PhD-FSTM-2021-005

The Faculty of Sciences, Technology and Medicine

DISSERTATION

Defence held on 27/01/2021 in Luxembourg

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG EN INFORMATIQUE

by

Saeid Sedighi

Born on 06 August 1987 in Tehran, (Iran)

Direction of Arrival Estimation and Localization Exploiting Sparse and One-Bit Sampling

Dissertation defense committee

Dr. Björn Ottersten, Dissertation Supervisor

Professor and Director of SnT, Université du Luxembourg, Luxembourg

Dr. Moeness G. Amin

Professor, Villanova University, The United States

Dr. Geert Leus, Vice Chairman

Professor, Delft University of Technology, The Netherlands

Dr. Marcus Völp, Chairman

Associate Professor, Université du Luxembourg, Luxembourg

Dr. Bhavani Shankar

Research Scientist, Université du Luxembourg, Luxembourg

Affidavit

I hereby confirm that the PhD thesis entitled "Direction of Arrival Estimation and Localization Exploiting Sparse and One-Bit Sampling" has been written independently and without any other sources than cited.

Luxembourg, _____ _____

Saeid Sedighi

Abstract

Data acquisition is a necessary first step in digital signal processing applications such as radar, wireless communications and array processing. Traditionally, this process is performed by uniformly sampling signals at a frequency above the Nyquist rate and converting the resulting samples into digital numeric values through high-resolution amplitude quantization. While the traditional approach to data acquisition is straightforward and extremely well-proven, it may be either impractical or impossible in many modern applications due to the existing fundamental trade-off between sampling rate, amplitude quantization precision, implementation costs, and usage of physical resources, e.g. bandwidth and power consumption. Motivated by this fact, system designers have recently proposed exploiting sparse and few-bit quantized sampling instead of the traditional way of data acquisition in order to reduce implementation costs and usage of physical resources in such applications. However, before transition from the tradition data acquisition method to the sparsely sampled and few-bit quntized data acquisition approach, a study on the feasibility of retrieving information from sparsely sampled and few-bit quantized data is first required to be conducted. This study should specifically seek to find the answers to the following fundamental questions:

1. Is the problem of retrieving the information of interest from sparsely sampled and few-bit quantized data an identifiable problem? If so, what are the identifiability conditions?

2. Under the identifiability conditions: what are the fundamental performance bounds for the problem of retrieving the information of interest from sparsely sampled and few-bit quantized data? and how close are these performance bounds to those of retrieving the same information from the data acquired through the traditional approach?

3. Does there exist any computationally efficient algorithm for retrieving the information of interest from sparsely sampled and few-bit quantized data capable of achieving the corresponding performance bounds?

My thesis focuses on finding the answers to the above fundamental questions for the problems of Direction of Arrival (DoA) estimation and localization, which are of the most important information retrieval problems in radar, wireless communication and array processing. Inthis regard, the first part of this thesis focuses on DoA estimation using Sparse Linear Arrays (SLAs). I consider this problem under three plausible scenarios from quantization perspective. Firstly, I assume that an SLA quantized the received signal to a large number of bits per samples such that the resulting quantization error

can be neglected. Although the literature presents a variety of estimators under such circumstances, none of them are (asymptotically) statistically efficient. Motivated by this fact, I introduce a novel estimator for the DoA estimation from SLA data employing the Weighted Least Squares (WLS) method. I analytically show that the large sample performance of the proposed estimator coincides with the Cramér-Rao Bound (CRB), thereby ensuring its asymptotic statistical efficiency. Next, I study the problem of DoA estimation from one-bit SLA measurements. The analytical performance of DoA estimation from one-bit SLA measurements has not yet been studied in the literature and performance analysis in the literature has be limited to simulations studies. Therefore, I study the performance limits of DoA estimation from one-bit SLA measurements through analyzing the identifiability conditions and the corresponding CRB. I also propose a new algorithm for estimating DoAs from one-bit quantized data. I investigate the analytical performance of the proposed method through deriving a closed-form expression for the covariance matrix of its asymptotic distribution and show that it outperforms the existing algorithms in the literature. Finally, the problem of DoA estimation from low-resolution multi-bit SLA measurements, e.g. 2 or 4 bit per sample, is studied. I develop a novel optimization-based framework for estimating DoAs from low-resolution multi-bit measurements. It is shown that increasing the sampling resolution to 2 or 4 bits per samples could significantly increase the DoA estimation performance compared to the one-bit sampling case while the power consumption and implementation costs are still much lower compared to the high-resolution sampling scenario.

In the second part of the thesis, the problem of target localization is addressed. Firstly, I consider the problem of passive target localization from one-bit data in the context of Narrowband Internet-of-Things (NB-IoT). In the recently proposed narrowband IoT (NB-IoT) standard, which trades off bandwidth to gain wide area coverage, the location estimation is compounded by the low sampling rate receivers and limited-capacity links. I address both of these NB-IoT drawbacks by consider a limiting case where each node receiver employs one-bit analog-to-digital-converters and propose a novel low-complexity nodal delay estimation method. Then, to support the low-capacity links to the fusion center (FC), the range estimates obtained at individual sensors are converted to one-bit data. At the FC, I propose a novel algorithm for target localization with the aggregated one-bit range vector. My overall one-bit framework not only complements the low NB-IoT bandwidth but also supports the design goal of inexpensive NB-IoT location sensing. Secondly, in order to reduce bandwidth usage for performing high precision time of arrival-based localization, I developed a novel sparsity-aware target localization algorithm with application to automotive radars.

The thesis concludes with summarizing the main research findings and some remarks on future directions and open problems.

Dedicated to my family

Acknowledgements

First of all, I would like to express my gratitude to my supervisor, Prof. Björn Ottersten, for his continuous support and encouragement throughout my PhD studies which was vital in making this thesis a reality. He always took time out of his busy schedule to give good advice and guidance. I admire his respect, encouragement and positive attitude towards his team. I am grateful to him for providing me with the opportunity to pursue my PhD studies at University of Luxembourg.

I would also like to thank my co-supervisor, Dr. Bhavani Shankar. His daily supervision and his positive attitude played a crucial role in the accomplishment of this thesis. Throughout my PhD, I had numerous meeting with Dr. Shankar discussing various issues ranging from my research to daily issues. I am grateful for his friendly personality; he was not only my co-supervisor, but also, sometimes acted as a big brother or a friend under different circumstances

Moreover, I would like to thank my PhD committee member, Prof. Geert Leus. His extensive knowledge and experience helped me to be on the right track with my research. I grateful to Prof. Moeness G. Amin and Prof. Marcus Völp for accepting to be in my defense committee.

I wish to thank Prof. Mojtaba Soltanalian and Dr. Vijay Mishra Kumar for their scientific collaboration during my PhD studies. I also wish to thank all my (current and past) colleagues of the SIGCOM group of SnT for creating an enjoyable working environment. Specially, I would like to thank Alireza, Arsham, Ehsan, Alberto, Christian and Hossein.

Finally, my foremost gratitude and recognition goes to my family for their unwavering support. Let me start with my fiancé Najva, thanks a lot for showing me how to love and for always being there for me. My deepest appreciation goes to my parents, Mansour and Manzar, and my sister, Atrin, for their unconditional and insatiable dedication and love.

Saeid Sedighi
Luxembourg, December 2020

Contents

List of Figures

List of Tables

Acronyms

3GPP	3rd Generation Partnership Project
ACBM	Augmented Covariance-Based Method
ACM	Augmented Covariance Matrix
ASCM	Augmented Sample Covariance Matrix
ANTARES	iterative joint RaNge-TARget location EStimation
ADC	Analog-to-Digital Converter
CAB-ESPRIT	Co-Array-Based ESPRIT
CAB-MUSIC	Co-Array-Based MUSIC
CDC	Continuous-to-Discrete Converter
CLT	Central Limit Theorem
CRB	Cramér-Rao Bound
DFT	Discrete Fourier Transform
DoA	Direction of Arrival
EOCAB-MUSIC	Enhanced One-bit Co-Array-Based MUSIC
ESPRIT	EStimation of Signal Parameters via Rotational Invariant Techniques
FIM	Fisher Information Matrix
FC	Fusion Center
GPS	Global Positioning System
ICAB-MUSIC	Infinite-bit Co-Array-Based MUSIC
IoT	Internet-of-Things
I-CRB	Infinite-bit Cramér-Rao Bound
KKT	Karush–Kuhn–Tucker
LS	Least Squares
MIMO	Multiple Input Multiple Output
MSE	Mean-Squared Error

ML	Maximum Likelihood
MRA	Minimum Redundancy Arrays
MUSIC	MUltiple SIgnal Classification
NB-IoT	Narrowband Internet-of-Things
N-RMSE	Normalized Root-Mean-Squared Error
OCAB-MUSIC	One-bit Co-Array-Based MUSIC
O-CRB	One-bit Cramér-Rao Bound
PDF	Probability Density Function
PSK	Phase Shift Keying
QPSK	Quadrature Phase Shift Keying
RMSE	Root-Mean-Squared Error
RSSI	Received Signal Strength Indicator
SCAB-MUSIC	Structured Covariance-Based Method
SDP	Semidefinite Programming
SS	signal strength
SBM	Sparsity-Based Method
SLA	Sparse Linear Array
SNR	Signal-to-Noise Ratio
SPA	Sparse and Parametric Approach
TDoA	Time Difference of Arrival
ToA	Time of Arrival
ULA	Uniform Linear Array
WLS	Weighted Least Squares

Notations

a	Scalar
$\mathbf{a}$	Column vector
$\mathbf{A}$	Matrix
$\mathbb{A}$	Set
$[\mathbf{a}]_i$	i^{th} entry of $\mathbf{a}$
$[\mathbf{A}]_{i,j}$	$(i,j)^{\text{th}}$ entry of $\mathbf{A}$
$\lvert a\rvert$	Absolute value of a
$\lVert\mathbf{a}\rVert_1$	ℓ_1-norm of $\mathbf{a}$
$\lVert\mathbf{a}\rVert_2$	ℓ_2-norm of $\mathbf{a}$
$\lVert\mathbf{A}\rVert_F$	Frobenius norm of $\mathbf{A}$
$\text{vec}\,(\mathbf{A})$	Vectorization of $\mathbf{A}$ by concatenating its columns
$\text{tr}(\mathbf{A})$	Trace of $\mathbf{A}$
$\text{rank}(\mathbf{A})$	Rank of $\mathbf{A}$
$\mathbf{A}^\dagger$	Pseudoinverse of $\mathbf{A}$
$\Pi_{\mathbf{A}}^{\perp}$	Projection matrix onto the null space of $\mathbf{A}^H$
$\Pi_{\mathbf{A}}$	Projection matrix onto the range space of $\mathbf{A}$
$\mathcal{R}(\mathbf{A})$	Range space of $\mathbf{A}$
$\mathcal{N}(\mathbf{A})$	Null space of $\mathbf{A}$
$\mathbf{A} \succeq \mathbf{0}$	Positive semidefinite matrix
$\mathbf{a} \succeq \mathbf{0}$	Vector with non-negative elements
$\text{diag}(\mathbf{a})$	Diagonal matrix composed of the elements of $\mathbf{a}$
$\lvert\mathbb{A}\rvert$	Cardinality of $\mathbb{A}$
$(a_1, a_2, \cdots, a_n)$	n-tuple with elements of $a_1, a_2, \cdots, a_n$
$\mathbf{I}_M$	$M \times M$ identity matrix
$\otimes$	Kronecker product

$\odot$	Khatri-Rao product
$\circ$	Hadamard product
$(\cdot)^T$	Transpose of $(\cdot)$
$(\cdot)^*$	Conjugate of $(\cdot)$
$(\cdot)^H$	Hermitian of $(\cdot)$
$\mathbb{E}\{\cdot\}$	Expectation of $\{\cdot\}$
$\Re\{\cdot\}$	Real part of $\{\cdot\}$
$\Im\{\cdot\}$	Imaginary part of $\{\cdot\}$
$(\cdot)^{(k)}$	Value of $(.)$ at the k^{th} iteration
$\delta[.]$	Kronecker's delta
sgn(.)	sign function
$\lceil . \rceil$	Floor function
$\lfloor . \rfloor$	Ceiling function
deg(.)	Degree of a polynomial
$\mathcal{CN}(\mathbf{a}, \mathbf{A})$	Circular complex Gaussian distribution with mean $\mathbf{a}$ and covariance matrix $\mathbf{A}$

"The moving finger writes; and, having writ, moves on: nor all thy piety nor wit shall lure it back to cancel half a line, nor all thy tears wash out a word of it."

Omar Khayyam

Chapter 1

Introduction

1.1 Motivation

The problem of inferring location information of animate and inanimate objects from radio sensor measurements has been a topic of research for decades. Direction of Arrival (DoA) estimation and source/target localization are two key aspects of this problem. DoA estimation refers to the process of finding the direction from which some electromagnetic waves arrive at a particular point from the output of a set of receiving antennas which form an array of sensors. A basic model for DoA estimation problem is illustrated in Fig. 1.1. Indeed, Fig. 1.1 depicts multiple sources radiate electromagnetic waves, which are received by a set of co-located equally-spaced linear antennas, termed a Uniform Linear Array (ULA). The ULA measurements are then processed by one of DoA estimation algorithms to find source DoAs, i.e., $\theta_k \forall k$ [1]. In Fig. 1.1, the sources are depicted in red empty circles and the wavefronts of the electromagnetic waves are indicated in red curves. In addition, the receiving antennas are shown in blue solid circles. Localization denotes the process of finding the location of some sources/targets, radiating electromagnetic waves, with respect to a reference point. Localization can be performed by using the measurements of either co-located sensors or distributed sensors. In the former, we simply need to estimate the distances between the sources/targets and the co-located sensor array in addition to the DoAs. The later exploit the spatial diversity, provided by the measurements of spatially separated sensors, to localize sources/targets. An illustrative example of a distributed localization scenario is given in Fig. 1.2. Fig. 1.2 shows a number of distributed sensors which receive the electromagnetic wave radiated by a source/target. The sensor measurements are individually processed to estimate the distances between the source/target and the sensors. Then, the distance estimates are used to localize the source/target. Fig. 1.2 shows graphically how the distance estimates

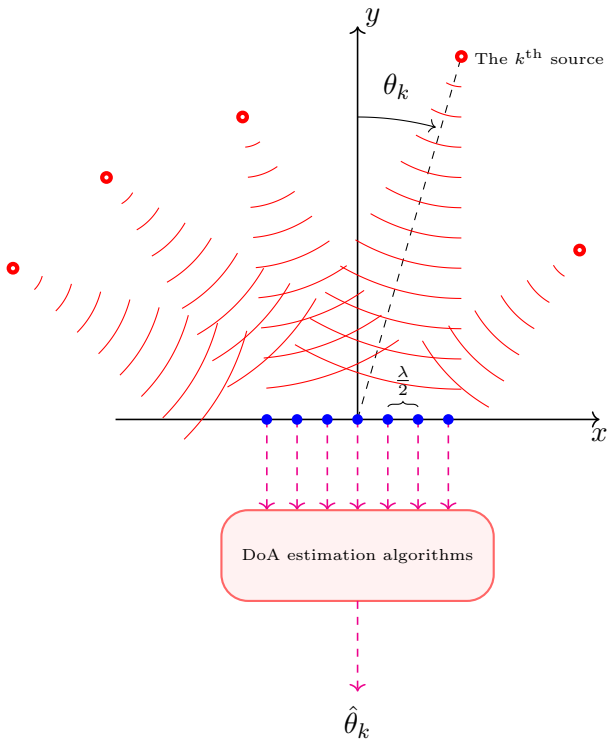


FIGURE 1.1: A basic model for DoA estimation problem.

can be employed for localizing the source/target. In Fig. 1.2, the distributed sensors are shown in blue solid circles and the source/target of interest is depicted in a red empty circle.

DoA estimation and target/source localization are two fundamental problems in signal processing with extensive applications in wireless communications, radar, sonar, remote sensing and speech processing [1–4]. Performing accurate DoA estimation and target/-source localization with low resource usage and low implementation costs has been always a serious challenge amongst all the aforementioned applications. Exploiting sparse and low-resolution sampling strategies could address this challenge to a great extent.

The concept of sparse sampling is based on the fact that the dimension of information of interest in many practical applications is much smaller than the dimension of the observed data. Under such circumstances, only a small portion of data suffices for recovering the information of interest. This fact allows for a significant reduction in usage of physical resources such as bandwidth, power, and the number of sensors through collecting only sparse low-dimensional samples of data. In consequence, sparse sampling has found applications in a large number of areas such as communication systems [5–7], radar [8–12], array processing [13, 13–17] and image processing [18–20].

Further, analog-to-digital conversion is an essential step in digital signal processing. Ideally, the analog-to-digital conversion requires an infinite number of bits to accurately

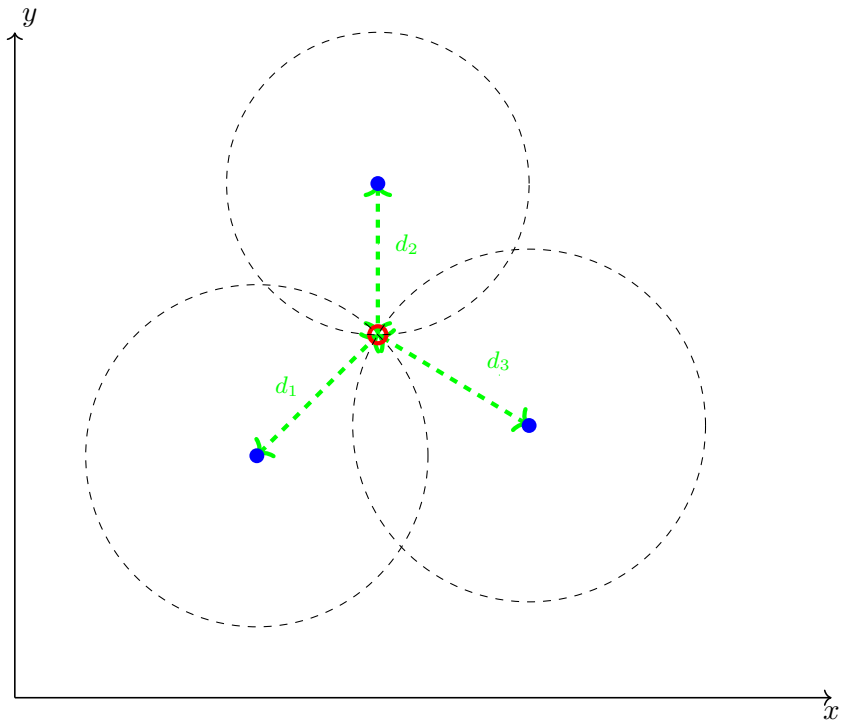


FIGURE 1.2: An illustrative example of a distributed localization scenario.

represent continuous-time signals in the digital domain. In practice, the signal is quantized to a finite number of bits leading to errors in the digital approximation of the original analog signal. If the sampling resolution is large enough, this error has negligible effect on digital signal processing. However, power consumption and production cost of Analog-to-Digital Converters (ADCs) increase exponentially with the number of quantization bits and sampling frequency [21]. Hence, deployment of high-resolution ADCs in many modern applications, e.g. cognitive radio [22], cognitive radars [23], automotive radars [24], radio astronomy[25] and massive multiple-input multiple-output (MIMO) systems[26], is not economically viable owing to limitations on power consumption and production costs. In order to reduce energy consumption and production cost in such applications, researchers and system designers have recently proposed using low-resolution ADCs. As an extreme case of low-resolution ADCs, one-bit ADCs, which convert an analog signal into digital data using a single bit per sample, has received significant attention in the literature [27–33]. One-bit ADCs offer an extremely high sampling rate at a low cost and very low energy consumption [21]. Additionally, they enjoy the benefits of relatively easy implementation due to their simple architecture [34].

This thesis aims at answering some of the fundamental open problems in the context of DoA estimation and source/target localization in presence of sparse and one-bit samples. In the first part of the thesis, we are interested in the DoA estimation problem. DoAs of electromagnetic waves/sources are typically retrieved through processing the data

collected a ULA. However, if the electromagnetic waves/sources impinging on a ULA are uncorrelated, the dimension of the parameters have to be estimated in the DoA estimation problem will be considerably smaller than the dimension of the data observed by the ULA [16]. As a result, based on sparse sampling concept, exploiting the prior information about the uncorrelatedness of sources allows for reducing the size of acquired data by removing antennas from ULAs. This leads to a new array configuration, referred to as a Sparse Linear Array (SLA). In the first part of this thesis, we focus on the problem of estimating DoAs of uncorrelated source signals from measurements collected by SLAs. This problem is investigated under three plausible scenarios from quantization perspective:

1. The analog signal received by an SLA is converted into digital data using a large number of bits per sample such that the resulting quantization errors can be neglected.
2. The analog received signal by an SLA is converted into digital data using a single bit per sample.
3. The analog signal received by an SLA is converted into digital data using p bit per sample. Here p is assumed to be an arbitrary positive integer number but not large enough such that the resulting quantization errors can be neglected.

In case of scenario 1, numerous algorithms for DoA estimation has been proposed in the literature [15, 16, 35–38]. However, none of them are able to achieve the CRB performance. A detail review of the state of the art in this area is provided in Chapter 2. Motivated by this fact, in Chapter 3, we go beyond the state of the art in this area by introducing a new algorithm for DoA estimation which is proved to be asymptotically statistically efficient. Concerning scenario 2, the analytical performance of DoA estimation from one-bit SLA measurements has not yet been studied in the literature and performance analysis in the literature has be limited to simulations studies. The reader is referred to Chapter 2 for a detailed literature review [32, 39–43]. Therefore, we study the performance limits of DoA estimation from one-bit SLA measurements through analyzing the identifiability conditions and the corresponding Cramér-Rao Bound (CRB) in Chapter 4. We also introduce a new DoA estimator with enhanced performance compared to the state of the art in this case. To the best of our knowledge, scenario 3 has never been investigated in the literature. In Chapter 5, we establish an optimization framework whereby DoA estimation from low-resolution multi-bit measurements, say 2 or 4 bits per sample, becomes possible.

In the second part of the thesis, the problem of distributed target localization is addressed. In Chapter 6, we consider the problem of passive target localization from one-bit data

in the context of Narrowband Internet-of-Things (NB-IoT). In this context, apart from application to NB-IoT localization, our work is the first work in the context of one-bit sampling in a passive and distributed radar array setting. Then, in order to reduce bandwidth usage for performing high precision time of arrival-based localization, with application to automotive radars, we developed a novel sparsity-aware target localization algorithm in Chapter 7.

1.2 Thesis Outline

This thesis is comprised of two major parts. The first part of the thesis, including Chapters 3, 4 and 5, is dedicated to DoA estimation from SLA measurements. The second part of the thesis, including Chapters 6 and 7, studies distributed target localization in presence of sparse and one-bit samples. It is noted that all the chapters of the thesis are self-contained. The thesis outline is as follows:

- Chapter 2 conducts a review on the most relevant state-of-the-art works in the literature and highlight the contributions of the thesis compared to the literature.
- Chapter 3 considers DoA estimation from infinite-bit SLA measurements. Although the literature presents a variety of estimators in the context of DoA estimation from SLA measurements, none of them are proven to be statistically efficient. This chapter introduces a novel estimator for DoA estimation employing the Weighted Least Squares (WLS) method, which is shown to be asymptotically statistically efficient.
- In Chapter 4, we study the performance limits of DoA estimation from one-bit SLA measurements through analyzing the identifiability conditions and the corresponding Cramér-Rao Bound (CRB). We also introduce a new DoA estimator with enhanced performance compared to the state of the art in this case.
- Chapter 5 considers the problem of DoA estimation from low-resolution multi-bit SLA measurements. We develop a novel optimization-based framework for recovering the covaraince matrix of unquantized data from low-resolution multi-bit measurements. Then, MUSIC is applied to an augmented version of the recovered covariance matrix to find the DoAs of interest.
- In Chapter 6, we investigate the problem of passive target localization from one-bit measurements in the context of Internet-of-Things (IoT) applications and propose a novel algorithm for target localization with one-bit measurements.

- Chapter 7 presents a sparsity-based waveform for the problem of distributed target localization with application to automotive radar. The proposed waveform enable us to achieve a high precision localization performance without increasing the required bandwidth.
- Chapter 8 summarizes the thesis and discusses future work in the considered domain.

Appendices

- Appendices A, B and C include the detailed proofs of the Theorems and Lemmas given in Chapters 3, 4 and 6, respectively.

1.3 Publications

The work presented in this thesis has resulted in a number of peer-reviewed journal and conference papers, currently published or under revision. The publications included in this thesis are listed here below.

Journals

- J1: **S. Sedighi**, B. Shankar, B. Ottersten, "An Asymptotically Efficient Weighted Least Squares Estimator for Co-Array-Based DoA Estimation," *IEEE Transactions on Signal Processing*, vol. 68, pp. 589-604, 2020.
- J2: **S. Sedighi**, K. V. Mishra, B. Shankar, B. Ottersten, "Localization with One-Bit Passive Radars in Narrowband Internet-of-Things using Multivariate Polynomial Optimization," *IEEE Transactions on Signal Processing*, vol. 69, pp. 2525-2540, 2021.
- J3: **S. Sedighi**, M. Soltanalian, B. Shankar, B. Ottersten, "DoA Estimation Using Low-Resolution Multi-Bit Sparse Array Measurements," *IEEE Signal Processing Letters*, vol. 28, pp. 1400-1404, 2021.
- J4: **S. Sedighi**, B. Shankar, M. Soltanalian, B. Ottersten, "On the Performance of One-Bit DoA Estimation via Sparse Linear Arrays," *IEEE Transactions on Signal Processing*, accepted, in press.
- J5: **S. Sedighi**, B. Shankar, B. Ottersten, "Performance Analysis of DoA Estimation from One-Bit Sample Covarinace Matrix," *IEEE Signal Processing Letters*, to be submitted.

Conferences

- C1: **S. Sedighi**, B. Shankar, Sina Maleki, B. Ottersten, "Multi-Target localization in asynchronous MIMO radars using sparse sensing," in *Proc. IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Curacao, 2017, pp. 1-5.

- C2: **S. Sedighi**, B. Shankar, Sina Maleki, B. Ottersten, "Consistent Least Squares Estimator for Co-Array-Based DOA Estimation," in *Proc. IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Sheffield, 2018, pp. 524-528.

- C3: **S. Sedighi**, B. Shankar, B. Ottersten, "A Statistically Efficient Estimator for Co-array.Based DoA Estimation," in *Proc. 52nd Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, 2018, pp. 880-883.

- C4: **S. Sedighi**, K. V. Mishra, B. Shankar, B. Ottersten, "Localization Perfomance of 1-Bit Passive Radars in NB-IoT Applications," in *Proc. IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Le gosier, Guadeloupe, 2019, pp. 156-160.

- C5: **S. Sedighi**, B. Shankar, M. Soltanalian, B. Ottersten, "DoA Estimation via Sparse Arrays and One-Bit sampling," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, 2020, pp. 9135-9139.

- C6: **S. Sedighi**, B. Shankar, M. Soltanalian, B. Ottersten, "On the Asymptotic Performance of One-Bit Co-Array-Based MUSIC," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, ON, Canada, 2021, pp. 4635 - 4639.

1.4 Publications not included in this thesis

The following publications, carried out in the context of the PhD work, are not included in this thesis to keep it consistent.

- C7: **S. Sedighi**, K. V. Mishra, B. Shankar, B. Ottersten, "Optimum Design for Sparse FDA-MIMO Automotive Radar," in *Proc. 53rd Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, 2019, pp. 913-918.

- C8: M. Alaee, **S. Sedighi**, B. Shankar, B. Ottersten, "Designing (In)Finite-Alphabet Sequence via Shaping the Radar Ambiguity Function," in *Proc. IEEE*

International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, United Kingdom, 2019, pp. 4295-4299.

- C9: S. Dokhanchi, B. Shankar, Y. Nijsure, T. Stifler, **S. Sedighi**, B. Ottesrten, "Joint automotive radar communications waveform design,"" in *Proc. IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Montreal, QC, 2017, pp. 1-7.

Chapter 2

Background and Contributions

In this chapter, we first conduct a literature review on Direction of Arrival (DoA) estimation and distributed target/source localization with a focus on the most relevant research works to the contributions of this thesis. Then, we highlight the contribution of the thesis compared to the literature. The literature review is divided into two parts. The first part is dedicated to DoA estimation and the distributed localization is considered in the second part.

2.1 DoA Estimation

The problem of Direction of Arrival (DoA) estimation is of central importance in the field of array processing with many applications in radar, sonar, and wireless communications [1–3]. Conventionally, source DoAs are retrieved through processing the output of equally-spaced linear sensor arrays, termed Uniform Linear Arrays (ULAs). Estimating DoAs using ULAs is well investigated in the literature; a number of algorithms Maximum Likelihood (ML) estimation, MUSIC, ESPRIT and subspace fitting have been presented and their performance thoroughly analyzed [44–47]. However, it is widely known that ULAs are not capable of identifying more sources than the number of physical elements in the array [3, 46].

To transcend this limitation, exploitation of Sparse Linear Arrays (SLAs) with particular geometries, such as Minimum Redundancy Arrays (MRAs) [48], co-prime arrays [17], nested arrays [16] and their variants [49–51], has been proposed. These architectures can dramatically boost the degrees of freedom of the array when source signals are uncorrelated such that a significantly larger number of sources than the number of physical elements in the array can be identified. In addition, the enhanced degrees of

freedom provided by these SLAs can improve the resolution performance appreciably compared to ULAs [16]. These features have spurred further research on DoA estimation using SLAs in recent years. A detailed study on DoA estimation via SLAs through an analysis of the Cramér-Rao Bound (CRB) has been conducted in [52]. The findings in [52] provide valuable insights into the performance limits of DoA estimation via SLAs. For example, it has been shown that, when there are more sources than the number of array elements, the CRB does not drop to zero as the Signal-to-Noise Ratio (SNR) goes to infinity. Further, a number of DoA estimators based on the difference co-array of SLAs have been proposed in the literature. In general, existing co-array-based estimators can be classified under two main groups: 1. Sparsity-Based Methods (SBMs); 2. Augmented Covariance-Based Methods (ACBMs).

2.1.1 DOA Estimation Using SLAs

In this section, after providing a quick description of the system model, we will review DoA estimation from SLA measurements using SBMs and ACBMs.

2.1.1.1 System Model

Consider an SLA with M elements located at positions $\left(m_1\frac{\lambda}{2}, m_2\frac{\lambda}{2}, \cdots, m_M\frac{\lambda}{2}\right)$ with $m_i \in \mathbb{M}$. Here $\mathbb{M}$ is a set of integers with cardinality $|\mathbb{M}| = M$, and λ denotes the wavelength of the incoming signals. It is assumed that K narrowband *uncorrelated* signals with distinct DoAs $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_K]^T \in [-\pi/2, \pi/2]^{K\times 1}$ impinge on the SLA from far field. The signal received at the array at time instance t can be modeled as

$$\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}(t) + \mathbf{n}(t) \in \mathbb{C}^{M\times 1}, \quad t = 0, \cdots, N-1, \tag{2.1}$$

where $\mathbf{s}(t) \in \mathbb{C}^{K\times 1}$ denotes the vector of source signals, $\mathbf{n}(t) \in \mathbb{C}^{M\times 1}$ is additive noise, and $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \cdots, \mathbf{a}(\theta_K)] \in \mathbb{C}^{M\times K}$ represents the SLA steering matrix with

$$\mathbf{a}(\theta_k) = [e^{j\pi \sin\theta_k m_1},\ e^{j\pi \sin\theta_k m_2},\ \cdots,\ e^{j\pi \sin\theta_k m_M}]^T, \tag{2.2}$$

being the SLA manifold vector for the i^{th} signal. Further, the following assumptions are made on source signals and noise:

A1 $\mathbf{n}(t)$ follows a zero-mean circular complex Gaussian distribution with the covariance matrix $\mathbb{E}\{\mathbf{n}(t)\mathbf{n}^H(t)\} = \sigma^2\mathbf{I}_M$.

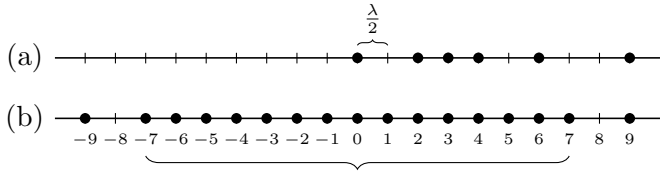


FIGURE 2.1: Array geometry of a co-prime array with $M = 6$ elements: (a) physical array with $\mathbb{M} = \{0, 2, 3, 4, 6, 9\}$; (b) difference co-array with $\mathbb{D} = \{0, 1, 2, 3, 4, 5, 6, 7, 9\}$ and $v = 8$.

A2 The source signal vector is modeled as a zero-mean circular complex Gaussian random vector with covariance matrix $\mathbb{E}\{\mathbf{s}(t)\mathbf{s}^H(t)\} = \text{diag}(\mathbf{p})$ where $\mathbf{p} = [p_1, p_2, \cdots, p_K]^T \in \mathbb{R}_{>0}^{K\times 1}$ (i.e., $p_k > 0,\ \forall k$).

A3 Source and noise vectors are mutually independent.

A4 There is no temporal correlation between the snapshots, i.e., $\mathbb{E}\{\mathbf{n}(t_1)\mathbf{n}^H(t_2)\} = \mathbb{E}\{\mathbf{s}(t_1)\mathbf{s}^H(t_2)\} = \mathbf{0}$ when $t_1 \neq t_2$ and $\mathbf{0}$ is an all-zero matrix of appropriate dimensions.

Based on the above assumptions, the covariance matrix of $\mathbf{y}(t)$ is expressed as

$$\mathbf{R} = \mathbf{A}(\boldsymbol{\theta})\text{diag}(\mathbf{p})\mathbf{A}^H(\boldsymbol{\theta}) + \sigma^2\mathbf{I}_M \in \mathbb{C}^{M\times M}. \tag{2.3}$$

Vectorizing $\overline{\mathbf{R}}$ leads to [38, 52, 53]

$$\mathbf{r} \doteq \text{vec}(\mathbf{R}) = \mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})\mathbf{p} + \sigma^2\mathbf{J}\mathbf{e} \in \mathbb{C}^{M^2\times 1}, \tag{2.4}$$

where $\mathbf{A}_d(\boldsymbol{\theta}) \in \mathbb{C}^{(2D-1)\times K}$ corresponds to the steering matrix of the difference co-array of the SLA whose elements are located at $(-\ell_{D-1}\frac{\lambda}{2}, \cdots, 0, \cdots, \ell_{D-1}\frac{\lambda}{2})$ with $\ell_i \in \mathbb{D} = \{|m_p - m_q| : m_p, m_q \in \mathbb{M}\}$ and $D = |\mathbb{D}|$. Moreover, $\mathbf{e} \in \{0, 1\}^{(2D-1)\times 1}$ is a column vector with $[\mathbf{e}]_i = \delta[i - D]$, and the selection matrix $\mathbf{J} \in \{0, 1\}^{M^2\times(2D-1)}$ is represented as follows [52]:

$$\mathbf{J} = \left[\text{vec}(\mathbf{L}_{D-1}^T),\ \cdots,\ \text{vec}(\mathbf{L}_0),\ \cdots,\ \text{vec}(\mathbf{L}_{D-1})\right], \tag{2.5}$$

where $[\mathbf{L}_n]_{p,q} = \begin{cases} 1, & \text{if } m_p - m_q = \ell_n, \\ 0, & \text{otherwise,} \end{cases}$ with $1 \leq p, q \leq M$ and $0 \leq n \leq D - 1$. The steering matrix of the difference co-array typically includes a contiguous ULA segment around the origin with the size of $2v - 1$ where v is the largest integer such that $\{0, 1, \cdots, v - 1\} \subseteq \mathbb{D}$. An illustrative example of an SLA, the corresponding difference co-array, and its contiguous ULA segment is presented in Fig. 2.1.

2.1.1.2 SBMs

SBMs can be categorized to on-grid, off-grid and gridless methods. On-grid methods discretize the angular domain into a set of grid points $\overline{\boldsymbol{\theta}} = \{\bar{\theta}_1, \bar{\theta}_2, \cdots, \bar{\theta}_L\}$ where $L \gg 2D-1$. This means that the DoAs can only take the values given in $\overline{\boldsymbol{\theta}}$. Hence, the model (2.4) can be equivalently expressed as follows:

$$\mathbf{r} = \mathbf{J}\mathbf{A}_d(\overline{\boldsymbol{\theta}})\widetilde{\mathbf{p}} + \sigma^2\mathbf{J}\mathbf{e} \tag{2.6}$$

where $\mathbf{A}_d(\overline{\boldsymbol{\theta}}) \in \mathbb{C}^{(2D-1)\times L}$ is a known dictionary matrix and $\widetilde{\mathbf{p}}$ is a sparse vector with only K non-zero elements defined as

$$[\widetilde{\mathbf{p}}]_i = \begin{cases} p_k, & \text{if } \bar{\theta}_i = \theta_k, \\ 0, & \text{otherwise}, \end{cases} \tag{2.7}$$

It is observed from (2.7) that the DoAs are encoded in the support of $\widetilde{\mathbf{p}}$. Thus, one can find DoAs by recovering the support of $\widetilde{\mathbf{p}}$. This can be done by solving the following convex optimization problem [15, 35–37, 54]:

$$\underset{\widetilde{\mathbf{p}},\sigma^2}{\text{minimize}} \quad \mu \, \|\widehat{\mathbf{r}} - \mathbf{J}\mathbf{A}_d(\overline{\boldsymbol{\theta}})\widetilde{\mathbf{p}} - \sigma^2\mathbf{J}\mathbf{e}\|_2^2 + \tfrac{1}{2}\|\widetilde{\mathbf{p}}\|_1, \tag{2.8}$$

where $\widehat{\mathbf{r}} = \text{vec}(\widehat{\mathbf{R}})$ with $\widehat{\mathbf{R}} = \frac{1}{N}\sum_{t=0}^{N-1}\mathbf{y}(t)\mathbf{y}^H(t)$ and μ is a regularization parameter. Such estimators are susceptible to grid mismatch leading to significant performance degradation when DoAs do not lie on the predefined grid [55]. To alleviate this problem, off-grid methods approximate the array manifold at a specific DoA θ_k by using a first order Taylor series expansion as follows:

$$\mathbf{a}_d(\theta_k) \simeq \mathbf{a}_d(\bar{\theta}_k) + \dot{\mathbf{a}}_d(\bar{\theta}_k)\Delta\theta_k \tag{2.9}$$

where $\bar{\theta}_k$ is the nearest grid point to θ_k and $\dot{\mathbf{a}}_d(\bar{\theta}_k)$ denotes the derivative of $\mathbf{a}(\theta_k)$ with respect to θ_k, computed at $\bar{\theta}_k$. From (2.9), the model (2.4) reduces to

$$\mathbf{r} = \mathbf{J}\left[\mathbf{A}_d(\overline{\boldsymbol{\theta}}) + \dot{\mathbf{A}}_d(\overline{\boldsymbol{\theta}})\text{diag}(\boldsymbol{\beta})\right]\widetilde{\mathbf{p}} + \sigma^2\mathbf{J}\mathbf{e} \tag{2.10}$$

where $\dot{\mathbf{A}}_d(\overline{\boldsymbol{\theta}}) = [\dot{\mathbf{a}}_d(\bar{\theta}_1), \dot{\mathbf{a}}_d(\bar{\theta}_2), \cdots, \dot{\mathbf{a}}_d(\bar{\theta}_L)]^T$,

$$[\boldsymbol{\beta}]_i = \begin{cases} \Delta\theta_k, & \text{if } \bar{\theta}_i = \theta_k, \\ 0, & \text{otherwise}, \end{cases} \tag{2.11}$$

and $\widetilde{\mathbf{p}}$ is given in (2.7). Considering sparsity of $\widetilde{\mathbf{p}}$ and $\boldsymbol{\beta}$, DoAs and mismatch variables can be jointly estimated as follows [56]:

$$\underset{\widetilde{\mathbf{p}},\boldsymbol{\beta},\sigma^2}{\text{minimize}} \frac{1}{2}\|\widehat{\mathbf{r}} - \mathbf{J}\left[\mathbf{A}_d(\overline{\boldsymbol{\theta}}) + \dot{\mathbf{A}}_d(\overline{\boldsymbol{\theta}})\text{diag}(\boldsymbol{\beta})\right]\widetilde{\mathbf{p}} - \sigma^2\mathbf{J}\mathbf{e}\|_2^2 + \mu_1\ \|\widetilde{\mathbf{p}}\|_1 + \mu_2\ \|\boldsymbol{\beta}\|_1. \quad (2.12)$$

Although the off-grid methods show an improvement over on-grid methods, they are still restricted by higher-order mismatch terms. The shortcoming of discretization approach is overcome by a grid-less sparsity-based algorithm in [57]. This algorithm, named as Sparse and Parametric Approach (SPA), employs the covariance fitting criteria and semi-definite programming. Although SPA does not suffer from grid mismatch, its asymptotic performance (for the large number of snapshots) is not guaranteed to achieve the CRB.

2.1.1.3 ACBMs

ACBMs estimate DoAs by applying conventional subspace methods such as MUSIC and ESPRIT on an Augmented Covariance Matrix (ACM) constructed from $\mathbf{r}$ [16, 38, 58]. Two different ways of constructing the ACM are given in the literature, namely, 1. the direct augmentation approach [59]; 2. spatial smoothing approach [16]. Both approaches first construct v overlapping uniform linear sub-arrays of size v from the contiguous ULA segment of the difference co-array as follows:

$$\mathbf{h}_i = \mathbf{T}_i\mathbf{J}^\dagger\mathbf{r}. \quad (2.13)$$

where $\mathbf{T}_i$ is a selection matrix defined as

$$\mathbf{T}_i = \begin{bmatrix}\mathbf{0}_{v\times(i+D-v-1)} & \mathbf{I}_v & \mathbf{0}_{v\times(D-i)}\end{bmatrix} \in \{0,1\}^{v\times(2D-1)}. \quad (2.14)$$

Then, the ACMs corresponding to the direct augmentation approach and the spatial smoothing approach are respectively given by

$$\mathbf{R}_{v1} = \begin{bmatrix}\mathbf{h}_v & \mathbf{h}_{v-1} & \cdots & \mathbf{h}_1\end{bmatrix} = \mathbf{A}_v(\boldsymbol{\theta})\text{diag}(\mathbf{p})\mathbf{A}_v^H(\boldsymbol{\theta}) + \overline{\sigma}^2\mathbf{I}_v, \in \mathbb{C}^{v\times v}, \quad (2.15)$$

$$\mathbf{R}_{v2} = \frac{1}{v}\sum_{i=1}^{v}\mathbf{h}_i\mathbf{h}_i^H = \frac{1}{v}(\mathbf{A}_v(\boldsymbol{\theta})\text{diag}(\mathbf{p})\mathbf{A}_v^H(\boldsymbol{\theta}) + \overline{\sigma}^2\mathbf{I}_v)^2, \in \mathbb{C}^{v\times v} \quad (2.16)$$

where $\mathbf{A}_v(\boldsymbol{\theta}) = [\mathbf{a}_v(\theta_1), \mathbf{a}_v(\theta_2), \cdots, \mathbf{a}_v(\theta_K)] \in \mathbb{C}^{v\times K}$ denotes the steering matrix of a contiguous ULA with v elements located at $(0, \frac{\lambda}{2}, \cdots, (v-1)\frac{\lambda}{2})$. Considering the structure of $\mathbf{R}_{v1}$ and $\mathbf{R}_{v2}$, DoAs can be estimated by applying subspace algorithms such as MUSIC and ESPRIT to them. We note that, in practice, the true values of $\mathbf{R}_{v1}$ and $\mathbf{R}_{v2}$ are unknown and they should be replaced with their consistent estimates, called

Augmented Sample Covariance Matrix (ASCM), obtained by replacing $\mathbf{R}$ with $\widehat{\mathbf{R}}$ where $\widehat{\mathbf{R}} = \frac{1}{N}\sum_{t=0}^{N-1}\mathbf{y}(t)\mathbf{y}^H(t)$ in (2.4) and (2.13). It was shown that both approaches span identical subspaces resulting in the same estimation performance [38], but the former is computationally efficient. A main drawback of this family of algorithms is the need for prior knowledge on the exact number of sources, which may not be available in practice. On the contrary, SBMs only need information about an upper bound on the number of sources. The performance of such estimators, called Co-Array-Based MUSIC (CAB-MUSIC) and Co-Array-Based ESPRIT (CAB-ESPRIT), has been rigorously analyzed in [38] and [58], respectively. The analyses reveal existence of a considerable gap between their performance and the CRB especially when the number of sources is greater than one.

Further, CAB-MUSIC and CAB-ESPRIT suffer from an inherent performance loss when applied to SLAs with holes in their difference co-arrays, such as co-prime arrays, since they discard the information contained in the non-contiguous segment of the difference co-array. To avoid this performance loss, array interpolation-based algorithms has been used for DoA estimation with co-prime arrays in [60–62]. These algorithms work with a virtual ULA obtained by interpolating extra sensors into the holes of the difference co-array. Firstly, the Toeplitz covariance matrix of the interpolated virtual ULA, denoted by $\mathbf{T}(\mathbf{u})$, is estimated from the co-prime array measurements through solving the following convex optimization problem:

$$\begin{aligned} \underset{\mathbf{u}}{\text{minimize}} \quad & \tfrac{1}{2}\|\widetilde{\mathbf{T}}(\mathbf{u}) - \widehat{\mathbf{R}}_{v1}\|_F^2 + \mu\ \mathrm{tr}(\mathbf{T}(\mathbf{u})) \\ \text{s.t.} \quad & \mathbf{T}(\mathbf{u}) \succeq \mathbf{0}, \end{aligned} \tag{2.17}$$

where $\widetilde{\mathbf{T}}(\mathbf{u})$ is a sub-matrix of $\mathbf{T}(\mathbf{u})$ corresponding to the elements exists in the difference co-array and μ is a regularization parameter. Indeed, the optimization problem (2.17) finds the Toeplitz matrix with the least fitting error to the co-prime array ASCM. Once $\mathbf{T}(\mathbf{u})$ is determined, subspace methods such as mMUSIC and ESPRIT is applied to it to find the DoAs. The performance of interpolation-based algorithms has been analyzed in [63]. It was shown that although the proposed estimators in [60, 61] succeed in dealing with the performance loss due to omission of the non-contiguous segment of difference co-array, they are still unable to achieve the CRB performance.

2.1.2 DOA Estimation from One-Bit Measurements

The aforementioned techniques for DoA estimation in the previous section rest on the assumption that the analog array measurements are digitally represented by a significantly large number of bits per sample such that the resulting quantization errors can be

disregarded. However, the production costs and energy consumption of Analog-to-Digital Converters (ADCs) escalate dramatically as the number of quantization bits and sampling rate increase [21]. In consequence, deployment of high-resolution ADCs in many modern applications, e.g. cognitive radio [22], cognitive radars [23], automotive radars [24], radio astronomy[25] and massive multiple-input multiple-output (MIMO) systems[26], is not economically viable owing to their very high bandwidth. In order to reduce energy consumption and production cost in such applications, researchers and system designers have recently proposed using low-resolution ADCs. As an extreme case of low-resolution ADCs, one-bit ADCs, which convert an analog signal into digital data using a single bit per sample, has received significant attention in the literature. One-bit ADCs offer an extremely high sampling rate at a low cost and very low energy consumption [21]. Additionally, they enjoy the benefits of relatively easy implementation due to their simple architecture [34]. In the past few years, numerous studies were conducted to investigate the impact of using one-bit sampling on various applications such as massive MIMO systems [27, 28, 64–66], dictionary learning [67], radar [29–31, 68–70], and array processing [32, 33].

The problem of DoA estimation from one-bit quantized data has been studied in the literature presuming both the deterministic signal model [71] and the stochastic signal model [46]. The studies in [39–43] presuppose the deterministic signal model. The authors in [39] developed an algorithm for reconstruction of the unquantized array measurements from one-bit samples followed by MUSIC to determine DOAs. The ML estimation was deployed in [40] for finding DoAs from one-bit data. In [43], the authors utilized a sparse Bayesian learning algorithm to solve the DoA estimation problem from one-bit samples. Two sparsity-based approaches were also proposed in [41, 42]. Further, DoA estimation from one-bit data assuming the stochastic signal model has been discussed in [32, 33, 72, 73]. In the special case of a two-sensor array, the exact CRB expression for the DoA estimation problem from one-bit quantized data was derived in [32].Moreover, an approach for estimating DoAs was proposed in [32] which is based on reconstruction of the covariance matrix of unquantized data from one-bit ULA samples [74]. In contrast to the approach employed in [75] which relies on the covaraince matrix reconstruction of unquantized data, the DoA estimation was performed in [73] by directly applying MUSIC on the sample covariance matrix of one-bit ULA data. The numerical simulations demonstrated that the approach proposed in [73] performs similar to the algorithm proposed in [32] in the low Signal-to-Noise Ratio (SNR) regime. An upper bound on the CRB of estimating a single source DoA from one-bit ULA measurements was derived in [33].

The aforementioned research works considered using ULAs for one-bit DoA estimation. Exploitation of SLAs for one-bit DoA estimation has been studied in [75–78]. The authors

in [75] deployed the arcsine law [74] to reconstruct the ASCM from one-bit SLA data. According to the arcsine law, the following relationship exists between the covariance matrix of one-bit data and the normalized covariance matrix unquantized data

$$\overline{\mathbf{R}} = \sin(\frac{\pi}{2}\mathbf{R}_{\mathbf{x}}), \tag{2.18}$$

where $\mathbf{R}_{\mathbf{x}}$ denotes the covariance matrix of one-bit data, $[\text{arcsine}(\overline{\mathbf{R}})]_{m,n} = \arcsin(\Re\{[\overline{\mathbf{R}}]_{m,n}\}) + j\arcsin(\Im\{[\overline{\mathbf{R}}]_{m,n}\})$ and

$$\overline{\mathbf{R}} = \frac{1}{\sigma^2 + \sum_{k=1}^{K} p_k}\mathbf{R}, \tag{2.19}$$

is the normalized covariance matrix of unquantized data. Equation (2.18) implies that a consistent estimate of $\overline{\mathbf{R}}$ is obtained by replacing $\mathbf{R}_{\mathbf{x}}$ with the sample covariance matrix of one-bit data. Then, the ASCM is reconstructed from the derived consistent estimate of $\overline{\mathbf{R}}$ followed by MUSIC to estimate DoAs. It was shown in [75] that the performance degradation due to one-bit quantization can, to some extent, be compensated using SLAs. An array interpolation-based algorithm was employed in [78] to estimate DoAs from one-bit data received by co-prime arrays. Cross-dipoles sparse arrays were deployed in [77] to develop a method for one-bit DoA estimation which is robust against polarization states. In [76], the authors proposed an approach to jointly estimate DoAs and array calibration errors from one-bit data.

Nonetheless, the analytical performance of DoA estimation from one-bit SLA measurements has not yet been studied in the literature and performance analysis in the literature has be limited to simulations studies. Therefore, fundamental performance limitations of DoA estimation form one-bit SLA measurements have not well understood.

2.2 Localization

Localization is one of the fundamental tasks of communications and radar systems and it has received a considerable attention in the last few years [79–81]. In general, localization techniques can be categorized under two main approaches with regard to the sensing nodes architecture. The first approach develops the localization algorithms based on measurements received by a fully-synchronized collocated sensors [82, 83]. Although localization with fully-synchronized collocated sensors is able to achieve high localization performance through providing the waveform diversity [84], they demand a complicated transceiver structure with expensive devices to enable the coherent processing required. However, many commercial applications, such as automotive radar and Internet-of-Things

(IoT), are required to perform localization using cheap devices. This capability is enabled by the second approach based on distributed localization techniques where simple, independent nodes are used instead of a complicated collocated system. Distributed localization techniques exploit the angular (spatial) diversity, provided by spatially separated sensing nodes, to estimate target/source locations [84]. Generally, the distributed localization techniques can be classified into three groups based on the information they use for estimating the source/target location, namely Direction of arrival (DoA) based localization, signal strength (SS) based localization, and Time of Arrival (ToA) based localization [85]. In the following, we briefly review each of these techniques and discuss their advantages and disadvantages.

2.2.1 DoA-based localization

In a DoA-based localization technique, sensing nodes measure the DoA of the signal emitted by the source/target of interest compared to a reference direction. The reference direction is a fixed direction and against that all DoAs are measured. In order to measure the DoA of the received signals, the sensing nodes are required to be equipped with either an antenna array or a directional antenna. Once the DoAs are determined, the DoA measurements are gathered by a Fusion Center (FC) at which the location of the corresponding target/source is estimated by triangulation technique [86–88]. It was shown that, to identify the location of a target/source in a two-dimensional space, at least two DoAs are required [86]. Obviously, the localization accuracy in DoA-based localization technique highly depends on the quality of DoA measurements at sensing nodes. On the other hand, attaining accurate DoA measurements requires an array with a large number of antenna at each sensing node, which will be very costly. This is considered as a main drawback of DoA-based localization technique.

2.2.2 SS-based Localization

In SS-based localization, firstly, the distances between sensing nodes and the signal emitted by the source/target of interest are calculated by measuring the energy of the received signal by exploiting a path loss model. Once the distances between sensing nodes and the signal emitted by the source/target are estimated, they are sent to a FC which determines the source/target location by making use of triangulation technique [89, 90]. The SS-Based technique requires at least three sensing nodes to determine the source/target location in a two-dimensional space [89]. As opposed to DoA-based technique, SS-based localization does not need costly receivers, but they need a knowledge of the channel characteristics. Therefore, SS-based technique is very sensitive to channel

parameter estimates. Further, the accuracy of the SS-Based technique does not improve by increasing the SNR and, moreover, the range information obtained from SS measurements is very coarse compared to that obtained from the other techniques.

2.2.3 ToA-based Localization

In ToA-Based localization, similar to the SS-based technique, the localization is performed by measuring the distances between sensing nodes and the signal emitted by the source/-target of interest. However, here the range information is obtained from calculating travel times of signals between sensing nodes and the source/target of interest [91, 92]. Indeed, if sensing nodes are synchronized with the emitted signal by the source/target, then they are able to determine the ToA of the incoming signal through matched-filtering. In ToA-Based localization, clock synchronization between the sensing nodes and the incoming signal is an important factor which highly influences ToA estimation accuracy. In case the incoming signal and the sensing nodes are asynchronous, but sensing nodes share the same clock, then the Time-Difference-of-Arrival (TDoA) technique can be employed. In this case, the TDoA of the signals received by two sensing nodes is calculated, which determines the location of the node on a hyperbola with foci at the locations of two sensing nodes [93, 94]. Intersecting the hyperbolas obtained from TDoA measurements determines the source/target location. We should note that the accuracy of ToA-based localization techniques is highly affected by the effective signal bandwidth such that a more accurate ToA-based localization demands a higher bandwidth. This could pose a problem in applications where high precision localization is desired. We will address this challenge in Chapter 7 through introducing a sparsity-based waveform.

2.3 Thesis Contributions

The main contributions of this thesis are briefly outlined in the following:

2.3.1 Chapter 3: An Asymptotically Efficient Weighted Least Squares Estimator for Co-Array-Based DoA Estimation

Although the literature presents a variety of estimators in the context of DoA estimation from SLA measurements, none of them are proven to be statistically efficient. This chapter introduces a novel estimator for the co-array-based DoA estimation employing the Weighted Least Squares (WLS) method. An analytical expression for the large sample performance of the proposed estimator is derived. Then, an optimal weighting is obtained

so that the asymptotic performance of the proposed WLS estimator coincides with the Cramér-Rao Bound (CRB), thereby ensuring asymptotic statistical efficiency of resulting WLS estimator. This implies that the proposed WLS estimator has a significantly better performance compared to existing methods. Numerical simulations are provided to validate the analytical derivations and corroborate the improved performance. The content of this chapter appears in the following publications:

- J1: **S. Sedighi**, B. Shankar, B. Ottersten, "An Asymptotically Efficient Weighted Least Squares Estimator for Co-Array-Based DoA Estimation," *IEEE Transactions on Signal Processing*, vol. 68, pp. 589-604, 2020.
- C2: **S. Sedighi**, B. Shankar, Sina Maleki, B. Ottersten, "Consistent Least Squares Estimator for Co-Array-Based DOA Estimation," in *Proc. IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Sheffield, 2018, pp. 524-528.
- C3: **S. Sedighi**, B. Shankar, B. Ottersten, "A Statistically Efficient Estimator for Co-array.Based DoA Estimation," in *Proc. 52nd Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, 2018, pp. 880-883.

2.3.2 Chapter 4: On the Performance of One-Bit DoA Estimation via Sparse Linear Arrays

In this chapter, we study the problem of DoA estimation from one-bit measurements received by an SLA. Specifically, we first investigate the identifiability conditions for the DoA estimation problem from one-bit SLA data and establish an equivalency with the case when DoAs are estimated from infinite-bit unquantized measurements. Towards determining the performance limits of DoA estimation from one-bit quantized data, we derive a pessimistic approximation of the corresponding Cramér-Rao Bound (CRB). This pessimistic CRB is then used as a benchmark for assessing the performance of one-bit DoA estimators. We also propose a new algorithm for estimating DoAs from one-bit quantized data. We investigate the analytical performance of the proposed method through deriving a closed-form expression for the covariance matrix of the asymptotic distribution of the DoA estimation errors and show that it outperforms the existing algorithms in the literature. Numerical simulations are provided to validate the analytical derivations and corroborate the resulting performance improvement. The content of this chapter appears in the following publications:

- J4: **S. Sedighi**, B. Shankar, M. Soltanalian, B. Ottersten, "On the Performance of One-Bit DoA Estimation via Sparse Linear Arrays," *IEEE Transactions on Signal Processing*, Major Revision submitted, 19 July 2021.

- J5: **S. Sedighi**, B. Shankar, B. Ottersten, "Performance Analysis of DoA Estimation from One-Bit Sample Covarinace Matrix," *IEEE Signal Processing Letters*, To be Submitted.

- C6: **S. Sedighi**, B. Shankar, M. Soltanalian, B. Ottersten, "On the Asymptotic Performance of One-Bit Co-Array-Based MUSIC," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, ON, Canada, 2021, pp. 4635 - 4639.

2.3.3 Chapter 5: DoA Estimation Using Low-Resolution Multi-Bit Sparse Array Measurements

In previous works, the problem of Direction of Arrival (DoA) estimation was studied under two extreme scenarios for analog-to-digital conversion, i.e., infinite-bit quantization and one-bit quantization. In this chapter, the problem of DoA estimation from low-resolution multi-bit SLA measurements, e.g. 2 or 4 bit per sample, is studied. In such cases, contrary to the one-bit case, the well known arcsine law cannot be employed to reconstruct the covaraince matrix of unquantized data. Instead, we develop a novel optimization-based framework for recovering the covaraince matrix of unquantized data from low-resolution multi-bit measurements. The MUSIC algorithm is then applied to an augmented version of the recovered covariance matrix to find the DoAs of interest. The simulation results show that increasing the sampling resolution to 2 or 4 bits per samples could significantly increase the DoA estimation performance compared to the one-bit sampling regime while the power consumption and implementation costs is still much lower in comparison to the high-resolution sampling implementation. The content of this chapter appears in the following publications:

- J3: **S. Sedighi**, M. Soltanalian, B. Shankar, B. Ottersten, "DoA Estimation Using Low-Resolution Multi-Bit Sparse Array Measurements," *IEEE Signal Processing Letters*, vol. 28, pp. 1400-1404, 2021.

- C5: **S. Sedighi**, B. Shankar, M. Soltanalian, B. Ottersten, "DoA Estimation via Sparse Arrays and One-Bit sampling," in *Proc. IEEE IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, 2020, pp. 9135-9139.

2.3.4 Chapter 6: Localization with One-Bit Passive Radars in Narrowband Internet-of-Things using Multivariate Polynomial Optimization

Several Internet-of-Things (IoT) applications provide location-based services, wherein it is critical to obtain accurate position estimates by aggregating information from individual sensors. In the recently proposed narrowband IoT (NB-IoT) standard, which trades off bandwidth to gain wide coverage, the location estimation is compounded by the low sampling rate receivers and limited-capacity links. We address both of these NB-IoT drawbacks in the framework of passive sensing devices that receive signals from the target-of-interest. We consider the limiting case where each node receiver employs one-bit analog-to-digital-converters and propose a novel low-complexity nodal delay estimation method using constrained-weighted least squares minimization. To support the low-capacity links to the fusion center (FC), the range estimates obtained at individual sensors are then converted to one-bit data. At the FC, we propose target localization with the aggregated one-bit range vector using both optimal and sub-optimal techniques. The computationally expensive former approach is based on Lasserre's method for multivariate polynomial optimization while the latter employs our less complex iterative joint r*an*ge-*tar*get location *es*timation (ANTARES) algorithm. Our overall one-bit framework not only complements the low NB-IoT bandwidth but also supports the design goal of inexpensive NB-IoT location sensing. Numerical experiments demonstrate feasibility of the proposed one-bit approach with a 0.6% increase in the normalized localization error for the small set of 20-60 nodes over the full-precision case. When the number of nodes is sufficiently large (> 80), the one-bit methods yield the same performance as the full precision. The content of this chapter appears in the following publications:

- J2: **S. Sedighi**, K. V. Mishra, B. Shankar, B. Ottersten, "Localization with One-Bit Passive Radars in Narrowband Internet-of-Things using Multivariate Polynomial Optimization," *IEEE Transactions on Signal Processing*, vol. 69, pp. 2525-2540, 2021.
- C4: **S. Sedighi**, K. V. Mishra, B. Shankar, B. Ottersten, "Localization Perfomance of 1-Bit Passive Radars in NB-IoT Applications," in *Proc. IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Le gosier, Guadeloupe, 2019, pp. 156-160.

2.3.5 Chapter 7: Sparsity-Aided Localization in asynchronous MIMO radar

Target localization, warranted in emerging applications like autonomous driving, requires targets to be perfectly detected in the distributed nodes with accurate range measurements. This implies that high range resolution is crucial in distributed localization in the considered scenario. This chapter proposes a new framework for target localization, addressing the demand for the high range resolution in automotive applications without increasing the required bandwidth. In particular, it employs sparse stepped frequency waveform and infers the target ranges by exploiting sparsity in target scene. The range measurements are then sent to a fusion center where direction of arrival estimation is undertaken. Numerical results illustrate the impact of range resolution on multi-target localization and the performance improvement arising from the proposed algorithm in such scenarios. The content of this chapter appears in:

- C1: **S. Sedighi**, B. Shankar, Sina Maleki, B. Ottersten, "Multi-Target localization in asynchronous MIMO radars using sparse sensing," in *Proc. IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Curacao, 2017, pp. 1-5.

Part I

Direction of Arrival Estimation

Chapter 3

An Asymptotically Efficient Weighted Least Squares Estimator for Co-Array-Based DoA Estimation

3.1 Introduction

The problem of Direction of Arrival (DoA) estimation is of central importance in the field of array processing with many applications in radar, sonar, and wireless communications [1–3]. Estimating DoAs using Uniform Linear Arrays (ULAs) is well-investigated in the literature; a number of algorithms such as the Maximum Likelihood (ML) estimator, MUSIC, ESPRIT and subspace fitting were presented and their performance thoroughly analyzed [44–47, 71]. However, it is widely known that ULAs are not capable of identifying more sources than the number of physical elements in the array [3, 46].

To transcend this limitation, exploitation of Sparse Linear Arrays (SLAs) with particular geometries, such as Minimum Redundancy Arrays (MRAs) [48], co-prime arrays [17], nested arrays [16] and their variants [49–51], has been proposed. These architectures can dramatically boost the degrees of freedom of the array for uncorrelated source signals such that a significantly larger number of sources than the number of physical elements in the array can be identified. In addition, the enhanced degrees of freedom provided by these SLAs can improve the resolution performance appreciably compared to ULAs [16]. These features have spurred further research on DoA estimation using SLAs in recent years. A detailed study on DoA estimation via SLAs through an analysis of the Cramér-Rao

Bound (CRB) has been conducted in [52]. The findings in [52] provide valuable insights into the performance limits of DoA estimation via SLAs and are useful for benchmarking. Further, a number of DoA estimators based on the difference co-array of SLAs have been proposed in the literature. In general, existing co-array-based estimators can be classified under two main groups: 1. Sparsity-Based Methods (SBMs); 2. Augmented Covariance-Based Methods (ACBMs).

SBMs first discretize the angular domain into a grid and then estimate DoAs by imposing sparsity constraints on source profiles and exploiting the compressive sensing recovery techniques [15, 35–37, 54]. Such estimators are susceptible to grid mismatch leading to significant performance degradation when DoAs do not lie on the predefined grid [55]. To alleviate this problem, the authors in [56] include a first order approximation of grid mismatch in the model through Taylor series expansion and then estimate DoAs and mismatch variables jointly. Although the algorithm proposed in [56] shows an improvement over that of conventional sparsity-based methods, it is still restricted by higher-order mismatch terms. The shortcoming of discretization approach is overcome by a grid-less sparsity-based algorithm in [57]. This algorithm, named as Sparse and Parametric Approach (SPA), employs the covariance fitting criteria and semidefinite programming. Although SPA does not suffer from grid mismatch, its asymptotic performance (for the large number of snapshots) is not guaranteed to achieve the CRB.

In the second approach, DoAs are estimated by applying conventional subspace methods such as MUSIC, ESPRIT on an Augmented Sample Covariance Matrix (ASCM) obtained from the original sample covariance matrix by exploiting the difference co-array structure [16, 38, 58]. Two different ways of constructing the ASCM are given in the literature, namely, 1. the direct augmentation approach [59]; 2. spatial smoothing approach [16]. Both approaches span identical subspaces resulting in the same estimation performance, but the former is computationally efficient. A main drawback of this family of algorithms is the need for prior knowledge on the exact number of sources, which may not be available in practice. On the contrary, SBMs need information on an upper bound on the number of sources. The performance of such estimators, called Co-Array-Based MUSIC (CAB-MUSIC) and Co-Array-Based ESPRIT (CAB-ESPRIT), has been rigorously analyzed in [38] and [58], respectively. An existence of a considerable gap between their performance and the CRB is revealed when the number of sources is greater than one. Further, CAB-MUSIC and CAB-ESPRIT suffer from an inherent performance loss when applied to SLAs with holes in their difference co-arrays, such as co-prime arrays, since they discard the information contained in the non-contiguous segment of the difference co-array. To avoid this performance loss, array interpolation-based algorithms has been used for co-prime arrays in [60–62] where a convex optimization problem is formulated to recover the covariance matrix of a virtual ULA interpolated from the co-prime array.

The performance of interpolation-based algorithms has been analyzed in [63]. In addition, the authors of this paper have recently proposed a least squares estimator capable of exploiting the information contained in the non-contiguous ULA segment of the difference co-array [95]. Although the proposed estimators in [60, 61, 95] succeed in dealing with the performance loss due to omission of the non-contiguous segment of difference co-array, they are still unable to achieve the CRB performance.

It is known that the covariance matrix of uncorrelated signals received by a linear array is structured, e.g, Hermitian Toeplitz for ULA [96]. The structure in the covariance matrix is shown to be highly beneficial in obtaining an enhanced covariance matrix estimate compared to the conventional sample covariance matrix [96]. This, in turn, could yield better DoA estimates through an application of MUSIC. While the discussions in [96] are restricted to ULAs, the approach can be straightforwardly used for SLAs to obtain an enhanced covariance matrix estimate and subsequently construct the ASCM. The performance of such an estimator, which we call it Structured CAB-MUSIC (SCAB-MUSIC), has been never investigated in the literature for SLAs. However, our results for this method, reported in the ensuing Section 3.5, reveal that SCAB-MUSIC does not attain the CRB.

The performance gap between the estimators available in the literature and the CRB motivates the current work on designing an asymptotically statistically efficient estimator for co-array-based DoA estimation via SLAs. To close this gap, in this chapter, we propose a Weighted Least Squares (WLS) approach to DoA estimation using SLAs. We analytically prove that the proposed approach can yield an estimator that asymptotically achieves the corresponding CRB for any SLA configuration. As a consequence, the resulting WLS estimator exhibits enhanced performance compared to the existing algorithms in the literature. Accordingly, the contributions of this chapter are described as follows:

- For any given feasible weighting matrix, we formulate the WLS approach towards estimating the DoAs and the ancillary variables − source powers and noise variance.
- We first provide a consistent estimate of the noise variance which is applicable to the difference co-array model. Making use of this consistent noise variance estimate, we derive WLS estimates of the signal powers and concentrate the WLS objective on the DoAs.
- The proposed estimator is the minimizer of the aforementioned concentrated WLS objective. Key attributes of this WLS estimator are studied for any feasible weighting matrix by proving consistency, asymptotic unbiasedness and then deriving a closed-form expression for the asymptotic covariance matrix of DoA estimation errors.

- Considering the earlier asymptotic unbiasedness of the proposed WLS estimator and noting that the covariance matrix of any unbiased estimator is lower bounded by the CRB, the optimal weighting matrix should be the one that renders the resulting covariance matrix of DoA estimation errors and CRB identical. To derive this optimal weighting matrix, we reformulate the CRB expression given in [52] in a form suitable for establishing equality of the CRB and the covariance matrix of DoA estimation errors.

- The new expression for the CRB is exploited to analytically obtain the optimal weighting that results in the equivalence of the asymptotic performance of the proposed WLS estimator and the CRB.

- With the framework to obtain asymptotically efficient WLS estimate provided, we now consider the key aspect of implementing the minimization of the WLS objective that leads to the proposed WLS estimate. This, typically needs computationally complex minimization of a multimodal objective function. The quality of the solutions of the iterative algorithms used for minimizing such multimodal functions highly depends on the initialization such that the global minima potentially achieved in case a very good initial point, which is close enough to the global minima, is available. This motivates us to introduce two efficient algorithms for solving the optimization problem. The first algorithm is applicable to SLAs with hole-free co-arrays, such as MRA and nested arrays. This method recasts the optimization problem as a quadratic optimization problem followed by rooting a polynomial function. This leads to a significant reduction in computations, rendering the complexity of the proposed estimator comparable to that of the other techniques such as CAB-MUSIC, CAB-ESPIRIT and SCAB-MUSIC while the WLS estimator enjoys a better performance compared to them. The second algorithm can be used for SLAs with holes in their co-arrays such as co-prime arrays. This algorithm recasts the optimization problem as a polynomial optimization problem followed by rooting a polynomial function where the global minima of the introduced polynomial optimization problem is guaranteed to be attained by using Lasserre's Semidefinite Programming (SDP) relaxation given in [97].

- Further, we validate the analytical derivations through numerical simulations and compare the performance of the proposed WLS estimator with those proposed in the literature. Numerical results confirm asymptotic efficiency of the proposed WLS estimator and illustrate its superior performance in terms of estimation accuracy and resolution compared to the existing estimators in the literature.

Chapter organization: Section 3.2 describes the co-array system model. In Section 3.3, the proposed WLS framework is presented and the form of the WLS estimates of DoAs

is derived. The performance of the proposed WLS estimator is analytically evaluated and its asymptotic statistical efficiency is proved in Section 3.4. The simulation results and related discussions are included in Section 3.5. Finally, Section 3.6 concludes the chapter.

3.2 Co-Array System Model

We consider an SLA with M elements located at positions $\left(m_1\frac{\lambda}{2},\ m_2\frac{\lambda}{2},\cdots,m_M\frac{\lambda}{2}\right)$ with $m_i \in \mathbb{M}$. Here $\mathbb{M}$ is a set of integers with cardinality $|\mathbb{M}| = M$, and λ represents the wavelength of the incoming signals. It is assumed K narrowband signals with distinct DoAs $\boldsymbol{\theta} = \begin{bmatrix}\theta_1 & \theta_2 & \cdots & \theta_K\end{bmatrix}^T$ impinge on the SLA from far field. Accordingly, the vector of signals received by the SLA at time instance t can be modeled as

$$\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{x}(t) + \mathbf{n}(t) \in \mathbb{C}^{M\times 1}, \qquad t = 1,\cdots,N, \tag{3.1}$$

where $\mathbf{x}(t) \in \mathbb{C}^{K\times 1}$ denotes the vector of source signals, $\mathbf{n}(t) \in \mathbb{C}^{M\times 1}$ is additive noise, and $\mathbf{A}(\boldsymbol{\theta}) = \begin{bmatrix}\mathbf{a}(\theta_1), & \mathbf{a}(\theta_2), & \cdots & \mathbf{a}(\theta_K)\end{bmatrix} \in \mathbb{C}^{M\times K}$ represents the SLA steering matrix where

$$\mathbf{a}(\theta_i) = \begin{bmatrix}e^{j\pi\sin\theta_i m_1} & e^{j\pi\sin\theta_i m_2} & \cdots & e^{j\pi\sin\theta_i m_M}\end{bmatrix}^T, \tag{3.2}$$

is the SLA manifold vector for the i^{th} signal. Further, the following assumptions are made on source signals and noise:

A1 The noise vector follows a zero-mean circular complex Gaussian distribution with the covariance matrix, $\mathbb{E}\{\mathbf{n}(t)\mathbf{n}^H(t)\} = \sigma^2\mathbf{I}_M$.

A2 The source signals are modeled as zero-mean *uncorrelated* circular complex Gaussian random variables with covariance matrix $\mathbb{E}\{\mathbf{s}(t)\mathbf{s}^H(t)\} = \text{diag}(\mathbf{p})$ where $\mathbf{p} = [p_1, p_2, \cdots, p_K]^T \in \mathbb{R}_{>0}^{K\times 1}$ (i.e., $p_k > 0,\ \forall k$).

A3 Source and noise vectors are mutually independent.

A4 There is no temporal correlation between the snapshots, i.e., $\mathbb{E}\{\mathbf{n}(t_1)\mathbf{n}^H(t_2)\} = \mathbb{E}\{\mathbf{x}(t_1)\mathbf{x}^H(t_2)\} = \mathbf{0}$ when $t_1 \neq t_2$ and $\mathbf{0}$ is an all zero matrix of appropriate dimensions.

Based on the above assumptions, the covariance matrix of the received signals, i.e., $\mathbf{R} = \mathbb{E}\{\mathbf{y}(t)\mathbf{y}^H(t)\}$, is given by

$$\mathbf{R} = \mathbf{A}(\boldsymbol{\theta})\text{diag}(\mathbf{p})\mathbf{A}^H(\boldsymbol{\theta}) + \sigma^2\mathbf{I}_M \in \mathbb{C}^{M\times M}. \tag{3.3}$$

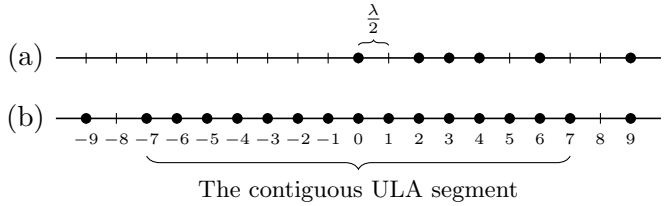


FIGURE 3.1: Array geometry of a co-prime array with $M = 6$ elements: (a) physical array with $\mathbb{M} = \{0, 2, 3, 4, 6, 9\}$; (b) difference co-array with $\mathbb{D} = \{0, 1, 2, 3, 4, 5, 6, 7, 9\}$ and $v = 8$.

Following [16, 38, 52], the difference co-array model of the SLA is obtained by vectorizing the covariance matrix in (3.3), which results in

$$\begin{aligned}\mathbf{r} \doteq \text{vec}(\mathbf{R}) &= (\mathbf{A}^*(\boldsymbol{\theta}) \odot \mathbf{A}(\boldsymbol{\theta}))\, \mathbf{p} + \sigma^2 \text{vec}(\mathbf{I}_M), \\ &= \mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})\mathbf{p} + \sigma^2 \mathbf{J}\mathbf{e} \in \mathbb{C}^{M^2 \times 1}, \end{aligned} \tag{3.4}$$

where $\mathbf{A}_d(\boldsymbol{\theta}) \in \mathbb{C}^{(2D-1)\times K}$ corresponds to the steering matrix of the difference co-array whose elements are located at $(-\ell_{D-1}\frac{\lambda}{2}, \cdots, 0, \cdots, \ell_{D-1}\frac{\lambda}{2})$ with $\ell_i \in \mathbb{D} = \{|m_p - m_q| | m_p, m_q \in \mathbb{M}\}$ and $D = |\mathbb{D}|$. Further, $\mathbf{e} \in \{0, 1\}^{(2D-1)\times 1}$ is a column vector with $[\mathbf{e}]_i = \delta[i - D]$, and the selection matrix $\mathbf{J}$ is represented as follows

Definition 3.1. The binary matrix $\mathbf{J} \in \{0, 1\}^{M^2 \times (2D-1)}$ is defined as [52]

$$\mathbf{J} = \Big[\text{vec}(\mathbf{L}_{D-1}^T) \quad \cdots \quad \text{vec}(\mathbf{L}_0) \quad \cdots \quad \text{vec}(\mathbf{L}_{D-1}),\Big], \tag{3.5}$$

where $[\mathbf{L}_n]_{p,q} = \begin{cases} 1, & \text{if } \; m_p - m_q = \ell_n, \\ 0, & \text{otherwise,} \end{cases}$ with $1 \le p, q \le M$ and $0 \le n \le D - 1$.

The difference co-array model in (3.4) can be perceived to be the response of a virtual array whose steering matrix is given by $\mathbf{A}_d(\boldsymbol{\theta})$ to the parameter vector with signal powers $\mathbf{p}$ in presence of the noise vector $\sigma^2\text{vec}(\mathbf{I}_M)$. This virtual array includes a contiguous ULA segment around the origin with the size of $2v - 1$ where v is the largest integer such that $\{0, 1, \cdots, v - 1\} \subseteq \mathbb{D}$. An illustrative example of an SLA, the corresponding difference co-array and its contiguous ULA segment is provided in Fig. 3.1. It has been shown in [16, 17, 52] that the size of the contiguous ULA segment of the difference co-array plays a crucial role in the number of identifiable sources such that K distinct sources are identifiable if $K \le v - 1$. Hence, in case the SLA is designed properly such that $v > M$, we are able to identify more sources than the number of physical elements in the SLA, exploiting the source signal covariance matrix structure efficiently.

3.3 Co-array-based WLS estimator

The problem under consideration is the estimation of the unknown parameters in (3.4) − DoAs, signal powers and the noise variance − using array observations, i.e., $\{\mathbf{y}(t)\}_{t=1}^{N}$. Of these, the DoAs are of primary interest and the other parameters are of subordinate interest. However, the estimation of the secondary parameters is essential for accurate DoA estimation. In what follows, we first propose a heuristic, but, consistent estimate of the noise variance. We then derive the WLS estimates of DoAs and source signal powers exploiting the proposed consistent estimate of the noise variance.

3.3.1 Estimation of the Noise Variance

Let $\hat{\mathbf{R}}$ denote the sample covariance matrix, defined as

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{t=1}^{N}\mathbf{y}(t)\mathbf{y}^{H}(t) \in \mathbb{C}^{M\times M}, \tag{3.6}$$

and $\hat{\mathbf{r}} = \text{vec}(\hat{\mathbf{R}})$ denote its vectorized form. In addition, let $\hat{\mathbf{R}}_v$ be the augmented sample covariance matrix, which is constructed as follows[59]

$$\hat{\mathbf{R}}_v = \begin{bmatrix}\mathbf{T}_v\mathbf{J}^{\dagger}\hat{\mathbf{r}} & \mathbf{T}_{v-1}\mathbf{J}^{\dagger}\hat{\mathbf{r}} & \cdots & \mathbf{T}_1\mathbf{J}^{\dagger}\hat{\mathbf{r}}\end{bmatrix} \in \mathbb{C}^{v\times v}, \tag{3.7}$$

where $\mathbf{T}_i$ is a selection matrix defined as

$$\mathbf{T}_i = \begin{bmatrix}\mathbf{0}_{v\times(i+D-v-1)} & \mathbf{I}_v & \mathbf{0}_{v\times(D-i)}\end{bmatrix} \in \{0,1\}^{v\times(2D-1)}, \tag{3.8}$$

Then, we are able to obtain a consistent estimate of the noise variance, as stated in the following lemma.

Lemma 3.1. *If $K \leq v-1$, a consistent estimate of the noise variance is given by*

$$\widehat{\sigma^2} = \frac{\text{vec}^{H}(\hat{\mathbf{U}}_n\hat{\mathbf{U}}_n^{H})\mathbf{T}\mathbf{J}^{\dagger}\hat{\mathbf{r}}}{v-K}, \tag{3.9}$$

where $\hat{\mathbf{U}}_n$ represents the eigenvectors of the augmented sample covariance matrix $\hat{\mathbf{R}}_v$ corresponding to its $v-K$ smallest eigenvalues and $\mathbf{T} = \begin{bmatrix}\mathbf{T}_v^T & \mathbf{T}_{v-1}^T & \cdots & \mathbf{T}_1^T\end{bmatrix}^T \in \mathbb{C}^{v^2\times(2D-1)}$.

Proof. See Appendix A.1 □

Remark 3.1 (Efficiency of the Noise Estimate)**.** It can be demonstrated that $\widehat{\sigma^2}$, while being consistent, is not statistically efficient, meaning that $\frac{\mathbb{E}\{(\widehat{\sigma^2}-\sigma^2)^2\}}{CRB(\sigma^2)} > 1$ where

$CRB(\sigma^2)$ represents the CRB of σ^2. We will not dwell on this further since estimation of σ^2 is not the main aim here. A consistent estimate of σ^2 suffices for our purpose.

Remark 3.2. Employing Lemma A.5 in Appendix A.7, it can easily be shown that $\widehat{\sigma^2}^* = \widehat{\sigma^2}$, implying that $\widehat{\sigma^2}$ is real-valued. It is not, however, guaranteed to be positive for a small number of snapshots. Nonetheless, considering the fact that $\widehat{\sigma^2}$ is a consistent estimate of $\sigma^2 > 0$ ensures that $\widehat{\sigma^2}$ is positive when the number of snapshots is adequately large. As a consequence, the asymptotic performance of the DoA estimator, which will be introduced in the next subsection, will not be affected.

3.3.2 WLS Estimates of DoAs

To estimate source DoAs from (3.4), it is possible to formulate the co-array-based LS estimates of $\boldsymbol{\theta}$, $\mathbf{p}$ as

$$\begin{bmatrix}\hat{\boldsymbol{\theta}}_{ls}\\ \hat{\mathbf{p}}_{ls}\end{bmatrix} = \underset{\boldsymbol{\theta},\mathbf{p}}{\operatorname{argmin}} \|\hat{\mathbf{r}} - \mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})\mathbf{p} - \widehat{\sigma^2}\operatorname{vec}(\mathbf{I}_M)\|_2^2. \tag{3.10}$$

However, our investigations, presented in [95], indicate that the LS estimates of DoAs do not show a significant performance improvement in terms of MSE compared to the existing algorithms. Thus it would be useful to introduce a weighting in the above criterion to achieve better performance. Hence, we propose the following WLS estimator instead

$$\begin{bmatrix}\hat{\boldsymbol{\theta}}_{wls}\\ \hat{\mathbf{p}}_{wls}\end{bmatrix} = \underset{\boldsymbol{\theta},\mathbf{p}}{\operatorname{argmin}} \left\| \mathbf{W}^{\frac{1}{2}} \left(\hat{\mathbf{r}} - \mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})\mathbf{p} - \widehat{\sigma^2}\operatorname{vec}(\mathbf{I}_M) \right) \right\|_2^2. \tag{3.11}$$

where $\mathbf{W}$ is a positive definite weighting matrix. The weighting matrix $\mathbf{W}$ should be determined to minimize the MSE of DoA estimates. For the time being, we defer problem of finding the optimal weighting matrix until Section 3.4.2 and proceed with the derivation of the WLS estimator for DoAs.

Inserting (3.9) into (3.11) and performing certain standard algebraic manipulations leads to

$$\begin{bmatrix}\hat{\boldsymbol{\theta}}_{wls}\\ \hat{\mathbf{p}}_{wls}\end{bmatrix} = \underset{\boldsymbol{\theta},\mathbf{p}}{\operatorname{argmin}} \|\mathbf{W}^{\frac{1}{2}}(\hat{\mathbf{Q}}\hat{\mathbf{r}} - \mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})\mathbf{p})\|_2^2, \text{ where} \tag{3.12}$$

$$\hat{\mathbf{Q}} \doteq \mathbf{I}_{M^2} - \frac{\operatorname{vec}(\mathbf{I}_M)\operatorname{vec}^H(\hat{\mathbf{U}}_n\hat{\mathbf{U}}_n^H)\mathbf{T}\mathbf{J}^\dagger}{v-K}. \tag{3.13}$$

Solving (3.12) with respect to $\mathbf{p}$ yields

$$\hat{\mathbf{p}}_{wls} = \left(\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})\right)^{\dagger}\mathbf{W}^{\frac{1}{2}}\hat{\mathbf{Q}}\hat{\mathbf{r}}. \tag{3.14}$$

Remark 3.3 (Consistency of Signal Power Estimates)**.** Based on (A.6) and (A.5) in Appendix A.2, it is readily deduced that $\hat{\mathbf{p}}_{wls}$ is a consistent estimator of $\mathbf{p}$ iff $\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})$ has full column rank. Clearly, $\mathbf{W}$ is positive definite by definition and it was shown in [52] that $\mathbf{J}$ has full column rank. These imply that $\hat{\mathbf{p}}_{wls}$ is consistent iff $\mathbf{A}_d(\boldsymbol{\theta})$ has full column rank. Following the same approach used in the proof of Lemma A.1 in Appendix A.2, it can be shown that $\mathbf{A}_d(\boldsymbol{\theta})$ has full column rank if $K \leq 2v-1$. This condition is weaker than the identifiability condition, i.e., $K \leq v-1$, given in the literature [16, 17, 52]. Hence, the consistency of $\hat{\mathbf{p}}_{wls}$ is guaranteed in practice.

Remark 3.4. Making use of Lemma A.6 in Appendix A.7, it can be shown that $\hat{\mathbf{p}}_{wls}$ is a real number, i.e., $\hat{\mathbf{p}}^{*}_{wls} = \hat{\mathbf{p}}_{wls}$. Hence, the same considerations mentioned in Remarks 3.2 for $\widehat{\sigma^2}$ are applicable to $\hat{\mathbf{p}}_{wls}$ as well.

Finally, inserting (3.14) into (3.12), concentrates the WLS objective on DoAs and the WLS estimator of $\boldsymbol{\theta}$ follows as

$$\hat{\boldsymbol{\theta}}_{wls} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \|\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}}\hat{\mathbf{Q}}\hat{\mathbf{r}}\|_2^2. \tag{3.15}$$

In general, the above problem can be solved iteratively by using either gradient descent or Newton's methods [98]. The gradient and Hessian of the objective function, needed for implementing the aforementioned methods, are given in Appendix A.3. However, finding the global minimum in (3.15) through these algorithms is not guaranteed due to multimodality of the objective function. The quality of the solution is susceptible to the initial point with the global minima potentially achieved in case a very good initial point, which is close enough to the global minima, is available. This motivates us to introduce two efficient algorithms for solving the optimization problem (3.15), which are presented in next. The first algorithm is applicable to SLAs with hole-free co-arrays, such as MRA and nested arrays, while the second one can be used for SLAs with holes in their co-arrays such as co-prime arrays.

Remark 3.5. We should remark that the proposed estimator, like ACBMs, requires uncorrelated sources and the exact knowledge of their number.

3.3.3 WLS Implementation for SLAs With Hole-Free Co-arrays

When the SLA has no holes in its difference co-array, it is possible to recast (3.15) as a quadratic optimization problem followed by rooting a polynomial through reparameterization of the objective function. The main idea is similar to the technique used in [99, 100]. Indeed, for such kind of SLAs, the objective function in (3.15) can be reparameterized in terms of the coefficient of the following polynomial

$$f(z) = \sum_{n=0}^{K} \gamma_{K-n} z^n = \gamma_0 \prod_{k=1}^{K} (z_k - e^{j\pi \sin\theta_k}). \tag{3.16}$$

To show that, let define

$$\mathbf{\Gamma}_{A_d}^H = \begin{bmatrix} \gamma_K & \cdots & \gamma_1 & \gamma_0 & 0 & \cdots & 0 \\ 0 & \gamma_K & \cdots & \gamma_1 & \gamma_0 & \cdots & 0 \\ \vdots & \ddots & \ddots & & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \gamma_K & \cdots & \gamma_1 & \gamma_0 \end{bmatrix} \in \mathbb{C}^{(2D-1-K)\times(2D-1)}, \tag{3.17}$$

and $\mathbf{J}_\perp \in \mathbb{R}^{M^2\times(M^2-2D+1)}$ obtained from QR-factorization of $\mathbf{J}$ as $\mathbf{J} = \begin{bmatrix} \mathbf{J}_\parallel & \mathbf{J}_\perp \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda} \\ \mathbf{0} \end{bmatrix}$. It is readily confirmed that $\begin{bmatrix} \mathbf{J}_\perp & \mathbf{J}^{\dagger H}\mathbf{\Gamma}_{A_d} \end{bmatrix}^H \mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}) = 0$ due to the Vadermonde structure of $\mathbf{A}_d(\boldsymbol{\theta})$ for the aforementioned SLAs. This implies that the columns of $\mathbf{\Gamma} \doteq \begin{bmatrix} \mathbf{J}_\perp & \mathbf{J}^{\dagger H}\mathbf{\Gamma}_{A_d} \end{bmatrix}$ span the null space of $\mathbf{A}_d^H(\boldsymbol{\theta})\mathbf{J}^H$. Hence, considering the fact that $\Pi^\perp_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})} = \Pi_{\mathbf{W}^{-\frac{1}{2}}\mathbf{\Gamma}}$ [101], the objective in (3.15) can be rewritten as

$$\hat{\mathbf{r}}^H\hat{\mathbf{Q}}^H\mathbf{\Gamma}(\mathbf{\Gamma}^H\mathbf{W}^{-1}\mathbf{\Gamma})^{-1}\mathbf{\Gamma}^H\hat{\mathbf{Q}}\hat{\mathbf{r}}. \tag{3.18}$$

Minimization of (3.18) with respect to the free parameters in $\mathbf{\Gamma}$ leads to the estimates of $\gamma_0, \gamma_1, \cdots, \gamma_K$ from which the WLS etimates of DoAs can be obtained through finding the roots of the polynomial (3.16). However, the reparameterized optimization problem is still complicated due to multimodality of the objective function and, moreover, the constraint on $\gamma_0, \gamma_1, \cdots, \gamma_K$ arising from the fact that the roots of the polynomial (3.16) should lie on the unit circle.

The multimodal objective function (3.18) can be relaxed through replacing $(\mathbf{\Gamma}^H\mathbf{W}^{-1}\mathbf{\Gamma})^{-1}$ with its consistent estimate. This relaxation does not affect the asymptotic behavior of the objective function[44, 99, 100, 102], but converts the objective function to a quadratic function with respect to $\gamma_0, \gamma_1, \cdots, \gamma_K$. A consistent estimate of $(\mathbf{\Gamma}^H\mathbf{W}^{-1}\mathbf{\Gamma})^{-1}$ can be obtained in two ways: 1. making use of CAB-MUSIC or CAB-ESPIRIT to derive an initial consistent estimate of $\boldsymbol{\theta}$; 2. through minimizing $\|\mathbf{\Gamma}^H\hat{\mathbf{Q}}\hat{\mathbf{r}}\|^2 = \|\mathbf{\Gamma}_{A_d}^H\mathbf{J}^\dagger\hat{\mathbf{Q}}\hat{\mathbf{r}}\|^2$ with

respect to $\gamma_0, \gamma_1, \cdots, \gamma_K$. Following the same arguments provided in Appendix A.2 and the fact that there is a one-to-one mapping from $\boldsymbol{\theta}$ to $\gamma_0, \gamma_1, \cdots, \gamma_K$ [99], it can easily be shown that minimizing $\|\boldsymbol{\Gamma}_{A_d}^H \mathbf{J}^\dagger \hat{\mathbf{Q}} \hat{\mathbf{r}}\|^2$ with respect to $\gamma_0, \gamma_1, \cdots, \gamma_K$ gives consistent estimates of $\gamma_0, \gamma_1, \cdots, \gamma_K$. Once consistent estimates of $\gamma_0, \gamma_1, \cdots, \gamma_K$ are given, an initial consistent estimates of $\boldsymbol{\theta}$ can be obtained by solving the polynomial equation (3.16).

In addition, we need to ensure that the roots of the polynomial lie on the unit circle. Following the methodology in [99, 100], this is addressed by imposing the conjugate symmetric constraint, i.e., $\gamma_n = \gamma_{K-n}^*$ for $n = 0, 1, \cdots, K$. While this constraint is only necessary, following [99], this relaxation tends to be tight in the asymptotic regimes. Further, to avoid $\gamma_i = 0$, $\forall i$, an additional constraint is required. Herein, we employ the linear constraints, i.e., $\Re\{\gamma_0\} = 1$ or $\Im\{\gamma_0\} = 1$ [99, 100] for simplicity.

Accordingly, the proposed procedure for estimating $\boldsymbol{\theta}$ can be summarized as follows:

1. Compute a consistent estimate of $\boldsymbol{\theta}$ by using CAB-MUSIC, CAB-ESPRIT or minimizing the quadratic function $\|\boldsymbol{\Gamma}_{A_d}^H \mathbf{J}^\dagger \hat{\mathbf{Q}} \hat{\mathbf{r}}\|$.

2. Based on the initial consistent estimate of $\boldsymbol{\theta}$ and the sample covarince matrix, i.e., $\hat{\mathbf{R}}$, calculate a consistent estimate of $(\boldsymbol{\Gamma}^H \mathbf{W}^{-1} \boldsymbol{\Gamma})^{-1}$.

3. Minimize the quadratic criterion

$$\hat{\mathbf{r}}^H \hat{\mathbf{Q}}^H \boldsymbol{\Gamma} (\hat{\boldsymbol{\Gamma}}^H \hat{\mathbf{W}}^{-1} \hat{\boldsymbol{\Gamma}})^{-1} \boldsymbol{\Gamma}^H \hat{\mathbf{Q}} \hat{\mathbf{r}}. \tag{3.19}$$

   with respect to $\gamma_0, \cdots, \gamma_K$ considering the conjugate symmetric constraint, i.e., $\gamma_n = \gamma_{K-n}^*$ for $n = 0, \cdots, K$, besides the linear constraints $\Re\{\gamma_0\} = 1$ or $\Im\{\gamma_0\} = 1$.

4. Obtain the WLS estimate of $\boldsymbol{\theta}$ by rooting $f(z)$ in (3.16).

3.3.4 WLS Implementation for SLAs With Holes in Co-arrays

For these SLAs, it is possible to recast (3.15) as a polynomial optimization problem followed by rooting $f(z)$ given in (3.16). This can be done, similar to Section 3.3.3, through reparameterization of (3.15) in terms of the coefficients of $f(z)$, i.e., $\gamma_0, \gamma_1, \cdots, \gamma_K$, by finding a set of bases spanning the null space of $\mathbf{A}_d^H(\boldsymbol{\theta}) \mathbf{J}^H$. Then, the introduced polynomial optimization problem can be globally solved by using the algorithm proposed in [97].

Let define $\tilde{\mathbf{\Gamma}}_{A_d} = \begin{bmatrix}\tilde{\mathbf{\Gamma}} & \mathbf{\Delta}_1 & \mathbf{\Delta}_2\end{bmatrix} \in \mathbb{C}^{(2D-1)\times(2D-1-K)}$ such that the matrices $\tilde{\mathbf{\Gamma}} \in \mathbb{C}^{(2D-1)\times(2v-1-K)}$, $\mathbf{\Delta}_1 \in \mathbb{C}^{(2D-1)\times(D-v)}$ and $\mathbf{\Delta}_2 \in \mathbb{C}^{(2D-1)\times(D-v)}$ are given by

$$\tilde{\mathbf{\Gamma}}^H = \begin{bmatrix} 0 & \cdots & 0 & \gamma_K & \cdots & \gamma_1 & \gamma_0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \gamma_K & \cdots & \gamma_1 & \gamma_0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \ddots & & \ddots & \ddots & \ddots & \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \gamma_K & \cdots & \gamma_1 & \gamma_0 & 0 & \cdots & 0 \end{bmatrix}, \tag{3.20}$$

$$\mathbf{\Delta}_1^H = \begin{bmatrix} \delta_K^1 & \cdots & \delta_1^1 & \delta_0^1 & 0 & \cdots & & 0 & 0 & \cdots & 0 \\ 0 & \delta_K^2 & \cdots & \delta_1^2 & \delta_0^2 & \cdots & & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & & \ddots & \ddots & & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \delta_K^{D-v} & \cdots & \delta_1^{D-v} & \delta_0^{D-v} & 0 & \cdots & 0 \end{bmatrix}, \tag{3.21}$$

$$\mathbf{\Delta}_2^H = \begin{bmatrix} 0 & \cdots & 0 & \delta_K^1 & \cdots & \delta_1^1 & \delta_0^1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \delta_K^2 & \cdots & \delta_1^2 & \delta_0^2 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \delta_K^{D-v} & \cdots & \delta_1^{D-v} & \delta_0^{D-v} \end{bmatrix}, \tag{3.22}$$

where $\gamma_0, \cdots, \gamma_K$ are the coefficients of $f(z)$ given in (3.16) and $\delta_0^i, \cdots, \delta_K^i$ are the coefficients of the following polynomial

$$q_i(z) = \sum_{n=0}^{K} \delta_{K-n}^i z^{\ell_{a_n^i}} = f(z) g_i(z) \tag{3.23}$$

with $a_n^i = D - 1 - K - i + n$ for $i \in \{1, \cdots, D-v\}$ and $g_i(z) = \frac{\sum_{n=0}^{K} \delta_{K-n}^i z^{\ell_{a_n^i}}}{f(z)}$. From (3.16), it is observed that $f(e^{j\pi \sin\theta_k}) = 0$ for $k \in \{1, \cdots, K\}$, which in turn implies that $q_i(e^{j\pi \sin\theta_k}) = 0$ for $k \in \{1, \cdots, K\}$. Hence, it is easily checked that $\begin{bmatrix}\mathbf{J}_\perp & \mathbf{J}^{\dagger H}\mathbf{\Gamma}_{A_d}\end{bmatrix}^H \mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}) = 0$, indicating that the columns of $\mathbf{\Gamma} \doteq \begin{bmatrix}\mathbf{J}_\perp & \mathbf{J}^{\dagger H}\mathbf{\Gamma}_{A_d}\end{bmatrix}$ span the null space of $\mathbf{A}_d^H(\boldsymbol{\theta})\mathbf{J}^H$. Hence, considering $\Pi^\perp_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})} = \Pi_{\mathbf{W}^{-\frac{1}{2}}\tilde{\mathbf{\Gamma}}}$ [101], the objective in (3.15) can be rewritten as

$$\hat{\mathbf{r}}^H \hat{\mathbf{Q}}^H \tilde{\mathbf{\Gamma}} (\tilde{\mathbf{\Gamma}}^H \mathbf{W}^{-1} \tilde{\mathbf{\Gamma}})^{-1} \tilde{\mathbf{\Gamma}}^H \hat{\mathbf{Q}} \hat{\mathbf{r}}. \tag{3.24}$$

Here the minimization should be done with respect to the free parameters in $\tilde{\mathbf{\Gamma}}$, i.e., $\gamma_0, \cdots, \gamma_K$ and $\delta_0^i, \cdots, \delta_K^i$ $\forall i$ from which the WLS etimates of DoAs can be obtained through finding the roots of the polynomial (3.16). Akin to Section 3.3.3, (3.24) can be relaxed through replacing $(\tilde{\mathbf{\Gamma}}^H \mathbf{W}^{-1} \tilde{\mathbf{\Gamma}})^{-1}$ with its consistent estimate without affecting the asymptotic behavior of the objective function[44, 99, 100, 102]. However, the reparameterized optimization problem is still complicated despite this relaxation due to the existing constraints on $\gamma_0, \cdots, \gamma_K$ and $\delta_0^i, \cdots, \delta_K^i$ $\forall i$. It can be further simplified if the parameters $\delta_0^i, \cdots, \delta_K^i$ $\forall i$ are somehow expressed in terms of $\gamma_0, \cdots, \gamma_K$. This parametrizes

the optimization problem only in terms of the desired parameters, i.e., $\gamma_0, \cdots, \gamma_K$ and thereby eliminates the constraints corresponding to $\delta_0^i, \cdots, \delta_K^i \ \forall i$. Towards this, in what follows, we use the fact that the remainder $h_i(z)$ generated by division of $q_i(z)$ by $f(z)$ is zero according to (3.23).

Let $\tilde{n}$ be an integer such that $\ell_{a_{\tilde{n}}^i} < K$ and $\ell_{a_{\tilde{n}+1}^i} \geq K$. It is then possible to rewrite $q_i(z)$ as $q_i(z) = \sum_{n=0}^{\tilde{n}} \delta_{K-n}^i z^{a_n^i} + \sum_{n=\tilde{n}+1}^{K} \delta_{K-n}^i z^{a_n^i}$. Making use of polynomial long division and after some tedious calculations, the remainder $h_i(z)$ resulting from division of $q_i(z)$ by $f(z)$ takes the form

$$h_i(z) = \sum_{n=0}^{K-1} c_{K-1-n}^i z^n, \tag{3.25}$$

where

$$c_{K-1-n}^i = \begin{cases} \delta_{K-n}^i - \tilde{a}_n(\gamma_0, \cdots, \gamma_K, \delta_{K-\tilde{n}-1}^i, \cdots, \delta_0^i) & \text{if} \quad n \in \mathbb{S} \\ \bar{a}_n(\gamma_0, \cdots, \gamma_K, \delta_{K-\tilde{n}-1}^i, \cdots, \delta_0^i) & \text{if} \quad n \notin \mathbb{S} \end{cases} \tag{3.26}$$

with $\mathbb{S} = \{\ell_{a_0^i}, \cdots, \ell_{a_{\tilde{n}}^i}\}$, $\tilde{a}_n(.)$ and $\bar{a}_n(.)$ being linear functions of $\delta_{K-\tilde{n}-1}^i, \cdots, \delta_0^i$ for $n = 0, \cdots, K-1$ whose coefficients are obtained during long division. $h_i(z)$ is identically zero $\forall i$ if and only if $c_n^i = 0 \ \forall n, i$. Letting c_n^i's equal to zero for each i results in K linear equations with respect to $\delta_K^i, \cdots, \delta_0^i$. Considering the fact that δ_0^i can be chosen arbitrarily, the solution of these K linear equations provides us with the values of $\delta_K^i, \cdots, \delta_1^i$ based on $\gamma_0, \cdots, \gamma_K$ and δ_0^i. Through an appropriate choice of δ_0^i, the parameters $\delta_K^i, \cdots, \delta_1^i$ can be expressed as polynomial functions of $\gamma_0, \cdots, \gamma_K$. Hence, $\tilde{\mathbf{\Gamma}}$ is parameterized only in terms of the coefficients $\gamma_0, \cdots, \gamma_K$.

Consequently, the optimization problem (3.23) is converted to a polynomial optimization with respect to $\gamma_0, \gamma_1, \cdots, \gamma_K$. Similar to Section 3.3.3, conjugate symmetry constraint is imposed on $\{\gamma_i\}$ towards ensuring roots of $f(z)$ on the unit circle. Then, the resulting polynomial optimization can be solved by using the Lasserre's SDP relaxation given in [97]. It is proved in [97] that the Lasserre's SDP relaxation attains the global minima of the polynomial optimization if the order of relaxation is big enough. However, evidently, this solution exhibits higher complexity than the descent algorithms mentioned earlier due to the nature of operations and increased number of variables.

Accordingly, the proposed procedure for estimating $\boldsymbol{\theta}$ can be summarized as follows:

1. Compute a consistent estimate of $\boldsymbol{\theta}$ by using CAB-MUSIC, CAB-ESPRIT.
2. Based on the initial consistent estimate of $\boldsymbol{\theta}$ and the sample covariance matrix, i.e., $\hat{\mathbf{R}}$, calculate a consistent estimate of $(\mathbf{\Gamma}^H \mathbf{W}^{-1} \mathbf{\Gamma})^{-1}$.

3. Compute the coefficients c_n^i's $\forall i, n$, using polynomial long division.

4. Let c_n^i's equal to zero for each i and solve the K resulting linear equations with respect to $\delta_K^i, \cdots, \delta_0^i$. This step gives us the values of $\delta_K^i, \cdots, \delta_1^i$ based on $\gamma_0, \cdots, \gamma_K$ and δ_0^i.

5. Choose the free variable δ_0^i such that δ_0^i, $\delta_K^i, \cdots, \delta_0^i$ are obtained as polynomial functions of $\gamma_0, \cdots, \gamma_K$.

6. Find the minima of the following polynomial criterion

$$\hat{\mathbf{r}}^H \hat{\mathbf{Q}}^H \tilde{\mathbf{\Gamma}} (\hat{\tilde{\mathbf{\Gamma}}}^H \hat{\mathbf{W}}^{-1} \hat{\tilde{\mathbf{\Gamma}}})^{-1} \tilde{\mathbf{\Gamma}}^H \hat{\mathbf{Q}} \hat{\mathbf{r}}. \tag{3.27}$$

   with respect to $\gamma_0, \cdots, \gamma_K$ considering the conjugate symmetric constraint and the linear constraints $\Re\{\gamma_0\} = 1$ or $\Im\{\gamma_0\} = 1$ by using the Lasserre's SDP relaxation.

7. Obtain the WLS estimate of $\boldsymbol{\theta}$ by rooting $f(z)$ in (3.16).

3.4 Asymptotic Performance Analysis

The asymptotic behavior of the proposed WLS estimator for a large number of samples is analyzed in this section and its asymptotic statistical efficiency when an optimal weighting matrix being selected is shown.

3.4.1 Asymptotic Performance

In this subsection, we analyze the key attributes of the proposed WLS estimator including consistency, bias and estimation errors for an arbitrary weighting matrix. Asymptotics are resorted to yield tractable analytical results. We start with proving the consistency of the DoA estimates provided by the proposed WLS estimator.

Theorem 3.1. *$\hat{\boldsymbol{\theta}}_{wls}$ is a consistent estimate of $\boldsymbol{\theta}$ if $K \leq v - 1$.*

Proof. See Appendix A.2 □

Remark 3.6 (Unbiasedness)**.** It readily follows from Theorem 3.1 that $\hat{\boldsymbol{\theta}}_{wls}$ is asymptotically unbiased as well.

Remark 3.7. Note that the sufficient condition for the consistency of $\hat{\boldsymbol{\theta}}_{wls}$, i.e., $K \leq v-1$, given in Theorem 3.1, is equivalent to the sufficient condition for source identifiability given in [16, 17, 52].

The consistency of $\hat{\boldsymbol{\theta}}_{wls}$ can be utilized to obtain the covariance matrix of DoA estimation errors through a Taylor series expansion approach when $N \to \infty$. This is detailed in the following theorem.

Theorem 3.2. *Let $\mathbf{K}_M$ denote the commutation matrix defined according to Definition A.1 in Appendix A.7. If $K \leq v-1$ and $\mathbf{K}_M\mathbf{W} = \mathbf{W}^*\mathbf{K}_M$, the asymptotic ($N \to \infty$) covariance matrix of the WLS estimator $\widehat{\boldsymbol{\theta}}_{wls}$ is given by (3.28),*

$$\begin{aligned}\mathbf{C}_{wls} =& \mathbb{E}\left\{(\hat{\boldsymbol{\theta}}_{wls}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_{wls}-\boldsymbol{\theta})^H\right\} = \frac{1}{\pi^2 N}\mathrm{diag}^{-1}(\mathbf{p})\left(\boldsymbol{\Omega}^H\mathbf{W}^{\frac{1}{2}}\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}}\boldsymbol{\Omega}\right)^{-1} \\ &\times\left(\boldsymbol{\Omega}^H\mathbf{W}^{\frac{1}{2}}\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}}\mathbf{Q}\mathbf{M}^2\mathbf{Q}^H\mathbf{W}^{\frac{1}{2}}\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}}\boldsymbol{\Omega}\right) \\ &\times\left(\boldsymbol{\Omega}^H\mathbf{W}^{\frac{1}{2}}\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}}\boldsymbol{\Omega}\right)^{-1}\mathrm{diag}^{-1}(\mathbf{p}),\end{aligned} \tag{3.28}$$

where

$$\mathbf{Q} \doteq \mathbf{I}_{M^2} - \frac{\mathrm{vec}(\mathbf{I}_M)\mathbf{b}^H}{v-K}, \tag{3.29}$$

$$\mathbf{b} \doteq \mathbf{J}^{\dagger H}\mathbf{T}^H\mathrm{vec}\left(\mathbf{U}_n\mathbf{U}_n^H\right), \tag{3.30}$$

$$\mathbf{M}^2 = \mathbf{R}^T \otimes \mathbf{R}, \tag{3.31}$$

$$\boldsymbol{\Omega} \doteq \mathbf{J}\mathrm{diag}\left(\mathbf{d}\right)\mathbf{A}_d(\boldsymbol{\theta})\boldsymbol{\Phi}(\boldsymbol{\theta}), \tag{3.32}$$

with $\mathbf{d} = \begin{bmatrix}-\ell_{D-1} & \cdots & \ell_0 & \cdots & \ell_{D-1}\end{bmatrix}^T$, $\mathbf{U}_n$ being given in Appendix A.1, and $\boldsymbol{\Phi}(\boldsymbol{\theta}) = \mathrm{diag}(\begin{bmatrix}\cos\theta_1 & \cdots & \cos\theta_K\end{bmatrix}^T)$.

Proof. See Appendix A.3 □

Remark 3.8. In Theorem 3.2, it is assumed that $\mathbf{K}_M\mathbf{W} = \mathbf{W}^*\mathbf{K}_M$ because it simplifies the expression for the covariance matrix of DoA estimation errors. Further, it is shown in Appendix A.5 (See Lemma A.7) that the optimal weighting matrix, to be introduced later in Theorem 3.4, fulfills the constraint. Thus the constraint is not restrictive. However, it is fairly straightforward to also obtain the covariance matrix of $\widehat{\boldsymbol{\theta}}_{wls}$ for the weighting matrices not satisfying the aforementioned constraint by exploiting the derivations given in Appendix A.3.

3.4.2 Optimal Weighting Matrix and Achieving CRB

The results presented in Section 3.4.1 are valid for an arbitrary weighting matrix. However, it is of interest to find an optimal weighting matrix providing the best DoA estimation performance in terms of MSE. In this regard, we resort to the fact that $\mathbf{C}_{wls}$, given in Theorem 3.2, is lower bounded by the CRB. Thus, a good way of finding the optimal

weighting matrix is to seek a weighting matrix rendering $\mathbf{C}_{wls}$ to be identical to the CRB. Accordingly, in what follows, we first reformulate the CRB expression given in [52] in a form suitable for establishing equality of $\mathbf{C}_{wls}$ and the CRB in Theorem 3.3. Then, we show in Theorem 3.4 that there is a weighting matrix enabling $\mathbf{C}_{wls}$ to coincide with the CRB.

Theorem 3.3. *The CRB expression given in [52, Eq. (49)] can be reformulated as the following form*

$$CRB(\boldsymbol{\theta}) = \left(\text{diag}(\mathbf{p})\boldsymbol{\Omega}^H\mathbf{F}\left(\mathbf{F}^H\mathbf{H}\mathbf{F}\right)^{-1}\mathbf{F}^H\boldsymbol{\Omega}\text{diag}(\mathbf{p})\right)^{-1}, \tag{3.33}$$

where $\mathbf{F} \in \mathbb{C}^{M^2\times(M^2-K)}$ *is any matrix whose columns span the null space of* $\mathbf{A}_d^H(\boldsymbol{\theta})\mathbf{J}^H$ *and*

$$\mathbf{H} = \mathbf{Q}\mathbf{M}^2\mathbf{Q}^H + \frac{\mathbf{M}^2\mathbf{b}\mathbf{b}^H\mathbf{M}^2}{\mathbf{b}^H\mathbf{M}^2\mathbf{b}}. \tag{3.34}$$

Proof. See Appendix A.4 □

Theorem 3.4. *If the weighting matrix is selected as follows*

$$\mathbf{W}_{\text{opt}} = \left(\Pi^{\perp}_{\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{S}\Pi^{\perp}_{\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})} + \mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})\mathbf{A}_d^H(\boldsymbol{\theta})\mathbf{J}^T\right)^{-1}, \tag{3.35}$$

where

$$\mathbf{S} = \mathbf{Q}\mathbf{M}^2\mathbf{Q}^H + \mathbf{b}\mathbf{b}^H, \tag{3.36}$$

then we have $\mathbf{C}_{wls} = CRB(\boldsymbol{\theta})$.

Proof. See Appendix A.5 □

Remark 3.9. We note that the optimal weighting matrix given in Theorem 3.4 depends on the true value of the parameters. However, in practice, it can be replaced with a consistent estimate without affecting the asymptotic performance of the WLS estimator[44, 99, 100, 102]. To this end, we can first use any other consistent estimator like CAB-MUSIC or CAB-ESPRIT to obtain an initial estimate of $\boldsymbol{\theta}$. Then, we compute a consistent estimate of the optimal weighting matrix based on the initial estimate of $\boldsymbol{\theta}$ and the sample covariance matrix $\hat{\mathbf{R}}$. Finally, we use the WLS estimator given in (3.15) to derive asymptotically statistically efficient estimates of DoAs. This procedure may be iterated with $\mathbf{W}_{\text{opt}}$ and $\boldsymbol{\theta}$ alternatively estimated. This may enhance the estimation accuracy especially at low snapshots.

We are also required to make sure that the optimal weighting matrix is positive definite, and moreover, its estimate, obtained using the approach explained in Remark 3.9, is always nonsingular regardless of the available number of snapshots. The following Lemma addresses this concern.

Lemma 3.2. $\mathbf{W}_{\text{opt}}$ *is positive definite and its estimate, obtained through either CAB-MUSIC or CAB-ESPRIT, is nonsingular regardless of the available number of snapshots.*

Proof. See Appendix A.6 □

3.5 Simulation Results

In this section, we provide some numerical results to validate the analytical results obtained in Section 3.3 as well as to assess the performance of the proposed estimator. Further, we compare the performance of the WLS estimator proposed in this chapter with that of CAB-MUSIC [16, 38], CAB-ESPRIT [58], SPA [57] and SCAB-MUSIC [96]; and we will show that the WLS estimator yields better performance in terms of resolution, estimation accuracy and statistical efficiency.

3.5.1 General Set-up

In all experiments, each simulated point has been computed by 5000 Monte Carlo repetitions. In addition, it is assumed that the K independent sources are located at $\{-60^\circ + 120^\circ(k-1)/(K-1) | k = 0, 1, \cdots, K-1\}$. All sources have equal powers, i.e., $p_k = p \ \forall k$, and the SNR is defined as $10 \log \frac{p}{\sigma^2}$. Throughout this section, we use three different types of SLAs with $M = 6$ physical elements and the following geometries:

$$\mathbb{M}_{\text{nested}} : \{1, 2, 3, 4, 8, 12\}, \tag{3.37}$$

$$\mathbb{M}_{\text{co-prime}} : \{0, 2, 3, 4, 6, 9\}, \tag{3.38}$$

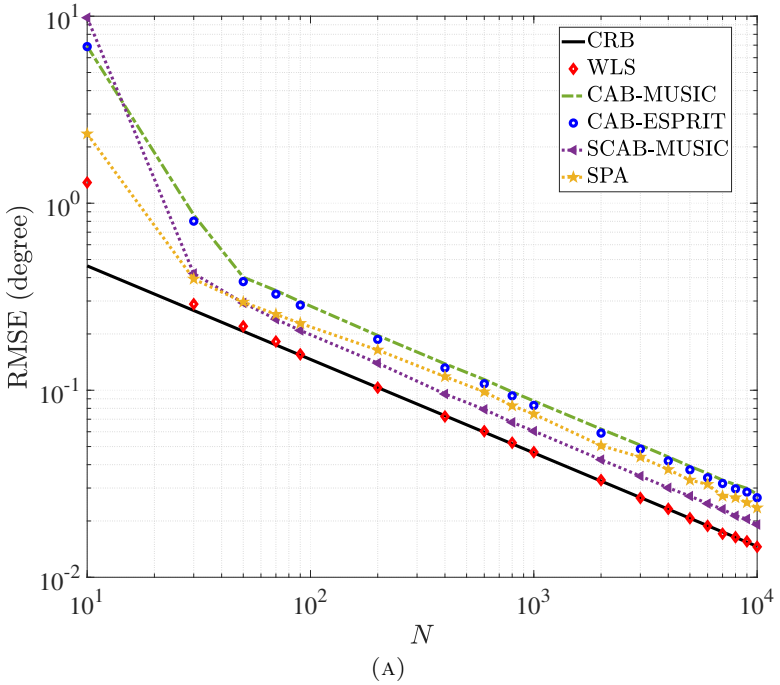
$$\mathbb{M}_{\text{MRA}} : \{0, 1, 6, 9, 11, 13\}. \tag{3.39}$$

These SLAs generate the difference co-arrays as:

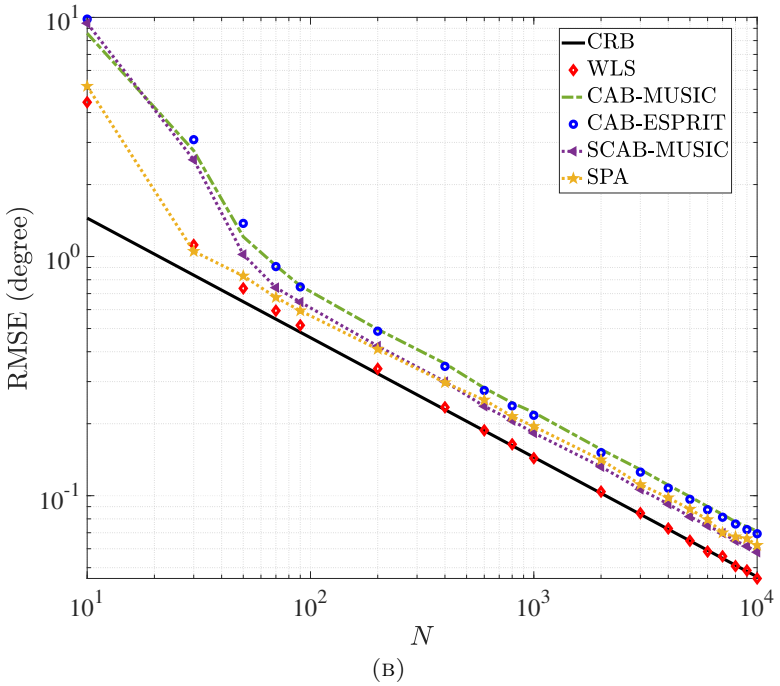
$$\mathbb{D}_{\text{nested}} : \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}, \tag{3.40}$$

$$\mathbb{D}_{\text{co-prime}} : \{0, 1, 2, 3, 4, 5, 6, 7, 9\}, \tag{3.41}$$

$$\mathbb{D}_{\text{MRA}} : \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13\}. \tag{3.42}$$



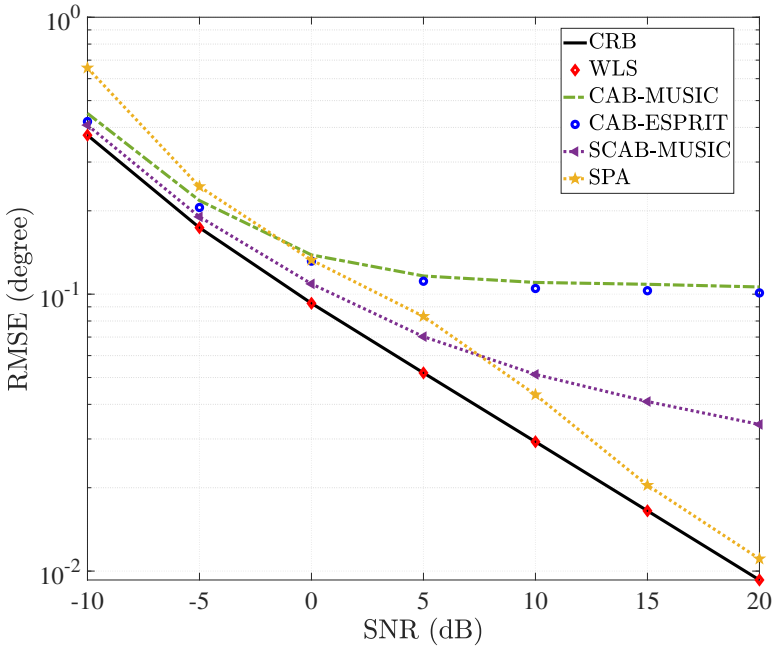
(A)



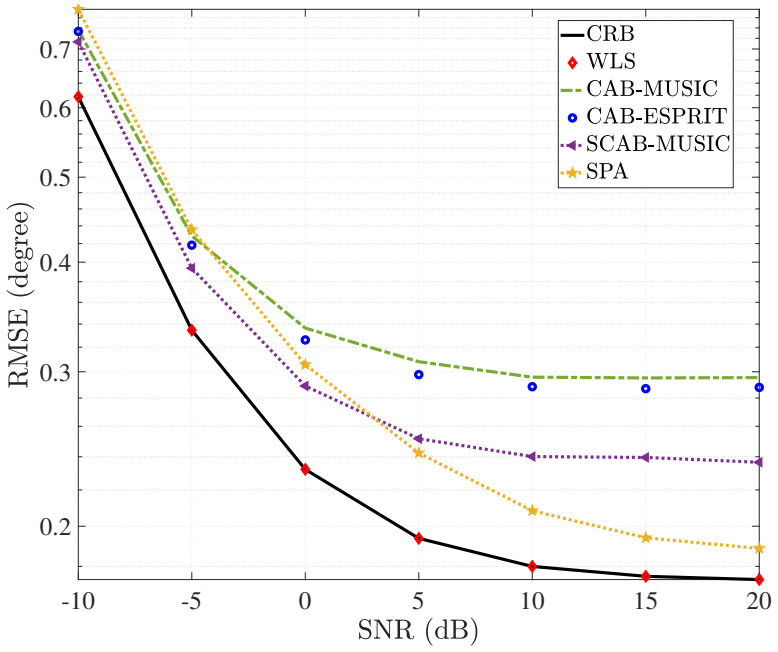
(B)

FIGURE 3.2: RMSE in degree for θ_2 versus the number of snapshots for a nested array with $M = 6$ elements and configuration given in (3.37), SNR = 3 dB, and: (A) $K = 4 < M$; (B) $K = 7 > M$.

The optimization problem (3.15) for MRAs and nested arrays is solved through the algorithm described in Section 3.3.3; and for co-prime arrays it is solved by using the algorithm given in Section 3.3.4. In both cases, CAB-ESPRIT is used to derive a consistent estimate of $(\mathbf{\Gamma}^H \mathbf{W}^{-1} \mathbf{\Gamma})^{-1}$. Further, we take the grid from $-90°$ to $90°$ with step size $0.001°$ to implement CAB-MUSIC and SCAB-MUSIC. Moreover, all estimators



(A)



(B)

FIGURE 3.3: RMSE in degree for θ_2 versus SNR for a nested array with $M = 6$ elements and configuration given in (3.37), $N = 500$, and: (A) $K = 4 < M$; (B) $K = 7 > M$.

but SPA need an exact knowledge of the exact number of sources. Hence, for a fair comparison, SPA is also assumed to know the exact number of sources in all simulations. For this end, it is implemented by applying MUSIC on the augmented covaraince matrix estimate obtained from the SPA algorithm.

3.5.2 MSE vs. the Number of Snapshots

Fig. 3.2 depicts the Root-Mean-Squares-Error (RMSE) for θ_2 in degree versus the number of snapshots for the nested array in (3.37). The SNR is assumed to be 3 dB. In addition, noting $M = 6$, two different scenarios are considered: (A) $K = 4 < M$, and (B) $K = 7 > M$. Fig. 3.2 illustrates a close agreement between the RMSE of the proposed WLS estimator and the CRB when about 70 or more snapshots are available, indicating asymptotic statistical efficiency of the WLS estimator. Further, a considerable gap is observed between the performance of CAB-MUSIC (CAB-ESPRIT) and that of the WLS estimator (the CRB). For instance, at $N = 400$, Figs 3.3a and 3.3b show a performance gain of roughly 2.6 dB and 2 dB, respectively, in terms of the RMSE when the WLS estimator is used. It is also observed that SCAB-MUSIC and SPA outperform CAB-MUSIC CAB-ESPRIT, but their performance is inferior to that of the WLS estimator and they are unable to attain the CRB.

Fig. 3.2 also shows that when a small number of snapshots is available, for example less than 70, all estimators are confronted with substantial performance degradation. Performance loss of the subspace methods, i.e., CAB-MUSIC, CAB-ESPRIT and SCAB-MUSIC, is justified by the subspace swap arising from the inaccurate estimate of the resulting augmented covariance matrix is this case. Further, the underlying reasons for performance degradation of the WLS estimator in such a regime are twofold. Firstly, as mentioned in Remarks 1 and 4, the estimates of σ^2 and $\mathbf{p}$ are not precise and might even yield negative values in this case. Consequently, the value of $\hat{\mathbf{Q}}\hat{\mathbf{r}}$ significantly deviates from its asymptotic value, i.e., $\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})\mathbf{p}$, which, in turn, causes the minimizer of (15) to diverge from the true value of $\boldsymbol{\theta}$. Secondly, there exists a poor estimate of the optimal weighting matrix, i.e., $\mathbf{W}_{\text{opt}}$. This has a detrimental effect on the performance of the WLS estimator. However, it is seen that the proposed WLS estimator still has superior performance compared to the other estimators even in low snapshot paradigm.

3.5.3 MSE vs. SNR

Fig. 3.3 shows the RMSE for θ_2 in degree versus SNR for the same setup used for Fig. 3.2. The number of snapshots is considered to be $N = 500$. It is seen in Figs. 3.3a and Fig. 3.3b that the RMSE of the WLS estimator perfectly matches the CRB for the considered range of SNR as a consequence of its asymptotic statistical efficiency. However, the other estimators are not capable of attaining the CRB.

Fig. 3.3a demonstrates that for $K = 4 < M$ the RMSEs of the WLS estimator and SPA, like the CRB, tend to decay to zero as SNR increases while the RMSEs of CAB-MUSIC,

CAB-ESPRIT and SCAB-MUSIC tend to get saturated at the high SNR regime. For instance, the RMSEs of CAB-MUSIC and CAB-ESPRIT get saturated at SNR around 5 dB. The underlying cause for this saturation behavior of CAB-MUSIC and CAB-ESPRIT was already explained in [20, Corollary 2] and [21], respectively, through an analysis of their MSE expressions. This saturation behaviour of CAB-MUSIC, CAB-ESPRIT and SCAB-MUSIC renders them highly inefficient as SNR increases while the WLS estimator remains statistically efficient for the considered range of SNR.

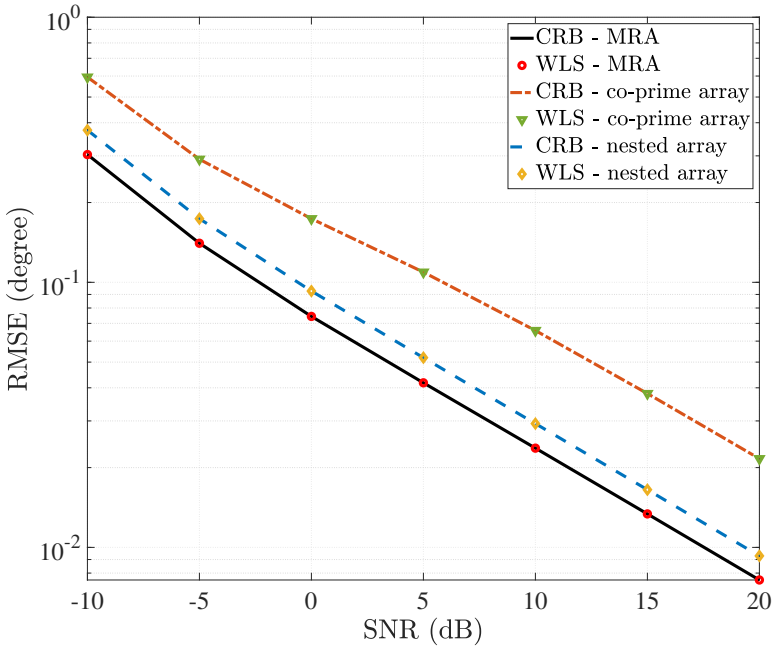
Fig. 3.3b shows that when $K = 7 < M$, the RMSEs of all the estimators as well as the CRB get saturated at the high SNR regime. The saturation point for the WLS estimator, SPA and the CRB is at the SNR around 15 dB while for CAB-MUSIC, CAB-ESPRIT and SCAB-MUSIC it happens at the SNR around 10 dB. Nonetheless, the WLS estimator still pefroms better than all the other estimators under this condition. For example, at SNR $= 15$, the performance gains of about 2.2 dB and 1.3 dB are attained in terms of RMSE compared to CAB-MUSIC (CAB-ESPRIT) and SCAB-MUSIC, respectively.

3.5.4 Impact of Different SLA Configurations

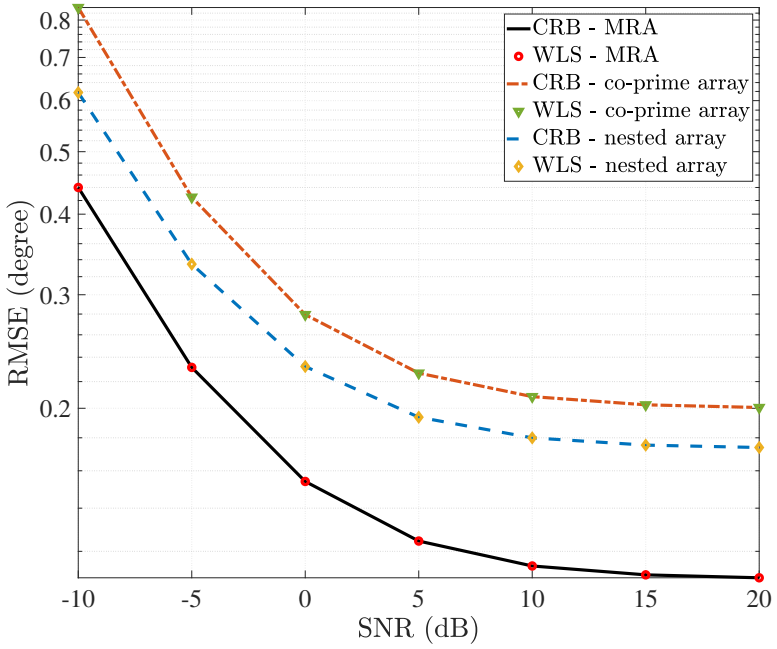
In Fig. 3.4, we plot the RMSE for θ_2 in degree versus SNR for different types of SLAs given in (3.37)-(3.39). The rest of parameters are equal to those in Fig. 3.3. It is readily observed that there is a good agreement between the RMSE of the WLS estimator and the CRB regardless of the array geometry. These simulations corroborate the analytical results where the asymptotic equality of the CRB and the MSE of the WLS estimator is shown considering a generic SLA. Another observation is that amongst these three SLAs, MRA is endowed with the least RMSE followed by the nested and co-prime arrays, respectively. This follows from the distinction between the size of their corresponding difference co-array in comparison to each other. The difference co-arrays for these SLAs are given in (3.40)-(3.42). Indeed, the array with a bigger difference co-array size brings about the lower RMSE.

3.5.5 Resolution Probability

Figs. 3.5 and 3.6 depict the probability of resolution versus SNR for the proposed WLS estimator, CAB-MUSIC, CAB-ESPRIT, SCAB-MUSIC and SPA. The co-prime and nested arrays with the configurations given in (3.38) and (3.37) are considered in Figs. 3.5 and 3.6, respectively. The number of snapshots is considered to be $N = 500$. In addition, we consider two sources with equal powers, located at $\theta_1 = 20^\circ - \frac{\Delta\theta}{2}$ and $\theta_2 = 20^\circ + \frac{\Delta\theta}{2}$ with: (A) $\Delta\theta = 1^\circ$, and (B) $\Delta\theta = 2^\circ$. We define the two sources as being resolvable if



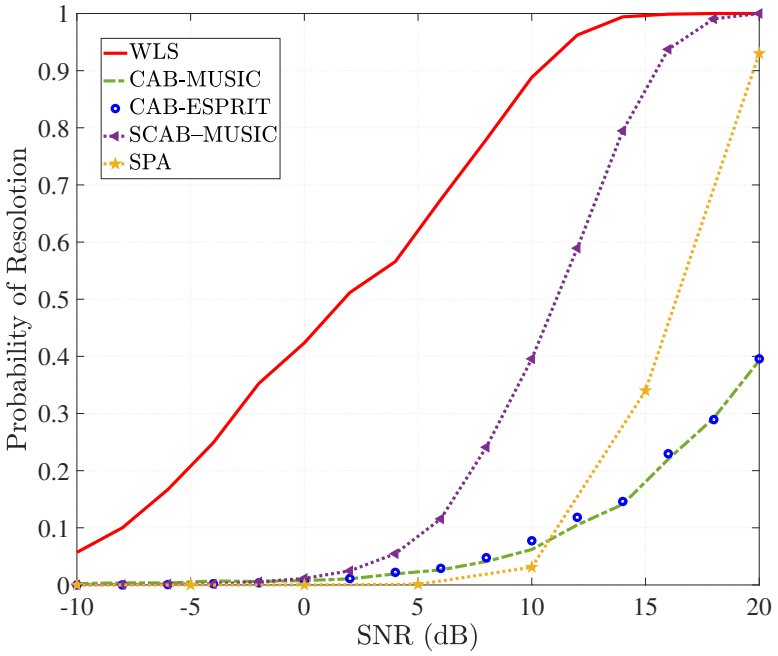
(A)



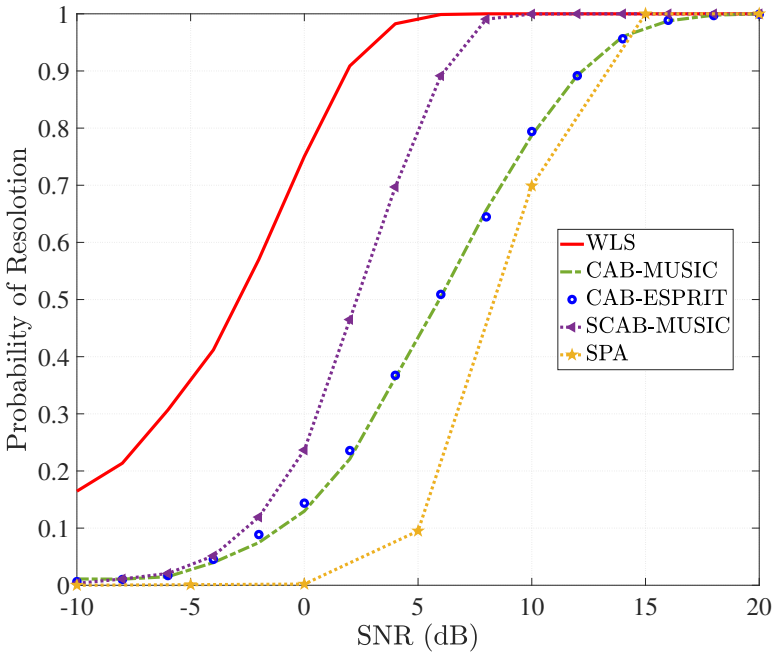
(B)

FIGURE 3.4: RMSE in degree for θ_2 versus SNR for SLAs with $M = 6$ elements and different configurations, $N = 500$, and: (A) $K = 4 < M$; (B) $K = 7 > M$.

$\max_{i\in\{1,2\}} |\hat{\theta}_i - \theta_i| < \frac{\Delta\theta}{2}$ [103]. Figs. 3.5 and 3.6 demonstrate that the WLS estimator has the best resolution performance amongst all the estimators while the SPA resolution performance is inferior to that of all the other ones. Furthermore, CAB-MUSIC and CAB-ESPRIT perform almost equivalently and SCAB-MUIC performs slightly better than them. When $\Delta\theta = 1°$, all the estimators but the WLS are unable to resolve the sources with a probability of 1 even at SNR = 17 dB while the WLS estimator could



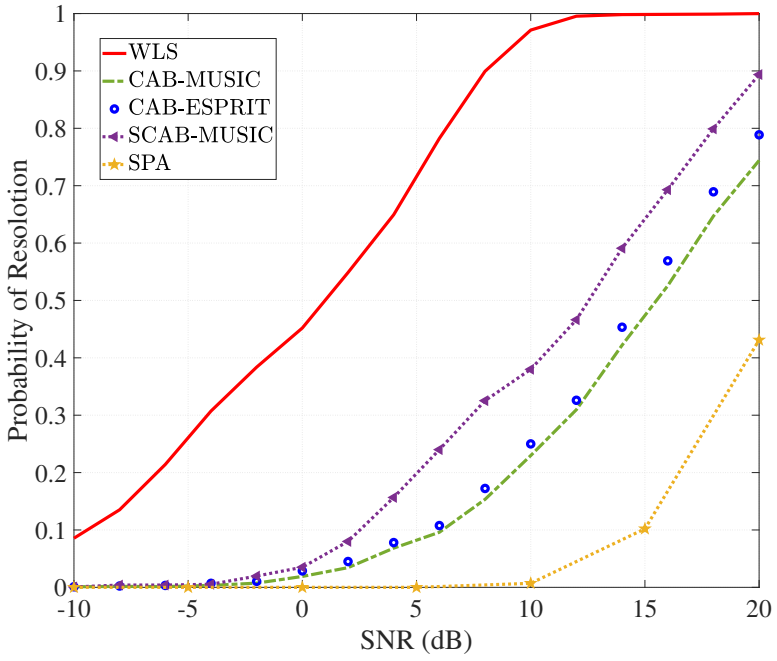
(A)



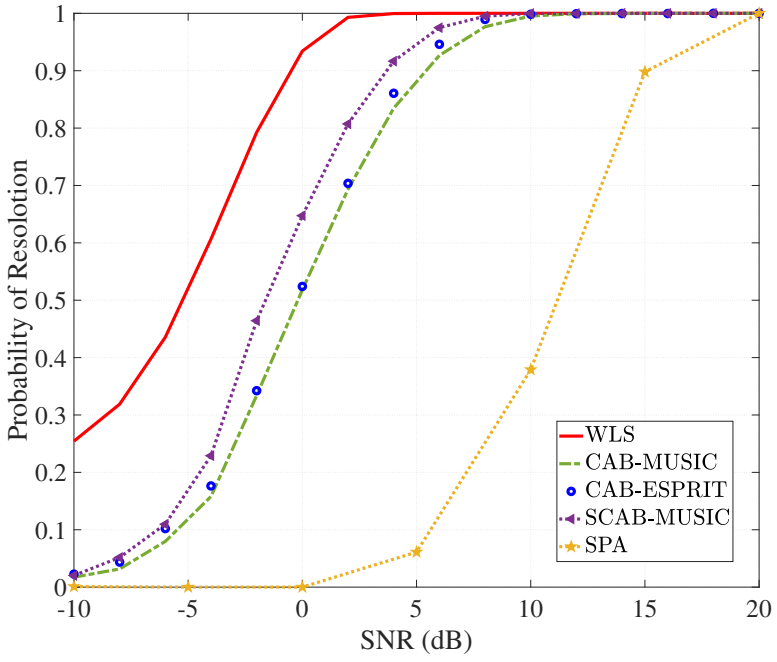
(B)

FIGURE 3.5: Probability of resolution versus SNR for a co-prime array with $M = 6$ elements and configuration given in (3.38), $N = 500$, and: (A) $\Delta\theta = 1°$; (B) $\Delta\theta = 2°$.

achieve a resolution probability of 1 at SNR = 13 dB. In case $\Delta\theta$ is increased to 2°, the WLS achieves SNR gains of 3 dB and 6 dB compared to SCAB-MUSIC, to attain a resolution probability of 1, when co-prime and nested arrays are used respectively. A comparison of Figs. 3.5 and 3.6 indicates that the probability of resolution for the WLS, CAB-MUSIC and CAB-ESPRIT increases when the co-prime array is replaced with the nested array. This can be justified comparing the aperture size of the co-prime and nested



(A)



(B)

FIGURE 3.6: Probability of resolution versus SNR for a nested array with $M = 6$ elements and configuration given in (3.37), $N = 500$, and: (A) $\Delta\theta = 1°$; (B) $\Delta\theta = 2°$.

arrays. Indeed, the nested array enjoys a bigger aperture compared to the co-prime array; thus it is expected to demonstrate a better probability of resolution. However, it is observed that the resolution probabilities of SCAB-MUSIC and SAP decline as the nested array is used instead of the co-prime array. This behavior arises from the structure exhibited by the covarinace matrix in co-prime arrays compared to that of nested arrays. Hence, SCAB-MUSIC and SAP, utilizing a structured estimate of the covariance matrix,

are capable of providing a better estimate of the covariance matrix when the co-prime array is used. This naturally leads to a better covaraince matrix estimate and hence better DoA estimates.

3.6 Conclusion

In this chapter, a novel WLS estimator for the co-array-based DoA estimation via SLAs was proposed and its performance is thoroughly analyzed. It was shown that the proposed WLS estimator provides consistent estimates of DoAs of identifiable sources for any SLAs. Further, an asymptotic closed-form expression for the resulting covariance matrix of DoA estimation errors was derived and it was analytically proved that it asymptotically coincides with the CRB in case the optimal weighting matrix is selected. This implies that the proposed WLS estimator is asymptotically statistically efficient. It thus closes an important gap in the co-array-based DoA estimation. Simulation results demonstrated superior performance of the proposed WLS estimator compared to the existing algorithms in the literature in terms of estimation accuracy and resolution.

Chapter 4

On the Performance of One-Bit DoA Estimation via Sparse Linear Arrays

4.1 Introduction

The problem of Direction of Arrival (DoA) estimation is of central importance in the field of array processing with many applications in radar, sonar, and wireless communications [1–3]. Estimating DoAs using Uniform Linear Arrays (ULAs) is well-investigated in the literature; a number of algorithms such as the Maximum Likelihood (ML) estimator, MUSIC, ESPRIT and subspace fitting were presented and their performance thoroughly analyzed [44–47, 71**?**]. However, it is widely known that ULAs are not capable of identifying more sources than the number of physical elements in the array [3, 46].

To transcend this limitation, exploitation of Sparse Linear Arrays (SLAs) with particular geometries, such as Minimum Redundancy Arrays (MRAs) [48], co-prime arrays [17] and nested arrays [16] has been proposed. These architectures can dramatically boost the degrees of freedom of the array for uncorrelated source signals such that a significantly larger number of sources than the number of physical elements in the array can be identified. In addition, the enhanced degrees of freedom provided by these SLAs can improve the resolution performance appreciably compared to ULAs [16]. These features have spurred further research on DoA estimation using SLAs in recent years. A detailed study on DoA estimation via SLAs through an analysis of the Cramér-Rao Bound (CRB) was conducted in [52]. Further, a number of approaches to estimating DoAs from SLA measurements were proposed in the literature. In general, the proposed approaches can be classified under two main groups: 1. Sparsity-Based Methods (SBMs);

2. Augmented Covariance-Based Methods (ACBMs). SBMs estimate DoAs by imposing sparsity constraints on source profiles and exploiting the compressive sensing recovery techniques [15, 35–37, 55–57]. However, in ACBMs, DoAs are estimated by applying conventional subspace methods such as MUSIC and ESPRIT on an Augmented Sample Covariance Matrix (ASCM) developed from the original sample covariance matrix by exploiting the difference co-array structure [16, 38, 95]. In addition, the authors of this paper recently proposed a Weighted Least Squares (WLS) estimator capable of asymptotically achieving the corresponding CRB for DoA estimation from SLA data [53, 104].

The aforementioned techniques for DoA estimation from SLA data rest on the assumption that the analog array measurements are digitally represented by a significantly large number of bits per sample such that the resulting quantization errors can be disregarded. However, the production costs and energy consumption of Analog-to-Digital Converters (ADCs) escalate dramatically as the number of quantization bits and sampling rate increase [21]. In consequence, deployment of high-resolution ADCs in many modern applications, e.g. cognitive radio [22], cognitive radars [23], automotive radars [24], radio astronomy[25] and massive multiple-input multiple-output (MIMO) systems[26], is not economically viable owing to their very high bandwidth. In order to reduce energy consumption and production cost in such applications, researchers and system designers have recently proposed using low-resolution ADCs. As an extreme case of low-resolution ADCs, one-bit ADCs, which convert an analog signal into digital data using a single bit per sample, has received significant attention in the literature. One-bit ADCs offer an extremely high sampling rate at a low cost and very low energy consumption [21]. Additionally, they enjoy the benefits of relatively easy implementation due to their simple architecture [34]. In the past few years, numerous studies were conducted to investigate the impact of using one-bit sampling on various applications such as massive MIMO systems [27, 28, 64–66], dictionary learning [67], radar [29, 68**?** –70], and array processing [32, 33].

4.1.1 Relevant Works

The problem of DoA estimation from one-bit quantized data has been studied in the literature presuming both the deterministic signal model [71] and the stochastic signal model [46]. The studies in [39–43] presuppose the deterministic signal model. The authors in [39] developed an algorithm for reconstruction of the unquantized array measurements from one-bit samples followed by MUSIC to determine DOAs. The ML estimation was deployed in [40] for finding DoAs from one-bit data. In [43], the authors utilized a sparse Bayesian learning algorithm to solve the DoA estimation problem from one-bit

samples. Two sparsity-based approaches were also proposed in [41, 42]. Further, DoA estimation from one-bit data assuming the stochastic signal model has been discussed in [32, 33, 72, 73]. In the special case of a two-sensor array, the exact CRB expression for the DoA estimation problem from one-bit quantized data was derived in [32]. Moreover, an approach for estimating DoAs from one-bit ULA samples was proposed in [32] which is based on reconstruction of the covariance matrix of unquantized data using the arcsine law [74]. In contrast to the approach employed in [75] which relies on the covaraince matrix reconstruction of unquantized data, the DoA estimation was performed in [73] by directly applying MUSIC on the sample covariance matrix of one-bit ULA data. The numerical simulations demonstrated that the approach proposed in [73] performs similar to the algorithm proposed in [32] in the low Signal-to-Noise Ratio (SNR) regime. An upper bound on the CRB of estimating a single source DoA from one-bit ULA measurements was derived in [33].

The aforementioned research works considered using ULAs for one-bit DoA estimation. Exploitation of SLAs for one-bit DoA estimation has been studied in [75–78]. The authors in [75] deployed the arcsine law [74] to reconstruct the ASCM from one-bit SLA data. Then, they applied MUSIC on the reconstructed ASCM to estimate DoAs. It was shown in [75] that the performance degradation due to one-bit quantization can, to some extent, be compensated using SLAs. An array interpolation-based algorithm was employed in [78] to estimate DoAs from one-bit data received by co-prime arrays. Cross-dipoles sparse arrays were deployed in [77] to develop a method for one-bit DoA estimation which is robust against polarization states. In [76], the authors proposed an approach to jointly estimate DoAs and array calibration errors from one-bit data.

Nonetheless, the analytical performance of DoA estimation from one-bit SLA measurements has not yet been studied in the literature and performance analysis in the literature has been limited to simulations studies. Therefore, fundamental performance limitations of DoA estimation form one-bit SLA measurements have not well understood.

4.1.2 Our Contributions

It is of great importance to analytically investigate the performance of DoA estimation from one-bit SLA measurements. Such a performance analysis not only provides us with valuable insights into the performance of DoA estimation from one-bit SLA data but also enables us to compare its performance with that of DoA estimation using infinite-bit (unquantized) SLA data. Hence, as one of the contributions of this paper, we conduct a rigorous study on the performance of estimating source DoAs from one-bit SLA samples. Furthermore, we propose a new algorithm for estimating source DoAs from one-bit SLA

measurements and analyze its asymptotic performance. Specifically, the contributions of this paper are described as follows:

- **Identifiability Analysis:** We study the identifiability conditions for the DoA estimation from one-bit SLA data. We first show that the identifiability condition for estimating DoAs from one-bit SLA data is equivalent to the case when DoAs are estimated from infinite-bit (unquantized) SLA data. Then, we determine a sufficient condition for global identifiablity of DoAs from one-bit data based on the relationship between the number of source and array elements.

- **CRB Derivation and Analysis:** We derive a pessimistic approximation of the CRB of DoA estimation using one-bit data received by an SLA. This pessimistic CRB approximation provides a benchmark for the performance of DoA estimation algorithms from one-bit data. Additionally, it helps us to spell out the condition under which the Fisher Information Matrix (FIM) of one-bit data is invertible, and thus, the CRB is a valid bound for one-bit DoA estimators. Further, we derive the performance limits of one-bit DoA estimation using SLAs at different conditions.

- **Novel One-bit DoA Estimator:** We propose a new MUSIC-based algorithm for estimating DoAs from one-bit SLA measurements. In this regard, we first construct an enhanced estimate of the normalized covariance matrix of infinite-bit (unquantized) data by exploiting the structure of the normalized covariance matrix efficiently. Then, we apply MUSIC to an augmented version of the enhanced normalized covariance matrix estimate to determine the DoAs.

- **Performance Analysis of the Proposed Estimator:** We derive a closed-form expression for the second-order statistics of the asymptotic distribution (for the large number of snapshots) of the proposed algorithm. Our asymptotic performance analysis shows that the proposed estimator outperform its counterparts in the literature and that its performance is very close to the proposed pessimistic approximation of the CRB. Moreover, the asymptotic performance analysis of the proposed DoA estimator enables us to provide valuable insights on its performance. For examples, we observe that the Mean Square Error (MSE) depends on both the physical array geometry and the co-array geometry. In addition, we observe that the MSE does not drop to zero even if the SNR approaches infinity.

- **Wider Applicability of the derived performance Analysis:** We provide a closed-form expression for the large sample performance of the one-bit DoA estimator in [75] as a byproduct of the performance analysis of our proposed DoA estimator.

Chapter organization: Section 4.2 describes the system model. In Section 4.3, the identifiability condition for DoA estimation problem from one-bit quantized data is discussed. Section 4.4 presents the pessimistic approximation of the CRB and related discussions. In Section 4.5, the proposed algorithm for DoA estimation from one-bit measurements is given and its performance is analyzed. The simulation results and related discussions are included in Section 4.6. Finally, Section 4.7 concludes the chapter.

4.2 System Model

We consider an SLA with M elements located at positions $\left(m_1\frac{\lambda}{2}, m_2\frac{\lambda}{2}, \cdots, m_M\frac{\lambda}{2}\right)$ with $m_i \in \mathbb{M}$. Here $\mathbb{M}$ is a set of integers with cardinality $|\mathbb{M}| = M$, and λ denotes the wavelength of the incoming signals. It is assumed that K narrowband signals with distinct DoAs $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_K]^T \in [-\pi/2, \pi/2]^{K\times 1}$ impinge on the SLA from far field. The signal received at the array at time instance t can be modeled as

$$\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}(t) + \mathbf{n}(t) \in \mathbb{C}^{M\times 1}, \quad t = 0, \cdots, N-1, \tag{4.1}$$

where $\mathbf{s}(t) \in \mathbb{C}^{K\times 1}$ denotes the vector of source signals, $\mathbf{n}(t) \in \mathbb{C}^{M\times 1}$ is additive noise, and $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \cdots, \mathbf{a}(\theta_K)] \in \mathbb{C}^{M\times K}$ represents the SLA steering matrix with

$$\mathbf{a}(\theta_k) = [e^{j\pi \sin\theta_k m_1}, \ e^{j\pi \sin\theta_k m_2}, \ \cdots, \ e^{j\pi \sin\theta_k m_M}]^T, \tag{4.2}$$

being the SLA manifold vector for the i^{th} signal. Further, the following assumptions are made on source signals and noise:

- **A1** $\mathbf{n}(t)$ follows a zero-mean circular complex Gaussian distribution with the covariance matrix $\mathbb{E}\{\mathbf{n}(t)\mathbf{n}^H(t)\} = \sigma^2\mathbf{I}_M$.
- **A2** The source signals are modeled as zero-mean *uncorrelated* circular complex Gaussian random variables with covariance matrix $\mathbb{E}\{\mathbf{s}(t)\mathbf{s}^H(t)\} = \text{diag}(\mathbf{p})$ where $\mathbf{p} = [p_1, p_2, \cdots, p_K]^T \in \mathbb{R}_{>0}^{K\times 1}$ (i.e., $p_k > 0, \ \forall k$).
- **A3** Source and noise vectors are mutually independent.
- **A4** There is no temporal correlation between the snapshots, i.e., $\mathbb{E}\{\mathbf{n}(t_1)\mathbf{n}^H(t_2)\} = \mathbb{E}\{\mathbf{s}(t_1)\mathbf{s}^H(t_2)\} = \mathbf{0}$ when $t_1 \neq t_2$ and $\mathbf{0}$ is an all-zero matrix of appropriate dimensions.

Based on the above assumptions, the covariance matrix of $\mathbf{y}(t)$ is expressed as

$$\mathbf{R} = \mathbf{A}(\boldsymbol{\theta})\text{diag}(\mathbf{p})\mathbf{A}^H(\boldsymbol{\theta}) + \sigma^2\mathbf{I}_M \in \mathbb{C}^{M\times M}. \tag{4.3}$$

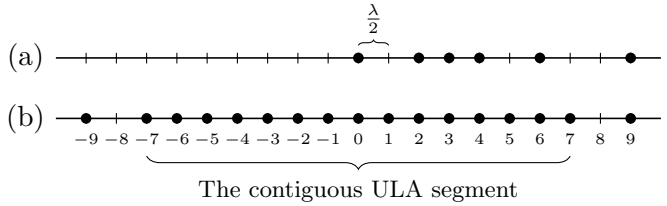


FIGURE 4.1: Array geometry of a co-prime array with $M = 6$ elements: (a) physical array with $\mathbb{M} = \{0, 2, 3, 4, 6, 9\}$; (b) difference co-array with $\mathbb{D} = \{0, 1, 2, 3, 4, 5, 6, 7, 9\}$ and $v = 8$.

Vectorizing $\overline{\mathbf{R}}$ leads to [38, 52, 53]

$$\begin{aligned}\mathbf{r} \doteq \text{vec}(\mathbf{R}) &= (\mathbf{A}^*(\boldsymbol{\theta}) \odot \mathbf{A}(\boldsymbol{\theta}))\, \mathbf{p} + \sigma^2 \text{vec}(\mathbf{I}_M), \\ &= \mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})\mathbf{p} + \sigma^2 \mathbf{J}\mathbf{e} \in \mathbb{C}^{M^2 \times 1},\end{aligned} \tag{4.4}$$

where $\mathbf{A}_d(\boldsymbol{\theta}) \in \mathbb{C}^{(2D-1)\times K}$ corresponds to the steering matrix of the difference co-array of the SLA whose elements are located at $(-\ell_{D-1}\frac{\lambda}{2}, \cdots, 0, \cdots, \ell_{D-1}\frac{\lambda}{2})$ with $\ell_i \in \mathbb{D} = \{|m_p - m_q| : m_p, m_q \in \mathbb{M}\}$ and $D = |\mathbb{D}|$. Moreover, $\mathbf{e} \in \{0, 1\}^{(2D-1)\times 1}$ is a column vector with $[\mathbf{e}]_i = \delta[i - D]$, and the selection matrix $\mathbf{J} \in \{0, 1\}^{M^2 \times (2D-1)}$ is represented as follows [52]:

$$\mathbf{J} = \Big[\text{vec}(\mathbf{L}_{D-1}^T),\ \cdots,\ \text{vec}(\mathbf{L}_0),\ \cdots,\ \text{vec}(\mathbf{L}_{D-1})\Big], \tag{4.5}$$

where $[\mathbf{L}_n]_{p,q} = \begin{cases} 1, & \text{if } m_p - m_q = \ell_n, \\ 0, & \text{otherwise,} \end{cases}$ with $1 \le p, q \le M$ and $0 \le n \le D-1$. The steering matrix of the difference co-array includes a contiguous ULA segment around the origin with the size of $2v-1$ where v is the largest integer such that $\{0, 1, \cdots, v-1\} \subseteq \mathbb{D}$. The size of the contiguous ULA segment of the difference co-array plays a crucial role in the number of identifiable sources such that K distinct sources are identifiable if $K \le v - 1$. Hence, in case the SLA is designed properly such that $v > M$, we are able to identify more sources than the number of physical elements in the SLA; exploiting the resulting structure of $\mathbf{R}$ efficiently[16, 17, 52, 53]. An illustrative example of an SLA, the corresponding difference co-array, and its contiguous ULA segment is presented in Fig. 4.1.

Here it is assumed that each array element is connected to a one-bit ADC which directly converts the received analog signal into binary data by comparing the real and imaginary parts of the received signal individually with zero. In such a case, the one-bit measurements at the m^{th} array element are given by

$$[\mathbf{x}(t)]_m = \frac{1}{\sqrt{2}}\text{sgn}\left(\Re\{[\mathbf{y}(t)]_m\}\right) + \frac{j}{\sqrt{2}}\ \text{sgn}\left(\Im\{[\mathbf{y}(t)]_m\}\right). \tag{4.6}$$

The problem under consideration is the estimation of source DoAs, i.e., $\boldsymbol{\theta}$, from one-bit quantized measurements, i.e., $\mathbf{X} = \Big[\mathbf{x}(0), \quad \mathbf{x}(1), \quad \cdots, \quad \mathbf{x}(N-1)\Big]$, collected by the SLA.

4.3 Identifiability Conditions

Note that there is a significant information loss expected when gping from infinite-bit (unquantized) data, i.e., $\mathbf{Y} = [\mathbf{y}(0), \mathbf{y}(1), \cdots, \mathbf{y}(M)]$, to one-bit data, i.e., $\mathbf{X}$. This information loss may affect the attractive capability of SLAs to identify a larger number of uncorrelated sources than the number of array elements. To address this concern, we will consider the identifiability conditions for DoA estimation from one-bit SLA measurements in this section. Before proceeding further, we first need to give a clear definition of indetifiability for this problem.

Definition 4.1 (Identifiability)**.** Let $f(\mathbf{X} \mid \boldsymbol{\theta}, \mathbf{p}, \sigma^2)$ denote the Probability Density Function (PDF) of $\mathbf{X}$ parameterized by $\boldsymbol{\theta}$, $\mathbf{p}$ and σ^2. Then, the source DoAs are said to be identifiable from $\mathbf{X}$ at point $\boldsymbol{\theta}_0 \in [-\pi/2, \pi/2]^{K\times 1}$ if there exist no $\breve{\boldsymbol{\theta}} \neq \boldsymbol{\theta}_0 \in [-\pi/2, \pi/2]^{K\times 1}$ such that $f(\mathbf{X} \mid \boldsymbol{\theta}_0, \mathbf{p}, \sigma^2) = f(\mathbf{X} \mid \breve{\boldsymbol{\theta}}, \breve{\mathbf{p}}, \breve{\sigma}^2)$ for any arbitrary values of $\mathbf{p} \in \mathbb{R}_{>0}^{K\times 1}$, $\breve{\mathbf{p}} \in \mathbb{R}_{>0}^{K\times 1}$, σ^2 and $\breve{\sigma}^2$ [105, Ch. 1, Definition 5.2] [106, pp. 62].

Remark 4.1. The above definition can be used for identifiabilty of $\boldsymbol{\theta}_0$ from $\mathbf{Y}$ by replacing $f(\mathbf{X}\,|\,\boldsymbol{\theta}, \mathbf{p}, \sigma^2)$ with $f(\mathbf{Y}\,|\,\boldsymbol{\theta}, \mathbf{p}, \sigma^2)$.

Based on the above definition, the necessary and sufficient condition for a particular DoA point to be identifiable from one-bit SLA data is given in the following Theorem.

Theorem 4.1. *The source DoAs are identifiable from* $\mathbf{X}$ *at* $\boldsymbol{\theta}_0 \in [-\pi/2, \pi/2]^{K\times 1}$ *if and only if they are identifiable from* $\mathbf{Y}$ *at* $\boldsymbol{\theta}_0$*.*

Proof. See Appendix B.1. □

The above Theorem shows that the identifiability condition for the DoA estimation problem from one-bit SLA measurements is equivalent to that for the DoA estimation problem from infinite-bit (unquantized) SLA measurements. Hence, the information loss arises from one-bit quantization does not influence the number of identifiable sources. However, Theorem 4.1 simply spells out the identifiability condition of a single DoA point. A sufficient condition for global identifiablity of source DoAs from one-bit data is given in the following theorem.

Definition 4.2 (Global identifiability)**.** The source DoAs are said to be globally identifiable from $\mathbf{X}$ if there exists no distinct $\boldsymbol{\theta} \in [-\pi/2, \pi/2]^{K\times 1}$ and $\breve{\boldsymbol{\theta}} \in [-\pi/2, \pi/2]^{K\times 1}$ such that $f(\mathbf{X} \mid \boldsymbol{\theta}, \mathbf{p}, \sigma^2) = f(\mathbf{X} \mid \breve{\boldsymbol{\theta}}, \breve{\mathbf{p}}, \breve{\sigma}^2)$ for any arbitrary values of $\mathbf{p} \in \mathbb{R}_{>0}^{K\times 1}$, $\mathbf{p} \in \mathbb{R}_{>0}^{K\times 1}$, σ^2 and σ^2.

Theorem 4.2. *The sufficient conditions for global indentifiability and global non-indentifiability of source DoAs from one-bit SLA data are given as follows:*

- **S1** *The source DoAs are globally identifiable (with probability one) from* $\mathbf{X}$ *for any value of* $\boldsymbol{\theta} \in [-\pi/2, \pi/2]^{K\times 1}$ *if* $K \leq v - 1$.
- **S2** *The source DoAs are globally unidentifiable from* $\mathbf{X}$ *for any value of* $\boldsymbol{\theta} \in [-\pi/2, \pi/2]^{K\times 1}$ *if* $K \geq D$.

Proof. See Appendix B.2. □

Having revealed that one-bit quantization does not affect the indentifiability conditions of source DoAs, we will investigate the performance of DoA estimation from one-bit SLA data through a CRB analysis in the next section.

4.4 Cramér-Rao Bound Analysis

It is well-known that the CRB offers a lower bound on the covariance of any unbiased estimator [107]. Hence, it is considered as a standard metric for evaluating the performance of estimators. In particular, the CRB can provide valuable insights into the fundamental limits of estimation for specific problems as well as the dependence of the estimation performance on various system parameters. Deriving a closed-form expression for the CRB requires knowledge of the data distribution. However, the data distribution may not be known for some problems. In such cases, the Gaussian assumption is a natural choice which leads to the largest CRB in a general class of data distributions [108].

In the problem of DoA estimation from one-bit SLA measurements, the true PDF of one-bit data is obtained from the orthant probabilities [109] of Gaussian distribution, for which a closed-form expression is not available in general. Motivated by this fact, in what follows, we derive a pessimistic closed-form approximation for the CRB of the DoA estimation problem from one-bit SLA data through considering a Gaussian distribution for $\mathbf{x}(t)$. This pessimistic closed-form approximation is used for benchmarking the performance of one-bit DoA estimators as well as for investigating the performance limits of the DoA estimation problem from one-bit data. Making use of assumptions **A1-A4**,

it is readily confirmed that $\mathbb{E}\{\mathbf{x}(t)\} = \mathbf{0}$. Further, the arcsine law [74] establishes the following relationship between $\mathbf{R}$ and $\mathbf{R}_\mathbf{x}$:

$$\mathbf{R}_\mathbf{x} = \mathbb{E}\{\mathbf{x}(t)\mathbf{x}^H(t)\} = \frac{2}{\pi}\text{arcsine}(\overline{\mathbf{R}}), \tag{4.7}$$

where $[\text{arcsine}(\overline{\mathbf{R}})]_{m,n} = \arcsin(\Re\{[\overline{\mathbf{R}}]_{m,n}\}) + j\arcsin(\Im\{[\overline{\mathbf{R}}]_{m,n}\})$ and

$$\begin{aligned}\overline{\mathbf{R}} &= \frac{1}{\sigma^2 + \sum_{k=1}^K p_k}\mathbf{R} \\ &= \mathbf{A}(\boldsymbol{\theta})\text{diag}(\overline{\mathbf{p}})\mathbf{A}^H(\boldsymbol{\theta}) + (1 - \sum_{k=1}^K \overline{p}_k)\mathbf{I}_M,\end{aligned} \tag{4.8}$$

is the normalized covariance matrix of $\mathbf{y}(t)$ with $\overline{\mathbf{p}} = [\overline{p}_1, \overline{p}_2, \cdots, \overline{p}_K]^T$ and $\overline{p}_k = \frac{p_k}{\sigma^2 + \sum_{k=1}^K p_k}$. It follows from (4.7) and (4.8) that $\mathbf{R}_\mathbf{x}$ is a function of the parameters $\boldsymbol{\theta}$ and $\overline{\mathbf{p}}$. Let $\boldsymbol{\varrho} = [\boldsymbol{\theta}, \overline{\mathbf{p}}]^T$ denote the vector of unknown parameters. Then, considering the Gaussian assumption, the worst-case Fisher Information Matrix (FIM) $\mathcal{I}_w(\boldsymbol{\varrho})$ is given by [107]

$$\begin{aligned}[\mathcal{I}_w(\boldsymbol{\varrho})]_{m,n} &= N\text{tr}(\mathbf{R}_\mathbf{x}^{-1}\frac{\partial \mathbf{R}_\mathbf{x}}{\partial [\boldsymbol{\varrho}]_m}\mathbf{R}_\mathbf{x}^{-1}\frac{\partial \mathbf{R}_\mathbf{x}}{\partial [\boldsymbol{\varrho}]_n}) \\ &= N\frac{\partial \mathbf{r}_\mathbf{x}^H}{\partial [\boldsymbol{\varrho}]_m}(\mathbf{R}_\mathbf{x}^{-T} \otimes \mathbf{R}_\mathbf{x}^{-1})\frac{\partial \mathbf{r}_\mathbf{x}}{\partial [\boldsymbol{\varrho}]_n},\end{aligned} \tag{4.9}$$

where $\mathbf{r}_\mathbf{x} = \text{vec}(\mathbf{R}_\mathbf{x})$ and the last equality is obtained by using the relation $\text{tr}(\mathbf{C}_1\mathbf{C}_2\mathbf{C}_3\mathbf{C}_4) = \text{vec}^H(\mathbf{C}_2^H)(\mathbf{C}_1^T \otimes \mathbf{C}_3)\text{vec}(\mathbf{C}_4^H)$. From (4.4), (4.7) and (4.9), we obtain

$$\begin{aligned}\mathbf{r}_\mathbf{x} &= \frac{2}{\pi}\text{arcsine}(\text{vec}(\overline{\mathbf{R}})) \\ &= \frac{2}{\pi}\mathbf{J}\text{arcsine}\left(\mathbf{A}_d(\boldsymbol{\theta})\overline{\mathbf{p}} + (1 - \sum_{k=1}^K \overline{p}_k)\mathbf{e}\right).\end{aligned} \tag{4.10}$$

Computing the derivative of $\mathbf{r}_\mathbf{x}$ with respect to θ_k and $\overline{p}_k$ yields

$$\frac{\partial \mathbf{r}_\mathbf{x}}{\partial \theta_k} = j\pi\cos(\theta_k)\overline{p}_k\mathbf{J}\text{diag}(\mathbf{d})\left[\text{diag}(\overline{\mathbf{h}})\Re\{\mathbf{a}_d(\theta_k)\} + j\text{diag}(\mathbf{h})\Im\{\mathbf{a}_d(\theta_k)\}\right], \tag{4.11}$$

$$\frac{\partial \mathbf{r}_\mathbf{x}}{\partial \overline{p}_k} = \mathbf{J}\left[\text{diag}(\mathbf{h})\Re\{\mathbf{a}_d(\theta_k)\} + j\text{diag}(\overline{\mathbf{h}})\Im\{\mathbf{a}_d(\theta_k)\}\right], \tag{4.12}$$

where $\mathbf{h}$ and $\overline{\mathbf{h}}$ are given by

$$\mathbf{h} = \left[\frac{1}{\sqrt{1-|\Re\{\sum_{k=1}^{K}\overline{p}_k e^{-j\pi\sin\theta_k \ell_{D-1}}\}|^2}} \quad \cdots \quad 0 \quad \cdots \quad \frac{1}{\sqrt{1-|\Re\{\sum_{k=1}^{K}\overline{p}_k e^{j\pi\sin\theta_k \ell_{D-1}}\}|^2}}\right]^T, \tag{4.13}$$

$$\overline{\mathbf{h}} = \left[\frac{1}{\sqrt{1-|\Im\{\sum_{k=1}^{K}\overline{p}_k e^{-j\pi\sin\theta_k \ell_{D-1}}\}|^2}} \quad \cdots \quad 0 \quad \cdots \quad \frac{1}{\sqrt{1-|\Im\{\sum_{k=1}^{K}\overline{p}_k e^{j\pi\sin\theta_k \ell_{D-1}}\}|^2}}\right]^T, \tag{4.14}$$

$\mathbf{a}_d(\theta_k)$ denotes the k^{th} column of $\mathbf{A}_d(\boldsymbol{\theta})$ and $\mathbf{d} = [-\ell_{D-1}, \cdots, \ell_0, \cdots, \ell_{D-1}]^T$. It follows from (4.9), (4.11) and (4.11) that

$$\mathcal{I}_w(\boldsymbol{\varrho}) = N \begin{bmatrix}\mathbf{G}^H \\ \mathbf{V}^H\end{bmatrix} \mathbf{J}^H (\mathbf{R}_{\mathbf{x}}^{-T} \otimes \mathbf{R}_{\mathbf{x}}^{-1}) \mathbf{J} \begin{bmatrix}\mathbf{G} & \mathbf{V}\end{bmatrix}, \tag{4.15}$$

where

$$\mathbf{G} = j\pi \text{diag}(\mathbf{d}) \big[\text{diag}(\overline{\mathbf{h}})\Re\{\mathbf{A}_d(\boldsymbol{\theta})\} + j\text{diag}(\mathbf{h})\Im\{\mathbf{A}_d(\boldsymbol{\theta})\}\big] \boldsymbol{\Phi}(\boldsymbol{\theta}) \text{diag}(\overline{\mathbf{p}}), \tag{4.16}$$

$$\mathbf{V} = \text{diag}(\mathbf{h})\Re\{\mathbf{A}_d(\boldsymbol{\theta})\} + j\text{diag}(\overline{\mathbf{h}})\Im\{\mathbf{A}_d(\boldsymbol{\theta})\}, \tag{4.17}$$

with $\boldsymbol{\Phi}(\boldsymbol{\theta}) = \text{diag}([\cos\theta_1, \cos\theta_2, \cdots, \cos\theta_K]^T)$. If $\mathcal{I}_w(\boldsymbol{\varrho})$ is non-singular, a pessimistic approximation for the CRB of estimating DoAs from one-bit SLA data can be obtained through inverting $\mathcal{I}_w(\boldsymbol{\varrho})$. Hence, we need to first establish the non-singularity of $\mathcal{I}_w(\boldsymbol{\varrho})$.

Lemma 4.1. *Define* $\boldsymbol{\Upsilon} = \begin{bmatrix}\boldsymbol{\Delta} & F\end{bmatrix} \in \mathbb{C}^{(2D-1)\times 2K}$, *where*

$$\boldsymbol{\Delta} = \text{diag}(\mathbf{d}) \big[\text{diag}(\overline{\mathbf{h}})\Re\{\mathbf{A}_d(\boldsymbol{\theta})\} + j\text{diag}(\mathbf{h})\Im\{\mathbf{A}_d(\boldsymbol{\theta})\}\big], \tag{4.18}$$

$$F = \text{diag}(\mathbf{h})\Re\{\mathbf{A}_d(\boldsymbol{\theta})\} + j\text{diag}(\overline{\mathbf{h}})\Im\{\mathbf{A}_d(\boldsymbol{\theta})\}. \tag{4.19}$$

Then, $\mathcal{I}_w(\boldsymbol{\varrho})$ *is non-singular if and only if the matrix* $\boldsymbol{\Upsilon}$ *is full column rank.*

Proof. See Appendix B.3 □

Remark 4.2. Assuming $\mathcal{I}(\varrho)$ to be the true FIM, it follows from $\mathcal{I}(\varrho) \succeq \mathcal{I}_w(\boldsymbol{\varrho})$ that $\boldsymbol{\Upsilon}$ being full column rank is also a sufficient condition for the non-sigularity of $\mathcal{I}(\varrho)$.

Theorem 4.3. *Let* $CRB(\boldsymbol{\theta})$ *denote the the CRB for source DoAs* $\boldsymbol{\theta}$ *from* X. *If* $\mathcal{I}_w(\boldsymbol{\varrho})$ *is non-singular, then a pessimistic approximation of* $CRB(\boldsymbol{\theta})$, *denoted by* $CRB_w(\boldsymbol{\theta})$, *is given by*

$$CRB(\boldsymbol{\theta}) \preceq CRB_w(\boldsymbol{\theta}) = \frac{1}{4N}(\mathbf{Q}^H \Pi^{\perp}_{\mathbf{M}^{\frac{1}{2}}\mathbf{V}} \mathbf{Q})^{-1}, \tag{4.20}$$

where

$$\mathbf{M} = \mathbf{J}^H \left(\text{arcsine}(\overline{\mathbf{R}}^T) \otimes \text{arcsine}(\overline{\mathbf{R}})\right)^{-1} \mathbf{J}, \tag{4.21}$$

$$\mathbf{\Omega} = \frac{1}{\pi}\mathbf{G}, \tag{4.22}$$

$$\mathbf{Q} = \mathbf{M}^{\frac{1}{2}} \text{diag}(\mathbf{d}) \mathbf{\Omega} \mathbf{\Phi}(\boldsymbol{\theta}) \text{diag}(\overline{\mathbf{p}}), \tag{4.23}$$

with $\mathbf{G}$ *and* $\mathbf{V}$ *being given in* (4.16) *and* (4.17)*, respectively.*

Proof. See Appendix B.4 □

Remark 4.3. We note that $CRB_w(\boldsymbol{\theta})$ bears a superficial resemblance to the CRB expression for DoA estimation from unquantized data, given by [52, Theorem 2]

$$CRB_I(\boldsymbol{\theta}) = \frac{1}{4N\pi^2} (\widetilde{\mathbf{Q}}^H \Pi^{\perp}_{\widetilde{\mathbf{M}}^{\frac{1}{2}}\widetilde{\mathbf{V}}} \widetilde{\mathbf{Q}})^{-1}, \tag{4.24}$$

where

$$\widetilde{\mathbf{M}} = \mathbf{J}^H \left(\overline{\mathbf{R}}^T \otimes \overline{\mathbf{R}}\right)^{-1} \mathbf{J}, \tag{4.25}$$

$$\widetilde{\mathbf{Q}} = \widetilde{\mathbf{M}}^{\frac{1}{2}} \text{diag}(\mathbf{d}) \mathbf{A}_d(\boldsymbol{\theta}) \mathbf{\Phi}(\boldsymbol{\theta}) \text{diag}(\overline{\mathbf{p}}), \tag{4.26}$$

$$\widetilde{\mathbf{V}} = \begin{bmatrix} \mathbf{A}_d(\boldsymbol{\theta}) & \mathbf{e} \end{bmatrix}. \tag{4.27}$$

Theorem 4.4. *Assume all sources have equal power* p *and* $SNR = p/\sigma^2$*. Then, we have*

$$\lim_{SNR \to \infty} CRB_w(\boldsymbol{\theta}) \succ \mathbf{0}, \tag{4.28}$$

Proof. See Appendix B.5 □

Remark 4.4. Theorem 4.4 implies that the $CRB_w(\boldsymbol{\theta})$ does not go to zero as the SNR increases. As a consequence, in the one-bit DoA estimation problem, we may not be able to render estimation errors arbitrarily small by increasing the SNR.

4.5 Proposed One-Bit DoA Estimator

In this section, we first derive an enhanced estimate of the normalized covariance matrix of $\mathbf{y}(t)$, i.e., $\overline{\mathbf{R}}$, from one-bit SLA measurements through exploiting the structure of $\overline{\mathbf{R}}$. Then, we obtain DoA estimates by applying Co-Array-Based MUSIC (CAB-MUSIC) [38, 59] to the enhanced estimate of $\overline{\mathbf{R}}$. Further, we investigate the analytical performance of the proposed method for estimating DoAs from one-bit measurements.

4.5.1 Enhanced One-Bit Co-Array-Based MUSIC

It is deduced from the strong law of large numbers [110, ch. 8] that the sample covariance matrix of one-bit measurements provides a consistent estimate of $\mathbf{R_x}$ with probability 1, i.e.,

$$\Pr\left(\lim_{N\to\infty} \widehat{\mathbf{R}}_{\mathbf{x}} = \mathbf{R_x}\right) = 1, \tag{4.29}$$

where $\widehat{\mathbf{R}}_{\mathbf{x}} = \frac{1}{N}\mathbf{X}\mathbf{X}^H$. In addition, reformulating (4.7) provides $\overline{\mathbf{R}}$ based on the covariance matrix of one-bit data as follows:

$$\overline{\mathbf{R}} = \text{sine}(\frac{\pi}{2}\mathbf{R_x}), \tag{4.30}$$

where $[\text{sine}(\frac{\pi}{2}\mathbf{R_x})]_{m,n} = \sin(\frac{\pi}{2}\Re\{[\overline{\mathbf{R}}]_{m,n}\}) + j\sin(\frac{\pi}{2}\Im\{[\overline{\mathbf{R}}]_{m,n}\})$. Accordingly, a consistent estimate of $\overline{\mathbf{R}}$ is obtained as

$$\widetilde{\overline{\mathbf{R}}} = \text{sine}(\frac{\pi}{2}\widehat{\mathbf{R}}_{\mathbf{x}}). \tag{4.31}$$

Most of the algorithms in the literature employ $\widetilde{\overline{\mathbf{R}}}$ for estimating DoAs from one-bit measurements [32, 75]. However, an enhanced estimate of $\overline{\mathbf{R}}$ compared to $\widetilde{\overline{\mathbf{R}}}$ can be found if the structure of $\overline{\mathbf{R}}$ is taken into account. This enhanced estimate could in turn yield a better DoA estimation performance. In what follows, we introduce such an enhanced estimate of $\overline{\mathbf{R}}$ by exploiting its structure. Then, we use this enhanced estimate to improve the DoA estimation performance from one-bit data.

It is readily known from (4.8) that $\overline{\mathbf{R}}$ has the following structure

$$\overline{\mathbf{R}} = \mathbf{I}_M + \sum_{n=1}^{D-1} u_n \mathbf{L}_n + \sum_{n=1}^{D-1} u_n^* \mathbf{L}_n^T, \tag{4.32}$$

where $u_n = \sum_{k=1}^{K} \overline{p}_k e^{j\pi \sin\theta_k \ell_n}$ and $\mathbf{L}_n$ is given after eq. (4.5) for $1 \le n \le D-1$. It can be observed from (4.32) that the diagonal elements of $\overline{\mathbf{R}}$ are all one while the off-diagonal elements are parameterized by the vector $\mathbf{u} = [u_1, \cdots, u_{D-1}]^T \in \mathbb{C}^{(D-1)\times 1}$. This means that there exist only $2D-2$ free real parameters in $\overline{\mathbf{R}}$. Let $\ddot{\mathbf{r}} \in \mathbb{C}^{(M^2-M)\times 1}$ be the vector containing the off-diagonal elements of $\overline{\mathbf{R}}$, obtained by removing the diagonal elements of $\overline{\mathbf{R}}$ from $\text{vec}(\overline{\mathbf{R}})$. Evidently, $\ddot{\mathbf{r}}$ is given by

$$\ddot{\mathbf{r}} = \overline{\mathbf{J}} \begin{bmatrix} \mathbf{u}^* \\ \mathbf{u} \end{bmatrix} = \overline{\mathbf{J}}\boldsymbol{\Psi}\boldsymbol{\phi}, \tag{4.33}$$

where $\boldsymbol{\phi} = [\Re\{\mathbf{u}\}^T, \Im\{\mathbf{u}\}^T]^T \in \mathbb{R}^{(2D-1\times 1)}$,

$$\boldsymbol{\Psi} = \begin{bmatrix} \mathbf{I}_{D-1} & -j\mathbf{I}_{D-1} \\ \mathbf{I}_{D-1} & j\mathbf{I}_{D-1} \end{bmatrix}. \tag{4.34}$$

and $\overline{\mathbf{J}} \in \{0,1\}^{(M^2-M)\times(2D-2)}$ is obtained by removing the D-th column as well as the rows with indices $(i-1)M+1$ for all $1 \le i \le M$ from $\mathbf{J}$. It follows from (4.33) that $\overline{\mathbf{R}}$ is parameterized by the real-valued vector $\boldsymbol{\phi}$. We wish to find $\boldsymbol{\phi} \in \mathbb{E}_{\boldsymbol{\phi}} = \{\boldsymbol{\phi} \mid \overline{\mathbf{R}}(\boldsymbol{\phi}) \succeq \mathbf{0}\}$ from $\widehat{\mathbf{R}}_{\mathbf{x}}$. To this end, let $\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}} \in \mathbb{R}^{(M^2-M)\times 1}$ denote the vector made from the off-diagonal elements of $\widehat{\mathbf{R}}_{\mathbf{x}}$. For large N, it follows from the Central Limit Theorem (CLT) [110, ch. 8] that the distribution of $\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}$ asymptotically approaches a complex proper Gaussian distribution, i.e., $\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}} \xrightarrow{D} \mathcal{CN}(\ddot{\mathbf{r}}_{\mathbf{x}}, \frac{4}{\pi^2 N}\boldsymbol{\Sigma})$, where $\ddot{\mathbf{r}}_{\mathbf{x}}$ is the vector obtained from stacking the off-diagonal elements of $\mathbf{R}_{\mathbf{x}}$ and $\boldsymbol{\Sigma} = \frac{\pi^2 N}{4}\mathbb{E}\{(\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}} - \ddot{\mathbf{r}}_{\mathbf{x}})(\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}} - \ddot{\mathbf{r}}_{\mathbf{x}})^H\} \in \mathbb{C}^{(M^2-M)\times(M^2-M)}$. The closed-form expressions for the elements of $\boldsymbol{\Sigma}$ are provided in Appendix B.11. It is observed that the elements of $\boldsymbol{\Sigma}$ are functions of $\ddot{\mathbf{r}}$, thereby parameterized by $\boldsymbol{\phi}$ as well. Considering the transformation (4.31), the asymptotic distribution of the off-diagonal elements of $\widetilde{\overline{\mathbf{R}}}$, denoted by $\widetilde{\ddot{\mathbf{r}}} \in \mathbb{C}^{(M^2-M)\times 1}$, is given by

$$\begin{aligned} f(\widetilde{\ddot{\mathbf{r}}} \mid \boldsymbol{\phi}) = &\left(\frac{N^{M^2-M}}{(2\pi)^{M^2-M}\det(\boldsymbol{\Sigma}(\boldsymbol{\phi}))}\right) \\ &\times \frac{\exp\{-N[\text{arcsine}(\widetilde{\ddot{\mathbf{r}}}) - \overline{\mathbf{J}}\boldsymbol{\Psi}\arcsin(\boldsymbol{\phi})]^H\boldsymbol{\Sigma}^{-1}(\boldsymbol{\phi})[\text{arcsine}(\widetilde{\ddot{\mathbf{r}}}) - \overline{\mathbf{J}}\boldsymbol{\Psi}\arcsin(\boldsymbol{\phi})]\}}{\Pi_{n=1}^{D-1}(1-[\boldsymbol{\phi}]_n^2)^{\nu_n}(1-[\boldsymbol{\phi}]_{n+D-1}^2)^{\nu_n}}. \end{aligned} \tag{4.35}$$

Hence, the asymptotic ML estimation of $\boldsymbol{\phi}$ from $\widetilde{\ddot{\mathbf{r}}}$ is derived as follows;

$$\widehat{\boldsymbol{\phi}} = \underset{\boldsymbol{\phi}\in\mathbb{E}_{\boldsymbol{\phi}}}{\text{argmin}}\ L(\boldsymbol{\phi}), \tag{4.36}$$

where the cost function $L(\boldsymbol{\phi})$ is given by

$$\begin{aligned} L(\boldsymbol{\phi}) = &\ln\det(\boldsymbol{\Sigma}(\boldsymbol{\phi})) - \sum_{n=1}^{D-1}\nu_n \ln(1-[\boldsymbol{\phi}]_n^2)(1-[\boldsymbol{\phi}]_{n+D-1}^2) \\ &+ N[\text{arcsine}(\widetilde{\ddot{\mathbf{r}}}) - \overline{\mathbf{J}}\boldsymbol{\Psi}\arcsin(\boldsymbol{\phi})]^H\boldsymbol{\Sigma}^{-1}(\boldsymbol{\phi})[\text{arcsine}(\widetilde{\ddot{\mathbf{r}}}) - \overline{\mathbf{J}}\boldsymbol{\Psi}\arcsin(\boldsymbol{\phi})], \end{aligned} \tag{4.37}$$

with $\nu_n = \|\text{vec}(\mathbf{L}_n)\|^2$. However, the minimization of (4.37) with respect to $\boldsymbol{\phi}$ is very complicated owing to the nonlinearity of the cost function as well as the constraint $\boldsymbol{\phi} \in \mathbb{E}_{\boldsymbol{\phi}}$. To make the problem computationally tractable, we first find an asymptotic equivalent approximation of $L(\boldsymbol{\phi})$ which is much simpler to minimize. Let $\boldsymbol{\gamma} \in \mathbb{E}_{\boldsymbol{\gamma}} \subset \mathbb{R}^{(M^2-M)\times 1}$ be the $(M^2-M)\times 1$ vector containing the real and imaginary parts of the elements of $\overline{\mathbf{R}}$ above its main diagonal elements. Obviously, there is the following relationship

between $\boldsymbol{\phi}$ and $\boldsymbol{\gamma}$:

$$\boldsymbol{\gamma} = \mathbf{F}\overline{\mathbf{J}}\boldsymbol{\Psi}\boldsymbol{\phi} \in \mathbb{E}_{\boldsymbol{\gamma}}, \forall \boldsymbol{\phi} \in \mathbb{E}_{\boldsymbol{\phi}}, \tag{4.38}$$

where $\mathbf{F} = \frac{1}{2}\begin{bmatrix}\ddot{\mathbf{F}}^T & j\widetilde{\mathbf{F}}^T\end{bmatrix}^T \in \{0,1\}^{(M^2-M)\times(M^2-M)}$ such that for all $1 \leq p < q \leq M$:

1. the $\left((p-1)M + q - \frac{p(p+1)}{2}\right)$-th rows of $\ddot{\mathbf{F}} \in \{0,1\}^{\frac{(M^2-M)}{2}\times(M^2-M)}$ is obtained by removing the elements with indices $(i-1)M+1$ for all $1 \leq i \leq M$ from $\overline{\mathbf{e}}_p^T \otimes \overline{\mathbf{e}}_q^T + \overline{\mathbf{e}}_q^T \otimes \overline{\mathbf{e}}_p^T$ with $[\overline{\mathbf{e}}_p]_n = \delta[p-n]$ for $1 \leq n \leq M$.
2. the $\left((p-1)M + q - \frac{p(p+1)}{2}\right)$-th rows of $\widetilde{\mathbf{F}} \in \{0,1\}^{\frac{(M^2-M)}{2}\times(M^2-M)}$ is obtained by removing the elements with indices $(i-1)M+1$ for all $1 \leq i \leq M$ from $\overline{\mathbf{e}}_p^T \otimes \overline{\mathbf{e}}_q^T - \overline{\mathbf{e}}_q^T \otimes \overline{\mathbf{e}}_p^T$ with $[\overline{\mathbf{e}}_p]_n = \delta[p-n]$ for $1 \leq n \leq M$.

Lemma 4.2. *The matrices* $\mathbf{F}$, $\boldsymbol{\Psi}$ *and* $\overline{\mathbf{J}}$ *are full rank.*

Proof. See Appendix B.6. □

The mapping from $\boldsymbol{\phi} \in \mathbb{E}_{\boldsymbol{\phi}}$ to $\boldsymbol{\gamma} \in \mathbb{E}_{\boldsymbol{\gamma}}$ is one-to-one due to the full rankness of $\mathbf{F}$, $\boldsymbol{\Psi}$ and $\overline{\mathbf{J}}$. Hence, it is possible to equivalently reparameterize (4.37) in terms of $\boldsymbol{\gamma}$ instead of $\boldsymbol{\phi}$. This can be done by simply replacing $\boldsymbol{\phi}$ with $\boldsymbol{\Psi}^{-1}\overline{\mathbf{J}}^{\dagger}\mathbf{F}^{-1}\boldsymbol{\gamma}$. To achieve computational simplification, we make use of the fact that a consistent estimate of $\boldsymbol{\gamma}$ can be obtained as $\widetilde{\boldsymbol{\gamma}} = \mathbf{F}\widetilde{\ddot{\mathbf{r}}}$. We can see that $\widetilde{\boldsymbol{\gamma}} \in \mathbb{R}^{(M^2-M)\times 1} \notin \mathbb{E}_{\boldsymbol{\gamma}}$ with probability one, since the $\mathbb{E}_{\boldsymbol{\gamma}}$ is a zero-measure subset of $\mathbb{R}^{(M^2-M)\times 1}$. Now, considering the Taylor series expansion of $L(\boldsymbol{\gamma})$ around $\widetilde{\boldsymbol{\gamma}}$, we obtain

$$\begin{aligned} L(\boldsymbol{\gamma}) =& L(\widetilde{\boldsymbol{\gamma}}) + (\boldsymbol{\gamma} - \widetilde{\boldsymbol{\gamma}})^H \nabla_{\boldsymbol{\gamma}} L(\widetilde{\boldsymbol{\gamma}}) \\ &+ \frac{1}{2}(\widetilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma})^H \nabla^2_{\boldsymbol{\gamma}} L(\widetilde{\boldsymbol{\gamma}})(\widetilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) + \cdots, \end{aligned} \tag{4.39}$$

where $\nabla_{\boldsymbol{\gamma}} L(\widetilde{\boldsymbol{\gamma}})$ and $\nabla^2_{\boldsymbol{\gamma}} L(\widetilde{\boldsymbol{\gamma}})$ denote the gradient vector and the Hessian matrix of $L(\boldsymbol{\gamma})$ with respect to $\boldsymbol{\gamma}$, computed at $\widetilde{\boldsymbol{\gamma}}$, respectively. The first term in (4.39) is constant and, moreover, the higher other term can be neglected for large N considering the fact that $\widetilde{\boldsymbol{\gamma}}$ is a consistent estimate of $\boldsymbol{\gamma}$. Consequently, making use of (4.38) and the fact that $\widetilde{\boldsymbol{\gamma}} = \mathbf{F}\widetilde{\ddot{\mathbf{r}}}$, we have

$$\begin{aligned} \widehat{\boldsymbol{\phi}} \simeq \underset{\boldsymbol{\phi}\in\mathbb{E}_{\boldsymbol{\phi}}}{\operatorname{argmin}}\ & (\overline{\mathbf{J}}\boldsymbol{\Psi}\boldsymbol{\phi} - \widetilde{\ddot{\mathbf{r}}})^H \mathbf{F}^H \nabla_{\boldsymbol{\gamma}} L(\widetilde{\boldsymbol{\gamma}}) \\ &+ (\widetilde{\ddot{\mathbf{r}}} - \overline{\mathbf{J}}\boldsymbol{\Psi}\boldsymbol{\phi})^H \mathbf{F}^H \nabla^2_{\boldsymbol{\gamma}} L(\widetilde{\boldsymbol{\gamma}}) \mathbf{F}(\widetilde{\ddot{\mathbf{r}}} - \overline{\mathbf{J}}\boldsymbol{\Psi}\boldsymbol{\phi}). \end{aligned} \tag{4.40}$$

The above quadratic optimization problem is asymptotically equivalent to (4.36) but is much more convenient to work with. Relaxing the constraint $\boldsymbol{\phi} \in \mathbb{E}_{\boldsymbol{\phi}}$ with $\boldsymbol{\phi} \in \mathbb{R}^{2D-2}$

yields the following closed-form solution for $\widehat{\boldsymbol{\phi}}$

$$\begin{aligned}\widehat{\boldsymbol{\phi}} \simeq &\boldsymbol{\Psi}^{-1}\left(\overline{\mathbf{J}}^H\mathbf{F}^H\nabla^2_{\boldsymbol{\gamma}}L(\widetilde{\boldsymbol{\gamma}})\mathbf{F}\overline{\mathbf{J}}\right)^{-1}\overline{\mathbf{J}}^H \\ &\times\left[\mathbf{F}^H\nabla^2_{\boldsymbol{\gamma}}L(\widetilde{\boldsymbol{\gamma}})\mathbf{F}\widetilde{\ddot{\mathbf{r}}}-\mathbf{F}^H\nabla_{\boldsymbol{\gamma}}L(\widetilde{\boldsymbol{\gamma}})\right].\end{aligned} \tag{4.41}$$

To derive the final expression for $\widehat{\boldsymbol{\phi}}$, we need to calculate $\nabla_{\boldsymbol{\gamma}}L(\widetilde{\boldsymbol{\gamma}})$ and $\nabla^2_{\boldsymbol{\gamma}}L(\widetilde{\boldsymbol{\gamma}})$. It is straightforward to derive $L(\boldsymbol{\gamma})$ by making use of (4.38). It follows that

$$\nabla_{\boldsymbol{\gamma}}L(\widetilde{\boldsymbol{\gamma}}) = \mathbf{g}(\widetilde{\boldsymbol{\gamma}}), \tag{4.42}$$

where $[\mathbf{g}(\boldsymbol{\gamma})]_n = \frac{4[\boldsymbol{\gamma}]_n}{1-|[\boldsymbol{\gamma}]_n|^2} + \frac{\partial \ln\det(\boldsymbol{\Sigma}(\boldsymbol{\gamma}))}{\partial[\boldsymbol{\gamma}]_n}$, for $1 \le n \le M^2 - M$. Additionally, the Hessian matrix at $\widetilde{\boldsymbol{\gamma}}$ is obtained as

$$\nabla^2_{\boldsymbol{\gamma}}L(\widetilde{\boldsymbol{\gamma}}) = N\mathrm{diag}(\widehat{\mathbf{b}})\mathbf{F}^{-H}\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{F}^{-1}\mathrm{diag}(\widehat{\mathbf{b}}) + \mathbf{E}(\widetilde{\boldsymbol{\gamma}}), \tag{4.43}$$

where $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\widetilde{\boldsymbol{\gamma}})$, $[\widehat{\mathbf{b}}]_n = \frac{1}{\sqrt{1-|[\widetilde{\boldsymbol{\gamma}}]_n|^2}}$, for $1 \le n \le M^2 - M$ and $[\mathbf{E}(\boldsymbol{\phi})]_{n,l} = \frac{2\nu_n(1+|[\boldsymbol{\phi}]_n|^2)}{(1-|[\boldsymbol{\phi}]_n|^2)^2} + \frac{\partial^2 \ln\det(\boldsymbol{\Sigma}(\boldsymbol{\gamma}))}{\partial[\boldsymbol{\gamma}]_n\partial[\boldsymbol{\gamma}]_m}$. Inserting (4.42) and (4.43) into (4.41) leads to

$$\begin{aligned}\widehat{\boldsymbol{\phi}} \simeq &\boldsymbol{\Psi}^{-1}\left(\overline{\mathbf{J}}^H\mathbf{F}^H\mathrm{diag}(\widehat{\mathbf{b}})\mathbf{F}^{-H}\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{F}^{-1}\mathrm{diag}(\widehat{\mathbf{b}})\mathbf{F}\overline{\mathbf{J}} + \overbrace{\frac{\mathbf{E}(\widetilde{\boldsymbol{\gamma}})}{N}}^{\hbar}\right)^{-1} \\ &\times\overline{\mathbf{J}}^H\Bigg(\mathbf{F}^H\mathrm{diag}(\widehat{\mathbf{b}})\mathbf{F}^{-H}\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{F}^{-1}\mathrm{diag}(\widehat{\mathbf{b}})\mathbf{F}\widetilde{\ddot{\mathbf{r}}} + \underbrace{\frac{\mathbf{F}^H\mathbf{E}(\widetilde{\boldsymbol{\gamma}})\mathbf{F}\widetilde{\ddot{\mathbf{r}}} - \mathbf{F}^H\mathbf{g}(\widetilde{\boldsymbol{\gamma}})}{N}}_{\aleph}\Bigg).\end{aligned} \tag{4.44}$$

In the above equation, the terms $\hbar$ and $\aleph$ can be neglected for large N, thus (4.44) may be simplified as

$$\begin{aligned}\widehat{\boldsymbol{\phi}} \simeq &\boldsymbol{\Psi}^{-1}\left(\overline{\mathbf{J}}^H\mathbf{F}^H\mathrm{diag}(\widehat{\mathbf{b}})\mathbf{F}^{-H}\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{F}^{-1}\mathrm{diag}(\widehat{\mathbf{b}})\mathbf{F}\overline{\mathbf{J}}\right)^{-1} \\ &\times\overline{\mathbf{J}}^H\mathbf{F}^H\mathrm{diag}(\widehat{\mathbf{b}})\mathbf{F}^{-H}\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{F}^{-1}\mathrm{diag}(\widehat{\mathbf{b}})\mathbf{F}\widetilde{\ddot{\mathbf{r}}}.\end{aligned} \tag{4.45}$$

Hence, from (4.33), an enhanced consistent estimate of $\overline{\mathbf{r}} = \mathrm{vec}(\overline{\mathbf{R}})$ is derived as follows

$$\widehat{\overline{\mathbf{r}}} = \mathbf{J}\begin{bmatrix}\mathbf{0} & \mathbf{I}_{D-1} & -j\mathbf{I}_{D-1}\\ 1 & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{D-1} & j\mathbf{I}_{D-1}\end{bmatrix}\begin{bmatrix}1\\ \widehat{\boldsymbol{\phi}}\end{bmatrix}. \tag{4.46}$$

Remark 4.5. Considering $\lim_{N\to\infty}\widetilde{\ddot{\mathbf{r}}} = \ddot{\mathbf{r}}$, it is readily observed from (4.33) and (4.45) that $\widehat{\boldsymbol{\phi}}$ is a consistent estimate of $\boldsymbol{\phi}$. This in turn implies that $\widehat{\overline{\mathbf{r}}}$ is also a consistent estimate of $\overline{\mathbf{r}}$.

Algorithm 1 EOCAB-MUSIC

Input: SLA one-bit observations, i.e., $\mathbf{X}$.
Output: The estimates of source DoAs.

1: Compute the sample covariance matrix of one-bit data as $\widehat{\mathbf{R}}_{\mathbf{x}} = \frac{1}{N}\mathbf{X}\mathbf{X}^H$.
2: Compute $\widetilde{\overline{\mathbf{R}}}$ from (4.31).
3: Form $\ddot{\widetilde{\mathbf{r}}}$ by removing the diagonal elements of $\widetilde{\overline{\mathbf{R}}}$ from vec($\widetilde{\overline{\mathbf{R}}}$).
4: Compute $\widetilde{\boldsymbol{\gamma}}$ from $\widetilde{\boldsymbol{\gamma}} = \mathbf{F}\ddot{\widetilde{\mathbf{r}}}$.
5: Compute $\widehat{\mathbf{b}}$ using $[\widehat{\mathbf{b}}]_n = \frac{1}{\sqrt{1-|[\widetilde{\boldsymbol{\gamma}}]_n|^2}}$, for $1 \le n \le M^2 - M$.
6: Compute $\widehat{\boldsymbol{\Sigma}}$ by replacing $\overline{\mathbf{R}}$ with $\widetilde{\overline{\mathbf{R}}}$ in the closed-form expressions for the elements of $\boldsymbol{\Sigma}$ given in Appendix K.
7: Compute $\widehat{\boldsymbol{\phi}}$ from (4.45).
8: Compute $\widehat{\overline{\mathbf{r}}}$ from (4.46).
9: Compute $\widehat{\widehat{\mathbf{R}}}_v$ from (4.47).
10: Apply MUSIC to $\widehat{\widehat{\mathbf{R}}}_v$ to estimate DoAs.

To estimate DoAs using $\widehat{\overline{\mathbf{r}}}$, we resort to CAB-MUSIC [75]. Specifically, we first construct the normalized augmented covariance matrix as

$$\widehat{\widehat{\mathbf{R}}}_v = \begin{bmatrix} \mathbf{T}_v\mathbf{J}^\dagger\widehat{\overline{\mathbf{r}}} & \mathbf{T}_{v-1}\mathbf{J}^\dagger\widehat{\overline{\mathbf{r}}} & \cdots & \mathbf{T}_1\mathbf{J}^\dagger\widehat{\overline{\mathbf{r}}} \end{bmatrix} \in \mathbb{C}^{v\times v}, \tag{4.47}$$

where $\mathbf{T}_i$ is a selection matrix, defined as

$$\mathbf{T}_i = \begin{bmatrix} \mathbf{0}_{v\times(i+D-v-1)} & \mathbf{I}_v & \mathbf{0}_{v\times(D-i)} \end{bmatrix} \in \{0,1\}^{v\times(2D-1)}. \tag{4.48}$$

It follows from the consistency of $\widehat{\overline{\mathbf{r}}}$ that

$$\begin{aligned} \lim_{N\to\infty} \widehat{\widehat{\mathbf{R}}}_v &= \begin{bmatrix} \mathbf{T}_v\mathbf{J}^\dagger\overline{\mathbf{r}} & \mathbf{T}_{v-1}\mathbf{J}^\dagger\overline{\mathbf{r}} & \cdots & \mathbf{T}_1\mathbf{J}^\dagger\overline{\mathbf{r}} \end{bmatrix} \in \mathbb{C}^{v\times v} \\ &= \mathbf{A}_v(\boldsymbol{\theta})\mathrm{diag}(\overline{\mathbf{p}})\mathbf{A}_v^H(\boldsymbol{\theta}) + \overline{\sigma}^2\mathbf{I}_v, \end{aligned} \tag{4.49}$$

where $\mathbf{A}_v(\boldsymbol{\theta}) = [\mathbf{a}_v(\theta_1), \mathbf{a}_v(\theta_2), \cdots, \mathbf{a}_v(\theta_K)] \in \mathbb{C}^{v\times K}$ denotes the steering matrix of a contiguous ULA with v elements located at $(0, \frac{\lambda}{2}, \cdots, (v-1)\frac{\lambda}{2})$. Hence, we can apply MUSIC to $\widehat{\widehat{\mathbf{R}}}_v$ to estimate the DoAs. We call the proposed method Enhanced One-bit CAB-MUSIC (EOCAB-MUSIC). Algorithm 1 summarizes the steps of EOCAB-MUSIC.

Remark 4.6. The computational complexity of each step of Algorithm 1 is separately specified in Table 4.1 where $\mathcal{G}(n)$, $\mathcal{K}(n)$ and $\mathcal{Z}$ denote the complexity of the chosen algorithm for multiplication of two n-digit numbers, the complexity of integration in (108) and the number of grid point of the MUSIC algorithm, respectively. Considering that D and v are typically in the order of M^2 and, moreover, n and M are normally very smaller than $\mathcal{Z}$, it follows from Table 4.1 that the complexity of EOCAB-MUSIC is in the order of $\mathcal{O}(MN + M^2(\mathcal{G}(n)(\mathcal{Z}+M^4) + \mathcal{K}(n)M^2))$. On the other hand implementation of OCAB-MUSIC needs only steps 1, 2, 9 and 10 in algorithm 1. Hence, its complexity

TABLE 4.1: Complexity of the steps of Algorithm 1

Step order	Complexity
1	$\mathcal{O}(MN)$
2	$\mathcal{O}(\mathcal{G}(n)\sqrt{n}M^2)$
3	$\mathcal{O}(M^2)$
4	$\mathcal{O}(\mathcal{G}(n)M^4)$
5	$\mathcal{O}(\mathcal{G}(n)M^2)$
6	$\mathcal{O}(\mathcal{K}(n)M^4)$
7	$\mathcal{O}(\mathcal{G}(n)(DM^4+M^6+D^3))$
8	$\mathcal{O}(\mathcal{G}(n)(D^2M^2)$
9	$\mathcal{O}(\mathcal{G}(n)(M^2v(2D-1+v))$
10	$\mathcal{O}(\mathcal{G}(n)(\mathcal{Z}M^2+M^3))$

is given by $\mathcal{O}(MN + M^2\mathcal{G}(n)(\mathcal{Z}+M^4))$. Typically, we have $\mathcal{G}(n)(\mathcal{Z}+M^4) \gg \mathcal{K}(n)M^2$, implying that the complexity of EOCAB-MUSIC is almost in the same order as that of OCAB-MUSIC.

4.5.2 Asymptotic Performance Analysis

In this section, we investigate the asymptotic performance of the proposed estimator through the derivation of a closed-form expression for the second-order statistics of the asymptotic distribution (as $N \to \infty$) of the DoA estimation errors. Our main results are summarized in Theorem 4.5, Corollary 4.1, Corollary 4.2 and Theorem 4.6.

Lemma 4.3. *$\hat{\boldsymbol{\theta}}$ obtained by EOCAB-MUSIC is a consistent estimate of $\boldsymbol{\theta}$ if $K \leq v-1$.*

Proof. See Appendix B.7 □

Theorem 4.5. *The closed-form expression for the covariance of the asymptotic distribution (as $N \to \infty$) of the DoA estimation errors obtained by EOCAB-MUSIC is given by*

$$\mathcal{E}_{\theta_{k_1},\theta_{k_2}} = \frac{(\sigma^2+\sum_{k=1}^{K}p_k)^2\Re\{\mathbf{z}_{k_1}^T\overline{\mathbf{T}}(\overline{\mathbf{J}}^H\mathbf{W}\overline{\mathbf{J}})^{-1}\overline{\mathbf{J}}^H\mathbf{W}\boldsymbol{\Gamma}\mathbf{W}\overline{\mathbf{J}}(\overline{\mathbf{J}}^H\mathbf{W}\overline{\mathbf{J}})^{-1}\overline{\mathbf{T}}^H\mathbf{z}_{k_2}^*\}}{N\pi^2 p_{k_1}p_{k_2}q_{k_1}q_{k_2}\cos\theta_{k_1}\cos\theta_{k_2}}, \quad (4.50)$$

where

$$\mathbf{z}_k = \boldsymbol{\beta}_k \otimes \boldsymbol{\alpha}_k, \tag{4.51}$$

$$\boldsymbol{\beta}_k = \Pi^{\perp}_{\mathbf{A}_v(\boldsymbol{\theta})} \mathrm{diag}(\mathbf{v}) \mathbf{a}_v(\theta_k), \tag{4.52}$$

$$\boldsymbol{\alpha}_k = \mathbf{A}_v^{\dagger T}(\boldsymbol{\theta}) \boldsymbol{\imath}_k, \tag{4.53}$$

$$q_k = \mathbf{a}_v^H(\theta_k) \mathrm{diag}(\mathbf{v}) \Pi^{\perp}_{\mathbf{A}_v} \mathrm{diag}(\mathbf{v}) \mathbf{a}_v(\theta_k), \tag{4.54}$$

$$\mathbf{W} = \mathbf{F}^H \mathrm{diag}(\mathbf{b}) \mathbf{F}^{-H} \boldsymbol{\Sigma}^{-1} \mathbf{F}^{-1} \mathrm{diag}(\mathbf{b}) \mathbf{F}, \tag{4.55}$$

$$\begin{aligned}[\boldsymbol{\Gamma}]_{p,q} =& \frac{1}{2}\Bigg(\sqrt{1 - [\Re\{[\ddot{\mathbf{r}}]_p\}]^2} \times \sqrt{1 - [\Re\{[\ddot{\mathbf{r}}]_q\}]^2} \\ &+ \sqrt{1 - [\Im\{[\ddot{\mathbf{r}}]_p\}]^2} \times \sqrt{1 - [\Im\{[\ddot{\mathbf{r}}]_q\}]^2} \Bigg) \Re\{[\boldsymbol{\Sigma}]_{p,q}\} \\ &+ \frac{j}{2}\Bigg(\sqrt{1 - [\Im\{[\ddot{\mathbf{r}}]_p\}]^2} \times \sqrt{1 - [\Re\{[\ddot{\mathbf{r}}]_q\}]^2} \\ &+ \sqrt{1 - [\Re\{[\ddot{\mathbf{r}}]_p\}]^2} \times \sqrt{1 - [\Im\{[\ddot{\mathbf{r}}]_q\}]^2} \Bigg) \Im\{[\boldsymbol{\Sigma}]_{p,q}\}, \end{aligned} \tag{4.56}$$

with $\mathbf{v} = [0, 1, 2, \cdots, v-1]^T$, $[\mathbf{b}]_n = \frac{1}{\sqrt{1-|[\boldsymbol{\gamma}]_n|^2}}$ *for* $1 \le n \le M^2 - M$, $\boldsymbol{\Sigma} \in \mathbb{C}^{(M^2-M)\times(M^2-M)}$ *as given in Appendix B.11,* $\overline{\mathbf{T}} \in \mathbb{C}^{v^2 \times (2D-2)}$ *as defined in* (B.38) *in Appendix B.8, and* $\boldsymbol{\imath}_k$ *being the* k^{th} *column of* $\mathbf{I}_K$.

Proof. See Appendix B.8 □

corollary 4.1. *The asymptotic MSE expression (as* $N \to \infty$*) for the DoA estimates obtained by EOCAB-MUSIC is given by*

$$\begin{aligned}\mathcal{E}_{\theta_k} &= \mathbb{E}\{(\theta_{k_1} - \widehat{\theta}_k)^2\} \\ &= \frac{(\sigma^2 + \sum_{k'=1}^{K} p_{k'})^2 \Re\{\mathbf{z}_k^T \overline{\mathbf{T}} (\overline{\mathbf{J}}^H \mathbf{W} \overline{\mathbf{J}})^{-1} \overline{\mathbf{J}}^H \mathbf{W} \boldsymbol{\Gamma} \mathbf{W} \overline{\mathbf{J}} (\overline{\mathbf{J}}^H \mathbf{W} \overline{\mathbf{J}})^{-1} \overline{\mathbf{T}}^H \mathbf{z}_k^*\}}{N \pi^2 p_k^2 q_k^2 \cos^2 \theta_k}. \end{aligned} \tag{4.57}$$

corollary 4.2. *The covariance of the asymptotic distribution (as* $N \to \infty$*) of the DoA estimation errors and the asymptotic MSE expression (as* $N \to \infty$*) for the one-bit DoA estimator given in [75], named as One-bit CAB-MUSIC (OCAB-MUSIC), is easily obtained by replacing* $\mathbf{W}$ *with* $\mathbf{I}_{M^2-M}$ *in* (4.50) *and* (4.57)*, respectively.*

Proof. See Appendix B.9 □

Remark 4.7. It is concluded from Corollary 4.1 and Corollary 4.2 that, similar to Infinite-bit Co-Array-Based MUSIC (ICAB-MUSIC) [38], the MSEs of EOCAB-MUSIC and OCAB-MUSIC also depend on both the physical and the virtual array geometries through $\mathbf{A}_v(\theta)$ and $\overline{\mathbf{R}}$, respectively.

Remark 4.8. Another interesting implication of Corollary 4.1 is that the MSEs of EOCAB-MUSIC and OCAB-MUSIC reduce at the same rate as that of ICAB-MUSIC [38] with respect to N; i.e. $\mathcal{E}_{\theta_k} \propto \frac{1}{N}$ for both.

Remark 4.9. It is readily clear from the definition that $\overline{\mathbf{r}}$ is a function of the SNR, and not $\mathbf{p}$ and σ^2. This indicates that $\mathbf{W}$ and $\mathbf{\Gamma}$ are also functions of the SNR instead of $\mathbf{p}$ and σ^2. Further, multiplying the numerator and denominator of $(\sigma^2+\sum_{k'=1}^{K} p_{k'})^2/p_k^2$ by $1/\sigma^4$ reformulates it as a function of the SNR. These observations imply that the MSEs of EOCAB-MUSIC and OCAB-MUSIC are functions of the SNR instead of $\mathbf{p}$ and σ^2. This fact can also be deduced directly from system model where we have

$$\begin{aligned}[\mathbf{x}(t)]_m &= \frac{1}{\sqrt{2}}\text{sgn}\left(\Re\{[\mathbf{y}(t)]_m\}\right)+\frac{j}{\sqrt{2}}\ \text{sgn}\left(\Im\{[\mathbf{y}(t)]_m\}\right)\\ &= \frac{1}{\sqrt{2}}\text{sgn}\left(\Re\{\frac{[\mathbf{y}(t)]_m}{\sigma}\}\right)+\frac{j}{\sqrt{2}}\ \text{sgn}\left(\Im\{\frac{[\mathbf{y}(t)]_m}{\sigma}\}\right).\end{aligned} \tag{4.58}$$

for $\sigma > 0$. This implies that, without loss of generality, the power of each source can consider to be equal to the SNR for that source and the noise variance can be consider to be equal to 1.

Theorem 4.6. *Assume all sources have equal power p and $SNR = p/\sigma^2$. Then, for a sufficiently large SNR, the MSE of EOCAB-MUSIC converges to the following constant value:*

$$\begin{aligned}\lim_{SNR\to\infty} \mathcal{E}_{\theta_k} =& \frac{K^2}{N\pi^2 q_k^2 \cos^2\theta_k}\\ &\times \Re\{\mathbf{z}_k^T\overline{\mathbf{T}}(\overline{\mathbf{J}}^H\mathbf{W}_\infty\overline{\mathbf{J}})^{-1}\overline{\mathbf{J}}^H\mathbf{W}_\infty\mathbf{\Gamma}_\infty\mathbf{W}_\infty\overline{\mathbf{J}}(\overline{\mathbf{J}}^H\mathbf{W}_\infty\overline{\mathbf{J}})^{-1}\overline{\mathbf{T}}^H\mathbf{z}_k^*\} > 0,\end{aligned} \tag{4.59}$$

where $\mathbf{W}_\infty$ and $\mathbf{\Gamma}_\infty$ are obtained by replacing $\overline{\mathbf{R}}$, $\ddot{\mathbf{r}}$ and $\boldsymbol{\gamma}$ in the definitions of $\mathbf{W}$ and $\mathbf{\Gamma}$ (kindly refer to Theorem 4.5) with $\overline{\mathbf{R}}_\infty$, $\boldsymbol{\gamma}_\infty$ and $\ddot{\mathbf{r}}_\infty$, respectively, where

$$\overline{\mathbf{R}}_\infty = \frac{1}{K}\mathbf{A}(\boldsymbol{\theta})\mathbf{A}^H(\boldsymbol{\theta}) + (1-\frac{1}{K})\mathbf{I}_M, \tag{4.60}$$

$\boldsymbol{\gamma}_\infty$ is the $(M^2-M)\times 1$ vector containing the real and imaginary parts of the elements of $\overline{\mathbf{R}}_\infty$ above its main diagonal elements and $\ddot{\mathbf{r}}_\infty = \mathbf{\Psi}^{-1}\overline{\mathbf{J}}^\dagger\mathbf{F}^{-1}\boldsymbol{\gamma}_\infty$.

Proof. See Appendix B.10. □

Remark 4.10. It follows from Theorem 4.6 that it is not possible to make the MSEs of EOCAB-MUSIC and OCAB-MUSIC arbitrarily small by increasing the SNR.

4.6 Simulation Results

In this section, we provide some numerical results to validate the analytical results obtained in previous sections as well as to assess the performance of the proposed DoA estimator. Specifically, we will show that the proposed estimator yields better performance in terms of estimation accuracy and resolution compared to the approach given in [75]. In the rest of this section, we will refer to: 1. the CRB for DoA estimation from infinite-bit measurements as Infinite-bit CRB (I-CRB), whose expression is given in Remark 4.3; 2. the pessimistic approximation of the CRB for DoA estimation from one-bit measurements as One-bit CRB (O-CRB); 3. CAB-MUSIC using infinite-bit measurements as Infinite-bit CAB-MUSIC (ICAB-MUSIC); 4. the DoA estimator given in [75] as one-bit CAB-MUSIC (OCAB-MUSIC); 5. the proposed estimator in this paper as Enhanced One-bit CAB-MUSIC (EOCAB-MUSIC).

4.6.1 General Set-up

In all experiments, each simulated point has been computed by 5000 Monte Carlo repetitions. Unless the source locations are specified for a particular result, it is assumed that the K independent sources are equally spaced in the angular domain $[-60^\circ, 60^\circ]$ such that $\theta = -60^\circ$ when $K = 1$. Further, all sources are assumed to have equal powers, i.e., $p_k = p$ for all k, and the SNR is defined as $10 \log \frac{p}{\sigma^2}$. For our numerical investigation, we use four different types of arrays with $M = 10$ physical elements and the following geometries:

$$\mathbb{M}_{\text{nested}} : \{1, 2, 3, 4, 5, 6, 12, 18, 24, 30\}, \tag{4.61}$$

$$\mathbb{M}_{\text{co-prime}} : \{0, 3, 5, 6, 9, 10, 12, 15, 20, 25\}, \tag{4.62}$$

$$\mathbb{M}_{\text{MRA}} : \{0, 1, 3, 6, 13, 20, 27, 31, 35, 36\}, \tag{4.63}$$

$$\mathbb{M}_{\text{ULA}} : \{0, 1, 2, \cdots, 9\}. \tag{4.64}$$

These arrays generate the difference co-arrays:

$$\mathbb{D}_{\text{nested}} : \{0, 1, 2, \cdots, 29\}, \tag{4.65}$$

$$\mathbb{D}_{\text{co-prime}} : \{0, 1, 2, \cdots, 22, 25\}, \tag{4.66}$$

$$\mathbb{D}_{\text{MRA}} : \{0, 1, 2, \cdots, 36\}, \tag{4.67}$$

$$\mathbb{D}_{\text{ULA}} : \{0, 1, 2, \cdots, 9\}. \tag{4.68}$$

Further, we generate the grid from -90° to 90° with step size 0.001° to implement MUSIC.

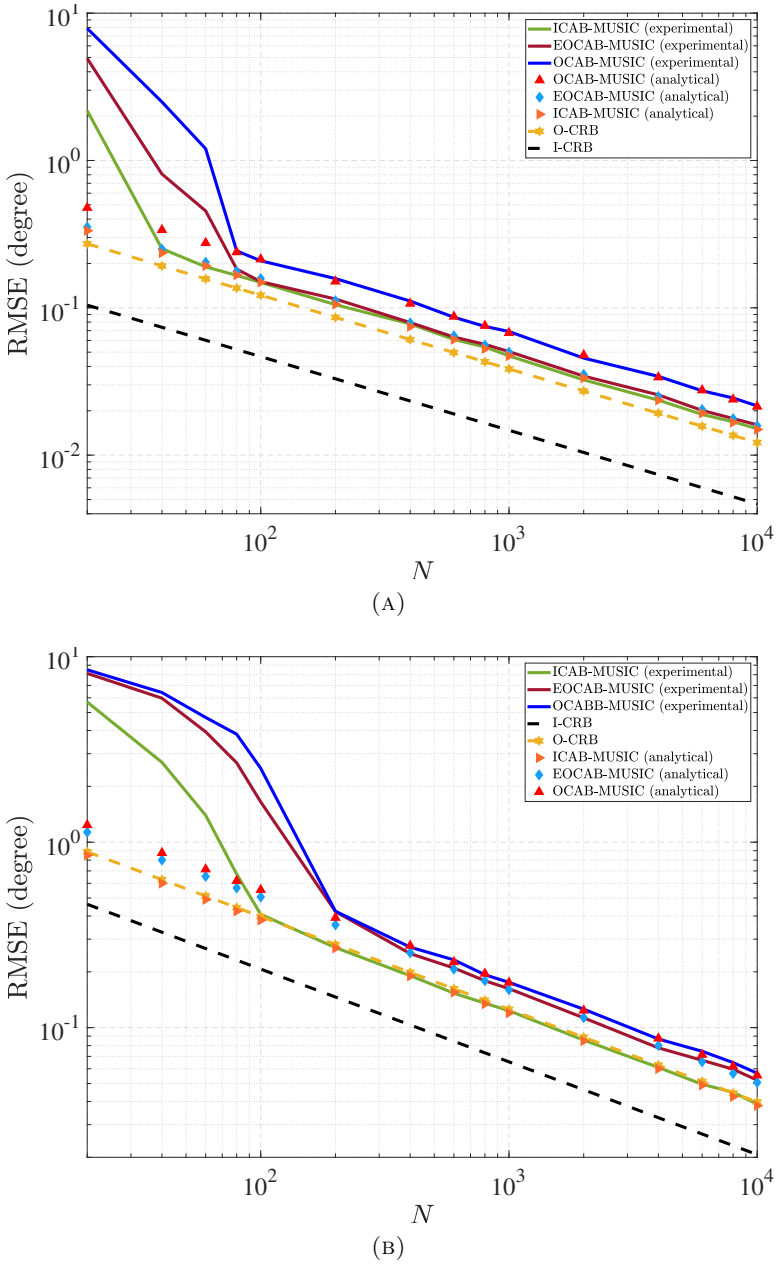


FIGURE 4.2: RMSE in degrees for θ_2 versus the number of snapshots for a nested array with $M = 10$ elements and configuration given in (4.61), SNR = 3 dB, and: (A) $K = 5 < M$; (B) $K = 12 > M$.

4.6.2 MSE vs. the Number of Snapshots

Fig. 4.2 depicts the Root-Mean-Squares-Error (RMSE) for θ_2 in degree versus the number of snapshots when the nested array in (4.61) is used. The SNR is assumed to be 3 dB. In addition, noting $M = 10$, two different scenarios are considered: (A) $K = 5 < M$, and (B) $K = 12 > M$. Fig. 4.2 illustrates a close agreement between the numerical simulations and analytical expression derived for RMSEs of OCAB-MUSIC and EOCAB-MUSIC when about 200 or more snapshots are available. Further, a considerable gap is

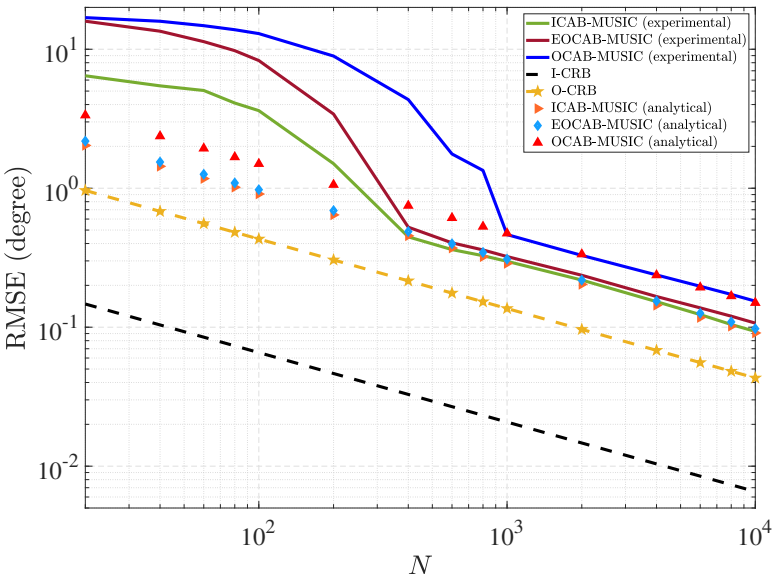


FIGURE 4.3: RMSE in degrees for θ_2 versus N for a nested array with $M = 10$ elements and configuration given in (4.61) when $K = 3$, $\theta_1 = 2°$, $\theta_2 = 3°$, $\theta_3 = 75°$, $\text{SNR}_1 = 20$ dB, $\text{SNR}_2 = 8$ and $\text{SNR}_3 = 22$.

observed between the performance of OCAB-MUSIC and that of the EOCAB-MUSIC. For instance, at $N = 400$, Figs. 4.2a and 4.2b show a performance gain of roughly 3 dB and 1 dB, respectively, in terms of the RMSE when the EOCAB-MUSIC is used. It is also observed that EOCAB-MUSIC performs as well as ICAB-MUSIC when $K = 5 < M$. Further, it is observed that the RMSE of EOCAB-MUSIC is very close to O-CRB when $K = 5 < M$ but we see a gap between them when $K = 12 > M$.

Fig. 4.2 also shows that when a small number of snapshots is available, e.g. less than 1000, all estimators are confronted with substantial performance degradation. This performance loss is justified by the subspace swap arising from the inaccurate estimate of the normalized covariance matrix of $\mathbf{y}(t)$, i.e. $\overline{\mathbf{R}}$, in this case. However, it is seen that the proposed estimator still has superior performance compared to OCAB-MUSIC, even in the low snapshot paradigm.

Fig. 4.3 depicts the RMSE θ_2 in degree versus the number of snapshots when $K = 3$ and the sources powers are unequal. Specifically, It is assumed that $\theta_1 = 2°$, $\theta_2 = 3°$, $\theta_3 = 75°$, $\text{SNR}_1 = 20$ dB, $\text{SNR}_2 = 8$ and $\text{SNR}_3 = 22$. Comparing Fig. 4.2 with Fig. 4.3 reveals that a high difference between the SNRs of the closely-spaced source signals do not have a meaningful impact on the relative asymptotic performance of ICAB-MUSIC, OCAB-MUSIC and EOCAB-MUSIC, however, by increasing the difference between SNRs, OCAB-MUSIC needs more number of snapshots to achieve its asymptotic performance compared to EOCAB-MUSIC and ICAB-MUSIC.

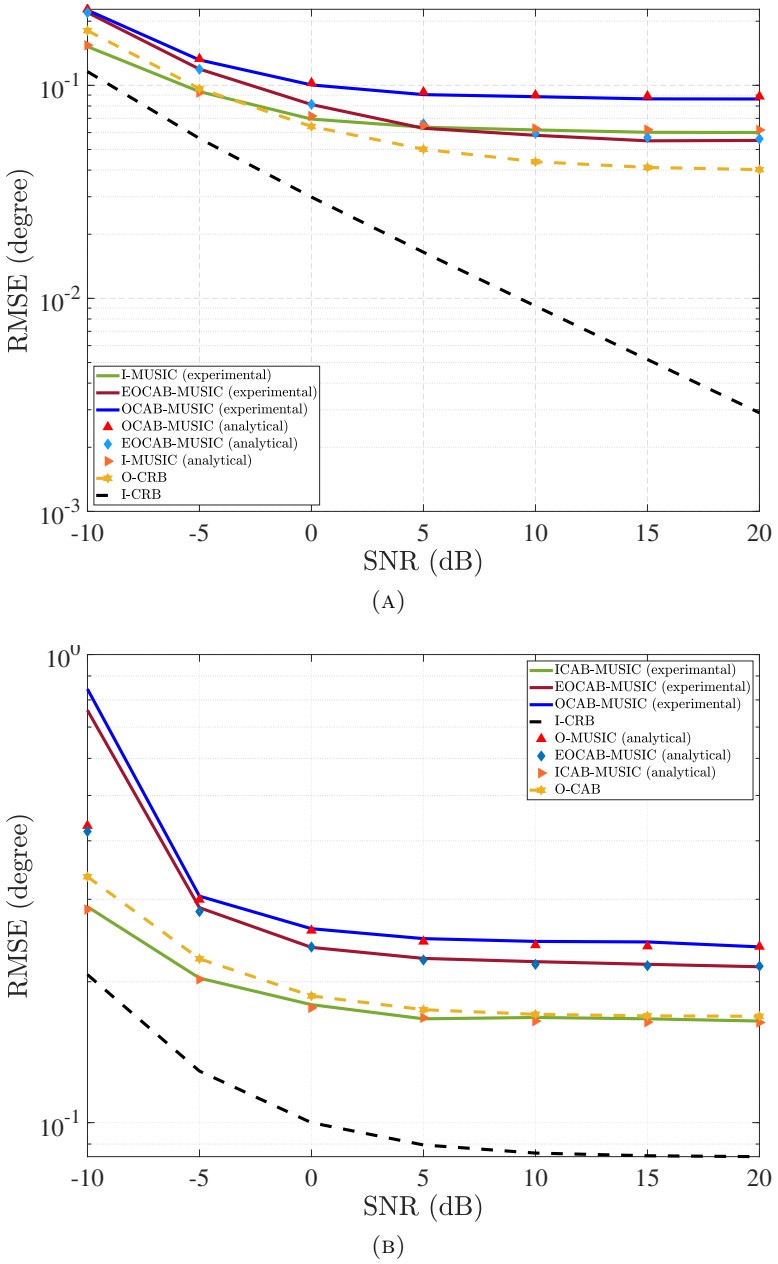
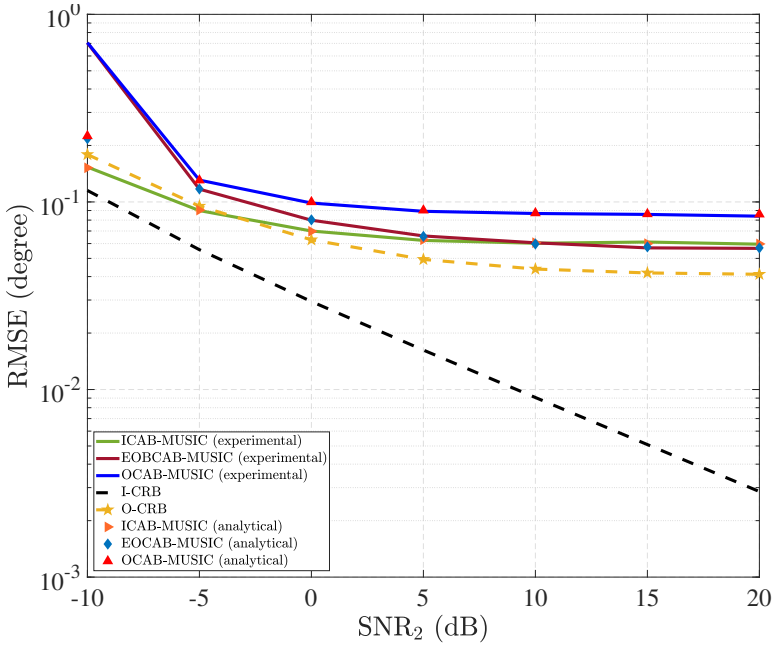


FIGURE 4.4: RMSE in degrees for θ_2 versus SNR for a nested array with $M = 10$ elements and configuration given in (4.61), $N = 500$, and: (A) $K = 5 < M$; (B) $K = 14 > M$.

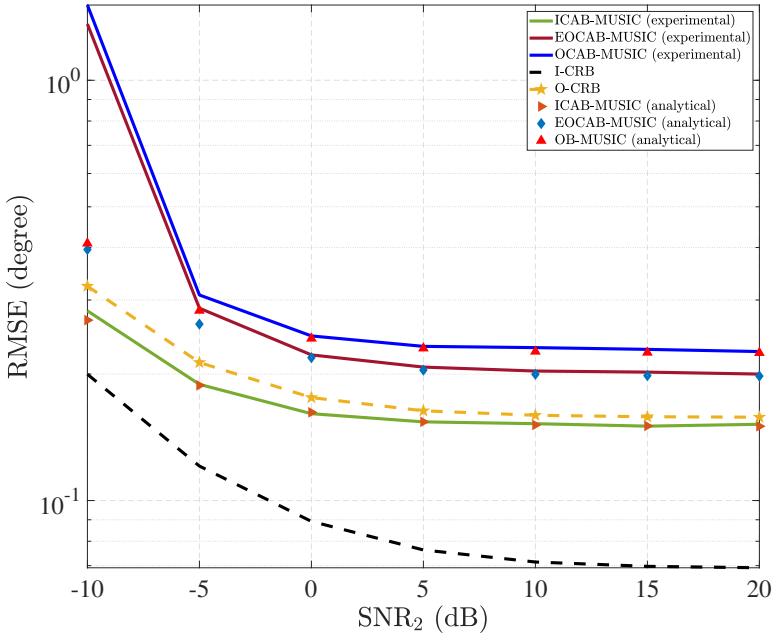
4.6.3 MSE vs. SNR

Fig. 4.4 shows the RMSE for θ_2 in degrees versus SNR for the same setup used for Fig. 4.2. The number of snapshots is considered to be $N = 500$. It is seen in Figs. 4.4a and Fig. 4.4b that the RMSEs of OCAB-MUSIC and EOCAB-MUSIC perfectly match with their asymptotic analytical RMSEs given in Corollary 4.1 and Corollary 4.2.

Fig. 4.4 demonstrates that the I-CRB tends to decay to zero as the SNR increases when $K = 4 < M$ while it gets saturated as the SNR increases when $K = 12 > M$. However,



(A)



(B)

FIGURE 4.5: RMSE in degrees for θ_2 versus SNR wen the source powers are unequal for a nested array with $M = 10$ elements and configuration given in (4.61), $N = 500$, and: (a) $K = 5 < M$; (b) $K = 12 > M$.

as opposed to the I-CRB, O-CRB tends to converge to a constant non-zero value at the high SNR regime for both the cases $K = 4 < M$ and $K = 12 > M$. This behavior of O-CRB was already predicted by Theorem 4.4. In addition, as shown in Theorem 4.6, the RMSEs of OCAB-MUSIC and EOCAB-MUSIC also converge to a constant non-zero value as the SNR increases for both $K = 4 < M$ and $K = 12 > M$.

We observe from Fig. 4.4 that EOCAB-MUSIC preforms better than OCAB-MUSIC in both scenarios $K = 4 < M$ and $K = 12 > M$. For example, at SNR $= 5$, EOCAB-MUSIC leads to performance gains of about 3.7 dB and 1.15 dB in terms of RMSE compared to OCAB-MUSIC. Further, it is seen that EOCAB-MUSIC even outperforms ICAB-MUSIC at high SNR regime when $K = 4 < M$. Another interesting observation is that the RMSE of O-CRB is either better or equal to that of ICAB-MUSIC.

Fig. 4.5 shows the RMSE for θ_2 in degrees versus SNR when the sources powers are unequal and DoAs are not exactly on the grid as opposed to Fig. 4.4. The number of snapshots is considered to be $N = 500$. In case of $K = 5 < M$, the sources are located at $\theta_1 = -49.4551°, \theta_2 = -30.1443°, \theta_3 = -2.4525°, \theta_4 = 26.8293°$ and $\theta_5 = 56.5149°$. Further, the source SNRs are assumed to be $\text{SNR}_1 = 0.75 \times \text{SNR}_2, \text{SNR}_3 = 1.22 \times \text{SNR}_2, \text{SNR}_4 = 0.92 \times \text{SNR}_2$ and $\text{SNR}_5 = 0.66 \times \text{SNR}_2$ while SNR_2 varies from 10 dB to 20 dB as shown in Fig. 4.5a. Further, in case of $K = 12 > M$, the sources are located at $\theta_1 = -56.3351°, \theta_2 = -36.2628°, \theta_3 = -19.9004°, \theta_4 = -2.4093°, \theta_5 = 0.0027°, \theta_6 = 13.1840°, \theta_7 = 23.8495°, \theta_8 = 25.8044°, \theta_9 = 29.2889°, \theta_10 = 40.9107°, \theta_11 = 48.4465°$ and $\theta_12 = 48.5667°$. The source SNRs are assumed to be $\text{SNR}_1 = 1.34 \times \text{SNR}_2, \text{SNR}_3 = 0.84 \times \text{SNR}_2, \text{SNR}_4 = 0.83 \times \text{SNR}_2, \text{SNR}_5 = 0.67 \times \text{SNR}_2, \text{SNR}_6 = 0.69 \times \text{SNR}_2, \text{SNR}_7 = 0.95 \times \text{SNR}_2, \text{SNR}_8 = 0.61 \times \text{SNR}_2, \text{SNR}_9 = 0.79 \times \text{SNR}_2, \text{SNR}_10 = 0.56 \times \text{SNR}_2, \text{SNR}_10 = 0.82 \times \text{SNR}_2$ and $\text{SNR}_12 = 0.88 \times \text{SNR}_2$ while SNR_2 varies from 10 dB to 20 dB as shown in Fig. 4.5b. Comparing Fig. 4.5 with Fig. 4.4 reveals that unequal source powers do not have remarkable impact on the estimation accuracy particularly in high-SNR regime.

4.6.4 CRB vs. the Number of Source Signals

Fig. 4.6 plots the I-CRB and the O-CRB for θ_2 in degree versus the number of source signals for SNR $= 3$ dB and $N = 500$ and different type of SLAs given in (4.61), (4.62), (4.63) and and (4.64). The values of D and v for the different types of arrays are as: 1. MRA: $D = 37$ and $v = 37$; 2. nested array: $D = 30$ and $v = 30$; 3. co-prime array: $D = 26$ and $v = 23$; ULA: $D = 10$ and $v = 10$. Fig. 4.6 indicates that both the I-CRB and the O-CRB increase as the number of source signals increases. Moreover, it is observed that the I-CRB and the O-CRB are quite small for all the SLAs as long as $1 \leq K \leq v - 1$, but they escalate dramatically when K approaches values that are equal to or larger than D. This observation is in compliance with Theorem 4.2 which indicates that the DoA estimation problem is globally identifiable when $1 \leq K \leq v - 1$ and is globally non-identifiable when $K \geq D$.

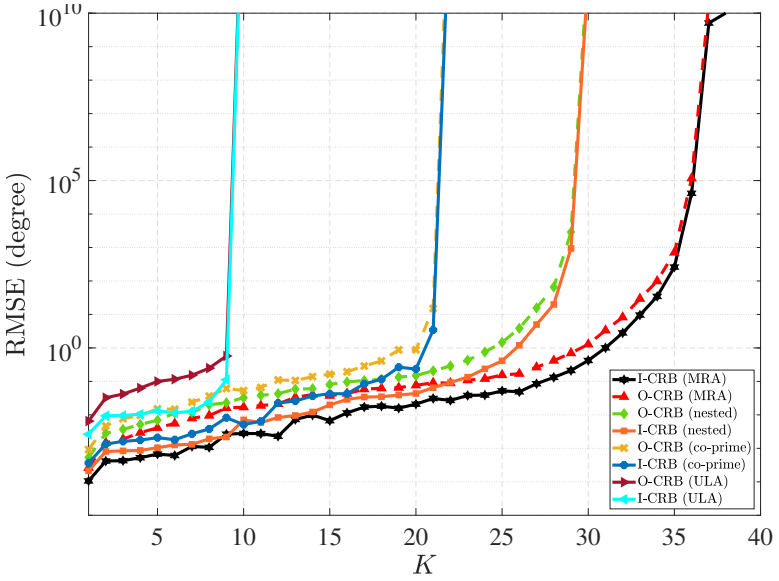


FIGURE 4.6: The CRB versus the number of sources K for various array configurations given from (4.61) to (4.63), $N = 500$ and SNR $= 3$ dB.

4.6.5 Resolution Probability

Fig. 4.7 depicts the probability of resolution versus the source separation for ICAB-MUSIC, EOCAB-MUSIC and OCAB-MUSIC when the nested array given in (4.61) is employed. The number of snapshots and the SNR are considered to be $N = 500$ and 0 dB, respectively. In addition, we consider two sources with equal powers, located at $\theta_1 = 20^\circ - \frac{\Delta\theta}{2}$ and $\theta_2 = 20^\circ + \frac{\Delta\theta}{2}$. We define the two sources as being resolvable if $\max_{i\,\in\,\{1,2\}} |\hat{\theta}_i - \theta_i| < \frac{\Delta\theta}{2}$ [103]. According to this definition and making use of two-dimensional Chebychev's bound [111], the probability of resolution can be lower bounded as

$$\begin{aligned} &\mathbb{P}(\max_{i\,\in\,\{1,2\}} |\hat{\theta}_i - \theta_i| < \frac{\Delta\theta}{2}) \\ &= \mathbb{P}(|\hat{\theta}_1 - \theta_1| < \frac{\Delta\theta}{2}, |\hat{\theta}_2 - \theta_2| < \frac{\Delta\theta}{2}) \geq 1 - \frac{2[\mathcal{E}(\theta_1) + \mathcal{E}(\theta_2)]}{\Delta\theta^2} \\ &+ \frac{2\sqrt{\mathcal{E}_{\theta_1}^2 + \mathcal{E}_{\theta_2}^2 + 2\mathcal{E}_{\theta_1}\mathcal{E}_{\theta_2} - 4\mathcal{E}_{\theta_1,\theta_2}^2}}{\Delta\theta^2}, \end{aligned} \tag{4.69}$$

where $\mathcal{E}(\theta_1)$, $\mathcal{E}(\theta_2)$ and $\mathcal{E}(\theta_1, \theta_2)$ are given in (4.57) and (4.50). The analytical expression on the right-hand side of (4.69) enables us to predict the minimum source separation required for achieving a particular probability of resolution. For example, Fig. 4.6 shows the predicted values for the the minimum source separation to achieve a probability of resolution greater than 0.9, obtained from (4.69), for ICAB-MUSIC, OCAB-MUSIC and EOCAB-MUSIC. It is observed that the predicted values of the minimum source separation for ICAB-MUSIC, EOCAB-MUSIC and OCAB-MUSIC, which are respectively $\Delta\theta = 1.2^\circ$, $\Delta\theta = 1.4^\circ$ and $\Delta\theta = 1.5^\circ$, are in a good agreement with the values obtained

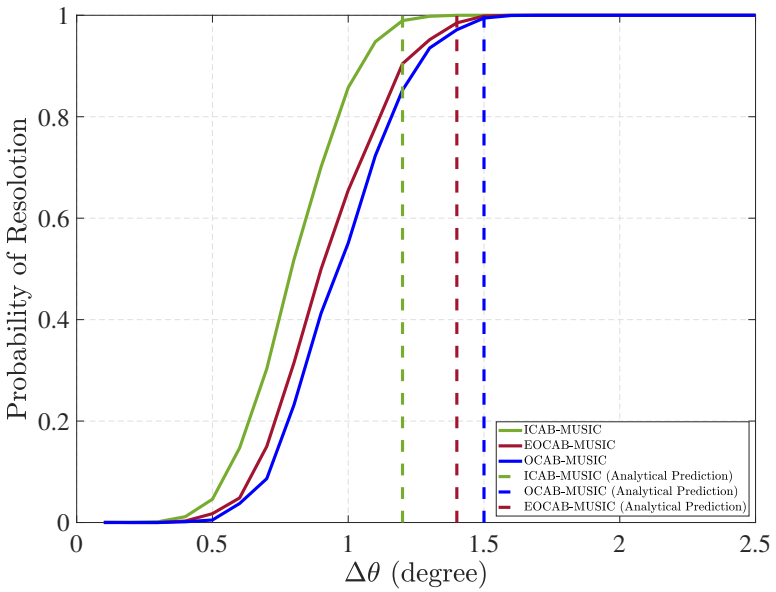


FIGURE 4.7: Probability of resolution versus source separation in degree for a nested array with $M = 10$ elements and configuration given in (4.61), $N = 500$ and SNR $= 0$ dB.

from the numerical simulations, which are respectively $\Delta\theta = 1.1^\circ$, $\Delta\theta = 1.2^\circ$ and $\Delta\theta = 1.3^\circ$. Additionally, Fig. 4.7 demonstrates the resolution performance of EOCAB-MUSIC is superior to that of OCAB-MUSIC while ICAB-MUSIC outperforms both.

4.7 Conclusion

In this paper, we considered the problem of DoA estimation from one-bit measurements received by an SLA. We showed that the idetifiability condition for the DoA estimation problem from one-bit SLA data is equivalent to that for the case when DoAs are estimated from infinite-bit unquantized measurements. Then, we derived a pessimistic approximation of the corresponding CRB. This pessimistic CRB was used as a benchmark for assessing the performance of one-bit DoA estimators. Further, it provides us with valuable insights on the performance limits of DoA estimation from one-bit quantized data. For example, it was shown that the DoA estimation errors in one-bit scenario reduces at the same rate as that of infinite-bit case with respect to the number of samples and, moreover, that the DoA estimation errors in one-bit scenario converges to a constant value by increasing the SNR. We also proposed a new algorithm for estimating DoAs from one-bit quantized data. We investigated the analytical performance of the proposed method through deriving a closed-form expression for the second-order statistics of its asymptotic distribution (for the large number of snapshots) and show that it outperforms

the existing algorithms in the literature. Numerical simulations were provided to validate the analytical derivations and corroborate the improvement in estimation performance.

Chapter 5

DoA Estimation Using Low-Resolution Multi-Bit Sparse Array Measurements

5.1 Introduction

Direction of Arrival (DoA) estimation from Uniform Linear Array (ULA) measurements is extensively studied in the literature [44–46]. However, the number of identifiable sources with ULAs is limited to the number of array elements minus one [3, 46]. Deployment of Sparse Linear Arrays (SLAs), e.g. Minimum Redundancy Arrays (MRAs) [48], co-prime arrays [17] and nested arrays [16], allows for transcending this limitation under the assumption of uncorrelated source signals such that the number of identifiable sources can go considerably beyond the number array elements. A detailed study on the performance of DoA estimation via SLAs has been conducted in [52] through an analysis of the Cramér-Rao Bound (CRB). Further, a variety of algorithms for estimating DoAs from SLA data have been presented in the literature [15, 16, 36, 38, 53, 56, 57, 95].

Most of the algorithms developed for estimating DoAs from SLA measurements are based on the assumption that quantization errors are negligible as a result of using high-resolution Analog-to-Digital Converters (ADCs). However, use of high-resolution ADCs is typically expensive and power-hungry [21]. Hence, to reduce energy consumption and production costs, DoA estimation with binary measurements collected by one-bit ADCs has been recently proposed and discussed in the literature [39–43, 72, 73, 75–78]. One-bit ADCs represent each sample of the analog array observations with only a single bit offering, an exceedingly high sampling rate at a low production cost and very low power consumption [21]. The analytical performance bounds for DoA estimation from

one-bit data have been studied in [32, 33, 112]. Further, a number of one-bit DoA estimators have been provided in [73, 75, 76, 78], which rest on retrieving the covariance matrix of unquantized array observations using the well-known Bussgang theorem [74].

In this paper, as opposed to the previous works which have studied the problem of DoA estimation under two extreme scenarios for analog-to-digital conversion, i.e., infinite-bit quantization and one-bit quantization, we aim to investigate the problem of estimating DoAs from low-resolution few-bit SLA measurements. In such cases, contrary to the one-bit quantization case, the Bussgang theorem may not be directly employed to retrieve the covaraince matrix of array unquantized observations. Instead, we develop a novel optimization-based framework for retrieving the covaraince matrix of unquantized array observations from low-resolution multi-bit measurements. Then, we apply the Co-Array-Based MUSIC (CAB-MUSIC) [16, 38] to the recovered covariance matrix to find the DoAs of interest. The simulation results show that increasing the sampling resolution with a few bits per samples could significantly improve the DoA estimation performance compared to the one-bit sampling case while the power consumption and implementation costs are still much lower than the high-resolution scenario.

Chapter organization: The system model is described in Section 5.2. Section 5.3 presents the proposed algorithm for estimating DoAs from few-bit data. Simulation results are shown and discussed in Section 5.4. Finally, conclusions are drawn in Section 5.5.

5.2 System Model

We consider an SLA with M elements located at positions $m_1\frac{\lambda}{2}, m_2\frac{\lambda}{2}, \cdots, m_M\frac{\lambda}{2}$ with $m_i \in \mathbb{M}$. Here λ denotes the wavelength of the incoming signals and $\mathbb{M}$ is a set of integers with a cardinality of M. It is assumed that K narrowband signals with distinct DoAs $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_K]^T$ impinge on the SLA from far-field. The signal received by the array at time instance t can be modeled as

$$\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}(t) + \mathbf{n}(t) \in \mathbb{C}^{M\times 1}, \quad t = 0, \cdots, N-1, \tag{5.1}$$

where $\mathbf{s}(t) \in \mathbb{C}^{K\times 1}$ denotes the vector of K source signals, $\mathbf{n}(t) \in \mathbb{C}^{M\times 1}$ is additive noise, and $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \cdots, \mathbf{a}(\theta_K)] \in \mathbb{C}^{M\times K}$ represents the SLA steering matrix with

$$\mathbf{a}(\theta_k) = [e^{j\pi \sin\theta_k m_1}, e^{j\pi \sin\theta_k m_2}, \cdots, e^{j\pi \sin\theta_k m_M}]^T, \tag{5.2}$$

being the SLA manifold vector for the k^{th} signal. Further, the following assumptions are made on source and noise signals:

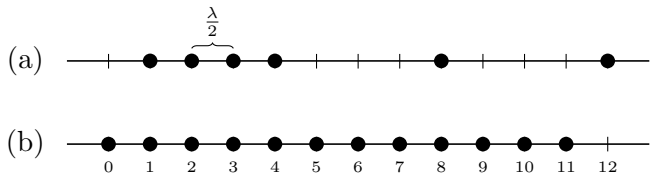


FIGURE 5.1: (a) An SLA with $\mathbb{M} = \{1, 2, 3, 4, 8, 12\}$; (b) corresponding difference co-array with $\mathbb{D} = \{0, 1, \cdots, 11\}$.

A1 $\mathbf{n}(t)$ follows a zero-mean circular complex Gaussian distribution with the covariance matrix $\mathbb{E}\{\mathbf{n}(t)\mathbf{n}^H(t)\} = \sigma^2\mathbf{I}_M$.

A2 The source signals are modeled as zero-mean *uncorrelated* circular complex Gaussian random variables with covariance matrix $\mathbb{E}\{\mathbf{s}(t)\mathbf{s}^H(t)\} = \text{diag}(\mathbf{p})$ where $\mathbf{p} = [p_1, p_2, \cdots, p_K]^T \in \mathbb{R}_{>0}^{K\times 1}$ (i.e., $p_k > 0,\ \forall k$).

A3 No temporal correlation is assumed between the snapshots, i.e., $\mathbb{E}\{\mathbf{n}(t_1)\mathbf{n}^H(t_2)\} = \mathbb{E}\{\mathbf{s}(t_1)\mathbf{s}^H(t_2)\} = \mathbf{0}$ when $t_1 \neq t_2$ and $\mathbf{0}$ is an all-zero matrix of appropriate dimensions.

Based on the above assumptions, the covariance matrix of $\mathbf{y}(t)$ is given by

$$\mathbf{R} = \mathbb{E}\{\mathbf{y}(t)\mathbf{y}^H(t)\} = \mathbf{A}(\boldsymbol{\theta})\text{diag}(\mathbf{p})\mathbf{A}^H(\boldsymbol{\theta}) + \sigma^2\mathbf{I}_M \in \mathbb{C}^{M\times M}. \tag{5.3}$$

It is readily verified that $\mathbf{R}$ is a structured matrix with only $2D-1$ free parameters where $D = |\mathbb{D}|$ with $\mathbb{D} = \{|m_p - m_q| : m_p, m_q \in \mathbb{M}\}$. The set $\mathbb{D}$ is called the difference co-array [52, 53]. Noticing the structure in $\mathbf{R}$, it can be rewritten as follows

$$\mathbf{R}(\mathbf{u}) = u_0\mathbf{L}_0 + \sum_{n=1}^{D-1} u_n\mathbf{L}_n + \sum_{n=1}^{D-1} u_n^*\mathbf{L}_n^T, \tag{5.4}$$

where $u_0 = \sigma^2 + \sum_{k=1}^K p_k$, $u_n = \sum_{k=1}^K p_k e^{j\pi \sin\theta_k \ell_n}$ and

$$[\mathbf{L}_n]_{p,q} = \begin{cases} 1, & \text{if } \ m_p - m_q = \ell_n, \\ 0, & \text{otherwise}, \end{cases} \tag{5.5}$$

with $\ell_n \in \mathbb{D}$, $m_p, m_q \in \mathbb{M}$, $1 \leq p, q \leq M$ and $0 \leq n \leq D-1$. A proper design of SLA allows for identifying more uncorrelated source signals than the number of array elements by exploiting the resulting structure of $\mathbf{R}$ efficiently[16, 17, 52, 53]. Fig. 5.1 illustrates an SLA along with its difference co-array.

In the classical mode, the received signals are sampled at Nyquist rate and processed assuming full-precision analog-to-digital conversion. On the other hand, herein, we assume that each array sensor is equipped with a low-resolution multi-bit ADC converting the

received analog signal into digital data using q bits per sample. A generic q-bit ADC has $2^q - 1$ threshold levels where $\alpha_1 < \alpha_2 < \cdots < 0 < \cdots < \alpha_{2^q} < \alpha_{2^q-1}$. The q-bit ADC at the m^{th} array element transforms the real and imaginary parts of $[\mathbf{y}(t)]_m$ into one of the $2^q - 1$ prescribed qunatization levels $\{\gamma_1, \gamma_2, \cdots, \gamma_{2^q-1}\}$ by comparing them individually with the threshold levels. Particularly, the q-bit quantized output signal at the m^{th} array element is expressed as

$$[\mathbf{x}(t)]_m = Q([\Re\{\mathbf{y}(t)\}]_m) + jQ([\Im\{\mathbf{y}(t)\}]_m), \tag{5.6}$$

where $Q(.)$ denotes the q-bit quantization operation defined as

$$Q(a) = \gamma_h \quad \text{if} \quad \alpha_h \leq a < \alpha_{h+1}. \tag{5.7}$$

We are interested in estimating DoAs from q-bit quantized output signals of the SLA, i.e., $\mathbf{X} = \Big[\mathbf{x}(0), \quad \mathbf{x}(1), \quad \cdots, \quad \mathbf{x}(N-1)\Big]$.

5.3 Multi-Bit DoA estimation with Sparse Arrays

In this section, we first formulate an optimization problem whose solution provides us with an estimate of the covariance matrix of $\mathbf{y}(t)$, i.e., $\mathbf{R}$, using q-bit quantized array measurements, i.e., $\mathbf{X}$. Subsequently, we apply the CAB-MUSIC [16, 38] to obtain DoA estimates from the estimate of $\mathbf{R}$.

It follows from (5.4) that $\mathbf{R}$ is fully described by the complex vector $\mathbf{u} = [u_0, u_1, \cdots, u_{D-1}]^T$. Hence, for a given $\mathbf{Y} = \Big[\mathbf{y}(0), \quad \mathbf{y}(1), \quad \cdots, \quad \mathbf{y}(N-1)\Big]$, $\mathbf{R}$ can be obtained from the solution of the following optimization problem

$$\begin{aligned} \underset{\mathbf{u}}{\text{minimize}} \quad & \|\mathbf{R}(\mathbf{u}) - \mathbf{Y}\mathbf{Y}^H\|^2 \\ \text{subject to} \quad & \mathbf{R}(\mathbf{u}) \succeq \mathbf{0}. \end{aligned} \tag{5.8}$$

However, $\mathbf{Y}$ is unknown here, and instead, we only have access to its q-bit quantized values, i.e., $\mathbf{X}$. It follows from (5.6) and (5.7) that each element of the observation matrix $\mathbf{X}$ determines a lower and an upper bound for the real and imaginary parts of the corresponding element in $\mathbf{Y}$. Putting these lower and upper bound into the matrices $\mathbf{\Gamma}_l$

and $\mathbf{\Gamma}_u$, an optimization problem for joint estimation of $\mathbf{u}$ and $\mathbf{Y}$ can be cast as follows:

$$\begin{aligned} \underset{\mathbf{u},\mathbf{Y}}{\text{minimize}} \quad & \|\mathbf{R}(\mathbf{u}) - \mathbf{Y}\mathbf{Y}^H\|_F^2 \\ \text{subject to} \quad & \mathbf{R}(\mathbf{u}) \succeq \mathbf{0}, \\ & \text{vec}(\Re\{\mathbf{Y}\} - \Re\{\mathbf{\Gamma}_l\}) \geq \mathbf{0}, \\ & \text{vec}(\Im\{\mathbf{Y}\} - \Im\{\mathbf{\Gamma}_l\}) \geq \mathbf{0}, \\ & -[\text{vec}(\Re\{\mathbf{Y}\} - \Re\{\mathbf{\Gamma}_u\})] \geq \mathbf{0}, \\ & -[\text{vec}(\Im\{\mathbf{Y}\} - \Im\{\mathbf{\Gamma}_u\})] \geq \mathbf{0}. \end{aligned} \tag{5.9}$$

where the last four constraints in (5.9) aim to enforce the consistency of $\mathbf{Y}$ with the q-bit measurements by ensuring that the elements of $\mathbf{Y}$ lie in the regions determined by the observation matrix $\mathbf{X}$. The above optimization problem is non-convex as its objective is a quartic function with respect to $\mathbf{Y}$. In what follows, we first present an equivalent reformulation for (5.9), which paves the way for iteratively solving this non-convex optimization problem.

Theorem 5.1. *Consider slack variables* $\mathbf{G} \in \mathbb{C}^{(M+N)\times M}$, $\mathbf{W} \in \mathbb{C}^{M\times M}$ *and* $\xi \in \mathbb{R}$. *The optimization problem* (5.9) *is equivalent to*

$$\begin{aligned} \underset{\mathbf{u},\mathbf{Y},\mathbf{W},\mathbf{G},\xi}{\textit{minimize}} \quad & \|\mathbf{R}(\mathbf{u}) - \mathbf{W}\|_F^2 + \eta\xi \\ \textit{subject to} \quad & \mathbf{R}(\mathbf{u}) \succeq \mathbf{0}, \\ & \text{vec}(\Re\{\mathbf{Y}\} - \Re\{\mathbf{\Gamma}_l\}) \geq \mathbf{0}, \\ & \text{vec}(\Im\{\mathbf{Y}\} - \Im\{\mathbf{\Gamma}_l\}) \geq \mathbf{0}, \\ & -[\text{vec}(\Re\{\mathbf{Y}\} - \Re\{\mathbf{\Gamma}_u\})] \geq \mathbf{0}, \\ & -[\text{vec}(\Im\{\mathbf{Y}\} - \Im\{\mathbf{\Gamma}_u\})] \geq \mathbf{0}, \\ & \mathbf{T} \succeq \mathbf{0}, \\ & \xi\mathbf{I}_M - \mathbf{G}^H\mathbf{T}\mathbf{G} \succeq \mathbf{0}, \\ & \mathbf{G}^H\mathbf{G} = \mathbf{I}_M, \end{aligned} \tag{5.10}$$

where $\mathbf{T} = \begin{bmatrix} \mathbf{I}_N & \mathbf{Y}^H \\ \mathbf{Y} & \mathbf{W} \end{bmatrix} \in \mathbb{C}^{(M+N)\times(M+N)}$ *and* η *is a regularization parameter.*

Proof. Consider the slack variables $\mathbf{W}=\mathbf{Y}\mathbf{Y}^H$. Then it is readily seen that the optimization problem (5.9) is equivalent to:

$$\begin{aligned}
\underset{\mathbf{u},\mathbf{Y},\mathbf{W}}{\text{minimize}} \quad & \|\mathbf{R}(\mathbf{u})-\mathbf{W}\|_F^2 \\
\text{subject to} \quad & \mathbf{R}(\mathbf{u}) \succeq \mathbf{0}, \\
& \text{vec}(\Re\{\mathbf{Y}\}-\Re\{\mathbf{\Gamma}_l\}) \geq \mathbf{0}, \\
& \text{vec}(\Im\{\mathbf{Y}\}-\Im\{\mathbf{\Gamma}_l\}) \geq \mathbf{0}, \\
& -[\text{vec}(\Re\{\mathbf{Y}\}-\Re\{\mathbf{\Gamma}_u\})] \geq \mathbf{0}, \\
& -[\text{vec}(\Im\{\mathbf{Y}\}-\Im\{\mathbf{\Gamma}_u\})] \geq \mathbf{0}, \\
& \mathbf{W}=\mathbf{Y}\mathbf{Y}^H.
\end{aligned} \tag{5.11}$$

It is readily confirmed that $\mathbf{W}=\mathbf{Y}\mathbf{Y}^H$ if and only if $\text{rank}(\mathbf{W}-\mathbf{Y}\mathbf{Y}^H)=0$. Further, $\text{rank}(\mathbf{W}-\mathbf{Y}\mathbf{Y}^H)=0$ can be equivalently expressed as $\text{rank}(\mathbf{I}_N)+\text{rank}(\mathbf{W}-\mathbf{Y}\mathbf{Y}^H)=N$. Since $\mathbf{I}_N$ is positive definite, it follows from the Guttman rank additivity formula [113] that $\text{rank}(\mathbf{I}_N)+\text{rank}(\mathbf{W}-\mathbf{Y}\mathbf{Y}^H)=\text{rank}(\mathbf{T})$. Moreover, it follows from $\mathbf{W}-\mathbf{Y}\mathbf{Y}^H=\mathbf{0}$ and $\mathbf{I}_N \succ \mathbf{0}$ that $\mathbf{T}$ has to be positive semi-definite. These imply that the equality constraint in (5.11) can be replaced with a rank constraint on a semi-definite matrix. Hence, the optimization problem (5.11), and equivalently (5.9), can be recast as follows:

$$\begin{aligned}
\underset{\mathbf{u},\mathbf{Y},\mathbf{W}}{\text{minimize}} \quad & \|\mathbf{R}(\mathbf{u})-\mathbf{W}\|^2 \\
\text{subject to} \quad & \mathbf{R}(\mathbf{u}) \succeq \mathbf{0}, \\
& \text{vec}(\Re\{\mathbf{Y}\}-\Re\{\mathbf{\Gamma}_l\}) \geq \mathbf{0}, \\
& \text{vec}(\Im\{\mathbf{Y}\}-\Im\{\mathbf{\Gamma}_l\}) \geq \mathbf{0}, \\
& -[\text{vec}(\Re\{\mathbf{Y}\}-\Re\{\mathbf{\Gamma}_u\})] \geq \mathbf{0}, \\
& -[\text{vec}(\Im\{\mathbf{Y}\}-\Im\{\mathbf{\Gamma}_u\})] \geq \mathbf{0}, \\
& \mathbf{T} \succeq \mathbf{0}, \\
& \text{rank}(\mathbf{T})=N.
\end{aligned} \tag{5.12}$$

The constraint $\text{rank}(\mathbf{T})=N$ in (5.12) is equivalent to imposing the constraint that the M smallest eigenvalues of $\mathbf{T}$ are all zero. This constraint on the M smallest eigenvalues of $\mathbf{T}$ can be formulated by introducing the new slack variables $\mathbf{G}\in\mathbb{C}^{(M+N)\times M}$ and $\xi\in\mathbb{R}$ with $\mathbf{G}^H\mathbf{G}=\mathbf{I}_M$. Indeed, in what follows, we will show the M smallest eigenvalues of T are all zero if $\xi\mathbf{I}_M-\mathbf{G}^H\mathbf{T}\mathbf{G}\succeq\mathbf{0}$ and $\xi\to 0$. Let $\rho_1\leq\rho_2\leq\cdots\leq\rho_{M+N}$ and $\nu_1\leq\nu_2\leq\cdots\leq\nu_M$ denote the eigenvalues of $\mathbf{T}$ and $\mathbf{G}^H\mathbf{T}\mathbf{G}$, respectively. From $\xi\mathbf{I}_M-\mathbf{G}^H\mathbf{T}\mathbf{G}\succeq\mathbf{0}$, we have $\nu_i\leq\xi$ for $i=1,2,\cdots,M$. Additionally, it follows from [114, Corollary 4.3.16] that $0\leq\rho_i\leq\nu_i$ for $i=1,2,\cdots,M$. Hence, we observe that

$$\mathbf{0}\preceq\text{diag}([\rho_1,\rho_2,\cdots,\rho_M]^T)\preceq\text{diag}([\nu_1,\nu_2,\cdots,\nu_M]^T)\preceq\xi\mathbf{I}_M. \tag{5.13}$$

It easily observed from (5.13) that $\xi\to 0$ leads the M smallest eigenvalues of $\mathbf{T}$ to go to

zero. Accordingly, we can deduce that, by properly selecting η in (5.13) such that the optimum value of ξ goes to zero, the constraints $\xi\mathbf{I}_M - \mathbf{G}^H\mathbf{T}\mathbf{G} \succeq \mathbf{0}$ and $\mathbf{G}^H\mathbf{G} = \mathbf{I}_M$ in (5.13) will be equivalent to the rank constraint in (5.12). This implies that (5.13) is equivalent to (5.12) and thus to (5.9). This completes the proof. □

The optimization problem (5.10) can be solved iteratively by alternating between $\mathbf{G}$ and the other parameters, i.e., $\mathbf{u}$, $\mathbf{Y}$, $\mathbf{W}$ and ξ. Let $\mathbf{G}^{(k)}$, $\mathbf{u}^{(k)}$, $\mathbf{Y}^{(k)}$, $\mathbf{W}^{(k)}$ and $\xi^{(k)}$ be the values of the parameters $\mathbf{G}$, $\mathbf{u}$, $\mathbf{Y}$, $\mathbf{W}$ and ξ at the k-th iteration, respectively. Given $\mathbf{G}^{(k-1)}$, the optimization problem with respect to $\mathbf{u}$, $\mathbf{Y}$, $\mathbf{W}$ and ξ at the k-th iteration becomes

$$\begin{aligned}
\underset{\mathbf{u}^{(k)},\mathbf{Y}^{(k)},\mathbf{W}^{(k)},\xi^{(k)}}{\text{minimize}} \quad & \|\mathbf{R}(\mathbf{u}^{(k)}) - \mathbf{W}^{(k)}\|_F^2 + \eta\xi^{(k)} \\
\text{subject to} \quad & \mathbf{R}(\mathbf{u}^{(k)}) \succeq \mathbf{0}, \\
& \text{vec}(\Re\{\mathbf{Y}^{(k)}\} - \Re\{\mathbf{\Gamma}_l\}) \geq \mathbf{0}, \\
& \text{vec}(\Im\{\mathbf{Y}^{(k)}\} - \Im\{\mathbf{\Gamma}_l\}) \geq \mathbf{0}, \\
& -\left[\text{vec}(\Re\{\mathbf{Y}^{(k)}\} - \Re\{\mathbf{\Gamma}_u\})\right] \geq \mathbf{0}, \\
& -\left[\text{vec}(\Im\{\mathbf{Y}^{(k)}\} - \Im\{\mathbf{\Gamma}_u\})\right] \geq \mathbf{0}, \\
& \mathbf{T}^{(k)} \succeq \mathbf{0}, \\
& \xi^{(k)}\mathbf{I}_M - \mathbf{G}^{(k-1)^H}\mathbf{T}^{(k)}\mathbf{G}^{(k-1)} \succeq \mathbf{0}, \\
& \xi^{(k)} \leq \xi^{(k-1)}.
\end{aligned} \tag{5.14}$$

Once $\mathbf{T}^{(k)}$, $\mathbf{u}^{(k)}$ and $\xi^{(k)}$ are found by solving (5.14), $\mathbf{G}^{(k)}$ can be obtained by seeking an $(M+N)\times M$ matrix with orthonormal columns such that $\mathbf{G}^{(k)^H}\mathbf{T}^{(k)}\mathbf{G}^{(k)} \preceq \xi^{(k)}\mathbf{I}_M$. Choosing $\mathbf{G}^{(k)}$ to be equal to the matrix composed of the eigenvectors of $\mathbf{T}^{(k)}$ corresponding to its M smallest eigenvalues, and following similar arguments provided after (5.12), we have

$$\begin{aligned}
&\mathbf{G}^{(k)^H}\mathbf{T}^{(k)}\mathbf{G}^{(k)} = \text{diag}([\rho_1^{(k)}, \rho_2^{(k)}, \cdots, \rho_M^{(k)}]^T) \\
&\preceq \text{diag}([\nu_1^{(k-1)}, \nu_2^{(k-1)}, \cdots, \nu_M^{(k-1)}]^T) \preceq \xi^{(k)}\mathbf{I}_M,
\end{aligned} \tag{5.15}$$

where $\rho_1^{(k)} \leq \rho_2^{(k)} \leq \cdots \leq \rho_{M+N}^{(k)}$ and $\nu_1^{(k-1)} \leq \nu_2^{(k-1)} \leq \cdots \leq \nu_M^{(k-1)}$ denote the eigenvalues of $\mathbf{T}^{(k)}$ and $\mathbf{G}^{(k-1)^H}\mathbf{T}^{(k)}\mathbf{G}^{(k-1)}$, respectively. It follows from (5.15) that the matrix composed of the eigenvectors of $\mathbf{T}^{(k)}$ corresponding to its M smallest eigenvalue is a right choice of $\mathbf{G}^{(k)}$. Accordingly, at each iteration of the proposed algorithm, we need to solve a Semi-Definite Program (SDP), which can be solved efficiently, followed by an Eigenvalue Decomposition (ED). The alternating optimization procedure is repeated until either the objective or the optimization variables converge to a constant value. Algorithm 2 summarizes the steps of the aforementioned iterative approach to solving (5.9). Further, to initialize the algorithm, $\mathbf{G}^{(0)}$ can be found through the ED of $\mathbf{T}^{(0)}$

Algorithm 2 Covarinace Matrix Estimatiton from Low-Resolution Few-Bit Data

Input: The problem information $\mathbf{\Gamma}_l$, $\mathbf{\Gamma}_u$, η, ϵ_1, ϵ_2, ϵ_3 and ϵ_4.
Output: The estimate of the covariance matrix of the full-precision data.
1: **initialization:** Set $k = 0$ and obtain $\mathbf{G}^{(0)}$ by dropping the rank constraint.
2: **while** $\|\mathbf{u}^{(k)} - \mathbf{u}^{(k-1)}\|_2 \geq \epsilon_1$, $\|\mathbf{W}^{(k)} - \mathbf{W}^{(k-1)}\|_{\mathrm{F}} \geq \epsilon_2$, $\|\mathbf{Y}^{(k)} - \mathbf{Y}^{(k-1)}\|_{\mathrm{F}} \geq \epsilon_3$ and $\xi^{(k)} \geq \epsilon_4$ **do**
3: Increase k by one.
4: Find $\mathbf{u}^{(k)}$, $\mathbf{W}^{(k)}$, $\mathbf{Y}^{(k)}$ and $\xi^{(k)}$ by solving (5.14).
5: Compute the ED of $\mathbf{T}^{(k)}$.
6: $\mathbf{G}^{(k)}$ equals the matrix composed of the eigenvectors of $\mathbf{T}^{(k)}$ corresponding to its M smallest eigenvalues.
7: **end while**

obtained from solving (5.12) without considering the rank constraint. We note that the proposed algorithm, which is based on alternating optimization method, is guaranteed to converge to at least a local minimum of (5.10) [115]. Once $\mathbf{R}$ is retrieved from Algorithm 2, the CAB-MUSIC [16, 38] is applied to the retrieved $\mathbf{R}$ to estimate DoAs.

5.4 Simulation Results

In this section, numerical results are provided for assessing the performance of the proposed algorithm for estimating DoAs from low-resolution few-bit SLA output. In all experiments, each simulated point has been computed by 1000 Monte Carlo repetitions over noise realizations. In addition, the K independent sources with an equal power p are equispaced in the angular domain $[-60°, 60°]$ with respect to a 8-sensor nested array with $\mathbb{M} : \{1, 2, 3, 4, 5, 10, 15, 20\}$. The SNR is also defined as $10 \log \frac{p}{\sigma^2}$.

Fig. 5.2 depicts the Root-Mean-Squares-Error (RMSE) for θ_2 in degree versus SNR for different bit-width when $N = 300$, $M = 8$ and: (a) $K = 4 < M$; (b) $K = 10 > M$. Fig. 5.2 demonstrates that increasing the number of quantization bits from one to two and then to four leads to a considerable performance improvement. Further, it is observed that the RMSE of 4-bit DoA estimation is very close to that of DoA estimates obtained from the unquantized array observations. For instance, when $K = 4$, the performance loss arising from quantization, defined as $10 \log(\mathrm{RMSE}_{\mathrm{quantized}}/\mathrm{RMSE}_{\mathrm{unquantized}})$, at SNR $= 5$ dB are about 3.33 and 1.39 dB in case of 1-bit and 2-bit quantization, respectively, while it is almost zero in case of 4-bit quantization. However, the implementation costs and power consumption of 4-bit and 2-bit ADCs are still much lower compared to high-resolution ADCs. For example, at sampling frequency of 10 MHz, a 14-bit ADC consumes roughly 10^3 times more power than 2-bit and 4-bit ADCs [116]. The gap between the power consumption of low- and high-resolution ADCs further increases with higher sampling frequencies, e.g. at sampling frequency of 1 GHz, a 14-bit ADC consumes roughly 10^5 times more power than 2-bit and 4-bit ADCs [116]. Further, it is

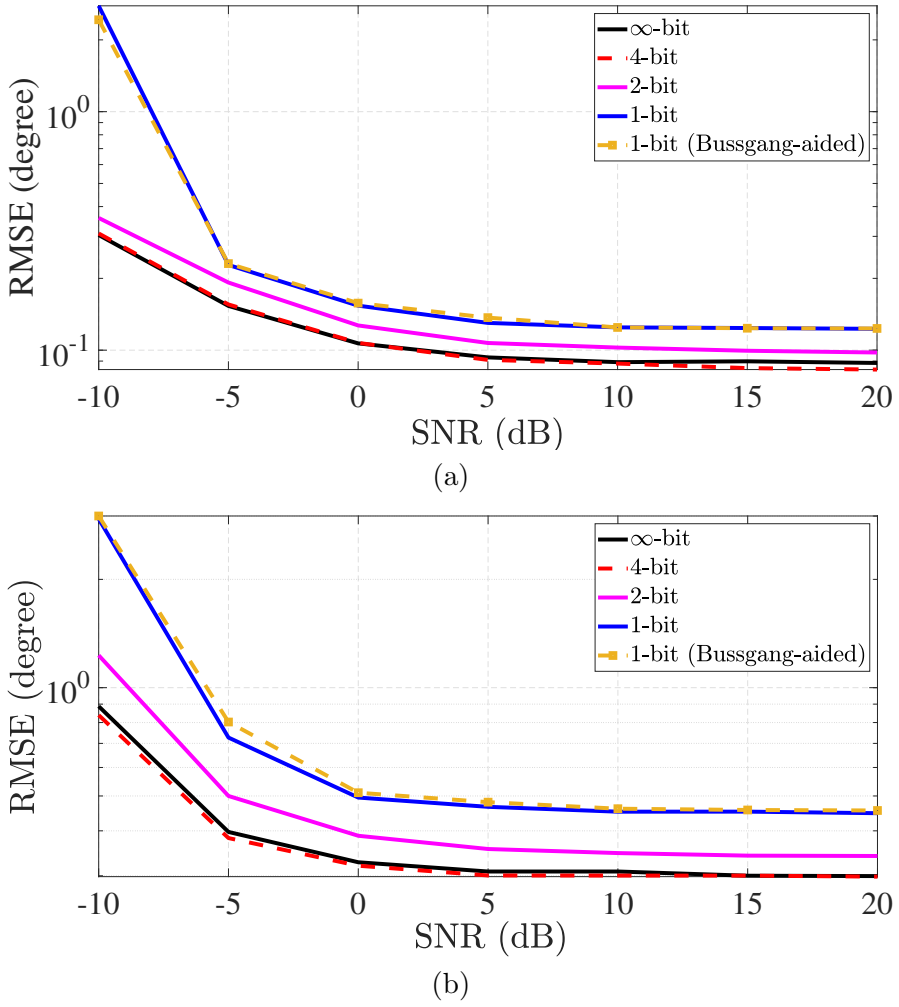


FIGURE 5.2: RMSE in degree for θ_2 versus SNR for a 8-sensor nested array with $\mathbb{M} : \{1, 2, 3, 4, 5, 10, 15, 20\}$., $N = 300$, and: (a) $K = 4 < M$; (b) $K = 10 > M$.

relatively easy to implement 4-bit and 2-bit ADCs even at very high sampling frequencies while implementation feasibility of high-resolution ADCs moves from difficult at sampling frequencies of ~ 1 MHz to infeasible beyond those sampling frequencies [116]. Moreover, it is seen that the proposed method in case of 1-bit quantization performs as well as the one-bit DoA estimator in [75], which relies on estimating the covariance matrix of unquantized array observations directly from one-bit data using the Bussgang theorem.

Fig. 5.3 plots the RMSE for θ_2 in degree versus the number of snapshots for SNR $= 0$ dB and: (a) $K = 4 < M$, and (b) $K = 10 > M$. Fig. 5.3 shows that, to achieve an RMSE of 0.1 for example, infinite-bit, 4-bit, 2-bit and one-bit cases need 300, 300, 500 and 800 samples when $K = 4$, respectively. This indicates that the total number of bits required to achieve an RMSE of 0.1 is, respectively, 1200, 1000 and 800 bits for 4-bit, 2-bit and one-bit sampling scenarios.

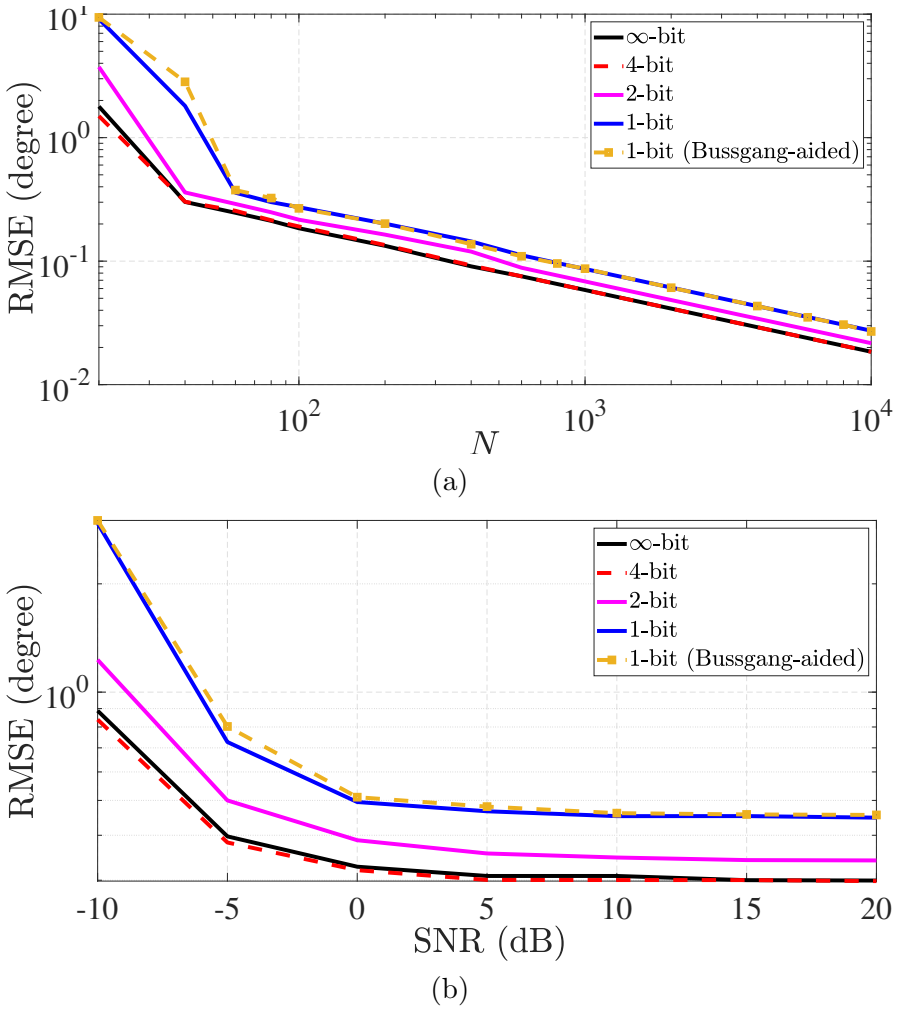


FIGURE 5.3: RMSE in degree for θ_2 versus the number of snapshots for a 8-sensor nested array with $\mathbb{M} : \{1, 2, 3, 4, 5, 10, 15, 20\}$., SNR $= 0$ dB, and: (a) $K = 4 < M$; (b) $K = 10 > M$.

5.5 Conclusion

The problem of DoA estimation from low-resolution few-bit SLA data was investigated. Firstly, the covariance matrix of unquantized array observations was retrieved from low-resolution few-bit SLA data by employing an iterative optimization-based algorithm. Then, DoAs were estimated by applying CAB-MUSIC to the recovered covariance matrix of unquantized array observations. The simulation results showed that increasing the sampling resolution to 2 or 4 bits per samples could significantly increase the DoA estimation performance compared to the one-bit sampling case while the power consumption and implementation costs are still much lower than the high-resolution sampling scenario.

Part II

Distributed Target Localization

Chapter 6

Localization with One-Bit Passive Radars in Narrowband Internet-of-Things using Multivariate Polynomial Optimization

6.1 Introduction

Recent industry estimates project that nearly 75 billion devices will be connected in the Internet-of-Things (IoT) by the year 2025 [117]. The IoT is envisioned to connect the physical and digital world through extensive instrumentation with sensing, wearable, and intelligent devices [118]. A common IoT application is to provide various *localization-based services* [119, 120], wherein a large network of devices collects and transmits data to determine the position of entities-of-interest with respect to a node or sensor within the IoT. The location information is critical in order to gather crucial inference from physical measurements in applications such as military surveillance [121], physiological sensors [122], smart homes [123], disaster response [124], and environmental monitoring [125].

Global Positioning System (GPS) devices are quite reliable in providing localization measurements in other applications. However, GPS deployment at every IoT node is very expensive in terms of cost and power, especially for networks with massive number of devices. Further, GPS performs poorly in indoor environments. Therefore, many

alternative IoT localization methods have been proposed in recent studies [120]. A promising technology is passive sensor tags that augment existing IoT deployments through backscatter communications [126]. These tags do not have any *active* radio-frequency (RF) chain components thereby leading to huge savings in cost and energy. This is also a practical approach because it is difficult to re-purpose the preset IoT network sensing modalities (usually fixed before the deployment), especially when it comprises millions of devices [127]. On the other hand, addition of passive sensors does not require changing the deployed IoT hardware or placement of new communications and power sources [128].

Since the IoT framework is defined by a massive number of largely battery-powered devices, that also transmit or receive data, the underlying challenges for any communications link in this setting are low power, low data rate, wide coverage, and scalability [129]. In this context, the 3rd generation partnership project (3GPP) recently introduced narrowband IoT (NB-IoT) system specifications to support wide coverage area, long user lifetime, and low power/cost devices over a narrow bandwidth of 180 kHz [130]. While not fully backward compatible with existing 3GPP devices, the NB-IoT harmoniously coexists with legacy networks by reusing the functionalities of the latter's design. The reduced NB-IoT bandwidth implies higher transmit power spectral density within the existing 3GPP specifications. This, combined with a soft re-transmission strategy [131], enhances the coverage of NB-IoT over conventional IoT solutions. The ultra-low complexity and low power consumption features of NB-IoT are advantageous for location-based services such as smart parking, smart tracking, and smart home [132]. In this paper, we focus on passive localization in NB-IoT networks.

While NB-IoT networks benefit from low bandwidth to enhance their coverage, the same feature imposes challenges in localization by severely limiting the data rate. Commonly used ranging-based localization techniques lose accuracy because of low data rates [133]. In NB-IoT devices, low battery-power is insufficient to handle high sampling rates required to attain necessary localization accuracy [134–137]. A popular alternative NB-IoT localization technique is to employ *fingerprinting*, wherein the received signal strength indicator (RSSI) measurements are collected at specified locations during the training phase and then compared with online measurements to determine the location of the target [90, 134]. However, this approach requires prior knowledge of a detailed RSSI database which may be unavailable or unattainable. Hence, recent NB-IoT studies explore RSSI-independent signal processing methods such as successive interference cancellation [135], maximum likelihood estimation [136], frequency hopping [137] and machine learning [138]. Our proposed technique is inspired by localization in passive radar [93] not requiring prior RSSI measurements.

The aforementioned works assume that measurements at each node are digitally represented by a large number of bits per sample such that the resulting quantization errors can be neglected. Further, when nodal measurements are sent to a fusion center (FC) for an aggregate decision, full capacity links are assumed. In this paper, contrary to these works, we consider the limiting case wherein the receivers at each node employ one-bit analog-to-digital converters (ADCs), which directly convert node measurements into *complex* data with binary components, each containing one-bit information, by comparing the real and imaginary parts of the node measurements with appropriate thresholds separately and noting the sign. This leads to one-bit per component measurements. Considering the fact that the cost and power consumption of ADCs increase exponentially with the number of quantization bits and sampling frequency [139], the use of one-bit ADCs supports the low-cost and low-power-consumption features of NB-IoT. We then leverage the recent advances in one-bit signal processing [140] to estimate the target *range/delay* with respect to a specific node. To cope with the capacity limitations of the nodal links, we assume that, prior to transmission to FC, the receive sensors quantize nodal estimates to one-bit data. The FC then performs target *localization*, i.e. determination of target's position with respect to the entire network, using the one-bit range vector aggregated from the estimates sent by all the nodes.

Converting analog signals into digital data using a single bit per sample leads to significant errors in the digital approximation of the original analog signals. This necessitates development of new algorithms for information retrieval from one-bit samples. One-bit sampling has a rich heritage of research in statistical signal processing [141–143] and signal reconstruction [144]. It was shown in [144] that, for band-limited bounded-amplitude square-integrable input signals, a sufficient number of one-bit samples lead to recovery of full-precision data with locally bounded point-wise error, resulting in an exponentially decaying distortion-rate characteristic. In the past few years, one-bit signal processing has received significant attention in numerous modern applications such as array processing [145, 146], massive multiple-input multiple-output (MIMO) [147], deep learning [148], dictionary learning [67], and radar [29]. Most of these works are based on either well-known Bussgang's Theorem [145, 147, 149] or compressive sensing techniques [29, 67, 148, 150]. Further, there are some elegant works on colocated one-bit radar and array processing [31, 146] which formulate the parameter estimation from one-bit measurements as an optimization problem with linear constraints which can be solved by polynomial-time algorithms. Contrary to previous works on colocated one-bit radar [31], our proposed method investigates widely separated radar setting.

We first formulate the problem of range/time-delay estimation in a clutter-free environment from one-bit samples received by each NB-IoT sensor as a sparse recovery problem. The formulation and approach of the clutter-free scenario is effectively applicable in

a weak clutter environment but the impact of strong clutter is unexamined and left for the future work. We show that, unlike infinite precision sampling, oversampling could improve the range/delay estimation performance in one-bit sampling. Further, oversampling leads our proposed approach to be able to achieve a considerably high resolution for time-delay estimation despite the narrow bandwidth used in NB-IoT. Toward dealing with the capacity limitations of the backhaul links, we assume that each sensor forwards an one-bit conversion of their range measurements to the FC. Collecting these one-bit measurements at the FC, we formulate the passive localization problem using the bistatic range-difference model. Note that the passive localization with NB-IoT sensors has a model similar to that of a passive radar [93]. The passive radar localization has been considered in [93] in the high-resolution ADC framework in which full-precision range measurements are assumed. This usually results in a system of several equations that are solved conventionally by the least squares (LS) method. In this context, apart from application to NB-IoT localization, ours is the first work in the context of one-bit sampling in a passive and distributed radar setting.

In our bistatic range-difference model, recovering locations from one-bit samples requires minimizing a cost function that is a non-negative polynomial in range measurement variables and subjected to polynomial inequalities defined by the positive-valued samples (the one-bit range measurements). The general approach to solving this problem is to re-cast the feasibility of this finite system of polynomial constraints in terms of an equivalent polynomial that involves squares of (unknown) polynomials [151]. However, it is rather difficult to express a non-negative multivariate polynomial as a sum-of-squares. To address this, we employ Lasserre's general solution approach for polynomial optimization problems via semi-definite programming (SDP) using methods based on moment theory [152]. Our novel formulation jointly estimates the full-precision data as well as the target location. While this method could attain the global minimum, its computational complexity grows considerably with increase in the number of NB-IoT sensors. In order to reduce the computational complexity, we trade accuracy with complexity by proposing a novel sub-optimal iterative joint r*an*ge-*tar*get location *es*timation (ANTARES) algorithm. We also derive the Cramér-Rao bound (CRB) for localization with one-bit nodal range measurements and use it as benchmark for assessing the estimation performance of the proposed optimal and sub-optimal algorithms. Numerical results show that when sufficiently large number of NB-IoT nodes are available, the optimal approach yields same performance as the full-precision and ANTARES leads to only 0.43% increase in the normalized localization error. Further, the normalized localization error rises minimally by 2.2% and 0.6% for a smaller set of 20-60 nodes using ANTARES and optimal algorithm, respectively, over the full precision case.

Preliminary results of this work appeared in our conference publication [153], where performance analysis was not included and only Lasserre's approach was considered. In this paper, we also investigate the one-bit time-delay estimation for the oversampled scenario and present ANTARES algorithm. In summary, our work provides a robust framework for location-based services in NB-IoT, does not require prior RSSI measurements, performs target delay estimation with one-bit samples, yields localization using limited capacity links, and is computationally efficient. Further, our work also has connections with the recent developments in spectrum sharing and joint radar-communications (JRC) design [139, 154]. Unlike some recent works [155] where new waveforms are developed for distributed JRC, our work exploits existing NB-IoT signaling for a sensing application.

Chapter organization: In section 6.2, we describe the system and signal model of the passive localization problem via the NB-IoT sensors. We introduce our one-bit nodal range estimation algorithm in Section 6.3. Then, using these estimates, we localize the target at FC in Section 6.4 through a polynomial optimization. We validate our models and methods through numerical experiments in Section 6.5 before concluding in Section 6.6.

6.2 System Model

Consider a source, say, a communications base-station whose location in Cartesian coordinates is $\begin{bmatrix}\delta_b^x & \delta_b^y & \delta_b^z\end{bmatrix}^T \in \mathbb{R}^{3\times 1}$. The source transmits a known baseband single-tone NB-IoT signal $s(t) \in \mathbb{C}$ with bandwidth B. As per NB-IoT specifications, the signal has spectrum limited to 180 kHz. It is similar to LTE with fewer (1, 3, 6, or 12) subcarriers with normal cyclic prefix [129, 156] and employs rotated phase shift keying (PSK) constellations, either $\pi/2$ binary PSK ($\pi/2$-BPSK) or $\pi/4$ quadrature PSK ($\pi/4$-QPSK). The resulting signal is

$$s(t) = \sum_{k=0}^{N_c-1} a_k e^{\mathrm{j}k\frac{\pi}{M}} g(t - kT_c), \quad 0 \le t < T, \tag{6.1}$$

where $a_k \in \{\pm 1\}$ for $\pi/2$-BPSK and $a_k \in \{\pm 1, \pm \mathrm{j}\}$ for $\pi/4$-QPSK are known pilot symbols, M is the alphabet size (2 for $\pi/2$-BPSK and 4 for $\pi/4$-QPSK), N_c is the maximum number of symbols allowed during the transmission, T denotes the observation interval, T_c is the symbol period, and $g(t)$ is the pulse shaping filter impulse response with bandwidth B.

The transmit signal is bounced off from the target-of-interest located at $\begin{bmatrix}\delta^x & \delta^y & \delta^z\end{bmatrix}^T \in \mathbb{R}^{3\times 1}$. In a typical NB-IoT setting, a target could be a subject carrying a mobile phone,

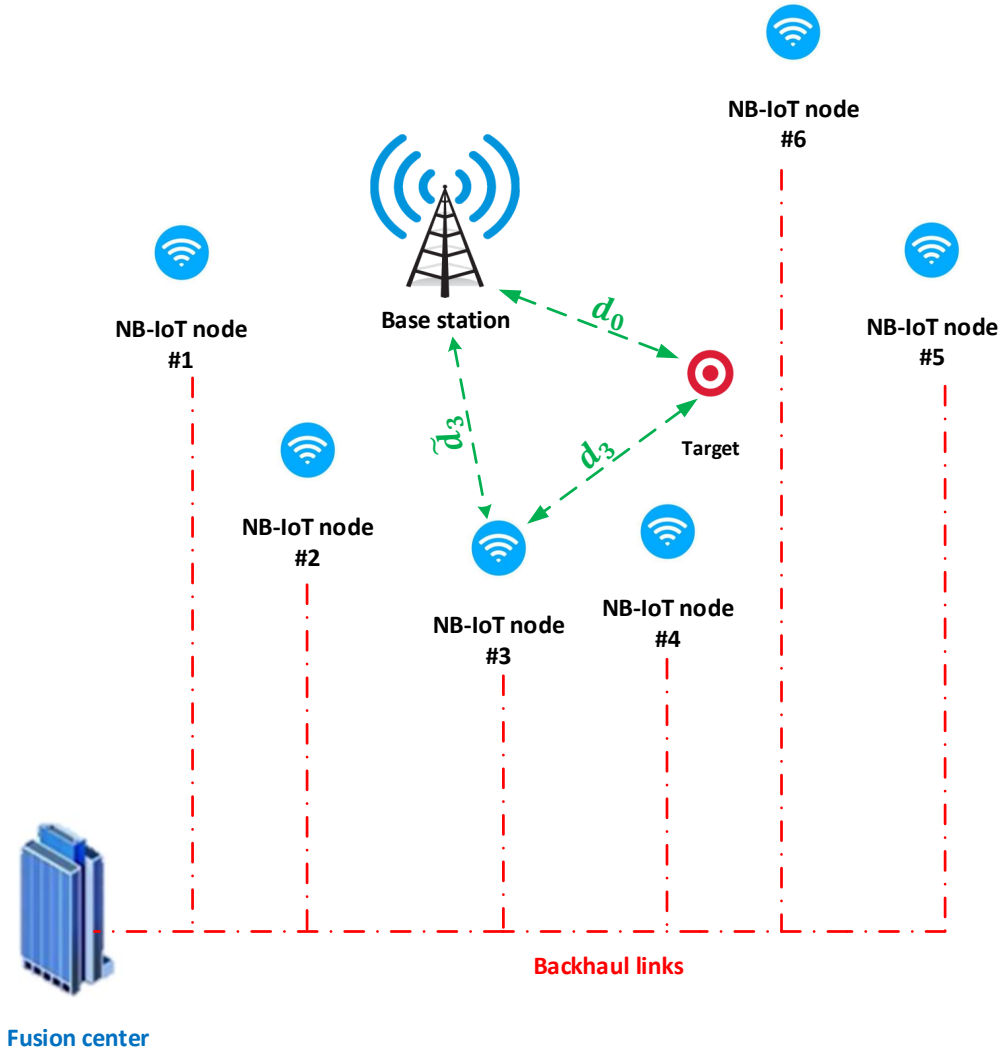


FIGURE 6.1: Illustration of the localization scenario. The NB-IoT #1, #2, $\cdots$, nodes (blue) are passive sensors (located at distances $\tilde{d}_1, \tilde{d}_2, \cdots, \tilde{d}_6$ from the base station). The nodes receive the signal from the source bounced off from a target-of-interest (red) located at distances $d_1, d_1, \cdots, d_6$ from the nodes and d_0 from the base station. In our proposed model, the nodes employ one-bit ADCs to sample the received signal and estimate the range. The estimated range at each node is quantized and then forwarded to the FC for an aggregated estimate.

an intelligent vehicle or a robot. The backscattered signal is then received by M distinct NB-IoT sensor nodes. The location of the m-th node is $\begin{bmatrix}\delta_m^x & \delta_m^y & \delta_m^z\end{bmatrix}^T \in \mathbb{R}^{3\times 1}, m \in \mathbb{M} \doteq \{1, 2, \cdots, M\}$. These nodes are synchronized with the base-station (Fig. 6.1). Synchronization could be provided by sending a periodic synchronization signal from the base-station to the NB-IoTs, including timing information of the base-station, while the base-station maintains a constant clock using either receiving a reference time from GPS or an atomic clock. After receiving the the base-station timing information,

NB-IoTs are able to accurately synchronize their clocks with the base-station clock [157–160]. More detailed information about the periodic synchronization signal and the synchronization mechanism in NB-IoT systems are provided in [159, 160], and the references therein. Synchronization may be also achieved through the use of protocols such as IEEE 1588 generic precision time protocol (gPTP) [161], network time protocol (NTP) [162] and wireless PTP [163]. These cost-effective clock synchronization protocols are also popular in other applications, including electrical grid networks, cellular base-station synchronization, industrial control, and vehicular systems [164, 165].

If the distance between the source and the target is d_0 and that between the target and the m-th NB-IoT node is

$$d_m = \sqrt{(\delta_m^x - \delta^x)^2 + (\delta_m^y - \delta^y)^2 + (\delta_m^z - \delta^z)^2},\ 1 \leq m \leq M, \tag{6.2}$$

then the true target range with respect to the m-th NB-IoT node is

$$r_m = d_m + d_0,\ 1 \leq m \leq M. \tag{6.3}$$

The propagation is non-dispersive and the base-station signal received by the NB-IoT nodes includes a direct line-of-sight (LoS) path from the base-station to the nodes and an indirect non-LoS (NLoS) path from the base-station to the target and then to the nodes. The demodulated baseband analog signal received at m-th sensor is

$$\breve{y}_m(t) = \widetilde{\alpha}_m s(t - \widetilde{\tau}_m) + \alpha_m s(t - \tau_m) + \breve{n}_m(t), \tag{6.4}$$

where $\widetilde{\alpha}_m \in \mathbb{C}$ ($\alpha_m \in \mathbb{C}$) and $\widetilde{\tau}_m \in \mathbb{R}$ ($\tau_m \in \mathbb{R}$) are the attenuation coefficient and time-delay of the propagation channel for the direct (indirect) path, respectively; and $\breve{n}_m(t) \in \mathbb{C}$ denotes additive white noise following a circular-symmetric complex Gaussian distribution with variance $N_m > 0$. The unknown time delay τ_m is linearly proportional to r_m, i.e. $\tau_m = r_m/c$ where $c = 3 \times 10^8$ m/s is the speed of light. The unknown direct path delay $\widetilde{\tau}_m$ is also linearly proportional to the distance between the m-th node and the base station. i.e., $\widetilde{\tau}_m = \widetilde{d}_m/c$ where $\widetilde{d}_m = \sqrt{(\delta_m^x - \delta_b^x)^2 + (\delta_m^y - \delta_b^y)^2 + (\delta_m^z - \delta_b^z)^2}$ denotes the distance between the m-th node and the base station.

The baseband signal is filtered by an ideal low-pass filter with bandwidth B and frequency response

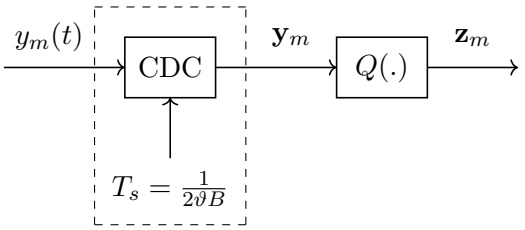


FIGURE 6.2: Conceptual representation of the oversampled one-bit ADC. The CDC block represents the digitizer operating at sampling rate of $1/T_s$. A quantizer $Q(\cdot)$ then converts the digital samples into a one-bit data stream.

$$H(\Omega) = \begin{cases} 1, & |\Omega| \leq 2\pi B, \\ 0, & \text{otherwise.} \end{cases} \tag{6.5}$$

This low-pass filtering of the signal $\breve{y}_m(t)$ yields

$$y_m(t) = \widetilde{\alpha}_m s(t - \widetilde{\tau}_m) + \alpha_m s(t - \tau_m) + n_m(t), \tag{6.6}$$

where $n_m(t)$ is the filtered noise trail whose auto-correlation is

$$\begin{aligned} R_{n_m}(t_1 - t_2) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} N_m |H(\Omega)|^2 e^{-j\Omega(t_1 - t_2)} \mathrm{d}\Omega \\ &= 2BN_m \mathrm{sinc}(2B(t_1 - t_2)), \end{aligned} \tag{6.7}$$

where $\mathrm{sinc}(u) = \frac{\sin(\pi u)}{\pi u}$.

Each NB-IoT node is equipped with a one-bit ADC which admits binary samples of the corresponding $y_m(t)$ during the observation interval $[0, T)$. The ADC sampling frequency $f_s = \frac{1}{T_s} = 2\vartheta B$, where ϑ is an integer greater than or equal to one, referred to as the oversampling factor. Figure 6.2 conceptually depicts a one-bit ADC which comprises a Continuous-to-Discrete Converter (CDC) with sampling frequency $f_s = 2\vartheta B$ followed by a one-bit quantizer. The CDC produces $L = \frac{T}{T_s} = 2\vartheta BT$ discrete samples of $y_m(t)$ during the time interval $[0, T)$. Stacking all discrete samples produces a $\mathbb{C}^{L \times 1}$ vector

$$\mathbf{y}_m = \widetilde{\alpha}_m \mathbf{s}(\widetilde{\tau}_m) + \alpha_m \mathbf{s}(\tau_m) + \mathbf{n}_m, \tag{6.8}$$

where $[\mathbf{y}_m]_l = y_m((l-1)T_s)$, $[\mathbf{s}(\widetilde{\tau}_m)]_l = s((l-1)T_s - \widetilde{\tau}_m)$, $[\mathbf{s}(\tau_m)]_l = s((l-1)T_s - \tau_m)$, and $[\mathbf{n}_m]_l = n_m((l-1)T_s)$ for $l = 1, 2, \cdots, L$. From (6.7) and Gaussianity of $n_m(t)$, vector $\mathbf{n}_m$ follows a zero-mean complex Gaussian distribution with the covariance

$$\mathbb{E}\{\mathbf{n}_m \mathbf{n}_m^H\} = \sigma_m^2 \mathbf{\Sigma} \quad \in \mathbb{C}^{L \times L} \tag{6.9}$$

where $[\mathbf{\Sigma}]_{i,j} = \text{sinc}\left(\frac{|i-j|}{\vartheta}\right)$ and $\sigma_m^2 = 2BN_m$.

The quantizer, represented by a function $Q(\cdot)$, converts the discrete samples into binary data by comparing each sample to a known threshold and then measuring the sign of the real and imaginary parts of the resulting difference. These one-bit measurements at the m-th NB-IoT node are

$$\mathbf{z}_m = Q(\mathbf{y}_m), \tag{6.10}$$

where the l-th element of $Q(\mathbf{y}_m)$ is

$$\begin{aligned} &[Q(\mathbf{y}_m)]_l \\ &= \frac{1}{\sqrt{2}}\text{sgn}(\Re\{[\mathbf{y}_m]_l - [\boldsymbol{\gamma}_m]_l\}) + \frac{j}{\sqrt{2}}\ \text{sgn}(\Im\{[\mathbf{y}_m]_l - [\boldsymbol{\gamma}_m]_l\}). \end{aligned} \tag{6.11}$$

with $\boldsymbol{\gamma}_m \in \mathbb{C}^{L\times 1}$ are known thresholds levels.

The nodal processing at each NB-IoT receiver entails estimation of the target time-delays, and hence the range, from one-bit samples $\mathbf{z}_m$. In the next section, we devise a method for one-bit time-delay estimation.

6.3 Time-Delay Estimation with One-Bit Samples

Several approaches have been proposed in the literature to estimate range (time-delay) of targets from one-bit samples with most formulating this as an optimization problem. For example, the covariance matrix formulation of [31] employs cyclic optimization method to extract the range along with other parameters. Other recent works using only one sensor exploit sparsity of the target scenario to estimate unknown parameters by applying techniques such as ℓ_1-norm minimization [166] and log-relaxation [167] to solve the resulting optimization. In our passive NB-IoT sensor set-up, the objective function is a variation of weighted least squares (WLS) that we minimize via ℓ_1-norm regularization to estimate τ_m using the one-bit quantized observations, i.e., $\mathbf{z}_m$. In conventional passive radars, direct and indirect path signals are recorded in separate reference and surveillance channels, respectively. However, the direct signal may seep into the surveillance channel and mask the relatively weaker indirect signal. In such cases, adaptive filters are employed to first suppress the direct signal in the surveillance channel [168]. However, our NB-IoT scenario is an opportunistic sensing application where the receivers are not equipped to record separate channels. Moreover, as explained next, the (additive) overlap of direct signal with the target echo is useful because the former is used to estimate the latter in

our formulation. Here, we also remark that there are passive radar applications where direct signal suppression is not crucial. For example, this requirement is often relaxed in passive sensing using communications satellites because of the relatively weak power of the direct path satellite signal than, say, commonly used broadcasting signals [169].

6.3.1 Constrained-Weighted Least Squares Minimization

Equation (6.8) can be transformed to the frequency domain by multiplying both sides by an $L \times L$ Discrete Fourier Transform (DFT) matrix $\mathbf{F}$, whose (n,k)-th entry is $e^{\frac{-\mathrm{j}2\pi nk}{L}}$. This yields

$$\mathbf{F}\mathbf{y}_m = \widetilde{\alpha}_m \mathrm{diag}(\overline{\mathbf{s}}_{\widetilde{\tau}_m})\mathbf{a}(\widetilde{\tau}_m) + \alpha_m \mathrm{diag}(\overline{\mathbf{s}}_{\tau_m})\mathbf{a}(\tau_m) + \overline{\mathbf{n}}_m, \tag{6.12}$$

where $\overline{\mathbf{n}}_m = \mathbf{F}\mathbf{n}_m$, $[\mathbf{a}(u)]_l = e^{-j2\pi \frac{(l-1)u}{LT_s}}$ for $0 \leq l \leq L-1$ and $\overline{\mathbf{s}}_u = \mathbf{F}\mathbf{s}_u$ with

$$[\mathbf{s}_u]_l = \begin{cases} s((l-1)T_s) & 1 \leq l \leq L - \lfloor \frac{u}{L} \rfloor, \\ 0 & \text{otherwise.} \end{cases} \tag{6.13}$$

Let us discretize the continuous space of the time delay, i.e., $[0, T)$, into a given set of $N \geq L$ grid points, i.e., $\{\overline{\tau}_{m,1}, \cdots, \overline{\tau}_{m,N}\}$ [170]. This discretization transforms (6.12) into the following sparse model

$$\mathbf{F}\mathbf{y}_m = [\overline{\mathbf{S}} \odot \mathbf{A}(\overline{\boldsymbol{\tau}}_m)]\overline{\boldsymbol{\alpha}}_m + \overline{\mathbf{n}}_m \tag{6.14}$$

where $\mathbf{A}(\overline{\boldsymbol{\tau}}_m) = \begin{bmatrix} \mathbf{a}(\overline{\tau}_{m,1}) & \cdots & \mathbf{a}(\overline{\tau}_{m,N}) \end{bmatrix} \in \mathbb{C}^{L \times N}$, $\overline{\mathbf{S}} = \begin{bmatrix} \overline{\mathbf{s}}_{\overline{\tau}_{m,1}} & \cdots & \overline{\mathbf{s}}_{\overline{\tau}_{mN1}} \end{bmatrix} \in \mathbb{C}^{L \times N}$ and $\overline{\boldsymbol{\alpha}}_m = \begin{bmatrix} \overline{\alpha}_{m,1} & \cdots & \overline{\alpha}_{m,N} \end{bmatrix} \in \mathbb{C}^{N \times 1}$ is a sparse vector with

$$[\overline{\boldsymbol{\alpha}}_m]_k = \begin{cases} \alpha_m, & \text{if } \overline{\tau}_{m,k} = \tau_m, \\ \widetilde{\alpha}_m, & \text{if } \overline{\tau}_{m,k} = \widetilde{\tau}_m, \\ 0, & \text{otherwise.} \end{cases} \tag{6.15}$$

The waveform $\mathbf{s}$ is known at NB-IoT receiver. Hence, the problem is to find $\mathbf{y}_m$ and a sparse vector $\overline{\boldsymbol{\alpha}}_m$ which are consistent with the model in (6.15) as well as one-bit measurments $\mathbf{z}_m$. In consequence, the time-delay estimation problem can be formulated as follows [166]

$$\begin{aligned} \underset{\mathbf{y}_m, \overline{\boldsymbol{\alpha}}_m}{\text{minimize}} \quad & \|\overline{\boldsymbol{\alpha}}_m\|_1 + \rho \|\mathbf{W}\left[\mathbf{F}\mathbf{y}_m - [\overline{\mathbf{S}} \odot \mathbf{A}(\overline{\boldsymbol{\tau}}_m)]\overline{\boldsymbol{\alpha}}_m\right]\|_2^2 \\ \text{s.t.} \quad & \Re\{\mathbf{z}_m\} \odot \Re\{\mathbf{y}_m - \boldsymbol{\gamma}_m\} \succeq \mathbf{0}, \\ & \Im\{\mathbf{z}_m\} \odot \Im\{\mathbf{y}_m - \boldsymbol{\gamma}_m\} \succeq \mathbf{0}. \end{aligned} \tag{6.16}$$

where ρ is a regularization parameter and $\mathbf{W} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{F}^H$ is a weighting matrix. The first term in the objective of (6.16) promotes sparsity in $\overline{\boldsymbol{\alpha}}_m$ while the second term is a WLS criterion that penalizes the model mismatch in (6.14) considering the fact that the additive noise in (6.14) follows a circular-symmetric complex Gaussian distribution with the covariance matrix $\sigma_m \mathbf{F}\boldsymbol{\Sigma}\mathbf{F}^H$. Further, linear constraints arise because one-bit quantized and discrete samples must share the same sign. Introducing a slack variable $\mathbf{x}_m = \boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{F}^H\left[\mathbf{F}\mathbf{y}_m - [\overline{\mathbf{S}} \odot \mathbf{A}(\overline{\boldsymbol{\tau}}_m)]\overline{\boldsymbol{\alpha}}_m\right]$, (6.16) becomes

$$\begin{aligned} \underset{\mathbf{x}_m, \overline{\boldsymbol{\alpha}}_m}{\text{minimize}} \quad & \|\overline{\boldsymbol{\alpha}}_m\|_1 + \rho \|\mathbf{x}_m\|_2^2 \\ \text{s.t.} \quad & \Re\{\mathbf{z}_m\} \odot \Re\{\mathbf{F}^H[\overline{\mathbf{S}} \odot \mathbf{A}(\overline{\boldsymbol{\tau}}_m)]\overline{\boldsymbol{\alpha}}_m + \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{x}_m - \boldsymbol{\gamma}_m\} \succeq \mathbf{0}, \\ & \Im\{\mathbf{z}_m\} \odot \Im\{\mathbf{F}^H[\overline{\mathbf{S}} \odot \mathbf{A}(\overline{\boldsymbol{\tau}}_m)]\overline{\boldsymbol{\alpha}}_m + \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{x}_m - \boldsymbol{\gamma}_m\} \succeq \mathbf{0}. \end{aligned} \tag{6.17}$$

The above problem comprises minimization of a convex objective function with linear constraints and can be solved efficiently [98].

The solution of (6.17) yields estimate of $\overline{\boldsymbol{\alpha}}_m$ which has two non-zero elements at indices k_1 and k_2. From this, we find $\widehat{\boldsymbol{\tau}}_m = \left[\frac{(k_1-1)T}{N}, \frac{(k_2-1)T)}{N}\right]^T$. The estimated unknown time delay corresponding to the indirect path is then

$$\widehat{\tau}_m = \max\{[\widehat{\boldsymbol{\tau}}_m]_1, [\widehat{\boldsymbol{\tau}}_m]_2\}, \tag{6.18}$$

Lemma 6.1. *$\widehat{\tau}_m$ is a consistent estimate of τ_m.*

Proof. See Appendix C.1. □

Hence, the a consistent estimate of the range of the target is given by $\widehat{r}_m = c\widehat{\tau}_m$.

6.3.2 Improved Performance with Oversampling

It is possible to improve the recovery performance if the one-bit ADCs sample at a rate higher than the Nyquist. Note that the samples are still quantized to only single bits. In this section, we analyze the effect of oversampling.

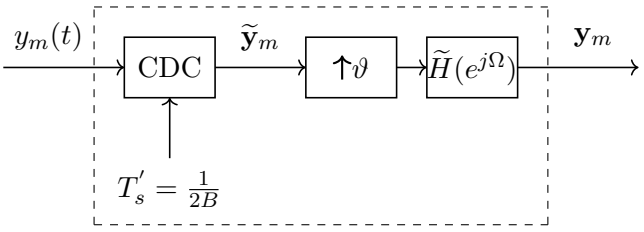


FIGURE 6.3: An equivalent representation of Fig. 6.3 to show both oversampled $\mathbf{y}_m$ and Nyquist-sampled $\widetilde{\mathbf{y}}_m$.

In case of oversampling, let replace the CDC module in Fig. 6.2 with an equivalent system (Fig. 6.3) composed of a CDC that samples $y_m(t)$ at the Nyquist rate followed by an ϑ-fold upsampling. A low-pass filter with frequency response

$$\widetilde{H}(e^{j\Omega}) = \begin{cases} \vartheta, & |\Omega| \leq \frac{\pi}{\vartheta}, \\ 0, & \text{Otherwise}, \end{cases} \tag{6.19}$$

outputs the oversampled data $\mathbf{y}_m$. The oversampled $\mathbf{y}_m$ and Nyquist-sampled $\widetilde{\mathbf{y}}_m$ (see Fig. 6.3) are related as [171]

$$\begin{aligned} [\mathbf{y}_m]_l &= \sum_{p=1}^{L/\vartheta} [\widetilde{\mathbf{y}}_m]_p \text{sinc}\left(\frac{l-1}{\vartheta} - p + 1\right) \\ &= \begin{cases} [\widetilde{\mathbf{y}}_m]_p, & \text{if } l = (p-1)\vartheta + 1,\ 1 \leq p \leq L/\vartheta, \\ \sum_{p=1}^{L/\vartheta} [\widetilde{\mathbf{y}}_m]_p \text{sinc}\left(\frac{l-1}{\vartheta} - p + 1\right), & \text{otherwise.} \end{cases} \end{aligned} \tag{6.20}$$

Indeed, (6.20) implies that $\frac{L}{\vartheta}$ elements of $\mathbf{y}_m$ are exactly equal to those of $\widetilde{\mathbf{y}}_m$; and the other elements of $\mathbf{y}_m$ are obtained from linear combinations of the elements of $\widetilde{\mathbf{y}}_m$. Let $[\overline{\mathbf{y}}_m]_l = [\mathbf{y}_m]_l$ for $l \neq (p-1)\vartheta + 1$ and $1 \leq p \leq L/\vartheta$ and $\boldsymbol{\mathcal{I}}(.|\boldsymbol{\theta})$ denote the Fisher Information Matrix (FIM) with respect to the parameter vector $\boldsymbol{\theta}$. The linear dependence of $\overline{\mathbf{y}}_m$ and $\widetilde{\mathbf{y}}_m$ implies that $\boldsymbol{\mathcal{I}}(\overline{\mathbf{y}}_m|\widetilde{\mathbf{y}}_m, \boldsymbol{\tau}_m, \boldsymbol{\alpha}_m) = \mathbf{0}$. Hence, it follows from the chain rule of FIM [172] that

$$\boldsymbol{\mathcal{I}}(\mathbf{y}_m|\boldsymbol{\tau}_m, \boldsymbol{\alpha}_m) = \boldsymbol{\mathcal{I}}(\widetilde{\mathbf{y}}_m|\boldsymbol{\tau}_m, \boldsymbol{\alpha}_m). \tag{6.21}$$

This means that oversampling has no impact on the accuracy of the time-delay estimation using full-precision data in our model.

Now let us consider the effect of oversampling on the accuracy of the time-delay estimation using one-bit data. Substituting (6.20) into (6.10) yields

$$[\mathbf{z}_m]_l = Q([\mathbf{y}_m]_l) \tag{6.22}$$
$$= \begin{cases} [\widetilde{\mathbf{z}}_m]_p, & \text{if } l=(p-1)\vartheta+1,\ 1\leq p\leq L, \\ Q\left(\sum_{p=1}^{L}[\widetilde{\mathbf{y}}_m]_p \text{sinc}\left(\frac{l-1}{\vartheta}-p+1\right)\right), & \text{otherwise,} \end{cases}$$

where $\widetilde{\mathbf{z}}_m = Q(\widetilde{\mathbf{y}}_m)$ contains the one-bit data at the Nyquist rate. From (6.22), we deduce that whereas $\frac{L}{\vartheta}$ elements of $\mathbf{z}_m$ are exactly equal to those of $\widetilde{\mathbf{z}}_m$, the remaining elements of $\mathbf{z}_m$, denoted by $\overline{\mathbf{z}}_m \in \mathbb{C}^{(1-\frac{1}{\vartheta})L\times 1}$, can not be constructed from linear combinations of the elements of $\widetilde{\mathbf{z}}_m$ like the full-precision case. In other words, (6.22) indicates that while $\widetilde{\mathbf{z}}_m$ provides information about only the signs of $\widetilde{\mathbf{y}}_m$, $\overline{\mathbf{z}}_m$ provides additional information on the signs of the linear combinations of $\widetilde{\mathbf{y}}_m$. Therefore, in general, $\boldsymbol{\mathcal{I}}(\overline{\mathbf{z}}_m|\widetilde{\mathbf{z}}_m, \boldsymbol{\tau}_m, \boldsymbol{\alpha}_m) \succeq \mathbf{0}$. From the chain rule of FIM [172], we have

$$\boldsymbol{\mathcal{I}}(\mathbf{z}_m|\boldsymbol{\tau}_m, \boldsymbol{\alpha}_m) = \boldsymbol{\mathcal{I}}(\widetilde{\mathbf{z}}_m|\boldsymbol{\tau}_m, \boldsymbol{\alpha}_m) + \boldsymbol{\mathcal{I}}(\overline{\mathbf{z}}_m|\widetilde{\mathbf{z}}_m, \boldsymbol{\tau}_m, \boldsymbol{\alpha}_m). \tag{6.23}$$

Considering (6.23) and $\boldsymbol{\mathcal{I}}(\overline{\mathbf{z}}_m|\widetilde{\mathbf{z}}_m, \boldsymbol{\tau}_m, \boldsymbol{\alpha}_m) \succeq \mathbf{0}$, we observe

$$\boldsymbol{\mathcal{I}}(\mathbf{z}_m|\boldsymbol{\tau}_m, \boldsymbol{\alpha}_m) \succeq \boldsymbol{\mathcal{I}}(\widetilde{\mathbf{z}}_m|\boldsymbol{\tau}_m, \boldsymbol{\alpha}_m) \tag{6.24}$$

This implies that oversampling could enhance the parameter estimation performance when one-bit quantized data is used.

6.4 Target Localization with One-Bit Samples

In order to comply with bandwidth and power limitations, each of the M sensors converts its nodal range measurements into a binary sample w_m by comparing it to a positive threshold $\lambda_m > 0$, i.e.,

$$w_m = \text{sgn}(r_m - \lambda_m). \tag{6.25}$$

All nodes forward this binary range and the corresponding thresholds to the FC which localizes the target using the binary range measurements from all nodes. We first present a framework for target localization with full precision (or infinite-bit) range measurements and follow it with our methods for one-bit data.

6.4.1 Localization with Full-Precision Range Estimates

Recall the expressions of d_m and r_m in (6.2) and (6.3), respectively. Without loss of generality, consider the first ($m = 1$) sensor as the reference sensor. The difference

between the true range with respect to reference sensor and any of the remaining m-th ($m > 1$) sensor is

$$r_m - r_1 = d_m - d_1, \tag{6.26}$$

Rearranging (6.26) as $r_m - r_1 + d_1 = d_m$, and squaring both sides produces

$$((r_m - r_1) + d_1)^2 = d_m^2 = (\delta_m^x - \delta_1^x)^2 + (\delta_m^y - \delta_1^y)^2 + (\delta_m^z - \delta_1^z)^2, \tag{6.27}$$

where the last equality follows after substituting d_m from (6.2). Simplifying yields

$$\begin{aligned} &(\delta^x - \delta_1^x)(\delta_m^x - \delta_1^x) + (\delta^y - \delta_1^y)(\delta_m^y - \delta_1^y) + (\delta^z - \delta_1^z)(\delta_m^z - \delta_1^z) \\ &+ (r_m - r_1)d_1 = \\ &\frac{1}{2}\left[(\delta_m^x - \delta_1^x)^2 + (\delta_m^y - \delta_1^y)^2 + (\delta_m^z - \delta_1^z)^2 - (r_m - r_1)^2\right], \end{aligned} \tag{6.28}$$

which are linear in the target coordinates $\begin{bmatrix}\delta^x & \delta^y & \delta^z\end{bmatrix}^T$. Denote the unknown parameter vector

$$\boldsymbol{\theta} = \begin{bmatrix}\delta^x - \delta_1^x & \delta^y - \delta_1^y & \delta^z - \delta_1^z & d_1\end{bmatrix}^T \in \mathbb{R}^{4\times 1}. \tag{6.29}$$

Then, collecting all linear equations specified by (6.28) for $m = 2, \cdots, M$, we obtain the following compact matrix form

$$\mathbf{G}\boldsymbol{\theta} = \mathbf{h}, \tag{6.30}$$

where

$$\mathbf{G} = \begin{bmatrix} (\delta_2^x - \delta_1^x) & (\delta_2^y - \delta_1^y) & (\delta_2^z - \delta_1^z) & r_2 - r_1 \\ \vdots & \vdots & \vdots & \vdots \\ (\delta_M^x - \delta_1^x) & (\delta_M^y - \delta_1^y) & (\delta_M^z - \delta_1^z) & r_M - r_1 \end{bmatrix} \in \mathbb{R}^{(M-1)\times 4}, \tag{6.31}$$

and

$$\mathbf{h} = \frac{1}{2}\begin{bmatrix} (\delta_2^x - \delta_1^x)^2 + (\delta_2^y - \delta_1^y)^2 + (\delta_2^z - \delta_1^z)^2 - (r_2 - r_1)^2 \\ \vdots \\ (\delta_m^x - \delta_1^x)^2 + (\delta_M^y - \delta_1^y)^2 + (\delta_M^z - \delta_1^z)^2 - (r_M - r_1)^2 \end{bmatrix} \in \mathbb{R}^{(M-1)\times 1}. \tag{6.32}$$

In practice, every true m-th sensor range r_m is unknown. As explained in the previous section, we employ constrained WLS to obtain the estimate $\widehat{r}_m$. Assume $\widehat{r}_m = r_m + e_m$, where e_m is the estimation error due to the receiver noise. Then, the equality in (6.30) does not hold and the resulting perturbed system of equations takes the form

$$\boldsymbol{\epsilon} = \mathbf{G}\boldsymbol{\theta} - \mathbf{h}, \tag{6.33}$$

where $\boldsymbol{\epsilon}$ denotes the perturbation term. Assuming $\mathbf{G}$ is full column rank, the least squares (LS) solution of the system of linear equations in (6.33) yields

$$\widehat{\boldsymbol{\theta}} = \mathbf{G}^{\dagger}\mathbf{h}. \tag{6.34}$$

Then, the target location is obtained as

$$\begin{bmatrix} \delta^x & \delta^y & \delta^z \end{bmatrix}^T = \begin{bmatrix} [\widehat{\boldsymbol{\theta}}]_1 + \delta_1^x & [\widehat{\boldsymbol{\theta}}]_2 + \delta_1^y & [\widehat{\boldsymbol{\theta}}]_3 + \delta_1^z \end{bmatrix}^T. \tag{6.35}$$

Remark 6.1. Contrary to range estimation, WLS is not applicable for estimating $\boldsymbol{\theta}$ in (6.33) because the covariance matrix of perturbation $\boldsymbol{\epsilon}$ is unknown. This is apparent from the fact that the covariance matrix of the perturbation term is a function of the variances of the range estimation errors, i.e., $e_1, e_2, \cdots, e_M$, as well as the unknown target location. Under such circumstances, the best choice for the weighting matrix is the identity matrix, which reduces WLS to LS.

When the FC receives the full-precision nodal range estimates, i.e., $\widehat{r}_m$ for $1 \leq m \leq M$, the aforementioned LS solution in (6.35) is quite effective. However, when the nodal range estimates are quantized to one-bit as in (6.25), the LS approach is no longer applicable at the FC.

6.4.2 Optimal Localization with One-Bit Nodal Range Estimates

We first develop an optimal approach for localization with one-bit quantized range measurements from the M nodes denoted by $\mathbf{w} = \begin{bmatrix} w_1, w_2, \cdots, w_M \end{bmatrix}^T$. We show that this optimal approach achieves the global minimum.

Consider $\overline{\mathbf{r}} = \begin{bmatrix} r_2 & r_3 & \cdots & r_M \end{bmatrix}^T \in \mathbb{R}^{(M-1)\times 1}$ and denote $\mathbf{1}$ as a $(M-1)\times 1$ vector with all ones as its elements. Define

$$\mathbf{V} = \begin{bmatrix} (\delta_2^x - \delta_1^x) & (\delta_2^y - \delta_1^y) & (\delta_2^z - \delta_1^z) \\ \vdots & \vdots & \vdots \\ (\delta_M^x - \delta_1^x) & (\delta_M^y - \delta_1^y) & (\delta_M^z - \delta_1^z) \end{bmatrix} \in \mathbb{R}^{(M-1)\times 3}, \tag{6.36}$$

and

$$\mathbf{b} = \frac{1}{2} \begin{bmatrix} (\delta_2^x - \delta_1^x)^2 + (\delta_2^y - \delta_1^y)^2 + (\delta_2^z - \delta_1^z)^2 \\ \vdots \\ (\delta_M^x - \delta_1^x)^2 + (\delta_M^y - \delta_1^y)^2 + (\delta_M^z - \delta_1^z)^2 \end{bmatrix} \in \mathbb{R}^{(M-1)\times 1}. \tag{6.37}$$

Both $\mathbf{V}$ and $\mathbf{b}$ are known *a priori*. Then,

$$\mathbf{G} = \begin{bmatrix} \mathbf{V} & \overline{\mathbf{r}} - r_1\mathbf{1} \end{bmatrix}, \tag{6.38}$$

$$\mathbf{h} = \mathbf{b} - \frac{1}{2}(\overline{\mathbf{r}} - r_1\mathbf{1}) \odot (\overline{\mathbf{r}} - r_1\mathbf{1}). \tag{6.39}$$

We jointly estimate the unknown $\boldsymbol{\theta}$ and $\mathbf{r}$ by solving the optimization

$$\begin{aligned} \underset{\mathbf{r},\boldsymbol{\theta}}{\text{minimize}} \quad & \|\mathbf{G}\boldsymbol{\theta} - \mathbf{h}\|_2^2 \\ \text{s.t.} \quad & \mathbf{w} \odot (\mathbf{r} - \boldsymbol{\lambda}) \succeq \mathbf{0}, \\ & \mathbf{r} \succeq \mathbf{0}, \end{aligned} \tag{6.40}$$

where $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \cdots, \lambda_M]^T$. The first linear constraint in (6.40), similar to the formulation in Section 6.3, arises because the one-bit quantized data and the elements of $\mathbf{r} - \boldsymbol{\lambda}$ must share the same sign; and the second constraint indicates that range values are non-negative. Reformulate the objective function $\mathcal{L}(\mathbf{r}, \boldsymbol{\theta}) \triangleq \|\mathbf{G}\boldsymbol{\theta} - \mathbf{h}\|_2^2$ as

$$\mathcal{L}(\mathbf{r}, \boldsymbol{\theta}) \triangleq \left\| \begin{bmatrix} \mathbf{V} & \overline{\mathbf{r}} - r_1\mathbf{1} \end{bmatrix} \boldsymbol{\theta} - \mathbf{b} + \frac{1}{2}(\overline{\mathbf{r}} - r_1\mathbf{1}) \odot (\overline{\mathbf{r}} - r_1\mathbf{1}) \right\|_2^2. \tag{6.41}$$

When $\mathbf{r}$ is fixed, the LS solution for $\boldsymbol{\theta}$ is given by (6.34). Substituting (6.34) into (6.41) yields

$$\begin{aligned} \mathcal{L}(\mathbf{r}) = \mathcal{L}(\mathbf{r}, \widehat{\boldsymbol{\theta}}) &\triangleq \|\mathbf{G}\mathbf{G}^\dagger\mathbf{h} - \mathbf{h}\|_2^2 = \|\Pi_{\mathbf{G}}^{\perp}\mathbf{h}\|_2^2 \\ &= \left\| \left[\Pi_{\mathbf{V}}^{\perp} - \Pi_{\Pi_{\mathbf{V}}^{\perp}(\overline{\mathbf{r}} - r_1\mathbf{1})}\right] \left[\mathbf{b} - \frac{1}{2}(\overline{\mathbf{r}} - r_1\mathbf{1}) \odot (\overline{\mathbf{r}} - r_1\mathbf{1})\right] \right\|_2^2, \end{aligned} \tag{6.42}$$

where the last equality is obtained by substituting (6.38)-(6.39) and using $\Pi_{\mathbf{G}}^{\perp} = \Pi_{\mathbf{V}}^{\perp} - \Pi_{\Pi_{\mathbf{V}}^{\perp}(\overline{\mathbf{r}} - r_1\mathbf{1})}$ following the projection decomposition theorem [173]. Since $\Pi_{\mathbf{V}}^{\perp}(\overline{\mathbf{r}} - r_1\mathbf{1}) \in \mathcal{N}(\mathbf{V}^H)$, it is easily confirmed that $\Pi_{\mathbf{V}}^{\perp}\Pi_{\Pi_{\mathbf{V}}^{\perp}(\overline{\mathbf{r}} - r_1\mathbf{1})} = \Pi_{\Pi_{\mathbf{V}}^{\perp}(\overline{\mathbf{r}} - r_1\mathbf{1})}$ simplifying (6.42) to

$$\mathcal{L}(\mathbf{r}) = \left[\mathbf{b} - \frac{1}{2}(\overline{\mathbf{r}} - r_1\mathbf{1}) \odot (\overline{\mathbf{r}} - r_1\mathbf{1})\right]^T \left[\Pi_{\mathbf{V}}^{\perp} - \Pi_{\Pi_{\mathbf{V}}^{\perp}(\overline{\mathbf{r}} - r_1\mathbf{1})}\right] \times \left[\mathbf{b} - \frac{1}{2}(\overline{\mathbf{r}} - r_1\mathbf{1}) \odot (\overline{\mathbf{r}} - r_1\mathbf{1})\right]. \tag{6.43}$$

Expanding $\Pi_{\Pi_{\mathbf{V}}^{\perp}(\overline{\mathbf{r}} - r_1\mathbf{1})}$ yields

$$\begin{aligned}\Pi_{\Pi_{\mathbf{V}}^{\perp}(\overline{\mathbf{r}} - r_1\mathbf{1})} &= \Pi_{\mathbf{V}}^{\perp}(\overline{\mathbf{r}} - r_1\mathbf{1})\Pi_{\mathbf{V}}^{\perp}(\overline{\mathbf{r}} - r_1\mathbf{1})^{\dagger} \\ &= \frac{\Pi_{\mathbf{V}}^{\perp}(\overline{\mathbf{r}} - r_1\mathbf{1})(\overline{\mathbf{r}} - r_1\mathbf{1})^T\Pi_{\mathbf{V}}^{\perp}}{\|\Pi_{\mathbf{V}}^{\perp}(\overline{\mathbf{r}} - r_1\mathbf{1})\|_2^2}.\end{aligned} \tag{6.44}$$

Note that the fact that $\mathbf{G}$ is full column rank guarantees $\|\Pi_{\mathbf{V}}^{\perp}(\overline{\mathbf{r}} - r_1\mathbf{1})\|_2^2 \neq 0$. Substituting (6.44) in (6.43), the $\mathcal{L}(\mathbf{r})$ takes the rational form $\frac{\mathcal{F}(\mathbf{r})}{\mathcal{J}(\mathbf{r})}$ where

$$\begin{aligned}\mathcal{F}(\mathbf{r}) =& \|\Pi_{\mathbf{V}}^{\perp}(\overline{\mathbf{r}} - r_1\mathbf{1})\|_2^2 \Big(\|\Pi_{\mathbf{V}}^{\perp}\mathbf{b}\|_2^2 + \frac{1}{4}\|\Pi_{\mathbf{V}}^{\perp}\left[(\overline{\mathbf{r}} - r_1\mathbf{1}) \odot (\overline{\mathbf{r}} - r_1\mathbf{1})\right]\|_2^2 \\ &- \mathbf{b}^T\Pi_{\mathbf{V}}^{\perp}\left[(\overline{\mathbf{r}} - r_1\mathbf{1}) \odot (\overline{\mathbf{r}} - r_1\mathbf{1})\right] \Big) \\ &- \left(\mathbf{b}^T\Pi_{\mathbf{V}}^{\perp}(\overline{\mathbf{r}} - r_1\mathbf{1})\right)^2 - \frac{1}{4}\left(\left[(\overline{\mathbf{r}} - r_1\mathbf{1}) \odot (\overline{\mathbf{r}} - r_1\mathbf{1})\right]^T\Pi_{\mathbf{V}}^{\perp}(\overline{\mathbf{r}} - r_1\mathbf{1})\right)^2 \\ &+ \mathbf{b}^T\Pi_{\mathbf{V}}^{\perp}(\overline{\mathbf{r}} - r_1\mathbf{1})(\overline{\mathbf{r}} - r_1\mathbf{1})^T\Pi_{\mathbf{V}}^{\perp}\left[(\overline{\mathbf{r}} - r_1\mathbf{1}) \odot (\overline{\mathbf{r}} - r_1\mathbf{1})\right],\end{aligned} \tag{6.45}$$

is a polynomial of degree 6 and

$$\mathcal{J}(\mathbf{r}) = \|\Pi_{\mathbf{V}}^{\perp}(\overline{\mathbf{r}} - r_1\mathbf{1})\|_2^2, \tag{6.46}$$

is a polynomial of degree 2. Hence, (6.40) becomes

$$\begin{aligned}\underset{\mathbf{r}}{\text{minimize}} \quad & \frac{\mathcal{F}(\mathbf{r})}{\mathcal{J}(\mathbf{r})} \\ \text{s.t.} \quad & \mathbf{w} \odot (\mathbf{r} - \boldsymbol{\lambda}) \succeq \mathbf{0}, \\ & \mathbf{r} \succeq \mathbf{0}.\end{aligned} \tag{6.47}$$

The optimization problem in (6.47) is non-convex. In order to relax this fractional structure, we decouple the numerator and the denominator as stated in the following theorem.

Theorem 6.1. *The optimization problem in* (6.47) *is equivalent to*

$$\begin{aligned}\underset{v,\mathbf{r}}{\textit{minimize}} \quad & v \\ \text{s.t.} \quad & v\mathcal{J}(\mathbf{r}) - \mathcal{F}(\mathbf{r}) \geq 0, \\ & \mathbf{w} \odot (\mathbf{r} - \boldsymbol{\lambda}) \succeq \mathbf{0}, \\ & \mathbf{r} \succeq \mathbf{0},\end{aligned} \tag{6.48}$$

where v is a slack variable.

Proof. See Appendix C.2. □

The objective in the optimization problem (6.48) is not rational. However, it is still non-convex because of the polynomial constraint $v\mathcal{J}(\mathbf{r}) - \mathcal{F}(\mathbf{r}) \geq 0$ of degree 6. To reformulate the problem to an equivalent SDP, we employ Lasserre's multivariate polynomial optimization [152].

Definition 6.1 (Monomial basis of degree p)**.** The vector $\mathbf{g}_p(\mathbf{u})$ is called the monomial basis of degree p if it contains all monomials $u_1^{\nu_1}u_2^{\nu_2}\cdots u_q^{\nu_q}$ such that $\sum_{i=1}^{q}\nu_i \leq p$ with ν_i's being integers.

For example, $\mathbf{g}_2(u_1, u_2)$ is the monomial basis of degree 2 if

$$\mathbf{g}_2([u_1, u_2]^T) = \begin{bmatrix} 1 & u_1 & u_2 & u_1^2 & u_1u_2 & u_2^2 \end{bmatrix}^T. \tag{6.49}$$

To parametrize the first constraint of (6.48), substituting (6.45)-(6.46) in $v\mathcal{J}(\mathbf{r}) - \mathcal{F}(\mathbf{r})$, and expanding the resulting equation, we obtain

$$\begin{aligned}
v\mathcal{J}(\mathbf{r})-\mathcal{F}(\mathbf{r}) &= \sum_{m=1}^{M}\psi_{mm}r_m^2 v + (\kappa_m^2-\chi\psi_{mm})r_m^2 + \sum_{\substack{m=1\\m\neq n}}^{M}\sum_{n=1}^{M}\psi_{mn}r_m r_n v + \frac{(\psi_{mn}^2-\psi_{mm}\psi_{nn})}{4} \\
&\times (r_m^4 r_n^2 - r_m^3 r_n^3) + (\kappa_m\psi_{mn}-\psi_{mm}\psi_{nn})(r_m^3 r_n - r_m^2 r_n^2) + (\kappa_m\kappa_n - \chi\psi_{mn})r_m r_n \\
&+ \sum_{\substack{m=1\\m\neq n\neq k}}^{M}\sum_{n=1}^{M}\sum_{k=1}^{M}\frac{(\psi_{mn}\psi_{mk}-\psi_{mm}\psi_{nk})}{4}(r_m^4 r_n r_k - 2r_m^3 r_n^2 r_k + r_m^2 r_n^2 r_k^2) \\
&+ (\psi_{mm}\psi_{nk} - \psi_{mn}\kappa_k)r_m^2 r_n r_k + \sum_{\substack{m=1\\m\neq n\neq k\neq q}}^{M}\sum_{n=1}^{M}\sum_{k=1}^{M}\sum_{q=1}^{M}\frac{(\psi_{mk}\psi_{nq}-\psi_{mn}\psi_{kq})}{4}r_m^2 r_n^2 r_k r_q \\
&+ \sum_{\substack{m=2\\m\neq n\neq k}}^{M}\sum_{n=2}^{M}\sum_{k=2}^{M}(\psi_{1m}\psi_{nk}-3\psi_{mn}\psi_{1k})r_1^3 r_m r_n r_k + (4\psi_{mn}\psi_{mk}+3\psi_{mm}\psi_{nk})r_m^3 r_n r_k r_1 \\
&+ (\psi_{mn}\kappa_k - 2\psi_{mm}\psi_{nk})r_m r_n r_k r_1 + 3\sum_{\substack{m=2\\m\neq n\neq k\neq q}}^{M}\sum_{n=2}^{M}\sum_{k=2}^{M}\sum_{q=2}^{M}\psi_{mn}\psi_{kq}r_m^2 r_n r_k r_q,
\end{aligned} \tag{6.50}$$

where

$$\psi_{mn}=\begin{cases}\sum_{i=1}^{M-1}\sum_{j=1}^{M-1}[\Pi_{\mathbf{V}}^{\perp}]_{i,j}, & \text{if } m=n=1,\\ \left[\Pi_{\mathbf{V}}^{\perp}\right]_{m-1,m-1}, & \text{if } 2\leq m=n\leq M,\\ -\sum_{i=1}^{M-1}[\Pi_{\mathbf{V}}^{\perp}]_{i,m-1}, & \text{if } m=1, 2\leq n\leq M\\ \left[\Pi_{\mathbf{V}}^{\perp}\right]_{m-1,n-1}, & \text{if } 2\leq m\neq n\leq M,\end{cases} \tag{6.51}$$

$$\kappa_{m}=\begin{cases}-\sum_{i=1}^{M-1}\sum_{j=1}^{M-1}[\Pi_{\mathbf{V}}^{\perp}]_{i,j}[\mathbf{b}]_{j}, & \text{if } m=1,\\ \sum_{i=1}^{M-1}[\Pi_{\mathbf{V}}^{\perp}]_{i,m-1}[\mathbf{b}]_{j}, & \text{if } 2\leq m\leq M,\end{cases} \tag{6.52}$$

and $\chi=\|\Pi_{\mathbf{V}}^{\perp}\mathbf{b}\|_2^2$. Using Definition 6.1, we parameterize the polynomial in the first constraint of (6.48) as

$$v\mathcal{J}(\mathbf{r})-\mathcal{F}(\mathbf{r})=\boldsymbol{\phi}^T\mathbf{g}_6([\mathbf{r},v]^T) \tag{6.53}$$

where $\boldsymbol{\phi}$ is the vector of the coefficients corresponding to the monomial basis $\mathbf{g}_6([\mathbf{r},v]^T)$, which is readily obtained from (6.50). We state the SDP equivalent of (6.48) in the following theorem.

Theorem 6.2. *Given the scalars $r_1, r_2, \cdots, r_M$ and integers $\{\nu_i\}_{i=1}^{M}$, define $\mathcal{K}:\mathbb{R}^{M+1}\rightarrow\mathbb{R}$ as $\mathcal{K}(r_1^{\nu_1}r_2^{\nu_2}\cdots r_M^{\nu_M}v^{\nu_{M+1}})=\mu_{\nu_1\nu_2\cdots\nu_{M+1}}$ such that $\mathcal{K}(1)=\mu_{00\cdots0}=1$. Construct the matrices*

$$\mathbf{T}_p(\boldsymbol{\mu})=\mathcal{K}\big(\mathbf{g}_p([\mathbf{r},v]^T)\mathbf{g}_p^T([\mathbf{r},v]^T)\big), \tag{6.54}$$

$$\mathbf{T}_{p-1}^{m}(\boldsymbol{\mu})= \tag{6.55}$$

$$\begin{cases}\mathcal{K}\big(\mathbf{g}_{p-1}([\mathbf{r},v]^T)\mathbf{g}_{p-1}^T([\mathbf{r},v]^T)w_m(r_m-\lambda_m)\big), & \text{if } 1\leq m\leq M,\\ \mathcal{K}\big(\mathbf{g}_{p-1}([\mathbf{r},v]^T)\mathbf{g}_{p-1}^T([\mathbf{r},v]^T)r_m\big), & \text{if } M+1\leq m\leq 2M,\\ \mathcal{K}\big(\mathbf{g}_{p-1}([\mathbf{r},v]^T)\mathbf{g}_{p-1}^T([\mathbf{r},v]^T)(v_{\max}-v)\big) & \text{if } m=2M+1,\end{cases}$$

and

$$\mathbf{T}_{p-3}(\boldsymbol{\mu})=\mathcal{K}\big(\mathbf{g}_{p-3}([\mathbf{r},v]^T)\mathbf{g}_{p-3}^T([\mathbf{r},v]^T)\boldsymbol{\phi}^T\mathbf{g}_6([\mathbf{r},v]^T)\big). \tag{6.56}$$

Then, there exists an integer $p\geq 3$ for which the optimization problem (6.48) is equivalent to

$$\begin{aligned}\underset{\boldsymbol{\mu}}{\text{minimize}}\quad & \boldsymbol{\mu}_{00\cdots01}\\ \text{s.t.}\quad & \mathbf{T}_p(\boldsymbol{\mu})\succeq\mathbf{0},\\ & \mathbf{T}_{p-3}(\boldsymbol{\mu})\succeq\mathbf{0},\\ & \mathbf{T}_{p-1}^{m}(\boldsymbol{\mu})\succeq\mathbf{0},\quad 1\leq m\leq 2M+1,\end{aligned} \tag{6.57}$$

such that the minimizer of (6.48) *is*

$$[r_1^\star, r_2^\star, \cdots, r_M^\star, v^\star]^T = [\mu_{10\cdots00}^\star, \mu_{01\cdots00}^\star, \cdots, \mu_{00\cdots10}^\star, \mu_{00\cdots01}^\star]^T. \tag{6.58}$$

Proof. See Appendix C.3. □

Remark 6.2. Note that the number of optimization variables in (6.57) is equal to $\binom{M+2p+1}{2p}$ which could be very large even for moderate values of the number of sensors M and the relaxation order p. Therefore, even though this method is able to attain the global minimum, it could become computationally expensive in the practical scenarios.

6.4.3 Sub-Optimal Localization with One-Bit Nodal Range Estimates

It is possible to reduce the computational complexity of the Lasserre's SDP method by trading off the optimality. We now present such a sub-optimal approach by iteratively solving (6.40) through alternating minimizations over $\boldsymbol{\theta}$, r_1 and $\bar{\mathbf{r}}$. Although this method, that we call ANTARES standing for iterative joint r*AN*ge-*TAR*get location *E*Stimation, achieves only a local minimum, its computationally efficiency is significantly higher than SDP.

Denote $\boldsymbol{\theta}^{(k)}$, $r_1^{(k)}$ and $\bar{\mathbf{r}}^{(k)}$ to be the values of the parameters $\boldsymbol{\theta}$, r_1 and $\bar{\mathbf{r}}$ at the k-th iteration, respectively. Given $\boldsymbol{\theta}^{(k)}$ and $r_1^{(k)}$, using (6.41), the problem in (6.40) with respect to $\bar{\mathbf{r}}$ at the $(k+1)$-th iteration becomes

$$\begin{aligned} \underset{\bar{\mathbf{r}}}{\text{minimize}} \quad & \sum_{m=2}^{M} \left(\frac{(r_m - r_1^{(k)})^2}{2} + [\boldsymbol{\theta}^{(k)}]_4 (r_m - r_1^{(k)}) + \zeta_m^{(k)} \right)^2 \\ \text{s.t.} \quad & w_m(r_m - \lambda_m) \geq 0, \quad 2 \leq m \leq M, \\ & r_m \geq 0, \qquad\qquad\quad 2 \leq m \leq M, \end{aligned} \tag{6.59}$$

where $\zeta_m^{(k)} = [\mathbf{V}\overline{\boldsymbol{\theta}}^{(k)}]_{m-1} - [\mathbf{b}]_{m-1}$ with $\overline{\boldsymbol{\theta}}^{(k)} = \begin{bmatrix} [\boldsymbol{\theta}^{(k)}]_1 & [\boldsymbol{\theta}^{(k)}]_2 & [\boldsymbol{\theta}^{(k)}]_3 \end{bmatrix}^T$. The global minimizer of (6.59) gives the update of $\bar{\mathbf{r}}^{(k)}$ as $\bar{\mathbf{r}}^{(k+1)}$ to be used in the next iteration. Observe this optimization problem is separable in $r_2, r_3, \cdots, r_M$. Hence, we convert it into $M-1$ parallel optimization problems, each of which is

$$\begin{aligned} \underset{r_m}{\text{minimize}} \quad & \tfrac{1}{4} r_m^4 + \beta_m^{(k)} r_m^3 + \varsigma_m^{(k)} r_m^2 + \omega_m^{(k)} r_m + \eta_m^{(k)} \\ \text{s.t.} \quad & w_m(r_m - \lambda_m) \geq 0, \\ & r_m \geq 0, \end{aligned} \tag{6.60}$$

where

$$\beta_m^{(k)} = [\boldsymbol{\theta}^{(k)}]_4 - r_1^{(k)}, \tag{6.61a}$$

$$\varsigma_m^{(k)} = \frac{3(r_1^{(k)})^2}{2} - 3[\boldsymbol{\theta}^{(k)}]_4 r_1^{(k)} + ([\boldsymbol{\theta}^{(k)}]_4)^2 + \zeta_m^{(k)}, \tag{6.61b}$$

$$\omega_m^{(k)} = -(r_1^{(k)})^3 + 3[\boldsymbol{\theta}^{(k)}]_4 (r_1^{(k)})^2 - 2\left(([\boldsymbol{\theta}^{(k)}]_4)^2 + \zeta_m^{(k)}\right) r_1^{(k)} + 2[\boldsymbol{\theta}^{(k)}]_4 \zeta_m^{(k)}, \tag{6.61c}$$

$$\eta_m^{(k)} = \frac{(r_1^{(k)})^4}{4} - [\boldsymbol{\theta}^{(k)}]_4 (r_1^{(k)})^3 + \left(([\boldsymbol{\theta}^{(k)}]_4)^2 + \zeta_m^{(k)}\right) (r_1^{(k)})^2 - 2[\boldsymbol{\theta}^{(k)}]_4 \zeta_m^{(k)} r_1^{(k)} + (\zeta_m^{(k)})^2. \tag{6.61d}$$

Since the objective and constraints in (6.60) are differentiable, the global minimizer of (6.60) belongs to a set of points which satisfy the following Karush-Kuhn-Tucker (KKT) conditions [98]:

$$r_m^3 + 3\beta_m^{(k)} r_m^2 + 2\varsigma_m^{(k)} r_m + \omega_m^{(k)} - \varrho_1 w_m - \varrho_2 = 0, \tag{6.62a}$$

$$w_m(r_m - \lambda_m) \geq 0, \tag{6.62b}$$

$$r_m \geq 0, \tag{6.62c}$$

$$\varrho_1 w_m (r_m - \lambda_m) = 0, \tag{6.62d}$$

$$\varrho_2 r_m = 0, \tag{6.62e}$$

$$\varrho_1 \geq 0, \tag{6.62f}$$

$$\varrho_2 \geq 0. \tag{6.62g}$$

where ϱ_1 and ϱ_2 are the KKT multipliers. From (6.62b)-(6.62g), there are three possibilities:

(i) $\varrho_1 > 0$ and $\varrho_2 = 0$: From (6.62e), under this condition, r_m must be equal to λ_m. Considering $r_m = \lambda_m$ and $\varrho_2 = 0$, it follows from (6.62a) that

$$\varrho_1 = w_m(\lambda_m^3 + 3\beta_m^{(k)} \lambda_m^2 + 2\varsigma_m^{(k)} \lambda_m + \omega_m^{(k)}). \tag{6.63}$$

Further, from $\varrho_1 > 0$, the point $r_m = \lambda_m$ satisfies the KKT conditions if

$$w_m(\lambda_m^3 + 3\beta_m^{(k)} \lambda^2 + 2\varsigma_m^{(k)} \lambda_m + \omega_m^{(k)}) > 0. \tag{6.64}$$

(ii) $\varrho_1 = 0$ and $\varrho_2 > 0$: From (6.62f), r_m must be zero under this scenario. Considering $r_m = 0$ and $\varrho_1 = 0$, it follows from (6.62a) and (6.62b) that $\varrho_2 = \omega_m^{(k)}$ and $w_m \leq 0$.

Hence, when $\varrho_2 > 0$, the point $r_m = 0$ satisfies the KKT conditions if

$$\begin{cases} \omega_m^{(k)} > 0, \\ w_m \leq 0. \end{cases} \tag{6.65}$$

(iii) $\varrho_1 = 0$ and $\varrho_2 = 0$: Under this scenario, the KKT conditions imply that r_m must be equal to the non-negative real roots of the following cubic equation

$$r_m^3 + 3\beta_m^{(k)} r_m^2 + 2\varsigma_m^{(k)} r_m + \omega_m^{(k)} = 0, \tag{6.66}$$

which satisfy (6.62b). The roots of (6.66) are given by

$$\digamma_q = -\frac{1}{3}\left(3\beta_m^{(k)} + \xi^q \Delta_2 + \frac{\Delta_0}{\xi^q \Delta_2}\right), \quad q \in \{0, 1, 2\}, \tag{6.67}$$

where $\xi = \frac{-1+j\sqrt{3}}{2}$, $\Delta_2 = \sqrt[3]{\frac{\Delta_1 \pm \sqrt{\Delta_1^2 - 4\Delta_0^3}}{2}}$, $\Delta_0 = 9(\beta_m^{(k)})^2 - 6\varsigma_m^{(k)}$ and $\Delta_1 = 54(\beta_m^{(k)})^3 - 54\beta_m^{(k)}\varsigma_m^{(k)} + 27\omega_m^{(k)}$. Further, it is well-known that amongst the KKT-compatible non-negative real roots of (6.66), only those which also satisfy the following second-order sufficient condition

$$3\digamma_q^2 + 6\beta_m^{(k)}\digamma_q + 2\varsigma_m^{(k)} \geq 0, \tag{6.68}$$

act as the minimizers of (6.60) [98]. As a result, we only consider the non-negative real root of (6.66) for which (6.62b) and (6.68) hold true.

Accordingly, the set of points which are the minimizers of (6.60) is derived by following (i) to (iii) above. Then, the global minimizer of (6.60) is the point in this set at which the value of the objective in (6.60) is the smallest.

Once $\bar{\mathbf{r}}^{(k+1)}$ is found, the problem (6.40) with respect to r_1 at the $(k+1)$-th iteration is cast as

$$\begin{aligned} \underset{r_1}{\text{minimize}} \quad & \tfrac{1}{4} r_1^4 + \beta_1^{(k)} r_1^3 + \varsigma_1^{(k)} r_1^2 + \omega_1^{(k)} r_1 + \eta_1^{(k)} \\ \text{s.t.} \quad & w_1(r_1 - \lambda_1) \geq 0, \\ & r_1 \geq 0, \end{aligned} \tag{6.69}$$

where

$$\beta_1^{(k)} = \frac{-1}{M-1}\sum_{m=2}^{M} r_m^{(k+1)} - [\boldsymbol{\theta}^{(k)}]_4, \tag{6.70a}$$

$$\varsigma_1^{(k+1)} = \frac{1}{M-1}\sum_{m=2}^{M} \frac{3}{2}(r_m^{(k+1)})^2 + 3[\boldsymbol{\theta}^{(k)}]_4 r_m^{(k+1)} + \zeta_m^{(k)} + ([\boldsymbol{\theta}^{(k)}]_4)^2, \tag{6.70b}$$

$$\begin{aligned}\omega_1^{(k)} =& \frac{-1}{M-1}\sum_{m=2}^{M}(r_m^{(k+1)})^3 + 3[\boldsymbol{\theta}^{(k)}]_4 (r_m^{(k+1)})^2 \\ &+ 2\left(([\boldsymbol{\theta}^{(k)}]_4)^2 + \zeta_m^{(k)}\right) r_m^{(k+1)} + 2[\boldsymbol{\theta}^{(k)}]_4 \zeta_m^{(k)},\end{aligned} \tag{6.70c}$$

$$\begin{aligned}\eta_1^{(k)} =& \frac{1}{M-1}\sum_{m=2}^{M} \frac{(r_m^{(k+1)})^4}{4} + [\boldsymbol{\theta}^{(k)}]_4 (r_m^{(k+1)})^3 \\ &+ \left(([\boldsymbol{\theta}^{(k)}]_4)^2 + \zeta_m^{(k)}\right)(r_m^{(k+1)})^2 + 2[\boldsymbol{\theta}^{(k)}]_4 \zeta_m^{(k)} r_m^{(k+1)} + (\zeta_m^{(k)})^2.\end{aligned} \tag{6.70d}$$

The global minimizer of (6.69) is attained by following a procedure similar to that of (6.60). From $\overline{\mathbf{r}}^{(k+1)}$ and $r_1^{(k+1)}$, the update of $\boldsymbol{\theta}^{(k)}$ at $(k+1)$-th iteration is

$$\boldsymbol{\theta}^{(k+1)} = \mathbf{G}^{\dagger^{(k+1)}} \mathbf{h}^{(k+1)}, \tag{6.71}$$

where $\mathbf{G}^{\dagger^{(k+1)}}$ and $\mathbf{h}^{(k+1)}$ are computed by substituting $\overline{\mathbf{r}}^{(k+1)}$ and $r_1^{(k+1)}$ for $\overline{\mathbf{r}}$ and r_1 in (6.31) and (6.32), respectively.

Algorithm 3 summarizes the steps of aforementioned ANTARES for joint estimation of $\boldsymbol{\theta}$ and $\mathbf{r}$. Note that each iteration of ANTARES requires solving one-dimensional optimizations, each of which has a closed-form solution. Further, the optimizations with respect to $r_2, r_3, \cdots, r_m$ are solved in parallel at each iteration. Hence, ANTARES is computationally highly efficient compared to (6.57).

6.4.4 CRB for Localization with One-Bit Nodal Range Estimates

We employ the CRB as a benchmark for assessing the estimation performance of the proposed optimal and sub-optimal algorithms. This is also useful for demonstrating the performance loss of one-bit quantization over the unquantized processing.

Assume that the estimation error term in $\widehat{r}_m = r_m + e_m$, i.e., e_m, follows a zero-mean Gaussian distribution with variance υ_m^2, $1 \le m \le M$. Then, $\widehat{r}_m$ is distributed as a Gaussian random variable with mean r_m and variance υ_m^2, $1 \le m \le M$. The $\widehat{r}_1, \widehat{r}_2, \cdots, \widehat{r}_M$ are statistically independent. Hence, the conditional probability density

Algorithm 3 Iterative joint r*ange*-*tar*get location *e*stimation (ANTARES)

Input: one-bit samples $\mathbf{w}$, threshold vector $\boldsymbol{\lambda}$, optimality tolerance parameters ε_1 and ε_2.

Output: Target location estimate $\widehat{\boldsymbol{\theta}}$, range estimate $\widehat{\mathbf{r}}$.

Initialization: Set $k=0$, $\boldsymbol{\theta}^{(0)} \in \mathbb{R}^{4\times 1}$ arbitrarily and $r_1^{(0)} \geq 0$ such that $w_1(r_1^{(0)} - \lambda_1) > 0$.
while $\|\boldsymbol{\theta}^{(k+1)} - \boldsymbol{\theta}^{(k)}\|_2^2 \geq \varepsilon_1$ and $\|\mathbf{r}^{(k+1)} - \mathbf{r}^{(k)}\|_2^2 \geq \varepsilon_2$ **do**
  if $2 \leq m \leq M$ **then**
    $\mathbb{S} \leftarrow \{\varnothing\}$.
    if (6.64) is fulfilled **then**
      $\mathbb{S} \leftarrow \{\lambda_m\} \cup \mathbb{S}$.
    else
      $\mathbb{S} \leftarrow \mathbb{S}$.
    end if
    if (6.65) is fulfilled **then**
      $\mathbb{S} \leftarrow \{0\} \cup \mathbb{S}$.
    else
      $\mathbb{S} \leftarrow \mathbb{S}$.
    end if
    for $q \leftarrow 0$ to 2 **do**
      $\mathbb{D} \leftarrow \{\varnothing\}$.
      Find $\digamma_q$ from (6.67).
      if $w_m(\digamma_q - \lambda_m) \geq 0$, $\digamma_q \geq 0$, $\Im\{\digamma_q\} = 0$ and $3\digamma_q^2 + 6\beta_m^{(k)}\digamma_q + 2\varsigma_m^{(k)} \geq 0$ **then**
        $\mathbb{D} \leftarrow \mathbb{D} \cup \digamma_q$.
      end if
    end for
    $\mathbb{S} \leftarrow \mathbb{D} \cup \mathbb{S}$.
    Find $r_{\text{opt}} \in \mathbb{S}$ at which the objective of (6.59) is minimized.
    $r_m^{(k+1)} \leftarrow r_{\text{opt}}$.
  end if
  Follow steps 4-17 to solve (6.69) for $r_1^{(k+1)}$.
  $\boldsymbol{\theta}^{(k+1)} \leftarrow \mathbf{G}^{\dagger(k+1)}\mathbf{h}^{(k+1)}$.
end while
$\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{(k+1)}$ and $\widehat{\mathbf{r}} = \mathbf{r}^{(k+1)}$.

function of $\mathbf{w}$ given $\mathbf{q} = [\delta^x, \delta^y, \delta^z, d_0, \upsilon_1, \upsilon_2, \cdots, \upsilon_M]^T \in \mathbb{R}^{(M+4)\times 1}$ is

$$f(\mathbf{w} \mid \mathbf{q}) = \prod_{m=1}^{M} \Phi\left(\frac{w_m(r_m - \lambda_m)}{\upsilon_m}\right), \tag{6.72}$$

where $\Phi(x) = \frac{1}{\sqrt{2}} \int_{\infty}^{x} e^{-u^2/2} \mathrm{d}u$. The CRB is the inverse of the Fisher Information Matrix (FIM) $\mathbf{I}(\mathbf{q})$, whose (i,j)-th element is [107]

$$[\mathbf{I}(\mathbf{q})]_{i,j} = \mathbb{E}\left\{ \frac{\partial \log f(\mathbf{w} \mid \mathbf{q})}{\partial [\mathbf{q}]_i} \frac{\partial \log f(\mathbf{w} \mid \mathbf{q})}{\partial [\mathbf{q}]_j} \right\}. \tag{6.73}$$

From (6.72), (6.2) and (6.3), the partial derivatives of the log-likelihood $\log f(\mathbf{w} \mid \mathbf{q})$ are

$$\frac{\partial \log f(\mathbf{w} \mid \mathbf{q})}{\partial \delta^x} = \frac{1}{\sqrt{2\pi}} \sum_{m=1}^{M} \frac{w_m(\delta^x - \delta_m^x) e^{-\frac{(r_m-\lambda_m)^2}{2\upsilon_m^2}}}{\upsilon_m d_m \Phi(\frac{w_m(r_m-\lambda_m)}{\upsilon_m})}, \tag{6.74}$$

$$\frac{\partial \log f(\mathbf{w} \mid \mathbf{q})}{\partial \delta^y} = \frac{1}{\sqrt{2\pi}} \sum_{m=1}^{M} \frac{w_m(\delta^y - \delta_m^y) e^{-\frac{(r_m-\lambda_m)^2}{2\upsilon_m^2}}}{\upsilon_m d_m \Phi(\frac{w_m(r_m-\lambda_m)}{\upsilon_m})}, \tag{6.75}$$

$$\frac{\partial \log f(\mathbf{w} \mid \mathbf{q})}{\partial \delta^z} = \frac{1}{\sqrt{2\pi}} \sum_{m=1}^{M} \frac{w_m(\delta^z - \delta_m^z) e^{-\frac{(r_m-\lambda_m)^2}{2\upsilon_m^2}}}{\upsilon_m d_m \Phi(\frac{w_m(r_m-\lambda_m)}{\upsilon_m})}, \tag{6.76}$$

$$\frac{\partial \log f(\mathbf{w} \mid \mathbf{q})}{\partial d_0} = \frac{1}{\sqrt{2\pi}} \sum_{m=1}^{M} \frac{w_m e^{-\frac{(r_m-\lambda_m)^2}{2\upsilon_m^2}}}{\upsilon_m d_m \Phi(\frac{w_m(r_m-\lambda_m)}{\upsilon_m})}, \tag{6.77}$$

$$\frac{\partial \log f(\mathbf{w} \mid \mathbf{q})}{\partial \upsilon_m} = -\frac{w_m(r_m - \lambda_m) e^{-\frac{(r_m-\lambda_m)^@}{2\upsilon_m^2}}}{\upsilon_m^2 \Phi(\frac{w_m(r_m-\lambda_m)}{\upsilon_m})}, \; 1 \le m \le M. \tag{6.78}$$

Inserting (6.74) to (6.78) into (6.74) and exploiting the statistical independence of $w_1, w_2, \cdots, w_M$, the elements of the FIM are

$$[\mathbf{I}(\mathbf{q})]_{1,1} = \sum_{m=1}^{M} \frac{(\delta_m^x - \delta^x)^2}{2\pi \upsilon_m^2 d_m^2} \left[\frac{e^{-\frac{(r_m-\lambda_m)^2}{\upsilon_m^2}}}{\Phi(\frac{r_m-\lambda_m}{\upsilon_m})} + \frac{e^{-\frac{(r_m-\lambda_m)^2}{\upsilon_m^2}}}{\Phi(\frac{-r_m+\lambda_m}{\upsilon_m})} \right], \tag{6.79}$$

$$[\mathbf{I}(\mathbf{q})]_{2,2} = \sum_{m=1}^{M} \frac{(\delta_m^y - \delta^y)^2}{2\pi \upsilon_m^2 d_m^2} \left[\frac{e^{-\frac{(r_m-\lambda_m)^2}{\upsilon_m^2}}}{\Phi(\frac{r_m-\lambda_m}{\upsilon_m})} + \frac{e^{-\frac{(r_m-\lambda_m)^2}{\upsilon_m^2}}}{\Phi(\frac{-r_m+\lambda_m}{\upsilon_m})} \right], \tag{6.80}$$

$$[\mathbf{I}(\mathbf{q})]_{3,3} = \sum_{m=1}^{M} \frac{(\delta_m^z - \delta^z)^2}{2\pi \upsilon_m^2 d_m^2} \left[\frac{e^{-\frac{(r_m-\lambda_m)^2}{\upsilon_m^2}}}{\Phi(\frac{r_m-\lambda_m}{\upsilon_m})} + \frac{e^{-\frac{(r_m-\lambda_m)^2}{\upsilon_m^2}}}{\Phi(\frac{-r_m+\lambda_m}{\upsilon_m})} \right], \tag{6.81}$$

$$[\mathbf{I}(\mathbf{q})]_{1,2} = \sum_{m=1}^{M} \frac{(\delta_m^x - \delta^x)(\delta_m^y - \delta^y) e^{-\frac{(r_m-\lambda_m)^2}{\upsilon_m^2}}}{2\pi \upsilon_m^2 d_m^2} \times \left[\frac{1}{\Phi(\frac{r_m-\lambda_m}{\upsilon_m})} + \frac{1}{\Phi(\frac{-r_m+\lambda_m}{\upsilon_m})} \right], \tag{6.82}$$

$$[\mathbf{I}(\mathbf{q})]_{1,3} = \sum_{m=1}^{M} \frac{(\delta_m^x - \delta^x)(\delta_m^z - \delta^z) e^{-\frac{(r_m-\lambda_m)^2}{\upsilon_m^2}}}{2\pi \upsilon_m^2 d_m^2} \times \left[\frac{1}{\Phi(\frac{r_m-\lambda_m}{\upsilon_m})} + \frac{1}{\Phi(\frac{-r_m+\lambda_m}{\upsilon_m})} \right], \tag{6.83}$$

$$[\mathbf{I}(\mathbf{q})]_{2,3} = \sum_{m=1}^{M} \frac{(\delta_m^y - \delta^y)(\delta_m^z - \delta^z) e^{-\frac{(r_m-\lambda_m)^2}{\upsilon_m^2}}}{2\pi \upsilon_m^2 d_m^2} \times \left[\frac{1}{\Phi(\frac{r_m-\lambda_m}{\upsilon_m})} + \frac{1}{\Phi(\frac{-r_m+\lambda_m}{\upsilon_m})} \right], \tag{6.84}$$

$$[\mathbf{I}(\mathbf{q})]_{4,4} = \sum_{m=1}^{M} \frac{e^{-\frac{(r_m-\lambda_m)^2}{\upsilon_m^2}}}{2\pi\upsilon_m^2 d_m^2} \left[\frac{1}{\Phi(\frac{r_m-\lambda_m}{\upsilon_m})} + \frac{1}{\Phi(\frac{-r_m+\lambda_m}{\upsilon_m})} \right], \tag{6.85}$$

$$[\mathbf{I}(\mathbf{q})]_{1,4} = \sum_{m=1}^{M} \frac{(\delta^x - \delta_m^x)}{2\pi\upsilon_m^2 d_m^2} \left[\frac{e^{-\frac{(r_m-\lambda_m)^2}{\upsilon_m^2}}}{\Phi(\frac{r_m-\lambda_m}{\upsilon_m})} + \frac{e^{-\frac{(r_m-\lambda_m)^2}{\upsilon_m^2}}}{\Phi(\frac{-r_m+\lambda_m}{\upsilon_m})} \right], \tag{6.86}$$

$$[\mathbf{I}(\mathbf{q})]_{2,4} = \sum_{m=1}^{M} \frac{(\delta^y - \delta_m^y)}{2\pi\upsilon_m^2 d_m^2} \left[\frac{e^{-\frac{(r_m-\lambda_m)^2}{\upsilon_m^2}}}{\Phi(\frac{r_m-\lambda_m}{\upsilon_m})} + \frac{e^{-\frac{(r_m-\lambda_m)^2}{\upsilon_m^2}}}{\Phi(\frac{-r_m+\lambda_m}{\upsilon_m})} \right], \tag{6.87}$$

$$[\mathbf{I}(\mathbf{q})]_{3,4} = \sum_{m=1}^{M} \frac{(\delta^z - \delta_m^z)}{2\pi\upsilon_m^2 d_m^2} \left[\frac{e^{-\frac{(r_m-\lambda_m)^2}{\upsilon_m^2}}}{\Phi(\frac{r_m-\lambda_m}{\upsilon_m})} + \frac{e^{-\frac{(r_m-\lambda_m)^2}{\upsilon_m^2}}}{\Phi(\frac{-r_m+\lambda_m}{\upsilon_m})} \right], \tag{6.88}$$

$$[\mathbf{I}(\mathbf{q})]_{m+4,m+4} = \frac{(r_m-\lambda_m)^2}{2\pi\upsilon_m^4} \left[\frac{e^{-\frac{(r_m-\lambda_m)^2}{\upsilon_m^2}}}{\Phi(\frac{r_m-\lambda_m}{\upsilon_m})} + \frac{e^{-\frac{(r_m-\lambda_m)^2}{\upsilon_m^2}}}{\Phi(\frac{-r_m+\lambda_m}{\upsilon_m})} \right], \ 1 \le m \le M, \tag{6.89}$$

$$[\mathbf{I}(\mathbf{q})]_{m+4,m'+4} = 0, \ 1 \le m \ne m' \le M, \tag{6.90}$$

$$[\mathbf{I}(\mathbf{q})]_{1,m+4} = \frac{(\delta_m^x - \delta^x)(r_m-\lambda_m)}{2\pi\upsilon_m^4} \left[\frac{e^{-\frac{(r_m-\lambda_m)^2}{\upsilon_m^2}}}{\Phi(\frac{r_m-\lambda_m}{\upsilon_m})} + \frac{e^{-\frac{(r_m-\lambda_m)^2}{\upsilon_m^2}}}{\Phi(\frac{-r_m+\lambda_m}{\upsilon_m})} \right], \ 1 \le m \le M, \tag{6.91}$$

$$[\mathbf{I}(\mathbf{q})]_{2,m+4} = \frac{(\delta_m^y - \delta^y)(r_m-\lambda_m)}{2\pi\upsilon_m^4} \left[\frac{e^{-\frac{(r_m-\lambda_m)^2}{\upsilon_m^2}}}{\Phi(\frac{r_m-\lambda_m}{\upsilon_m})} + \frac{e^{-\frac{(r_m-\lambda_m)^2}{\upsilon_m^2}}}{\Phi(\frac{-r_m+\lambda_m}{\upsilon_m})} \right], \ 1 \le m \le M, \tag{6.92}$$

$$[\mathbf{I}(\mathbf{q})]_{3,m+4} = \frac{(\delta_m^z - \delta^z)(r_m-\lambda_m)}{2\pi\upsilon_m^4} \left[\frac{e^{-\frac{(r_m-\lambda_m)^2}{\upsilon_m^2}}}{\Phi(\frac{r_m-\lambda_m}{\upsilon_m})} + \frac{e^{-\frac{(r_m-\lambda_m)^2}{\upsilon_m^2}}}{\Phi(\frac{-r_m+\lambda_m}{\upsilon_m})} \right], \ 1 \le m \le M, \tag{6.93}$$

$$[\mathbf{I}(\mathbf{q})]_{4,m+4} = \frac{(r_m-\lambda_m)}{2\pi\upsilon_m^4} \left[\frac{e^{-\frac{(r_m-\lambda_m)^2}{\upsilon_m^2}}}{\Phi(\frac{r_m-\lambda_m}{\upsilon_m})} + \frac{e^{-\frac{(r_m-\lambda_m)^2}{\upsilon_m^2}}}{\Phi(\frac{-r_m+\lambda_m}{\upsilon_m})} \right], \ 1 \le m \le M. \tag{6.94}$$

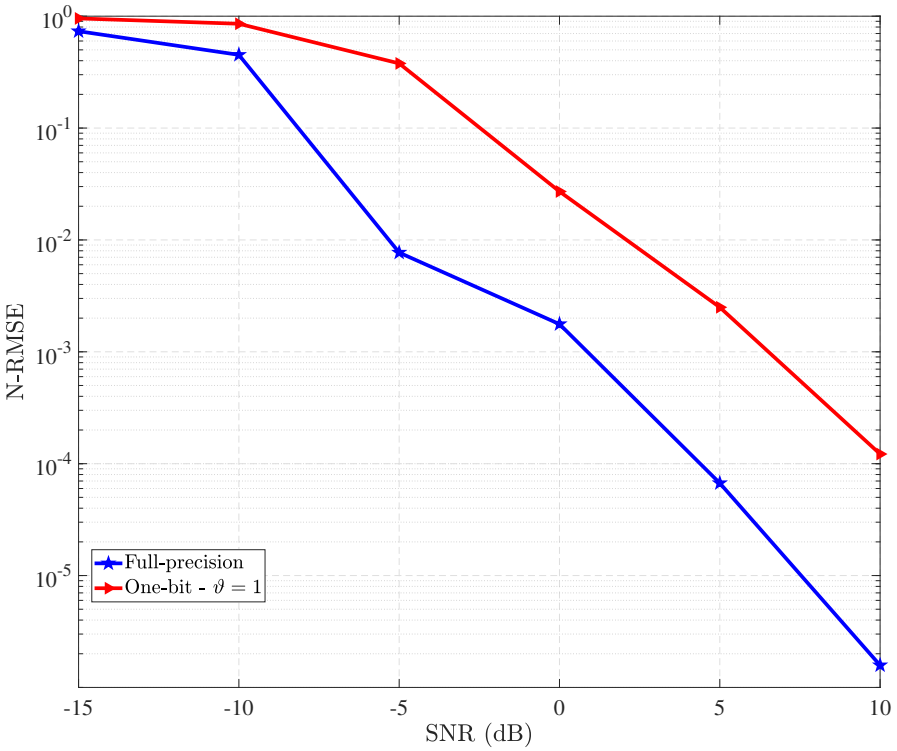


FIGURE 6.4: N-RMSE of the time-delay estimates versus the SNR with $L = 100$ and $\vartheta = 1$. The signal $s(t)$ is a $\pi/2$-BPSK modulated signal with bandwidth $B = 180$ KHz.

6.5 Numerical Experiments

We investigated the performance of our proposed method through numerical simulations. We also compared the performance of one-bit processing with full precision measurements. We used MATLAB CVX package to solve optimizations in (6.17) and (6.57) [174]. All the experiments are conducted under identical conditions under Matlab R2018a on a PC equipped with an operating system of Windows 10 64-bit, an Intel i7-6820HQ 2.70GHz CPU, and a 8GB RAM. Throughout all the experiments, we define signal-to-noise ratio (SNR) (in dB) at the m-th node as

$$\mathrm{SNR}_m = 10 \log_{10} \frac{|\alpha_m|^2 \|\mathbf{s}(\tau_m)\|^2}{\sigma_m^2}. \tag{6.95}$$

One-bit time-delay estimation: For 100 digital samples obtained at the Nyquist rate, i.e. $L = 100$ and $\vartheta = 1$, Fig. 6.4 shows the normalized root-mean-squared-error (N-RMSE) of the time-delay estimates, computed over 1000 Monte Carlo trials, with respect to SNR. This *estimation N-RMSE* is $\frac{\sqrt{\sum_{j=1}^{J}(\widehat{\tau}_{m,j}-\tau_m)^2}}{\tau_m J}$ where $\widehat{\tau}_{m,j}$ denotes the time-delay estimate at the j-th Monte Carlo trial and J is the number of Monte Carlo trials. We assume $s(t)$ to be $\pi/2$-BPSK-modulated with a raised cosine shaping filter of

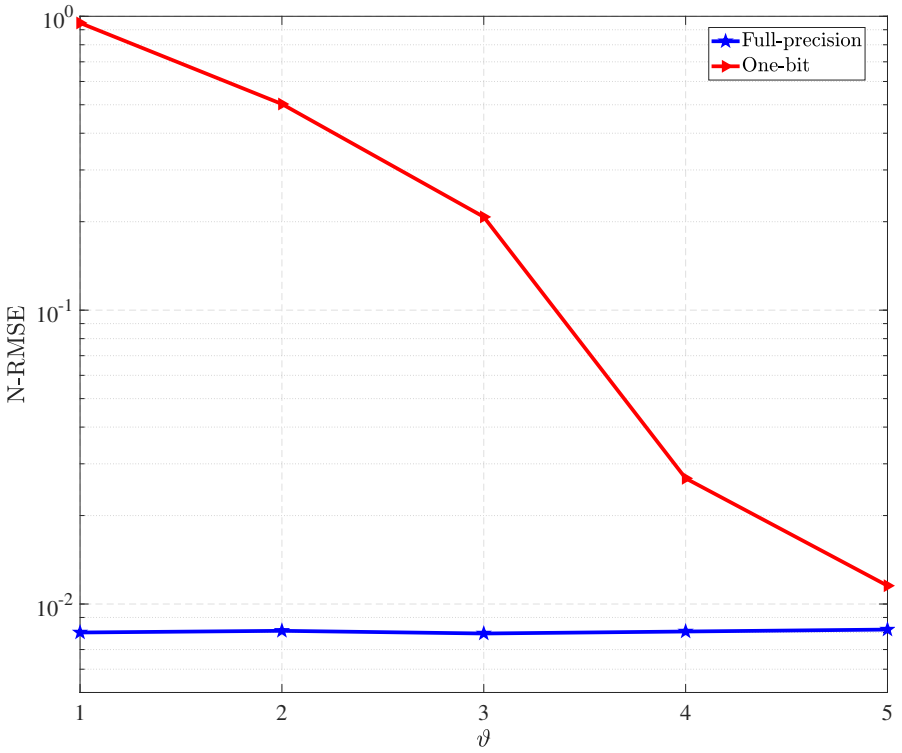


FIGURE 6.5: N-RMSE of the time-delay estimates versus the the oversampling factor ϑ with $L = 100$ and SNR $= -5$ dB. The signal $s(t)$ is a $\pi/2$-BPSK modulated signal with bandwidth $B = 180$ KHz.

the bandwidth 180 KHz and the roll-off factor 1. The temporal threshold $\boldsymbol{\gamma}_m$ is randomly drawn from a uniform distribution with support $[-A_{\max}, A_{\max}]$, where $A_{\max}$ denotes the maximum amplitude of the received signal at NB-IoT nodes. We observe that to achieve the same N-RMSE, the SNR should be about 5 dB higher for one-bit processing than the full-precision case.

Effect of oversampling: As discussed in Section 6.3.2, oversampling compensates the performance loss arising from the one-bit quantization scheme. Fig. 6.5 shows the N-RMSE of the time-delay estimates versus the oversampling factor, i.e., ϑ, at SNR $= -5$ dB. As predicted in theory, the N-RMSE of oversampled one-bit processing with $\vartheta = 5$ approaches that of the full-precision processing.

Localization with different node geometries: Next, we investigate our proposed localization method for various node placements. We consider three node geometries: uniform circular (Fig. 6.6), uniform linearly-spaced in an L-shape (Fig. 6.7), and random (Fig. 6.8). To show the performance over different ranges, we consider the performance of these geometries over small ($[-800$ m$, 800$ m$] \times [-800$ m$, 800$ m$]$), large ($[-2000$ m$, 2000$ m$] \times [-2000$ m$, 2000$ m$]$), and mid-size ($[-1200$ m$, 1200$ m$] \times$

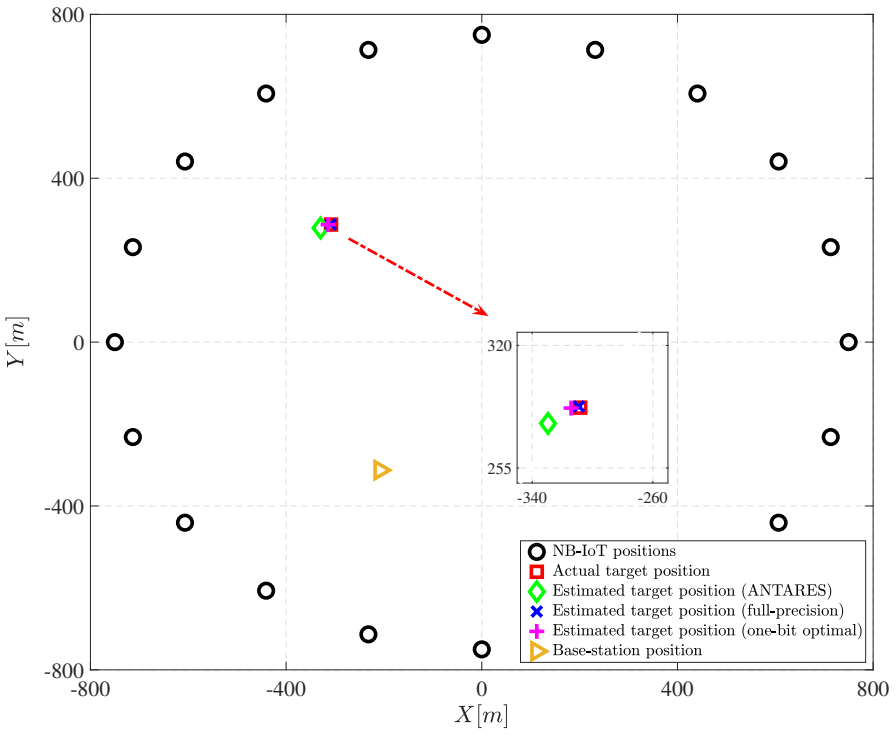


FIGURE 6.6: Localization with $M = 20$ NB-IoT nodes (black circles) uniformly spaced on a circle with radius of 800 m. The target-of-interest is randomly placed at $(-309 \text{ m}, 287 \text{ m})$. The SNR at all the NB-IoT nodes is 0 dB.

$[-1200 \text{ m}, 1200 \text{ m}]$) areas, respectively. In Fig. 6.6, the nodes were spaced on a circle with radius of 800 m and the target and the base-station were randomly placed at $[-309 \text{ m}, 287 \text{ m}]$ and $[-208, \text{ m}, -312 \text{ m}]$ (in X-Y Cartesian coordinate system), respectively. When the nodes were configured in L-shape and randomly, the target was randomly placed at $[371.7 \text{ m}, -338.4 \text{ m}]$ and $[-615.8 \text{ m}, -753.8 \text{ m}]$ and the base station was randomly located at $[-98 \text{ m}, 1112 \text{ m}]$ and $[-87 \text{ m}, 53 \text{ m}]$, respectively.

To consider the impact of the relative distances of the different nodes to the target of interest on the SNR, we generate the SNR at the m-th node ($m > 1$) as $\text{SNR}_m = \text{SNR}_1(\frac{d_m}{d_1})^2$ where SNR_1 denotes the SNR at the reference node, which is assumed to be 0 dB in Figs. 6.6, 6.7, and 6.8. The temporal thresholds and $s(t)$ are generated similar to Fig. 6.4. The maximum detectable range by NB-IoT nodes, i.e., $r_{\max}$, was considered to be 4000 m. The positive thresholds λ_m's were randomly drawn from 8 predetermined values over the interval $(0, r_{\max}]$. These thresholds are encoded with 3 bits and transmitted to the FC along with one-bit range information.

Our ANTARES algorithm estimates the target location with errors of 22.89, 23.87, and 21.52 m for circular, L-shape, and random geometries, respectively. This is very close to that of the optimal method given in Theorem 6.2, wherein the corresponding errors are 6, 9.4, and 7.81 m, respectively; the errors in the full-precision methods are 1 m, 1.2, and 1.06 m, respectively. This indicates the robustness of our method against distribution in

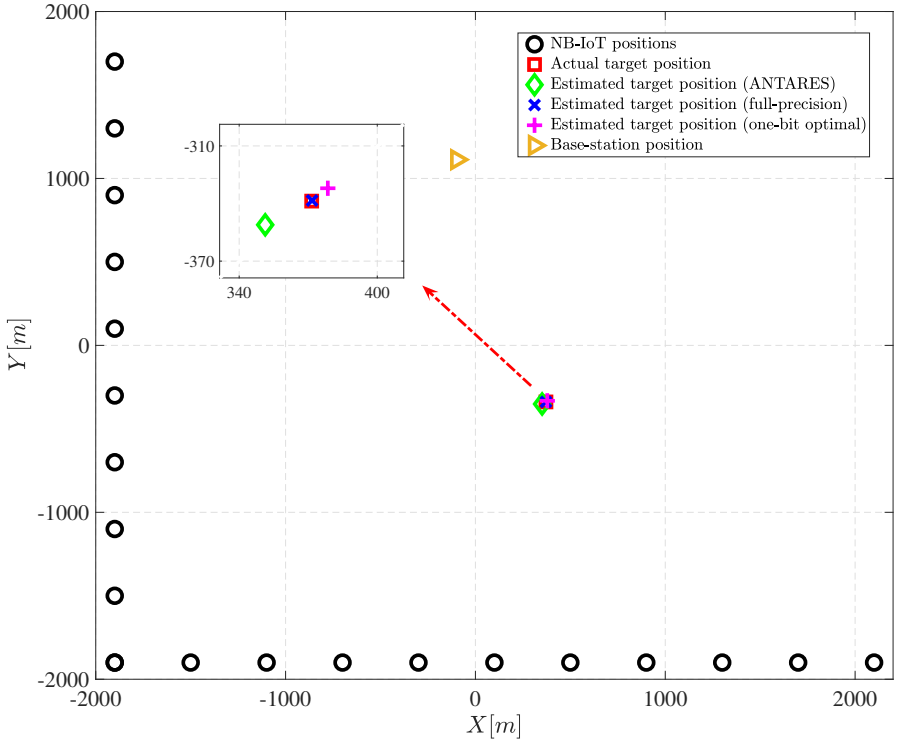


FIGURE 6.7: Localization with $M = 20$ NB-IoT nodes (black circles) linearly spaced in an L-shape. The target-of-interest is randomly placed at (371 m, −338 m). The SNR at the m-th node ($m > 1$) is $\text{SNR}_m = \text{SNR}_1 \left(\frac{d_m}{d_1}\right)^2$ with $\text{SNR}_1 = 0$ dB.

of NB-IoT nodes. In order to draw a comparison between the computational complexities of ANTARES and the optimal method, we take account of their corresponding run-times for the investigated scenarios in Figs. 6.6, 6.7, and 6.8, which are, respectively, 3.27 s, 3.63 s, and 3.91 s for ANTARES besides 81.39 s, 88.53 s, and 85.74 s for the optimal method. This implies that ANTARES is considerably more computationally efficient than the optimal method in Theorem 6.2.

Next, for the random geometry, we show the effect of decreasing SNR_1 to -5 dB (Fig. 6.9). The error with ANTARES algorithm now degrades to 59.85 m compared to 12.4 and 3.4 m observed in the optimal and full-precision approaches.

Statistical performance: Figs. 6.10a illustrates the *localization N-RMSE*, i.e. N-RMSE in the estimation of the target location, with respect to the number NB-IoT nodes M, defined as $\frac{\sqrt{\sum_{j=1}^{J}(\delta^x-\widehat{\delta}_j^x)^2+(\delta^y-\widehat{\delta}_j^y)^2}}{J\sqrt{\delta^{x\,2}+\delta^{y\,2}}}$, where $[\widehat{\delta}_j^x, \widehat{\delta}_j^y]^T$ denotes the target location estimate at the j-th Monte Carlo trial and J is the number of Monte Carlo trials. Figs. 6.10a plots the normalized-root-localization-CRB, i.e., $\sqrt{\frac{[\mathbf{I}^{-1}(\mathbf{q})]_{1,1}+[\mathbf{I}^{-1}(\mathbf{q})]_{2,2}}{\delta^{x\,2}+\delta^{y\,2}}}$ where $\mathbf{I}(\mathbf{q})$ is specified in Section 6.4.4. The nodes and targets were placed randomly over $([-800\text{ m}, 800\text{ m}] \times [-800\text{ m}, 800\text{ m}])$ area during each of the 200 Monte Carlo trials. The

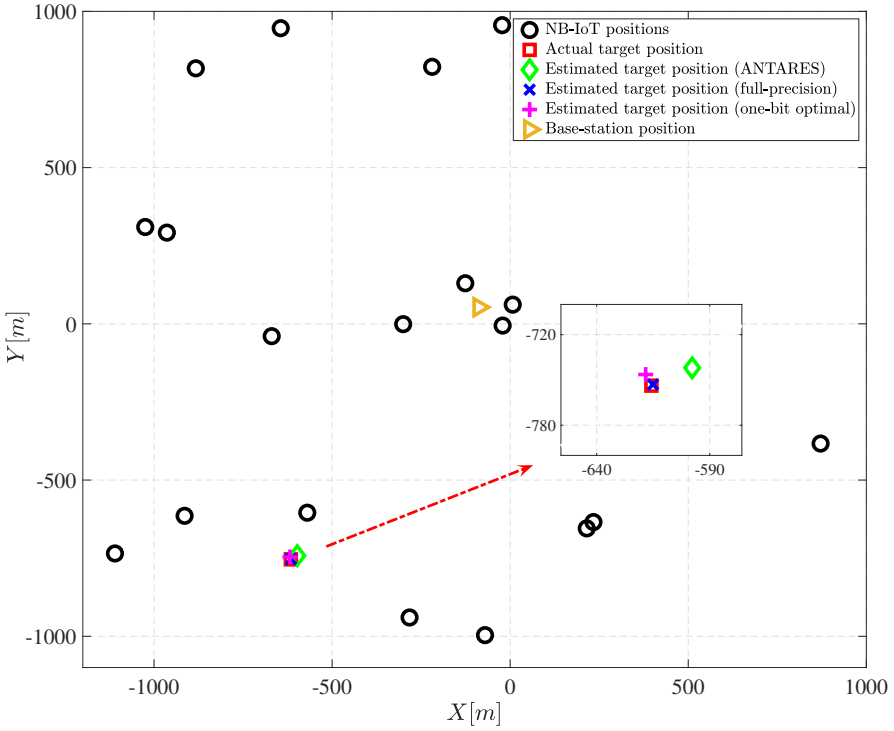


FIGURE 6.8: Localization with $M = 20$ NB-IoT nodes (black circles) randomly distributed over the area $[-1200\text{ m}, 1200\text{ m}] \times [-1200\text{ m}, 1200\text{ m}]$. The target-of-interest is randomly placed at $(1160\text{ m}, -340\text{ m})$. The SNR at the m-th node $(m > 1)$ is $\text{SNR}_m = \text{SNR}_1 \left(\frac{d_m}{d_1}\right)^2$ with $\text{SNR}_1 = 0$ dB.

SNR at the m-th node $(m > 1)$ is assumed to be $\text{SNR}_m = \text{SNR}_1 \left(\frac{d_m}{d_1}\right)^2$ with $\text{SNR}_1 = -2$ dB. Further, the temporal thresholds, $s(t)$ and λ_m's are generated similar to Figs. 6.4 and 6.7. We observe that the N-RMSEs of the proposed optimal and ANTARES methods improve with increase in M. The N-RMSE for the optimal method is very close to the normalized root of the CRB and it approaches to that of the full-precision when $M > 80$. It is also seen that the normalized CRB tends to the N-RMSEs of the full-precision at the high number of sensors. In addition, Fig 6.10b shows the relative N-RMSE, namely the difference in N-RMSE of the optimal and ANTARES methods as well as the normalized CRB relative to that of full-precision. We observe that the relative N-RMSE rises by 2.2%, 0.6% and 0.3% in case of ANTARES, optimal methods and the CRB, respectively, over the full-precision approach when $M = 20$. The observed difference in the estimation performance of ANTARES and optimal approaches arises from the fact that the alternating approach employed for ANTARES is guaranteed to converge to only a local minimum of the optimization problem in (44) [115], while the optimal method always provides the global minimum of (44).

The temporal thresholds were randomly generated in all experiments. Comparing the localization accuracy in Figs. 6.6-6.10 show that variations in temporal thresholds do not have considerable influence on the overall localization performance.

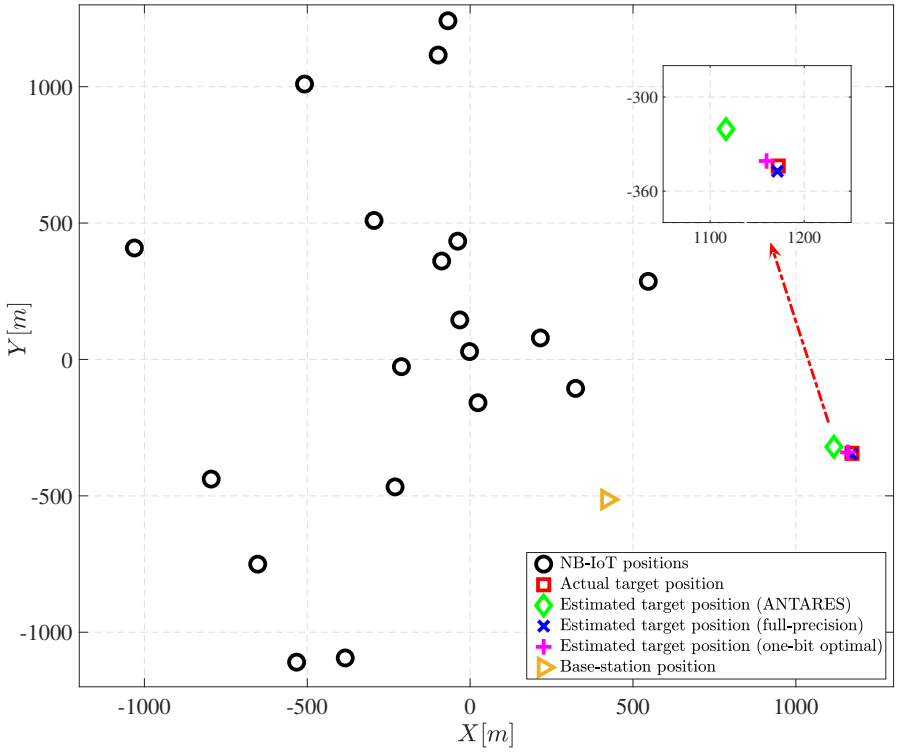
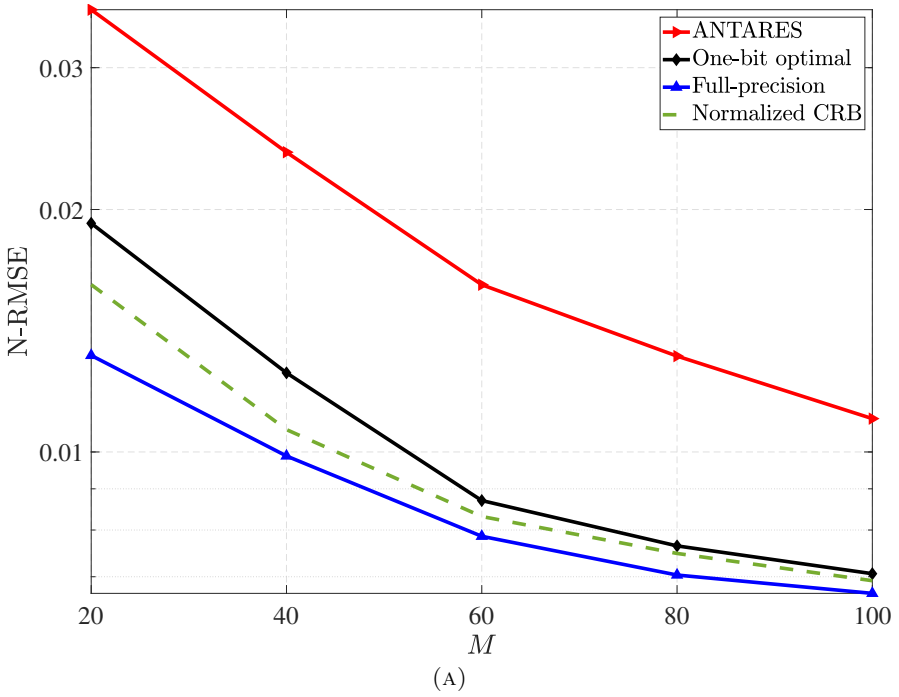


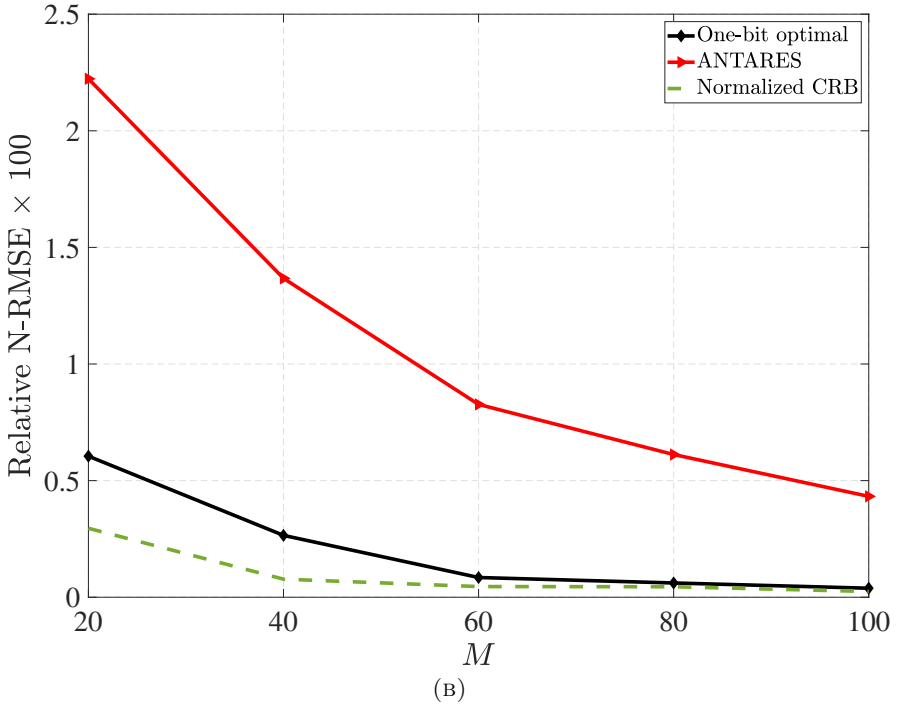
FIGURE 6.9: Localization with $M = 20$ NB-IoT nodes (black circles) randomly distributed within the area $[-1200\text{ m}, 1200\text{ m}] \times [-1200\text{ m}, 1200\text{ m}]$. The target-of -nterest is randomly placed at $(-618\text{ m}, -338\text{ m})$. The SNR at the m-th node ($m > 1$) is $\text{SNR}_m = \text{SNR}_1 \left(\frac{d_m}{d_1}\right)^2$ with $\text{SNR}_1 = -5$ dB.

6.6 Summary

In summary, the one-bit sampling offers an attractive solution to the challenges posed by the NB-IoT for location-based services. The one-bit samplers are integral to developing low cost and low power devices. We proposed a one-bit passive sensor array formulation to estimate the time-of-arrival in an NB-IoT network. The quantized samples of the estimates are then forwarded to an FC. We propose a novel method that casts the localization problem from aggregated quantized nodal estimates as a multivariate fractional optimization problem that we solve using the optimal Lasserre's SDP relaxation. We also propose the ANTARES algorithm as an alternative sub-optimal method with reduced computational complexity compared to Lasserre's. Our approach is helpful in addressing the problem of maintaining high localization accuracy while deploying reduced-rate ADCs at the nodes as well as limited-capacity NB-IoT links.



(A)



(B)

FIGURE 6.10: (a) N-RMSE and (b) Relative N-RMSE in the estimated target location with respect to the number of IoT devices M. The SNR at the m-th node ($m > 1$) is $\text{SNR}_m = \text{SNR}_1 \left(\frac{d_m}{d_1}\right)^2$ with $\text{SNR}_1 = -2$ dB.

Chapter 7

Sparsity-Aided Localization in asynchronous MIMO radar

7.1 Introduction

Target localization is one of the fundamental tasks of radar systems and it has received a considerable attention in the last few years [79–81]. In general, target localization techniques can be categorized under two main approaches with regard to the radar system architecture. The first approach develops the localization algorithms based on measurements received by a fully-synchronized collocated radar sensors [82, 83]. Although localization with fully-synchronized collocated is able to achieve high localization performance through providing the waveform diversity [84], they demand a complicated transceiver structure with expensive devices to enable the coherent processing required. However, many commercial applications, such as automotive radar, are required to localize the targets using cheap devices. This capability is enabled by the second approach based on distributed localization techniques where simple, independent radar nodes are used instead of a complicated collocated radar system. Distributed localization techniques exploit the angular (spatial) diversity, provided by spatially separated radar nodes, to estimate targets locations [84].

In addition to lack of support for coherent processing, the independent radar nodes may only provide coarse local Direction of Arrival (DoA) information. This is indeed the case when each of the nodes is equipped with a single or a few antenna elements, as envisaged in automotive applications [175], due to cost and implementation issues. Thus, the radar nodes have access only to fine range, or equivalently time delay, measurements. Several approaches have been proposed in the literature to estimate the targets locations in a

distributed manner using range measurements provided by distributed radar nodes [176–179]. Best Linear Unbiased Estimates (BLUE) of the target locations have been derived in [176] by linearizing the elliptic equations corresponding to the target positions. A two-stage weighted least square approach is proposed in [178], which is further demonstrated to achieve Cramer-Rao Lower bound (CRLB) in the high Signal-to-Noise-Ratio (SNR) regime.

Evidently, the performance of targer distributed localization relies heavily on perfect target detection at each of the nodes. Erroneous target detection can significantly degrade the localization performance regardless of the algorithm used. Further, since the radar nodes are closely located in automotive applications (e.g. on the bumper), small errors in range measurements can affect the localization. Therefore, high range resolution is a must in targer distributed localization for automotive application. However, achieving such a high range resolution using conventional methods demands high bandwidths which may not be affordable in practice. Thus, in such applications, the need for developing new algorithms, which are able to provide the desired range resolution without additional requirements on bandwidth, is seriously felt.

Towards achieving enhanced targer localization for automotive application, we propose a framework based on Sparse Stepped Frequency Waveforms (SSFW) and sparse sensing. Specifically, exploiting the sparsity in the target scene, the framework uses a two-step decentralized procedure for localization. In the first step, the target detection and super-resolution range estimation are undertaken. The link between range resolution and the bandwidth is broken by proposing a SSFW and then estimating the target ranges using the sparse sensing technique. It is shown that the proposed approach provides super-resolution estimates of the targets ranges enabling multi-target resolution with affordable bandwidths. The waveform design parameters offer additional degrees-of-freedom for efficient sparse sensing. In the second step, the precise range measurements are transmitted to a Fusion Center (FC), where the DoA estimation is performed. The Maximum Likelihood Estimation (MLE) is used in the FC for the DoA estimation and subsequent localization.

Chapter organization: Section 7.2 describes the system and basic assumptions. The proposed algorithm for sparsity-aided distributed target localization is given in Section 7.3. We validate our proposed localization algorithm through numerical experiments in Section Section 7.4. Finally, Section 7.5 concludes the chapter.

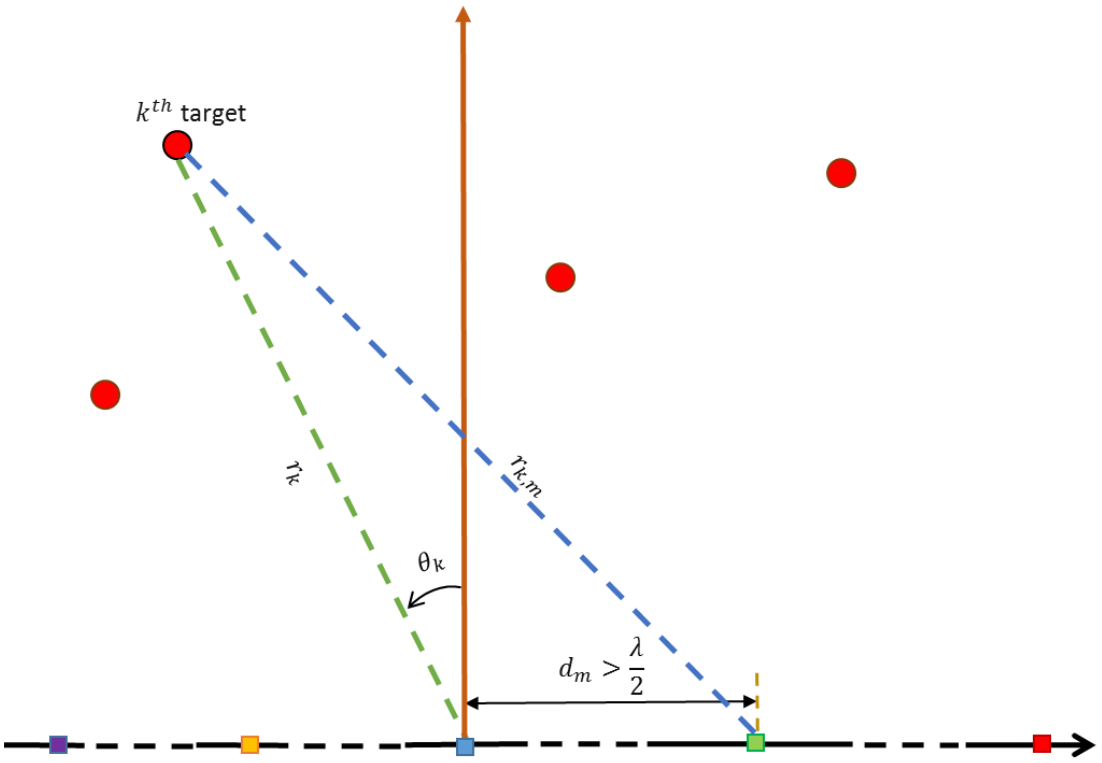


FIGURE 7.1: Distributed radar system model

7.2 Basic Assumptions and System Model

We consider a distributed radar system (see Fig. 7.1) including M mono-static radar nodes which are linearly distributed so that the m^{th} radar node is located at distance d_m from some reference point. The reference point can also be one of the radar nodes. The distances of the radar nodes from the reference point is assumed to be greater than half of the wavelength of operation. Each radar node is composed of D patch antennas embedded in an integrated chip, which beamform the transmit signal in a particular direction at each sensing time. The transmit beams are designed such that the beams of a subset $\mathcal{A}$ of the set of radar nodes intersect at each sensing time, with $|\mathcal{A}| \geq \kappa$. No synchronization is assumed between the radar nodes, although the patch antennas in each node are assumed to be synchronous. It is also supposed that K stationary point targets are present in the area covered by the beams of radar nodes within the subset $\mathcal{A}$. The k^{th} is located at distance r_k and azimuth angle θ_k with respect to the reference point. The targets are assumed to be in the far-field, i.e., $r_k \gg d_m$. Such a system model is valid, for example, in automotive applications where the radar transceivers are installed on car-bumpers. We note that the data association problem, arising in multi-target scenario, is not considered in this work. Indeed, it is assumed that the signals backscattered from different targets are properly associated in a preceding processing step.

We employ a SSFW for target illumination by each of the nodes. The SSFW is a variation of Stepped Frequency Waveforms (SFWs). A classical SFW (CSFW) is composed of P consecutive monotone narrowband pulses with duration T_c so that the frequency of the p^{th} pulse is assumed to be $f_p = f_c + p\triangle f$, where $p \in \mathcal{P} \doteq \{1, 2, \cdots, P\}$, f_c is the carrier frequency and $\triangle f$ denotes the frequency step size. The bandwidth and Coherent Processing Interval (CPI) of the CSFW are $P\Delta f$ and PT_c, respectively. The SSFW can be made out of the CSFW by transmitting only portions of the monotone narrowband pulses of the CSFW in a CPI. Indeed, the frequency of the n^{th} pulse of the SSFW is $f_n = f_c + \mathcal{S}(n)\triangle f$, where $\mathcal{S} \subseteq \mathcal{P}$ with $|\mathcal{P}| = N < P$ and $\mathcal{S}(s)$ denotes the n^{th} element of $\mathcal{S}$. The subset $\mathcal{S}$ can be selected from $\mathcal{P}$ in a deterministic or random fashion so as to achieve a desired inference performance [180]. The bandwidth and CPI of the SSFW are respectively $N\Delta f$ and NT_c, which are considerably smaller than those of the CSFW.

Without loss of generality, it can be supposed that all transmit nodes use the same carrier frequency[1]. Hence, the n^{th} pulse transmitted by the m^{th} radar node, where $m \in \mathcal{A}$, can be modeled as

$$T_m(n,t) = s_m\left(t - nT_c\right) \exp\left\{j2\pi\left(f_c + \mathcal{S}(n)\triangle f\right)\left(t - nT_c\right)\right\}, \tag{7.1}$$

where $0 \leq n \leq N-1$, and $s_m(t)$, $0 < t < T_c$, denotes the complex envelope of the monotone narrowband pulse transmitted by the m^{th} radar node. The transmitted waveforms are assumed to be orthogonal across the M transmit nodes, i.e., $\frac{1}{T_c}\int_0^{T_c} s_m(t)s^*_{m'}(t-\tau)dt$, $m \neq m'$, is assumed to be negligible compared to $\frac{1}{T_c}\int_0^{T_c} s_m(t)s^*_m(t-\tau)dt$. Assuming that the propagation is non-dispersive, the received signal at the location of the k^{th} target equals

$$v_k(n,t) = \sum_{m=1}^{M} s_m\left(t - nT_c - \tau_{m,k}\right) \exp\left\{j2\pi\left(f_c + \mathcal{S}(n)\triangle f\right)\left(t - nT_c - \tau_{k,m}\right)\right\}, \tag{7.2}$$

where $\tau_{k,m} = \frac{r_{k,m}}{c}$ and c denotes the speed of light in vacuum. The backscattered signal from K targets at the i^{th} radar node, where $m \in \mathcal{A}$, is then described by

$$\begin{aligned} y_i(n,t) = &\sum_{k=1}^{K}\sum_{m=1}^{M} \alpha_k s_m\left(t - nT_c - \tau_{k,m} - \tau_{k,i}\right) \\ &\times \exp\left\{j2\pi\left(f_c + \mathcal{S}(n)\triangle f\right)\left(t - nT_c - \tau_{k,m} - \tau_{k,i}\right)\right\} + \eta_i(n,t), \end{aligned} \tag{7.3}$$

where α_k is referred to as the complex amplitude proportional to the radar cross section (RCS) of the k^{th} target, and $\eta_i(n,t)$ is the zero-mean additive white Gaussian noise at

[1]It should be noted that the radar nodes are assumed to use the same carrier frequency for notational brevity. However, in general, there may be some offset between the carrier frequencies of radar nodes, which does not affect the validity of the ensuing derivations and system model.

i^{th} radar node, whose variance is σ^2.

The received signal at the i^{th} radar node, i.e., $y_i(n,t)$, is downconverted into the baseband through being multiplied by $e^{-j2\pi(f_c+\mathcal{S}(n)\triangle f)(t-nT_c)}$, and is then cross-correlated with the complex envelope of the monotone narrowband pulse transmitted by the same node at time $t \in [nT_c, (n+1)T_c]$, i.e., $s_i(t-nT_c)$, yielding

$$z_i(n) = \frac{1}{T_c}\int_{nT_c}^{(n+1)T_c} \tilde{y}_i(n,t)s_i^*(t-nT_c)\mathrm{d}t$$
$$= \overbrace{\sum_{k=1}^{K} \frac{\alpha_k \exp\left\{-j4\pi\left(f_c+\mathcal{S}(n)\triangle f\right)\tau_{k,i}\right\}}{T_c}\int_{nT_c}^{(n+1)T_c} s_i\left(t-nT_c-2\tau_{k,i}\right)s_i^*(t-nT_c)\mathrm{d}t}^{A_1} +$$
$$\overbrace{\sum_{k=1}^{K}\sum_{m\neq i} \frac{\alpha_k \exp\left\{-j2\pi\left(f_c+\mathcal{S}(n)\triangle f\right)\left(\tau_{k,m}+\tau_{k,i}\right)\right\}}{T_c}\int_{nT_c}^{(n+1)T_c} s_m\left(t-nT_c-\tau_{k,m}-\tau_{k,i}\right)s_i^*(t-nT_c)\mathrm{d}t}^{A_2}$$
$$+\tilde{\eta}_i(n). \tag{7.4}$$

where $\tilde{\eta}_i(n) = \frac{1}{T_c}\int_{nT_c}^{(n+1)T_c}\eta_i(t,n)s_i^*(t-nT_c)\mathrm{d}t$. Due to the orthogonality of the transmitted waveforms across the radar nodes, the term A_2 in (7.4) is negligible compared to the term A_1. Therefore, (7.4) can be simplified as

$$z_i(n) = \sum_{k=1}^{K}\beta_{k,i}\exp\left\{-j2\pi\left(\frac{2r_{k,i}\triangle f}{c}\mathcal{S}(n)\right)\right\} + \tilde{\eta}_i(n), \tag{7.5}$$

where $\beta_{k,i} = \alpha_k a(\frac{2r_{k,i}}{c})\exp\left\{-j2\pi\left(\frac{2r_{k,i}}{\lambda_c}\right)\right\}$, $a(\tau) = \frac{1}{T_c}\int_{nT_c}^{(n+1)T_c} s_i\left(t-nT_c-\tau\right)s_i^*(t-nT_c)dt$, and λ_c denotes the wavelength of the transmitted signal.

7.3 Proposed Localization Algorithm

In this section, we aim to determine the locations of the targets with respect to the reference point. We propose a two-stage decentralized localization algorithm. In the first step, the target detection and precise range estimation are separately done in each radar node. Then, the range measurements are transmitted to a FC for DoA estimation and subsequent localization.

7.3.1 Range Estimation

Let us discretize the range plane on a fine grid so that

$$2r_{k,i} = l\vartheta, \qquad 0 \leq l \leq L-1, \tag{7.6}$$

where ϑ and $\triangle f$ are chosen so that $\frac{\vartheta \triangle f}{c} = \frac{1}{L}$ with $L \gg N$. We can then rewrite (7.5) as

$$z_i(n) = \sum_{l=0}^{L-1} x_l \exp\{-j2\pi l\mathcal{S}(n)/L\} + \tilde{\eta}_i(n), \tag{7.7}$$

where $x_l = \alpha_k a(l\vartheta/c)e^{-j2\pi l\vartheta \triangle f/c}$ if a target is present at the range $l\vartheta$ and zero otherwise. Stacking $z_i(n)$'s into the vector $\mathbf{z}_i = [z_1(0), \cdots, z_1(N-1)]^T$, we can express (7.7) in the matrix form as

$$\mathbf{z}_i = \mathbf{\Psi}\mathbf{x}_i + \tilde{\eta}_i, \tag{7.8}$$

where $\mathbf{x}_i = [x_0, \cdots, x_{L-1}]^T$ and $\mathbf{\Psi}$ is a Vandermonde matrix with $[\mathbf{\Psi}]_{n,l} = e^{\frac{-j2\pi l\mathcal{S}(n)}{N}}$, i.e., it is composed of N rows of the $L \times L$ Discrete Fourier Transform (DFT) matrix. The estimate of target ranges can be found by solving the sparse problem (7.8) such that non-zero index l denotes a target with delay $l\vartheta$. In principle, the sparse vector $\mathbf{x}$ can be recovered by solving the ℓ_1-norm optimization problem

$$\hat{\mathbf{x}}_i = \min\|\mathbf{x}_i\|_1 \quad \text{s.t.} \quad \|\mathbf{z}_i - \mathbf{\Psi}\mathbf{x}_i\|_2 \leq \varepsilon, \tag{7.9}$$

which is also referred to as the least absolute shrinkage and selection operator (LASSO). The LASSO is a convex problem and converge to a global solution with a high probability in polynomial time if the matrix $\mathbf{\Psi}$ satisfies the Restricted Isometry Property (RIP) of order $2K$ with the $2K$-Restricted Isometry Constant (RIC) $\delta_{2K} < \frac{2}{3+\sqrt{7/4}}$ [181]. The K-RIC of a matrix $\mathbf{A}$, i.e., δ_K , is the smallest number such that the inequality

$$(1-\delta_K)\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1+\delta_K)\|\mathbf{x}\|_2^2, \tag{7.10}$$

holds for all K-sparse vectors $\mathbf{x}$ and a matrix $\mathbf{A}$ fulfills the RIP of order K if $\delta_K < 1$ [181]. Hence, in order to the range recovery is guaranteed, the subset $\mathcal{S}$ should be chosen such that the RIP of order $2K$ [182] is satisfied. As mentioned in Section 7.2, the subset $\mathcal{S}$ can be chosen from the set $\mathcal{P}$ in a deterministic or random way. In current work, we select the subset $\mathcal{S}$ uniformly at random out of the set $\mathcal{P}$. In this case, it has been shown in [182] that the matrix $\mathbf{\Psi}$ obeys the RIP of order $2K$ with a high probability if $N \geq CK\log(L/K)$, where C is some constant.

7.3.2 DoA Estimation

After the distributed sparse processing, the recovered ranges are transmitted to a FC for DoA estimation. According to the law of cosine, we have

$$r_{k,m} = \sqrt{r_k^2 + d_m^2 \pm 2 r_k d_m \sin\theta_k}, \tag{7.11}$$

where $+$ and $-$ hold for the nodes on the left and right hand side of the reference point, respectively. Considering the far-field assumption, i.e. $r_k \gg d_m$, and the Taylor series expansion, (7.11) can be approximated by

$$r_{k,m} \simeq r_k \pm d_m \sin\theta_k. \tag{7.12}$$

Thus, if $\hat{r}_{k,m}$ denotes the recovered range of the kth target with respect to the m^{th} node and $\hat{r}_k$ denote the recovered range of the same target with respect to the reference point, we can argue that,

$$q_{k,i} = \hat{r}_{k,m} - \hat{r}_k = \pm d_m \sin\theta_k + e_m, \tag{7.13}$$

where e_m is the error proceeding from quantization as well as estimation error which can be generally modeled as a Gaussian random variable with variance ρ_m^2. Staking all $q_{k,m}$'s from different antenna for a particular target k into a vector $\mathbf{q}_k = [q_{k,1}, \cdots, q_{k,M-1}]$, we have

$$\mathbf{q}_k = \mathbf{w} \sin\theta_k + \mathbf{e}, \tag{7.14}$$

where $\mathbf{e} = [e_1, \cdots, e_{M-1}]^T$ and $\mathbf{w} = [w_1, \cdots, w_{M-1}]$ with $w_m = d_m$ if the i^{th} radar node is on the left hind side of the reference point and $w_m = -d_m$ otherwise. We can then recast the Maximum Likelihood Estimation (MLE) of $\gamma \doteq \sin\theta_k$ as the following optimization problem,

$$\begin{aligned} \underset{\gamma}{\text{argmin}} \quad & (\mathbf{q}_k - \mathbf{w}\gamma)^T \mathbf{\Sigma}^{-1} (\mathbf{q}_k - \mathbf{w}\gamma), \\ \text{s.t.} \quad & -1 \leq \gamma \leq 1, \end{aligned} \tag{7.15}$$

where $\mathbf{\Sigma} = \mathrm{E}\{\mathbf{e}\mathbf{e}^H\} = \mathrm{diag}\{\rho_1^2, \cdots, \rho_{M-1}^2\}$. The above constrained optimization problem is convex. Therefore, applying the Karush-Kuhn-Tucker (KKT) optimality conditions

[98] leads us to the following estimate of ν

$$\hat{\nu} = \begin{cases} \frac{\mathbf{w}^T\boldsymbol{\Sigma}^{-1}\mathbf{q}}{\mathbf{w}^T\boldsymbol{\Sigma}^{-1}\mathbf{w}} & -1 \leq \frac{\mathbf{w}^T\boldsymbol{\Sigma}^{-1}\mathbf{q}}{\mathbf{w}^T\boldsymbol{\Sigma}^{-1}\mathbf{w}} \leq 1 \\ 1 & \frac{\mathbf{w}^T\boldsymbol{\Sigma}^{-1}\mathbf{q}}{\mathbf{w}^T\boldsymbol{\Sigma}^{-1}\mathbf{w}} > 1 \\ -1 & \frac{\mathbf{w}^T\boldsymbol{\Sigma}^{-1}\mathbf{q}}{\mathbf{w}^T\boldsymbol{\Sigma}^{-1}\mathbf{w}} < -1 \end{cases} . \tag{7.16}$$

Taking account of the invariance property of the MLE [107], the estimate of θ is obtained as

$$\hat{\theta} = \arcsin \hat{\nu}. \tag{7.17}$$

Therefore, the location of the k^{th} in the polar coordinate system with respect to the reference point is obtained as $(\hat{r}_k, \hat{\theta})$

7.4 Simulation Results

In this section, we present numerical results to illustrate the main contributions of the chapter which relate to the enhanced performance of proposed localization algorithm arising from sparse sensing.

7.4.1 General Set

We consider an automotive scenario with 5 radar nodes, separated by a distance of 0.1 m, and each containing 4 phased array elements. The phased array in each node creates a wide beam with a width of $\pi/3$. We consider multiple point targets in the coverage of the beams ensuring that the overlap requirement mentioned in Section II is satisfied. The carrier frequency and the frequency step size are $f_c = 79$ GHz and $\Delta f = 10$ MHz, respectively. Further, we consider a challenging situation with the SNR being 0 dB. The SSFW is generated by selecting a subset $\mathcal{S}$, with $|\mathcal{S}| = \frac{\text{bandwidth}}{\Delta f}$, uniformly at random out of the set $P = \{1, 2, \cdots, 1000\}$. The number of grid points, i.e., L, is 1000. This set up is simulated using the phased-array toolbox of Matlab.

7.4.2 Enhanced Range Resolution for given bandwidth

We consider the support recovery error [183] as figure of merit towards illustrating the ability of the proposed scheme in achieving higher range resolution for a given bandwidth. The support recovery error is defined as the error event when at least one target is estimated erroneously [183]. Fig. 7.2 depicts the support recovery error

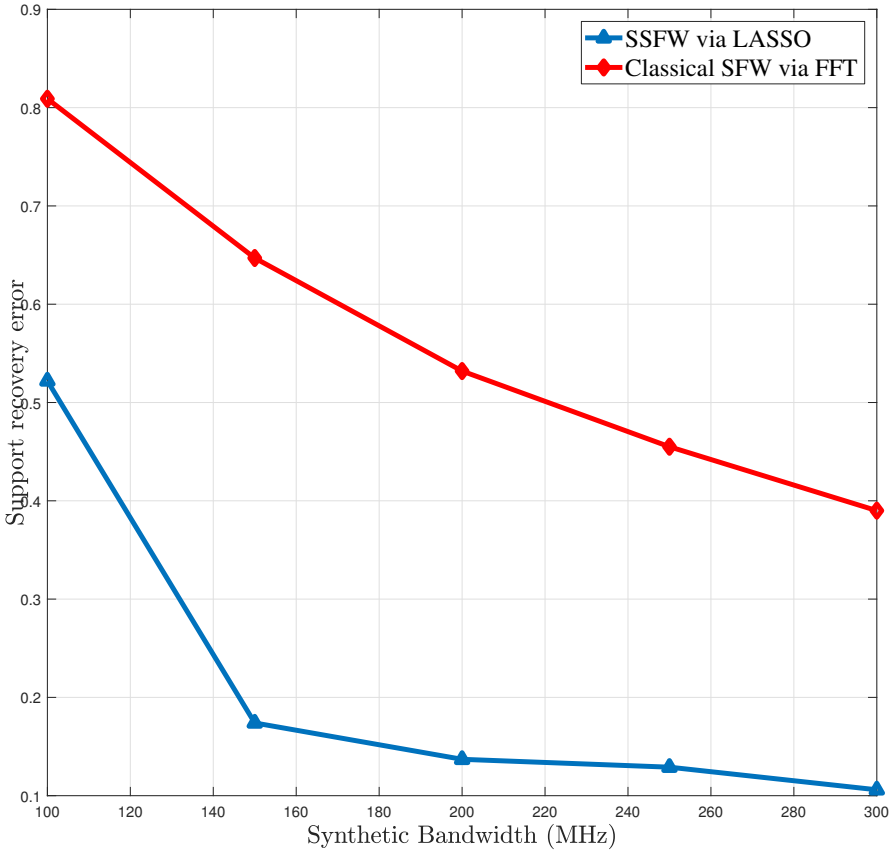


FIGURE 7.2: Support recovery error as a function of the synthetic bandwidth.

versus bandwidth for the conventional and proposed technique. Fig. 7.3 compares the localization performance of the SSFW and the SFW. We assumed that two targets are present and the synthetic bandwidth of the SFW is 7 times of the SSFW. It can be seen that the SSFW localizes the two targets properly, while using much lower synthetic bandwidth than the SFW. On the other hand, SFW estimates the location of one of the targets erroneously and fails to detect the second target although benefits from higher synthetic bandwidth. 5000 Monte Carlo lo repetitions were performed using independent realization of $K = 5$ targets locations, their RCS, the subset $\mathcal{S}$, and noise. It can be seen that the radar system exploiting SSFW and LASSO to recover the target ranges exhibits much better performance than a CSFW radar system using DFT processing [184]. For 1.5 GHz of bandwidth, the proposed technique achieves a 73% improvement in support recovery error.

7.4.3 Targer localization

Fig. 7.3 compares the localization performance of the SSFW and the CSFW. We consider two targets at locations (8.1 m, $\pi/6$) and (8.2 m, $\pi/4$) in the polar coordinate system, as shown in Fig. 7.3. Further, we consider the bandwidth of CSFW to be seven times that

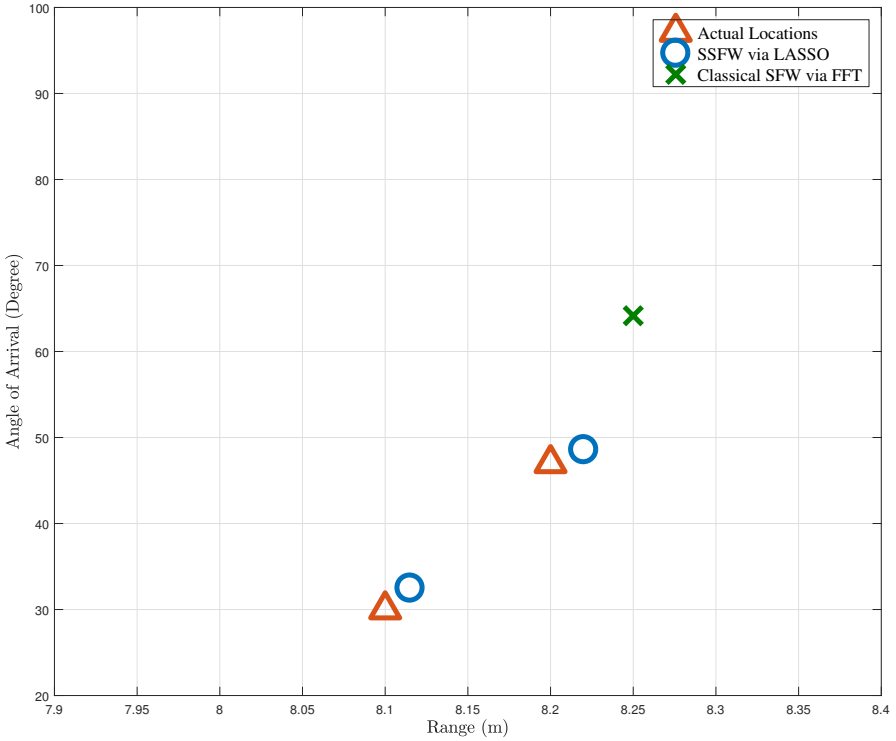


FIGURE 7.3: Localization of two targets.

of SSFW. It can be seen that the SSFW localizes the two targets properly, while using much lower bandwidth than the CSFW. On the other hand, CSFW estimates the range of one of the targets erroneously and fails to detect the second target, resulting in very poor localization.

7.5 conclusion

The chapter considered the problem of targer localization using asynchronous MIMO radars with applications to automotive scenario. Arguing the need for high range resolution for accurate localization with affordable bandwidth in such scenarios, the chapter considered a new framework based on SSFW and the sparse sensing paradigm. A two stage decentralized procedure for localization is proposed within this framework. The proposed scheme is shown to localize the targets better than the conventional schemes for given bandwidths and offers avenues for further waveform optimization. Enhanced performance and the limited resources consumed, makes the proposed scheme attractive to the industry.

Chapter 8

Conclusions and Future Work

8.1 Summary and conclusions

In this chapter, the main conclusions of the thesis are summarized, and also the possible research directions for the future works are identified and discussed. Broadly speaking, the problems of Direction of Arrival (DoA) estimation and target localization by exploiting sparse and one-bit samples has been investigated in this thesis. The works in this thesis go beyond the state of the art in these areas by proposing novel algorithms. Moreover, the performance of the proposed algorithms has been analyzed and then assessed through comparing them with the Cramér-Rao Bound (CRB) as well as with the state-of-the-art and conventional algorithms.

The first part of the thesis focuses the problem of DoA estimation from Sparse Linear Array (SLA) measurements. Chapter 3 proposed a novel estimator for DoA estimation from SLA measurements by deploying the Wieghted Least Squares technique. The performance of the proposed estimator is analytically calculated and it was shown that it provides consistent estimates of DoAs of identifiable sources for any SLAs. Further, an asymptotic closed-form expression for the resulting covariance matrix of DoA estimation errors was derived and it was analytically proved that it asymptotically coincides with the CRB in case the optimal weighting matrix is selected. This implies that the proposed WLS estimator is asymptotically statistically efficient. It thus closes an important gap in the co-array-based DoA estimation. Simulation results demonstrated superior performance of the proposed WLS estimator compared to the existing algorithms in the literature in terms of estimation accuracy and resolution. The problem of DoA estimation from one-bit SLA measurements was studied in Chapter 4. It was showed that the idetifiability condition for the DoA estimation problem from one-bit SLA data is equivalent to that for the case when DoAs are estimated from infinite-bit unquantized

measurements. Then, a pessimistic approximation of the corresponding CRB was derived. This pessimistic CRB was used as a benchmark for assessing the performance of one-bit DoA estimators. Further, it provides us with valuable insights on the performance limits of DoA estimation from one-bit quantized data. For example, it was shown that the DoA estimation errors in one-bit scenario reduces at the same rate as that of infinite-bit case with respect to the number of samples and, moreover, that the DoA estimation errors in one-bit scenario converges to a constant value by increasing the SNR. A new algorithm for estimating DoAs from one-bit quantized data was also proposed. and its analytical performance of the proposed method was investigated through deriving a closed-form expression for its asymptotic MSE and show that it outperforms the existing algorithms in the literature. Numerical simulations were provided to validate the analytical derivations and corroborate the improvement in estimation performance. In Chapter 5, the problem of DoA estimation from low-resolution multi-bit SLA measurements, say 2 or 4 bits per sample, was investigated. We proposed a novel optimization-based framework for estimating the covariance matrix of unquantized data. Then, Co-Array-Based MUSIC (CAB-MUSIC) was used for estimating DoAs of interest. The simulation results showed that increasing the sampling resolution to 2 or 4 bits per samples could significantly increase the DoA estimation performance compared to the one-bit sampling case while the power consumption and implementation costs is still much lower beside the high-resolution sampling scenario.

The second part of the thesis focuses on target localization problem using sparse and one-bit measurements. It was shown in Chapter 6 that the one-bit sampling offers an attractive solution to the challenges posed by the NB-IoT for location-based services. The one-bit samplers are integral to developing low cost and low power devices. We proposed a one-bit passive sensor array formulation to estimate the time-of-arrival in an NB-IoT network. The quantized samples of the estimates are then forwarded to an FC. We propose a novel method that casts the localization problem from aggregated quantized nodal estimates as a multivariate fractional optimization problem that we solve using the optimal Lasserre's SDP relaxation. We also propose the ANTARES algorithm as an alternative sub-optimal method with reduced computational complexity compared to Lasserre's. Our approach is helpful in addressing the problem of maintaining high localization accuracy while deploying reduced-rate ADCs at the nodes as well as limited-capacity NB-IoT links. Chapter 7 considered the problem of targer localization using asynchronous MIMO radars with applications to automotive scenario. Arguing the need for high range resolution for accurate localization with affordable bandwidth in such scenarios, the chapter considered a new framework based on SSFW and the sparse sensing paradigm. A two stage decentralized procedure for localization is proposed within this framework. The proposed scheme is shown to localize the targets better than

the conventional schemes for given bandwidths and offers avenues for further waveform optimization. Enhanced performance and the limited resources consumed, makes the proposed scheme attractive to the industry.

8.2 Future directions

The work carried out in this thesis can be extended in several directions. The main issues left for future work are discussed below:

- **DoA Estimation from One-Bit SLAs Measurements with Varying Thresholds:** Prior studies on one-bit DoA estimation presuppose that all one-bit ADCs use an identical threshold, equal to zero, during the sensing time as well as across the array elements to quantize the received analog signal. This is a convenient choice but could be far away from the optimum. Hence, as a possible future work, one can consider the problem of DoA estimation from one SLA data under conditions that one-bit ADC thresholds vary during the sensing time as well as across the array elements. This way of selecting one-bit ADC thresholds gives a boost to the DoA estimation performance. The varying thresholds could be selected in either a random or a systematic manner. In the latter case, the varying thresholds can be selected for example by making use of $\Sigma\Delta$ sampling architecture in spatial and/or time domain. This approach allows for alleviating the effect of quantization noise to some extent.

  We should note that the arcsine law will not be applicable anymore when the one-bit ADC thresholds are not equal to zero. In consequence, we are not able to use arcesine-law-based algorithms under such scenarios. The optimization framework proposed in Chapter 5 after some modification could be a suitable approach for DoA estimation in such cases.

- **Sub-Bit Samplin:** In the context of low-resolution sampling, it is also possible to go even below one bit per sample. Let call this sampling strategy sub-bit sampling. Sub-bit sampling can be done, e.g., by linearly mixing of N samples and use K one-bit ADCs for sampling with $K < N$ (non-linear mixing can also be considered). I believe sub-bit sampling could contribute significantly to developing very cheap ubiquitous signal processing devices.

- **Colored Noise:** Throughout Part I, noise is assumed to be white across the array elements. However, in practice, this assumption may not be necessarily true. It would be of great interest to investigate the performance bound of DoA

estimation from both infinite-bit and one-bit SLA data as well as to devise proper DoA estimators for the case when the white noise assumption is relaxed.

- **One-Bit Distributed Detection:** Chapter 6 investigates distributed target localization using one-bit quntized measurements in passive radars. The distributed detection problem using one-bit quntized measurements in passive radars still remains to be investigated.

Part III

Appendices

Appendix A

Appendices of Chapter 3

A.1 Proof of Lemma 3.1

It is well-known that the sample covariance $\hat{\mathbf{R}}$ is a consistent estimate of $\mathbf{R}$ under the current assumption [185], implying that $\lim_{N\to\infty} \hat{\mathbf{r}} = \mathbf{r}$. As a consequence, considering (3.7), we obtain

$$\lim_{N\to\infty} \hat{\mathbf{R}}_v = \begin{bmatrix} \mathbf{T}_v\mathbf{J}^\dagger\mathbf{r} & \mathbf{T}_{v-1}\mathbf{J}^\dagger\mathbf{r} & \cdots & \mathbf{T}_1\mathbf{J}^\dagger\mathbf{r} \end{bmatrix}. \tag{A.1}$$

On the other hand, it has been proved in [59] that

$$\lim_{N\to\infty} \hat{\mathbf{R}}_v = \mathbf{A}_v(\boldsymbol{\theta})\mathrm{diag}(\mathbf{p})\mathbf{A}_v^H(\boldsymbol{\theta}) + \sigma^2\mathbf{I}_v \doteq \mathbf{R}_v, \tag{A.2}$$

where the matrix $\mathbf{R}_v \in \mathbb{C}^{v\times v}$ has the same structure as the covariance matrix of signals received by a contiguous ULA whose elements are located at $\left(0, \frac{\lambda}{2}, \lambda, \cdots, (v-1)\frac{\lambda}{2}\right)$ and $\mathbf{A}_v(\boldsymbol{\theta}) \in \mathbb{C}^{v\times K}$ is its corresponding steering matrix. Therefore, in case $K \leq v-1$, exploiting the eigendecomposition, it is possible for (A.2) to be expressed as $\mathbf{R}_v = \mathbf{U}_s\boldsymbol{\Lambda}_s\mathbf{U}_s^H + \sigma^2\mathbf{U}_n\mathbf{U}_n^H$, where $\mathbf{U}_s$ and $\mathbf{U}_n$ represent the eigenvectors of $\mathbf{R}_v$ corresponding to its K largest and $v-K$ smallest eigenvalues, respectively. From (A.2), $\hat{\mathbf{R}}_v$ can be deemed to be a perturbed version of $\mathbf{R}_v$. Therefore, we have

$$\lim_{N\to\infty} \hat{\mathbf{U}}_n\hat{\mathbf{U}}_n^H = \mathbf{U}_n\mathbf{U}_n^H. \tag{A.3}$$

Eventually, making use of (A.1), (A.2), (A.3) and the fact that $\mathbf{U}_n^H \mathbf{A}_v(\boldsymbol{\theta}) = 0$, we obtain

$$\begin{aligned}\lim_{N\to\infty} \hat{\sigma}^2 &= \frac{\mathrm{vec}^H(\mathbf{U}_n\mathbf{U}_n^H)\mathbf{T}\mathbf{J}^\dagger\mathbf{r}}{v-K} \\ &= \frac{\mathrm{vec}^H(\mathbf{U}_n\mathbf{U}_n^H)\mathrm{vec}\left(\begin{bmatrix}\mathbf{T}_v\mathbf{J}^\dagger\mathbf{r} & \mathbf{T}_{v-1}\mathbf{J}^\dagger\mathbf{r} & \cdots & \mathbf{T}_1\mathbf{J}^\dagger\mathbf{r}\end{bmatrix}\right)}{v-K} \\ &= \frac{\sigma^2\mathrm{tr}(\mathbf{U}_n^H\mathbf{U}_n)}{v-K} = \sigma^2. \end{aligned} \tag{A.4}$$

A.2 Proof of Theorem 3.1

Let define $L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}}) = \|\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}}\hat{\mathbf{Q}}\hat{\mathbf{r}}\|_2^2$, $\tilde{L}(\boldsymbol{\theta}) = \lim_{N\to\infty} L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$ and use $\boldsymbol{\theta}_0$ to distinguish the actual DoA vector from a generic vector $\boldsymbol{\theta}$. Since the derivatives of $\mathbf{A}_d(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ are bounded, $L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$ converges uniformly to $\tilde{L}(\boldsymbol{\theta})$ as $N \to \infty$ [102]. Thus, $\hat{\boldsymbol{\theta}}_{wls}$ also converges to the minimizing argument of $\tilde{L}(\boldsymbol{\theta})$ as $N \to \infty$.

It readily follows from (A.3) that $\lim_{N\to\infty} \hat{\mathbf{Q}} = \mathbf{Q}$. In addition, from (A.4) and (3.4)

$$\mathbf{Q}\mathbf{r} = \mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}_0)\mathbf{p} \tag{A.5}$$

is readily checked. Hence, considering the fact that $\lim_{N\to\infty} \hat{\mathbf{r}} = \mathbf{r}$ and making use of continuous differentiability of $L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$, we obtain

$$\tilde{L}(\boldsymbol{\theta}) = L(\boldsymbol{\theta}, \lim_{N\to\infty}\hat{\mathbf{r}}, \lim_{N\to\infty}\hat{\mathbf{Q}}) = \|\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}_0)\mathbf{p}\|_2^2. \tag{A.6}$$

It is evident from (A.6) that $\tilde{L}(\boldsymbol{\theta}_0) = 0$ and $\tilde{L}(\boldsymbol{\theta}) \geq 0$, implying $\tilde{L}(\boldsymbol{\theta})$ has a global minima at $\boldsymbol{\theta}_0$. Consequently, consistency of $\hat{\boldsymbol{\theta}}_{wls}$ follows if $\boldsymbol{\theta}_0$ is the unique solution to $\tilde{L}(\boldsymbol{\theta}) = 0$ with respect to $\boldsymbol{\theta}$. According to (A.6), $\tilde{L}(\boldsymbol{\theta})$ is equal to zero if and only if

$$\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}_0)\mathbf{p} = 0. \tag{A.7}$$

In what follows, we employ the method of proof by contradiction to complete the proof. Let assume that $\boldsymbol{\theta}_1$ is a solution to (A.7) but $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_0$. This means that $\boldsymbol{\theta}_1$ could differ from $\boldsymbol{\theta}_0$ at q DoAs where $1 \leq q \leq K$. Substituting $\boldsymbol{\theta}_1$ into (A.7) yields

$$\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}_1)}\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}_0')\mathbf{p}' = 0. \tag{A.8}$$

where $\boldsymbol{\theta}_0' \in \mathbb{C}^{q\times 1}$ consists of those elements of $\boldsymbol{\theta}_0$ which are not shared with $\boldsymbol{\theta}_1$, and $\mathbf{A}_d(\boldsymbol{\theta}_0') \in \mathbb{C}^{q\times 1}$ and $\mathbf{p}' \in \mathbb{C}^{q\times 1}$ are the corresponding blocks of $\mathbf{A}_d(\boldsymbol{\theta}_0)$ and $\mathbf{p}$. The

expression on the left side of the above equation can be rewritten as follows

$$\mathbf{W}^{\frac{1}{2}}\mathbf{J}\underbrace{\begin{bmatrix}\mathbf{A}_d(\boldsymbol{\theta}_0^{'}) & \mathbf{A}_d(\boldsymbol{\theta}_1)\end{bmatrix}\begin{bmatrix}\mathbf{I}_K \\ -\left(\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}_1)\right)^{\dagger}\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}_0^{'})\end{bmatrix}\mathbf{p}^{'}}_{\aleph}. \tag{A.9}$$

Since the weighting matrix $\mathbf{W}$ is positive definite by definition and $\mathbf{J}$ a is full column rank matrix[52], it is concluded that (A.7) is zero if and only if the term $\aleph$ in (A.9) is zero.

Lemma A.1. $\begin{bmatrix}\mathbf{A}_d(\boldsymbol{\theta}_0^{'}) & \mathbf{A}_d(\boldsymbol{\theta}_1)\end{bmatrix}$ *has full column rank if* $K \leq v-1$.

Proof. Let $\mathbf{A}_\vartheta(\boldsymbol{\theta}) \in \mathbb{C}^{(2v-1)\times K}$ denote the steering matrix corresponding to the contiguous ULA segment of the difference co-array. Since $\begin{bmatrix}\mathbf{A}_\vartheta(\boldsymbol{\theta}_0^{'}) & \mathbf{A}_\vartheta(\boldsymbol{\theta}_1)\end{bmatrix}$ is a sub-matrix of $\begin{bmatrix}\mathbf{A}_d(\boldsymbol{\theta}_0^{'}) & \mathbf{A}_d(\boldsymbol{\theta}_1)\end{bmatrix}$, it is sufficient to show that $\begin{bmatrix}\mathbf{A}_\vartheta(\boldsymbol{\theta}_0^{'}) & \mathbf{A}_\vartheta(\boldsymbol{\theta}_1)\end{bmatrix}$ is full column rank instead of $\begin{bmatrix}\mathbf{A}_d(\boldsymbol{\theta}_0^{'}) & \mathbf{A}_d(\boldsymbol{\theta}_1)\end{bmatrix}$.

It is possible to decompose $\begin{bmatrix}\mathbf{A}_\vartheta(\boldsymbol{\theta}_0^{'}) & \mathbf{A}_\vartheta(\boldsymbol{\theta}_1)\end{bmatrix}$ as follows

$$\begin{bmatrix} 1 & \cdots & 1 & 1 & \cdots & 1 \\ \alpha_1 & \cdots & \alpha_q & \beta_1 & \cdots & \beta_K \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \alpha_1^{2(v-1)} & \cdots & \alpha_q^{2(v-1)} & \beta_1^{2(v-1)} & \cdots & \beta_K^{2(v-1)} \end{bmatrix} \times \operatorname{diag}(\begin{bmatrix}\alpha_1^{1-v} & \cdots & \alpha_q^{1-v} & \beta_1^{1-v} & \cdots & \beta_K^{1-v}\end{bmatrix}). \tag{A.10}$$

where $\alpha_i = e^{j\pi\sin[\boldsymbol{\theta}_0^{'}]_i}$ and $\beta_i = e^{j\pi\sin[\boldsymbol{\theta}_1]_i}$. The second matrix in (A.10) is a $(K+q)\times(K+q)$ diagonal matrix and thus full rank. However, the first one is a $(2v-1)\times(K+q)$ Vandermonde matrix which has full column rank for distinct DoAs iff $K+q \leq 2v-1$. This condition is fulfilled for all admissible q iff $K \leq v-1$. Hence, it follows that $\begin{bmatrix}\mathbf{A}_\vartheta(\boldsymbol{\theta}_0^{'}) & \mathbf{A}_\vartheta(\boldsymbol{\theta}_1)\end{bmatrix}$ and, in turn, $\begin{bmatrix}\mathbf{A}_d(\boldsymbol{\theta}_0^{'}) & \mathbf{A}_d(\boldsymbol{\theta}_1)\end{bmatrix}$ have full column rank if $K \leq v-1$. □

According to Lemma A.1, on condition that $K \leq v-1$, the term $\aleph$ in (A.9) is equal to zero if and only if

$$\begin{bmatrix}\mathbf{I}_K \\ -\left(\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\tilde{\boldsymbol{\theta}})\right)^{\dagger}\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}_0)\end{bmatrix}\mathbf{p}^{'} = 0, \tag{A.11}$$

implying that $\mathbf{p}^{'} = 0$, which is in contradiction with the definition of $\mathbf{p}$ given in **A2**. Therefore, it is concluded that $\boldsymbol{\theta}_0$ has to be the unique solution to $\tilde{L}(\boldsymbol{\theta}) = 0$ with respect to $\boldsymbol{\theta}$ if $K \leq v - 1$, which indicates the consistency of $\hat{\boldsymbol{\theta}}_{wls}$. ■

A.3 Proof of Theorem 3.2

The proof is composed of four steps. First, a closed-form error expression for the DoA estimates is given through a Taylor series expansion method. The given closed-form expression involves the gradient and Hessian of $L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$. Hence, the corresponding gradient and Hessian is computed at the second and third steps, respectively. Finally, the covariance matrix of DoA estimation errors is obtained by combining the net results of preceding steps.

A.3.1 Closed-form expression for DoA estimation errors

From (3.15), we know that $\hat{\boldsymbol{\theta}}_{wls}$ is a critical point of $L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$, thus we have $\nabla_{\boldsymbol{\theta}} L(\hat{\boldsymbol{\theta}}_{wls}, \hat{\mathbf{r}}, \hat{\mathbf{Q}}) = \mathbf{0}$ where $\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$ denotes the gradient of $L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$ with respect to $\boldsymbol{\theta}$. $\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$ is a real-valued[1] function on $\mathbb{R}^K$, thereby applying Taylor theorem [186, Ch. 6, Theorem 12] around the true value of $\boldsymbol{\theta}$ yields

$$\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}}) + \nabla^2_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})(\hat{\boldsymbol{\theta}}_{wls} - \boldsymbol{\theta}) + \left(\mathbf{I}_K \otimes (\hat{\boldsymbol{\theta}}_{wls} - \boldsymbol{\theta})^T\right) \mathbf{H}(\boldsymbol{\theta}^*, \hat{\mathbf{r}}, \hat{\mathbf{Q}})(\hat{\boldsymbol{\theta}}_{wls} - \boldsymbol{\theta}) = \mathbf{0}, \tag{A.12}$$

where $\nabla^2_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$ denotes the Hessian matrix of $L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$ with respect to $\boldsymbol{\theta}$, the matrix $\mathbf{H}(\boldsymbol{\theta}^*, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$ is given in [186, Ch. 6, Definition 2], and $\boldsymbol{\theta}^* = \boldsymbol{\theta}(1-t) + t\hat{\boldsymbol{\theta}}_{wls}$ for some $t \in (0, 1)$. From (A.12), we have

$$\hat{\boldsymbol{\theta}}_{wls} - \boldsymbol{\theta} = -\left(\nabla^2_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}}) + \overbrace{\left(\mathbf{I}_K \otimes (\hat{\boldsymbol{\theta}}_{wls} - \boldsymbol{\theta})^T\right) \mathbf{H}(\boldsymbol{\theta}^*, \hat{\mathbf{r}}, \hat{\mathbf{Q}})}^{\hbar}\right)^{-1} \nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}}). \tag{A.13}$$

Noting that $\hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H = \mathbf{U}_n \mathbf{U}_n^H + \mathcal{O}(\frac{1}{\sqrt{N}})$ [71], it is readily concluded that $\hat{\mathbf{Q}} = \mathbf{Q} + \mathcal{O}(\frac{1}{\sqrt{N}})$ for large N. Consequently, considering the fact that $\hat{\mathbf{r}} = \mathbf{r} + \mathcal{O}(\frac{1}{\sqrt{N}})$ for large N [71, 107] and making use of continuous differentiability of $L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$, it can readily be shown that $\nabla^2_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}}) = \nabla^2_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \mathbf{r}, \mathbf{Q}) + \mathcal{O}(\frac{1}{\sqrt{N}})$ and $\mathbf{H}(\boldsymbol{\theta}^*, \hat{\mathbf{r}}, \hat{\mathbf{Q}}) = \mathbf{H}(\boldsymbol{\theta}^*, \mathbf{r}, \mathbf{Q}) + \mathcal{O}(\frac{1}{\sqrt{N}})$ for large N. On the other hand, since $\hat{\boldsymbol{\theta}}_{wls}$ is a consistent estimate of $\boldsymbol{\theta}$ according to Theorem 3.1, there exists $a > 0$ such that $\hat{\boldsymbol{\theta}}_{wls} - \boldsymbol{\theta} = \mathcal{O}(\frac{1}{N^a})$ for large N. Hence, it follows that

[1] The fact that $\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$ is a real-valued function will be shown later in (A.16)

the Hessian term in (A.13) converges to a constant value as $N \to \infty$ while the term $\hbar$ converges to zero as $N \to \infty$. Therefor, it is possible to neglect the term $\hbar$ compared to the Hessian term in (A.13) in the asymptotic regime ($N \to \infty$), leading to

$$\hat{\boldsymbol{\theta}}_{wls} - \boldsymbol{\theta} \simeq -\left(\nabla^2_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \mathbf{r}, \mathbf{Q})\right)^{-1} \nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}}). \tag{A.14}$$

A.3.2 Derivation of the Gradient Vector

Taking derivative of $L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$ with respect to $\boldsymbol{\theta}$ and exploiting the following expression for the derivative of projection matrix $\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}$ [187]

$$\begin{aligned}\frac{\partial \Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}}{\partial \theta_k} =& -(\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}))^{\dagger H}(\mathbf{J}\frac{\partial \mathbf{A}_d(\boldsymbol{\theta})}{\partial \theta_k})^H \mathbf{W}^{\frac{1}{2}} \Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})} \\ & - \Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})} \mathbf{W}^{\frac{1}{2}}\mathbf{J}\frac{\partial \mathbf{A}_d(\boldsymbol{\theta})}{\partial \theta_k}(\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}))^{\dagger},\end{aligned} \tag{A.15}$$

we get

$$\begin{aligned}\frac{\partial L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})}{\partial \theta_k} =& -2\Re\{\hat{\mathbf{r}}^H \hat{\mathbf{Q}}^H \mathbf{W}^{\frac{1}{2}}(\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}))^{\dagger H} \\ & \times (\mathbf{J}\frac{\partial \mathbf{A}_d(\boldsymbol{\theta})}{\partial \theta_k})^H \mathbf{W}^{\frac{1}{2}} \Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})} \mathbf{W}^{\frac{1}{2}} \hat{\mathbf{Q}}\hat{\mathbf{r}}\}.\end{aligned} \tag{A.16}$$

Making use of Lemma A.6 in Appendix A.7, it can be readily shown that the term inside the $\Re\{.\}$ operator is real-valued in case $\mathbf{K}_M\mathbf{W} = \mathbf{W}^*\mathbf{K}_M$. Hence, considering (3.14), (A.16) can be written as follows

$$\begin{aligned}\frac{\partial L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})}{\partial \theta_k} &= -2\hat{\mathbf{p}}_{ls}^H (\mathbf{J}\frac{\partial \mathbf{A}_d(\boldsymbol{\theta})}{\partial \theta_k})^H \mathbf{W}^{\frac{1}{2}} \Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})} \mathbf{W}^{\frac{1}{2}} \hat{\mathbf{Q}}\hat{\mathbf{r}} \\ &= -j2\pi [\hat{\mathbf{p}}_{ls}^H]_k \cos\theta_k \mathbf{a}_d^H(\theta_k) \mathrm{diag}(\mathbf{d}) \mathbf{J}^H \mathbf{W}^{\frac{1}{2}} \Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})} \mathbf{W}^{\frac{1}{2}} \hat{\mathbf{Q}}\hat{\mathbf{r}},\end{aligned} \tag{A.17}$$

where

$$\mathbf{a}_d(\theta_k) = \begin{bmatrix} e^{-j\pi \sin\theta_k \ell_{D-1}} & \cdots & 1 & \cdots & e^{j\pi \sin\theta_k \ell_{D-1}} \end{bmatrix}^T. \tag{A.18}$$

From (A.17) and using $\lim_{N\to\infty} \hat{\mathbf{p}}_{ls} = \mathbf{p}$ and $\lim_{N\to\infty} \hat{\mathbf{Q}} = \mathbf{Q}$, the gradient of $L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$ with respect to $\boldsymbol{\theta}$ is given by

$$\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}}) \simeq -j2\pi \mathrm{diag}(\mathbf{p}) \boldsymbol{\Phi}(\boldsymbol{\theta}) \mathbf{A}_d^H(\boldsymbol{\theta}) \mathrm{diag}(\mathbf{d}) \mathbf{J}^H \mathbf{W}^{\frac{1}{2}} \Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})} \mathbf{W}^{\frac{1}{2}} \mathbf{Q}\hat{\mathbf{r}}. \tag{A.19}$$

where $\boldsymbol{\Phi}(\boldsymbol{\theta}) = \mathrm{diag}(\begin{bmatrix} \cos\theta_1 & \cdots & \cos\theta_K \end{bmatrix}^T)$.

A.3.3 Derivation of the Hessian Matrix

Taking the derivative of (A.16) with respect to θ_l and making use of (A.15) and the following expression for the derivative of $(\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}))^{\dagger H}$

$$\begin{aligned}\frac{\partial(\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}))^{\dagger H}}{\partial\theta_l} =&\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}}\mathbf{J}\frac{\partial\mathbf{A}_d(\boldsymbol{\theta})}{\partial\theta_l}(\mathbf{A}_d^H(\boldsymbol{\theta})\mathbf{J}^H\mathbf{W}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}))^{-1}\\ &-(\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}))^{\dagger H}(\mathbf{J}\frac{\partial\mathbf{A}_d(\boldsymbol{\theta})}{\partial\theta_l})^H\mathbf{W}^{\frac{1}{2}}(\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}))^{\dagger H},\end{aligned} \tag{A.20}$$

leads to

$$\begin{aligned}\frac{\partial^2 L(\boldsymbol{\theta},\hat{\mathbf{r}},\hat{\mathbf{Q}})}{\partial\theta_l\partial\theta_k} =&-2\hat{\mathbf{r}}^H\hat{\mathbf{Q}}^H\mathbf{W}^{\frac{1}{2}}\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}}\mathbf{J}\frac{\partial\mathbf{A}_d(\boldsymbol{\theta})}{\partial\theta_l}(\mathbf{A}_d^H(\boldsymbol{\theta})\mathbf{J}^H\mathbf{W}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}))^{-1}\\ &\times(\mathbf{J}\frac{\partial\mathbf{A}_d(\boldsymbol{\theta})}{\partial\theta_k})^H\mathbf{W}^{\frac{1}{2}}\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}}\hat{\mathbf{Q}}\hat{\mathbf{r}}\\ &+2\hat{\mathbf{r}}^H\hat{\mathbf{Q}}^H\mathbf{W}^{\frac{1}{2}}(\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}))^{\dagger H}(\mathbf{J}\frac{\partial\mathbf{A}_d(\boldsymbol{\theta})}{\partial\theta_l})^H\mathbf{W}^{\frac{1}{2}}(\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}))^{\dagger H}\\ &\times(\mathbf{J}\frac{\partial\mathbf{A}_d(\boldsymbol{\theta})}{\partial\theta_k})^H\mathbf{W}^{\frac{1}{2}}\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}}\hat{\mathbf{Q}}\hat{\mathbf{r}}\\ &-2\hat{\mathbf{r}}^H\hat{\mathbf{Q}}^H\mathbf{W}^{\frac{1}{2}}(\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}))^{\dagger H}(\mathbf{J}\frac{\partial^2\mathbf{A}_d(\boldsymbol{\theta})}{\partial\theta_l\partial\theta_k})^H\mathbf{W}^{\frac{1}{2}}\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}}\hat{\mathbf{Q}}\hat{\mathbf{r}}\\ &+2\hat{\mathbf{r}}^H\hat{\mathbf{Q}}^H\mathbf{W}^{\frac{1}{2}}(\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}))^{\dagger H}(\mathbf{J}\frac{\partial\mathbf{A}_d(\boldsymbol{\theta})}{\partial\theta_k})^H\mathbf{W}^{\frac{1}{2}}(\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}))^{\dagger H}\\ &\times(\mathbf{J}\frac{\partial\mathbf{A}_d(\boldsymbol{\theta})}{\partial\theta_l})^H\mathbf{W}^{\frac{1}{2}}\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}}\hat{\mathbf{Q}}\hat{\mathbf{r}}\\ &+2\hat{\mathbf{r}}^H\hat{\mathbf{Q}}^H\mathbf{W}^{\frac{1}{2}}(\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}))^{\dagger H}(\mathbf{J}\frac{\partial\mathbf{A}_d(\boldsymbol{\theta})}{\partial\theta_k})^H\mathbf{W}^{\frac{1}{2}}\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\\ &\times\mathbf{W}^{\frac{1}{2}}\mathbf{J}\frac{\partial\mathbf{A}_d(\boldsymbol{\theta})}{\partial\theta_l}(\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}))^{\dagger}\hat{\mathbf{Q}}\hat{\mathbf{r}}.\end{aligned} \tag{A.21}$$

Given the fact that $\tilde{L}(\boldsymbol{\theta}_0)=0$, it is possible to neglect the first four terms on the right had side of (A.21) compared to the last term as $N\to\infty$. Thus, by replacing $(\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}))^{\dagger}\mathbf{W}^{\frac{1}{2}}\hat{\mathbf{Q}}\hat{\mathbf{r}}$ from (3.14), we have

$$\begin{aligned}\frac{\partial^2 L(\boldsymbol{\theta},\hat{\mathbf{r}},\hat{\mathbf{Q}})}{\partial\theta_l\partial\theta_k} \simeq&-2\pi^2[\hat{\mathbf{p}}_{ls}^H]_k\cos\theta_k\mathbf{a}_d^H(\theta_k)\mathrm{diag}(\mathbf{d})\mathbf{J}^H\mathbf{W}^{\frac{1}{2}}\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}}\\ &\times\mathbf{J}\mathrm{diag}(\mathbf{d})\mathbf{a}_d(\theta_l)\cos\theta_l[\hat{\mathbf{p}}_{ls}]_l.\end{aligned} \tag{A.22}$$

Based on (A.22) and considering $\lim_{N\to\infty}\hat{\mathbf{p}}_{ls}=\mathbf{p}$ and $\lim_{N\to\infty}\hat{\mathbf{Q}}=\mathbf{Q}$, the Hessian matrix of $L(\boldsymbol{\theta},\hat{\mathbf{r}},\hat{\mathbf{Q}})$ with respect to $\boldsymbol{\theta}$ is obtained as follows

$$\begin{aligned}\lim_{N\to\infty}\nabla^2_{\boldsymbol{\theta}}L(\boldsymbol{\theta},\hat{\mathbf{r}},\hat{\mathbf{Q}})=&\nabla^2_{\boldsymbol{\theta}}L(\boldsymbol{\theta},\mathbf{r},\mathbf{Q})\\ \simeq&-2\pi^2\mathrm{diag}(\mathbf{p})\boldsymbol{\Phi}(\boldsymbol{\theta})\mathbf{A}_d^H(\boldsymbol{\theta})\mathrm{diag}(\mathbf{d})\mathbf{J}^H\mathbf{W}^{\frac{1}{2}}\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}}\\ &\times\mathbf{J}\mathrm{diag}(\mathbf{d})\mathbf{A}_d(\boldsymbol{\theta})\boldsymbol{\Phi}(\boldsymbol{\theta})\mathrm{diag}(\mathbf{p}).\end{aligned}\tag{A.23}$$

A.3.4 Calculation of the error covariance matrix

Combining (A.14), (A.19) and (A.23), it is possible to derive an asymptotic ($N\to\infty$) expression for the covariance matrix of $\widehat{\boldsymbol{\theta}}_{wls}$ as follows

$$\begin{aligned}&\mathbb{E}\left\{(\hat{\boldsymbol{\theta}}_{wls}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_{wls}-\boldsymbol{\theta})^H\right\}\\ \simeq&\mathbb{E}\left\{(\nabla^2_{\boldsymbol{\theta}}L(\boldsymbol{\theta}_0,\hat{\mathbf{r}},\hat{\mathbf{Q}}))^{-1}\nabla_{\boldsymbol{\theta}}L(\boldsymbol{\theta}_0,\hat{\mathbf{r}},\hat{\mathbf{Q}})(\nabla_{\boldsymbol{\theta}}L(\boldsymbol{\theta}_0,\hat{\mathbf{r}},\hat{\mathbf{Q}}))^H(\nabla^2_{\boldsymbol{\theta}}L(\boldsymbol{\theta}_0,\hat{\mathbf{r}},\hat{\mathbf{Q}}))^{-1}\right\}\\ =&\frac{1}{\pi^2}(\mathrm{diag}(\mathbf{p})\boldsymbol{\Phi}^H(\boldsymbol{\theta})\mathbf{A}_d^H(\boldsymbol{\theta})\mathrm{diag}(\mathbf{d})\mathbf{J}^H\mathbf{W}^{\frac{1}{2}}\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathrm{diag}(\mathbf{d})\mathbf{A}_d(\boldsymbol{\theta})\boldsymbol{\Phi}(\boldsymbol{\theta})\mathrm{diag}(\mathbf{p}))^{-1}\\ &\times\mathrm{diag}(\mathbf{p})\boldsymbol{\Phi}^H(\boldsymbol{\theta})\mathbf{A}_d^H(\boldsymbol{\theta})\mathrm{diag}(\mathbf{d})\mathbf{J}^H\mathbf{W}^{\frac{1}{2}}\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}}\mathbf{Q}\mathbb{E}\{\hat{\mathbf{r}}\hat{\mathbf{r}}^H\}\mathbf{Q}^H\mathbf{W}^{\frac{1}{2}}\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\\ &\times\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathrm{diag}(\mathbf{d})\mathbf{A}_d(\boldsymbol{\theta})\boldsymbol{\Phi}(\boldsymbol{\theta})\mathrm{diag}(\mathbf{p})\\ &\times(\mathrm{diag}(\mathbf{p})\boldsymbol{\Phi}^H(\boldsymbol{\theta})\mathbf{A}_d^H(\boldsymbol{\theta})\mathrm{diag}(\mathbf{d})\mathbf{J}^H\mathbf{W}^{\frac{1}{2}}\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathrm{diag}(\mathbf{d})\mathbf{A}_d(\boldsymbol{\theta})\boldsymbol{\Phi}(\boldsymbol{\theta})\mathrm{diag}(\mathbf{p}))^{-1}.\end{aligned}\tag{A.24}$$

It is shown in [188] that

$$\mathbb{E}\{\hat{\mathbf{r}}\hat{\mathbf{r}}^H\}=\mathbf{r}\mathbf{r}^H+\frac{1}{N}\left(\mathbf{R}^T\otimes\mathbf{R}\right).\tag{A.25}$$

Inserting (A.25) into (A.24) and exploiting (A.5) gives (3.28).

A.4 Proof of Theorem 3.3

Before proceeding to the main proof, let us first introduce the following preliminary lemmas.

Lemma A.2. *The vector* $\mathbf{b}$ *belongs to the range space of the matrix* $\mathbf{F}$*, i.e.,* $\mathbf{b}\in\mathcal{R}(\mathbf{F})$.

Proof. Recall $\mathbf{A}_v(\boldsymbol{\theta})$ from Appendix A.1, which is defined as the steering matrix of a ULA whose elements are located at $\left(0,\frac{\lambda}{2},\lambda,\cdots,(v-1)\frac{\lambda}{2}\right)$. It is observe that there is a selection matrix $\mathbf{Z}\in\{0,1\}^{K^2\times K}$ such that

$$\mathbf{A}_v^*(\boldsymbol{\theta})\odot\mathbf{A}_v(\boldsymbol{\theta})=(\mathbf{A}_v^*(\boldsymbol{\theta})\otimes\mathbf{A}_v(\boldsymbol{\theta}))\mathbf{Z}.\tag{A.26}$$

Using (A.73), (A.26) and the fact that $\mathbf{A}_v^H(\boldsymbol{\theta})\mathbf{U}_n = 0$ leads to

$$\mathbf{b}^H\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}) = \mathrm{vec}^H(\mathbf{A}_v^H(\boldsymbol{\theta})\mathbf{U}_n\mathbf{U}_n^H\mathbf{A}_v(\boldsymbol{\theta}))\mathbf{Z} = 0. \tag{A.27}$$

(A.27) shows that $\mathbf{b}$ belongs to the null space of $\mathbf{A}_d^H(\boldsymbol{\theta})\mathbf{J}^H$. We know $\mathcal{N}(\mathbf{A}_d^H(\boldsymbol{\theta})\mathbf{J}^H) = \mathcal{R}(\mathbf{F})$. This concludes the proof. □

Lemma A.3. *The vector* $\mathbf{b}$ *belongs to the null space of* $\boldsymbol{\Omega}^H \doteq \boldsymbol{\Phi}^H(\boldsymbol{\theta})\mathbf{A}_d^H(\boldsymbol{\theta})\mathrm{diag}(\mathbf{d})\mathbf{J}^H$, *i.e.,* $\mathbf{b} \in \mathcal{N}(\boldsymbol{\Omega}^H)$.

Proof. Let $\mathbf{a}_v(\theta_k) \doteq \begin{bmatrix} 1 & e^{j\pi\sin\theta_k} & \cdots & e^{j\pi\sin\theta_k(v-1)} \end{bmatrix}^T$ denote the k^{th} column of $\mathbf{A}_v(\boldsymbol{\theta})$ and $\mathbf{v} \doteq \begin{bmatrix} 0 & 1 & \cdots & v-1 \end{bmatrix}^T$. Exploiting (A.18) and (A.73), we can show that

$$\begin{aligned}
\mathbf{T}\mathrm{diag}(\mathbf{d})\mathbf{A}_d(\boldsymbol{\theta})\boldsymbol{\Phi}(\boldsymbol{\theta}) &= \frac{-j}{\pi}\begin{bmatrix} \frac{\partial \mathbf{T}\mathbf{a}_d(\theta_1)}{\partial\theta_1} & \frac{\partial \mathbf{T}\mathbf{a}_d(\theta_2)}{\partial\theta_2} & \cdots & \frac{\partial \mathbf{T}\mathbf{a}_d(\theta_K)}{\partial\theta_K} \end{bmatrix} \\
&= \frac{-j}{\pi}\begin{bmatrix} \frac{\partial \mathbf{a}_v^*(\theta_1)\otimes\mathbf{a}_v(\theta_1)}{\partial\theta_1} & \frac{\partial \mathbf{a}_v^*(\theta_2)\otimes\mathbf{a}_v(\theta_2)}{\partial\theta_2} & \cdots & \frac{\partial \mathbf{a}_v^*(\theta_K)\otimes\mathbf{a}_v(\theta_K)}{\partial\theta_K} \end{bmatrix} \\
&= \frac{-j}{\pi}\left(\mathrm{diag}(\mathbf{v})\mathbf{A}_v^*(\boldsymbol{\theta}) \odot \mathbf{A}_v(\boldsymbol{\theta}) + \mathbf{A}_v^*(\boldsymbol{\theta}) \odot \mathrm{diag}(\mathbf{v})\mathbf{A}_v(\boldsymbol{\theta})\right)\boldsymbol{\Phi}(\boldsymbol{\theta}).
\end{aligned} \tag{A.28}$$

Exploiting (A.26), (A.28) and the fact that $\mathbf{A}_v^H(\boldsymbol{\theta})\mathbf{U}_n = 0$ gives

$$\begin{aligned}
\mathbf{b}^H\boldsymbol{\Omega} = &\frac{1}{j\pi}\big(\mathrm{vec}^H(\mathbf{A}_v^H(\boldsymbol{\theta})\mathbf{U}_n\mathbf{U}_n^H\mathrm{diag}(\mathbf{v})\mathbf{A}_v(\boldsymbol{\theta}))\mathbf{Z}\boldsymbol{\Phi}(\boldsymbol{\theta}) \\
&+ \mathrm{vec}^H(\mathbf{A}_v^H(\boldsymbol{\theta})\mathrm{diag}(\mathbf{v})\mathbf{U}_n\mathbf{U}_n^H\mathbf{A}_v(\boldsymbol{\theta}))\mathbf{Z}\boldsymbol{\Phi}(\boldsymbol{\theta})\big) = 0.
\end{aligned} \tag{A.29}$$

This completes the proof. □

Now, let define $\boldsymbol{\Psi} \doteq \begin{bmatrix} \mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}) & \mathrm{vec}(\mathbf{I}_M) \end{bmatrix}$, the CRB expression given in [52, Theorem 2] can then be rewritten as follows

$$CRB^{-1}(\boldsymbol{\theta}) = \frac{1}{\pi^2 N}\mathrm{diag}(\mathbf{p})\boldsymbol{\Omega}^H\mathbf{M}^{-1}\Pi^{\perp}_{\mathbf{M}^{-1}\boldsymbol{\Psi}}\mathbf{M}^{-1}\boldsymbol{\Omega}\mathrm{diag}(\mathbf{p}). \tag{A.30}$$

Based on the projection decomposition theorem [101], we have

$$\begin{aligned}
\Pi^{\perp}_{\mathbf{M}^{-1}\boldsymbol{\Psi}} &= \Pi^{\perp}_{\mathbf{M}^{-1}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})} - \Pi_{\Pi^{\perp}_{\mathbf{M}^{-1}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{M}^{-1}\mathrm{vec}(\mathbf{I}_M)} \\
&= \Pi_{\mathbf{MF}} - \Pi_{\Pi_{\mathbf{MF}}\mathbf{M}^{-1}\mathrm{vec}(\mathbf{I}_M)},
\end{aligned} \tag{A.31}$$

where the last equality is obtained by using the fact that $\Pi^{\perp}_{\mathbf{M}^{-1}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})} = \Pi_{\mathbf{MF}}$ [101]. Substituting (A.31) into (A.30) yields

$$CRB^{-1}(\boldsymbol{\theta}) = \frac{1}{\pi^2 N}\mathrm{diag}(\mathbf{p})\boldsymbol{\Omega}^H\mathbf{F}\Bigg[(\mathbf{F}^H\mathbf{M}^2\mathbf{F})^{-1} - \frac{(\mathbf{F}^H\mathbf{M}^2\mathbf{F})^{-1}\mathbf{F}^H\mathrm{vec}(\mathbf{I}_M)\mathrm{vec}^H(\mathbf{I}_M)\mathbf{F}(\mathbf{F}^H\mathbf{M}^2\mathbf{F})^{-1}}{\mathrm{vec}^H(\mathbf{I}_M)\mathbf{F}(\mathbf{F}^H\mathbf{M}^2\mathbf{F})^{-1}\mathbf{F}^H\mathrm{vec}(\mathbf{I}_M)}\Bigg]\mathbf{F}^H\boldsymbol{\Omega}\mathrm{diag}(\mathbf{p}). \tag{A.32}$$

Let decompose the vector $\mathrm{vec}(\mathbf{I}_M)$ as a sum of two vectors $\mathbf{h}_{\|}$ and $\mathbf{h}_{\perp}$ where

$$\mathbf{h}_{\|} \doteq \mathrm{vec}(\mathbf{I}_M) - \frac{(v-K)\mathbf{M}^2\mathbf{b}}{\mathbf{b}^H\mathbf{M}^2\mathbf{b}}, \tag{A.33}$$

$$\mathbf{h}_{\perp} \doteq \frac{(v-K)\mathbf{M}^2\mathbf{b}}{\mathbf{b}^H\mathbf{M}^2\mathbf{b}}. \tag{A.34}$$

It follows from (3.30), (A.47) and the definitions of $\mathbf{h}_{\|}$ and $\mathbf{h}_{\perp}$, given above, that

$$\begin{aligned}\mathbf{h}_{\perp}^H\mathbf{M}^{-2}\mathbf{h}_{\|} &= \frac{(v-K)\mathbf{b}^H\mathrm{vec}(\mathbf{I}_M)}{\mathbf{b}^H\mathbf{M}^2\mathbf{b}} - \frac{(v-K)^2\mathbf{b}^H\mathbf{M}^2\mathbf{b}}{(\mathbf{b}^H\mathbf{M}^2\mathbf{b})^2}\\ &= \frac{(v-K)\mathrm{vec}^H(\mathbf{U}_n\mathbf{U}_n^H)\mathbf{T}\mathbf{g}}{\mathbf{b}^H\mathbf{M}^2\mathbf{b}} - \frac{(v-K)^2}{\mathbf{b}^H\mathbf{M}^2\mathbf{b}}\\ &= \frac{(v-K)\mathrm{vec}^H(\mathbf{U}_n\mathbf{U}_n^H)\mathrm{vec}(\mathbf{I}_v)}{\mathbf{b}^H\mathbf{M}^2\mathbf{b}} - \frac{(v-K)^2}{\mathbf{b}^H\mathbf{M}^2\mathbf{b}} = 0.\end{aligned} \tag{A.35}$$

Accordingly and making use of the fact that $\Pi^{\perp}_{\mathbf{M}^{-1}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})} = \Pi_{\mathbf{MF}}$, we obtain

$$\begin{aligned}\mathrm{vec}^H(\mathbf{I}_M)\mathbf{F}(\mathbf{F}^H\mathbf{M}^2\mathbf{F})^{-1}\mathbf{F}^H\mathrm{vec}(\mathbf{I}_M) =& \mathrm{vec}^H(\mathbf{I}_M)\mathbf{M}^{-1}\Pi^{\perp}_{\mathbf{M}^{-1}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{M}^{-1}\mathrm{vec}(\mathbf{I}_M)\\ =& \mathrm{vec}^H(\mathbf{I}_M)\mathbf{M}^{-2}\mathrm{vec}(\mathbf{I}_M) - \mathbf{h}_{\|}^H\mathbf{M}^{-1}\Pi_{\mathbf{M}^{-1}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{M}^{-1}\mathbf{h}_{\|}\\ &- \mathbf{h}_{\perp}^H\mathbf{M}^{-1}(\mathbf{M}^{-1}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}))^{\dagger H}\mathbf{A}_d(\boldsymbol{\theta})^H\mathbf{J}^H\mathbf{M}^{-2}\mathbf{h}_{\perp}\\ &- \mathbf{h}_{\|}^H\mathbf{M}^{-1}(\mathbf{M}^{-1}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}))^{\dagger H}\mathbf{A}_d(\boldsymbol{\theta})^H\mathbf{J}^H\mathbf{M}^{-2}\mathbf{h}_{\perp}\\ &- \mathbf{h}_{\perp}^H\mathbf{M}^{-2}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})(\mathbf{M}^{-1}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}))^{\dagger}\mathbf{M}^{-1}\mathbf{h}_{\perp}.\end{aligned} \tag{A.36}$$

From the definition of $\mathbf{h}_{\perp}$ in (A.34) and Lemma A.2, we have

$$\mathbf{A}_d^H(\boldsymbol{\theta})\mathbf{J}^H\mathbf{M}^{-2}\mathbf{h}_{\perp} = \frac{(v-K)\mathbf{A}_d^H(\boldsymbol{\theta})\mathbf{J}^H\mathbf{b}}{\mathbf{b}^H\mathbf{M}^2\mathbf{b}} = \mathbf{0}. \tag{A.37}$$

Inserting (A.37) into (A.36) leads the last three terms in (A.36) to vanish. Then, exploiting $\Pi^{\perp}_{\mathbf{M}^{-1}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})} = \Pi_{\mathbf{MF}}$ once again yields

$$\begin{aligned}&\mathrm{vec}^H(\mathbf{I}_M)\mathbf{F}(\mathbf{F}^H\mathbf{M}^2\mathbf{F})^{-1}\mathbf{F}^H\mathrm{vec}(\mathbf{I}_M)\\ &= \mathrm{vec}^H(\mathbf{I}_M)\mathbf{M}^{-2}\mathrm{vec}(\mathbf{I}_M) - \mathbf{h}_{\|}^H\mathbf{M}^{-1}\Pi^{\perp}_{\mathbf{MF}}\mathbf{M}^{-1}\mathbf{h}_{\|}.\end{aligned} \tag{A.38}$$

Given $\text{vec}(\mathbf{I}_M) = \mathbf{h}_{\|} + \mathbf{h}_{\perp}$ and from (A.35), it is observed that

$$\text{vec}^H(\mathbf{I}_M)\mathbf{M}^{-2}\text{vec}(\mathbf{I}_M) = \mathbf{h}_{\|}^H\mathbf{M}^{-2}\mathbf{h}_{\|} + \mathbf{h}_{\perp}^H\mathbf{M}^{-2}\mathbf{h}_{\perp}. \tag{A.39}$$

Inserting (A.39) into (A.38) leads to

$$\begin{aligned}&\text{vec}^H(\mathbf{I}_M)\mathbf{F}(\mathbf{F}^H\mathbf{M}^2\mathbf{F})^{-1}\mathbf{F}^H\text{vec}(\mathbf{I}_M)\\ &= \frac{(v-K)^2}{\mathbf{b}^H\mathbf{M}^2\mathbf{b}} + \mathbf{h}_{\|}^H\mathbf{F}(\mathbf{F}^H\mathbf{M}^2\mathbf{F})^{-1}\mathbf{F}^H\mathbf{h}_{\|}.\end{aligned} \tag{A.40}$$

Further, making use of (A.37) and Lemma A.3, we can show that

$$\begin{aligned}&\boldsymbol{\Omega}^H\mathbf{F}(\mathbf{F}^H\mathbf{M}^2\mathbf{F})^{-1}\mathbf{F}^H\mathbf{h}_{\perp} = \boldsymbol{\Omega}^H\mathbf{M}^{-1}\Pi^{\perp}_{\mathbf{M}^{-1}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{M}^{-1}\mathbf{h}_{\perp}\\ &= \boldsymbol{\Omega}^H\mathbf{M}^{-2}\mathbf{h}_{\perp} - \boldsymbol{\Omega}^H\mathbf{M}^{-1}(\mathbf{M}^{-1}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}))^{\dagger H}\mathbf{A}_d(\boldsymbol{\theta})^H\mathbf{J}^H\mathbf{M}^{-2}\mathbf{h}_{\perp}\\ &= \frac{(v-K)\boldsymbol{\Omega}^H\mathbf{b}}{\mathbf{b}^H\mathbf{M}^2\mathbf{b}} = \mathbf{0}.\end{aligned} \tag{A.41}$$

Substituting (A.38) and (A.41) into (A.32) gives

$$\begin{aligned}CRB^{-1}(\boldsymbol{\theta}) =& \frac{1}{\pi^2 N}\text{diag}(\mathbf{p})\boldsymbol{\Omega}^H\mathbf{F}\Bigg[(\mathbf{F}^H\mathbf{M}^2\mathbf{F})^{-1}\\ &- \frac{(\mathbf{F}^H\mathbf{M}^2\mathbf{F})^{-1}\mathbf{F}^H\mathbf{h}_{\|}\mathbf{h}_{\|}^H\mathbf{F}(\mathbf{F}^H\mathbf{M}^2\mathbf{F})^{-1}}{\frac{(v-K)^2}{\mathbf{b}^H\mathbf{M}^2\mathbf{b}} + \mathbf{h}_{\|}^H\mathbf{F}(\mathbf{F}^H\mathbf{M}^2\mathbf{F})^{-1}\mathbf{F}^H\mathbf{h}_{\|}}\Bigg]\mathbf{F}^H\boldsymbol{\Omega}\text{diag}(\mathbf{p})\\ =& \frac{1}{\pi^2 N}\text{diag}(\mathbf{p})\boldsymbol{\Omega}^H\mathbf{F}\left(\mathbf{F}^H\left(\mathbf{M}^2 + \frac{\mathbf{b}^H\mathbf{M}^2\mathbf{b}}{(v-K)^2}\mathbf{h}_{\|}\mathbf{h}_{\|}^H\right)\mathbf{F}\right)^{-1}\mathbf{F}^H\boldsymbol{\Omega}\text{diag}(\mathbf{p}),\end{aligned} \tag{A.42}$$

where the last equality is obtained by using the matrix inversion lemma [187]. Now, given the definition of $\mathbf{h}_{\|}$ in (A.33), it is possible to show that

$$\mathbf{M}^2 + \frac{\mathbf{b}^H\mathbf{M}^2\mathbf{b}}{(v-K)^2}\mathbf{h}_{\|}\mathbf{h}_{\|}^H = \mathbf{Q}\mathbf{M}^2\mathbf{Q} + \frac{\mathbf{M}^2\mathbf{b}\mathbf{b}^H\mathbf{M}^2}{\mathbf{b}^H\mathbf{M}^2\mathbf{b}}. \tag{A.43}$$

Inserting (A.43) into (A.42) completes the proof.

A.5 Proof of Theorem 3.4

The proof is compromised of two steps. The first step involves simplification of the covariance matrix of DoA estimation errors through inserting the optimal weighting matrix given in Theorem 3.4 into (3.28). At the second step, we simplify the CRB expression given in Theorem 3.3 through doing some algebraic manipulations and show that the CRB coincides with the simplified covariance matrix of DoA estimation errors given in the first step.

A.5.1 Simplification of the errors covariance matrix

Considering the expression for $\mathbf{W}_{\text{opt}}$ from Theorem 3.4 and exploiting $\mathbf{F}^H\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}) = 0$ and $\Pi^{\perp}_{\mathbf{W}_{\text{opt}}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})} = \Pi_{\mathbf{W}_{\text{opt}}^{-\frac{1}{2}}\mathbf{F}}$ results in

$$\mathbf{W}_{opt}^{\frac{1}{2}}\Pi^{\perp}_{\mathbf{W}_{\text{opt}}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{W}_{\text{opt}}^{\frac{1}{2}} = \mathbf{F}(\mathbf{F}^H\mathbf{S}\mathbf{F})^{-1}\mathbf{F}^H. \tag{A.44}$$

Let now introduce the following Lemma, whereby we can proceed further with simplification of (A.44).

Lemma A.4. *The matrix* $\mathbf{Q}\mathbf{M}^2\mathbf{Q}^H \in \mathbb{C}^{M^2\times M^2}$ *can be decomposed as* $\mathbf{V}\mathbf{V}^H$*, where* $\mathbf{V}^H \in \mathbb{C}^{M^2\times M^2}$ *is a singular matrix with rank* M^2-1 *whose null space is spanned by* $\mathbf{b}$*.*

Proof. For any given matrices $\mathbf{B}$ and $\mathbf{C}$, it is known that $\text{rank}(\mathbf{B}-\mathbf{C}) \geq |\text{rank}(\mathbf{B}) - \text{rank}(\mathbf{C})|$ [187]. Hence, recalling the definition of $\mathbf{Q}$ in (3.29) , we have

$$\text{rank}(\mathbf{Q}) \geq M^2 - 1. \tag{A.45}$$

On the other hand, given (3.30), we obtain

$$\mathbf{b}^H\mathbf{Q} = \mathbf{b}^H - \frac{\text{vec}^H\left(\mathbf{U}_n\mathbf{U}_n^H\right)\mathbf{T}\mathbf{g}}{v-K}\mathbf{b}^H. \tag{A.46}$$

Let $\mathbf{g}' \in \{0,1\}^{(2v-1)\times 1}$ be a column vector with $[\mathbf{g}']_i = \delta[i-v]$. Utilizing (A.75), (A.72) and [52, Corollary 3], we observe

$$\mathbf{T}\mathbf{g} = \mathbf{T}'\mathbf{g}' = \text{vec}(\mathbf{I}_v). \tag{A.47}$$

where $\mathbf{T}' \in \{0,1\}^{v^2\times 2v-1}$ is defined in (A.72) in Appendix A.7. Inserting (A.47) into (A.46) gives

$$\mathbf{b}^H\mathbf{Q} = \mathbf{b}^H - \frac{\text{tr}(\mathbf{U}_n\mathbf{U}_n^H)}{v-K}\mathbf{b}^H = \mathbf{b}^H - \mathbf{b}^H = 0. \tag{A.48}$$

which implies that

$$\text{rank}(\mathbf{Q}) \leq M^2 - 1. \tag{A.49}$$

Comparing (A.45) and (A.49) concludes that $\text{rank}(\mathbf{Q}) = M^2 - 1$.

The fact $\text{rank}(\mathbf{Q}) = M^2 - 1$ implies that the dimension of the null space of $\mathbf{Q}^H$ is equal to 1. In addition, it follows from (A.46) that $\mathbf{b} \in \mathcal{N}(\mathbf{Q}^H)$. This means that the vector $\mathbf{b}$ spans the null space of $\mathbf{Q}^H$.

Let now define the matrix $\mathbf{V} = \mathbf{Q}\mathbf{M}$. Since $\mathbf{R}$ is positive definite, the matrix $\mathbf{M} = \left(\mathbf{R}^T \otimes \mathbf{R}\right)^{\frac{1}{2}}$ is also positive definite, implying that $\mathcal{R}(\mathbf{V}) = \mathcal{R}(\mathbf{Q})$. Hence, $\mathbf{V}^H$ is also a matrix of rank $M^2 - 1$ whose null space is spanned by the vector $\mathbf{b}$. □

It follows from Lemma A.2 and Lemma A.4 that $\mathcal{N}(\mathbf{V}^H) \subset \mathcal{R}(\mathbf{F})$ and, in turn, $\mathcal{N}(\mathbf{F}^H) \subset \mathcal{R}(\mathbf{V})$. This implies that a K-dimensional subspace of $\mathcal{R}(\mathbf{V})$ is a spanning set for the null space of $\mathbf{F}^H$. Hence, making use of Lemma A.4 and recalling the definition of $\mathbf{S}$ in (3.36) result in

$$\begin{aligned}\mathbf{F}^H\mathbf{S}\mathbf{F} &= \mathbf{F}^H\Pi_{\mathbf{F}}\mathbf{V}\mathbf{V}^H\Pi_{\mathbf{F}}\mathbf{F} + \mathbf{F}^H\mathbf{b}\mathbf{b}^H\mathbf{F} \\ &= \mathbf{F}^H\begin{bmatrix}\bar{\mathbf{V}} & \mathbf{b}\end{bmatrix}\begin{bmatrix}\bar{\mathbf{V}}^H \\ \mathbf{b}^H\end{bmatrix}\mathbf{F}, \end{aligned} \tag{A.50}$$

where $\bar{\mathbf{V}} = \mathbf{U}_v\mathbf{\Lambda}_v \in \mathbb{C}^{M^2\times(M^2-K-1)}$ with $\mathbf{U}_v \in \mathbb{C}^{M^2\times(M^2-K-1)}$ being comprised of left-singular vectors of $\Pi_{\mathbf{F}}\mathbf{V}$ and the diagonal matrix $\mathbf{\Lambda}_v \in \mathbb{C}^{(M^2-K-1)\times(M^2-K-1)}$ its corresponding singular values, meaning that

$$\mathcal{R}(\bar{\mathbf{V}}) \subset \mathcal{R}(\mathbf{F}), \tag{A.51}$$

$$\mathcal{R}(\bar{\mathbf{V}}) \subset \mathcal{R}(\mathbf{V}), \tag{A.52}$$

$$\operatorname{rank}(\bar{\mathbf{V}}) = M^2 - K - 1. \tag{A.53}$$

It follows from (A.52) and Lemma A.4 that $\mathbf{b} \notin \mathcal{R}(\bar{\mathbf{V}})$. From (A.53) and the fact that $\mathbf{b} \notin \mathcal{R}(\bar{\mathbf{V}})$, it can be deduced that

$$\operatorname{rank}(\begin{bmatrix}\bar{\mathbf{V}} & \mathbf{b}\end{bmatrix}) = M^2 - K = \operatorname{rank}(\mathbf{F}). \tag{A.54}$$

Comparing (A.51), (A.53) and (A.54) proves that $\mathcal{R}(\begin{bmatrix}\bar{\mathbf{V}} & \mathbf{b}\end{bmatrix}) = \mathcal{R}(\mathbf{F})$. Hence, there is a full rank matrix $\mathbf{X} \in \mathbb{C}^{(M^2-K)\times(M^2-K)}$ such that

$$\begin{bmatrix}\bar{\mathbf{V}} & \mathbf{b}\end{bmatrix} = \mathbf{F}\mathbf{X}. \tag{A.55}$$

Inserting (A.50) and (A.55) into (A.44) yields

$$\begin{aligned}\mathbf{W}_{\text{opt}}^{\frac{1}{2}}\Pi^{\perp}_{\mathbf{W}_{\text{opt}}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{W}_{\text{opt}}^{\frac{1}{2}} &= \mathbf{F}^{\dagger H}\mathbf{X}^{-H}\mathbf{X}^{-1}\mathbf{F}^{\dagger} = (\mathbf{F}\mathbf{X})^{\dagger H}(\mathbf{F}\mathbf{X})^{\dagger} \\ &= \begin{bmatrix}\bar{\mathbf{V}}^{\dagger H} & \mathbf{b}^{\dagger H}\end{bmatrix}\begin{bmatrix}\bar{\mathbf{V}}^{\dagger} \\ \mathbf{b}^{\dagger}\end{bmatrix}. \end{aligned} \tag{A.56}$$

Considering (A.56), Lemma A.3 and $\mathcal{R}(\mathbf{b}^{\dagger H}) = \mathcal{R}(\mathbf{b})$ gives

$$\begin{aligned}\boldsymbol{\Omega}^H \mathbf{W}_{\text{opt}}^{\frac{1}{2}} \Pi^{\perp}_{\mathbf{W}_{\text{opt}}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})} \mathbf{W}_{\text{opt}}^{\frac{1}{2}} &= \boldsymbol{\Omega}^H \begin{bmatrix} \bar{\mathbf{V}}^{\dagger H} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{V}}^{\dagger} \\ \mathbf{b}^{\dagger} \end{bmatrix} \\ &= \boldsymbol{\Omega}^H (\bar{\mathbf{V}}\bar{\mathbf{V}}^H)^{\dagger}. \end{aligned} \tag{A.57}$$

Moreover, given (A.50) and using $\Pi^{\perp}_{\mathbf{W}_{\text{opt}}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})} = \Pi_{\mathbf{W}_{\text{opt}}^{-\frac{1}{2}}\mathbf{F}}$, we have

$$\begin{aligned}&\Pi^{\perp}_{\mathbf{W}_{\text{opt}}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})} \mathbf{W}_{\text{opt}}^{\frac{1}{2}} \mathbf{Q}\mathbf{M}^2\mathbf{Q}^H \mathbf{W}_{\text{opt}}^{\frac{1}{2}} \Pi^{\perp}_{\mathbf{W}_{\text{opt}}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})} \\ &= \mathbf{W}_{\text{opt}}^{-\frac{1}{2}} \mathbf{F}(\mathbf{F}^H \mathbf{W}_{\text{opt}}^{-1} \mathbf{F})^{-1} \mathbf{F}^H \bar{\mathbf{V}}\bar{\mathbf{V}}^H \mathbf{F}(\mathbf{F}^H \mathbf{W}_{\text{opt}}^{-1} \mathbf{F})^{-1} \mathbf{F}^H \mathbf{W}_{\text{opt}}^{-\frac{1}{2}} \\ &= \Pi^{\perp}_{\mathbf{W}_{\text{opt}}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})} \mathbf{W}_{\text{opt}}^{\frac{1}{2}} \bar{\mathbf{V}}\bar{\mathbf{V}}^H \mathbf{W}_{\text{opt}}^{\frac{1}{2}} \Pi^{\perp}_{\mathbf{W}_{\text{opt}}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}. \end{aligned} \tag{A.58}$$

Eventually, substituting (A.57) and (A.58) into (3.28), we find

$$\begin{aligned}\mathbf{C}_{wls} =& \frac{1}{\pi^2 N} \text{diag}^{-1}(\mathbf{p}) \left(\boldsymbol{\Omega}^H (\bar{\mathbf{V}}\bar{\mathbf{V}}^H)^{\dagger} \boldsymbol{\Omega} \right)^{-1} \\ &\times \left(\boldsymbol{\Omega}^H (\bar{\mathbf{V}}\bar{\mathbf{V}}^H)^{\dagger} \bar{\mathbf{V}}\bar{\mathbf{V}}^H (\bar{\mathbf{V}}\bar{\mathbf{V}}^H)^{\dagger} \boldsymbol{\Omega} \right) \left(\boldsymbol{\Omega}^H (\bar{\mathbf{V}}\bar{\mathbf{V}}^H)^{\dagger} \boldsymbol{\Omega} \right)^{-1} \text{diag}^{-1}(\mathbf{p}) \\ =& \frac{1}{\pi^2 N} \left(\text{diag}(\mathbf{p}) \boldsymbol{\Omega}^H (\bar{\mathbf{V}}\bar{\mathbf{V}}^H)^{\dagger} \boldsymbol{\Omega} \text{diag}(\mathbf{p}) \right)^{-1} \end{aligned} \tag{A.59}$$

A.5.2 Simplification of the CRB expression

Now, we will show that the CRB expression given in Lemma 3.3 is also reduced to (A.59). Let decompose the vector $\frac{\mathbf{M}^2\mathbf{b}}{\sqrt{\mathbf{b}^H\mathbf{M}^2\mathbf{b}}}$ as a sum of two vectors $\mathbf{q}_{\|}$ and $\mathbf{q}_{\perp}$ where

$$\mathbf{q}_{\|} = \frac{\sqrt{\mathbf{b}^H\mathbf{M}^2\mathbf{b}}}{\|\mathbf{b}\|^2}\mathbf{b}, \quad \mathbf{q}_{\perp} = \frac{\mathbf{M}^2\mathbf{b}}{\sqrt{\mathbf{b}^H\mathbf{M}^2\mathbf{b}}} - \frac{\sqrt{\mathbf{b}^H\mathbf{M}^2\mathbf{b}}}{\|\mathbf{b}\|^2}\mathbf{b}. \tag{A.60}$$

It is observed that $\mathbf{q}_{\|}$ is a scaled version of $\mathbf{b}$ and that $\mathbf{b}^H\mathbf{q}_{\perp} = 0$. Hence, recalling (3.34) and taking account of (A.50), we have

$$\begin{aligned}\mathbf{F}^H\mathbf{H}\mathbf{F} &= \mathbf{F}^H\mathbf{V}\mathbf{V}\mathbf{F} + \mathbf{F}^H(\mathbf{q}_{\|} + \mathbf{q}_{\perp})(\mathbf{q}_{\|} + \mathbf{q}_{\perp})^H\mathbf{F} \\ &= \mathbf{F}^H \begin{bmatrix} \bar{\mathbf{V}} & \mathbf{q}_{\|} + \bar{\mathbf{q}}_{\perp} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{V}}^H \\ \mathbf{q}_{\|}^H + \bar{\mathbf{q}}_{\perp}^H \end{bmatrix} \mathbf{F}, \end{aligned} \tag{A.61}$$

where $\bar{\mathbf{q}}_{\perp} = \Pi_{\mathbf{F}}\mathbf{q}_{\perp}$. By definition, it is evident that $\bar{\mathbf{q}}_{\perp} \in \mathcal{R}(\mathbf{F})$ and $\mathbf{b}^H\bar{\mathbf{q}}_{\perp} = 0$, meaning that $\bar{\mathbf{q}}_{\perp} \in \mathcal{R}(\bar{\mathbf{V}})$. In addition, since $\mathbf{b} \notin \mathcal{R}(\bar{\mathbf{V}})$, $\mathbf{q}_{\|} \notin \mathcal{R}(\bar{\mathbf{V}})$ in turn. In consequence, considering (A.51) and (A.53), it can be inferred that $\mathcal{R}(\begin{bmatrix} \bar{\mathbf{V}} & \mathbf{q}_{\|} + \bar{\mathbf{q}}_{\perp} \end{bmatrix}) = \mathcal{R}(\mathbf{F})$ implying

that there is a full rank matrix $\mathbf{D} \in \mathbb{C}^{(M^2-K)\times(M^2-K)}$ such that

$$\begin{bmatrix} \bar{\mathbf{V}} & \mathbf{q}_{\|} + \bar{\mathbf{q}}_{\perp} \end{bmatrix} = \mathbf{F}\mathbf{D}. \tag{A.62}$$

By inserting (A.62) into (A.61) and doing some calculations similar to (A.56), we obtain

$$\mathbf{F}(\mathbf{F}^H\mathbf{H}\mathbf{F})^{-1}\mathbf{F}^H = \begin{bmatrix} \bar{\mathbf{V}} & \mathbf{q}_{\|} + \bar{\mathbf{q}}_{\perp} \end{bmatrix}^{\dagger H} \begin{bmatrix} \bar{\mathbf{V}}^H \\ \mathbf{q}_{\|}^H + \bar{\mathbf{q}}_{\perp}^H \end{bmatrix}^{\dagger}. \tag{A.63}$$

Since $\begin{bmatrix} \bar{\mathbf{V}} & \mathbf{q}_{\|} + \bar{\mathbf{q}}_{\perp} \end{bmatrix}$ is a full column rank matrix, its pseudoinverse by definition can be computed as

$$\begin{bmatrix} \bar{\mathbf{V}} & \mathbf{q}_{\|} + \bar{\mathbf{q}}_{\perp} \end{bmatrix}^{\dagger} = \begin{bmatrix} \bar{\mathbf{V}}\bar{\mathbf{V}}^H & \bar{\mathbf{V}}^H\bar{\mathbf{q}}_{\perp} \\ \bar{\mathbf{q}}_{\perp}^H\bar{\mathbf{V}} & \|\mathbf{q}_{\|}\|^2 + \|\bar{\mathbf{q}}_{\perp}\|^2 \end{bmatrix}^{-1} \begin{bmatrix} \bar{\mathbf{V}}^H \\ \mathbf{q}_{\|}^H + \bar{\mathbf{q}}_{\perp}^H \end{bmatrix} =$$
$$\begin{bmatrix} \bar{\mathbf{V}}^{\dagger} + \frac{\bar{\mathbf{V}}^{\dagger}\bar{\mathbf{q}}_{\perp}\bar{\mathbf{q}}_{\perp}^H\Pi_{\bar{\mathbf{V}}} - \bar{\mathbf{V}}^{\dagger}\bar{\mathbf{q}}_{\perp}(\mathbf{q}_{\|}^H + \bar{\mathbf{q}}_{\perp}^H)}{\|\mathbf{q}_{\|}\|^2 + \|\bar{\mathbf{q}}_{\perp}\|^2 - \bar{\mathbf{q}}_{\perp}^H\Pi_{\bar{\mathbf{V}}}\bar{\mathbf{q}}_{\perp}} \\ \frac{-\bar{\mathbf{q}}_{\perp}^H\Pi_{\bar{\mathbf{V}}} + \mathbf{q}_{\|}^H + \bar{\mathbf{q}}_{\perp}^H}{\|\mathbf{q}_{\|}\|^2 + \|\bar{\mathbf{q}}_{\perp}\|^2 - \bar{\mathbf{q}}_{\perp}^H\Pi_{\bar{\mathbf{V}}}\bar{\mathbf{q}}_{\perp}} \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{V}}^{\dagger} - \frac{\bar{\mathbf{V}}^{\dagger}\bar{\mathbf{q}}_{\perp}\mathbf{q}_{\|}^H}{\|\mathbf{q}_{\|}\|^2} \\ \mathbf{q}_{\|}^{\dagger} \end{bmatrix}, \tag{A.64}$$

where that last equality is obtained by exploiting the fact that $\bar{\mathbf{q}}_{\perp}$ belongs to the range space of $\bar{\mathbf{V}}$. Now, inserting (A.63) and (A.64) into the CRB expression given in (3.33) gives

$$\begin{aligned} CRB(\boldsymbol{\theta}) &= (\mathrm{diag}(\mathbf{p})\boldsymbol{\Omega}^H \begin{bmatrix} \bar{\mathbf{V}}^{\dagger H} - \frac{\mathbf{q}_{\|}\bar{\mathbf{q}}_{\perp}^H\bar{\mathbf{V}}^{\dagger H}}{\|\mathbf{q}_{\|}\|^2} & \mathbf{q}_{\|}^{\dagger H} \end{bmatrix} \\ &\quad \times \begin{bmatrix} \bar{\mathbf{V}}^{\dagger} - \frac{\bar{\mathbf{V}}^{\dagger}\bar{\mathbf{q}}_{\perp}\mathbf{q}_{\|}^H}{\|\mathbf{q}_{\|}\|^2} \\ \mathbf{q}_{\|}^{\dagger} \end{bmatrix} \boldsymbol{\Omega}\mathrm{diag}(\mathbf{p}))^{-1} \\ &= \frac{1}{\pi^2 N}\left(\mathrm{diag}(\mathbf{p})\boldsymbol{\Omega}^H(\bar{\mathbf{V}}\bar{\mathbf{V}}^H)^{\dagger}\boldsymbol{\Omega}\mathrm{diag}(\mathbf{p})\right)^{-1}, \end{aligned} \tag{A.65}$$

where the last equality is obtained by using Lemma A.3 and the fact that $\mathbf{q}_{\|}$ is a scaled version of $\mathbf{b}$.

Eventually, comparing (A.59) into (A.65) concludes the proof.

A.6 Proof of Lemma 3.2

We first prove positive definiteness of $\mathbf{W}_{\mathrm{opt}}$. It follows from (A.50) to (A.55) in Appendix A.5.1 that

$$\Pi^{\perp}_{\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{S}\Pi^{\perp}_{\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})} = \Pi_{\mathbf{F}}\mathbf{S}\Pi_{\mathbf{F}} = \mathbf{F}\mathbf{X}\mathbf{X}^H\mathbf{F}^H. \tag{A.66}$$

Consequently, we have

$$\mathbf{W}_{\text{opt}}^{-1} = \begin{bmatrix}\mathbf{FX} & \mathbf{JA}_d(\boldsymbol{\theta})\end{bmatrix}\begin{bmatrix}\mathbf{FX} & \mathbf{JA}_d(\boldsymbol{\theta})\end{bmatrix}^H. \tag{A.67}$$

The matrices $\mathbf{FX} \in \mathbb{C}^{M^2\times(M^2-K)}$ and $\mathbf{JA}_d(\boldsymbol{\theta}) \in \mathbb{C}^{M^2\times K}$ are full column rank by definition. Further, it is easily checked that $\mathbf{FX}$ and $\mathbf{JA}_d(\boldsymbol{\theta})$ are orthogonal subspaces, i.e., $\mathbf{X}^H\mathbf{F}^H\mathbf{JA}_d(\boldsymbol{\theta}) = \mathbf{0}$. Hence, $\begin{bmatrix}\mathbf{FX} & \mathbf{JA}_d(\boldsymbol{\theta})\end{bmatrix} \in \mathbb{C}^{M^2\times M^2}$ is full rank. This implies that $\mathbf{W}_{\text{opt}}$ is also full rank and thereby positive definite.

Now we show that the estimate of $\mathbf{W}_{\text{opt}}$, obtained from either CAB-MUSIC or CAB-ESPRIT, is always positive definite regardless of the available number of snapshots. Let assume that an arbitrary number of snapshots is available and $\hat{\mathbf{S}} = \hat{\mathbf{Q}}\hat{\mathbf{M}}^2\hat{\mathbf{Q}}^H + \hat{\mathbf{b}}\hat{\mathbf{b}}^H$ denotes the estimate of $\mathbf{S}$, obtained based on the sample covarinace matrix. Further, let $\hat{\boldsymbol{\theta}}$ be an estimate of $\boldsymbol{\theta}$ given by CAB-MUSIC or CAB-ESPRIT and $\hat{\mathbf{F}}$ is the estimate of $\mathbf{F}$ obtained from $\hat{\boldsymbol{\theta}}$. Following similar kind of arguments and derivations provided from Lemma A.4 to (A.55) in Appendix A.5.1, it can readily be shown that

$$\Pi^{\perp}_{\mathbf{JA}_d(\hat{\boldsymbol{\theta}})}\hat{\mathbf{S}}\Pi^{\perp}_{\mathbf{JA}_d(\hat{\boldsymbol{\theta}})} = \Pi_{\hat{\mathbf{F}}}\hat{\mathbf{S}}\Pi_{\hat{\mathbf{F}}} = \hat{\mathbf{F}}\hat{\mathbf{X}}\hat{\mathbf{X}}^H\hat{\mathbf{F}}^H, \tag{A.68}$$

where $\hat{\mathbf{X}} \in \mathbb{C}^{(M^2-K)\times(M^2-K)}$ is a full rank matrix. Hence, using (A.68), we observe that

$$\hat{\mathbf{W}}_{opt}^{-1} = \begin{bmatrix}\hat{\mathbf{F}}\hat{\mathbf{X}} & \mathbf{JA}_d(\hat{\boldsymbol{\theta}})\end{bmatrix}\begin{bmatrix}\hat{\mathbf{F}}\hat{\mathbf{X}} & \mathbf{JA}_d(\hat{\boldsymbol{\theta}})\end{bmatrix}^H. \tag{A.69}$$

Once again, we note that $\hat{\mathbf{F}}\hat{\mathbf{X}} \in \mathbb{C}^{M^2\times(M^2-K)}$ and $\mathbf{JA}_d(\hat{\boldsymbol{\theta}}) \in \mathbb{C}^{M^2\times K}$ are full column rank by definition, and moreover, they span orthogonal subspaces. Hence, $\begin{bmatrix}\hat{\mathbf{F}}\hat{\mathbf{X}} & \mathbf{JA}_d(\hat{\boldsymbol{\theta}})\end{bmatrix} \in \mathbb{C}^{M^2\times M^2}$ and in turn $\hat{\mathbf{W}}_{\text{opt}}$ is full rank.

A.7 Commutation Matrix and Some Relevant Lemmas

Definition A.1. Let $\mathbf{B}$ be any matrix in $\mathbb{R}^{p\times p}$. Then, there exists a permutation matrix $\mathbf{K}_p \in \{0,1\}^{p^2\times p^2}$ such that $\text{vec}\left(\mathbf{B}^T\right) = \mathbf{K}_p\text{vec}\left(\mathbf{B}\right)$. This matrix, called the commutation matrix, is an involutory and symmetric matrix, i.e., $\mathbf{K}_p = \mathbf{K}_p^T = \mathbf{K}_p^{-1}$ [186].

Lemma A.5. $\mathbf{K}_M\mathbf{J}^{\dagger H}\mathbf{T}^H\mathbf{K}_v = \mathbf{J}^{\dagger H}\mathbf{T}^H$ *where* $\mathbf{K}_M \in \{0,1\}^{M^2\times M^2}$ *and* $\mathbf{K}_v \in \{0,1\}^{v^2\times v^2}$ *are commutation matrices defined according to Definition A.1.*

Proof. It has been proved in [52] that $\mathbf{J}$ has orthogonal columns. Further, recalling the definition of $\mathbf{J}$ given in Definition 3.1, since $\|\text{vec}(\mathbf{L}_i^T)\|^2 = \|\text{vec}(\mathbf{L}_i)\|^2$ for $0 < n < D-1$,

it is readily confirmed that

$$(\mathbf{J}^H\mathbf{J})^{-1} = \mathrm{diag}(\begin{bmatrix} \frac{1}{\|\mathrm{vec}(\mathbf{L}_{D-1})\|^2} & \cdots & \frac{1}{\|\mathrm{vec}(\mathbf{L}_0)\|^2} & \cdots & \frac{1}{\|\mathrm{vec}(\mathbf{L}_{D-1})\|^2} \end{bmatrix}^T). \tag{A.70}$$

Recalling $\mathbf{A}_\vartheta$ and $\mathbf{A}_v$ in the proof of Lemmas 3.1 and A.1, and based on [52, Appendix B], it is straightforward to show that

$$\mathbf{A}_v^*(\boldsymbol{\theta}) \odot \mathbf{A}_v(\boldsymbol{\theta}) = \mathbf{T}'\mathbf{A}_\vartheta(\boldsymbol{\theta}), \tag{A.71}$$

where $\mathbf{T}' \in \{0,1\}^{v^2\times 2v-1}$ can be defined like $\mathbf{J}$ as

$$\mathbf{T}' = \begin{bmatrix} \mathrm{vec}(\mathbf{G}_{v-1}^T) & \cdots & \mathrm{vec}(\mathbf{G}_0) & \cdots & \mathrm{vec}(\mathbf{G}_{v-1}) \end{bmatrix}, \tag{A.72}$$

where $[\mathbf{G}_n]_{p,q} = \begin{cases} 1, & \text{if } p-q=n, \\ 0, & \text{otherwise.} \end{cases}$. On the other hand, comparing the vectorized form of (A.1) and (A.2) gives

$$\mathbf{T}\mathbf{A}_d(\boldsymbol{\theta}) = \mathbf{A}_v^*(\boldsymbol{\theta}) \odot \mathbf{A}_v(\boldsymbol{\theta}). \tag{A.73}$$

Making use of (A.71), (A.73) and the fact that

$$\mathbf{A}_\vartheta(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{0}_{(2v-1)\times(D-v)} & \mathbf{I}_{2v-1} & \mathbf{0}_{(2v-1)\times(D-v)} \end{bmatrix} \mathbf{A}_d(\boldsymbol{\theta}) \tag{A.74}$$

results in

$$\mathbf{T} = \begin{bmatrix} \mathbf{0}_{v^2\times(D-v)} & \mathbf{T}' & \mathbf{0}_{v^2\times(D-v)} \end{bmatrix}. \tag{A.75}$$

Combining (3.5), (A.70), (A.72) and (A.75) gives

$$\begin{aligned} \mathbf{K}_M\mathbf{J}^{\dagger H}\mathbf{T}^H\mathbf{K}_v = \mathbf{J}^{\dagger H}\mathbf{T}^H &= \frac{\mathrm{vec}(\mathbf{L}_0)\mathrm{vec}^T(\mathbf{G}_0)}{\|\mathrm{vec}(\mathbf{L}_0)\|^2} \\ &+ \sum_{i=1}^{v-1} \frac{\mathrm{vec}(\mathbf{L}_i)\mathrm{vec}^T(\mathbf{G}_i) + \mathrm{vec}(\mathbf{L}_i^T)\mathrm{vec}^T(\mathbf{G}_i^T)}{\|\mathrm{vec}(\mathbf{L}_i)\|^2}. \end{aligned} \tag{A.76}$$

□

Lemma A.6. *Let $\mathbf{K}_M \in \{0,1\}^{M^2\times M^2}$ be the commutation matrix as defined according to Definition A.1 and $\mathbf{K}_M\mathbf{W} = \mathbf{W}^*\mathbf{K}_M$. Then, it follows that*

a) $\hat{\mathbf{Q}}^*\hat{\mathbf{r}}^* = \mathbf{K}_M\hat{\mathbf{Q}}\hat{\mathbf{r}}$,

b) $\hat{\mathbf{r}}^T\hat{\mathbf{Q}}^T\mathbf{K}_M = \hat{\mathbf{r}}^H\hat{\mathbf{Q}}^H$,

c) $(\mathbf{J}\frac{\partial\mathbf{A}_d(\boldsymbol{\theta})}{\partial\theta_k})^T = (\mathbf{J}\frac{\partial\mathbf{A}_d(\boldsymbol{\theta})}{\partial\theta_k})^H\mathbf{K}_M$,

d) $\mathbf{K}_M\mathbf{W}^{\frac{1}{2}*}(\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}))^{\dagger T} = \mathbf{W}^{\frac{1}{2}}(\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}))^{\dagger H}$,

e) $\mathbf{K}_M\mathbf{W}^{\frac{1}{2}*}\Pi^{\perp *}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}*}\mathbf{K}_M = \mathbf{W}^{\frac{1}{2}}\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}}$.

Proof. The proof of each item is given in the following:

a) and b): Since $\hat{\mathbf{R}}$ is Hermitian [185], it readily follows that $\hat{\mathbf{r}}^* = \mathbf{K}_M\hat{\mathbf{r}}$. Hence, recalling (3.13) and making use of Lemma A.5 and the fact that $\hat{\mathbf{U}}_n\hat{\mathbf{U}}_n^H$ is Hermitian, we obtain

$$\hat{\mathbf{Q}}^*\hat{\mathbf{r}}^* = \mathbf{K}_M\big(\hat{\mathbf{r}} - \frac{\mathrm{vec}(\mathbf{I}_M)\mathrm{vec}^H(\hat{\mathbf{U}}_n\hat{\mathbf{U}}_n^H)\mathbf{K}_v\mathbf{T}\mathbf{J}^\dagger\mathbf{K}_M\hat{\mathbf{r}}}{v-K}\big) = \mathbf{K}_M\hat{\mathbf{Q}}\hat{\mathbf{r}}. \tag{A.77}$$

Further, transposing (A.77) results in item b).

c): Let $\mathbf{a} \in \mathbf{C}^{K\times 1}$ be an arbitrary vector. From (3.4), we find

$$\mathbf{K}_M\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})\mathbf{a} = \mathrm{vec}\left(\mathbf{A}^*(\boldsymbol{\theta})\mathrm{diag}\left(\mathbf{a}\right)\mathbf{A}^T(\boldsymbol{\theta})\right) = \left(\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})\right)^*\mathbf{a}. \tag{A.78}$$

Since $\mathbf{a}$ is an arbitrary vector, it can be concluded that

$$\mathbf{K}_M\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}) = \left(\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})\right)^*. \tag{A.79}$$

Eventually, replacing both side of (A.79) with their conjugate transposes and taking derivative with respect to θ_k gives c).

d): Using (A.79) and $\mathbf{K}_M\mathbf{W} = \mathbf{W}^*\mathbf{K}_M$, it is observed that

$$\begin{aligned}\mathbf{K}_M\mathbf{W}^{\frac{1}{2}}(\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}))^{\dagger H} &= \mathbf{W}^*\left(\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})\right)^*\left(\left(\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})\right)^H\mathbf{W}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})\right)^{-1}\\ &= \mathbf{W}^{\frac{1}{2}*}(\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}))^{\dagger T}.\end{aligned} \tag{A.80}$$

Multiplying both sides of (A.80) by $\mathbf{K}_M$ leads to d).

e) and f): Making use of (A.79), transpose of (A.80) and the fact that $\mathbf{W} = \mathbf{K}_M\mathbf{W}^*\mathbf{K}_M$, we get

$$\begin{aligned}\mathbf{W}^{\frac{1}{2}}\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}} &= \mathbf{K}_M\mathbf{W}^*\mathbf{K}_M - \mathbf{K}_M\mathbf{W}^*(\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}))^*(\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}))^{\dagger *}\mathbf{W}^{\frac{1}{2}*}\mathbf{K}_M\\ &= \mathbf{K}_M\mathbf{W}^{\frac{1}{2}*}\Pi^{\perp *}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}*}\mathbf{K}_M.\end{aligned} \tag{A.81}$$

□

Lemma A.7. $\mathbf{K}_M\mathbf{W}_{\mathrm{opt}} = \mathbf{W}^*_{\mathrm{opt}}\mathbf{K}_M$

Proof. Since $\mathbf{K}_M$ is involutory and symmetric, to prove the lemma, we can equivalently show that $\mathbf{K}_M\mathbf{W}^*_{\mathrm{opt}}\mathbf{K}_M = \mathbf{W}_{\mathrm{opt}}$. From (A.79), it is readily observed that

$$\mathbf{K}_M\Pi^{\perp *}_{\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})} = \Pi^{\perp}_{\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{K}_M. \tag{A.82}$$

Now, it follows from (A.79) and (A.82) that

$$\mathbf{K}_M \mathbf{W}_{\text{opt}}^* \mathbf{K}_M = \left(\Pi^{\perp}_{\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})} \mathbf{K}_M \mathbf{S}^* \mathbf{K}_M \Pi^{\perp}_{\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})} + \mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}) \mathbf{A}_d^H(\boldsymbol{\theta}) \mathbf{J}^T \right)^{-1}. \tag{A.83}$$

In consequence, $\mathbf{K}_M \mathbf{W}_{\text{opt}}^* \mathbf{K}_M = \mathbf{W}_{\text{opt}}$ if and only if $\mathbf{K}_M \mathbf{S}^* \mathbf{K}_M = \mathbf{S}$. Exploiting Lemma A.5 and recalling (3.30), we find

$$\mathbf{K}_M \mathbf{b}^* = \mathbf{K}_M \mathbf{J}^{\dagger H} \mathbf{T}^H \mathbf{K}_v \text{vec}\left(\mathbf{U}_n \mathbf{U}_n^H \right) = \mathbf{b}. \tag{A.84}$$

Further, recalling (3.29) and using transpose of (A.84), we have

$$\mathbf{K}_M \mathbf{Q}^* = \mathbf{K}_M - \frac{\mathbf{K}_M \text{vec}(\mathbf{I}_M) \mathbf{b}^T}{v - K} = \mathbf{Q}\mathbf{K}_M \tag{A.85}$$

Now recalling (3.36) and making use of (A.84) and (A.85) gives

$$\begin{aligned} \mathbf{K}_M \mathbf{S}^* \mathbf{K}_M &= \mathbf{K}_M \mathbf{Q}^* \mathbf{M}^{2*} \mathbf{Q}^T \mathbf{K}_M + \mathbf{K}_M \mathbf{b}^* \mathbf{b}^T \mathbf{K}_M \\ &= \mathbf{Q}\mathbf{K}_M \mathbf{M}^{2*} \mathbf{K}_M \mathbf{Q}^H + \mathbf{b}\mathbf{b}^H = \mathbf{S}, \end{aligned} \tag{A.86}$$

where the last equality is obtained by using the fact that $\mathbf{K}_M \mathbf{M}^{2*} \mathbf{K}_M = \mathbf{K}_M (\mathbf{R} \otimes \mathbf{R}^T) \mathbf{K}_M = \mathbf{R}^T \otimes \mathbf{R} = \mathbf{M}^2$. □

Appendix B

Appendices of Chapter 4

B.1 Proof of Theorem 4.1

We first prove the sufficiency. Assume that $\boldsymbol{\theta}_0 \in [-\pi/2, \pi/2]^{K\times 1}$ is identifiable from $\mathbf{Y}$. This implies that $f(\mathbf{Y} \mid \boldsymbol{\theta}_0, \mathbf{p}, \sigma^2) \neq f(\mathbf{Y} \mid \breve{\boldsymbol{\theta}}, \breve{\mathbf{p}}, \breve{\sigma}^2)$ for any arbitrary values of $\breve{\boldsymbol{\theta}} \neq \boldsymbol{\theta}_0 \in [-\pi/2, \pi/2]^{K\times 1}$, $\mathbf{p} \in \mathbb{R}_{>0}^{K\times 1}$, $\breve{\mathbf{p}} \in \mathbb{R}_{>0}^{K\times 1}$, σ^2 and $\breve{\sigma}^2$. Hence, considering $\mathbf{y}(0), \mathbf{y}(1), \cdots, \mathbf{y}(N-1)$ are independent and identically distributed with $\mathbf{y}(t) \sim \mathcal{CN}(\mathbf{0}, \mathbf{R})$, we have

$$\mathbf{A}(\boldsymbol{\theta}_0)\text{diag}(\mathbf{p})\mathbf{A}^H(\boldsymbol{\theta}_0) + \sigma^2\mathbf{I}_M \neq \mathbf{A}(\breve{\boldsymbol{\theta}})\text{diag}(\breve{\mathbf{p}})\mathbf{A}^H(\breve{\boldsymbol{\theta}}) + \breve{\sigma}^2\mathbf{I}_M, \tag{B.1}$$

for all $\breve{\boldsymbol{\theta}} \neq \boldsymbol{\theta}_0 \in [-\pi/2, \pi/2]^{K\times 1}$, $\mathbf{p} \in \mathbb{R}_{>0}^{K\times 1}$, $\breve{\mathbf{p}} \in \mathbb{R}_{>0}^{K\times 1}$, σ^2 and $\breve{\sigma}^2$.

In what follows, we employ the method of proof by contradiction to prove the sufficiency. In particular, we assume that $\boldsymbol{\theta}_0 \in [-\pi/2, \pi/2]^{K\times 1}$ is non-identifiable from $\mathbf{X}$. Hence, there exists a $\breve{\boldsymbol{\theta}} \neq \boldsymbol{\theta}_0 \in [-\pi/2, \pi/2]^{K\times 1}$ at which $f(\mathbf{X} \mid \boldsymbol{\theta}_0, \widetilde{\mathbf{p}}, \widetilde{\sigma}^2) = f(\mathbf{X} \mid \breve{\boldsymbol{\theta}}, \dot{\mathbf{p}}, \dot{\sigma}^2)$ for some values of $\widetilde{\mathbf{p}} \in \mathbb{R}_{>0}^{K\times 1}$, $\dot{\mathbf{p}} \in \mathbb{R}_{>0}^{K\times 1}$, $\widetilde{\sigma}^2$ and $\dot{\sigma}^2$. It is readily clear from assumption **A4** and (4.6) that $\mathbb{E}\{\mathbf{x}(t_1)\mathbf{x}^H(t_2)\} = \mathbf{0}$ when $t_1 \neq t_2$. Accordingly, we have

$$\begin{aligned} &\mathbb{E}\left\{\mathbf{X}\mathbf{X}^H \,|\, \boldsymbol{\theta}_0, \mathbf{p}, \sigma^2\right\} = \mathbb{E}\left\{\mathbf{X}\mathbf{X}^H \,|\, \breve{\boldsymbol{\theta}}, \breve{\mathbf{p}}, \breve{\sigma}^2\right\}, \\ \Rightarrow &\sum_{t=0}^{N-1} \mathbb{E}\{\mathbf{x}(t)\mathbf{x}^H(t) \,|\, \boldsymbol{\theta}_0, \mathbf{p}, \sigma^2\} = \sum_{t=0}^{N-1} \mathbb{E}\{\mathbf{x}(t)\mathbf{x}^H(t) \,|\, \breve{\boldsymbol{\theta}}, \breve{\mathbf{p}}, \breve{\sigma}^2\} \end{aligned} \tag{B.2}$$

From (B.2), (4.7), (4.3) and the fact that the arcsine function is one-to-one when its argument is between -1 and 1, it follows that

$$\frac{1}{\widetilde{\sigma}^2+\sum_{k=1}^{K}\tilde{p}_k}\left[\mathbf{A}(\boldsymbol{\theta}_0)\mathrm{diag}(\widetilde{\mathbf{p}})\mathbf{A}^H(\boldsymbol{\theta}_0)+\widetilde{\sigma}^2\mathbf{I}_M\right]= \frac{1}{\dot{\sigma}^2+\sum_{k=1}^{K}\dot{p}_k}\left[\mathbf{A}(\breve{\boldsymbol{\theta}})\mathrm{diag}(\dot{\mathbf{p}})\mathbf{A}^H(\breve{\boldsymbol{\theta}})+\dot{\sigma}^2\mathbf{I}_M\right]. \tag{B.3}$$

Considering $\mathbf{p}=\frac{\widetilde{\mathbf{p}}}{\widetilde{\sigma}^2+\sum_{k=1}^{K}\tilde{p}_k}$, $\sigma^2=\frac{\widetilde{\sigma}^2}{\widetilde{\sigma}^2+\sum_{k=1}^{K}\tilde{p}_k}$, $\breve{\mathbf{p}}=\frac{\dot{\mathbf{p}}}{\dot{\sigma}^2+\sum_{k=1}^{K}\dot{p}_k}$ and $\breve{\sigma}^2=\frac{\dot{\sigma}^2}{\dot{\sigma}^2+\sum_{k=1}^{K}\dot{p}_k}$, we obtain

$$\mathbf{A}(\boldsymbol{\theta}_0)\mathrm{diag}(\mathbf{p})\mathbf{A}^H(\boldsymbol{\theta}_0)+\sigma^2\mathbf{I}_M=\mathbf{A}(\breve{\boldsymbol{\theta}})\mathrm{diag}(\breve{\mathbf{p}})\mathbf{A}^H(\breve{\boldsymbol{\theta}})+\breve{\sigma}^2\mathbf{I}_M. \tag{B.4}$$

which is in contradiction with (B.1). Hence, the initial assumption that $\boldsymbol{\theta}_0\in[-\pi/2,\pi/2]^{K\times 1}$ is non-identifiable from $\mathbf{X}$ cannot be true. This proves the sufficiency.

To show the necessity, let assume that $\boldsymbol{\theta}_0\in[-\pi/2,\pi/2]^{K\times 1}$ is non-identifiable from $\mathbf{Y}$. This implies that there exist some $\breve{\boldsymbol{\theta}}\in[-\pi,\pi]^{q\times 1}\neq\boldsymbol{\theta}_0$, $\mathbf{p}$, $\breve{\mathbf{p}}$, σ^2 and $\breve{\sigma}^2$ for which $f(\mathbf{Y}\mid\boldsymbol{\theta}_0,\mathbf{p},\sigma^2)=f(\mathbf{Y}\mid\breve{\boldsymbol{\theta}},\breve{\mathbf{p}},\breve{\sigma}^2)$. Since the true PDF of $\mathbf{X}$ is obtained from the orthant probabilities of $\mathbf{Y}$, it is readily deduced that $f(\mathbf{X}\mid\boldsymbol{\theta}_0,\mathbf{p},\sigma^2)=f(\mathbf{X}\mid\breve{\boldsymbol{\theta}},\breve{\mathbf{p}},\breve{\sigma}^2)$ as well. This proves that identifiability of $\boldsymbol{\theta}_0\in[-\pi/2,\pi/2]^{K\times 1}$ from $\mathbf{Y}$ is a necessary condition for identifiability of $\boldsymbol{\theta}_0\in[-\pi/2,\pi/2]^{K\times 1}$ from $\mathbf{X}$.

B.2 Proof of Theorem 4.2

We first prove **S1**. Consider arbitrary $\boldsymbol{\theta}\in[-\pi/2,\pi/2]^{K\times 1}$ and $\breve{\boldsymbol{\theta}}\in[-\pi/2,\pi/2]^{K\times 1}$ such that $[\boldsymbol{\theta}]_k$ and $[\breve{\boldsymbol{\theta}}]_k$ are distinct for $1\leq k\leq K$. Moreover, let $\mathbf{A}_v(\boldsymbol{\theta})$ be the steering matrix of a contiguous ULA with v elements located at $(0,\frac{\lambda}{2},\cdots,(v-1)\frac{\lambda}{2})$. Considering the fact that $\mathbf{A}_v(\boldsymbol{\theta})$ is a Vandermonde matrix, if $K\leq v-1$, it follows from Caratheodory-Fejer-Pisarenko decomposition [189] that

$$\mathbf{A}_v(\boldsymbol{\theta})\mathrm{diag}(\mathbf{p})\mathbf{A}_v^H(\boldsymbol{\theta})+\sigma^2\mathbf{I}_v\neq\mathbf{A}_v(\breve{\boldsymbol{\theta}})\mathrm{diag}(\breve{\mathbf{p}})\mathbf{A}_v^H(\breve{\boldsymbol{\theta}})+\breve{\sigma}^2\mathbf{I}_v \tag{B.5}$$

for any arbitrary values of $\mathbf{p}\in\mathbb{R}_{>0}^{K\times 1}$, $\breve{\mathbf{p}}\in\mathbb{R}_{>0}^{K\times 1}$, σ^2 and $\breve{\sigma}^2$. From [53, Eq. (113)], vectorizing both sides of (B.5) leads to

$$\mathbf{T}'\mathbf{A}_\vartheta(\boldsymbol{\theta})\mathbf{p}+\sigma^2\mathbf{T}'\mathbf{e}'\neq\mathbf{T}'\mathbf{A}_\vartheta(\breve{\boldsymbol{\theta}})\breve{\mathbf{p}}+\breve{\sigma}^2\mathbf{T}'\mathbf{e}' \tag{B.6}$$

where $\mathbf{A}_\vartheta(\boldsymbol{\theta})\in\mathbb{C}^{(2v-1)\times K}$ denotes the steering matrix corresponding to the contiguous ULA segment of the difference co-array, $\mathbf{T}'\in\{0,1\}^{v^2\times 2v-1}$ is a selection matrix defined in [53, Eq. (114)] and $\mathbf{e}'\in\{0,1\}^{(2v-1)\times 1}$ is a column vector with $[\mathbf{e}']_i=\delta[i-v]$. Considering

$\mathbf{T}'$ is full column rank [53], multiplying both sides of (B.6) by $\mathbf{T}'^{\dagger}$ and then moving all the terms to one side of the equation yields

$$\mathbf{A}_{\vartheta}(\boldsymbol{\theta})\mathbf{p} - \mathbf{A}_{\vartheta}(\breve{\boldsymbol{\theta}})\breve{\mathbf{p}} + (\sigma^2 - \breve{\sigma}^2)\mathbf{e}' \neq \mathbf{0}. \tag{B.7}$$

It follows from $\breve{\boldsymbol{\theta}} \neq \boldsymbol{\theta}_0$ that $\breve{\boldsymbol{\theta}}$ could differs from $\boldsymbol{\theta}_0$ at q DoAs for some integer $q \in [1, K]$. Noting this fact, (B.7) is simplified to

$$\begin{bmatrix} \mathbf{A}_{\vartheta}(\boldsymbol{\theta}) & \mathbf{A}_{\vartheta}(\ddot{\boldsymbol{\theta}}) & \mathbf{e}' \end{bmatrix} \begin{bmatrix} \mathbf{p} - \breve{\mathbf{p}} \odot \boldsymbol{\varepsilon} \\ -\ddot{\mathbf{p}} \\ \sigma^2 - \breve{\sigma}^2 \end{bmatrix} \neq \mathbf{0}, \tag{B.8}$$

where $\ddot{\boldsymbol{\theta}} \in [-\pi, \pi]^{q\times 1}$ consists of those elements of $\breve{\boldsymbol{\theta}}$ which do not intersect with those in $\boldsymbol{\theta}$, $\ddot{\mathbf{p}} \in \mathbb{R}_{>0}^{q\times 1}$ contains those elements of $\breve{\mathbf{p}}$ corresponding to $\ddot{\boldsymbol{\theta}}$ and

$$[\boldsymbol{\varepsilon}]_i = \begin{cases} 1, & [\boldsymbol{\theta}]_i = [\breve{\boldsymbol{\theta}}]_i, \\ 0, & \text{otherwise.} \end{cases} \tag{B.9}$$

Considering the fact that $\begin{bmatrix} \mathbf{A}_{\vartheta}(\boldsymbol{\theta}) & \mathbf{A}_{\vartheta}(\breve{\boldsymbol{\theta}}) & \mathbf{e}' \end{bmatrix} \in \mathbb{C}^{(2v-1)\times(2K+1)}$ is a sub-matrix of $\begin{bmatrix} \mathbf{A}_d(\boldsymbol{\theta}) & \mathbf{A}_d(\ddot{\boldsymbol{\theta}}) & \mathbf{e} \end{bmatrix} \in \mathbb{C}^{(2D-1)\times(2K+1)}$, obtained from $2v-1$ rows of $\begin{bmatrix} \mathbf{A}_d(\boldsymbol{\theta}) & \mathbf{A}_d(\ddot{\boldsymbol{\theta}}) & \mathbf{e} \end{bmatrix}$, it follows from (B.8) that

$$\begin{bmatrix} \mathbf{A}_d(\boldsymbol{\theta}) & \mathbf{A}_d(\ddot{\boldsymbol{\theta}}) & \mathbf{e} \end{bmatrix} \begin{bmatrix} \mathbf{p} - \breve{\mathbf{p}} \odot \boldsymbol{\varepsilon} \\ -\ddot{\mathbf{p}} \\ \sigma^2 - \breve{\sigma}^2 \end{bmatrix} \neq \mathbf{0}, \tag{B.10}$$

$$\Rightarrow \mathbf{A}_d(\boldsymbol{\theta})\mathbf{p} - \mathbf{A}_d(\breve{\boldsymbol{\theta}})\breve{\mathbf{p}} + (\sigma^2 - \breve{\sigma}^2)\mathbf{e} \neq \mathbf{0}. \tag{B.11}$$

Multiplying (B.11) by $\mathbf{J}$ and exploiting (4.3) and (4.4), after some algebraic manipulations, we obtain

$$\begin{aligned} &\operatorname{vec}(\mathbf{A}(\boldsymbol{\theta})\operatorname{diag}(\mathbf{p})\mathbf{A}^H(\boldsymbol{\theta}) + \sigma^2\mathbf{I}_M) \\ &\neq \operatorname{vec}(\mathbf{A}(\breve{\boldsymbol{\theta}})\operatorname{diag}(\breve{\mathbf{p}})\mathbf{A}^H(\breve{\boldsymbol{\theta}}) + \breve{\sigma}^2\mathbf{I}_M), \end{aligned} \tag{B.12}$$

which in turn implies that

$$\mathbf{A}(\boldsymbol{\theta})\operatorname{diag}(\mathbf{p})\mathbf{A}^H(\boldsymbol{\theta}) + \sigma^2\mathbf{I}_M \neq \mathbf{A}(\breve{\boldsymbol{\theta}})\operatorname{diag}(\breve{\mathbf{p}})\mathbf{A}^H(\breve{\boldsymbol{\theta}}) + \breve{\sigma}^2\mathbf{I}_M, \tag{B.13}$$

for all $\boldsymbol{\theta} \neq \breve{\boldsymbol{\theta}} \in [-\pi/2, \pi/2]^{K\times 1}$, $\mathbf{p} \in \mathbb{R}_{>0}^{K\times 1}$, $\breve{\mathbf{p}} \in \mathbb{R}_{>0}^{K\times 1}$, σ^2 and $\breve{\sigma}^2$. Considering $\mathbf{y}(0), \mathbf{y}(1), \cdots, \mathbf{y}(N-1)$ are independent and identically distributed with $\mathbf{y}(t) \sim \mathcal{CN}(\mathbf{0}, \mathbf{R})$, it follows from (B.13) that $f(\mathbf{Y} \mid \boldsymbol{\theta}_0, \mathbf{p}, \sigma^2) \neq f(\mathbf{Y} \mid \breve{\boldsymbol{\theta}}, \breve{\mathbf{p}}, \breve{\sigma}^2)$ for any arbitrary values of $\boldsymbol{\theta} \neq \breve{\boldsymbol{\theta}} \in [-\pi/2, \pi/2]^{K\times 1}$, $\mathbf{p} \in \mathbb{R}_{>0}^{K\times 1}$, $\breve{\mathbf{p}} \in \mathbb{R}_{>0}^{K\times 1}$, σ^2 and $\breve{\sigma}^2$ if $K \leq v-1$. Now, from Theorem 4.1, we conclude that $f(\mathbf{X} \mid \boldsymbol{\theta}_0, \mathbf{p}, \sigma^2) \neq f(\mathbf{X} \mid \breve{\boldsymbol{\theta}}, \breve{\mathbf{p}}, \breve{\sigma}^2)$. This completes the proof of **S1**.

We now prove **S2**. We know from Lemma 4.1 that the FIM is singular for any value of $\boldsymbol{\theta} \in [-\pi/2, \pi/2]^{K\times 1}$ if $K \geq D$. This means that the problem is not even locally indentifiable at any $\boldsymbol{\theta}$ [190]. Since the local identifiablity is a necessary condition for the identifiablity any particular point, the problem is not identifiable for any $\boldsymbol{\theta}$.

B.3 Proof of Lemma 4.1

Let $\mathbf{R}_{\mathbf{x}}^r$ and $\overline{\mathbf{R}}^r$ denote the equivalent real representation for $\mathbf{R}_{\mathbf{x}}$ and $\overline{\mathbf{R}}$, respectively, given as

$$\mathbf{R}_{\mathbf{x}}^r = \begin{bmatrix} \Re\{\mathbf{R}_{\mathbf{x}}\} & -\Im\{\mathbf{R}_{\mathbf{x}}\} \\ \Im\{\mathbf{R}_{\mathbf{x}}\} & \Re\{\mathbf{R}_{\mathbf{x}}\} \end{bmatrix}, \quad \overline{\mathbf{R}}^r = \begin{bmatrix} \Re\{\overline{\mathbf{R}}\} & -\Im\{\overline{\mathbf{R}}\} \\ \Im\{\overline{\mathbf{R}}\} & \Re\{\overline{\mathbf{R}}\}. \end{bmatrix} \tag{B.14}$$

Making use of (4.7) and Taylor expansion of arcsine function, we have

$$\begin{aligned} \mathbf{R}_{\mathbf{x}}^r =& \frac{2}{\pi}\arcsin(\overline{\mathbf{R}}^r) = \overline{\mathbf{R}}^r + \frac{1}{6}\overline{\mathbf{R}}^r \odot \overline{\mathbf{R}}^r \odot \overline{\mathbf{R}}^r \\ &+ \frac{3}{40}\overline{\mathbf{R}}^r \odot \overline{\mathbf{R}}^r \odot \overline{\mathbf{R}}^r \odot \overline{\mathbf{R}}^r \odot \overline{\mathbf{R}}^r + \cdots \\ =& \sum_{n=0}^{\infty} \frac{(2n)!}{(2^n n!)^2(2n+1)} \underbrace{\overline{\mathbf{R}}^r \odot \overline{\mathbf{R}}^r \odot \cdots \odot \overline{\mathbf{R}}^r}_{2n+1 \text{ times}}. \end{aligned} \tag{B.15}$$

It is clear from (4.8) that $\overline{\mathbf{R}}$ is positive definite, and so is $\overline{\mathbf{R}}^r$. Further, it follows from the Schur product theorem [191, Theorem 3.1], which establishes that the Hadamard product of two positive-definite matrices is also a positive-definite matrix, that the $2n+1$ times Hadamard products of $\overline{\mathbf{R}}^r$ by itself is also positive definite for any integer $n \in [0, \infty)$. Hence, it follows from (B.15) that $\mathbf{R}_{\mathbf{x}}^r$ is obtained from a weighted sum of positive definite matrices, and thus it is positive definite. Evidently, $\mathbf{R}_{\mathbf{x}}$ is also positive definite. This in turn indicates non-singularity of $(\mathbf{R}_{\mathbf{x}}^{-T} \otimes \mathbf{R}_{\mathbf{x}}^{-1})$. Hence, since $\mathbf{J}$ is also full column rank [52], we easily conclude that $\mathbf{J}^H(\mathbf{R}_{\mathbf{x}}^{-T} \otimes \mathbf{R}_{\mathbf{x}}^{-1})\mathbf{J}$ is full rank. This implies that $\mathcal{I}_w(\boldsymbol{\varrho})$ is non-singular if and only if $\begin{bmatrix} \mathbf{G} & \mathbf{V} \end{bmatrix} \in \mathbb{R}^{(2D-1)\times 2K}$ is full column rank. In other words, $\mathcal{I}_w(\boldsymbol{\varrho})$ is non-singular if and only if

$$\begin{bmatrix} \mathbf{G} & \mathbf{V} \end{bmatrix} \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_1 \end{bmatrix} \neq \mathbf{0}, \tag{B.16}$$

for any arbitrary non-zero $\mathbf{c} = [\mathbf{c}_1^T, \mathbf{c}_2^T]^T \in \mathbb{C}^{2K\times 1}$. Inserting (4.16) and (4.17) into (B.16) leads to

$$\begin{bmatrix} \boldsymbol{\Delta} & F \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{c}}_1 \\ \mathbf{c}_2 \end{bmatrix} \neq \mathbf{0}. \tag{B.17}$$

where $\widetilde{\mathbf{c}}_1 = j\pi\boldsymbol{\Phi}(\boldsymbol{\theta})\text{diag}(\overline{\mathbf{p}})\mathbf{c}_1$. This completes the proof.

B.4 Proof of Theorem 4.3

We know from Appendix B.3 that $\mathbf{M} = \mathbf{J}^H(\mathbf{R}_{\mathbf{x}}^{-T} \otimes \mathbf{R}_{\mathbf{x}}^{-1})\mathbf{J}$ is positive-definite. Hence, (4.15) can be rewritten as

$$\begin{aligned}\mathcal{I}_w(\boldsymbol{\varrho}) &= N \begin{bmatrix} \mathbf{G}^H\mathbf{M}^{\frac{1}{2}} \\ \mathbf{V}^H\mathbf{M}^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \mathbf{M}^{\frac{1}{2}}\mathbf{G} & \mathbf{M}^{\frac{1}{2}}\mathbf{V} \end{bmatrix} \\ &= N \begin{bmatrix} \mathbf{G}^H\mathbf{M}\mathbf{G} & \mathbf{G}^H\mathbf{M}\mathbf{V} \\ \mathbf{V}^H\mathbf{M}\mathbf{G} & \mathbf{V}^H\mathbf{M}\mathbf{V} \end{bmatrix}. \end{aligned} \tag{B.18}$$

The $CRB_w(\boldsymbol{\theta})$ is then obtained by block-wise inversion as follows:

$$\begin{aligned} CRB_w(\boldsymbol{\theta}) &= \frac{1}{N}\left(\mathbf{G}^H\mathbf{M}\mathbf{G} - \mathbf{G}^H\mathbf{M}\mathbf{V}\left(\mathbf{V}^H\mathbf{M}\mathbf{V}\right)^{-1}\mathbf{V}^H\mathbf{M}\mathbf{G}\right)^{-1} \\ &= \frac{1}{N}\left(\mathbf{G}^H\mathbf{M}\mathbf{G} - \mathbf{G}^H\mathbf{M}^{\frac{1}{2}}\Pi_{\mathbf{M}^{\frac{1}{2}}\mathbf{V}}\mathbf{M}^{\frac{1}{2}}\mathbf{G}\right)^{-1} \\ &= \frac{1}{N}\left(\mathbf{G}^H\mathbf{M}^{\frac{1}{2}}\Pi^{\perp}_{\mathbf{M}^{\frac{1}{2}}\mathbf{V}}\mathbf{M}^{\frac{1}{2}}\mathbf{G}\right)^{-1}. \end{aligned} \tag{B.19}$$

The facts that $\mathbf{G} = j\pi\text{diag}(\mathbf{d})\boldsymbol{\Omega}\boldsymbol{\Phi}(\boldsymbol{\theta})\text{diag}(\overline{\mathbf{p}})$ and $\mathbf{R}_{\mathbf{x}} = \frac{2}{\pi}\text{arcsine}(\overline{\mathbf{R}})$ will lead to (4.20). In addition, It follows from $\mathcal{I}(\varrho) \succeq \mathcal{I}_w(\boldsymbol{\varrho})$ that $CRB(\boldsymbol{\theta}) \preceq CRB_w(\boldsymbol{\theta})$.

B.5 Proof of Theorem 4.4

Recalling $\overline{p}_k = \frac{p_k}{\sigma^2 + \sum_{k=1}^K p_k}$ and assuming that all sources have equal power p, we conclude that

$$\lim_{SNR\to\infty} \overline{p}_k = \lim_{SNR\to\infty} \frac{SNR}{K \times SNR + 1} = \frac{1}{K}. \tag{B.20}$$

Making use of (B.20), it can be readily shown that

$$\lim_{SNR\to\infty} \overline{\mathbf{R}} = \frac{1}{K}\mathbf{A}(\boldsymbol{\theta})\mathbf{A}^H(\boldsymbol{\theta}) + (1 - \frac{1}{K})\mathbf{I}_M. \tag{B.21}$$

The above equation implies that $\lim_{SNR\to\infty} \overline{\mathbf{R}}$ is a positive-definite matrix independent of the SNR. Further, it follows from (B.21) that

$$\lim_{SNR\to\infty} \text{diag}(\overline{\mathbf{p}}) = \frac{1}{K}\mathbf{I}_K, \tag{B.22}$$

$$\lim_{SNR\to\infty} \mathbf{h} = \left[\frac{1}{\sqrt{1 - \frac{|\Re\{\sum_{k=1}^K e^{-j\pi\sin\theta_k \ell_{D-1}}\}|^2}{K^2}}}, \cdots, 1, \cdots, \frac{1}{\sqrt{1 - \frac{|\Re\{\sum_{k=1}^K e^{j\pi\sin\theta_k \ell_{D-1}}\}|^2}{K^2}}}\right]^T, \tag{B.23}$$

$$\lim_{SNR\to\infty} \overline{\mathbf{h}} = \left[\frac{1}{\sqrt{1 - \frac{|\Im\{\sum_{k=1}^K e^{-j\pi\sin\theta_k \ell_{D-1}}\}|^2}{K^2}}}, \cdots, 1, \cdots, \frac{1}{\sqrt{1 - \frac{|\Im\{\sum_{k=1}^K e^{j\pi\sin\theta_k \ell_{D-1}}\}|^2}{K^2}}}\right]^T. \tag{B.24}$$

Substituting (B.22), (B.23) and (B.24) back into (4.16) and (4.17) indicates that $\lim_{SNR\to\infty} \begin{bmatrix}\mathbf{G} & \mathbf{V}\end{bmatrix}$ is a full column rank matrix independent of the SNR. Hence, recalling (4.15), we can conclude that $\lim_{SNR\to\infty} \mathcal{I}_w(\varrho)$ is positive-definite and independent of the SNR. This in turn implies that $\lim_{SNR\to\infty} CRB_w(\boldsymbol{\theta})$, which is Schur complement of $\lim_{SNR\to\infty} \mathcal{I}_w(\varrho)$, is also positive-definite and independent of the SNR. This completes the proof.

B.6 Proof of Lemma 4.2

We start with showing that $\boldsymbol{\Psi}$ is full rank. Making use of relations $\det(\begin{bmatrix}\mathbf{C}_1 & \mathbf{C}_2\\ \mathbf{C}_3 & \mathbf{C}_4\end{bmatrix}) = \det(\mathbf{C}_1)\det(\mathbf{C}_4 - \mathbf{C}_3\mathbf{C}_1^{-1}\mathbf{C}_2)$, we obtain

$$\det(\boldsymbol{\Psi}) = \det(\mathbf{I}_{D-1})\det(2j\mathbf{I}_{D-1}) = (2j)^{D-1} \neq 0, \tag{B.25}$$

which implies full rankness of $\boldsymbol{\Psi}$.

Next, we proceed with proving that $\overline{\mathbf{J}}$ is full rank. Let $\ddot{\mathbf{J}}$ denote the matrix obtained after removing the D-th column from $\mathbf{J}$. $\ddot{\mathbf{J}}$ is full column rank since its columns are a sub-set of the columns of the full-column-rank matrix $\mathbf{J}$ [52]. Further, for $1 \leq i \leq M$, it is readily confirmed that the $((i-1)M+1)$-th row of $\text{vec}(\mathbf{L}_n)$ as well as $\text{vec}(\mathbf{L}_n^T)$ equals the i-th diagonal element of $\mathbf{L}_n$, which is obviously zero for $n \neq 0$ according to the definition given after (4.5). Given (4.5), this in turn implies that the rows of $\ddot{\mathbf{J}}$ with indices $(i-1)M+1$, for all $1 \leq i \leq M$, are zero vectors. As a result, the matrix obtained by removing these rows from $\ddot{\mathbf{J}}$, i.e., $\overline{\mathbf{J}}$, has the same column rank as $\ddot{\mathbf{J}}$. This completes the proof.

Finally, we show that $\mathbf{F}$ is full rank. It follows from the fact that $\overline{\mathbf{e}}_p^T\overline{\mathbf{e}}_q = 0$ for $p \neq q$ that

$$\begin{aligned}&(\overline{\mathbf{e}}_i^T \otimes \overline{\mathbf{e}}_j^T \pm \overline{\mathbf{e}}_j^T \otimes \overline{\mathbf{e}}_i^T)(\overline{\mathbf{e}}_p \otimes \overline{\mathbf{e}}_q \pm \overline{\mathbf{e}}_q \otimes \overline{\mathbf{e}}_p) = \overline{\mathbf{e}}_i^T\overline{\mathbf{e}}_p \otimes \overline{\mathbf{e}}_j^T\overline{\mathbf{e}}_q\\ &\pm \overline{\mathbf{e}}_i^T\overline{\mathbf{e}}_q \otimes \overline{\mathbf{e}}_j^T\overline{\mathbf{e}}_p \pm \overline{\mathbf{e}}_j^T\overline{\mathbf{e}}_p \otimes \overline{\mathbf{e}}_i^T\overline{\mathbf{e}}_q \pm \overline{\mathbf{e}}_j^T\overline{\mathbf{e}}_q \otimes \overline{\mathbf{e}}_i^T\overline{\mathbf{e}}_p = 0.\end{aligned} \tag{B.26}$$

for $1 \leq i < j \leq M$ and $1 \leq p < q \leq M$ when either p or q differs from i and j. In addition, in case $i = p$ and $j = q$, we have

$$\begin{aligned}&(\overline{\mathbf{e}}_i^T \otimes \overline{\mathbf{e}}_j^T + \overline{\mathbf{e}}_j^T \otimes \overline{\mathbf{e}}_i^T)(\overline{\mathbf{e}}_i \otimes \overline{\mathbf{e}}_j - \overline{\mathbf{e}}_j \otimes \overline{\mathbf{e}}_i) = \overline{\mathbf{e}}_i^T\overline{\mathbf{e}}_i \otimes \overline{\mathbf{e}}_j^T\overline{\mathbf{e}}_j\\ &- \overline{\mathbf{e}}_i^T\overline{\mathbf{e}}_j \otimes \overline{\mathbf{e}}_j^T\overline{\mathbf{e}}_i + \overline{\mathbf{e}}_j^T\overline{\mathbf{e}}_i \otimes \overline{\mathbf{e}}_i^T\overline{\mathbf{e}}_j - \overline{\mathbf{e}}_j^T\overline{\mathbf{e}}_j \otimes \overline{\mathbf{e}}_i^T\overline{\mathbf{e}}_i = 0.\end{aligned} \tag{B.27}$$

It is also observed that, for $1 \leq i \leq M$ and $1 \leq p < q \leq M$, the $((i-1)M+1)$-th element of $\overline{\mathbf{e}}_p^T \otimes \overline{\mathbf{e}}_q^T \pm \overline{\mathbf{e}}_p^T \otimes \overline{\mathbf{e}}_q^T$ is equal to the i-th diagonal element of $\overline{\mathbf{e}}_p\overline{\mathbf{e}}_q^T \pm \overline{\mathbf{e}}_p\overline{\mathbf{e}}_q^T$, which is obviously zero for $p \neq q$. Consequently, the row vectors obtained by removing the elements with indices $(i-1)M+1$ for all $1 \leq i \leq M$ from $\overline{\mathbf{e}}_p^T \otimes \overline{\mathbf{e}}_q^T + \overline{\mathbf{e}}_q^T \otimes \overline{\mathbf{e}}_p^T$ and $\overline{\mathbf{e}}_p^T \otimes \overline{\mathbf{e}}_q^T - \overline{\mathbf{e}}_q^T \otimes \overline{\mathbf{e}}_p^T$ will be still orthogonal with each other. Hence, it is deduced that the square matrix $\mathbf{F}$ has orthogonal rows, thereby being full rank.

B.7 Proof of Lemma 4.3

Let define $E(\theta) = \mathbf{a}_v^H(\theta)\widehat{\mathbf{U}}_n\widehat{\mathbf{U}}_n^H\mathbf{a}_v(\theta)$ and $\breve{E}(\theta) = \mathbf{a}_v^H(\boldsymbol{\theta})\mathbf{U}_n\mathbf{U}_n^H\mathbf{a}_v(\boldsymbol{\theta})$ where $\widehat{\mathbf{U}}_n$ and $\mathbf{U}_n$ consist of, respectively, the eigenvectors of $\widehat{\overline{\mathbf{R}}}_v$ and $\mathbf{A}_v(\boldsymbol{\theta})\text{diag}(\overline{\mathbf{p}})\mathbf{A}_v^H(\boldsymbol{\theta}) + \overline{\sigma}^2\mathbf{I}_v$ corresponding to their $v-K$ smallest eigenvalues with $K \leq v-1$. We know that the elements of $\hat{\boldsymbol{\theta}}$ are equal to the minimizers of $E(\theta)$. Defining $E_n = \sup_\theta |E(\theta) - \breve{E}(\theta)|$, we have

$$\begin{aligned} E_n &= \sup_\theta \left|\mathbf{a}_v^H(\boldsymbol{\theta})(\widehat{\mathbf{U}}_n\widehat{\mathbf{U}}_n^H - \mathbf{U}_n\mathbf{U}_n^H)\mathbf{a}_v(\boldsymbol{\theta})\right| \\ &= \sup_\theta \left|\left(\mathbf{a}_v^T(\boldsymbol{\theta}) \otimes \mathbf{a}_v^H(\boldsymbol{\theta})\right)\text{vec}(\widehat{\mathbf{U}}_n\widehat{\mathbf{U}}_n^H - \mathbf{U}_n\mathbf{U}_n^H)\right| \\ &\leq \|\mathbf{a}_v^T(\boldsymbol{\theta}) \otimes \mathbf{a}_v^H(\boldsymbol{\theta})\|_2 \|\text{vec}(\widehat{\mathbf{U}}_n\widehat{\mathbf{U}}_n^H - \mathbf{U}_n\mathbf{U}_n^H)\|_2 \\ &= v^2\|\text{vec}(\widehat{\mathbf{U}}_n\widehat{\mathbf{U}}_n^H - \mathbf{U}_n\mathbf{U}_n^H)\|_2. \end{aligned} \tag{B.28}$$

It follows from (4.49) that $\lim_{N\to\infty}\widehat{\mathbf{U}}_n\widehat{\mathbf{U}}_n^H = \mathbf{U}_n\mathbf{U}_n^H$. Hence, $E_n \to 0$ as $N \to \infty$. This implies that $E(\theta)$ converges uniformly to $\breve{E}(\theta)$ as $N \to \infty$. Thus, the minimizers of $E(\theta)$, i.e., the elements of $\hat{\boldsymbol{\theta}}$, converge to the minimizers of $\breve{E}(\theta)$, i.e., $\theta_1, \theta_2, \cdots, \theta_K$, as $N \to \infty$. This complete the proof.

B.8 Proof of Theorem 4.5

Considering the consistency of $\hat{\boldsymbol{\theta}}$ and following the same arguments as in [38, App. B], for sufficiently large N, the asymptotic estimation error expression for EOCAB-MUSIC is given by

$$\hat{\theta}_k - \theta_k = -\frac{\Re\{\mathbf{z}_k^T\mathbf{T}\mathbf{J}^\dagger\Delta\bar{\mathbf{r}}\}}{\pi\overline{p}_k q_k \cos(\theta_k)}, \tag{B.29}$$

where $\Delta\bar{\mathbf{r}} = \widehat{\bar{\mathbf{r}}} - \bar{\mathbf{r}}$ and $\mathbf{T} = \begin{bmatrix}\mathbf{T}_v^T & \mathbf{T}_{v-1}^T & \cdots & \mathbf{T}_1^T\end{bmatrix}^T \in \mathbb{C}^{v^2\times(2D-1)}$. From (B.29), the asymptotic (as $N \to \infty$) covariance between the DoA estimation errors is given by

$$\begin{aligned} \mathcal{E}_{\theta_{k_1},\theta_{k_2}} &= \mathbb{E}\{(\hat{\theta}_{k_1} - \theta_{k_1})(\hat{\theta}_{k_2} - \theta_{k_2})\} \\ &= \frac{\mathbb{E}\left\{\Re\{\mathbf{z}_{k_1}^T\mathbf{T}\mathbf{J}^\dagger\Delta\bar{\mathbf{r}}\}\Re\{\mathbf{z}_{k_2}^T\mathbf{T}\mathbf{J}^\dagger\Delta\bar{\mathbf{r}}\}\right\}}{\pi^2\overline{p}_{k_1}\overline{p}_{k_2}q_{k_1}q_{k_2}\cos(\theta_{k_1})\cos(\theta_{k_2})}. \end{aligned} \tag{B.30}$$

Making use of the identity $\Re\{\mathbf{c}_1^H\mathbf{c}_2\}\Re\{\mathbf{c}_3^H\mathbf{c}_2\} = \frac{1}{2}\Re\{\mathbf{c}_1^H\mathbf{c}_2\mathbf{c}_2^H\mathbf{c}_3 + \mathbf{c}_1^H\mathbf{c}_2\mathbf{c}_2^T\mathbf{c}_3^*\}$, we obtain

$$\begin{aligned} &\mathbb{E}\left\{\Re\{\mathbf{z}_{k_1}^T\mathbf{T}\mathbf{J}^\dagger\Delta\bar{\mathbf{r}}\}\Re\{\mathbf{z}_{k_2}^T\mathbf{T}\mathbf{J}^\dagger\Delta\bar{\mathbf{r}}\}\right\} = \\ &\frac{1}{2}\mathbb{E}\{\Re\{\mathbf{z}_{k_1}^T\mathbf{T}\mathbf{J}^\dagger\Delta\bar{\mathbf{r}}\Delta\bar{\mathbf{r}}^H\mathbf{J}^{\dagger H}\mathbf{T}^H\mathbf{z}_{k_2}^* + \Re\{\mathbf{z}_{k_1}^T\mathbf{T}\mathbf{J}^\dagger\Delta\bar{\mathbf{r}}\Delta\bar{\mathbf{r}}^T\mathbf{J}^{\dagger H}\mathbf{T}^H\mathbf{z}_{k_2}\}\}. \end{aligned} \tag{B.31}$$

The matrix $\text{mat}_{M,M}(\mathbf{J}^{\dagger H}\mathbf{T}^H\mathbf{z}_k)$ is Hermitian [38, Lemma 6], thereby

$$\mathbf{J}^{\dagger H}\mathbf{T}^H\mathbf{z}_k^* = \mathbf{K}_M\mathbf{J}^{\dagger H}\mathbf{T}^H\mathbf{z}_k. \tag{B.32}$$

where $\mathbf{K}_M \in \{0,1\}^{M^2\times M^2}$ is the commutation matrix defined as $\text{vec}(\mathbf{C}^T) = \mathbf{K}_M\text{vec}(\mathbf{C})$ for any arbitrary matrix $\mathbf{C}$ [186]. In addition, since $\overline{\mathbf{R}}^H = \overline{\mathbf{R}}$, we have

$$\Delta\bar{\mathbf{r}}^T = \Delta\bar{\mathbf{r}}^H\mathbf{K}_M^H. \tag{B.33}$$

Inserting (B.32) and (B.33) into (B.31) and using the fact that $\mathbf{K}_M = \mathbf{K}_M^H = \mathbf{K}_M^{-1}$, we obtain

$$\begin{aligned}&\mathbb{E}\left\{\Re\{\mathbf{z}_{k_1}^T\mathbf{T}\mathbf{J}^\dagger\Delta\bar{\mathbf{r}}\}\Re\{\mathbf{z}_{k_2}^T\mathbf{T}\mathbf{J}^\dagger\Delta\bar{\mathbf{r}}\}\right\}\\&= \mathbb{E}\{\Re\{\mathbf{z}_{k_1}^T\mathbf{T}\mathbf{J}^\dagger\Delta\bar{\mathbf{r}}\Delta\bar{\mathbf{r}}^H\mathbf{J}^{\dagger H}\mathbf{T}^H\mathbf{z}_{k_2}^*\}\}.\end{aligned} \tag{B.34}$$

Recalling (4.46) and (4.32), we have

$$\begin{aligned}\mathbf{T}\mathbf{J}^\dagger\Delta\bar{\mathbf{r}} &= \mathbf{T}\mathbf{J}^\dagger\mathbf{J}\begin{bmatrix}\mathbf{0} & \mathbf{I}_{D-1} & -j\mathbf{I}_{D-1}\\ 1 & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{D-1} & j\mathbf{I}_{D-1}\end{bmatrix}\begin{bmatrix}0\\ \widehat{\boldsymbol{\phi}}-\boldsymbol{\phi}\end{bmatrix}\\&= \mathbf{T}\begin{bmatrix}\mathbf{0} & \mathbf{I}_{D-1} & -j\mathbf{I}_{D-1}\\ 0 & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{D-1} & j\mathbf{I}_{D-1}\end{bmatrix}\begin{bmatrix}0\\ \widehat{\boldsymbol{\phi}}-\boldsymbol{\phi}\end{bmatrix}.\end{aligned} \tag{B.35}$$

Additionally, from [53, Eq. (114) and Eq. (116)], we know

$$\mathbf{T} = \left[\mathbf{0}_{v^2\times(D-v)}, \text{vec}(\overline{\mathbf{L}}_{v-1}^T), \cdots, \text{vec}(\overline{\mathbf{L}}_0), \cdots, \text{vec}(\overline{\mathbf{L}}_{v-1}), \mathbf{0}_{v^2\times(D-v)}\right], \tag{B.36}$$

where $[\overline{\mathbf{L}}_n]_{p,q} = \begin{cases}1, & \text{if } p-q=n,\\ 0, & \text{otherwise.}\end{cases}$. Substituting (B.36) into (B.35) yields

$$\mathbf{T}\mathbf{J}^\dagger\Delta\bar{\mathbf{r}} = \overline{\mathbf{T}}\boldsymbol{\Psi}(\widehat{\boldsymbol{\phi}}-\boldsymbol{\phi}), \tag{B.37}$$

where

$$\overline{\mathbf{T}} = \left[\mathbf{0}_{v^2\times(D-v)}, \text{vec}(\overline{\mathbf{L}}_{v-1}^T), \cdots, \text{vec}(\overline{\mathbf{L}^T}_{-1}), \text{vec}(\overline{\mathbf{L}}_1), \cdots, \text{vec}(\overline{\mathbf{L}}_{v-1}), \mathbf{0}_{v^2\times(D-v)}\right]. \tag{B.38}$$

Inserting (B.37) into (B.34) gives

$$\mathbb{E}\left\{\Re\{\mathbf{z}_{k_1}^T\mathbf{T}\mathbf{J}^\dagger\Delta\bar{\mathbf{r}}\}\Re\{\mathbf{z}_{k_2}^T\mathbf{T}\mathbf{J}^\dagger\Delta\bar{\mathbf{r}}\}\right\} = \Re\{\mathbf{z}_{k_1}^T\overline{\mathbf{T}}\mathbb{E}\{\boldsymbol{\Psi}(\widehat{\boldsymbol{\phi}}-\boldsymbol{\phi})(\widehat{\boldsymbol{\phi}}-\boldsymbol{\phi})^H\boldsymbol{\Psi}^H\}\overline{\mathbf{T}}^H\mathbf{z}_{k_2}^*\}\}. \tag{B.39}$$

As a result, for sufficiently large N, using a first-order perturbation expansion leads to

$$\begin{aligned}&\mathbb{E}\{\boldsymbol{\Psi}(\widehat{\boldsymbol{\phi}}-\boldsymbol{\phi})(\widehat{\boldsymbol{\phi}}-\boldsymbol{\phi})^H\boldsymbol{\Psi}^H\} \simeq\\&\left(\overline{\mathbf{J}}^H\mathbf{F}^H\text{diag}(\mathbf{b})\mathbf{F}^{-H}\boldsymbol{\Sigma}^{-1}\mathbf{F}^{-1}\text{diag}(\mathbf{b})\mathbf{F}\overline{\mathbf{J}}\right)^{-1}\\&\times\overline{\mathbf{J}}^H\mathbf{F}^H\text{diag}(\mathbf{b})\mathbf{F}^{-H}\boldsymbol{\Sigma}^{-1}\mathbf{F}^{-1}\text{diag}(\mathbf{b})\mathbf{F}\mathbb{E}\{\tilde{\ddot{\mathbf{r}}}\tilde{\ddot{\mathbf{r}}}^H\}\\&\times\mathbf{F}^H\text{diag}(\mathbf{b})\mathbf{F}^{-H}\boldsymbol{\Sigma}^{-1}\mathbf{F}^{-1}\text{diag}(\mathbf{b})\mathbf{F}\overline{\mathbf{J}}\\&\times\left(\overline{\mathbf{J}}^H\mathbf{F}^H\text{diag}(\mathbf{b})\mathbf{F}^{-H}\boldsymbol{\Sigma}^{-1}\mathbf{F}^{-1}\text{diag}(\mathbf{b})\mathbf{F}\overline{\mathbf{J}}\right)^{-1} - \boldsymbol{\Psi}\boldsymbol{\phi}\boldsymbol{\phi}^H\boldsymbol{\Psi}^H,\end{aligned} \tag{B.40}$$

where $\mathbf{\Sigma} = \mathbf{\Sigma}(\boldsymbol{\gamma})$ given in B.11 and $[\mathbf{b}]_n = \frac{1}{\sqrt{1-|[\boldsymbol{\gamma}]_n|^2}}$ for $1 \leq n \leq M^2 - M$. It remains to compute $\mathbb{E}\{\widetilde{\ddot{\mathbf{r}}}\widetilde{\ddot{\mathbf{r}}}^H\}$. Making use of the relation $\widetilde{\ddot{\mathbf{r}}} = \text{sine}(\frac{\pi}{2}\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}})$, we obtain

$$\begin{aligned}
\mathbb{E}\{[\widetilde{\ddot{\mathbf{r}}}]_p[\widetilde{\ddot{\mathbf{r}}}]_q^*\} = \frac{1}{4}\mathbb{E}\Bigg\{ & e^{\frac{j\pi\left(\Re\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_p\}-\Re\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_q\}\right)}{2}} + e^{-\frac{j\pi\left(\Re\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_p\}-\Re\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_q\}\right)}{2}} \\
& - e^{\frac{j\pi\left(\Re\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_p\}+\Re\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_q\}\right)}{2}} - e^{-\frac{j\pi\left(\Re\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_p\}-\Re\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_q\}\right)}{2}} \\
& + e^{\frac{j\pi\left(\Im\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_p\}-\Im\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_q\}\right)}{2}} + e^{-\frac{j\pi\left(\Im\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_p\}-\Im\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_q\}\right)}{2}} \\
& - e^{\frac{j\pi\left(\Im\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_p\}+\Im\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_q\}\right)}{2}} - e^{-\frac{j\pi\left(\Im\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_p\}+\Im\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_q\}\right)}{2}} \Bigg\} \\
& + \frac{j}{4}\mathbb{E}\Bigg\{ e^{\frac{j\pi\left(\Im\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_p\}-\Re\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_q\}\right)}{2}} + e^{\frac{-j\pi\left(\Im\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_p\}-\Re\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_q\}\right)}{2}} \\
& - e^{\frac{j\pi\left(\Im\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_p\}+\Re\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_q\}\right)}{2}} - e^{\frac{-j\pi\left(\Im\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_p\}+\Re\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_q\}\right)}{2}} \\
& - e^{\frac{j\pi\left(\Re\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_p\}-\Im\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_q\}\right)}{2}} - e^{-\frac{j\pi\left(\Re\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_p\}-\Im\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_q\}\right)}{2}} \\
& + e^{\frac{j\pi\left(\Re\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_p\}+\Im\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_q\}\right)}{2}} + e^{-\frac{j\pi\left(\Re\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_p\}+\Im\{[\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}}]_q\}\right)}{2}} \Bigg\}. \qquad \text{(B.41)}
\end{aligned}$$

Considering that $\widehat{\ddot{\mathbf{r}}}_{\mathbf{x}} \overset{D}{\to} \mathcal{CN}(\ddot{\mathbf{r}}_{\mathbf{x}}, \frac{4}{\pi^2 N}\mathbf{\Sigma})$, the expectations in (B.41) can be computed using the characteristic function of the Gaussian distribution as follows:

$$\begin{aligned}
\mathbb{E}\{[\widetilde{\ddot{\mathbf{r}}}]_p[\widetilde{\ddot{\mathbf{r}}}]_q^*\} = & \frac{e^{\frac{-[\mathbf{\Sigma}]_{p,p}-[\mathbf{\Sigma}]_{q,q}}{4N}}}{2}\Bigg[\cos\left(\frac{\pi}{2}\left[\Re\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\} - \Re\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_q\}\right]\right) e^{\frac{\Re\{[\mathbf{\Sigma}]_{p,q}\}}{2N}} \\
& - \cos\left(\frac{\pi}{2}\left[\Re\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\} + \Re\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_q\}\right]\right) e^{-\frac{\Re\{[\mathbf{\Sigma}]_{p,q}\}}{2N}} \\
& + \cos\left(\frac{\pi}{2}\left[\Im\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\} - \Im\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_q\}\right]\right) e^{\frac{\Re\{[\mathbf{\Sigma}]_{p,q}\}}{2N}} \\
& - \cos\left(\frac{\pi}{2}\left[\Im\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\} + \Im\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_q\}\right]\right) e^{-\frac{\Re\{[\mathbf{\Sigma}]_{p,q}\}}{2N}} \\
& + j\cos\left(\frac{\pi}{2}\left[\Im\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\} - \Re\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_q\}\right]\right) e^{\frac{\Im\{[\mathbf{\Sigma}]_{p,q}\}}{2N}} \\
& - j\cos\left(\frac{\pi}{2}\left[\Im\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\} + \Re\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_q\}\right]\right) e^{-\frac{\Im\{[\mathbf{\Sigma}]_{p,q}\}}{2N}} \\
& - j\cos\left(\frac{\pi}{2}\left[\Re\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\} - \Im\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_q\}\right]\right) e^{-\frac{\Im\{[\mathbf{\Sigma}]_{p,q}\}}{2N}} \\
& + j\cos\left(\frac{\pi}{2}\left[\Re\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\} + \Im\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_q\}\right]\right) e^{\frac{\Im\{[\mathbf{\Sigma}]_{p,q}\}}{2N}}\Bigg]. \qquad \text{(B.42)}
\end{aligned}$$

Exploiting the Taylor expansion of the exponential function, (B.42) can be approximated for sufficiently large N as

$$\begin{aligned}
\mathbb{E}\{[\widetilde{\ddot{\mathbf{r}}}]_p[\widetilde{\ddot{\mathbf{r}}}]_q^*\} \simeq & \frac{1}{2}\Bigg[\cos\left(\frac{\pi}{2}\left[\Re\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\}-\Re\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_q\}\right]\right)\left(1+\frac{\Re\{[\boldsymbol{\Sigma}]_{p,q}\}}{2N}\right)\\
&-\cos\left(\frac{\pi}{2}\left[\Re\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\}+\Re\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_q\}\right]\right)\left(1-\frac{\Re\{[\boldsymbol{\Sigma}]_{p,q}\}}{2N}\right)\\
&+\cos\left(\frac{\pi}{2}\left[\Im\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\}-\Im\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_q\}\right]\right)\left(1+\frac{\Re\{[\boldsymbol{\Sigma}]_{p,q}\}}{2N}\right)\\
&-\cos\left(\frac{\pi}{2}\left[\Im\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\}+\Im\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_q\}\right]\right)\left(1-\frac{\Re\{[\boldsymbol{\Sigma}]_{p,q}\}}{2N}\right)\\
&+j\cos\left(\frac{\pi}{2}\left[\Im\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\}-\Re\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_q\}\right]\right)\left(1+\frac{\Im\{[\boldsymbol{\Sigma}]_{p,q}\}}{2N}\right)\\
&-j\cos\left(\frac{\pi}{2}\left[\Im\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\}+\Re\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_q\}\right]\right)\left(1-\frac{\Im\{[\boldsymbol{\Sigma}]_{p,q}\}}{2N}\right)\\
&-j\cos\left(\frac{\pi}{2}\left[\Re\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\}-\Im\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_q\}\right]\right)\left(1-\frac{\Im\{[\boldsymbol{\Sigma}]_{p,q}\}}{2N}\right)\\
&+j\cos\left(\frac{\pi}{2}\left[\Re\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\}+\Im\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_q\}\right]\right)\left(1+\frac{\Im\{[\boldsymbol{\Sigma}]_{p,q}\}}{2N}\right)\Bigg]\\
&=\sin(\frac{\pi}{2}\Re\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\})\sin(\frac{\pi}{2}\Re\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\})\\
&+\sin(\frac{\pi}{2}\Im\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\})\sin(\frac{\pi}{2}\Im\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\})\\
&+j\sin(\frac{\pi}{2}\Im\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\})\sin(\frac{\pi}{2}\Re\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\})\\
&-j\sin(\frac{\pi}{2}\Re\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\})\sin(\frac{\pi}{2}\Im\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\})\\
&+\frac{\Re\{[\boldsymbol{\Sigma}]_{p,q}\}}{2N}\Bigg[\cos(\frac{\pi}{2}\Re\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\})\cos(\frac{\pi}{2}\Re\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\})\\
&+\cos(\frac{\pi}{2}\Im\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\})\cos(\frac{\pi}{2}\Im\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\})\Bigg]\\
&+j+\frac{\Im\{[\boldsymbol{\Sigma}]_{p,q}\}}{2N}\Bigg[\cos(\frac{\pi}{2}\Im\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\})\cos(\frac{\pi}{2}\Re\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\})\\
&+\cos(\frac{\pi}{2}\Re\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\})\cos(\frac{\pi}{2}\Im\{[\ddot{\mathbf{r}}_{\mathbf{x}}]_p\})\Bigg].
\end{aligned}\tag{B.43}$$

Consequently, it follows from (4.30) that

$$\begin{aligned}
[\boldsymbol{\Gamma}]_{p,q} =& \mathbb{E}\{[\widetilde{\ddot{\mathbf{r}}}]_p[\widetilde{\ddot{\mathbf{r}}}]_q^*\} \simeq [\ddot{\mathbf{r}}]_p[\ddot{\mathbf{r}}]_q^*\\
&+\frac{1}{2N}\Bigg(\sqrt{1-[\Re\{[\ddot{\mathbf{r}}]_p\}]^2}\times\sqrt{1-[\Re\{[\ddot{\mathbf{r}}]_q\}]^2}\\
&+\sqrt{1-[\Im\{[\ddot{\mathbf{r}}]_p\}]^2}\times\sqrt{1-[\Im\{[\ddot{\mathbf{r}}]_q\}]^2}\Bigg)\Re\{[\boldsymbol{\Sigma}]_{p,q}\}\\
&+\frac{j}{2N}\Bigg(\sqrt{1-[\Im\{[\ddot{\mathbf{r}}]_p\}]^2}\times\sqrt{1-[\Re\{[\ddot{\mathbf{r}}]_q\}]^2}\\
&+\sqrt{1-[\Re\{[\ddot{\mathbf{r}}]_p\}]^2}\times\sqrt{1-[\Im\{[\ddot{\mathbf{r}}]_q\}]^2}\Bigg)\Im\{[\boldsymbol{\Sigma}]_{p,q}\}.
\end{aligned}\tag{B.44}$$

Inserting (B.44) into (B.40) and making use of (4.33) yields

$$\begin{aligned}
&\mathbb{E}\{\boldsymbol{\Psi}(\widehat{\boldsymbol{\phi}}-\boldsymbol{\phi})(\widehat{\boldsymbol{\phi}}-\boldsymbol{\phi})^H\boldsymbol{\Psi}^H\} \simeq \\
&\left(\overline{\mathbf{J}}^H\mathbf{F}^H\mathrm{diag}(\mathbf{b})\mathbf{F}^{-H}\boldsymbol{\Sigma}^{-1}\mathbf{F}^{-1}\mathrm{diag}(\mathbf{b})\mathbf{F}\overline{\mathbf{J}}\right)^{-1} \\
&\times \overline{\mathbf{J}}^H\mathbf{F}^H\mathrm{diag}(\mathbf{b})\mathbf{F}^{-H}\boldsymbol{\Sigma}^{-1}\mathbf{F}^{-1}\mathrm{diag}(\mathbf{b})\mathbf{F}\boldsymbol{\Gamma} \\
&\mathbf{F}^H\mathrm{diag}(\mathbf{b})\mathbf{F}^{-H}\boldsymbol{\Sigma}^{-1}\mathbf{F}^{-1}\mathrm{diag}(\mathbf{b})\mathbf{F}\overline{\mathbf{J}} \\
&\times \left(\overline{\mathbf{J}}^H\mathbf{F}^H\mathrm{diag}(\mathbf{b})\mathbf{F}^{-H}\boldsymbol{\Sigma}^{-1}\mathbf{F}^{-1}\mathrm{diag}(\mathbf{b})\mathbf{F}\overline{\mathbf{J}}\right)^{-1},
\end{aligned} \tag{B.45}$$

Finally, substituting (B.45) into (B.39) and considering that $\overline{p}_k = \frac{p_k}{\sigma^2+\sum_{k=1}^{K} p_k}$ concludes the proof of Theorem 4.5.

B.9 Proof of Corollary 4.2

OCAB-MUSIC employs $\widetilde{\overline{\mathbf{r}}} = \mathrm{vec}(\widetilde{\overline{\mathbf{R}}})$ instead of $\widehat{\overline{\mathbf{r}}}$. Hence, its asymptotic estimation error is obtained by replacing $\Delta\overline{\mathbf{r}}$ with $\widetilde{\overline{\mathbf{r}}} - \overline{\mathbf{r}}$ in (B.29). Following the same steps from (B.30) to (B.34), the covariance of the asymptotic distribution (as $N \to \infty$) of the DoA estimation errors for OCAB-MUSIC is obtained as

$$\mathcal{E}_{\theta_{k_1},\theta_{k_2}} = \Re\{\mathbf{z}_{k_1}^T\mathbf{T}\mathbf{J}^{\dagger}\mathbb{E}\{(\widetilde{\overline{\mathbf{r}}}-\overline{\mathbf{r}})(\widetilde{\overline{\mathbf{r}}}-\overline{\mathbf{r}})^H\}\mathbf{J}^{\dagger H}\mathbf{T}^H\mathbf{z}_{k_2}^*\}. \tag{B.46}$$

Considering the fact that the diagonal elements of $\widetilde{\overline{\mathbf{R}}}$ and $\overline{\mathbf{R}}$ are equal to one, (B.46) is simplified as

$$\mathcal{E}_{\theta_{k_1},\theta_{k_2}} = \Re\{\mathbf{z}_{k_1}^T\overline{\mathbf{T}\mathbf{J}}^{\dagger}\left[\mathbb{E}\{\widetilde{\ddot{\mathbf{r}}}\widetilde{\ddot{\mathbf{r}}}^H\} - \ddot{\mathbf{r}}\ddot{\mathbf{r}}^H\right]\overline{\mathbf{J}}^{\dagger H}\overline{\mathbf{T}}^H\mathbf{z}_{k_2}^*\}. \tag{B.47}$$

Substituting $\mathbb{E}\{\widetilde{\ddot{\mathbf{r}}}\widetilde{\ddot{\mathbf{r}}}^H\}$ from (B.44) completes the proof.

B.10 Proof of Theorem 4.6

To derive $\lim_{SNR\to\infty}\mathcal{E}_{\theta_k}$, we need to calculate $\lim_{SNR\to\infty}(\sigma^2+\sum_{k'=1}^{K}p_{k'})^2/p_k^2$, $\mathbf{W}_\infty = \lim_{SNR\to\infty}\mathbf{W}$ and $\boldsymbol{\Gamma}_\infty = \lim_{SNR\to\infty}\boldsymbol{\Gamma}$. It is obtained from (B.20) that

$$\lim_{SNR\to\infty}\frac{(\sigma^2+\sum_{k'=1}^{K}p_{k'})^2}{p_k^2} = \lim_{SNR\to\infty}\frac{1}{\overline{p}_k^2} = K^2. \tag{B.48}$$

In addition, it follows from (4.55) and (4.56) that $\mathbf{W}$ and $\boldsymbol{\Gamma}$ depend on SNR through $\overline{\mathbf{R}}$, $\boldsymbol{\gamma}$ and $\ddot{\mathbf{r}}$. Hence, for calculating $\mathbf{W}_\infty$ and $\boldsymbol{\Gamma}_\infty$, it is sufficient to first compute $\overline{\mathbf{R}}_\infty = \lim_{SNR\to\infty}\overline{\mathbf{R}}$, $\boldsymbol{\gamma}_\infty = \lim_{SNR\to\infty}\boldsymbol{\gamma}$ and $\ddot{\mathbf{r}}_\infty = \lim_{SNR\to\infty}\ddot{\mathbf{r}}$, and then insert them back into the expressions of $\mathbf{W}$ and $\boldsymbol{\Gamma}$ given in (4.55) and (4.56). $\overline{\mathbf{R}}_\infty$ is obtained in (B.20). Given $\overline{\mathbf{R}}_\infty$, $\boldsymbol{\gamma}_\infty = \lim_{SNR\to\infty}\boldsymbol{\gamma}$ is equal to the $(M^2-M)\times 1$ vector containing the real and imaginary parts of the elements of $\overline{\mathbf{R}}_\infty$ above

its main diagonal elements. Then, exploiting (4.38), we have $\ddot{\mathbf{r}}_\infty = \lim_{SNR\to\infty} \boldsymbol{\Psi}^{-1}\bar{\mathbf{J}}^\dagger\mathbf{F}^{-1}\boldsymbol{\gamma} = \boldsymbol{\Psi}^{-1}\bar{\mathbf{J}}^\dagger\mathbf{F}^{-1}\boldsymbol{\gamma}_\infty$. This completes the proof.

B.11

B.11.1 Preliminary to the Calculation of $\boldsymbol{\Sigma}$

Let us first introduce the following lemma which pares the way for calculating $\boldsymbol{\Sigma}$.

Lemma B.1. *Assume that $\varpi_1, \varpi_2, \varpi_3, \varpi_4$ are zero-mean jointly Gaussian random variables with correlation coefficients $\varphi_{i,j}$ where $1 \le i \ne j \le 4$. Define $\omega_n = \mathrm{sgn}(\varpi_n)$, for $1 \le n \le 4$. The expected value of $\omega_1\omega_2\omega_3\omega_4$ is then given by*

$$\mathbb{E}\{\omega_1\omega_2\omega_3\omega_4\} = \sum_{n=2}^{4}\int_0^1 \frac{\varphi_{1,n}}{\sqrt{1-\varphi_{1,n}^2u^2}}\arcsin\Big(\frac{\tau_n(u)}{\varsigma_n(u)\upsilon_n(u)}\Big)\mathrm{d}u, \tag{B.49}$$

where

$$\tau_1(u) = \varphi_{3,4} - \varphi_{2,3}\varphi_{2,4} - \big[\varphi_{1,3}\varphi_{1,4} + \varphi_{1,2}(\varphi_{1,2}\varphi_{3,4} - \varphi_{1,4}\varphi_{2,3} - \varphi_{1,3}\varphi_{2,4})\big]u^2, \tag{B.50}$$

$$\tau_2(u) = \varphi_{2,4} - \varphi_{2,3}\varphi_{3,4} - \big[\varphi_{1,2}\varphi_{1,4} + \varphi_{1,3}(\varphi_{1,3}\varphi_{2,4} - \varphi_{1,4}\varphi_{2,3} - \varphi_{1,2}\varphi_{3,4})\big]u^2, \tag{B.51}$$

$$\tau_3(u) = \varphi_{2,3} - \varphi_{2,4}\varphi_{3,4} - \big[\varphi_{1,2}\varphi_{1,3} + \varphi_{1,4}(\varphi_{1,4}\varphi_{2,3} - \varphi_{1,3}\varphi_{2,4} - \varphi_{1,2}\varphi_{3,4})\big]u^2, \tag{B.52}$$

$$\varsigma_2(u) = \varsigma_3(u) = \sqrt{1-\varphi_{2,3}^2 - \big[\varphi_{1,2}^2 + \varphi_{1,3}^2 - 2\varphi_{1,2}\varphi_{1,3}\varphi_{2,3}\big]u^2}, \tag{B.53}$$

$$\upsilon_2(u) = \varsigma_4(u) = \sqrt{1-\varphi_{2,4}^2 - \big(\varphi_{1,2}^2 + \varphi_{1,4}^2 - 2\varphi_{1,2}\varphi_{1,4}\varphi_{2,4}\big)u^2}, \tag{B.54}$$

$$\upsilon_3(u) = \upsilon_4(u) = \sqrt{1-\varphi_{3,4}^2 - \big(\varphi_{1,3}^2 + \varphi_{1,4}^2 - 2\varphi_{1,3}\varphi_{1,4}\varphi_{3,4}\big)u^2}. \tag{B.55}$$

Proof. We refer the reader to [192]. □

B.11.2 Calculation of $\boldsymbol{\Sigma}$

Define

$$\mathbf{H} = \frac{\pi^2 N}{4}\mathbb{E}\{(\widehat{\mathbf{r}}_{\mathbf{x}} - \mathbf{r}_{\mathbf{x}})(\widehat{\mathbf{r}}_{\mathbf{x}} - \mathbf{r}_{\mathbf{x}})^H\} \in \mathbb{C}^{M^2\times M^2}. \tag{B.56}$$

Then, it is readily observed that

$$[\boldsymbol{\Sigma}]_{p-\lceil\frac{p}{M+1}\rceil, q-\lceil\frac{q}{M+1}\rceil} = [\mathbf{H}]_{p,q}, \tag{B.57}$$

for $p, q \in \mathbb{W} = \{1, 2, \cdots, M^2\} \setminus \{k \mid k = (i-1)M + i, 1 \le i \le M\}$. Hence, instead of $\boldsymbol{\Sigma}$, we compute the elements of $\mathbf{H}$ for $p, q \in \mathbb{W}$ in the following.

The $(p,q)^{\text{th}}$ element of $\mathbf{H}$ is given as follows:

$$[\mathbf{H}]_{p,q} = \frac{\pi^2 N}{4}\mathbb{E}\{[\widehat{\mathbf{r}}_{\mathbf{x}}]_p[\widehat{\mathbf{r}}_{\mathbf{x}}]_q^*\} - \frac{\pi^2 N}{4}[\mathbf{r}_{\mathbf{x}}]_p[\mathbf{r}_{\mathbf{x}}]_q^*. \tag{B.58}$$

For any given $1 \leq p,q \leq M^2$, define $m = p - \lfloor \frac{p-1}{M} \rfloor M$, $i = 1 + \lfloor \frac{p-1}{M} \rfloor$, $n = q - \lfloor \frac{q-1}{M} \rfloor M$ and $l = 1 + \lfloor \frac{p-1}{M} \rfloor$. Then, considering $\hat{\mathbf{R}}_{\mathbf{x}} = \frac{1}{N}\mathbf{X}\mathbf{X}^H = \frac{1}{N}\sum_{t=0}^{N-1}\mathbf{x}(t)\mathbf{x}^H(t)$ yields

$$\begin{aligned}
\mathbb{E}\{[\widehat{\mathbf{r}}_{\mathbf{x}}]_p[\widehat{\mathbf{r}}_{\mathbf{x}}]_q^*\} =& \frac{1}{N^2}\mathbb{E}\left\{\left(\sum_{t=0}^{N-1}[\mathbf{x}(t)]_m[\mathbf{x}(t)]_i^*\right)\left(\sum_{t=0}^{N-1}[\mathbf{x}(t)]_n^*[\mathbf{x}(t)]_l\right)\right\} \\
=& \frac{1}{N^2}\mathbb{E}\left\{\sum_{t_1=0}^{N-1}\sum_{t_2=0}^{N-1}[\mathbf{x}(t_1)]_m[\mathbf{x}(t_1)]_i^*[\mathbf{x}(t_2)]_n^*[\mathbf{x}(t_2)]_l\right\} \\
=& \frac{1}{N}\mathbb{E}\left\{[\mathbf{x}(t)]_m[\mathbf{x}(t)]_i^*[\mathbf{x}(t)]_n^*[\mathbf{x}(t)]_l\right\} \\
&+ \frac{1}{N^2}\sum_{t_1=0}^{N-1}\sum_{t_2=0}^{N-1}\mathbb{E}\left\{[\mathbf{x}(t_1)]_m[\mathbf{x}(t_1)]_i^*\right\}\mathbb{E}\{[\mathbf{x}(t_2)]_n^*[\mathbf{x}(t_2)]_l\} \\
=& \frac{1}{N}\mathbb{E}\{[\mathbf{x}(t)]_m[\mathbf{x}(t)]_i^*[\mathbf{x}(t)]_n^*[\mathbf{x}(t)]_l\} + (1-\frac{1}{N})[\mathbf{r}_{\mathbf{x}}]_p[\mathbf{r}_{\mathbf{x}}]_q^*.
\end{aligned} \tag{B.59}$$

Inserting (B.59) into (B.58) and considering that $\mathbf{r}_{\mathbf{x}} = \frac{2}{\pi}\text{arcsine}(\bar{\mathbf{r}})$ results in

$$\begin{aligned}
[\mathbf{H}]_{p,q} =& \frac{\pi^2}{4}\mathbb{E}\{[\mathbf{x}(t)]_m[\mathbf{x}(t)]_i^*[\mathbf{x}(t)]_n^*[\mathbf{x}(t)]_l\} \\
&- \text{arcsine}([\bar{\mathbf{r}}]_p)\text{arcsine}([\bar{\mathbf{r}}]_q).
\end{aligned} \tag{B.60}$$

Calculation of the expectation in (B.60) needs computing the forth-order moment of $\mathbf{x}(t)$. We note that $p,q \in \mathbb{W}$ implies that $1 \leq m \neq i \leq M$ and $1 \leq n \neq l \leq M$. In the following, we calculate (B.60) through computing the forth-order moment of $\mathbf{x}(t)$ for all admissible values of m,i,n and l.

1. Consider $m = n$ and $i = l$. Then, we have

$$\begin{aligned}
[\mathbf{H}]_{p,p} =& \frac{\pi^2}{4}\mathbb{E}\{|[\mathbf{x}(t)]_m|^4\} - |\text{arcsine}([\bar{\mathbf{r}}]_p)|^2 \\
=& \frac{\pi^2}{4} - |\text{arcsine}([\bar{\mathbf{r}}]_p)|^2,
\end{aligned} \tag{B.61}$$

   where the last equality is obtained from the fact that $[\mathbf{x}(t)]_m = \pm 1/\sqrt{2} \pm j/\sqrt{2}$.

2. Consider $m = n$ and $i \neq l$. Then, we have

$$\begin{aligned}
[\mathbf{H}]_{p,q} =& \frac{\pi^2}{4}\mathbb{E}\{|[\mathbf{x}(t)]_m|^2[\mathbf{x}(t)]_i^*[\mathbf{x}(t)]_l\} - \text{arcsine}([\bar{\mathbf{r}}]_p)\text{arcsine}([\bar{\mathbf{r}}]_q) \\
=& \frac{\pi^2}{4}\mathbb{E}\{[\mathbf{x}(t)]_i^*[\mathbf{x}(t)]_l\} - \text{arcsine}([\bar{\mathbf{r}}]_p)\text{arcsine}([\bar{\mathbf{r}}]_q) \\
=& \frac{\pi}{2}\text{arcsine}([\overline{\mathbf{R}}^*]_{i,l}) - \text{arcsine}([\bar{\mathbf{r}}]_p)\text{arcsine}([\bar{\mathbf{r}}^*]_q),
\end{aligned}$$

   where the last equality is obtained from the arcsine law given in (4.7).

3. Consider $m \neq n$ and $i = l$. Then, we have

$$\begin{aligned}[\mathbf{H}]_{p,q} =& \frac{\pi^2}{4}\mathbb{E}\{|[\mathbf{x}(t)]_i|^2[\mathbf{x}(t)]_m[\mathbf{x}(t)]_n^*\} - \text{arcsine}([\overline{\mathbf{r}}]_p)\text{arcsine}([\overline{\mathbf{r}}]_q) \\ =& \frac{\pi^2}{4}\mathbb{E}\{[\mathbf{x}(t)]_m[\mathbf{x}(t)]_n^*\} - \text{arcsine}([\overline{\mathbf{r}}]_p)\text{arcsine}([\overline{\mathbf{r}}]_q) \\ =& \frac{\pi}{2}\text{arcsine}([\overline{\mathbf{R}}]_{m,n}) - \text{arcsine}([\overline{\mathbf{r}}]_p)\text{arcsine}([\overline{\mathbf{r}}^*]_q). \end{aligned} \tag{B.62}$$

4. Consider $m = l$ and $i = n$. Then, we have

$$\begin{aligned}[\mathbf{H}]_{p,q} =& \frac{\pi^2}{4}\mathbb{E}\left\{([\mathbf{x}(t)]_m)^2\left([\mathbf{x}(t)]_i^*\right)^2\right\} - \text{arcsine}([\overline{\mathbf{r}}]_p)\text{arcsine}([\overline{\mathbf{r}}]_q) \\ =& \frac{\pi^2}{4}\mathbb{E}\Big\{\left(|\Re\{[\mathbf{x}(t)]_m\}|^2 - |\Im\{[\mathbf{x}(t)]_m\}|^2\right. \\ &+ j2\Re\{[\mathbf{x}(t)]_m\}\Im\{[\mathbf{x}(t)]_m\}\big)\left(|\Re\{[\mathbf{x}(t)]_i\}|^2 - |\Im\{[\mathbf{x}(t)]_i\}|^2\right. \\ &\left.\left.+ 2j\Re\{[\mathbf{x}(t)]_i\}\Im\{[\mathbf{x}(t)]_i]\}\right)\Big\} - \text{arcsine}([\overline{\mathbf{r}}]_p)\text{arcsine}([\overline{\mathbf{r}}]_q) \\ =& \pi^2\mathbb{E}\{\Re\{[\mathbf{x}(t)]_m\}\Im\{[\mathbf{x}(t)]_m\}\Re\{[\mathbf{x}(t)]_i\}\Im\{[\mathbf{x}(t)]_i\}\} \\ &- \text{arcsine}([\overline{\mathbf{r}}]_p)\text{arcsine}([\overline{\mathbf{r}}]_q). \end{aligned} \tag{B.63}$$

Making use of Lemma B.1, we obtain

$$\begin{aligned}[\mathbf{H}]_{p,q} =& \int_0^1 \mu_{m,i}(u)\arcsin\left(\xi_{m,i}(u)\right)\mathrm{d}u - \int_0^1 \bar{\mu}_{m,i}(u)\arcsin\left(\bar{\xi}_{m,i}(u)\right)\mathrm{d}u \\ &- \text{arcsine}([\overline{\mathbf{r}}]_p)\text{arcsine}([\overline{\mathbf{r}}^*]_q), \end{aligned} \tag{B.64}$$

where

$$\mu_{m,i}(u) = \frac{\Re\{[\overline{\mathbf{R}}]_{m,i}\}}{\sqrt{1 - |\Re\{[\overline{\mathbf{R}}]_{m,i}\}|^2u^2}}, \tag{B.65}$$

$$\bar{\mu}_{m,i}(u) = \frac{\Im\{[\overline{\mathbf{R}}]_{m,i}\}}{\sqrt{1 - |\Im\{[\overline{\mathbf{R}}]_{m,i}\}|^2u^2}}, \tag{B.66}$$

$$\xi_{m,i}(u) = \Re\{[\overline{\mathbf{R}}]_{m,i}\}\sqrt{\frac{1 - \left[|\Re\{[\overline{\mathbf{R}}]_{m,i}\}|^2 + |\Im\{[\overline{\mathbf{R}}]_{m,i}\}|^2\right]u^2}{1 - |\Im\{[\overline{\mathbf{R}}]_{m,i}\}|^2 - |\Re\{[\overline{\mathbf{R}}]_{m,i}\}|^2u^2}},$$

$$\bar{\xi}_{m,i}(u) = \Im\{[\overline{\mathbf{R}}]_{m,i}\}\sqrt{\frac{1 - \left[|\Re\{[\overline{\mathbf{R}}]_{m,i}\}|^2 + |\Im\{[\overline{\mathbf{R}}]_{m,i}\}|^2\right]u^2}{1 - |\Re\{[\overline{\mathbf{R}}]_{m,i}\}|^2 - |\Im\{[\overline{\mathbf{R}}]_{m,i}\}|^2u^2}}.$$

5. Consider $m = l$ and $i \neq n$. Then, we obtain

$$\begin{aligned}
[\mathbf{H}]_{p,q} =& \frac{\pi^2}{4}\mathbb{E}\{([\mathbf{x}(t)]_m)^2[\mathbf{x}(t)]_i^*[\mathbf{x}(t)]_n^*\} - \text{arcsine}([\overline{\mathbf{r}}]_p)\text{arcsine}([\overline{\mathbf{r}}]_q) \\
=& \frac{\pi^2}{2}\mathbb{E}\Big\{\Re\{[\mathbf{x}(t)]_m\}\Im\{[\mathbf{x}(t)]_m\}\Re\{[\mathbf{x}(t)]_i\}\Im\{[\mathbf{x}(t)]_n \\
&+ \Im\{[\mathbf{x}(t)]_m\}\Re\{[\mathbf{x}(t)]_m\}\Im\{[\mathbf{x}(t)]_i\}\Re\{[\mathbf{x}(t)]_n\} \\
&+ j\Re\{[\mathbf{x}(t)]_m\}\Im\{[\mathbf{x}(t)]_m\}\Re\{[\mathbf{x}(t)]_i\}\Re\{[\mathbf{x}(t)]_n \\
&- j\Im\{[\mathbf{x}(t)]_m\}\Re\{[\mathbf{x}(t)]_m\}\Im\{[\mathbf{x}(t)]_i\}\Im\{[\mathbf{x}(t)]_n\Big\} \\
&- \text{arcsine}([\overline{\mathbf{r}}]_p)\text{arcsine}([\overline{\mathbf{r}}]_q).
\end{aligned} \tag{B.67}$$

Exploiting Lemma B.1 leads to

$$\begin{aligned}
[\mathbf{H}]_{p,q} =& \int_0^1 \mu_{m,i}(u)\arcsin\left(\frac{\vartheta_{m,i,n}(u)}{\delta_{m,i}(u)\kappa_{m,i,n}(u)}\right)\mathrm{d}u \\
&- \int_0^1 \bar{\mu}_{m,n}(u)\arcsin\left(\frac{\bar{\vartheta}_{m,i,n}(u)}{\bar{\delta}_{m,n}(u)\kappa_{m,i,n}(u)}\right)\mathrm{d}u \\
&+ j\int_0^1 \mu_{m,i}(u)\arcsin\left(\frac{\tilde{\vartheta}_{m,i,n}(u)}{\delta_{m,i}(u)\bar{\kappa}_{m,i,n}(u)}\right)\mathrm{d}u \\
&+ j\int_0^1 \mu_{m,n}(u)\arcsin\left(\frac{\ddot{\vartheta}_{m,i,n}(u)}{\tilde{\delta}_{m,n}(u)\bar{\kappa}_{m,i,n}(u)}\right)\mathrm{d}u - \text{arcsine}([\overline{\mathbf{r}}]_p)\text{arcsine}([\overline{\mathbf{r}}^*]_q),
\end{aligned} \tag{B.68}$$

where

$$\begin{aligned}
\vartheta_{m,i,n}(u) =& \Re\{[\overline{\mathbf{R}}]_{m,n}\} + \Im\{[\overline{\mathbf{R}}]_{m,i}\}\Im\{[\overline{\mathbf{R}}]_{i,n}\} - \Re\{[\overline{\mathbf{R}}]_{m,i}\} \\
&\times \left[\Re\{[\overline{\mathbf{R}}]_{m,i}\}\Re\{[\overline{\mathbf{R}}]_{m,n}\} + \Im\{[\overline{\mathbf{R}}]_{m,i}\}\Im\{[\overline{\mathbf{R}}]_{m,n}\}\right]u^2,
\end{aligned} \tag{B.69}$$

$$\begin{aligned}
\bar{\vartheta}_{m,i,n}(u) =& \Im\{[\overline{\mathbf{R}}]_{m,i}\} + \Re\{[\overline{\mathbf{R}}]_{m,n}\}\Im\{[\overline{\mathbf{R}}]_{i,n}\} - \Im\{[\overline{\mathbf{R}}]_{m,n}\} \\
&\times \left[\Im\{[\overline{\mathbf{R}}]_{m,n}\}\Im\{[\overline{\mathbf{R}}]_{m,i}\} + \Re\{[\overline{\mathbf{R}}]_{m,i}\}\Re\{[\overline{\mathbf{R}}]_{m,n}\}\right]u^2,
\end{aligned} \tag{B.70}$$

$$\begin{aligned}
\tilde{\vartheta}_{m,i,n}(u) =& \Im\{[\overline{\mathbf{R}}]_{m,n}\} - \Im\{[\overline{\mathbf{R}}]_{m,i}\}\Re\{[\overline{\mathbf{R}}]_{i,n}\} - \Re\{[\overline{\mathbf{R}}]_{m,i}\} \\
&\times \left[\Re\{[\overline{\mathbf{R}}]_{m,i}\}\Im\{[\overline{\mathbf{R}}]_{m,n}\} - \Re\{[\overline{\mathbf{R}}]_{m,n}\}\Im\{[\overline{\mathbf{R}}]_{m,i}\}\right]u^2,
\end{aligned} \tag{B.71}$$

$$\begin{aligned}
\ddot{\vartheta}_{m,i,n}(u) =& \Im\{[\overline{\mathbf{R}}]_{m,i}\} - \Im\{[\overline{\mathbf{R}}]_{m,n}\}\Re\{[\overline{\mathbf{R}}]_{i,n}\} - \Re\{[\overline{\mathbf{R}}]_{m,n}\} \\
&\times \left[\Re\{[\overline{\mathbf{R}}]_{m,n}\}\Im\{[\overline{\mathbf{R}}]_{m,i}\} - \Re\{[\overline{\mathbf{R}}]_{m,i}\}\Im\{[\overline{\mathbf{R}}]_{m,n}\}\right]u^2,
\end{aligned} \tag{B.72}$$

$$\delta_{m,i}(u) = \sqrt{1 - |\Im\{[\overline{\mathbf{R}}]_{m,i}\}|^2 - |\Re\{[\overline{\mathbf{R}}]_{m,i}\}|^2 u^2}, \tag{B.73}$$

$$\bar{\delta}_{m,n}(u) = \sqrt{1 - |\Re\{[\overline{\mathbf{R}}]_{m,n}\}|^2 - |\Im\{[\overline{\mathbf{R}}]_{m,n}\}|^2 u^2}, \tag{B.74}$$

$$\tilde{\delta}_{m,n}(u) = \sqrt{1 - |\Im\{[\overline{\mathbf{R}}]_{m,n}\}|^2 - |\Re\{[\overline{\mathbf{R}}]_{m,n}\}|^2 u^2}, \tag{B.75}$$

$$\begin{aligned}\kappa_{m,i,n}(u) =& \big[1 - |\Im\{[\overline{\mathbf{R}}]_{i,n}\}|^2 - \big[|\Re\{[\overline{\mathbf{R}}]_{m,i}\}|^2 \\ &+ |\Im\{[\overline{\mathbf{R}}]_{m,n}\}|^2 - 2\Re\{[\overline{\mathbf{R}}]_{m,i}\Im\{[\overline{\mathbf{R}}]_{m,n}\}\Im\{[\overline{\mathbf{R}}]_{i,n}\}u^2\big]^{1/2}\end{aligned} \tag{B.76}$$

$$\begin{aligned}\bar{\kappa}_{m,i,n}(u) =& \big[1 - |\Re\{[\overline{\mathbf{R}}]_{i,n}\}|^2 - \big[|\Re\{[\overline{\mathbf{R}}]_{m,i}\}|^2 \\ &+ |\Re\{[\overline{\mathbf{R}}]_{m,n}\}|^2 - 2\Re\{[\overline{\mathbf{R}}]_{m,i}\Re\{[\overline{\mathbf{R}}]_{m,n}\}\Re\{[\overline{\mathbf{R}}]_{i,n}\}u^2\big]^{1/2}.\end{aligned} \tag{B.77}$$

6. Consider $m \neq l$ and $i = n$. Then, we obtain

$$\begin{aligned}[\mathbf{H}]_{p,q} =& \frac{\pi^2}{4}\mathbb{E}\{[\mathbf{x}(t)]_m[\mathbf{x}(t)]_l([\mathbf{x}(t)]_i^*)^2\} - \text{arcsine}([\overline{\mathbf{r}}]_p)\text{arcsine}([\overline{\mathbf{r}}]_q) \\ =& \frac{\pi^2}{2}\mathbb{E}\Big\{\Re\{[\mathbf{x}(t)]_i\}\Im\{[\mathbf{x}(t)]_i\}\Re\{[\mathbf{x}(t)]_m\}\Im\{[\mathbf{x}(t)]_l \\ &+ \Im\{[\mathbf{x}(t)]_i\}\Re\{[\mathbf{x}(t)]_i\}\Im\{[\mathbf{x}(t)]_m\}\Re\{[\mathbf{x}(t)]_l\} \\ &- j\Re\{[\mathbf{x}(t)]_i\}\Im\{[\mathbf{x}(t)]_i\}\Re\{[\mathbf{x}(t)]_m\}\Re\{[\mathbf{x}(t)]_l \\ &- j\Im\{[\mathbf{x}(t)]_i\}\Re\{[\mathbf{x}(t)]_i\}\Im\{[\mathbf{x}(t)]_m\}\Im\{[\mathbf{x}(t)]_l\Big\} - \text{arcsine}([\overline{\mathbf{r}}]_p)\text{arcsine}([\overline{\mathbf{r}}]_q).\end{aligned} \tag{B.78}$$

Using Lemma B.1, we obtain

$$\begin{aligned}[\mathbf{H}]_{p,q} =& \int_0^1 \mu_{m,i}(u)\arcsin\left(\frac{\vartheta_{i,m,l}(u)}{\delta_{m,i}(u)\kappa_{i,m,l}(u)}\right)\mathrm{d}u \\ &- \int_0^1 \bar{\mu}_{i,l}(u)\arcsin\left(\frac{\bar{\vartheta}_{i,m,l}(u)}{\bar{\delta}_{i,l}(u)\kappa_{i,m,l}(u)}\right)\mathrm{d}u \\ &- j\int_0^1 \mu_{m,i}(u)\arcsin\left(\frac{\tilde{\vartheta}_{i,m,l}(u)}{\delta_{m,i}(u)\bar{\kappa}_{i,m,l}(u)}\right)\mathrm{d}u \\ &- j\int_0^1 \mu_{i,l}(u)\arcsin\left(\frac{\ddot{\vartheta}_{i,m,l}(u)}{\tilde{\delta}_{i,l}(u)\bar{\kappa}_{i,m,l}(u)}\right)\mathrm{d}u - \text{arcsine}([\overline{\mathbf{r}}]_p)\text{arcsine}([\overline{\mathbf{r}}^*]_q),\end{aligned} \tag{B.79}$$

where $\vartheta_{i,m,l}(u)$, $\bar{\vartheta}_{i,m,l}(u)$, $\tilde{\vartheta}_{i,m,l}(u)$, $\ddot{\vartheta}_{i,m,l}(u)$, $\delta_{m,i}(u)$, $\bar{\delta}_{m,i}(u)$, $\tilde{\delta}_{m,i}(u)$, $\kappa_{i,m,l}(u)$ and $\bar{\kappa}_{i,m,l}(u)$ are defined from (B.69) to (B.77), respectively.

7. If $m \neq l \neq i \neq n$, then we have

$$
\begin{aligned}
[\mathbf{H}]_{p,q} =& \frac{\pi^2}{4}\mathbb{E}\{[\mathbf{x}(t)]_m[\mathbf{x}(t)]_i^*[\mathbf{x}(t)]_n^*[\mathbf{x}(t)]_l\} - \text{arcsine}([\bar{\mathbf{r}}]_p)\text{arcsine}([\bar{\mathbf{r}}]_q) \qquad \text{(B.80)}\\
=& \frac{\pi^2}{4}\mathbb{E}\{\Re\{[\mathbf{x}(t)]_m\}\Re\{[\mathbf{x}(t)]_i\}\Re\{[\mathbf{x}(t)]_n\}\Re\{[\mathbf{x}(t)]_l\}\\
&+ \Im\{[\mathbf{x}(t)]_m\}\Im\{[\mathbf{x}(t)]_i\}\Im\{[\mathbf{x}(t)]_n\}\Im\{[\mathbf{x}(t)]_l\}\\
&+ \Re\{[\mathbf{x}(t)]_m\}\Re\{[\mathbf{x}(t)]_i\}\Im\{[\mathbf{x}(t)]_n\}\Im\{[\mathbf{x}(t)]_l\}\\
&+ \Im\{[\mathbf{x}(t)]_m\}\Im\{[\mathbf{x}(t)]_i\}\Re\{[\mathbf{x}(t)]_n\}\Re\{[\mathbf{x}(t)]_l\}\\
&+ \Re\{[\mathbf{x}(t)]_m\}\Im\{[\mathbf{x}(t)]_i\}\Re\{[\mathbf{x}(t)]_n\}\Im\{[\mathbf{x}(t)]_l\}\\
&- \Re\{[\mathbf{x}(t)]_m\}\Im\{[\mathbf{x}(t)]_i\}\Im\{[\mathbf{x}(t)]_n\}\Re\{[\mathbf{x}(t)]_l\}\\
&- \Im\{[\mathbf{x}(t)]_m\}\Re\{[\mathbf{x}(t)]_i\}\Re\{[\mathbf{x}(t)]_n\}\Im\{[\mathbf{x}(t)]_l\}\\
&+ \Im\{[\mathbf{x}(t)]_m\}\Re\{[\mathbf{x}(t)]_i\}\Im\{[\mathbf{x}(t)]_n\}\Re\{[\mathbf{x}(t)]_l\}\}\\
&+ \frac{j\pi^2}{4}\mathbb{E}\{\Re\{[\mathbf{x}(t)]_m\}\Re\{[\mathbf{x}(t)]_i\}\Re\{[\mathbf{x}(t)]_n\}\Im\{[\mathbf{x}(t)]_l\}\\
&- \Re\{[\mathbf{x}(t)]_m\}\Re\{[\mathbf{x}(t)]_i\}\Im\{[\mathbf{x}(t)]_n\}\Re\{[\mathbf{x}(t)]_l\}\\
&+ \Im\{[\mathbf{x}(t)]_m\}\Im\{[\mathbf{x}(t)]_i\}\Re\{[\mathbf{x}(t)]_n\}\Im\{[\mathbf{x}(t)]_l\}\\
&- \Im\{[\mathbf{x}(t)]_m\}\Im\{[\mathbf{x}(t)]_i\}\Im\{[\mathbf{x}(t)]_n\}\Re\{[\mathbf{x}(t)]_l\}\\
&+ \Im\{[\mathbf{x}(t)]_m\}\Re\{[\mathbf{x}(t)]_i\}\Re\{[\mathbf{x}(t)]_n\}\Re\{[\mathbf{x}(t)]_l\}\\
&+ \Im\{[\mathbf{x}(t)]_m\}\Re\{[\mathbf{x}(t)]_i\}\Im\{[\mathbf{x}(t)]_n\}\Im\{[\mathbf{x}(t)]_l\}\\
&- \Re\{[\mathbf{x}(t)]_m\}\Im\{[\mathbf{x}(t)]_i\}\Re\{[\mathbf{x}(t)]_n\}\Re\{[\mathbf{x}(t)]_l\}\\
&- \Re\{[\mathbf{x}(t)]_m\}\Im\{[\mathbf{x}(t)]_i\}\Im\{[\mathbf{x}(t)]_n\}\Im\{[\mathbf{x}(t)]_l\}\} - \text{arcsine}([\bar{\mathbf{r}}]_p)\text{arcsine}([\bar{\mathbf{r}}]_q).
\end{aligned}
$$

Making use of Lemma B.1, $[\mathbf{\Sigma}]_{p,q}$ is obtained as given in (B.81) as follows

$$
\begin{aligned}
[\mathbf{H}]_{p,q} = \frac{1}{2}\Bigg[& \int_0^1 \mu_{m,i}(u) \left[\arcsin\left(\frac{\rho_{m,i,n,l}(u)}{\eta_{m,i,n,l}(u)\eta_{m,i,l,n}(u)} \right) + \arcsin\left(\frac{\bar{\rho}_{m,i,n,l}(u)}{\bar{\eta}_{m,i,n,l}(u)\bar{\eta}_{m,i,l,n}(u)} \right) \right] \mathrm{d}u \\
& + \int_0^1 \bar{\mu}_{m,i}(u) \left[\arcsin\left(\frac{\tilde{\rho}_{m,n,i,l}(u)}{\bar{\eta}_{m,n,i,l}(u)\tilde{\eta}_{m,n,i,l}(u)} \right) - \arcsin\left(\frac{\tilde{\rho}_{m,l,i,n}(u)}{\bar{\eta}_{m,l,i,n}(u)\tilde{\eta}_{m,l,i,n}(u)} \right) \right] \mathrm{d}u \\
& + \int_0^1 \mu_{m,n}(u) \left[\arcsin\left(\frac{\rho_{m,n,i,l}(u)}{\eta_{m,i,n,l}(u)\eta_{m,n,l,i}(u)} \right) + \arcsin\left(\frac{\bar{\rho}_{m,n,i,l}(u)}{\bar{\eta}_{m,n,i,l}(u)\bar{\eta}_{m,n,l,i}(u)} \right) \right] \mathrm{d}u \\
& + \int_0^1 \bar{\mu}_{m,n}(u) \left[\arcsin\left(\frac{\tilde{\rho}_{m,i,n,l}(u)}{\bar{\eta}_{m,i,n,l}(u)\tilde{\eta}_{m,i,n,l}(u)} \right) - \arcsin\left(\frac{\tilde{\rho}_{m,l,n,i}(u)}{\bar{\eta}_{m,l,n,i}(u)\tilde{\eta}_{m,l,i,n}(u)} \right) \right] \mathrm{d}u \\
& + \int_0^1 \mu_{m,l}(u) \left[\arcsin\left(\frac{\rho_{m,l,i,n}(u)}{\eta_{m,i,l,n}(u)\eta_{m,n,l,i}(u)} \right) - \arcsin\left(\frac{\bar{\rho}_{m,l,i,n}(u)}{\bar{\eta}_{m,l,i,n}(u)\bar{\eta}_{m,l,n,i}(u)} \right) \right] \mathrm{d}u \\
& + \int_0^1 \bar{\mu}_{m,l}(u) \left[\arcsin\left(\frac{\tilde{\rho}_{m,i,l,n}(u)}{\bar{\eta}_{m,i,l,n}(u)\tilde{\eta}_{m,i,n,l}(u)} \right) + \arcsin\left(\frac{\tilde{\rho}_{m,n,l,i}(u)}{\bar{\eta}_{m,n,l,i}(u)\tilde{\eta}_{m,n,i,l}(u)} \right) \right] \mathrm{d}u \Bigg] \\
& + \frac{j}{2}\Bigg[\int_0^1 \mu_{m,i}(u) \left[\arcsin\left(\frac{\ddot{\rho}_{m,i,l,n}(u)(u)}{\eta_{m,i,l,n}(u)\bar{\eta}_{m,i,n,l}(u)} \right) - \arcsin\left(\frac{\ddot{\rho}_{m,i,n,l}(u)(u)}{\eta_{m,i,n,l}(u)\bar{\eta}_{m,i,l,n}(u)} \right) \right] \mathrm{d}u \\
& + \int_0^1 \bar{\mu}_{m,i}(u) \left[\arcsin\left(\frac{\dot{\rho}_{m,i,n,l}(u)(u)}{\tilde{\eta}_{m,l,i,n}(u)\tilde{\eta}_{m,n,i,l}(u)} \right) + \arcsin\left(\frac{\breve{\rho}_{m,n,l,i}(u)(u)}{\bar{\eta}_{m,n,i,l}(u)\bar{\eta}_{m,l,i,n}(u)} \right) \right] \mathrm{d}u \\
& + \int_0^1 \mu_{m,n}(u) \left[\arcsin\left(\frac{\ddot{\rho}_{m,n,l,i}(u)(u)}{\eta_{m,n,l,i}(u)\bar{\eta}_{m,n,i,l}(u)} \right) - \arcsin\left(\frac{\ddot{\rho}_{m,n,i,l}(u)(u)}{\eta_{m,i,n,l}(u)\bar{\eta}_{m,n,l,i}(u)} \right) \right] \mathrm{d}u \\
& + \int_0^1 \bar{\mu}_{m,n}(u) \left[\arcsin\left(\frac{\breve{\rho}_{m,i,l,n}(u)(u)}{\bar{\eta}_{m,i,n,l}(u)\bar{\eta}_{m,l,n,i}(u)} \right) + \arcsin\left(\frac{\dot{\rho}_{m,n,i,l}(u)(u)}{\tilde{\eta}_{m,l,i,n}(u)\tilde{\eta}_{m,i,n,l}(u)} \right) \right] \mathrm{d}u \\
& + \int_0^1 \mu_{m,l}(u) \left[\arcsin\left(\frac{\ddot{\rho}_{m,l,i,n}(u)(u)}{\eta_{m,i,l,n}(u)\bar{\eta}_{m,l,n,i}(u)} \right) + \arcsin\left(\frac{\ddot{\rho}_{m,l,n,i}(u)(u)}{\eta_{m,n,l,i}(u)\bar{\eta}_{m,l,i,n}(u)} \right) \right] \mathrm{d}u \\
& + \int_0^1 \bar{\mu}_{m,l}(u) \left[\arcsin\left(\frac{\dot{\rho}_{m,l,i,n}(u)(u)}{\tilde{\eta}_{m,n,i,l}(u)\tilde{\eta}_{m,i,n,l}(u)} \right) - \arcsin\left(\frac{\breve{\rho}_{m,i,n,l}(u)(u)}{\bar{\eta}_{m,i,l,n}(u)\bar{\eta}_{m,n,l,i}(u)} \right) \right] \mathrm{d}u \Bigg] \\
& - \mathrm{arcsine}([\bar{\mathbf{r}}]_p)\mathrm{arcsine}([\bar{\mathbf{r}}^*]_q),
\end{aligned}
\tag{B.81}
$$

where

$$\rho_{m,i,n,l}(u) = \Re\{[\overline{\mathbf{R}}]_{n,l}\} - \Re\{[\overline{\mathbf{R}}]_{i,n}\}\Re\{[\overline{\mathbf{R}}]_{i,l}\} - \big[\Re\{[\overline{\mathbf{R}}]_{m,n}\}\Re\{[\overline{\mathbf{R}}]_{m,l}\} + \Re\{[\overline{\mathbf{R}}]_{m,i}\}\left(\Re\{[\overline{\mathbf{R}}]_{m,i}\}\Re\{[\overline{\mathbf{R}}]_{n,l}\} - \Re\{[\overline{\mathbf{R}}]_{m,l}\}\Re\{[\overline{\mathbf{R}}]_{i,n}\} - \Re\{[\overline{\mathbf{R}}]_{m,n}\}\Re\{[\overline{\mathbf{R}}]_{i,l}\}\right)\big]u^2, \tag{B.82}$$

$$\bar{\rho}_{m,i,n,l}(u) = \Re\{[\overline{\mathbf{R}}]_{n,l}\} - \Im\{[\overline{\mathbf{R}}]_{i,n}\}\Im\{[\overline{\mathbf{R}}]_{i,l}\} - \big[\Im\{[\overline{\mathbf{R}}]_{m,n}\}\Im\{[\overline{\mathbf{R}}]_{m,l}\} + \Re\{[\overline{\mathbf{R}}]_{m,i}\}\left(\Re\{[\overline{\mathbf{R}}]_{m,i}\}\Re\{[\overline{\mathbf{R}}]_{n,l}\} - \Im\{[\overline{\mathbf{R}}]_{m,l}\}\Im\{[\overline{\mathbf{R}}]_{i,n}\} - \Im\{[\overline{\mathbf{R}}]_{m,n}\}\Im\{[\overline{\mathbf{R}}]_{i,l}\}\right)\big]u^2, \tag{B.83}$$

$$\tilde{\rho}_{m,i,n,l}(u) = \Im\{[\overline{\mathbf{R}}]_{i,l}\} - \Im\{[\overline{\mathbf{R}}]_{i,n}\}\Re\{[\overline{\mathbf{R}}]_{n,l}\} - \big[\Re\{[\overline{\mathbf{R}}]_{m,i}\}\Im\{[\overline{\mathbf{R}}]_{m,l}\} + \Im\{[\overline{\mathbf{R}}]_{m,n}\}\left(\Im\{[\overline{\mathbf{R}}]_{m,n}\}\Im\{[\overline{\mathbf{R}}]_{i,l}\} - \Im\{[\overline{\mathbf{R}}]_{m,l}\}\Im\{[\overline{\mathbf{R}}]_{i,l}\} - \Re\{[\overline{\mathbf{R}}]_{m,i}\}\Re\{[\overline{\mathbf{R}}]_{n,l}\}\right)\big]u^2, \tag{B.84}$$

$$\ddot{\rho}_{m,i,n,l}(u) = \Im\{[\overline{\mathbf{R}}]_{n,l}\} - \Re\{[\overline{\mathbf{R}}]_{i,n}\}\Im\{[\overline{\mathbf{R}}]_{i,l}\} - \big[\Re\{[\overline{\mathbf{R}}]_{m,n}\}\Im\{[\overline{\mathbf{R}}]_{m,l}\} + \Re\{[\overline{\mathbf{R}}]_{m,i}\}\left(\Re\{[\overline{\mathbf{R}}]_{m,i}\}\Im\{[\overline{\mathbf{R}}]_{n,l}\} - \Im\{[\overline{\mathbf{R}}]_{m,l}\}\Re\{[\overline{\mathbf{R}}]_{i,n}\} - \Re\{[\overline{\mathbf{R}}]_{m,n}\}\Im\{[\overline{\mathbf{R}}]_{i,l}\}\right)\big]u^2, \tag{B.85}$$

$$\dot{\rho}_{m,i,n,l}(u) = \Re\{[\overline{\mathbf{R}}]_{n,l}\} - \Re\{[\overline{\mathbf{R}}]_{i,n}\}\Re\{[\overline{\mathbf{R}}]_{i,l}\} - \big[\Im\{[\overline{\mathbf{R}}]_{m,n}\}\Im\{[\overline{\mathbf{R}}]_{m,l}\} + \Im\{[\overline{\mathbf{R}}]_{m,i}\}\left(\Im\{[\overline{\mathbf{R}}]_{m,i}\}\Re\{[\overline{\mathbf{R}}]_{n,l}\} - \Im\{[\overline{\mathbf{R}}]_{m,l}\}\Re\{[\overline{\mathbf{R}}]_{i,n}\} - \Im\{[\overline{\mathbf{R}}]_{m,n}\}\Re\{[\overline{\mathbf{R}}]_{i,l}\}\right)\big]u^2, \tag{B.86}$$

$$\breve{\rho}_{m,i,n,l}(u) = \Re\{[\overline{\mathbf{R}}]_{i,n}\} - \Im\{[\overline{\mathbf{R}}]_{i,l}\}\Im\{[\overline{\mathbf{R}}]_{n,l}\} - \big[\Re\{[\overline{\mathbf{R}}]_{m,i}\}\Re\{[\overline{\mathbf{R}}]_{m,n}\} + \Im\{[\overline{\mathbf{R}}]_{m,l}\}\left(\Im\{[\overline{\mathbf{R}}]_{m,l}\}\Re\{[\overline{\mathbf{R}}]_{i,n}\} - \Re\{[\overline{\mathbf{R}}]_{m,n}\}\Im\{[\overline{\mathbf{R}}]_{i,l}\} - \Re\{[\overline{\mathbf{R}}]_{m,i}\}\Im\{[\overline{\mathbf{R}}]_{n,l}\}\right)\big]u^2, \tag{B.87}$$

$$\eta_{m,i,n,l}(u) = \sqrt{1 - |\Re\{[\overline{\mathbf{R}}]_{i,n}\}|^2 - \left[|\Re\{[\overline{\mathbf{R}}]_{m,i}\}|^2 + |\Re\{[\overline{\mathbf{R}}]_{m,n}\}|^2 - 2\Re\{[\overline{\mathbf{R}}]_{m,i}\}\Re\{[\overline{\mathbf{R}}]_{m,n}\}\Re\{[\overline{\mathbf{R}}]_{i,n}\}\right]u^2}, \tag{B.88}$$

$$\bar{\eta}_{m,i,n,l}(u) = \sqrt{1 - |\Im\{[\overline{\mathbf{R}}]_{i,n}\}|^2 - \left[|\Re\{[\overline{\mathbf{R}}]_{m,i}\}|^2 + |\Im\{[\overline{\mathbf{R}}]_{m,n}\}|^2 - 2\Re\{[\overline{\mathbf{R}}]_{m,i}\}\Im\{[\overline{\mathbf{R}}]_{m,n}\}\Im\{[\overline{\mathbf{R}}]_{i,n}\}\right]u^2}, \tag{B.89}$$

$$\tilde{\eta}_{m,i,n,l}(u) = \sqrt{1 - |\Re\{[\overline{\mathbf{R}}]_{n,l}\}|^2 - \left[|\Im\{[\overline{\mathbf{R}}]_{m,n}\}|^2 + |\Im\{[\overline{\mathbf{R}}]_{m,l}\}|^2 - 2\Im\{[\overline{\mathbf{R}}]_{m,n}\}\Im\{[\overline{\mathbf{R}}]_{m,l}\}\Re\{[\overline{\mathbf{R}}]_{n,l}\}\right]u^2}. \tag{B.90}$$

Appendix C

Appendices of Chapter 6

C.1 Proof of Lemma 6.1

The optimization problems (6.16) and (6.17) are equivalent. Hence, it suffices to prove this for only (6.16). Define

$$\mathbf{y}_m^\circ = \widetilde{\alpha}_m^\circ \mathbf{s}(\widetilde{\tau}_m^\circ) + \alpha_m^\circ \mathbf{s}(\tau_m^\circ) + \mathbf{n}_m^\circ, \tag{C.1}$$

where $[\widetilde{\alpha}_m^\circ, \alpha_m^\circ, \widetilde{\tau}_m^\circ, \tau_m^\circ, \mathbf{n}_m^\circ]^T \neq [\widetilde{\alpha}_m, \alpha_m, \widetilde{\tau}_m, \tau_m, \mathbf{n}_m]^T$ and thus, $\mathbf{y}_m^\circ \neq \mathbf{y}_m$. It suffices to show that $[Q(\mathbf{y}_m^\circ)]_l \neq [Q(\mathbf{y}_m)]_l$ at least for one l as $L \to \infty$. The previous statement holds only if, at least for one l, the following occurs:

$$\begin{cases} \Re\{[\mathbf{y}_m]_l\} > \Re\{[\boldsymbol{\gamma}_m]_l\} > \Re\{[\mathbf{y}_m^\circ]_l\}, & \text{or,} \\ \Re\{[\mathbf{y}_m]_l\} < \Re\{[\boldsymbol{\gamma}_m]_l\} < \Re\{[\mathbf{y}_m^\circ]_l\}, & \text{or,} \\ \Im\{[\mathbf{y}_m]_l\} > \Im\{[\boldsymbol{\gamma}_m]_l\} > \Im\{[\mathbf{y}_m^\circ]_l\}, & \text{or,} \\ \Im\{[\mathbf{y}_m]_l\} < \Im\{[\boldsymbol{\gamma}_m]_l\} < \Im\{[\mathbf{y}_m^\circ]_l\}. \end{cases} \tag{C.2}$$

Let $\mathcal{A}$ denote the event described by (C.2) for a given l. In practice, the real and imaginary parts of $[\mathbf{y}_m]_l$ and $[\mathbf{y}_m^\circ]_l$ are upper bounded by, say, $A_{\max}$. Then, probability of $\mathcal{A}$ is [193]

$$\Pr(\mathcal{A}) = \frac{|\Re\{[\mathbf{y}_m]_l\} - \Re\{[\mathbf{y}_m^\circ]_l\}|}{2A_{\max}} + \frac{|\Im\{[\mathbf{y}_m]_l\} - \Im\{[\mathbf{y}_m^\circ]_l\}|}{2A_{\max}}. \tag{C.3}$$

The probability that (C.2) occurs at least for one l, denoted by $\mathcal{H}$, is

$$\Pr(\mathcal{H}) = 1 - \prod_{l=1}^{L} \Bigg(1 - \frac{|\Re\{[\mathbf{y}_m]_l\} - \Re\{[\mathbf{y}_m^\circ]_l\}|}{2A_{\max}} - \frac{|\Im\{[\mathbf{y}_m]_l\} - \Im\{[\mathbf{y}_m^\circ]_l\}|}{2A_{\max}} \Bigg). \tag{C.4}$$

From [193], $1 - x \leq e^{-x}, \forall x \in \mathbb{R}$. Hence, it follows that

$$\Pr(\mathcal{H}) \geq 1 - e^{-\sum_{l=1}^{L} \frac{|\Re\{[\mathbf{y}_m]_l\} - \Re\{[\mathbf{y}_m^\circ]_l\}|}{2A_{\max}} - \frac{|\Im\{[\mathbf{y}_m]_l\} + \Im\{[\mathbf{y}_m^\circ]_l\}|}{2A_{\max}}}. \tag{C.5}$$

But $\mathbf{y}_m^\circ \neq \mathbf{y}_m$. Thus, $-\sum_{l=1}^{L} \frac{|\Re\{[\mathbf{y}_m]_l\} - \Re\{[\mathbf{y}_m^\circ]_l\}|}{2A_{\max}} - \frac{|\Im\{[\mathbf{y}_m]_l\} + \Im\{[\mathbf{y}_m^\circ]_l\}|}{2A_{\max}} \to \infty$ as $L \to \infty$, and $\lim_{L\to\infty} \Pr(\mathcal{H}) = 1$. This implies that $\mathbf{y}_m$ is the only point which satisfies the constraints in (6.16) as $L \to \infty$. Accordingly, as $L \to \infty$, the optimization problem (6.16) reduces to the LASSO estimator which has been shown to be consistent [194]. This completes the proof.

C.2 Proof of Theorem6.1

To show that (6.48) is equivalent to (6.47), we first prove that the global minimum of (6.47) coincides with that of (6.48). Assume that $\mathbf{r}_o^\star$ and $\begin{bmatrix}\mathbf{r}_e^{\star T} & v^\star\end{bmatrix}^T$ are the minimizers of (6.47) and (6.48), respectively. Define a set $\mathbb{K} = \left\{\mathbf{r} \in \mathbb{R}_{\geq 0}^M \mid \mathbf{w} \odot (\mathbf{r} - \boldsymbol{\lambda}) \succeq \mathbf{0}\right\}$. Given $\mathcal{J}(\mathbf{r}) \geq 0$ for $\mathbf{r} \in \mathbb{K}$, it readily follows from the first constraint in (6.48) that $\frac{\mathcal{F}(\mathbf{r}_e^\star)}{\mathcal{J}(\mathbf{r}_e^\star)} \leq v^\star$. Considering that $\mathbf{r}_e^\star$ belongs to the feasible set of (6.47), i.e., $\mathbf{r}_e^\star \in \mathbb{K}$, we obtain

$$\frac{\mathcal{F}(\mathbf{r}_o^\star)}{\mathcal{J}(\mathbf{r}_o^\star)} \leq \frac{\mathcal{F}(\mathbf{r}_e^\star)}{\mathcal{J}(\mathbf{r}_e^\star)} \leq v^\star. \tag{C.6}$$

On the other hand, defining $v_o = \frac{\mathcal{F}(\mathbf{r}_o^\star)}{\mathcal{J}(\mathbf{r}_o^\star)}$ and considering $\mathbf{r}_o^\star \in \mathbb{K}$, it follows that $\begin{bmatrix}\mathbf{r}_o^{\star T} & v_o\end{bmatrix}^T$ is in the feasible set of (6.48). Therefore,

$$v^\star \leq v_o = \frac{\mathcal{F}(\mathbf{r}_o^\star)}{\mathcal{J}(\mathbf{r}_o^\star)}, \tag{C.7}$$

Now, comparing (C.6) and (C.7) implies that (6.47) and (6.48) share the same global minimum, i.e.,

$$v^\star = \frac{\mathcal{F}(\mathbf{r}_o^\star)}{\mathcal{J}(\mathbf{r}_o^\star)}. \tag{C.8}$$

Further deduction from (C.6) and (C.8) yields

$$\frac{\mathcal{F}(\mathbf{r}_o^\star)}{\mathcal{J}(\mathbf{r}_o^\star)} = \frac{\mathcal{F}(\mathbf{r}_e^\star)}{\mathcal{J}(\mathbf{r}_e^\star)}, \tag{C.9}$$

indicating $\mathbf{r}_e^\star$ is also a minimizer of (6.47). This completes the proof.

C.3 Proof of Theorem 6.2

C.3.1 Preliminaries to the Proof

Recall the definition of sum-of-squares (SOS) polynomial and a useful related result as follows.

Definition C.1 (Sum-of-squares)**.** A polynomial $\mathcal{P}(\mathbf{u})$ of degree $2q$ is sum-of-squares (SOS) if and only if there exist polynomials $\mathcal{Y}_1(\mathbf{u}), \cdots, \mathcal{Y}_I(\mathbf{u})$ of degree q such that $\mathcal{P}(\mathbf{u}) = \sum_{i=1}^{I} \mathcal{Y}_i^2(\mathbf{u})$.

Lemma C.1. *Given* $\mathbb{P}$ *as the set of SOS polynomials and polynomials* $\mathcal{E}_i(\mathbf{u})$ *for* $1 \leq i \leq I$*, define the sets*

$$\mathbb{W} = \{\mathbf{u} \in \mathbb{R}^n \mid \mathcal{E}_i(\mathbf{u}) \geq 0, \forall i \in \{1, 2, \cdots, I\}\} \tag{C.10}$$

$$\mathbb{G}_p = \left\{ \sum_{i=0}^{I} \mathcal{P}_i(\mathbf{u})\mathcal{E}_i(\mathbf{u}) \mid \mathcal{E}_0(\mathbf{u}) = 1, \mathcal{P}_i(\mathbf{u}) \in \mathbb{P}, \deg\left(\mathcal{P}_i(\mathbf{u})\mathcal{E}_i(\mathbf{u})\right) \leq 2p \right\}, \tag{C.11}$$

such that $\mathbb{W}$ *is compact and there exists a polynomial* $\mathcal{U}(\mathbf{u}) \in \mathbb{G}_p$ *where* $\{\mathbf{u} \in \mathbb{R}^n \mid \mathcal{U}(\mathbf{u}) \geq 0\}$ *is compact. Then, a polynomial* $\mathcal{B}(\mathbf{u})$ *of degree* q *is strictly positive on* $\mathbb{W}$*, i.e.,* $\mathcal{B}(\mathbf{u}) > 0 \ \forall \mathbf{u} \in \mathbb{W}$*, if and only if* $\mathcal{B}(\mathbf{u}) \in \mathbb{G}_p$ *for some integer* $p \geq \max\left(\lceil q \rceil, \max_i \left\lceil \frac{\deg(\mathcal{E}_i)}{2} \right\rceil\right)$.

Proof. We refer the reader to [195]. □

C.3.2 Proof of the Theorem

We first show that (6.48) satisfies the conditions stated in Lemma C.1 of Appendix C.3.1. In consequence, it can be reformulated as minimization of a positive polynomial function on a compact set. Lasserre has shown that minimizer of a positive polynomial function on a compact set can be obtained through solving an equivalent SDP [152, Theorem 4.2]. Thus, we ultimately resort to [152, Theorem 4.2] to recast the resulting optimization problem as an SDP.

Consider $\mathcal{E}_i$'s to be the inequality constraints of (6.48). Then, we need to prove the following three statements:

1. The feasible set of (6.48) is compact.
2. A polynomial $\mathcal{U}([\mathbf{r}, v]^T) \in \mathbb{G}_p$ exists such that $\{\mathbf{r} \in \mathbb{R}^M, v \in \mathbb{R} \mid \mathcal{U}([\mathbf{r}, v]^T) \geq 0\}$ is compact.
3. The objective function of (6.48) is strictly positive on its feasible set.

For the first statement, note that the feasible set contains all of its boundary points and is therefore closed. From Heine-Borel Theorem [196], to show compactness of the feasible set, it suffices to show that it is bounded. To this end, note the constraint on the value of $\mathbf{r}$ which is limited by the maximum detectable range $r_{\max} \in \mathbb{R}_{>0}$ of the NB-IoT nodes so that $r_m \leq r_{\max}$ for all $m \in \mathbb{M}$. This implies that the continuous function $\frac{\mathcal{F}(\mathbf{r})}{\mathcal{J}(\mathbf{r})}$ is bounded on $\mathbb{T} = \{\mathbf{r} \in \mathbb{R}^M \mid r_m \leq r_{\max}, \ \forall m \in \mathbb{M}\}$ [196, Theorem 4.16]. In other words, $\frac{\mathcal{F}(\mathbf{r})}{\mathcal{J}(\mathbf{r})} \leq \varphi$, where $\varphi = \underset{\mathbf{r} \in \mathbb{T}}{\text{maximize}} \ \frac{\mathcal{F}(\mathbf{r})}{\mathcal{J}(\mathbf{r})}$. The optimization problem in (6.48) is indeed a minimization of an upper bound of $\frac{\mathcal{F}(\mathbf{r})}{\mathcal{J}(\mathbf{r})}$, i.e. v. Without loss of generality, assume $v \leq v_{\max}$ where $v_{\max} \geq \varphi$. These practical constraints on $\mathbf{r}$ and v do not change the solution of (6.48) but guarantee boundedness and thereby compactness of the its feasible set. On the other hand, it is possible to show the boundedness of v, in turn, entails the boundedness of $\mathbf{r}$. To show that, let assume $\mathbb{B}$ to be an arbitrary subset of $\{1, \cdots, M\}$ and define $\mathbf{c}$ such that $[\mathbf{c}]_k = [\mathbf{r}]_k$ for $k \in \mathbb{B}$. When $v \leq v_{\max}$, from (6.45) and (6.46), we get

$$\begin{aligned}&\lim_{\mathbf{c}\to\infty} v\mathcal{J}(\mathbf{r})-\mathcal{F}(\mathbf{r})=\\&-\frac{1}{4}\lim_{\mathbf{c}\to\infty}\Big(\|\Pi_{\mathbf{V}}^{\perp}(\overline{\mathbf{r}}-r_1\mathbf{1})\|_2^2\|\Pi_{\mathbf{V}}^{\perp}\big[(\overline{\mathbf{r}}-r_1\mathbf{1})\odot(\overline{\mathbf{r}}-r_1\mathbf{1})\big]\|_2^2\\&-\Big(\big[(\overline{\mathbf{r}}-r_1\mathbf{1})\odot(\overline{\mathbf{r}}-r_1\mathbf{1})\big]^T\Pi_{\mathbf{V}}^{\perp}(\overline{\mathbf{r}}-r_1\mathbf{1})\Big)^2\Big),\end{aligned}\tag{C.12}$$

Using Cauchy–Schwarz inequality and idempotency of $\Pi_{\mathbf{V}}^{\perp}$, we have

$$\begin{aligned}&\|\Pi_{\mathbf{V}}^{\perp}(\overline{\mathbf{r}}-r_1\mathbf{1})\|_2^2\|\Pi_{\mathbf{V}}^{\perp}\big[(\overline{\mathbf{r}}-r_1\mathbf{1})\odot(\overline{\mathbf{r}}-r_1\mathbf{1})\big]\|_2^2\geq\\&\qquad\Big(\big[(\overline{\mathbf{r}}-r_1\mathbf{1})\odot(\overline{\mathbf{r}}-r_1\mathbf{1})\big]^T\Pi_{\mathbf{V}}^{\perp}(\overline{\mathbf{r}}-r_1\mathbf{1})\Big)^2.\end{aligned}\tag{C.13}$$

It follows from (C.12) and (C.13) that, when $v\leq v_{\max}$ and as each r_m approaches infinity, the constraint $v\mathcal{J}(\mathbf{r})-\mathcal{F}(\mathbf{r})$ becomes negative. Hence, when $v\leq v_{\max}$, to ensure $v\mathcal{J}(\mathbf{r})-\mathcal{F}(\mathbf{r})\geq 0$, the ranges r_m, $m\in\mathbb{M}$ must be bounded. This implies that $v\leq v_{\max}$ is sufficient for the compactness of the feasible set of (6.48). Accordingly, without loss of generality, the optimization problem (6.48) becomes

$$\begin{aligned}\underset{v,\mathbf{r}}{\text{minimize}}\quad & v\\ \text{s.t.}\quad & v\mathcal{J}(\mathbf{r})-\mathcal{F}(\mathbf{r})\geq 0,\\ & \mathbf{w}\odot(\mathbf{r}-\boldsymbol{\lambda})\succeq\mathbf{0},\\ & \mathbf{r}\succeq\mathbf{0},\\ & v_{\max}-v\geq 0,\end{aligned}\tag{C.14}$$

in which the feasible set is compact. Note that, in practice, the value of φ is unknown and, to satisfy the condition $v_{\max}\geq\varphi$, $v_{\max}$ should be selected sufficiently large.

For the second statement, consider

$$\mathcal{E}_i([\mathbf{r},v]^T)=\begin{cases}1 & \text{if}\ \ i=0,\\ v\mathcal{J}(\mathbf{r})-\mathcal{F}(\mathbf{r}), & \text{if}\ \ i=1,\\ w_{i-1}(r_{i-1}-\lambda_{i-1}), & \text{if}\ \ i=2,\cdots,M+1,\\ r_{i-M-1}, & \text{if}\ \ i=M+2,\cdots,2M+1,\\ v_{\max}-v, & \text{if}\ \ i=2M+2,\end{cases}\tag{C.15}$$

and that $\mathbb{G}_p$ is defined according to (C.11). Construct $\mathcal{P}_i([\mathbf{r},v]^T)=0$ for $i=0,1,\cdots,2M+1$ and $\mathcal{P}_{2M+2}([\mathbf{r},v]^T)=1$. It readily follows that $v_{\max}-v=\sum_{i=0}^{2M+2}\mathcal{P}_i([\mathbf{r},v]^T)\mathcal{E}_i([\mathbf{r},v]^T)$, thus $v_{\max}-v\in\mathbb{G}_p$ with $p\geq 1$. Further, the set $\{v\in\mathbb{R}\mid v_{\max}-v\geq 0\}$ is closed and bounded and, therefore, compact. This proves the second statement.

The third statement requires establishing the strict positiveness of the objective on the feasible set of (C.14), i.e., $\mathbb{W}=\{\mathbf{r}\in\mathbb{R}^M,\ v\in\mathbb{R}\mid\mathbf{r}\succeq\mathbf{0},\ \mathbf{w}\odot(\mathbf{r}-\boldsymbol{\lambda})\succeq\mathbf{0},\ v\mathcal{J}(\mathbf{r})-\mathcal{F}(\mathbf{r})\geq 0,\ v_{\max}-v\geq 0\}$. Considering $a\in\mathbb{R}_{>0}$ as a constant parameter independent of $\mathbf{r}$ and v, it is always possible to replace v with $v+a$ in the cost function of (C.14) without affecting its solution. Then, it follows

from (6.42) that $v \geq \mathcal{L}(\mathbf{r}) = \frac{\mathcal{F}(\mathbf{r})}{\mathcal{J}(\mathbf{r})} \geq 0$, thereby $v + a > 0$ on $\mathbb{W}$ for any constant $a \in \mathbb{R}_{>0}$. This proves the third statement.

Consequently, according to Lemma C.1, (6.48) is equivalent to minimization of the positive function $v+a$ on the compact set $\mathbb{W} = \{\mathbf{r} \in \mathbb{R}^M,\ v \in \mathbb{R} \mid \mathcal{E}_i([\mathbf{r}, v]^T) \geq 0, \forall i \in \{1, 2, \cdots, 2M+2\}\}$ where $\mathcal{E}_i$'s are given in (C.15). Now, resorting to [152, Theorem 4.2], the resulting minimization problem can be equivalently recast as the SDP in (6.57). This completes the proof.

Bibliography

[1] H. Van Trees, *Optimum Array Processing (Detection, Estimation, and Modulation Theory, Part IV)*. New York: John Wiley and Sons Inc., 2002.

[2] B. Ottersten, "Array processing for wireless communications," in *Proceedings of 8th Workshop on Statistical Signal and Array Processing*, Jun 1996, pp. 466–473.

[3] S. S. Haykin, J. Litva, , and T. J. Shepherd, Eds., *Radar Array Processing*. Berlin, Germany: Springer-Verlag, 1993.

[4] N. Dey and A. S. Ashour, *Direction of arrival estimation and localization of multi-speech sources*. Springer, 2018.

[5] S. F. Cotter and B. D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Transactions on Communications*, vol. 50, no. 3, pp. 374–377, 2002.

[6] C. Carbonelli, S. Vedantam, and U. Mitra, "Sparse channel estimation with zero tap detection," *IEEE Transactions on Wireless Communications*, vol. 6, no. 5, pp. 1743–1763, 2007.

[7] S. P. Chepuri and G. Leus, "Sparsity-promoting sensor selection for non-linear measurement models," *IEEE Transactions on Signal Processing*, vol. 63, no. 3, pp. 684–698, 2015.

[8] X. Wang, A. Hassanien, and M. G. Amin, "Dual-function mimo radar communications system design via sparse array optimization," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 3, pp. 1213–1226, 2019.

[9] E. Tohidi, M. Coutino, S. P. Chepuri, H. Behroozi, M. M. Nayebi, and G. Leus, "Sparse antenna and pulse placement for colocated mimo radar," *IEEE Transactions on Signal Processing*, vol. 67, no. 3, pp. 579–593, 2019.

[10] X. Wang, A. Hassanien, and M. G. Amin, "Sparse transmit array design for dual-function radar communications by antenna selection," *Digital Signal Processing*, vol. 83, pp. 223 – 234, 2018.

[11] D. Cohen, D. Cohen, and Y. C. Eldar, "High resolution fdma mimo radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, no. 4, pp. 2806–2822, 2020.

[12] A. De Maio, Y. C. Eldar, and A. M. Haimovich, *Compressed sensing in radar signal processing*. Cambridge University Press, 2019.

[13] G. Qin, M. G. Amin, and Y. D. Zhang, "Doa estimation exploiting sparse array motions," *IEEE Transactions on Signal Processing*, vol. 67, no. 11, pp. 3013–3027, 2019.

[14] S. A. Hamza and M. G. Amin, "Sparse array beamforming design for wideband signal models," *IEEE Transactions on Aerospace and Electronic Systems*, pp. 1–1, 2020.

[15] Y. D. Zhang, M. G. Amin, and B. Himed, "Sparsity-based DoA estimation using co-prime arrays," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 3967–3971.

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

[17] P. P. Vaidyanathan and P. Pal, "Sparse sensing with co-prime samplers and arrays," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 573–586, Feb 2011.

[18] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, 2006.

[19] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Transactions on Image Processing*, vol. 17, no. 1, pp. 53–69, 2008.

[20] M. Protter and M. Elad, "Image sequence denoising via sparse and redundant representations," *IEEE Transactions on Image Processing*, vol. 18, no. 1, pp. 27–35, 2009.

[21] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 4, pp. 539–550, April 1999.

[22] H. Sun, A. Nallanathan, C. Wang, and Y. Chen, "Wideband spectrum sensing for cognitive radio networks: a survey," *IEEE Wireless Communications*, vol. 20, no. 2, pp. 74–81, 2013.

[23] J. Lunden, V. Koivunen, and H. V. Poor, "Spectrum exploration and exploitation for cognitive radio: Recent advances," *IEEE Signal Processing Magazine*, vol. 32, no. 3, pp. 123–140, 2015.

[24] J. Hasch, E. Topak, R. Schnabel, T. Zwick, R. Weigel, and C. Waldschmidt, "Millimeter-wave technology for automotive radar sensors in the 77 GHz frequency band," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 3, pp. 845–860, 2012.

[25] B. F. Burke, F. Graham-Smith, and P. N. Wilkinson, *An introduction to radio astronomy*. Cambridge University Press, 2019.

[26] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, Oct 2014.

[27] A. Gokceoglu, E. Björnson, E. G. Larsson, and M. Valkama, "Spatio-temporal waveform design for multiuser massive MIMO downlink with 1-bit receivers," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 2, pp. 347–362, 2017.

[28] A. K. Saxena, I. Fijalkow, and A. L. Swindlehurst, "Analysis of one-bit quantized precoding for the multiuser massive MIMO downlink," *IEEE Transactions on Signal Processing*, vol. 65, no. 17, pp. 4624–4634, 2017.

[29] B. Zhao, L. Huang, J. Li, M. Liu, and J. Wang, "Deceptive SAR jamming based on 1-bit sampling and time-varying thresholds," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 3, pp. 939–950, 2018.

[30] X. Wang, G. Li, Y. Liu, and M. G. Amin, "Enhanced 1-bit radar imaging by exploiting two-level block sparsity," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 1131–1141, 2019.

[31] A. Ameri, A. Bose, J. Li, and M. Soltanalian, "One-bit radar processing with time-varying sampling thresholds," *IEEE Transactions on Signal Processing*, vol. 67, no. 20, pp. 5297–5308, 2019.

[32] O. Bar-Shalom and A. J. Weiss, "DoA estimation using one-bit quantized measurements," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 3, pp. 868–884, July 2002.

[33] M. Stein, K. Barbe, and J. A. Nossek, "DoA parameter estimation with 1-bit quantization bounds, methods and the exponential replacement," in *WSA 2016; 20th International ITG Workshop on Smart Antennas*, March 2016, pp. 1–6.

[34] M. J. Pelgrom, "Analog-to-digital conversion," in *Analog-to-Digital Conversion*. Springer, 2013, pp. 325–418.

[35] Q. Shen, W. Liu, W. Cui, and S. Wu, "Underdetermined DoA estimation under the compressive sensing framework: A review," *IEEE Access*, vol. 4, pp. 8865–8878, 2016.

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

[37] ——, "Correlation-aware techniques for sparse support recovery," in *2012 IEEE Statistical Signal Processing Workshop (SSP)*, Aug 2012, pp. 53–56.

[38] M. Wang and A. Nehorai, "Coarrays, MUSIC, and the Cramér-Rao bound," *IEEE Trans. Signal Process.*, vol. 65, no. 4, pp. 933–946, Feb 2017.

[39] X. Huang, S. Bi, and B. Liao, "Direction-of-arrival estimation based on quantized matrix recovery," *IEEE Communications Letters*, vol. 24, no. 2, pp. 349–353, 2020.

[40] I. Yoffe, N. Regev, and D. Wulich, "On direction of arrival estimation with 1-bit quantizer," in *2019 IEEE Radar Conference (RadarConf)*, 2019, pp. 1–6.

[41] C. Stöckle, J. Munir, A. Mezghani, and J. A. Nossek, "1-bit direction of arrival estimation based on compressed sensing," in *2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2015, pp. 246–250.

[42] X. Huang, P. Xiao, and B. Liao, "One-bit direction of arrival estimation with an improved fixed-point continuation algorithm," in *2018 10th International Conference on Wireless Communications and Signal Processing (WCSP)*, 2018, pp. 1–4.

[43] X. Meng and J. Zhu, "A generalized sparse bayesian learning algorithm for 1-bit doa estimation," *IEEE Communications Letters*, vol. 22, no. 7, pp. 1414–1417, 2018.

[44] A. Paulraj, B. Ottersten, R. Roy, A. Swindlehurst, G. Xu, and T. Kailath, "16 subspace methods for directions-of-arrival estimation," *Handbook of Statistics*, vol. 10, pp. 693–739, 1993.

[45] F. Li, H. Liu, and R. J. Vaccaro, "Performance analysis for DoA estimation algorithms: unification, simplification, and observations," *IEEE Trans. Aerosp. Electron. Syst*, vol. 29, no. 4, pp. 1170–1184, Oct 1993.

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

[47] M. Viberg, B. Ottersten, and T. Kailath, "Detection and estimation in sensor arrays using weighted subspace fitting," *IEEE Transactions on Signal Processing*, vol. 39, no. 11, pp. 2436–2449, Nov 1991.

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

[49] S. Qin, Y. D. Zhang, and M. G. Amin, "Generalized coprime array configurations for direction-of-arrival estimation," *IEEE Transactions on Signal Processing*, vol. 63, no. 6, pp. 1377–1390, 2015.

[50] S. Qin, Y. D. Zhang, M. G. Amin, and B. Himed, "Doa estimation exploiting a uniform linear array with multiple co-prime frequencies," *Signal Processing*, vol. 130, pp. 37 – 46, 2017.

[51] C. Liu and P. P. Vaidyanathan, "Super nested arrays: Linear sparse arrays with reduced mutual coupling—part i: Fundamentals," *IEEE Transactions on Signal Processing*, vol. 64, no. 15, pp. 3997–4012, 2016.

[52] C. L. Liu and P. P. Vaidyanathan, "Cramér-Rao bounds for coprime and other sparse arrays, which find more sources than sensors," *Digital Signal Processing*, vol. 61, pp. 43 – 61, 2017.

[53] S. Sedighi, B. S. M. R. Rao, and B. Ottersten, "An asymptotically efficient weighted least squares estimator for co-array-based DoA estimation," *IEEE Transactions on Signal Processing*, vol. 68, pp. 589–604, 2020.

[54] Q. Shen, W. Liu, W. Cui, S. Wu, Y. D. Zhang, and M. G. Amin, "Low-complexity direction-of-arrival estimation based on wideband co-prime arrays," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 9, pp. 1445–1456, 2015.

[55] Y. Chi, A. Pezeshki, L. Scharf, and R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, March 2010, pp. 3930–3933.

[56] Z. Tan and A. Nehorai, "Sparse direction of arrival estimation using co-prime arrays with off-grid targets," *IEEE Signal Processing Letters*, vol. 21, no. 1, pp. 26–29, Jan 2014.

[57] Z. Yang, L. Xie, and C. Zhang, "A discretization-free sparse and parametric approach for linear array signal processing," *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 4959–4973, Oct 2014.

[58] J. Steinwandt, F. Roemer, and M. Haardt, "Performance analysis of ESPRIT-type algorithms for co-array structures," in *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Dec 2017, pp. 1–5.

[59] C. L. Liu and P. P. Vaidyanathan, "Remarks on the spatial smoothing step in coarray MUSIC," *IEEE Signal Processing Letters*, vol. 22, no. 9, pp. 1438–1442, Sept 2015.

[60] C. Zhou, Y. Gu, X. Fan, Z. Shi, G. Mao, and Y. D. Zhang, "Direction-of-arrival estimation for coprime array via virtual array interpolation," *IEEE Transactions on Signal Processing*, vol. 66, no. 22, pp. 5956–5971, Nov 2018.

[61] C. Zhou, Y. Gu, Z. Shi, and Y. D. Zhang, "Off-grid direction-of-arrival estimation using coprime array interpolation," *IEEE Signal Processing Letters*, vol. 25, no. 11, pp. 1710–1714, Nov 2018.

[62] H. Qiao and P. Pal, "Gridless line spectrum estimation and low-rank toeplitz matrix compression using structured samplers: A regularization-free approach," *IEEE Transactions on Signal Processing*, vol. 65, no. 9, pp. 2221–2236, May 2017.

[63] ——, "Unified analysis of co-array interpolation for direction-of-arrival estimation," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017, pp. 3056–3060.

[64] S. Rao, A. Mezghani, and A. L. Swindlehurst, "Channel estimation in one-bit massive MIMO systems: Angular versus unstructured models," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 5, pp. 1017–1031, Sep. 2019.

[65] H. Pirzadeh, G. Seco-Granados, S. Rao, and A. L. Swindlehurst, "Spectral efficiency of one-bit sigma-delta massive MIMO systems," *IEEE Journal on Selected Areas in Communications*, pp. 1–1, 2020.

[66] Q. Wan, J. Fang, H. Duan, Z. Chen, and H. Li, "Generalized Bussgang LMMSE channel estimation for one-bit massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 19, no. 6, pp. 4234–4246, 2020.

[67] H. Zayyani, M. Korki, and F. Marvasti, "Dictionary learning for blind one-bit compressed sensing," *IEEE Signal Processing Letters*, vol. 23, no. 2, pp. 187–191, 2015.

[68] S. J. Zahabi, M. M. Naghsh, M. Modarres-Hashemi, and J. Li, "One-bit compressive radar sensing in the presence of clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, no. 1, pp. 167–185, 2020.

[69] F. Xi, Y. Xiang, S. Chen, and A. Nehorai, "Gridless parameter estimation for one-bit MIMO radar with time-varying thresholds," *IEEE Transactions on Signal Processing*, vol. 68, pp. 1048–1063, 2020.

[70] S. Sedighi, K. V. Mishra, M. R. B. Shankar, and B. Ottersten, "Localization with one-bit passive radars in narrowband internet-of-things using multivariate polynomial optimization," 2020. [Online]. Available: arXiv:2007.15108v1

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

[72] T. Chen, M. Guo, and X. Huang, "Direction finding using compressive one-bit measurements," *IEEE Access*, vol. 6, pp. 41 201–41 211, 2018.

[73] X. Huang and B. Liao, "One-bit MUSIC," *IEEE Signal Processing Letters*, vol. 26, no. 7, pp. 961–965, July 2019.

[74] J. H. Van Vleck and D. Middleton, "The spectrum of clipped noise," *Proceedings of the IEEE*, vol. 54, no. 1, pp. 2–19, 1966.

[75] C. Liu and P. P. Vaidyanathan, "One-bit sparse array DoA estimation," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017, pp. 3126–3130.

[76] K. N. Ramamohan, S. Prabhakar Chepuri, D. F. Comesaña, and G. Leus, "Blind calibration of sparse arrays for doa estimation with analog and one-bit measurements," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 4185–4189.

[77] Z. Cheng, S. Chen, Q. Shen, J. He, and Z. Liu, "Direction finding of electromagnetic sources on a sparse cross-dipole array using one-bit measurements," *IEEE Access*, vol. 8, pp. 83 131–83 143, 2020.

[78] C. Zhou, Y. Gu, Z. Shi, and M. Haardt, "Direction-of-arrival estimation for coprime arrays via coarray correlation reconstruction: A one-bit perspective," in *2020 IEEE 11th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, 2020, pp. 1–4.

[79] S. Gogineni and A. Nehorai, "Target estimation using sparse modeling for distributed mimo radar," *IEEE Transactions on Signal Processing*, vol. 59, no. 11, pp. 5315–5325, 2011.

[80] S. Xu and K. Doğançay, "Optimal sensor placement for 3-d angle-of-arrival target localization," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 3, pp. 1196–1211, 2017.

[81] S. Tomic, M. Beko, R. Dinis, and P. Montezuma, "A closed-form solution for rss/aoa target localization by spherical coordinates conversion," *IEEE Wireless Communications Letters*, vol. 5, no. 6, pp. 680–683, 2016.

[82] J. Li and P. Stoica, "Mimo radar with colocated antennas," *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 106–114, 2007.

[83] A. Hassanien, M. G. Amin, Y. D. Zhang, and F. Ahmad, "High-resolution single-snapshot doa estimation in mimo radar with colocated antennas," in *2015 IEEE Radar Conference (RadarCon)*, 2015, pp. 1134–1138.

[84] J. Li and P. Stoica, *MIMO radar signal processing*. John Wiley & Sons, 2008.

[85] J. J. Caffery, *Wireless location in CDMA cellular radio systems*. Springer Science & Business Media, 2000.

[86] C. Wang, F. Qi, G. Shi, and X. Wang, "Convex combination based target localization with noisy angle of arrival measurements," *IEEE Wireless Communications Letters*, vol. 3, no. 1, pp. 14–17, 2014.

[87] D. Niculescu and Badri Nath, "Ad hoc positioning system (aps) using aoa," in *IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No.03CH37428)*, vol. 3, 2003, pp. 1734–1743 vol.3.

[88] Y. He, A. Behnad, and X. Wang, "Accuracy analysis of the two-reference-node angle-of-arrival localization system," *IEEE Wireless Communications Letters*, vol. 4, no. 3, pp. 329–332, 2015.

[89] G. Mao, B. D. Anderson, and B. Fidan, "Path loss exponent estimation for wireless sensor network localization," *Computer Networks*, vol. 51, no. 10, pp. 2467–2483, 2007.

[90] H. Sallouha, A. Chiumento, and S. Pollin, "Localization in long-range ultra narrow band IoT networks using RSSI," in *IEEE International Conference on Communications*, 2017, pp. 1–6.

[91] S. Hu, X. Li, and F. Rusek, "On time-of-arrival estimation in nb-iot systems," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, 2019, pp. 1–6.

[92] K. Tong, X. Wang, A. Khabbazibasmenj, and A. Dounavis, "Optimum reference node deployment for toa-based localization," in *2015 IEEE International Conference on Communications (ICC)*, 2015, pp. 3252–3256.

[93] A. Noroozi and M. A. Sebt, "Target localization from bistatic range measurements in multi-transmitter multi-receiver passive radar," *IEEE Signal Processing Letters*, vol. 22, no. 12, pp. 2445–2449, 2015.

[94] Y. Zou and H. Liu, "Tdoa localization with unknown signal propagation speed and sensor position errors," *IEEE Communications Letters*, vol. 24, no. 5, pp. 1024–1027, 2020.

[95] S. Sedighi, R. B. S. Mysore, S. Maleki, and B. Ottersten, "Consistent least squares estimator for co-array-based DoA estimation," in *2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, July 2018, pp. 524–528.

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

[97] J. Nie, "An exact jacobian SDP relaxation for polynomial optimization," *Mathematical Programming*, vol. 137, no. 1-2, pp. 225–255, 2013.

[98] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[99] P. Stoica and A. Nehorai, "Mode, maximum likelihood and Cramér-Rao bound: conditional and unconditional results," in *International Conference on Acoustics, Speech, and Signal Processing*, Apr 1990, pp. 2715–2718.

[100] M. Jansson, B. Goransson, and B. Ottersten, "A subspace method for direction of arrival estimation of uncorrelated emitter signals," *IEEE Trans. Signal Process.*, vol. 47, no. 4, pp. 945–956, April 1999.

[101] H. Yanai, K. Takeuchi, and Y. Takane, *Projection Matrices, Generalized Inverse Matrices, and Singular Value decomposition*. USA: New York: Springer-Verlag, 2014.

[102] B. Ottersten, M. Viberg, and T. Kailath, "Performance analysis of the total least squares ESPRIT algorithm," *IEEE Transactions on Signal Processing*, vol. 39, no. 5, pp. 1122–1135, May 1991.

[103] M. Kaveh and A. Barabell, "The statistical performance of the MUSIC and the minimum-norm algorithms in resolving plane waves in noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 2, pp. 331–341, April 1986.

[104] S. Sedighi, M. R. Bhavani Shankar, and B. Ottersten, "A statistically efficient estimator for co-array based DoA estimation," in *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, 2018, pp. 880–883.

[105] E. L. Lehmann and G. Casella, *Theory of point estimation*. Springer Science & Business Media, 2006.

[106] A. W. Van der Vaart, *Asymptotic statistics*. Cambridge university press, 2000, vol. 3.

[107] S. M. Kay, *Fundamentals of statistical signal processing, Volume I: Estimation theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[108] P. Stoica and P. Babu, "The Gaussian data assumption leads to the largest Cramér-Rao bound [lecture notes]," *IEEE Signal Processing Magazine*, vol. 28, no. 3, pp. 132–133, 2011.

[109] I. Abrahamson *et al.*, "Orthant probabilities for the quadrivariate normal distribution," *The Annals of Mathematical Statistics*, vol. 35, no. 4, pp. 1685–1703, 1964.

[110] A. Papoulis, *Probability, random variables and stochastic processes*. New York: McGraw-hill, 2002.

[111] D. N. Lal, "A note on a form of Tchebycheff's inequality for two or more variables," *Sankhyā: The Indian Journal of Statistics (1933-1960)*, vol. 15, no. 3, pp. 317–320, 1955.

[112] S. Sedighi, M. R. B. Shankar, M. Soltanalian, and B. Ottersten, "On the performance of one-bit DoA estimation via sparse linear arrays," 2020. [Online]. Available: arXiv:2012.14051v1

[113] M. S. Gowda and R. Sznajder, "Schur complements, Schur determinantal and Haynsworth inertia formulas in Euclidean Jordan algebras," *Linear Algebra Appl*, vol. 432, pp. 1553–1559, 2010.

[114] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.

[115] J. C. Bezdek and R. J. Hathaway, "Convergence of alternating optimization," *Neural, Parallel & Scientific Computations*, vol. 11, no. 4, pp. 351–368, 2003.

[116] F. Rivet, Y. Deval, J.-B. Begueret, D. Dallet, P. Cathelin, and D. Belot, "The experimental demonstration of a SASP-based full software radio receiver," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 5, pp. 979–988, 2010.

[117] A. Ikpehai, B. Adebisi, K. M. Rabie, K. Anoh, R. E. Ande, M. Hammoudeh, H. Gacanin, and U. M. Mbanaso, "Low-power wide area network technologies for Internet-of-Things: A comparative review," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2225–2240, 2018.

[118] E. Sisinni, A. Saifullah, S. Han, U. Jennehag, and M. Gidlund, "Industrial Internet of Things: Challenges, opportunities, and directions," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 11, pp. 4724–4734, 2018.

[119] F. Khelifi, A. Bradai, A. Benslimane, P. Rawat, and M. Atri, "A survey of localization systems in Internet of Things," *Mobile Networks and Applications*, vol. 24, no. 3, pp. 761–785, 2019.

[120] R. C. Shit, S. Sharma, D. Puthal, and A. Y. Zomaya, "Location of things (LoT): A review and taxonomy of sensors localization in IoT infrastructure," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2028–2061, 2018.

[121] A. Kott, A. Swami, and B. J. West, "The internet of battle things," *Computer*, vol. 49, no. 12, pp. 70–75, 2016.

[122] Y. Rong, A. Chiriyath, A. Dutta, and D. Bliss, "Active breathing suppression for improved sleep monitoring heartbeat detection using UWB radar," in *IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, 2019, pp. 161–165.

[123] A. Dorri, S. S. Kanhere, R. Jurdak, and P. Gauravaram, "Blockchain for IoT security and privacy: The case study of a smart home," in *IEEE International Conference on Pervasive Computing and Communications Workshops*, 2017, pp. 618–623.

[124] S. Han, H. Huang, Z. Luo, and C. Foropon, "Harnessing the power of crowdsourcing and Internet of Things in disaster response," *Annals of Operations Research*, vol. 283, no. 1-2, pp. 1175–1190, 2019.

[125] K. V. Mishra, M. R. B. Shankar, and B. Ottersten, "Deep rainrate estimation from highly attenuated downlink signals of ground-based communications satellite terminals," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2020, pp. 9021–9025.

[126] D. Mishra and E. G. Larsson, "Multi-tag backscattering to MIMO reader: Channel estimation and throughput fairness," *IEEE Transactions on Wireless Communications*, vol. 18, no. 12, pp. 5584–5599, 2019.

[127] C. Pérez-Penichet, F. Hermans, A. Varshney, and T. Voigt, "Augmenting IoT networks with backscatter-enabled passive sensor tags," in *ACM Workshop on Hot Topics in Wireless*, 2016, pp. 23–27.

[128] J. F. Ensworth and M. S. Reynolds, "Every smart phone is a backscatter reader: Modulated backscatter compatibility with Bluetooth 4.0 Low Energy (BLE) devices," in *IEEE International Conference on RFID*, 2015, pp. 78–85.

[129] J. Xu, J. Yao, L. Wang, Z. Ming, K. Wu, and L. Chen, "Narrowband Internet of Things: Evolutions, technologies, and open issues," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1449–1462, 2017.

[130] W. Yang, M. Wang, J. Zhang, J. Zou, M. Hua, T. Xia, and X. You, "Narrowband wireless access for low-power massive Internet of Things: A bandwidth perspective," *IEEE Wireless Communications*, vol. 24, no. 3, pp. 138–145, 2017.

[131] L. Zhang, Y. Liang, and M. Xiao, "Spectrum sharing for Internet of Things: A survey," *IEEE Wireless Communications*, vol. 26, no. 3, pp. 132–139, 2019.

[132] B. Kellogg, A. Parks, S. Gollakota, J. R. Smith, and D. Wetherall, "Wi-Fi backscatter: Internet connectivity for RF-powered devices," *ACM SIGCOMM Computer Communication Review*, vol. 44, no. 4, pp. 607–618, 2015.

[133] S. Gezici, Z. Tian, G. B. Giannakis, H. Kobayashi, A. F. Molisch, H. V. Poor, and Z. Sahinoglu, "Localization via ultra-wideband radios: A look at positioning aspects of future sensor networks," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 70–84, 2005.

[134] Q. Song, S. Guo, X. Liu, and Y. Yang, "CSI amplitude fingerprinting-based NB-IoT indoor localization," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1494–1504, 2017.

[135] S. Hu, A. Berg, X. Li, and F. Rusek, "Improving the performance of OTDOA based positioning in NB-IoT systems," in *IEEE Global Communications Conference*, 2017, pp. 1–7.

[136] S. Hu, X. Li, and F. Rusek, "On time-of-arrival estimation in NB-IoT systems," in *IEEE Wireless Communications and Networking Conference*, 2019, pp. 1–6.

[137] W. S. Jeon, S. B. Seo, and D. G. Jeong, "Effective frequency hopping pattern for ToA estimation in NB-IoT random access," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 10, pp. 10 150–10 154, 2018.

[138] H. Sallouha, A. Chiumento, S. Rajendran, and S. Pollin, "Localization in ultra narrow band IoT networks: Design guidelines and tradeoffs," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 9375–9385, 2019.

[139] K. V. Mishra, M. R. Bhavani Shankar, V. Koivunen, B. Ottersten, and S. A. Vorobyov, "Toward millimeter wave joint radar communications: A signal processing perspective," *IEEE Signal Processing Magazine*, vol. 36, no. 5, pp. 100–114, 2019.

[140] Z. Li, W. Xu, X. Zhang, and J. Lin, "A survey on one-bit compressed sensing: Theory and applications," *Frontiers of Computer Science*, vol. 12, no. 2, pp. 217–230, 2018.

[141] O. Dabeer and E. Masry, "Multivariate signal parameter estimation under dependent noise from 1-bit dithered quantized data," *IEEE Transactions on Information Theory*, vol. 54, no. 4, pp. 1637–1654, 2008.

[142] A. Host-Madsen and P. Handel, "Effects of sampling and quantization on single-tone frequency estimation," *IEEE Transactions on Signal Processing*, vol. 48, no. 3, pp. 650–662, 2000.

[143] O. Dabeer and A. Karnik, "Signal parameter estimation using 1-bit dithered quantization," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5389–5405, 2006.

[144] Z. Cvetkovic, I. Daubechies, and B. F. Logan, "Single-bit oversampled A/D conversion with exponential accuracy in the bit rate," *IEEE Transactions on Information Theory*, vol. 53, no. 11, pp. 3979–3989, 2007.

[145] O. Bar-Shalom and A. J. Weiss, "DOA estimation using one-bit quantized measurements," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 3, pp. 868–884, 2002.

[146] S. Sedighi, B. Shankar, M. Soltanalian, and B. Ottersten, "One-bit DoA estimation via sparse linear arrays," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2020, pp. 9135–9139.

[147] H. Pirzadeh and A. L. Swindlehurst, "Spectral efficiency under energy constraint for mixed-adc mrc massive mimo," *IEEE Signal Processing Letters*, vol. 24, no. 12, pp. 1847–1851, 2017.

[148] A. M. Elbir and K. V. Mishra, "Joint antenna selection and hybrid beamformer design using unquantized and quantized deep learning networks," *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 1677–1688, 2020.

[149] J. J. Bussgang, "Crosscorrelation functions of amplitude-distorted Gaussian signals," Research Laboratory of Electronics, Massachusetts Institute of Technology, Tech. Rep. 216, 1952.

[150] K. Knudson, R. Saab, and R. Ward, "One-bit compressive sensing with norm estimation," *IEEE Transactions on Information Theory*, vol. 62, no. 5, pp. 2748–2758, 2016.

[151] N. Z. Shor, "Class of global minimum bounds of polynomial functions," *Cybernetics and Systems Analysis*, vol. 23, no. 6, pp. 731–734, 1987.

[152] J. B. Lasserre, "Global optimization with polynomials and the problem of moments," *SIAM Journal on Optimization*, vol. 11, no. 3, pp. 796–817, 2001.

[153] S. Sedighi, K. V. Mishra, M. R. Bhavani Shankar, and B. Ottersten, "Localization performance of 1-bit passive radars in NB-IoT applications," in *IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, 2019, pp. 156–160.

[154] S. H. Dokhanchi, B. S. Mysore, K. V. Mishra, and B. Ottersten, "A mmWave automotive joint radar-communications system," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 3, pp. 1241–1260, 2019.

[155] J. Liu, K. V. Mishra, and M. Saquib, "Co-designing statistical MIMO radar and in-band full-duplex multi-user MIMO communications," *arXiv preprint arXiv:2006.14774*, 2020.

[156] A. Loulou, J. Yli-Kaakinen, T. Levanen, V. Lehtinen, F. Schaich, T. Wild, M. Renfors, and M. Valkama, "Multiplierless filtered-OFDM transmitter for Narrowband IoT devices," *IEEE Internet of Things Journal*, vol. 7, no. 2, pp. 846–862, 2020.

[157] O. Simeone, U. Spagnolini, Y. Bar-Ness, and S. H. Strogatz, "Distributed synchronization in wireless networks," *IEEE Signal Processing Magazine*, vol. 25, no. 5, pp. 81–97, 2008.

[158] Y. Wu, Q. Chaudhari, and E. Serpedin, "Clock synchronization of wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 28, no. 1, pp. 124–138, 2011.

[159] A. Ali and W. Hamouda, "On the cell search and initial synchronization for nb-iot lte systems," *IEEE Communications Letters*, vol. 21, no. 8, pp. 1843–1846, 2017.

[160] J. Zhang, M. Wang, M. Hua, W. Yang, and X. You, "Robust synchronization waveform design for massive iot," *IEEE Transactions on Wireless Communications*, vol. 16, no. 11, pp. 7551–7559, 2017.

[161] *IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems*, 2008, IEEE Std 1588-2008 (Revision of IEEE Std 1588-2002).

[162] D. L. Mills, "Internet time synchronization: The network time protocol," *IEEE Trans. Commun.*, vol. 39, no. 10, pp. 1482–1493, 1991.

[163] A. Garg, A. Yadav, A. Sikora, and A. S. Sairam, "Wireless precision time protocol," *IEEE Commun. Lett.*, vol. 22, no. 4, pp. 812–815, 2017.

[164] A. K. Karthik and R. S. Blum, "Robust clock skew and offset estimation for IEEE 1588 in the presence of unexpected deterministic path delay asymmetries," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 5102–5119, 2020.

[165] G. Wang and K. V. Mishra, "Displaced sensor automotive radar imaging," *arXiv preprint arXiv:2010.04085*, 2020.

[166] S. J. Zahabi, M. M. Naghsh, M. Modarres-Hashemi, and J. Li, "One-bit compressive radar sensing in the presence of clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, no. 1, pp. 167–185, 2020.

[167] H. Zhu, X. Shang, and J. Li, "Target parameter estimation via one-bit PMCW radar," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2020, pp. 9145–9149.

[168] J. L. Garry, C. J. Baker, and G. E. Smith, "Evaluation of direct signal suppression for passive radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3786–3799, 2017.

[169] L. Daniel, S. Hristov, X. Lyu, A. G. Stove, M. Cherniakov, and M. Gashinova, "Design and validation of a passive radar concept for ship detection using communication satellite signals," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 6, pp. 3115–3134, 2017.

[170] Z. Yang, J. Li, P. Stoica, and L. Xie, "Sparse methods for direction-of-arrival estimation," in *Array, Radar and Communications Engineering*, ser. Academic Press Library in Signal Processing, R. Chellappa and S. Theodoridis, Eds. Academic Press, 2018, vol. 7, pp. 509–581.

[171] A. V. Oppenheim and R. W. Schafer, *Discrete-time signal processing*, 3rd ed. Prentice Hall, 2009.

[172] R. Zamir, "A proof of the Fisher information inequality via a data processing argument," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 1246–1250, 1998.

[173] H. Yanai, K. Takeuchi, and Y. Takane, *Projection matrices, generalized inverse matrices, and singular value decomposition*. Springer, 2011.

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

[175] F. Folster, H. Rohling, and U. Lubbert, "An automotive radar network based on 77 ghz fmcw sensors," in *IEEE International Radar Conference, 2005.*, 2005, pp. 871–876.

[176] H. Godrich, A. M. Haimovich, and R. S. Blum, "Target localization accuracy gain in mimo radar-based systems," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2783–2803, 2010.

[177] H. Jamali-Rad and G. Leus, "Sparsity-aware multi-source TDOA localization," *IEEE Transactions on Signal Processing*, vol. 61, no. 19, pp. 4874–4887, 2013.

[178] R. Amiri, F. Behnia, and H. Zamani, "Asymptotically efficient target localization from bistatic range measurements in distributed mimo radars," *IEEE Signal Processing Letters*, vol. 24, no. 3, pp. 299–303, 2017.

[179] C. Park and J. Chang, "Closed-form localization for distributed mimo radar systems using time delay measurements," *IEEE Transactions on Wireless Communications*, vol. 15, no. 2, pp. 1480–1490, 2016.

[180] S. P. Chepuri and G. Leus, "Sparse sensing for statistical inference," *Foundations and Trends® in Signal Processing*, vol. 9, no. 3–4, pp. 233–368, 2016. [Online]. Available: http://dx.doi.org/10.1561/2000000069

[181] H. Rauhut, "Compressive sensing and structured random matrices," *Theoretical foundations and numerical methods for sparse recovery*, vol. 9, pp. 1–92, 2010.

[182] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[183] M. Rossi, A. M. Haimovich, and Y. C. Eldar, "Spatial compressive sensing for mimo radar," *IEEE Transactions on Signal Processing*, vol. 62, no. 2, pp. 419–430, 2014.

[184] T. Feuillen, A. Mallat, and L. Vandendorpe, "Stepped frequency radar for automotive application: Range-doppler coupling and distortions analysis," in *MILCOM 2016 - 2016 IEEE Military Communications Conference*, 2016, pp. 894–899.

[185] M. Bilodeau and D. Brenner, *Theory of multivariate statistics*. New York: Springer-Verlag, 1999.

[186] J. R. Magnus and H. Neudecker, *Matrix differential calculus with applications in statistics and econometrics*. USA: New York: John Wiley and Sons, Inc., 1988.

[187] C. D. Meyer, *Matrix analysis and applied linear algebra*. Philadelphia: Society for Industrial and Applied Mathematics, 2000.

[188] A. J. Weiss and B. Friedlander, "Performance analysis of spatial smoothing with interpolated arrays," *IEEE Transactions on Signal Processing*, vol. 41, no. 5, pp. 1881–1892, May 1993.

[189] C. Carathéodory and L. Fejér, "Über den zusammenhang der extremen von harmonischen funktionen mit ihren koeffizienten und über den picard-landau'schen satz," *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, vol. 32, no. 1, pp. 218–239, 1911.

[190] T. J. Rothenberg, "Identification in parametric models," *Econometrica: Journal of the Econometric Society*, pp. 577–591, 1971.

[191] G. P. Styan, "Hadamard products and multivariate statistical analysis," *Linear algebra and its applications*, vol. 6, pp. 217–240, 1973.

[192] D. R. Childs, "Reduction of the multivariate normal integral to characteristic form," *Biometrika*, vol. 54, no. 1-2, pp. 293–300, 1967.

[193] C. Gianelli, L. Xu, J. Li, and P. Stoica, "One-bit compressive sampling with time-varying thresholds for sparse parameter estimation," in *IEEE Sensor Array and Multichannel Signal Processing Workshop*, 2016, pp. 1–5.

[194] A. Chatterjee and S. Lahiri, "Strong consistency of lasso estimators," *Sankhya A*, vol. 73, no. 1, pp. 55–78, 2011.

[195] J. Nie and M. Schweighofer, "On the complexity of Putinar's positivstellensatz," *Journal of Complexity*, vol. 23, no. 1, pp. 135–150, 2007.

[196] W. Rudin, *Principles of mathematical analysis*. McGraw-Hill, 1964.