussed for an array of AVSs in Section 3. However, this method cannot be applied to stationary sources since the covariance matrix for stationary sources is time-invariant. This chapter presents an approach to handle stationary sources based on a technique, which we call *spatial-velocity (SV) smoothing*. This approach is slightly different from the spatial smoothing techniques discussed in [5] and [13]. In order to exploit the degrees of freedom of the co-array, spatial smoothing in [5] builds a covariance matrix of rank D from \mathbf{z} since the full column rank of the array manifold is key to the application of subspace-based techniques. In [13], spatial smoothing is used on the measurement model (2.2) to decorrelate correlated sources. Here, the proposed approach is used to strictly enhance the rank of covariance matrix by smoothing along both the spatial and velocity dimensions. The basic principle is to divide the uniform linear co-array into overlapping subarrays of the same length and to average the resulting different covariance matrices.

4.1 Extraction of Rows

The method works as follows. First, the unique phase differences from $\mathbf{A}^* \circ \mathbf{A}$ are extracted while the velocity-related self-correlations $\mathbf{h}(\theta_d) \otimes \mathbf{h}(\theta_d)$, $d = 1, 2, \dots, D$ are retained. We define a new matrix $\hat{\mathbf{A}}$ whose columns are now given by

$$\begin{aligned} \hat{\mathbf{A}} &= [\tilde{\mathbf{g}}(\theta_1) \otimes \mathbf{h}(\theta_1) \otimes \mathbf{h}(\theta_1), \dots, \tilde{\mathbf{g}}(\theta_D) \otimes \mathbf{h}(\theta_D) \otimes \mathbf{h}(\theta_D)] \\ &= [\tilde{\mathbf{g}}(\theta_1) \otimes \hat{\mathbf{h}}(\theta_1), \dots, \tilde{\mathbf{g}}(\theta_D) \otimes \hat{\mathbf{h}}(\theta_D)] \in \mathbb{C}^{9\alpha \times D}, \end{aligned} \quad (4.1)$$

where $\hat{\mathbf{h}}(\theta_d) = \mathbf{h}(\theta_d) \otimes \mathbf{h}(\theta_d) \in \mathbb{R}^{9 \times 1}$ and we recall that $\tilde{\mathbf{g}}(\theta_d)$ contains the phase differences sorted from $-(\alpha - 1)/2$ to $(\alpha - 1)/2$. That is

$$\tilde{\mathbf{g}}(\theta_d) = [\psi_d^{-\frac{\alpha-1}{2}}, \dots, \psi_d^{\frac{\alpha-1}{2}}]^T \in \mathbb{C}^{\alpha \times 1}.$$

where $\psi_d = e^{j(2\pi/\lambda)d \cos \theta_d}$. We first illustrate the formation of these subarrays with an example.

4.2 Formation of Subarrays: Example

For the sake of exposition, we illustrate the construction of subarrays for $\alpha = 3$. Then, the d th column of $\widehat{\mathbf{A}}$, which we denote by $\widehat{\mathbf{a}}(\theta_d)$, is given by

$$\widehat{\mathbf{a}}(\theta_d) = \begin{bmatrix} \psi_d^{-1} \\ 1 \\ \psi_d \end{bmatrix} \otimes \begin{bmatrix} 1 \\ \cos \theta_d \\ \sin \theta_d \\ \hline \cos \theta_d \\ \cos^2 \theta_d \\ \cos \theta_d \sin \theta_d \\ \hline \sin \theta_d \\ \cos \theta_d \sin \theta_d \\ \sin^2 \theta_d \end{bmatrix} = \widetilde{\mathbf{g}}(\theta_d) \otimes \widehat{\mathbf{h}}(\theta_d).$$

The d th column of the first subarray, $\widehat{\mathbf{a}}_1(\theta_d)$, is now constructed as

$$\widehat{\mathbf{a}}_1(\theta_d) = \begin{bmatrix} 1 \\ \psi_d \end{bmatrix} \otimes \begin{bmatrix} 1 \\ \cos \theta_d \\ \sin \theta_d \end{bmatrix} \in \mathbb{C}^{6 \times 1}.$$

These rows correspond to the terms $[1, \cos \theta_d, \sin \theta_d]^T$ of $\widehat{\mathbf{h}}(\theta_d)$ and $[1, \psi_d]^T$ of $\widetilde{\mathbf{g}}(\theta_d)$. More specifically, these are the rows of $\widehat{\mathbf{a}}(\theta_d)$ which are given by the expression $9j + k$ where j and k span $\{1, 2\}$ and $\{1, 2, 3\}$ respectively. Similarly, the d th column of the second subarray, $\widehat{\mathbf{a}}_2(\theta_d)$, is constructed as

$$\widehat{\mathbf{a}}_2(\theta_d) = \begin{bmatrix} \psi_d^{-1} \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ \cos \theta_d \\ \sin \theta_d \end{bmatrix} = \psi_d^{-1} \widehat{\mathbf{a}}_1(\theta_d).$$

These are the rows $9(j - 1) + k$ of $\widehat{\mathbf{a}}(\theta_d)$ where j and k again span $\{1, 2\}$ and $\{1, 2, 3\}$ respectively. Extending this further, the d th columns of the third, fourth, fifth and sixth subarrays, respectively, are constructed as

$$\begin{aligned} \widehat{\mathbf{a}}_3(\theta_d) &= \begin{bmatrix} 1 \\ \psi_d \end{bmatrix} \otimes \begin{bmatrix} \cos \theta_d \\ \cos^2 \theta_d \\ \cos \theta_d \sin \theta_d \end{bmatrix} = \cos \theta_d \widehat{\mathbf{a}}_1(\theta_d), \\ \widehat{\mathbf{a}}_4(\theta_d) &= \begin{bmatrix} \psi_d^{-1} \\ 1 \end{bmatrix} \otimes \begin{bmatrix} \cos \theta_d \\ \cos^2 \theta_d \\ \cos \theta_d \sin \theta_d \end{bmatrix} = \psi_d^{-1} \widehat{\mathbf{a}}_3(\theta_d) = \cos \theta_d \psi_d^{-1} \widehat{\mathbf{a}}_1(\theta_d), \\ \widehat{\mathbf{a}}_5(\theta_d) &= \begin{bmatrix} 1 \\ \psi_d \end{bmatrix} \otimes \begin{bmatrix} \sin \theta_d \\ \cos \theta_d \sin \theta_d \\ \sin^2 \theta_d \end{bmatrix} = \sin \theta_d \widehat{\mathbf{a}}_1(\theta_d), \\ \widehat{\mathbf{a}}_6(\theta_d) &= \begin{bmatrix} \psi_d^{-1} \\ 1 \end{bmatrix} \otimes \begin{bmatrix} \sin \theta_d \\ \cos \theta_d \sin \theta_d \\ \sin^2 \theta_d \end{bmatrix} = \psi_d^{-1} \widehat{\mathbf{a}}_5(\theta_d) = \sin \theta_d \psi_d^{-1} \widehat{\mathbf{a}}_1(\theta_d). \end{aligned}$$

Thus, for $\widehat{\mathbf{a}}_i(\theta_d)$ with $i = 3, 4$, the rows that are selected from $\widehat{\mathbf{a}}(\theta_d)$ are given by $9(3 + j - i) + k$ where j and k span $\{1, 2\}$ and $\{4, 5, 6\}$ respectively. Whereas, for $\widehat{\mathbf{a}}_i(\theta_d)$ with $i = 5, 6$, the rows that are selected from $\widehat{\mathbf{a}}(\theta_d)$ are given by $9(5 + j - i) + k$ where j and k span $\{1, 2\}$ and $\{7, 8, 9\}$ respectively. These results are summarized in Table 4.1.

| Subarray i | $\widehat{\mathbf{a}}_i(\theta_d)$ | Rows of $\widehat{\mathbf{a}}(\theta_d)$ | j, k |
|--------------|---|--|---------------|
| 2 | $\psi_d^{-1}\widehat{\mathbf{a}}_1(\theta_d)$ | $9(j - 1) + k$ | $k = 1, 2, 3$ |
| 3 | $\cos\theta_d\widehat{\mathbf{a}}_1(\theta_d)$ | $9j + k$ | $k = 4, 5, 6$ |
| 4 | $\cos\theta_d\psi_d^{-1}\widehat{\mathbf{a}}_1(\theta_d)$ | $9(j - 1) + k$ | $k = 4, 5, 6$ |
| 5 | $\sin\theta_d\widehat{\mathbf{a}}_1(\theta_d)$ | $9j + k$ | $k = 7, 8, 9$ |
| 6 | $\sin\theta_d\psi_d^{-1}\widehat{\mathbf{a}}_1(\theta_d)$ | $9(j - 1) + k$ | $k = 7, 8, 9$ |

Table 4.1: Summary of subarray selection (Example)

4.3 Formation of Subarrays: General Procedure

Equipped with the above example, we may now generalize the method for any α . The array response for the i th subarray is denoted by $\widehat{\mathbf{A}}_i$, and it corresponds to certain rows of the matrix $\widehat{\mathbf{A}}$. More specifically, the array response of the first subarray is $\widehat{\mathbf{A}}_1 = [\widehat{\mathbf{a}}_1(\theta_1) \widehat{\mathbf{a}}_1(\theta_2) \cdots \widehat{\mathbf{a}}_1(\theta_D)] \in \mathbb{C}^{3(\alpha+1)/2 \times D}$ where the d th column of $\widehat{\mathbf{A}}_1$ is given by

$$\widehat{\mathbf{a}}_1(\theta_d) = [1, \psi_d, \dots, \psi_d^{\frac{(\alpha-1)}{2}}]^T \otimes [1, \cos\theta_d, \sin\theta_d]^T. \quad (4.2)$$

For any i , we may write

- For $i = 1, 2, \dots, (\alpha+1)/2$, the i th subarray is constructed from the $9(\frac{\alpha-1}{2} + j - i) + k$ rows of $\widehat{\mathbf{A}}$ where $j \in \{1, 2, \dots, (\alpha+1)/2\}$ and $k \in \{1, 2, 3\}$.
- For $i = (\alpha+1)/2 + 1, \dots, (\alpha+1)$, the i th subarray is constructed from the $9(\frac{\alpha-1}{2} + j - i + \frac{(\alpha+1)}{2}) + k$ rows of $\widehat{\mathbf{A}}$ where $j \in \{1, 2, \dots, (\alpha+1)/2\}$ and $k \in \{4, 5, 6\}$.
- For $i = (\alpha+2), \dots, 3(\alpha+1)/2$, the i th subarray is constructed from the $9(\frac{\alpha-1}{2} + j - i + (\alpha+1)) + k$ rows of $\widehat{\mathbf{A}}$ where $j \in \{1, 2, \dots, (\alpha+1)/2\}$ and $k \in \{7, 8, 9\}$.

When the subarrays are chosen in this manner, the i th subarray matrix $\widehat{\mathbf{A}}_i$ is related to $\widehat{\mathbf{A}}_1$ as

$$\widehat{\mathbf{A}}_i = \begin{cases} \widehat{\mathbf{A}}_1 \Phi^{i-1}, & 1 \leq i \leq (\alpha+1)/2 \\ \widehat{\mathbf{A}}_1 \Phi^{i - \frac{(\alpha+1)}{2} - 1} \Phi_c, & (\alpha+1)/2 < i \leq (\alpha+1) \\ \widehat{\mathbf{A}}_1 \Phi^{i - (\alpha+1) - 1} \Phi_s, & (\alpha+1) < i \leq 3(\alpha+1)/2, \end{cases}$$

where Φ , Φ_c and Φ_s , respectively, are given by

$$\begin{aligned}\Phi &= \begin{pmatrix} e^{-j(2\pi/\lambda)d \sin \theta_1} & & & \\ & e^{-j(2\pi/\lambda)d \sin \theta_2} & & \\ & & \ddots & \\ & & & e^{-j(2\pi/\lambda)d \sin \theta_D} \end{pmatrix}, \\ \Phi_c &= \begin{pmatrix} \cos \theta_1 & & & \\ & \cos \theta_2 & & \\ & & \ddots & \\ & & & \cos \theta_D \end{pmatrix}, \\ \Phi_s &= \begin{pmatrix} \sin \theta_1 & & & \\ & \sin \theta_2 & & \\ & & \ddots & \\ & & & \sin \theta_D \end{pmatrix}.\end{aligned}$$

4.4 Effective Model

With $\hat{\mathbf{A}}$ as the array manifold, the effective model of (2.8) becomes

$$\mathbf{z}_e = \hat{\mathbf{A}} \mathbf{r}_p \quad (4.3)$$

where \mathbf{z}_e contains the corresponding rows extracted from the vector \mathbf{z} .

Similar to the extraction of rows from $\hat{\mathbf{A}}$ for the i th subarray, the equivalent output of the i th subarray, denoted by $\hat{\mathbf{z}}_{ei}$ is formed. The same rows as the ones used to construct $\hat{\mathbf{A}}_i$ are removed from \mathbf{z}_e and stored in $\hat{\mathbf{z}}_{ei}$. More specifically,

$$\hat{\mathbf{z}}_{ei} = \hat{\mathbf{A}}_i \mathbf{r}_p \quad (4.4)$$

Then, for the i th subarray, $i = 1, 2, \dots, 3(\alpha + 1)/2$, we define $\hat{\mathbf{R}}_i$ as

$$\hat{\mathbf{R}}_i = \hat{\mathbf{z}}_{ei} \hat{\mathbf{z}}_{ei}^H = \hat{\mathbf{A}}_i \mathbf{r}_p \mathbf{r}_p^H \hat{\mathbf{A}}_i^H. \quad (4.5)$$

Averaging over the $3(\alpha + 1)/2$ subarrays, we define the smoothed matrix \mathbf{R}_{ss} as

$$\mathbf{R}_{ss} = \frac{1}{3(\alpha + 1)/2} \sum_{i=1}^{3(\alpha+1)/2} \hat{\mathbf{R}}_i. \quad (4.6)$$

Theorem 3. *The matrix \mathbf{R}_{ss} can be expressed as $\mathbf{R}_{ss} = \hat{\mathbf{R}}^2$ where*

$$\hat{\mathbf{R}} = \frac{1}{\sqrt{3(\alpha + 1)/2}} (\hat{\mathbf{A}}_1 \mathbf{R}_p \hat{\mathbf{A}}_1^H).$$

Proof. From (4.6), we have,

$$\begin{aligned}
\mathbf{R}_{\text{ss}} &= \frac{1}{3(\alpha+1)/2} \sum_{i=1}^{3(\alpha+1)/2} \widehat{\mathbf{R}}_i \\
&= \frac{1}{3(\alpha+1)/2} \widehat{\mathbf{A}}_1 \left(\sum_{i=1}^{(\alpha+1)/2} (\Phi^{i-1} \mathbf{r}_p \mathbf{r}_p^H (\Phi^{i-1})^H) \right. \\
&\quad + \sum_{i=(\alpha+1)/2+1}^{(\alpha+1)} (\Phi^{i-\frac{(\alpha+1)}{2}-1} \Phi_c \mathbf{r}_p \mathbf{r}_p^H \Phi_c (\Phi^{i-\frac{(\alpha+1)}{2}-1})^H) \\
&\quad \left. + \sum_{i=\alpha+2}^{3(\alpha+1)/2} (\Phi^{i-(\alpha+1)-1} \Phi_s \mathbf{r}_p \mathbf{r}_p^H \Phi_s (\Phi^{i-(\alpha+1)-1})^H) \right) \widehat{\mathbf{A}}_1^H \\
&= \frac{1}{3(\alpha+1)/2} \widehat{\mathbf{A}}_1 \left(\sum_{i=1}^{(\alpha+1)/2} \Phi^{i-1} (\mathbf{r}_p \mathbf{r}_p^H + \Phi_c \mathbf{r}_p \mathbf{r}_p^H \Phi_c + \Phi_s \mathbf{r}_p \mathbf{r}_p^H \Phi_s) (\Phi^{i-1})^H \right) \widehat{\mathbf{A}}_1^H \\
&= \frac{1}{3(\alpha+1)/2} \widehat{\mathbf{A}}_1 \left(\sum_{i=1}^{(\alpha+1)/2} \mathbf{R}_p \Psi_i \Psi_i^H \mathbf{R}_p \right) \widehat{\mathbf{A}}_1^H \\
&= \frac{1}{3(\alpha+1)/2} \widehat{\mathbf{A}}_1 \mathbf{R}_p \left(\sum_{i=1}^{(\alpha+1)/2} \Psi_i \Psi_i^H \right) \mathbf{R}_p \widehat{\mathbf{A}}_1^H
\end{aligned} \tag{4.7}$$

where Ψ_i is given by

$$\Psi_i = \begin{pmatrix} \psi_1^{-(i-1)} & \psi_1^{-(i-1)} \cos \theta_1 & \psi_1^{-(i-1)} \sin \theta_1 \\ \psi_2^{-(i-1)} & \psi_2^{-(i-1)} \cos \theta_2 & \psi_2^{-(i-1)} \sin \theta_2 \\ \vdots & \vdots & \vdots \\ \psi_D^{-(i-1)} & \psi_D^{-(i-1)} \cos \theta_D & \psi_D^{-(i-1)} \sin \theta_D \end{pmatrix} \in \mathbb{C}^{D \times 3}.$$

Let us define a new matrix Ψ given by

$$\begin{aligned}
&\Psi \\
&= \begin{pmatrix} 1 & \cos \theta_1 & \sin \theta_1 & \dots & \psi_1^{\frac{-(\alpha-1)}{2}} & \psi_1^{\frac{-(\alpha-1)}{2}} \cos \theta_1 & \psi_1^{\frac{-(\alpha-1)}{2}} \sin \theta_1 \\ 1 & \cos \theta_2 & \sin \theta_2 & \dots & \psi_2^{\frac{-(\alpha-1)}{2}} & \psi_2^{\frac{-(\alpha-1)}{2}} \cos \theta_2 & \psi_2^{\frac{-(\alpha-1)}{2}} \sin \theta_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \cos \theta_D & \sin \theta_D & \dots & \psi_D^{\frac{-(\alpha-1)}{2}} & \psi_D^{\frac{-(\alpha-1)}{2}} \cos \theta_D & \psi_D^{\frac{-(\alpha-1)}{2}} \sin \theta_D \end{pmatrix} \\
&= \widehat{\mathbf{A}}_1^H.
\end{aligned} \tag{4.8}$$

Noticing the structure of Ψ , we can write $\left(\sum_{i=1}^{(\alpha+1)/2} \Psi_i \Psi_i^H\right)$ as

$$\left(\sum_{i=1}^{(\alpha+1)/2} \Psi_i \Psi_i^H\right) = \Psi \Psi^H = \hat{\mathbf{A}}_1^H \hat{\mathbf{A}}_1. \quad (4.9)$$

Then, we can simplify (4.7) to

$$\begin{aligned} \mathbf{R}_{\text{ss}} &= \frac{1}{3(\alpha+1)/2} \hat{\mathbf{A}}_1 \mathbf{R}_p \hat{\mathbf{A}}_1^H \hat{\mathbf{A}}_1 \mathbf{R}_p \hat{\mathbf{A}}_1^H \\ &= \left(\frac{1}{\sqrt{3(\alpha+1)/2}} \hat{\mathbf{A}}_1 \mathbf{R}_p \hat{\mathbf{A}}_1^H\right)^2. \end{aligned} \quad (4.10)$$

■

It can be seen from Theorem 1 that the matrix $\hat{\mathbf{R}}$ has the form of a covariance matrix of an AVS-ULA with $3(\alpha+1)/2$ measurements. Further, since $\hat{\mathbf{R}}$ and \mathbf{R}_{ss} have the same set of eigenvectors, subspace based techniques can be used on \mathbf{R}_{ss} . The column space of \mathbf{R}_{ss} also spans the column space of $\hat{\mathbf{A}}_1$. As shown in [14], $\hat{\mathbf{A}}_1$ is full column rank when $D \leq 3(\alpha+1)/2$. Further, it can be shown that no two columns of $\hat{\mathbf{A}}_1$ with distinct DOAs are ambiguous. Thus, by applying the MUSIC on \mathbf{R}_{ss} , up to $3(\alpha+1)/2 - 1$ sources can be identified. We note here that although the degrees of freedom available from the co-array are higher than $3(\alpha+1)/2$, the spatial-velocity based smoothing reduces this number to $3(\alpha+1)/2$ due to the averaging of subarrays. this can be increased to $(9/4)(\alpha+1)$.

4.5 Simulation Results - Stationary Sources

For the sake of illustration, two topologies, an AVS-MRA and a two-level nested AVS array each of 6 sensors are used. For a two-level nested AVS array with $M = 6$, $5\alpha_{\text{nested}} = 5 \times 23$, but spatial-velocity smoothing reduces the number of sources that can be identified to $3(\alpha_{\text{nested}} + 1)/2 - 1 = 35$. Figure 4.1a shows the normalized MUSIC spectrum for the nested array after the SV-smoothing process. All the 32 sources are identified. Similarly, Figure 4.3a shows the normalized MUSIC spectrum for the nested array with $M = 6$. For the AVS-MRA, the number of sources that can be identified using SV-smoothing is $3(\alpha_{\text{MRA}} + 1)/2 - 1 = 41$. A total number of $N = 5000$ samples have been used, the SNR is assumed to be 10dB, and $\beta = 1$ for both cases.

Figure 4.1b shows the root mean squared error (RMSE) as a function of the SNR for a source at 33.4° . Since the two-level nested AVS array of $M = 6$ elements can effectively identify $3(\alpha+1)/2 - 1 = 35$ sources, an AVS-ULA of 12 elements which can also identify $12 \times 3 - 1 = 35$ sources has been used for comparison. The RMSE is also compared with that of an AVS-ULA of 6 elements. It is seen that the performance of MUSIC after spatial smoothing improves with SNR and comes close to that of the conventional MUSIC on the AVS-ULA of 12 elements. The AVS-ULA of 6 elements

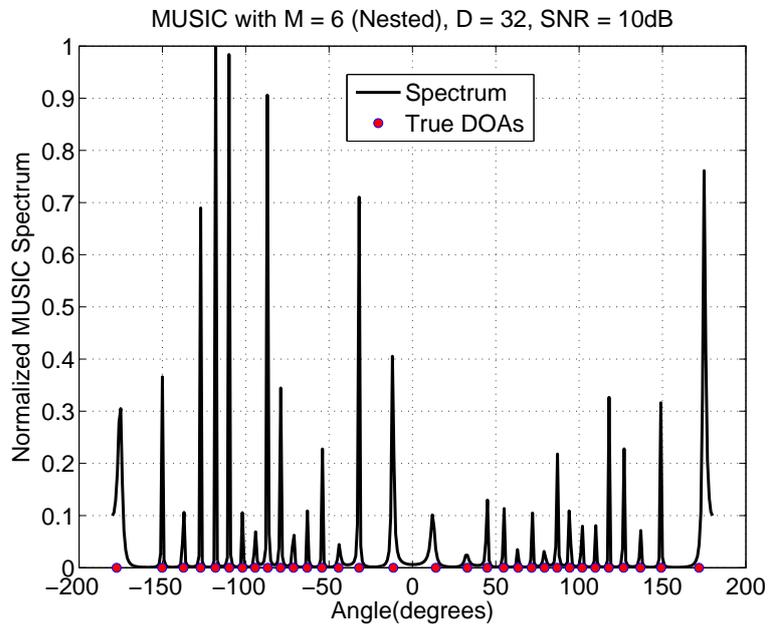
performs significantly worse than the other two array topologies. Similarly, Figure 4.3b illustrates the RMSE versus SNR for the MRA. Since the AVS-MRA of 6 elements can effectively identify $3(\alpha + 1)/2 - 1 = 41$ sources, an AVS-ULA of 14 elements is used since it also identifies $14 \times 3 - 1 = 41$ sources. Since the RMSE is evaluated for a single source, only $N = 500$ samples have been considered. The performance is also compared with that of the AVS-ULA of 6 elements.

As in Section 3.4, we also plot the probabilities of resolution and the probability of detection. To plot the probability of resolution, we consider two cases: (i) two sources with $\theta_1 = 31^\circ$ and $\theta_2 = 36^\circ$, $\Delta\theta = 5^\circ$, (ii) two sources with $\theta_1 = 31^\circ$ and $\theta_2 = 33^\circ$, $\Delta\theta = 2^\circ$. Figure 4.2a shows the probability of resolution for the two-level nested array. It is seen that the performance of the nested array is similar to that of the AVS-ULA of 12 elements. The AVS-ULA of 6 elements performs considerably worse than the nested AVS array as well as the AVS-ULA of 12 elements. Similarly, Figure 4.4a shows the probability of resolution for the MRA. For all the cases, $N = 1000$.

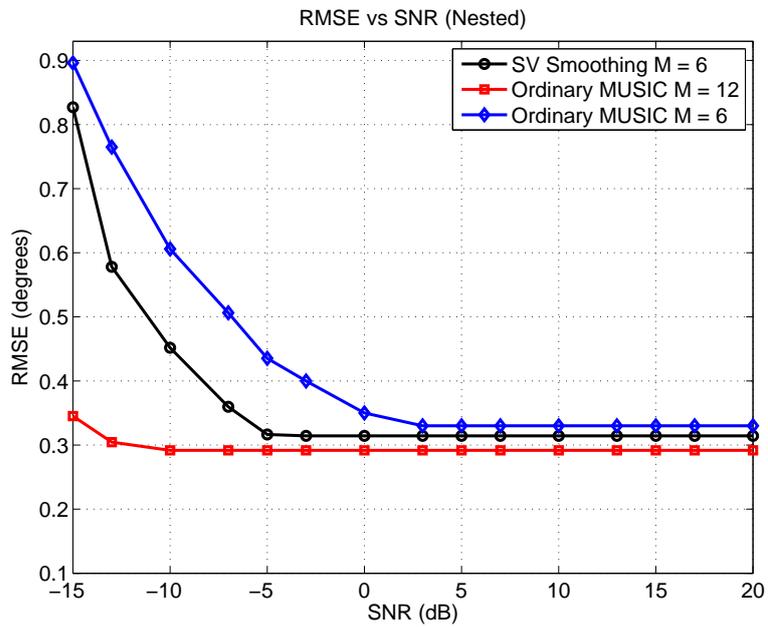
Figure 4.2b illustrates the probability of detection for the two-level nested AVS array. The probability of detection is calculated following Algorithm 1. Further, we only use the AVS-ULA of 12 elements for comparison since the AVS-ULA of 6 elements cannot identify 32 sources. Similarly Figure 4.4b illustrates the probability of detection versus SNR for the AVS-MRA and compare it to that of the AVS-ULA of 14 elements. Here, $N = 5000$.

4.6 Conclusions

In this chapter, we proposed a new technique to leverage the higher number of unique measurements that were available in the co-array. In the proposed spatial-velocity smoothing, subarrays were constructed by dividing the longer coarray. The smoothed covariance matrix was calculated by averaging the outputs of these subarrays. It was also seen that the smoothed matrix has the same form as that of an AVS-ULA. Hence, the traditional approaches to DOA estimation could be used. For the sparse AVS array, the maximum number of source DOAs that can be identified was upper bounded by $3(\alpha + 1)/2 - 1$ where α is the DOF of the sparse scalar sensor array. The simulation results show that with only 6 elements, the two level nested AVS array has a resolution and a detection performance that is very close to that of a longer AVS-ULA of 12 elements. Similarly, the AVS-MRA of 6 elements has a detection and a resolution performance similar to that of an AVS-ULA of 14 elements.

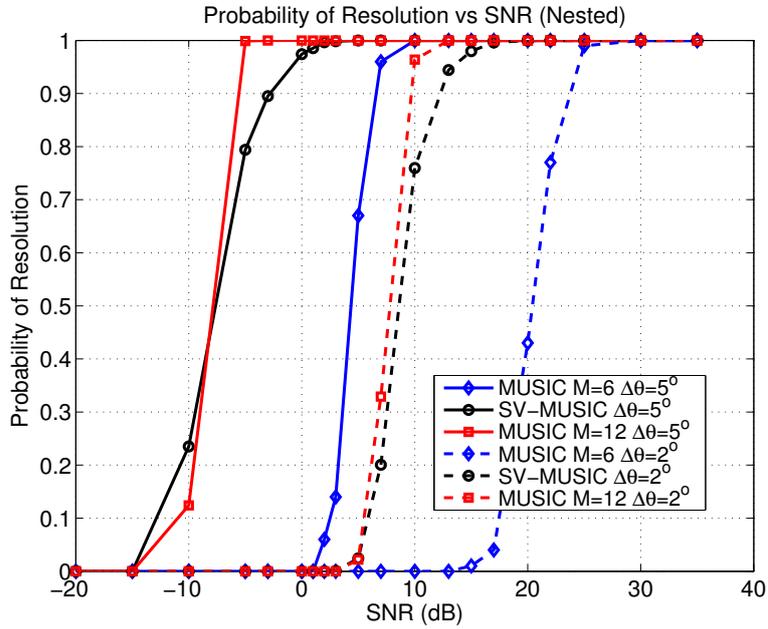


(a) SV Smoothing MUSIC for a two level Nested Array $M = 6$, $D = 32$.

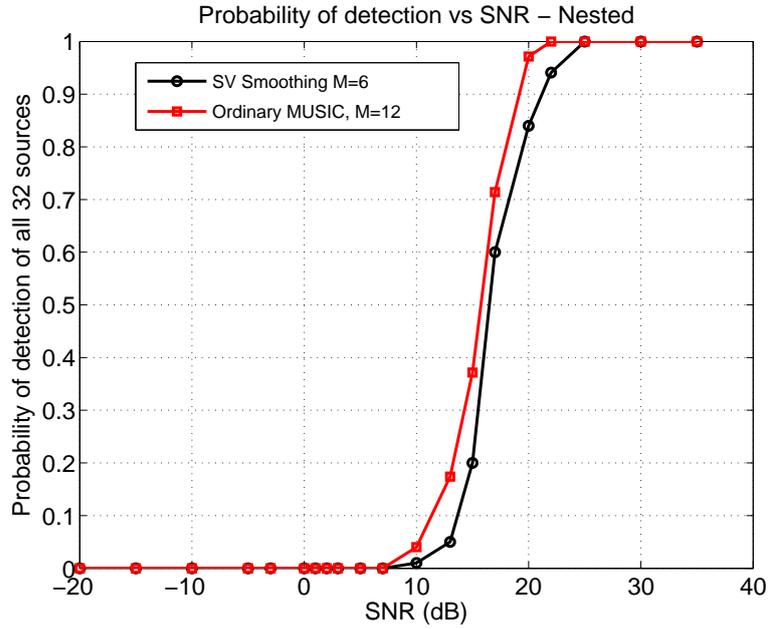


(b) RMSE (degrees) versus SNR (dB) - two level Nested Array, $\theta = 33.4^\circ$.

Figure 4.1: SV-Smoothing: MUSIC Spectrum and RMSE versus SNR, two-level nested array, $M = 6$.

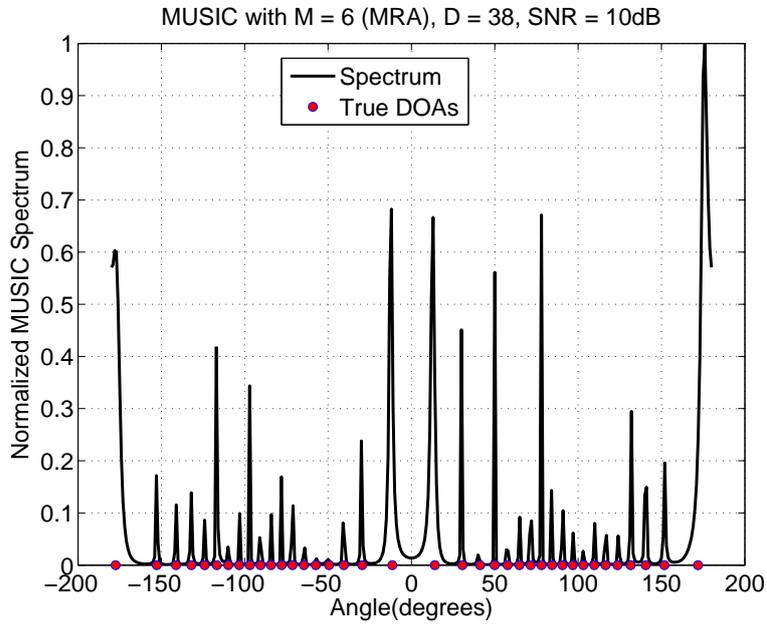


(a) SV-Smoothing: Probability of Resolution: Two level Nested Array.

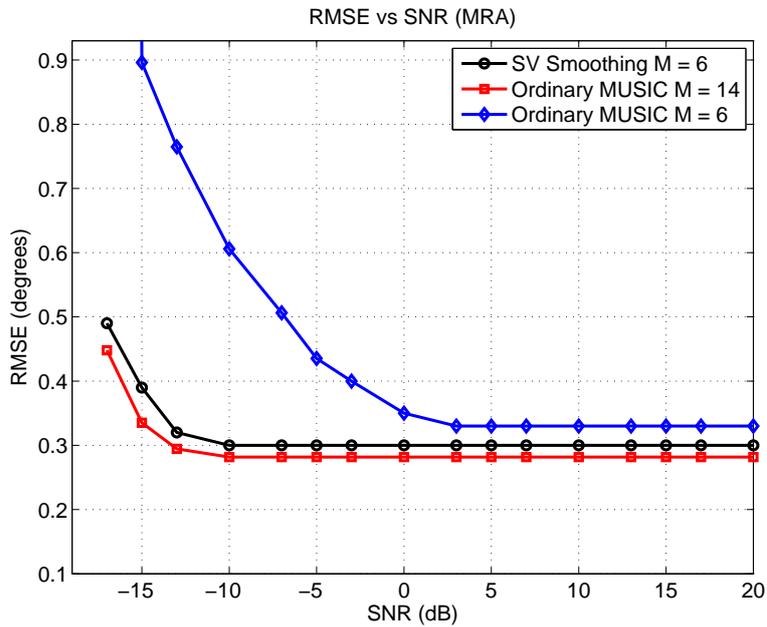


(b) SV-Smoothing: Probability of Detection of all 32 sources.

Figure 4.2: SV- Smoothing: Probability of resolution and probability of detection- two-level nested array.

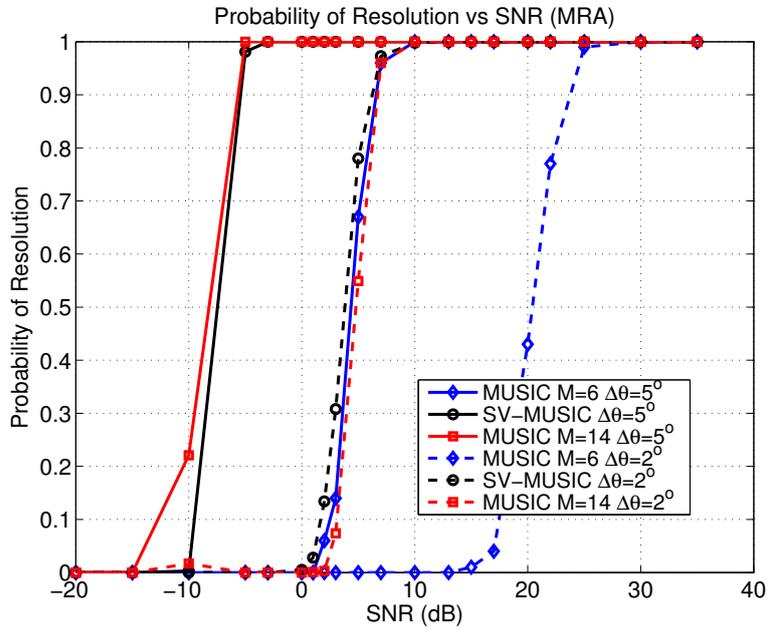


(a) SV Smoothing: MUSIC for an MRA $M = 6$, $D = 38$.

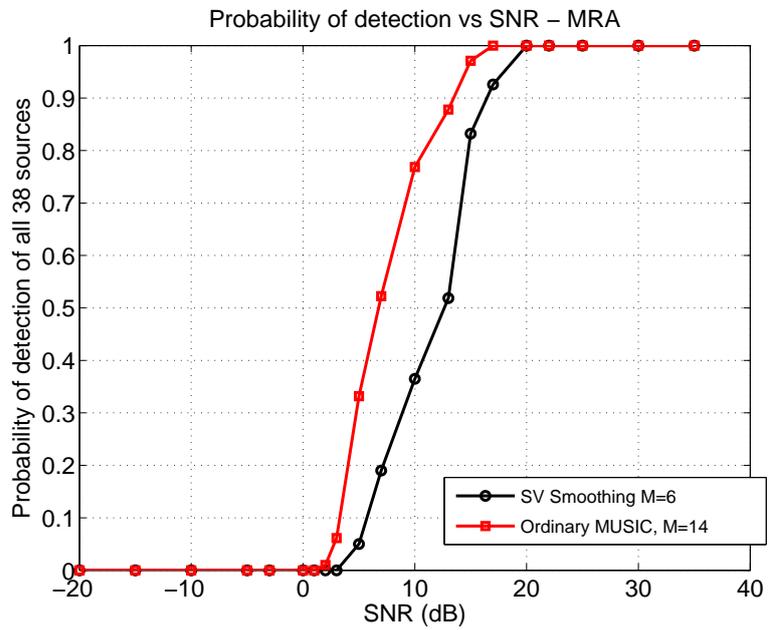


(b) SV Smoothing: RMSE (degrees) versus SNR (dB) for an MRA, $\theta = 33.4^\circ$.

Figure 4.3: SV Smoothing: MUSIC Spectrum and the RMSE versus SNR - MRA, $M = 6$.



(a) SV Smoothing: Probability of resolution, MRA $M = 6$.



(b) SV Smoothing: Probability of detection of all 38 sources, MRA $M = 6$.

Figure 4.4: SV Smoothing: Probability of resolution and probability of detection - MRA.

Part II

Sensor Selection for Non-linear Measurement Models

Greedy Sensor Selection for Non-Linear Models

5

Sensor networks have become popular in many applications, like localization, environmental monitoring, and remote-sensing, to list a few. Typically, a large number of sensors are involved in making observations related to a physical phenomenon, thereby making the sensing process expensive. Hence, it is of interest to select a subset of K sensors from M candidate sensors, where $K \ll M$. *Sensor selection* addresses the question of choosing the best subset of sensors so that the uncertainty in estimating the unknown parameter is minimized.

A straightforward method to solve the sensor selection problem is to perform a search over all possible $\binom{M}{K}$ combinations and select the subset that yields the best performance. Evidently, finding the optimal solution in this manner is computationally intensive and the computational cost involved is exponential. Thus, it is often suboptimally solved using two approaches depending on the cost function that specifies the uncertainty: convex optimization or greedy heuristics. Convex optimization techniques can be used to solve sensor selection if the cost function is a convex function of the selection variables. On the other hand, greedy heuristics aim at finding a computationally attractive solution. In order to ensure optimality of such greedy algorithms, submodular cost functions are required.

Typical functions that are used in convex optimization are related to the A-, D-, and E- optimal experiment designs [15]. In [16], non-linear measurement models are considered and scalar functions of the Cramér-Rao Bound (CRB) (as a generalization of the mean-squared error considered in [17]) are used as performance measures. Convex optimization approaches for sensor selection are computationally intensive. This is even more so for non-linear models due to the fact that the performance measure depends on the unknown parameter itself. Specifically, solving such convex optimization problems with interior point methods (using SeDuMi [18], for example) incurs cubic complexity. On the other hand, greedy algorithms have a linear complexity in the size of the problem, but require proxies of the mean-squared error (MSE) that are submodular. A number of such proxies are discussed in [19, 20]. A fundamental result by Nemhauser et al. [7] shows that a greedy algorithm which maximizes a normalized, monotone submodular function is near-optimal. In [21], it has been shown that the D-optimal criterion is submodular for linear measurement models. Another submodular function proposed in [20] uses a scalar property called frame potential (FP), which is related to the inner product of the rows of the measurement matrix. Hence, it can be used only for linear measurement models.

In this work, we propose two submodular costs for non-linear measurement models, which are independent of the unknown parameter, but use the knowledge of the domain where the parameter resides. The non-linear model is first linearized around points

obtained by gridding the unknown parameter space and this is used to build a weighted performance measure. In particular, we devise two cost functions that are submodular, namely, the *weighted FP* and *weighted log-det criterion*. Then, a greedy algorithm to optimize these costs is presented. The low complexity of the greedy algorithm allows us to densely grid the domain where the unknown parameter is assumed to be present.

5.1 Problem Statement

We consider a general non-linear model with additive Gaussian noise, i.e.,

$$y_m = g_m(\boldsymbol{\theta}) + n_m, \quad m = 1, 2, \dots, M, \quad (5.1)$$

where y_m denotes the m th sensor measurement, the function $g_m(\cdot)$ is non-linear, and n_m is the Gaussian noise process. Here, $\boldsymbol{\theta} \in \mathbb{R}^N$ is the unknown parameter. Let its estimate be denoted by $\hat{\boldsymbol{\theta}}$. The aim of sensor selection is to select a subset of K sensors out of M sensors so that a certain accuracy of $\hat{\boldsymbol{\theta}}$ is achieved. Mathematically, the optimization problem is formulated as

$$\min_{\mathcal{K}} F(\mathcal{K}), \quad \text{s. to } \mathcal{K} \subset \{1, 2, \dots, M\}, |\mathcal{K}| = K, \quad (5.2)$$

where $F(\mathcal{K})$ is the cost function characterized by the set of selected sensors \mathcal{K} and specifies the estimation quality. The number of selected sensors K , is assumed to be known.

The sensor selection problem in (5.2) is combinatorial. However, if $F(\mathcal{K})$ is a submodular function, (5.2) can be solved near-optimally with a lower complexity algorithm. More specifically, the solution of the linear-complexity greedy algorithm is within $(1 - 1/e)$ of the optimal solution, where e is Euler's number [7].

5.2 Performance Measures

Due to the non-linearity of the problem, any performance measure that specifies the estimation quality also depends on the unknown $\boldsymbol{\theta}$ [16]. Therefore, we grid the known parameter space to get $\mathcal{U} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_D\}$. In what follows, the non-linear function in (5.1) is first linearized around the obtained grid-points. This will enable us to extend the FP [20], [22], and log-det [17] cost functions to their weighted versions. These weighted cost functions are reasonable measures to optimize; they are related to the weighted MSE.

5.2.1 Taylor Approximation

If the first-order derivative of $g_m(\boldsymbol{\theta})$ exists $\forall \boldsymbol{\theta} \in \mathcal{U}$, it can be linearized around every $\boldsymbol{\theta}_d \in \mathcal{U}$ using a first-order Taylor approximation of (5.1) to get

$$y_m \approx g_m(\boldsymbol{\theta}_d) + \left. \frac{\partial g_m(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_d} (\boldsymbol{\theta} - \boldsymbol{\theta}_d) + n_m, \quad m = 1, 2, \dots, M,$$

where $n_m \sim \mathcal{N}(0, \sigma^2)$. Stacking y_m , n_m , and $g_m(\boldsymbol{\theta}_d)$ for $m = 1, 2, \dots, M$, in vectors \mathbf{y} , \mathbf{n} , and \mathbf{g} , respectively, we get

$$\mathbf{y} = \mathbf{g} + \mathbf{H}_d(\boldsymbol{\theta} - \boldsymbol{\theta}_d) + \mathbf{n}, \quad (5.3)$$

where $\mathbf{H}_d \in \mathbb{R}^{M \times N}$ with $\partial g_m(\boldsymbol{\theta}_d)/\partial \boldsymbol{\theta}_d^T$ being the m th row of \mathbf{H}_d .

Assuming the noise in (5.3) is i.i.d. Gaussian and that the unknown $\boldsymbol{\theta}$ corresponds to $\boldsymbol{\theta}_d$, the MSE (equal to the Cramér-Rao bound) of the least-squares estimate computed using the sensor subset \mathcal{K} , is given by

$$\text{MSE}(\mathcal{K}) = \mathbb{E} \left(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_d\|_2^2 \right) = \sigma^2 \text{Tr} \left(\sum_{i \in \mathcal{K}} \mathbf{h}_{i,d} \mathbf{h}_{i,d}^T \right)^{-1} \quad (5.4)$$

where $\mathbf{h}_{i,d}^T$ is the i th row of the matrix \mathbf{H}_d , and $\text{Tr}(\cdot)$ denotes the trace of its argument. Thus, the MSE of $\hat{\boldsymbol{\theta}}$ depends on the spread of the eigenvalues of $\mathbf{H}_d^T \mathbf{H}_d$.

5.2.2 Cost Function

For a set of D matrices $\{\mathbf{H}_d\}$, we may define the weighted MSE, denoted by $\text{MSE}_W(\cdot)$, over all the measurement matrices as follows:

$$\text{MSE}_W(\mathcal{K}) = \sigma^2 \sum_{d=1}^D \pi_d \text{Tr} \left(\sum_{i \in \mathcal{K}} \mathbf{h}_{i,d} \mathbf{h}_{i,d}^T \right)^{-1}, \quad (5.5)$$

where π_d is a known positive weight (for example, it could indicate the probability that the true $\boldsymbol{\theta}$ lies on the grid point $\boldsymbol{\theta}_d$, i.e., $\Pr(\boldsymbol{\theta} = \boldsymbol{\theta}_d) = \pi_d$ with $\pi_d > 0$). In practice, the true parameter might not lie on the assumed grid and might worsen the performance, but this is due to the fact that we have a non-linear model. We now consider two functions, namely, the weighted frame potential and weighted log-det as surrogates for the MSE.

Weighted FP

The frame potential of a matrix \mathbf{H}_d is defined as [20]

$$\text{FP}(\mathcal{K}) = \sum_{i,j \in \mathcal{K}} |\langle \mathbf{h}_{i,d}, \mathbf{h}_{j,d} \rangle|^2. \quad (5.6)$$

Thus, minimizing the frame potential leads to finding the rows that are the most orthogonal to each other. Since we have linearized the function g_m around the points in the domain \mathcal{U} , a joint optimization of the FP for all the measurement matrices is considered by choosing a weighted FP, denoted by $\text{FP}_W(\cdot)$, as follows

$$\text{FP}_W(\mathcal{K}) := \sum_{d=1}^D \pi_d \sum_{i,j \in \mathcal{K}} |\langle \mathbf{h}_{i,d}, \mathbf{h}_{j,d} \rangle|^2. \quad (5.7)$$

However, (5.7) is not submodular. In order to develop a greedy algorithm and use the results of [7], we require the objective to be normalized, monotonically increasing, and submodular. Therefore, we adapt (5.7) under change of variable $\mathcal{K} = \mathcal{M} \setminus \mathcal{S}$ to obtain

$$F(\mathcal{S}) := \text{FP}_W(\mathcal{M} \setminus \mathcal{S}) - \text{FP}_W(\mathcal{M}), \quad (5.8)$$

where $\mathcal{S} = \mathcal{M} \setminus \mathcal{K}$, \mathcal{M} is the available set of sensors, and $\text{FP}_W(\mathcal{M})$ is used to ensure that $F(\cdot)$ is zero for an empty input set. It can be verified that the function $-F(\cdot)$ is normalized and monotonically increasing, and the proof is shown in Appendix 5.A. Since $-F(\cdot)$ is a positive linear combination of the submodular cost function of [20], it is also submodular and can be greedily maximized.

Weighted log-det

Another popular proxy for the MSE is the D-optimality criterion, i.e., the log-volume of the confidence ellipsoid for a measurement matrix \mathbf{H}_d , given by

$$\text{LD}(\mathcal{K}) = \log \det \left(\sum_{i \in \mathcal{K}} \mathbf{h}_{i,d} \mathbf{h}_{i,d}^T \right)^{-1}.$$

Similar to the weighted FP, a weighted log-det criterion can be formed as

$$\text{LD}_W(\mathcal{K}) := \sum_{d=1}^D \pi_d \log \det \left(\sum_{i \in \mathcal{K}} \mathbf{h}_{i,d} \mathbf{h}_{i,d}^T \right)^{-1} \quad (5.9)$$

However, the above function is not defined for an empty set. Hence, we alter the above cost by adding a regularizer to obtain:

$$F(\mathcal{K}) = \sum_{d=1}^D \pi_d \left(\log \det \left(\sum_{i \in \mathcal{K}} \mathbf{h}_{i,d} \mathbf{h}_{i,d}^T + \epsilon \mathbf{I}_N \right)^{-1} + N \log \epsilon \right), \quad (5.10)$$

where $\epsilon > 0$ is a small positive constant, \mathbf{I}_N is the identity matrix of size N , and $N \log \epsilon$ ensures that the function $F(\cdot)$ is zero when the input is an empty set. The negative of the summand in the above cost is normalized, monotone, and submodular [21]. Consequently, $-F(\cdot)$ is also submodular as it is a positive linear combination of submodular functions. The proof is shown in Appendix 5.B.

5.3 Proposed Algorithm

The algorithm aims at finding a set of K sensors in $\{1, 2, \dots, M\}$, assuming that the matrices $\{\mathbf{H}_d, d = 1, 2, \dots, D\}$ are known, and is summarized in Algorithm 2. At each iteration, the row that maximizes the function $G(\cdot) = -F(\cdot)$, where $F(\cdot)$ is given by

either (5.8) or (5.10), is selected and assigned to a test set. Since (5.8) gives us the complement set of \mathcal{K} , the complement of the test set is taken to get \mathcal{K} .

Since the functions in (5.8) and (5.10) are normalized, monotone submodular functions, we can use a classic result of [7] to bound the performance of the greedy algorithm with respect to the solution obtained from an exhaustive search. More specifically, a result of [7] for submodular functions gives us the following:

$$F(\mathcal{K}) \leq \left(1 - \frac{1}{e}\right) F(\text{OPT}), \quad (5.11)$$

where $F(\text{OPT})$ is the optimal value of the function given by $F(\text{OPT}) = \min_{\mathcal{A} \subset \mathcal{M}, |\mathcal{A}|=K} F(\mathcal{A})$, and OPT is the optimal set of elements. The result of (5.11) shows that the solution of the greedy algorithm is always close to the optimal one with respect to the chosen cost function.

Algorithm 2 Greedy Algorithm

1: **Input:**

D matrices $\{\mathbf{H}_d\}$, number of sensors K , function $G(\cdot) = -F(\cdot)$, with $F(\cdot)$ from (5.8) or (5.10).

2: **Initialize:**

Test sensor set $\mathcal{T} = \emptyset$.

3: **Repeat:**

1. Find the row $r^* = \arg \max_{r \notin \mathcal{T}} G(\mathcal{T} \cup r)$.
 2. $\mathcal{T} \leftarrow \mathcal{T} \cup r^*$.
 3. For weighted FP cost:
 - (a) If $|\mathcal{T}| = M - K$, stop.
 - (b) Assign set of selected sensors $\mathcal{K} = \mathcal{M} \setminus \mathcal{T}$.
 4. For weighted log-det cost:
 - (a) If $|\mathcal{T}| = K$, stop.
 - (b) Assign set of selected sensors $\mathcal{K} = \mathcal{T}$.
-

5.4 Greedy vs. Convex Optimization

In comparison to greedy algorithms, state-of-the-art methods based on convex optimization for non-linear models can ensure that the performance constraints are met for every $\boldsymbol{\theta}_d \in \mathcal{U}$ [16]. Thus, they generally have a better performance. Constraining the performance for every $\boldsymbol{\theta}_d \in \mathcal{U}$ using the greedy algorithm is, however, not straightforward. Therefore, we chose comparatively weaker measures, viz., the weighted FP and weighted log-det, which are simply the cost functions averaged over the related points $\{\boldsymbol{\theta}_d\}$.

For convex optimization based sensor selection, the computational complexity depends on the convex solver used. For example, the complexity of the interior-point solvers

is typically cubic in M . To alleviate the computational complexity due to gridding, a weighted measure similar to the measures considered in this paper can be used, or the performance due to the worst case $\boldsymbol{\theta}_d \in \mathcal{U}$ can be constrained [16]. On the other hand, a greedy algorithm like the one in Algorithm 1 to optimize a weighted measure has a much lower complexity. That is, with convex optimization based sensor selection, worst-case, and average performance measures can be considered, whereas the greedy algorithm proposed here can handle only a weighted measure.

We evaluate the complexity of Algorithm 1. Consider the weighted log-det cost. The complexity of evaluating the determinant of an $N \times N$ matrix is $O(N^3)$ and there are $O(M)$ such matrices. Then, the determinant is summed for D grid points. Since Algorithm 1 has K steps, the total complexity of the greedy algorithm for the weighted log-det performance measure is $O(MDKN^3)$. Similarly, for the weighted FP, the inner product of an $N \times 1$ vector with another vector of the same size requires $O(N)$ operations. Over D points and $M - K$ steps of the greedy algorithm, the total complexity of the algorithm is $O(M^2NKD)$. On the other hand, convex solvers such as SeDuMi [18] have a complexity which is cubic in M [16].

5.5 Numerical Example

An example of sensor selection for source localization is considered. The performance of Algorithm 1 in terms of the root mean-squared error (RMSE) is compared with the solutions provided by convex optimization, using D- and E-optimality related constraints given in [16]. We assume that each sensor measures the distance to the target with coordinates $\boldsymbol{\theta}$ corrupted with additive Gaussian noise so that the model in (5.1) can be written as

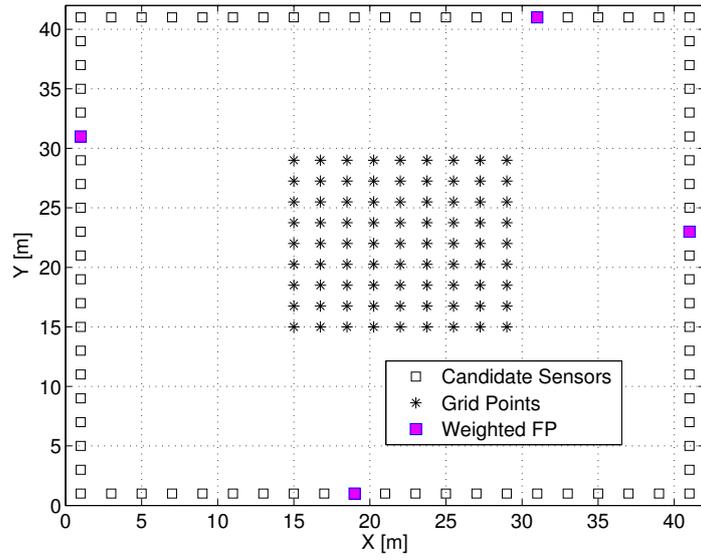
$$y_m = \overbrace{\|\boldsymbol{\theta} - \mathbf{a}_m\|_2}^{g_m(\boldsymbol{\theta})} + n_m, \quad m = 1, 2, \dots, M \quad (5.12)$$

where \mathbf{a}_m is the location of the m th sensor and $n_m \sim \mathcal{N}(0, \sigma^2)$.

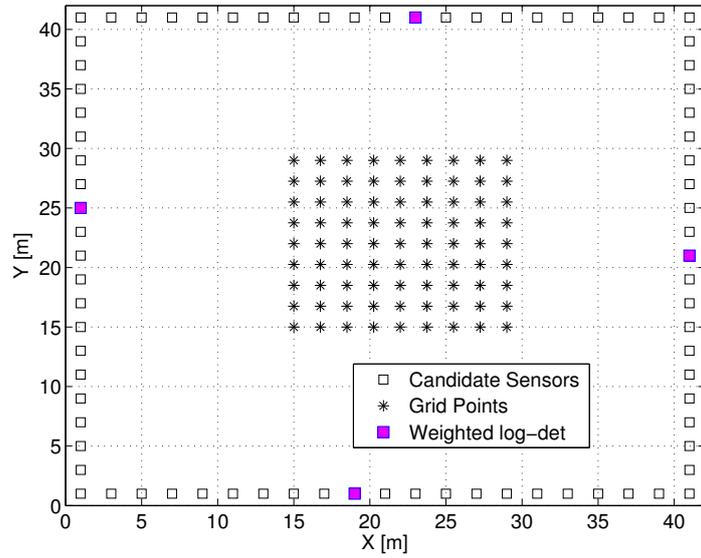
The domain \mathcal{U} where the target is expected (i.e., the surveillance area) is gridded to obtain $D = 81$ points and we use $\sigma^2 = 1$. The measurement setup with $M = 80$ sensors is shown in Fig. 5.1. Further, we consider equal combining weights, i.e., $\pi_d = 1$ for $d = 1, 2, \dots, D$.

Fig. 5.1a shows the sensors selected by the greedy algorithm for the weighted FP cost and Fig. 5.1b shows the sensors selected by the greedy algorithm for the weighted log-det cost for $K = 4$. It can be noticed that the sensor indices selected by the algorithm for the two costs are approximately the same. The RMSE of the estimated vector $\hat{\boldsymbol{\theta}}$ is plotted in Fig. 5.2b. To evaluate the RMSE, the target position at each iteration was generated so that each component of $\boldsymbol{\theta}$ was drawn from a uniform distribution in [15, 29], and the squared error was averaged over 1000 Monte-Carlo simulations.

It is seen that the greedy algorithm performs nearly as well as the convex optimization based approaches. Although the true parameter in this example might not lie exactly on the grid, the algorithm provides satisfactory results for both cost functions.



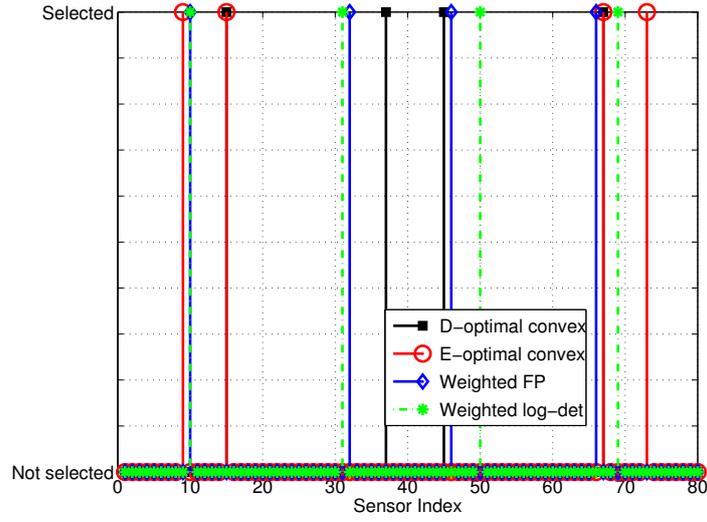
(a)



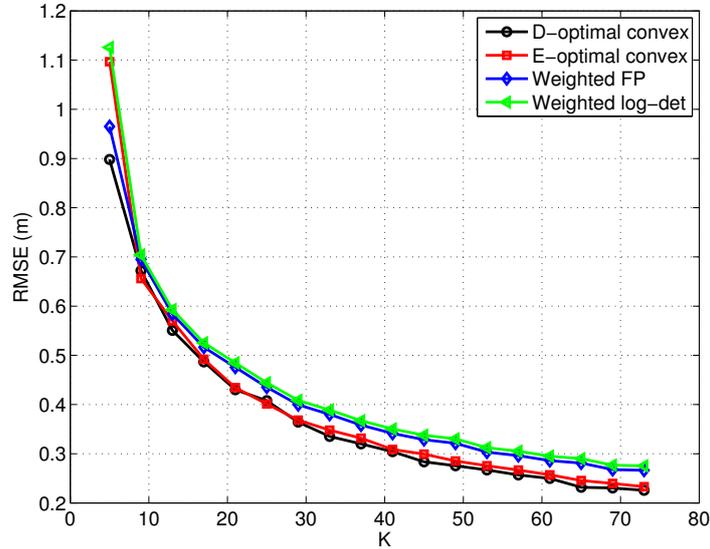
(b)

Figure 5.1: Sensor placement for localization. (a.) Placement based on the weighted FP cost. (b.) Placement based on the weighted log-det cost.

To illustrate the sensor placement with different costs, we also plot the selected sensors in Fig. 5.2a for $K = 4$. The threshold for the E-optimality related constraint in [16] is selected by fixing $R_e = 2.82\text{cm}$, and $P_e = 0.95$ where R_e and P_e specify a certain accuracy constraint. Similarly, for the D-optimality related constraint in [16], the threshold is selected by fixing $\bar{R}_e = 7.74\text{cm}$ and $P_e = 0.95$.



(a)



(b)

Figure 5.2: (a.) Sensor selection for the different costs for $K = 4$. (b.) Computational time versus K from an available set of $M = 80$ sensors. (c.) RMSE versus the number of selected sensors.

5.6 Conclusions

Sensor selection for a non-linear measurement model in additive noise was considered. The performance measure for non-linear models generally depends on the unknown parameter. In the proposed approach, the non-linear model was linearized around a set of points in the domain where the true parameter is assumed to reside. Two submodular costs, namely, the weighted frame potential and weighted log-det function were formulated. To optimize these costs, a near-optimal greedy algorithm was presented.

The proposed linear-complexity greedy algorithm is computationally attractive as compared to the state-of-the-art sensor selection solvers for non-linear models. Since the proposed cost functions are normalized, monotone and submodular, the solution of the greedy algorithm is always within $(1 - 1/e)$ of the optimal one.

Appendix

5.A Submodularity of the cost function - weighted FP

The set function G_{FP} given by

$$G_{\text{FP}}(\mathcal{S}) = \text{FP}_{\text{W}}(\mathcal{M}) - \text{FP}_{\text{W}}(\mathcal{M} \setminus \mathcal{S}),$$

is a normalized, monotone submodular function.

Proof. Before submodularity, we first show that $G_{\text{FP}}(\cdot)$ is normalized and non-decreasing. It can be seen easily that it is normalized since $G_{\text{FP}}(\emptyset) = 0$. To show that it is monotonic, we show that the increment of $G_{\text{FP}}(\cdot)$ with regard to set \mathcal{A} due to adding index $i \notin \mathcal{A}$ is positive. Assigning $\mathcal{A} \cup i = \bar{A}$,

$$\begin{aligned} & G_{\text{FP}}(\mathcal{A} \cup i) - G_{\text{FP}}(\mathcal{A}) \\ &= \sum_{d=1}^D \pi_d \text{FP}_{\text{W}}(\mathcal{M} \setminus \mathcal{A}) - \sum_{d=1}^D \pi_d \text{FP}_{\text{W}}(\mathcal{M} \setminus (\mathcal{A} \cup i)) \\ &= \sum_{d=1}^D \pi_d \sum_{n,m \in \bar{A} \cup i} |\langle \mathbf{h}_{n,d}, \mathbf{h}_{m,d} \rangle|^2 - \sum_{d=1}^D \pi_d \sum_{n,m \in \bar{A}} |\langle \mathbf{h}_{n,d}, \mathbf{h}_{m,d} \rangle|^2 \\ &= \sum_{d=1}^D \pi_d \left(\sum_{n,m \in \bar{A}} |\langle \mathbf{h}_{n,d}, \mathbf{h}_{m,d} \rangle|^2 + 2 \sum_{n \in \bar{A}} |\langle \mathbf{h}_{n,d}, \mathbf{h}_{i,d} \rangle|^2 + |\langle \mathbf{h}_{i,d}, \mathbf{h}_{i,d} \rangle|^2 \right) \\ &\quad - \sum_{d=1}^D \pi_d \sum_{n,m \in \bar{A}} |\langle \mathbf{h}_{n,d}, \mathbf{h}_{m,d} \rangle|^2 \\ &= 2 \sum_{d=1}^D \pi_d \sum_{n \in \bar{A}} |\langle \mathbf{h}_{n,d}, \mathbf{h}_{i,d} \rangle|^2 + |\langle \mathbf{h}_{i,d}, \mathbf{h}_{i,d} \rangle|^2 \\ &\geq 0 \end{aligned}$$

To show submodularity, we consider the difference in gain between adding an index to set \mathcal{A} and to a set \mathcal{B} . Without loss of generality, we assume $\mathcal{B} = \mathcal{A} \cup j$.

$$\begin{aligned}
& G_{\text{FP}}(\mathcal{A} \cup i) - G_{\text{FP}}(\mathcal{A}) - G_{\text{FP}}(\mathcal{B} \cup i) + G_{\text{FP}}(\mathcal{B}) \\
&= \sum_{d=1}^D \pi_d \{ \text{FP}_{\text{W}}(\mathcal{M} \setminus (\mathcal{A} \cup \{i, j\})) + \text{FP}_{\text{W}}(\mathcal{M} \setminus \mathcal{A}) \\
&\quad - \text{FP}_{\text{W}}(\mathcal{M} \setminus (\mathcal{A} \cup j)) - \text{FP}_{\text{W}}(\mathcal{M} \setminus (\mathcal{A} \cup i)) \} \\
&= \sum_{d=1}^D \pi_d \left(\sum_{n, m \in \mathcal{A} \cup \{i, j\}} |\langle \mathbf{h}_{n,d}, \mathbf{h}_{m,d} \rangle|^2 + \sum_{n, m \in \mathcal{A}} |\langle \mathbf{h}_{n,d}, \mathbf{h}_{m,d} \rangle|^2 \right. \\
&\quad \left. - \sum_{n, m \in \mathcal{A} \cup j} |\langle \mathbf{h}_{n,d}, \mathbf{h}_{m,d} \rangle|^2 - \sum_{n, m \in \mathcal{A} \cup i} |\langle \mathbf{h}_{n,d}, \mathbf{h}_{m,d} \rangle|^2 \right) \\
&= 2 |\langle \mathbf{h}_{i,d}, \mathbf{h}_{j,d} \rangle|^2 \\
&\geq 0
\end{aligned}$$

■

5.B Submodularity of the cost function - weighted log-det

The set function $G_{\text{WD}}(\cdot)$ given by

$$G_{\text{WD}}(\mathcal{K}) = \sum_{d=1}^D \pi_d \left(\log \det \left(\sum_{i \in \mathcal{K}} \mathbf{h}_{i,d} \mathbf{h}_{i,d}^T + \epsilon \mathbf{I}_N \right) - N \log \epsilon \right).$$

is a normalized, monotone submodular function.

Proof. Before submodularity, we first show that $G_{\text{WD}}(\cdot)$ is normalized and non-decreasing. It can be seen easily that it is normalized since $G_{\text{WD}}(\emptyset) = 0$. To show that it is monotonic, we show that the increment of $G_{\text{WD}}(\cdot)$ with regard to set \mathcal{X} due to adding index $i \notin \mathcal{A}$ is positive.

$$\begin{aligned}
& G_{\text{WD}}(\mathcal{A} \cup i) - G_{\text{WD}}(\mathcal{A}) \\
&= \sum_{d=1}^D \pi_d \log \det \left(\sum_{n \in \mathcal{A} \cup i} \mathbf{h}_{n,d} \mathbf{h}_{n,d}^T + \epsilon \mathbf{I}_N \right) - \sum_{d=1}^D \pi_d \log \det \left(\sum_{n \in \mathcal{A}} \mathbf{h}_{n,d} \mathbf{h}_{n,d}^T + \epsilon \mathbf{I}_N \right) \\
&= \sum_{d=1}^D \pi_d \left(\log \det \{ (\mathbf{H}_d^{\mathcal{A}})^T \mathbf{H}_d^{\mathcal{A}} + (\mathbf{H}_d^i)^T \mathbf{H}_d^i + \epsilon \mathbf{I}_N \} - \log \det \{ (\mathbf{H}_d^{\mathcal{A}})^T \mathbf{H}_d^{\mathcal{A}} + \epsilon \mathbf{I}_N \} \right).
\end{aligned}$$

In the above equations, we have used $\mathbf{H}_d^{\mathcal{A}}$ to denote the matrix selected by the sensor subset \mathcal{A} , i.e., $\mathbf{H}_d^{\mathcal{A}} = \sum_{n \in \mathcal{A}} \mathbf{h}_{n,d} \mathbf{h}_{n,d}^T$. Let $\{ (\mathbf{H}_d^{\mathcal{A}})^T \mathbf{H}_d^{\mathcal{A}} + \epsilon \mathbf{I}_N \} = \tilde{\mathbf{H}}_d^{\mathcal{A}}$. Since $\tilde{\mathbf{H}}_d^{\mathcal{A}}$ is positive semidefinite and non-singular,

$$G_{\text{WD}}(\mathcal{A} \cup i) - G_{\text{WD}}(\mathcal{A}) = \sum_{d=1}^D \pi_d \log \det \{ \mathbf{I}_N + (\tilde{\mathbf{H}}_d^{\mathcal{A}})^{-1} (\mathbf{H}_d^i)^T \mathbf{H}_d^i \} \geq 0$$

where the last inequality follows due to the fact that $(\mathbf{H}_d^i)^T \mathbf{H}_d^i$ is positive semidefinite. To show submodularity, we consider the difference in gain between adding an index to set \mathcal{A} and to a set \mathcal{B} . Without loss of generality, we assume $\mathcal{B} = \mathcal{A} \cup j$ and $\mathcal{A} \subset \mathcal{B}$.

$$\begin{aligned} & G_{\text{WD}}(\mathcal{A} \cup i) - G_{\text{WD}}(\mathcal{A}) - G_{\text{WD}}(\mathcal{B} \cup i) + G_{\text{WD}}(\mathcal{B}) \\ &= \sum_{d=1}^D \pi_d \left[\log \det \left(\sum_{n \in \mathcal{A} \cup i} \mathbf{h}_{n,d} \mathbf{h}_{n,d}^T + \epsilon \mathbf{I}_N \right) - \log \det \left(\sum_{n \in \mathcal{A}} \mathbf{h}_{n,d} \mathbf{h}_{n,d}^T + \epsilon \mathbf{I}_N \right) \right] \\ &\quad - \sum_{d=1}^D \pi_d \left[\log \det \left(\sum_{n \in \mathcal{A} \cup \{i,j\}} \mathbf{h}_{n,d} \mathbf{h}_{n,d}^T + \epsilon \mathbf{I}_N \right) - \log \det \left(\sum_{n \in \mathcal{A} \cup j} \mathbf{h}_{n,d} \mathbf{h}_{n,d}^T + \epsilon \mathbf{I}_N \right) \right] \\ &= \sum_{d=1}^D \pi_d \left(\log \det \{ (\mathbf{H}_d^{\mathcal{A}})^T \mathbf{H}_d^{\mathcal{A}} + (\mathbf{H}_d^i)^T \mathbf{H}_d^i + \epsilon \mathbf{I}_N \} - \log \det \{ (\mathbf{H}_d^{\mathcal{A}})^T \mathbf{H}_d^{\mathcal{A}} + \epsilon \mathbf{I}_N \} \right) \\ &\quad - \sum_{d=1}^D \pi_d \left(\log \det \{ (\mathbf{H}_d^{\mathcal{A} \cup j})^T \mathbf{H}_d^{\mathcal{A} \cup j} + (\mathbf{H}_d^i)^T \mathbf{H}_d^i + \epsilon \mathbf{I}_N \} - \log \det \{ (\mathbf{H}_d^{\mathcal{A} \cup j})^T \mathbf{H}_d^{\mathcal{A} \cup j} + \epsilon \mathbf{I}_N \} \right). \end{aligned}$$

As before, we assign $\{ (\mathbf{H}_d^{\mathcal{A}})^T \mathbf{H}_d^{\mathcal{A}} + \epsilon \mathbf{I}_N \} = \tilde{\mathbf{H}}_d^{\mathcal{A}}$ and $\{ (\mathbf{H}_d^{\mathcal{A} \cup j})^T \mathbf{H}_d^{\mathcal{A} \cup j} + \epsilon \mathbf{I}_N \} = \tilde{\mathbf{H}}_d^{\mathcal{A} \cup j}$. Substituting in the above equation, we get

$$\begin{aligned} & G_{\text{WD}}(\mathcal{A} \cup i) - G_{\text{WD}}(\mathcal{A}) - G_{\text{WD}}(\mathcal{B} \cup i) + G_{\text{WD}}(\mathcal{B}) \\ &= \sum_{d=1}^D \pi_d \log \det \{ \mathbf{I}_N + (\tilde{\mathbf{H}}_d^{\mathcal{A}})^{-1} (\mathbf{H}_d^i)^T \mathbf{H}_d^i \} - \sum_{d=1}^D \pi_d \log \det \{ \mathbf{I}_N + (\tilde{\mathbf{H}}_d^{\mathcal{A} \cup j})^{-1} (\mathbf{H}_d^i)^T \mathbf{H}_d^i \}. \end{aligned}$$

We notice that $\tilde{\mathbf{H}}_d^{\mathcal{A} \cup j} = \tilde{\mathbf{H}}_d^{\mathcal{A}} + (\mathbf{H}_d^j)^T \mathbf{H}_d^j$. We use Sherman-Morrison's identity to relate $(\tilde{\mathbf{H}}_d^{\mathcal{A} \cup j})^{-1}$ and $(\tilde{\mathbf{H}}_d^{\mathcal{A}})^{-1}$ [23]. That is,

$$(\tilde{\mathbf{H}}_d^{\mathcal{A} \cup j})^{-1} = (\tilde{\mathbf{H}}_d^{\mathcal{A}})^{-1} - \frac{(\tilde{\mathbf{H}}_d^{\mathcal{A}})^{-1} (\mathbf{H}_d^j)^T \mathbf{H}_d^j (\tilde{\mathbf{H}}_d^{\mathcal{A}})^{-1}}{1 + \mathbf{H}_d^j (\tilde{\mathbf{H}}_d^{\mathcal{A}})^{-1} (\mathbf{H}_d^j)^T}. \quad (5.13)$$

It can be verified that the term on the right of (5.13) is positive semidefinite. Then, using Corollary 2 of [24], we get

$$\det(\mathbf{I}_N + (\tilde{\mathbf{H}}_d^{\mathcal{A}})^{-1} (\mathbf{H}_d^i)^T \mathbf{H}_d^i) \geq \det(\mathbf{I}_N + (\tilde{\mathbf{H}}_d^{\mathcal{A} \cup j})^{-1} (\mathbf{H}_d^i)^T \mathbf{H}_d^i). \quad (5.14)$$

Therefore,

$$G_{\text{WD}}(\mathcal{A} \cup i) - G_{\text{WD}}(\mathcal{A}) - G_{\text{WD}}(\mathcal{B} \cup i) + G_{\text{WD}}(\mathcal{B}) \geq 0. \quad \blacksquare$$

6.1 Conclusion

Traditionally, the problem of direction of arrival (DOA) estimation has been studied using scalar sensors. The increasing need to measure more information has coincided with the development of acoustic vector sensors, which measure acoustic particle velocity in addition to the acoustic pressure. Additionally, a surge in the work on sparse scalar sensor arrays has resulted in the estimation of a much higher number of source DOAs than that possible with uniform linear arrays. This problem is also a subject of *covariance sensing*. Since a sparse linear array has a higher array aperture than that of a ULA with the same number of sensor elements, the sparse array also has a better resolution performance. In applications involving the AVSs, a sparse AVS array would be highly beneficial since it can be expected to localize even more source DOAs compared to that possible with an AVS-ULA of the same number of elements. Moreover, critical military applications require a much higher performance in parameter estimation accuracy. This is the classical sensor selection problem where one can specify a desired performance criterion to select a subset of the candidate sensor set. Although we do not consider an example of sensor selection for the AVS data model, the proposed framework can be extended to applications of source localization and DOA estimation using AVSs.

In this thesis, we addressed the following two questions:

- Covariance sensing using sparse AVS arrays - The upper bound on the number of source DOAs that could be identified using sparse AVS arrays was established. We considered two kinds of sources, namely, quasi-stationary and stationary sources.
- Greedy sensor placement for non-linear measurement models- Two submodular cost functions were formulated and a greedy algorithm to optimize these functions was presented.

In Chapter 3, we investigated DOA estimation of quasi-stationary sources. The time-variant nature of the source covariance matrix for quasi-stationary sources was leveraged in order to construct a covariance matrix on which subspace based techniques could be used. Further, the available degrees of freedom of the AVS- coarray was studied and the number of sources that could be identified was upper bounded using the linear independence of steering vectors of the coarray. Using the KR subspace approach, the maximum number of DOAs that can be identified is upper-bounded by $5\alpha - 1$, where α is the degrees of freedom of the sparse scalar sensor array. Furthermore, the resolution and detection performance of the sparse AVS arrays were very close to that of the longer

AVS-ULAs with an equivalent number of measurements. For instance, the AVS-MRA of a mere 6 elements has a performance which is very similar to that of a much longer AVS-ULA of 45 elements.

In Chapter 4, an alternative method for DOA estimation of stationary sources was proposed. This method was based on a spatial-velocity based smoothing to construct a rank- D covariance matrix, where D is the number of sources. The basic principle was to construct subarrays from the longer coarray and average the different resulting covariance matrices of these subarrays. However, the number of sources that could be identified was found to be lesser than that possible with the quasi-stationary assumption of Chapter 3. Using MUSIC on the coarray, the maximum number of source DOAs that could be uniquely identified was shown theoretically to be $3(\alpha + 1)/2 - 1$. However, the approach could handle the time-invariant nature of covariance matrices of stationary sources. Hence, the quasi-stationary assumption is not a pre-requisite for this approach. Similar to the results of Chapter 3, the error, resolution and detection performance using SV smoothing on the coarray of the sparse AVS array was very close to that of a longer AVS-ULA.

The second part of the thesis considered the problem of sensor selection. Recall that this problem was interesting because, in several applications, it is important to guarantee a certain estimation accuracy on the parameter of interest. Using the concept of submodularity, we proposed two cost functions- one based on the frame potential, called the weighted frame potential, and the other based on the log – det function, or the weighted D-optimal criterion. In order to formulate these costs, the non-linear function was first linearized using a first order Taylor approximation. Further, a greedy algorithm to maximize the two submodular costs was proposed in Chapter 5. Using Theorem 4, it was shown that the output of the greedy algorithm is $(1 - 1/e)$ within the optimal solution. The proposed approach was applied to an example of source localization and the performance of the greedy algorithm was compared with convex optimization based approaches. The results show that the error performance of the submodular costs comes close to that convex optimization based approaches while consuming substantially lower computational time. Moreover, convex optimization based costs have complexity which is cubic in the size of the problem but the greedy algorithm has only linear complexity. In the next section, we discuss the possible directions for future research.

6.2 Suggestions for Future Research

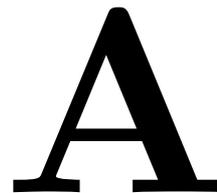
The following could be topics of future research:

- **Wideband source localization:** While our development for DOA estimation is primarily based on the narrowband array processing model, the proposed frameworks for quasi-stationary and stationary sources can be extended to the wideband scenario using the frequency domain approaches [25] [26]. The main idea of frequency domain processing is to divide the entire wide frequency band into several narrow frequency bands on which the spatial-velocity smoothing or the

KR-subspace method could be used. The subspaces at various frequencies can be combined using the incoherent signal subspace method (ISSM) of [25].

- **2D-DOA estimation using AVS arrays:** In this work, we focused on 1D-DOA estimation using sparse linear arrays, but sparse arrays such as the two dimensional nested array can be constructed [27] [28]. Moreover, since a single AVS is capable of measuring of both the azimuth and elevation angles, a one-dimensional linear array can also be used for 2D-DOA estimation. Preliminary results show that both approaches to 2D-DOA estimation, the spatial-velocity smoothing and Khatri-Rao subspace approach for quasi-stationary sources, provide satisfactory results. However, the upper bound on the number of sources that can be identified uniquely is more complicated. For identification of both azimuth and elevation angles, the conditions for full rank of the array manifold, \mathbf{A} , are harder to derive since linearly dependent columns taking two parameters θ and ϕ are easier to construct. However, stochastic characterizations for full-rank of \mathbf{A} can be used [29] [30]. Furthermore, for unique identifiability using subspace based approaches like MUSIC in two dimensions, it is neither necessary nor sufficient for \mathbf{A} to be full column rank [28] [31]. Thus, it would be interesting to investigate the necessary, sufficient and well-posed conditions for unique 2D-source identifiability.
- **Compressive sensing:** It is known from compressive sensing theory that recovery of parameters from sparse measurements is possible. For recovery of source DOAs which are sparse in the DOA-space using the ℓ_1 -norm, the sensing matrix is expected to satisfy certain properties. Similar to [32], a correlation-aware recovery method can be formulated on the correlation domain data model. By gridding the angular space at a pre-determined resolution, the source DOAs can be uniquely recovered from \mathbf{z} in (3.4) provided that certain properties like the restricted isometry property (RIP) and the mutual incoherence of $\mathbf{A}^* \circ \mathbf{A}$ are satisfied [33].
- **Other submodular costs for sensor selection:** In the context of sensor selection, there are other information theoretic costs like mutual information and entropy [19] [34] which assume a probabilistic model on the measurements. Costs similar to the weighted frame potential can be formulated since these costs are submodular. Greedy algorithms to optimize these functions could be used and it would be interesting to compare their performance to the weighted frame potential and the weighted D-optimal criterion, as well as the convex optimization based methods.

Preliminaries of Covariance Sensing



In this chapter, we provide the background for the first part thesis. We briefly describe the concept of sparse rulers, and spatial smoothing.

A.1 Sparse Arrays

In recent work, the problem of detecting more sources than the number of physical scalar sensors has been studied by considering non-uniform linear arrays [5], [4], [6]. The basic concept is to increase the spacing between certain antenna elements so that the number of redundant spacings in the co-domain, called the co-array, are reduced. We call the resulting array a *sparse array*. By constructing such sparse arrays smartly, a significant increase in the degrees of freedom (DOF) can be attained. all integer differences 0 to $\alpha - 1$ using only M marks where $M < \alpha$. The DOF is the number of unique measurements available, and this directly affects the number of source DOAs that can be estimated. More specifically, consider an array of M sensors. Let r_i denote the position of the i th element. The *co-array* set is defined by taking the distinct values from the difference set $\mathcal{S} = \{r_i - r_j\}, \forall i, j \in \{1, 2, \dots, M\}$. Thus, each element in the difference co-array set corresponds to the spatial correlation lag between an element pair with that separation. The DOF, denoted by α , is essentially the length of the co-array. However, the difference set may have some redundancies due to the fact that some of the spatial lags between array elements may be repeated. By reducing the number of redundant spacings in the difference set, the DOF can be increased. A *hole* in the co-array set refers to a missing lag. Thus, the aim is to design the array such that the co-array set does not contain any holes so that the co-array is a filled ULA and has minimal/no repeating correlation lags. The maximum number of sources that can be detected is then limited by the highest marking present in the co-array set. The role of the difference or co-array set naturally arises in several signal processing applications such as the computation of second-order moments as we will see in Chapters 3 and 4. We consider the following two non-linear arrays:

- *Minimum Redundancy Array (MRA)*: The minimum redundancy arrays (MRAs) or *minimum sparse rulers*, first introduced in [4], form a class of arrays which achieve the largest aperture with the constraint that the co-array is a ULA. That is, the MRA consists of the minimum number of non-uniformly spaced sensors such that the distinct elements of difference set \mathcal{S} forms a ULA. In other words, for an MRA with a given DOF, α , the co-array set is a ULA with the spatial lags ranging from $-(\alpha - 1)/2$ to $(\alpha - 1)/2$. They have the lowest redundancy for a given number of antennas M and have no holes in the co-array. Unfortunately,

a closed-form expression for the MRA topology and the achievable DOF for a given M is not available. These values can only be obtained through an expensive exhaustive search. Figure A.1a shows an illustration of one such MRA consisting of $M = 6$ elements. Here, $r_m, m = 1, 2, \dots, M$ refers to the position of the m th array element. The DOF for this array is equal to 25. Table A.1 lists some MRA configurations and their DOF compiled from [35]. The separations between the array elements are explicitly represented by the numbers in the configuration column of the table.

| M | DOF(α_{MRA}) | Configuration |
|-----|------------------------------|---------------|
| 2 | 3 | ·1· |
| 3 | 7 | ·1·2· |
| 4 | 13 | ·1·3·2· |
| 5 | 19 | ·1·3·3·2· |
| 6 | 27 | ·1·5·3·2·2· |

Table A.1: Some MRA configurations.

- *Nested Array*: Nested arrays are another class of non-uniform arrays which are also capable of increasing the degrees of freedom [5]. Similar to MRAs, the coarray set is a ULA with spatial lags ranging from $-(\alpha - 1)/2$ to $(\alpha - 1)/2$. They are constructed by nesting two or more ULAs and provide up to $O(M^2)$ degrees of freedom using only M elements. More specifically, a two-level nested array is constructed by the union of two sets, namely, $\mathcal{S}_{\text{inner}}$ containing M_1 elements and $\mathcal{S}_{\text{outer}}$ containing M_2 elements. The sets $\mathcal{S}_{\text{inner}}$ and $\mathcal{S}_{\text{outer}}$ are given by

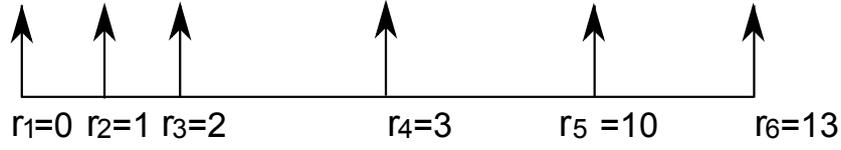
$$\begin{aligned}\mathcal{S}_{\text{inner}} &= \{md, m = 1, 2, \dots, M_1\}, \\ \mathcal{S}_{\text{outer}} &= \{m(M_1 + 1)d, m = 1, 2, \dots, M_2\},\end{aligned}\tag{A.1}$$

where d is the spacing of the inner array and $M_1 + M_2 = M$. The optimal distribution of the sensors for a two-level nested array to maximize the DOF is summarized in Table A.2 [5].

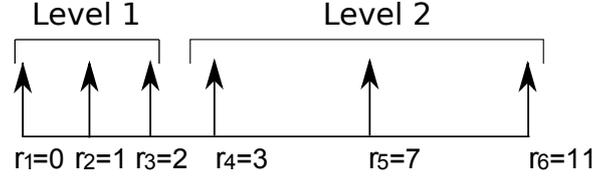
| M | M_1, M_2 | DOF(α_{nested}) |
|----------|--|---------------------------------|
| Even M | $M_1 = M_2 = \frac{M}{2}$ | $\frac{M^2}{2} + M - 1$ |
| Odd M | $M_1 = \frac{M-1}{2}, M_2 = \frac{M+1}{2}$ | $\frac{M^2-1}{2} + M$ |

Table A.2: Optimal two-level nested array configurations for even and odd M .

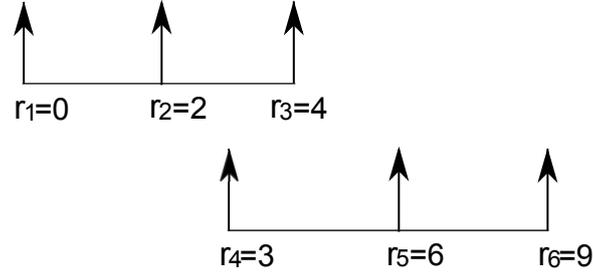
Unlike the MRAs, a closed-form expression can be obtained for the degrees of freedom of the proposed array as well as for the sensor locations. This is advantageous for large-scale arrays since the array is easily constructed. However, extending the nesting beyond two levels fails to produce a co-array which is a filled ULA, i.e. it has holes. For a given number of physical sensors, the MRA offers more DOF than the nested array. Figure A.1b shows an illustration of a



(a) Minimum Sparse Ruler Array, $M = 6$ $\text{DOF} = \alpha_{\text{MRA}} = 27$.



(b) Two-level Nested Array, $M = 6$, $\text{DOF} = \alpha_{\text{nested}} = \frac{M^2}{2} + M - 1 = 23$.



(c) Coprime Array, $J = 2$, $B = 3$, $M = B + 2J - 1 = 6$, $\text{DOF} = \alpha_{\text{coprime}} = 2JB + 1 = 13$.

Figure A.1: Sparse Arrays.

two-level nested array of $M = 6$ elements. In the subsequent sections, these array topologies are extended to AVS arrays and the increase in the degrees of freedom is studied.

- *Coprime Arrays*: Another approach to increase the degrees of freedom is presented in [6] which uses a pair of coprime samplers. More specifically, two uniform linear arrays with spacings Jd and Bd , where J and B are coprime. Moreover, coprime numbers have a unique property: if J and B are coprime with $J < B$, there exist integers $0 \leq j \leq 2J - 1$ and $0 \leq b \leq B - 1$ such that $Bj - Jb = p$ where p is an integer such that $0 \leq p \leq JB$. Thus by varying j and b , all integers in the range $[-JB, JB]$ can be produced. An example of a coprime array with $J = 2$ and $B = 3$ is shown in Figure A.1c.

A.2 Spatial Smoothing

Spatial Smoothing was first introduced by Shan et al. in [13] to circumvent problems encountered in DOA estimation. For correlated sources, the source covariance matrix \mathbf{R}_s is non-diagonal. Furthermore, if any two sources are fully coherent, the array

manifold and \mathbf{R}_z are also rank deficient. Hence, they propose a method to divide the ULA into overlapping subarrays. Then, a smoothed covariance matrix is constructed by averaging over the covariance matrices of the smaller subarrays. More specifically, if the underlying ULA is divided into N_{tot} subarrays, each having N_{sub} physical sensors, then the smoothed covariance matrix is given by

$$\bar{\mathbf{R}}_z = \frac{1}{N_{\text{tot}}} \sum_{i=1}^{N_{\text{tot}}} \mathbf{R}_i,$$

where \mathbf{R}_i is the $N_{\text{sub}} \times N_{\text{sub}}$ covariance matrix of the i th subarray. Since the number of sensors available is fixed and the number of coherent sources that can be identified is limited by the size of the subarray, there is evidently a trade-off between the size of each subarray and the number of such subarrays that can be averaged. An improvement of the above technique in [36], called the *forward-backward spatial smoothing*, further increases the number of coherent sources that can be identified. The spatial smoothing technique has also been in [5, 6, 37] since the underlying coarray for the nested, coprime and the MRA are ULAs. As we will see in the next chapters, these coarrays naturally occur from taking the second order statistics.

A.3 Related Work

Acoustic vector sensors have been shown to outperform the traditional scalar sensors in source localization [38]. The vector sensors also allow the construction of smaller array apertures while providing a high accuracy and a high resolution as well as resolving the left/right ambiguity in source localization. Some practical works on AVSs deal with, for example, the localization of wideband sources using a single AVS [39] and electromagnetic source localization [40].

A lot of theoretical work exists on acoustic vector sensors. In [41], a general expression for the Cramér-Rao bound (CRB) for DOA estimation using an AVS-array is derived. Conventional and Capon (MVDR) beamforming for a vector-sensor array are considered and advantages over a scalar-sensor array are presented in [38]. Root-MUSIC and ESPRIT have been applied to 1-D and 2-D arrays of velocity sensors in [42] and [43], respectively. It has been shown that a maximum of two sources can be identified using a single AVS [31] as compared to a single source that can be identified using a 2-element scalar-sensor array. DOA estimation of quasi-stationarity sources using a single AVS has been studied in [12]. However, no generalization exists for an array of AVSs. Further, DOA estimation using nested AVSs has been studied using the concept of tensor modeling in [44]. For 1D sources, the maximum number of DOAs that is identifiable with an AVS array of M sensors is shown theoretically to be $3M - 1$ in [14].

Preliminaries of Sensor Selection

B

In this chapter, we discuss the state-of-the art sensor selection. The non-linear measurement model is briefly introduced followed by a discussion on the performance measures for non-linear models.

For a general non-linear model with additive Gaussian noise as,

$$y_i = f_i(\boldsymbol{\theta}) + n_i, \quad i = 1, 2, \dots, M, \quad (\text{B.1})$$

where y_i denotes the i th sensor measurement, and the function $f_i(\cdot)$ is non-linear in general. The noise process is denoted by n_i and we assume that it is additive and white with a variance σ^2 . We consider the problem of estimating an unknown parameter $\boldsymbol{\theta} \in \mathbb{R}^N$ from a subset K sensors from the available number of sensors M . Let its estimate be denoted by $\hat{\boldsymbol{\theta}}$. Further, we use $\mathbf{w} \in \{0, 1\}^M$ to denote the sensor selection vector, where w_i encodes whether the i th measurement is selected from the available set of measurements. Let $\mathbf{E}(\mathbf{w})$ be the estimation error covariance matrix given by

$$\mathbf{E}(\mathbf{w}) = \mathbb{E}((\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T). \quad (\text{B.2})$$

The sensor selection problem is formulated as the following optimization problem: Mathematically, the optimization problem is formulated as

$$\begin{aligned} \arg \min_{\mathbf{w}} \quad & f(\mathbf{E}(\mathbf{w})) \\ \text{subject to} \quad & \|\mathbf{w}\|_0 = K, \quad w_i \in \{0, 1\}, \quad i = 1, 2, \dots, M, \end{aligned} \quad (\text{B.3})$$

where $\|\cdot\|_0$ represents the l_0 -norm, K is the number of sensors to be selected and $f(\cdot)$ is a scalar function on \mathbf{E} . Thus, it is desired to design \mathbf{w} so that a measure of \mathbf{E} is optimized. Thus, it is desired to design \mathbf{w} so that a measure of \mathbf{E} is optimized. This problem has been called the *experiment design* problem. Usually, scalarizations of the experiment design problem use a measure related to the η -confidence ellipsoid. The η -confidence ellipsoid of $\boldsymbol{\theta}$ is defined as

$$\boldsymbol{\xi} = \{\mathbf{z} \mid (\mathbf{z} - \hat{\boldsymbol{\theta}})^T \mathbf{E}^{-1}(\mathbf{z} - \hat{\boldsymbol{\theta}}) \leq \beta\} \quad (\text{B.4})$$

where β is a constant that depends on the confidence level η . The following scalarization choices could be adopted: [15]:

1. *D-optimal design*- The determinant of the error confidence ellipsoid is minimized. This amounts to minimizing the volume of $\boldsymbol{\xi}$ for a certain confidence level. That is, $f(\mathbf{E}(\mathbf{w})) = \det(\mathbf{E}(\mathbf{w}))$.

2. *E-optimal design*- The norm of \mathbf{E} , or the maximum eigenvalue of \mathbf{E} is minimized. This is equivalent to minimizing the major axis of $\boldsymbol{\xi}$ since the major axis is proportional to $\|\mathbf{E}\|_2^{1/2}$. That is, $f(\mathbf{E}(\mathbf{w})) = \lambda_{max}(\mathbf{E}(\mathbf{w}))$.
3. *A-optimal design*- The trace of the error covariance matrix \mathbf{E} is minimized. This is equivalent to minimizing the arithmetic mean of the radii of $\boldsymbol{\xi}$. That is, $f(\mathbf{E}(\mathbf{w})) = \text{tr}(\mathbf{E}(\mathbf{w}))$.

The problem in (B.3) is essentially combinatorial in nature since an exhaustive search is intractable for large values of M and K . A lot of literature exists on solving the above problem with approximations.

B.1 Existing Art

There is a lot of literature on the subject of sensor selection when the function f_i in (B.1) is linear. A method based on convex optimization for linear measurement models is studied in [17]. By relaxing the non-convex Boolean constraint $\{0, 1\}^M$ to a convex box constraint $[0, 1]^M$, the optimization is solved efficiently. However, the complexity of the convex optimization i.e., $O(M^3)$.

In [16], a general non-linear measurement model has been considered and the Cramér-Rao bound (CRB) is adopted as a performance measure. The additive property of the Fisher Information Matrix (FIM) for independent observations is exploited to derive the performance constraints. A number of convex solvers to alleviate the computational complexity of the optimization have been proposed.

Alternately, one may consider directly minimizing the mean squared error (MSE). However, since the MSE has several local minima, proxies to the MSE are often considered. Some information-theoretic measures are entropy [34] [45], mutual information [19] and cross entropy [46] [47] [48]. A number of sensor selection objectives also satisfy submodularity, which is intuitively related to the concept of diminishing returns. For example, adding a sensor to an existing set is less beneficial than adding the same sensor to a subset of the existing set. The MSE is not submodular in general. A large volume of work on minimizing submodular functions exists [49] [50]. However, a fundamental result by Nemhauser et al. [7] shows that a greedy algorithm which maximizes a submodular function is near-optimal in terms of the MSE. In the purview of submodularity and linear measurement models, an approach based on minimization of a measure related to orthogonality of the rows of a sensing matrix, called frame potential (FP), is proposed in [20]. A greedy algorithm to maximize a submodular function related to the frame potential is presented in [20]. For certain assumptions on the sampling matrix, it is also shown to be near-optimal in terms of the MSE. An extension of this work in [22] uses a weighted frame potential measure for a signal lying in a Union of Subspaces (UoS).

In [19], the sensor observations are modeled as Gaussian Processes. *Mutual Information*, another submodular function, is used as a design criterion to choose the sensors which are the most informative about unsensed locations. both entropy and mutual information are submodular. Since mutual information is also monotonic, guarantees

on the performance can be provided. However, entropy is not (even approximately) monotonic, thus prohibiting the use of the results of [7]. Maximizing the mutual information between sensor measurements results in a selection of the most informative sensors.

B.2 Performance Measures and Convex Optimization

The main result of the discussion on Fisher Information Matrix (FIM) from [16] is stated here for the sake of completeness. Given that the regularity condition holds for \mathbf{y} , the covariance of an unbiased estimator $\hat{\boldsymbol{\theta}}$ satisfies

$$\mathbb{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T] \succeq \mathbf{C}(\boldsymbol{\theta}) = \mathbf{F}^{-1}(\boldsymbol{\theta}) \quad (\text{B.5})$$

where $\mathbf{C}(\boldsymbol{\theta})$ is the CRB matrix and $\mathbf{F}(\boldsymbol{\theta})$ is the $N \times N$ FIM matrix. The FIM is given by

$$\mathbf{F}(\boldsymbol{\theta}) = \sum_{i=1}^M \mathbf{F}_i(\boldsymbol{\theta}) = \sum_{i=1}^M \frac{1}{\sigma^2} \left(\frac{\partial f_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial f_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \quad (\text{B.6})$$

where $\mathbf{F}_i(\boldsymbol{\theta})$ is the FIM of the i th measurement.

Performance Measures

The constrained optimization problem in (B.3) may also be cast as:

$$\begin{aligned} \arg \min_{\mathbf{w}} \quad & \|\mathbf{w}\|_0 \\ \text{subject to} \quad & f(\mathbf{E}(\mathbf{w})) \leq \lambda \\ & w_i \in \{0, 1\}, i = 1, 2, \dots, M \end{aligned} \quad (\text{B.7})$$

where the l_0 -norm represents the number of non-zero elements in \mathbf{w} . The CRB is used in [16] to bound the performance of the estimator. Since the subset of sensors that yield a lower CRB also yield a lower MSE. To estimate the unknown parameter $\boldsymbol{\theta}$ with sufficiently high accuracy, we require the estimation error $\mathbf{e} = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ to fall within an origin-centered circle of radius R_e with probability greater than P_e i.e., $\forall \boldsymbol{\theta}$, $\Pr(\|\mathbf{e}\|_2 \leq R_e) \geq P_e$. Although several performance measures based on the accuracy requirement are possible, two popular measures are:

- *E-optimal Related Design*: To satisfy the accuracy requirement at $\boldsymbol{\theta}$, one of the conditions is $\lambda_{\min}\{\mathbf{F}(\mathbf{w}, \boldsymbol{\theta})\} \geq \lambda_g$ where λ_{\min} refers to the minimum eigenvalue of $\mathbf{F}(\mathbf{w}, \boldsymbol{\theta})$ and λ_g is translated from the accuracy requirement defined by R_e and P_e [16], [51]. By ensuring a lower bound on each eigenvalue of the matrix \mathbf{F} , it is

ensured that the semi-major axis of the η -confidence ellipsoid is restricted to λ_g . This can be expressed as an equivalent Linear Matrix Inequality (LMI) [16] as

$$\sum_{i=1}^M w_i \mathbf{F}_i(\boldsymbol{\theta}) - \lambda_g \mathbf{I}_N \succeq \mathbf{0}_N, \quad \forall \boldsymbol{\theta}. \quad (\text{B.8})$$

- *D-optimal Related Design*: Another sufficient constraint, related to the *D-optimal* experiment design is the constraint which minimizes the volume of the η -confidence ellipsoid. More specifically [16]

$$\log \det\{\mathbf{C}(\mathbf{w}, \boldsymbol{\theta})\} = \log \det\left\{\left(\sum_{i=1}^M w_i \mathbf{F}_i(\boldsymbol{\theta})\right)^{-1}\right\} \leq \lambda_{\det} \quad (\text{B.9})$$

where λ_{\det} is a threshold derived from [16].

Convex Relaxation

The l_0 -norm in (B.7) is non-convex and NP-hard. Hence, the l_0 -norm is substituted by its best convex surrogate, i.e., the l_1 -norm. By also relaxing the Boolean constraint in (B.7) to a convex box constraint, the optimization problem for the E-optimal related design is [16]

$$\begin{aligned} \arg \min_{\mathbf{w}} \quad & \|\mathbf{w}\|_1 \\ \text{subject to} \quad & \sum_{i=1}^M w_i \mathbf{F}_i - \lambda_g \mathbf{I}_N \succeq \mathbf{0}_N, \\ & w_i \in [0, 1], \quad i = 1, 2, \dots, M. \end{aligned} \quad (\text{B.10})$$

A similar optimization problem can be obtained for a D-optimal related design. In practice, $\boldsymbol{\theta}$ is unknown but is known to take values within a certain domain \mathcal{U} . Then, the domain is assumed to consist of a set of D points obtained by gridding the domain. If $\mathcal{U} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_D\}$, then the constraints of (B.8) and (B.9) have to be satisfied for every $\boldsymbol{\theta}_d \in \mathcal{U}$, $d = 1, 2, \dots, D$.

The above problem is a standard SDP problem which can be solved using interior-point methods such as CVX [52] and SeDuMi [18]. However, the computational complexity of the chosen solver depends on the number of grid points. While we may choose a larger number of grid points to describe the unknown parameter $\boldsymbol{\theta}$ more accurately, the computational burden also increases. Although other convex solvers of [16] may be used, the computational complexity of the convex optimization using interior-point methods is $O(DM^3)$. Evidently, this method is computationally intensive for large values of M . In Chapter 5, we propose an approach to alleviate the computational burden by choosing simpler cost functions. Furthermore, we compare the performance of the proposed approach to the convex optimization problem of (B.10) and similarly for the D-optimal related design.

B.3 Submodular Cost Functions

There are some cost functions, often encountered, which have special properties. As mentioned earlier, some submodular functions are mutual information, entropy, and the D-optimal related function $\log \det(\cdot)$. The principal advantage of greedy algorithms comes from the fact that they are easy to implement and require less computing resources. Further, a result of Nemhauser et al. [7] states that for a normalized monotone submodular function, the solution from the greedy algorithm is always within $(1 - 1/e)$ of the optimal solution. We first define the concept of submodularity and then introduce the theorem by Nemhauser et. al.

Definition 1 (Submodular Function): Let \mathcal{A} and \mathcal{B} be two finite sets such that $\mathcal{A}, \mathcal{B} \subset \mathcal{M}$. A real-valued function $g(\cdot)$ defined on the set of the subsets of \mathcal{M} is submodular if:

$$g(\mathcal{A}) + g(\mathcal{B}) \geq g(\mathcal{A} \cup \mathcal{B}) + g(\mathcal{A} \cap \mathcal{B}) \quad (\text{B.11})$$

In other words, if $\mathcal{A} \subset \mathcal{B} \subset \mathcal{M}$ for an incremental value $j \in \mathcal{M} - \mathcal{B}$,

$$g(\mathcal{A} + j) - g(\mathcal{A}) \geq g(\mathcal{B} + j) - g(\mathcal{B}). \quad (\text{B.12})$$

Additionally, a function $g(\cdot)$ is non-decreasing if for all $\mathcal{A}, \mathcal{B} \subset \mathcal{M}$, $g(\mathcal{A}) \leq g(\mathcal{B})$ and normalized if $g(\emptyset) = 0$. Next, the theorem of [7] is briefly stated.

Theorem 4. [7] *Let \mathcal{O} be the optimal set of sensors, so that $\mathcal{O} = \max_{\mathcal{A} \subset \mathcal{M}, |\mathcal{A}|=K} g(\mathcal{A})$. If $g(\cdot)$ is a normalized, monotone submodular function over a finite set \mathcal{M} and \mathcal{K} is the set of K elements chosen from \mathcal{M} by the greedy algorithm, then*

$$g(\mathcal{K}) \geq \left(1 - \frac{1}{e}\right) g(\mathcal{O}) \quad (\text{B.13})$$

Thus, if $g(\cdot)$ is a normalized, monotone submodular function, a greedy algorithm provides a solution which is always near to the optimal solution. An important difference between the classical design criteria (A-, D- and E- optimal designs [15]) and information-theoretic measures (such as mutual information and entropy) is that while the former approaches are tailored to minimize the error of predicting the unknown parameter at sensor locations, the latter approaches try to decrease the uncertainty over the entire parameter space.

Bibliography

- [1] Microflow Technologies, “Acoustic Vector Sensor,” <http://microflow.com/products/standard-probes/usp-regular.html>, 2015, [Online; accessed 26-May-2015].
- [2] Microflow AVISA BV, “Velocity Sensor,” <http://microflow-avisa.com/acoustic-vector-sensor/>, 2015, [Online; accessed 26-May-2015].
- [3] R.O. Schmidt, “Multiple emitter location and signal parameter estimation,” *Antennas and Propagation, IEEE Transactions on*, vol. 34, no. 3, pp. 276–280, Mar 1986.
- [4] A. Moffet, “Minimum-redundancy linear arrays,” *Antennas and Propagation, IEEE Transactions on*, vol. 16, no. 2, pp. 172–175, Mar 1968.
- [5] P. Pal and P.P. Vaidyanathan, “Nested arrays: A novel approach to array processing with enhanced degrees of freedom,” *Signal Processing, IEEE Transactions on*, vol. 58, no. 8, pp. 4167–4181, Aug 2010.
- [6] P. Pal and P.P. Vaidyanathan, “Coprime sampling and the music algorithm,” in *Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop (DSP/SPE), 2011 IEEE*, Jan 2011, pp. 289–294.
- [7] George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher, “An analysis of approximations for maximizing submodular set functions - i,” *Mathematical Programming*, vol. 14, no. 1, pp. 265–294, 1978.
- [8] Wing-Kin Ma, Tsung-Han Hsieh, and Chong-Yung Chi, “DOA estimation of quasi-stationary signals via khatri-rao subspace,” in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, April 2009, pp. 2165–2168.
- [9] Malcolm Hawkes and Arye Nehorai, “Acoustic vector-sensor correlations in ambient noise,” *Oceanic Engineering, IEEE Journal of*, vol. 26, no. 3, pp. 337–347, 2001.
- [10] Mostafa Kaveh and Arthur J Barabell, “The statistical performance of the music and the minimum-norm algorithms in resolving plane waves in noise,” *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 34, no. 2, pp. 331–341, 1986.
- [11] Weiguo Chen, Kon Max Wong, and James P Reilly, “Detection of the number of signals: A predicted eigen-threshold approach,” *Signal Processing, IEEE Transactions on*, vol. 39, no. 5, pp. 1088–1098, 1991.

- [12] Shengkui Zhao, Tigran Saluev, and Douglas L Jones, “Underdetermined direction of arrival estimation using acoustic vector sensor,” *Signal Processing*, vol. 100, pp. 160–168, 2014.
- [13] Tie-Jun Shan, M. Wax, and T. Kailath, “On spatial smoothing for direction-of-arrival estimation of coherent signals,” *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 33, no. 4, pp. 806–811, Aug 1985.
- [14] Yanqun Wu, Zhengliang Hu, Hong Luo, and Yongming Hu, “Source number detectability by an acoustic vector sensor linear array and performance analysis,” *Oceanic Engineering, IEEE Journal of*, vol. 39, no. 4, pp. 769–778, Oct 2014.
- [15] Stephen Boyd and Lieven Vandenberghe, *Convex Optimization*, Cambridge University Press, New York, NY, USA, 2004.
- [16] S.P. Chepuri and G. Leus, “Sparsity-promoting sensor selection for non-linear measurement models,” *Signal Processing, IEEE Transactions on*, vol. 63, no. 3, pp. 684–698, Feb 2015.
- [17] S. Joshi and S. Boyd, “Sensor selection via convex optimization,” *Signal Processing, IEEE Transactions on*, vol. 57, no. 2, pp. 451–462, Feb 2009.
- [18] Jos F. Sturm, “Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones,” 1998.
- [19] Andreas Krause, Ajit Singh, and Carlos Guestrin, “Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies,” *J. Mach. Learn. Res.*, vol. 9, pp. 235–284, June 2008.
- [20] J. Ranieri, A. Chebira, and M. Vetterli, “Near-optimal sensor placement for linear inverse problems,” *Signal Processing, IEEE Transactions on*, vol. 62, no. 5, pp. 1135–1146, March 2014.
- [21] M. Shamaiah, S. Banerjee, and H. Vikalo, “Greedy sensor selection: Leveraging submodularity,” in *Decision and Control (CDC), 2010 49th IEEE Conference on*, Dec 2010, pp. 2572–2577.
- [22] Dalia El Badawy, Juri Ranieri, and Martin Vetterli, “Near-optimal Sensor Placement for Signals lying in a Union of Subspaces,” in *22nd European Signal Processing Conference (EUSIPCO 2014)*, 2014, Invited Paper.
- [23] Kaare Brandt Petersen, “The matrix cookbook,” .
- [24] Max A. Woodbury, *Inverting modified matrices*, Statistical Research Group, Memo. Rep. no. 42. Princeton University, Princeton, N. J., 1950.
- [25] Yeo-Sun Yoon, Lance M Kaplan, and James H McClellan, “Tops: new doa estimator for wideband signals,” *Signal Processing, IEEE Transactions on*, vol. 54, no. 6, pp. 1977–1989, 2006.

- [26] H. Wang and M. Kaveh, “Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources,” *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 33, no. 4, pp. 823–831, Aug 1985.
- [27] P. Pal and P.P. Vaidyanathan, “Nested arrays in two dimensions, part i: Geometrical considerations,” *Signal Processing, IEEE Transactions on*, vol. 60, no. 9, pp. 4694–4705, Sept 2012.
- [28] P. Pal and P.P. Vaidyanathan, “Nested arrays in two dimensions, part ii: Application in two dimensional array processing,” *Signal Processing, IEEE Transactions on*, vol. 60, no. 9, pp. 4706–4718, Sept 2012.
- [29] M.C. Vanderveen, A.-J. van der Veen, and A. Paulraj, “Estimation of multipath parameters in wireless communications,” *Signal Processing, IEEE Transactions on*, vol. 46, no. 3, pp. 682–690, Mar 1998.
- [30] A. Gershman and N. Sidiropoulos, *Space-Time Processing for MIMO Communications*, Wiley, 2005.
- [31] B. Hochwald and Arye Nehorai, “Identifiability in array processing models with vector-sensor applications,” *Signal Processing, IEEE Transactions on*, vol. 44, no. 1, pp. 83–95, Jan 1996.
- [32] P. Pal and P.P. Vaidyanathan, “Correlation-aware techniques for sparse support recovery,” in *Statistical Signal Processing Workshop (SSP), 2012 IEEE*, Aug 2012, pp. 53–56.
- [33] Emmanuel J Candès and Michael B Wakin, “An introduction to compressive sampling,” *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 21–30, 2008.
- [34] Hanbiao Wang, G. Pottie, K. Yao, and D. Estrin, “Entropy-based sensor selection heuristic for target localization,” in *Information Processing in Sensor Networks, 2004. IPSN 2004. Third International Symposium on*, April 2004, pp. 36–45.
- [35] John Leech, “On the representation of 1, 2,..., n by differences,” *Journal of the London Mathematical Society*, vol. 1, no. 2, pp. 160–169, 1956.
- [36] S.U. Pillai and B.H. Kwon, “Forward/backward spatial smoothing techniques for coherent signal identification,” *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 37, no. 1, pp. 8–15, Jan 1989.
- [37] S. Shakeri, D.D. Ariananda, and G. Leus, “Direction of arrival estimation using sparse ruler array design,” in *Signal Processing Advances in Wireless Communications (SPAWC), 2012 IEEE 13th International Workshop on*, June 2012, pp. 525–529.
- [38] M. Hawkes and Arye Nehorai, “Acoustic vector-sensor beamforming and capon direction estimation,” *Signal Processing, IEEE Transactions on*, vol. 46, no. 9, pp. 2291–2304, Sep 1998.

- [39] M. Hawkes and Arye Nehorai, “Wideband source localization using a distributed acoustic vector-sensor array,” *Signal Processing, IEEE Transactions on*, vol. 51, no. 6, pp. 1479–1491, June 2003.
- [40] Arye Nehorai and E. Paldi, “Vector-sensor array processing for electromagnetic source localization,” *Signal Processing, IEEE Transactions on*, vol. 42, no. 2, pp. 376–398, Feb 1994.
- [41] Arye Nehorai and E. Paldi, “Acoustic vector-sensor array processing,” *Signal Processing, IEEE Transactions on*, vol. 42, no. 9, pp. 2481–2491, Sep 1994.
- [42] K.T. Wong and M.D. Zoltowski, “Root-MUSIC-based azimuth-elevation angle-of-arrival estimation with uniformly spaced but arbitrarily oriented velocity hydrophones,” *Signal Processing, IEEE Transactions on*, vol. 47, no. 12, pp. 3250–3260, Dec 1999.
- [43] K.T. Wong and M.D. Zoltowski, “Extended-aperture underwater acoustic multisource azimuth/elevation direction-finding using uniformly but sparsely spaced vector hydrophones,” *Oceanic Engineering, IEEE Journal of*, vol. 22, no. 4, pp. 659–672, Oct 1997.
- [44] Keyong Han and A. Nehorai, “Nested vector-sensor array processing via tensor modeling,” *Signal Processing, IEEE Transactions on*, vol. 62, no. 10, pp. 2542–2553, May 2014.
- [45] Michael C Shewry and Henry P Wynn, “Maximum entropy sampling,” *Journal of Applied Statistics*, vol. 14, no. 2, pp. 165–170, 1987.
- [46] M. Naeem, S. Xue, and D.C. Lee, “Cross-entropy optimization for sensor selection problems,” in *Communications and Information Technology, 2009. ISCIT 2009. 9th International Symposium on*, Sept 2009, pp. 396–401.
- [47] Naren Ramakrishnan, Chris Bailey-kellogg, Satish Tadepalli, Varun N. Pandey, and Varun N. P, “Gaussian processes for active data mining of spatial aggregates,” in *In Proceedings of the SIAM International Conference on Data Mining*, 2005.
- [48] D MacKay, “Information-based objective functions for active data selection,” *Neural Computation*, vol. 4, no. 4, pp. 590–604, July 1992.
- [49] S. Fujishige, *Submodular Functions and Optimization: Second Edition*, Annals of Discrete Mathematics. Elsevier Science, 2005.
- [50] Alexander Schrijver, *Combinatorial optimization: polyhedra and efficiency*, vol. 24, Springer, 2003.
- [51] Tao Wang, G. Leus, and Li Huang, “Ranging energy optimization for robust sensor positioning based on semidefinite programming,” *Signal Processing, IEEE Transactions on*, vol. 57, no. 12, pp. 4777–4787, Dec 2009.
- [52] Michael Grant and Stephen Boyd, “CVX: Matlab software for disciplined convex programming, version 2.1,” <http://cvxr.com/cvx>, Mar 2014.