Summer 8-15-2018

Statistical Performance Analysis of Sparse Linear Arrays

Mianzhi Wang
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

School of Engineering and Applied Science

Department of Electrical & Systems Engineering

Dissertation Examination Committee:
Arye Nehorai, Chair
R. Martin Arthur
Matthew Lew
Nan Lin
Joseph O'Sullivan

Statistical Performance Analysis of Sparse Linear Arrays

by

Mianzhi Wang

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2018
Saint Louis, Missouri

Contents

List of Figures

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my research advisor, Dr. Arye Nehorai. Since I joined the lab in August 2014, he has continuously provided guidance and support for my research at Washington University. He opened the door to the field of statistical signal processing for me, and introduced me to a vast world of knowledge I could not imagine before. He has given me great freedom to explore various interesting research topics, and has been always been encouraging and understanding when I encountered obstacles during my pursuit.

I would like to thank my dissertation defense committee members, Dr. Joseph O'Sullivan, Dr. Matthew Lew, Dr. Nan Lin, and Dr. R. Martin Arthur, for their insightful comments on improving my dissertation. I wish to thank Dr. R. Martin Arthur and Dr. Joseph O'Sullivan for also being my course instructors and committee members of my preliminary exam. I also owe thanks to the course instructors at both Fudan University in China and Washington University here in the US, for preparing me with a solid background for my research.

I would like to thank my labmates, Peng, Keyong, Zhao, Jichuan, Alex, Mengxue, Prateek, Zhen, Yijian, Zhenqi, Yiqi, Hesam, and Eric for their inspiration and support. It has been

a joy to collaborate with them and exchange ideas with them in front of the white board in the lab. I also owe thanks to my friends, especially my classmates from Fudan University who are also here in the US, for leaving me countless unforgettable memories.

I would like to thank Mr. James Ballard at the Engineering Communication Center, for spending many hours helping me improve my English writing with great patience.

I am also deeply thankful to my parents for their endless love. I thank them for raising me with a love of science, and for providing continuous support during my hardships.

Mianzhi Wang

Washington University in Saint Louis

August 2018

Dedicated to my parents

ABSTRACT OF THE DISSERTATION

Statistical Performance Analysis of Sparse Linear Arrays

by

Mianzhi Wang

Doctor of Philosophy in Electrical Engineering

Washington University in St. Louis, 2018

Professor Arye Nehorai, Chair

Direction-of-arrival (DOA) estimation remains an important topic in sensor array signal processing. With uniform linear arrays (ULAs), traditional subspace-based methods can resolve only up to $M-1$ sources using M sensors. On the other hand, by exploiting their so-called difference coarray model, sparse linear arrays, such as co-prime and nested arrays, can resolve up to $O(M^2)$ sources using only $O(M)$ sensors. Various new sparse linear array geometries were proposed and many direction-finding algorithms were developed based on sparse linear arrays. However, the statistical performance of such arrays has not been analytically conducted. In this dissertation, we (i) study the asymptotic performance of the MUtiple SIgnal Classification (MUSIC) algorithm utilizing sparse linear arrays, (ii) derive and analyze performance bounds for sparse linear arrays, and (iii) investigate the robustness of sparse linear arrays in the presence of array imperfections. Based on our analytical results, we also propose robust direction-finding algorithms for use when data are missing.

We begin by analyzing the performance of two commonly used coarray-based MUSIC direction estimators. Because the coarray model is used, classical derivations no longer apply. By using an alternative eigenvector perturbation analysis approach, we derive a closed-form expression of the asymptotic mean-squared error (MSE) of both estimators. Our expression is computationally efficient compared with the alternative of Monte Carlo simulations. Using this expression, we show that when the source number exceeds the sensor number, the MSE remains strictly positive as the signal-to-noise ratio (SNR) approaches infinity. This finding theoretically explains the unusual "saturation" behavior of coarray-based MUSIC estimators that had been observed in previous studies.

We next derive and analyze the Cramér-Rao bound (CRB) for general sparse linear arrays under the assumption that the sources are uncorrelated. We show that, unlike the classical stochastic CRB, our CRB is applicable even if there are more sources than the number of sensors. We also show that, in such a case, this CRB remains strictly positive definite as the SNR approaches infinity. This unusual behavior imposes a strict lower bound on the variance of unbiased DOA estimators in the underdetermined case. We establish the connection between our CRB and the classical stochastic CRB and show that they are asymptotically equal when the sources are uncorrelated and the SNR is sufficiently high. We investigate the behavior of our CRB for co-prime and nested arrays with a large number of sensors, characterizing the trade-off between the number of spatial samples and the number of temporal samples. Our analytical results on the CRB will benefit future research on optimal sparse array designs.

We further analyze the performance of sparse linear arrays by considering sensor location errors. We first introduce the deterministic error model. Based on this model, we derive a closed-form expression of the asymptotic MSE of a commonly used coarray-based MUSIC

estimator, the spatial-smoothing based MUSIC (SS-MUSIC). We show that deterministic sensor location errors introduce a constant estimation bias that cannot be mitigated by only increasing the SNR. Our analytical expression also provides a sensitivity measure against sensor location errors for sparse linear arrays. We next extend our derivations to the stochastic error model and analyze the Gaussian case. We also derive the CRB for joint estimation of DOA parameters and deterministic sensor location errors. We show that this CRB is applicable even if there are more sources than the number of sensors.

Lastly, we develop robust DOA estimators for cases with missing data. By exploiting the difference coarray structure, we introduce three algorithms to construct an augmented covariance matrix with enhanced degrees of freedom. By applying MUSIC to this augmented covariance matrix, we are able to resolve more sources than sensors. Our method utilizes information from all snapshots and shows improved estimation performance over traditional DOA estimators.

Chapter 1

Introduction

Direction-of-arrival (DOA) estimation is an important topic in array signal processing, with wide applications in radar, sonar, audio and speech processing, geophysics, and communications [1–4]. To estimate the DOAs of the impinging signals, sensor arrays are deployed to collect spatial samples of these source signals. Given a sufficient number of samples, various algorithms can be applied to obtain the DOAs. In general, these algorithms can be divided into spectral-based algorithms and parametric model based algorithms. Spectral-based algorithms include conventional beamforming-based algorithms [5–8], and MUtiple SIgnal Classification (MUSIC) [9]. Parametric model based algorithms usually obtains the DOAs by solving maximum-likelihood (ML) problems. Typical ML-based algorithms include the conditional maximum likelihood estimator, the stochastic maximum-likelihood estimator, weighed-subspace fitting and their variants [3, 10–20]. There are also parametric model based algorithms that utilize the signal subspace, such as root-MUSIC [21–23], and the estimation of signal parameters via rotational invariance techniques algorithm (ESPRIT) [24, 25]. Recently, with the development of compressed sensing theory [26, 27], new DOA estimation methods have been developed based on sparse recovery [28–31], sparse Bayesian learning [32, 33], and super-resolution theory [34, 35].

The development of various DOA estimation algorithms is accompanied by the consideration of new linear array geometries. Traditionally, a uniform linear array (ULA) is deployed to uniformly sample the source signals in spaces. However, with conventional subspace-based methods, an M-sensor ULA can resolve only up to $O(M-1)$ sources [9, 15]. Additionally, many sensors are required to cover a large aperture. To tackle these issues, the concept of sparse linear arrays was developed. By utilizing their so-called difference coarray model, up to $O(M^2)$ uncorrelated sources can be resolved using only M sensors. In [36] and [37], the authors introduced minimum redundancy arrays (MRAs). However, these arrays do not have closed-form expressions for their geometries and cannot easily be generalized when the number of sensors is large. Recently, Pal and Vaidyanathan introduced nested arrays [38] and co-prime arrays [39], both of which have closed-form expressions and can resolve up to $O(M^2)$ uncorrelated sources using only M sensors. The introduction of nested and co-prime arrays has generated renewed interest in sparse linear arrays [31,40–44], leading to new sparse linear array geometries such as generalized co-prime arrays [45], and super nested arrays [46, 47]. There have also been extensions to 2D arrays [48, 49] and vector-sensor arrays [50, 51].

With the introduction of new sparse linear arrays, it became important to statistically analyze their performance. Previous performance analysis of such arrays relies on numerical simulations, which are computationally expensive. Most of the existing analytical performance analysis approaches are based on the array model of ULAs, and cannot be readily extended to the difference coarray model [7, 21, 25, 52, 53]. In this dissertation, we mainly focus on the statistical performance analysis of sparse linear arrays. By developing such a statistical performance analysis framework, we are able to gain more insights of the performance of sparse linear arrays without computationally expensive simulations. In practice, arrays will not always be perfectly calibrated, and various perturbations will exist. Our

analysis covers both the perturbation-free case and the perturbed case. Based on our theoretical results, we also develop robust DOA estimation algorithms for sparse linear arrays that utilize the difference coarray model.

1.1 Contributions of this work

In this dissertation, we provide a thorough statistical performance analysis of sparse linear arrays, and develop robust direction finding algorithms. Our contributions can be summarized as follows:

Performance analysis of coarray-based MUSIC algorithms: Coarray-based MUSIC algorithms resolve the DOAs by applying the classical MUSIC algorithm to the augmented covariance matrix constructed according to the difference coarray model. We investigate two common methods of constructing such an augmented covariance matrix, namely, the direct augmentation based approach (DAA) [54,55] and the spatial smoothing based approach [38]. We show that MUSIC yields the same asymptotic estimation error for both methods. Based on this finding, we are the first to derive a closed-form asymptotic mean-squared error (MSE) expression that is applicable to both methods. This expression is more computationally efficient than traditional Monte Carlo simulations, and facilitates the performance analysis of coarray-based MUSIC algorithms. Using this expression, we show that, when there are more sources than the number of sensors, the asymptotic MSE does not drop to zero even if the SNR approaches infinity. This result theoretically explains the "saturation" behavior of the coarray-based MUSIC algorithms in high SNR regions observed in previous studies.

Analyses of Cramér-Rao bounds (CRBs) for general sparse linear arrays: The CRB gives the lower bound on the minimum variance any unbiased estimator can achieve. The classical stochastic CRB for general linear arrays was derived and analyzed by Stoica et at. [52, 56]. This CRB is derived without the assumption that sources are uncorrelated, and does not exist when the number of sources exceeds the number of sensors. We derive the CRB for general sparse linear arrays under the assumption that the sources are uncorrelated, which is applicable even if the number of sources is greater than the number of sensors. We show that in high SNR regions our CRB is asymptotically equivalent to the classical stochastic CRB for uncorrelated sources. We also show that, when there are more sources than the number of sensors, our CRB is strictly nonzero as the SNR goes to infinity. We further analyze the behavior of our CRB for co-prime and nested arrays with a large number of sensors. We show that this CRB can decrease at a rate of $O(M^{-5})$ for large values of M for co-prime and nested arrays, but this rate is only $O(M^{-3})$ for an M-sensor ULA. This finding analytically demonstrates that co-prime and nested arrays can achieve better estimation performance when the number of sensors is a limiting factor. We also show that for a fixed aperture, co-prime and nested arrays require many more snapshots to achieve the same performance as ULAs. This finding illustrates the trade-off between the number of spatial samples and the number of temporal samples.

Perturbation analysis of the difference coarray model: The above results are based on the assumption that the arrays are perfectly calibrated. However, array imperfections exist and the difference coarray model may be perturbed. We introduce a signal model for sparse linear arrays in the presence of deterministic unknown location errors. Based on this signal model, we derive a closed-form expression of the asymptotic MSE of coarray-based MUSIC algorithms. With this expression, we show that the sensor location errors introduce a constant bias depending on both the physical array geometry and the coarray geometry,

which cannot be mitigated by only increasing the SNR. We also extend our analysis to cases when the sensor location errors are stochastic, and we investigate the Gaussian case. Additionally, we derive the Cramér-Rao bound for joint estimation of DOAs and sensor location errors for sparse linear arrays, which can be applicable even if the number of sources exceeds the number of sensors.

Direction finding in the presence of missing data: We investigate the problem of robust DOA estimation using sparse linear arrays in the case of missing data resulting from sensor failures. We introduce a signal model where sensor failures occur after a certain number of snapshots. Based on our signal model, we formulate a structured covariance estimation problem by exploiting the special geometry of sparse linear arrays. By utilizing the information in both complete measurements and incomplete measurements, our method achieves better estimation accuracy than the traditional method using only complete measurements. We also derive the CRB in the missing data case.

1.2 Organization of this dissertation

The rest of this dissertation is organized as follows. In Chapter 2, we present the background for sparse linear arrays, the concept of the difference coarray model, and the coarray-based MUSIC. In Chapter 3, we conduct detailed statistical performance analyses of sparse linear arrays. We first derive and analyze the asymptotic mean-squared error (MSE) for two commonly used coarray-based MUSIC algorithms, and then derive and analyze the CRB. In Chapter 4, we investigate the effect of sensor location errors on coarray-based MUSIC algorithms and the achievable performance bounds. Then in Chapter 5, we introduce a robust direction finding algorithm in the case of missing data resulting from sensor failures

and derive the corresponding CRB. Finally, in Chapter 6, we draw conclusions and propose potential future directions.

1.3 Notations

Given a matrix $\boldsymbol{A}$, we use $\boldsymbol{A}^T$, $\boldsymbol{A}^H$, and $\boldsymbol{A}^*$ to denote the transpose, the Hermitian transpose, and the conjugate of $\boldsymbol{A}$, respectively. We use A_{ij} to denote the (i,j)-th element of $\boldsymbol{A}$, and $\boldsymbol{a}_i$ to denote the i-th column of $\boldsymbol{A}$. If $\boldsymbol{A}$ is full column rank, we define its pseudo inverse as $\boldsymbol{A}^\dagger = (\boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{A}^H$. We also define the projection matrix onto the null space of $\boldsymbol{A}$ as $\boldsymbol{\Pi}_{\boldsymbol{A}}^\perp = \boldsymbol{I} - \boldsymbol{A}\boldsymbol{A}^\dagger$. Let $\boldsymbol{A} = [\boldsymbol{a}_1\, \boldsymbol{a}_2\, \ldots\, \boldsymbol{a}_N] \in \mathbb{C}^{M\times N}$, and we define the vectorization operation as $\mathrm{vec}(\boldsymbol{A}) = [\boldsymbol{a}_1^T\, \boldsymbol{a}_2^T\, \ldots\, \boldsymbol{a}_N^T]^T$, and $\mathrm{mat}_{M,N}(\cdot)$ as its inverse operation. We use $\otimes$, $\odot$, and $\circ$ to denote the Kronecker product, the Khatri-Rao product (i.e., the column-wise Kronecker product), and the Hadamard product (i.e., the element-wise product), respectively. We denote by $\Re(\boldsymbol{A})$ and $\Im(\boldsymbol{A})$ the real and the imaginary parts of $\boldsymbol{A}$. If $\boldsymbol{A}$ is a square matrix, we denote its trace by $\mathrm{tr}(\boldsymbol{A})$.

Chapter 2

Direction Finding Using Sparse Linear Arrays

In this chapter, we provide a comprehensive overview of direction finding methods utilizing sparse linear arrays. We first introduce the definition of sparse linear arrays and their relationship with ULAs. We then introduce the difference coarray model of general sparse linear arrays. Finally, we introduce the MUSIC algorithm and show how it can be applied to the difference coarray model to identify the DOAs. In this chapter, we assume that the arrays are perfectly calibrated.

2.1 ULAs and sparse linear arrays

Without loss of generality, we assume that the first sensor is placed at the origin. Let d_0 denote the smallest inter-element spacing. An M-sensor ULA is defined as follows:

Definition 2.1. *An M-sensor ULA is given by $\{0, 1, \ldots, M-1\}d_0$.*

Sparse linear arrays can be constructed by strategically removing sensors from ULAs. Typical sparse linear arrays include MRAs [36], co-prime arrays [39], and nested arrays [38]. Fig. 2.1 shows examples of a ULA plus three different types of sparse linear arrays. We can observe that all three sparse linear arrays, below the ULA, can be constructed by removing certain sensors from the ULA on top.

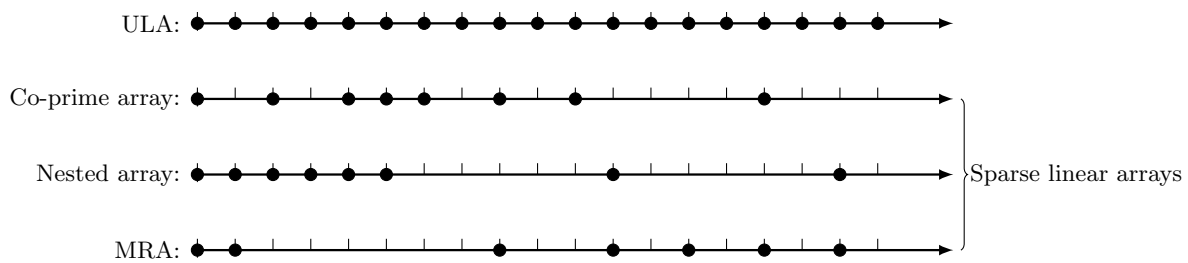


Figure 2.1: Examples of a ULA and three different types of sparse linear arrays.

MRAs do not have closed-form expressions, and a list of MRAs can be found in [37]. The definitions of co-prime[1] and nested arrays are stated in Definition 2.2 and Definition 2.3.

Definition 2.2. *A co-prime array generated by the co-prime pair* (N_1, N_2) *is given by* $\{0, N_1, \ldots, (N_2-1)N_1\}d_0 \cup \{N_2, 2N_2, \ldots, (2N_1-1)N_2\}d_0$.

Definition 2.3. *A nested array generated by the parameter pair* (N_1, N_2) *is given by* $\{0, 1, \ldots, N_1-1\}d_0 \cup \{N_1, 2N_1+1, \ldots, N_2N_1+N_2-1\}d_0$.

The structures of co-prime arrays and nested arrays are illustrated in Fig. 2.2 and Fig. 2.3. Both co-prime arrays and nested arrays consist of two subarrays with different inter-element spacings.

[1] In fact, given a co-prime pair (M, N), there are two difference co-prime array configurations, namely the "M" configuration and the "$2M$" configuration [39]. Throughout this dissertation, we will consider only the "$2M$" configuration, which is stated in Definition 2.2.

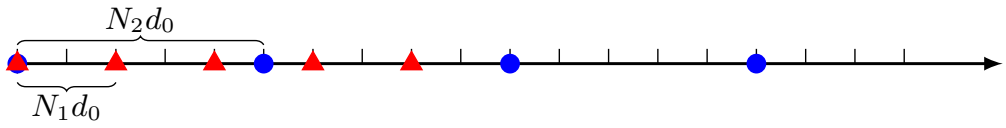


Figure 2.2: Illustration of the structure of a co-prime array. The red triangles represent the first subarray with inter-element spacing $N_1 d_0$, while the blue circles represent the second subarray with inter-element spacing $N_2 d_0$.

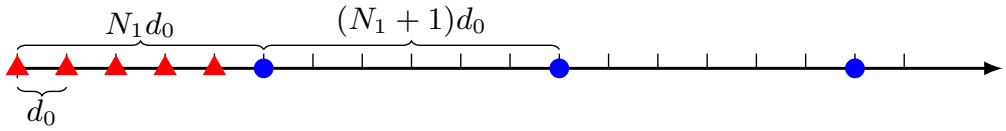


Figure 2.3: Illustration of the structure of a nested array. The red triangles represent the first subarray with inter-element spacing d_0, while the blue circles represent the second subarray with inter-element spacing $(N_1 + 1)d_0$.

2.2 Signal model

In this section, we introduce the stochastic signal model and the difference coarray model of sparse linear arrays.

2.2.1 The stochastic signal model

We consider a sparse linear array consisting of M sensors whose locations are given by $\mathcal{D} = \{d_1, d_2, \ldots, d_M\}$. Each sensor location d_i is chosen to be an integer multiple of d_0. Therefore we can also represent the sensor locations using the integer set $\bar{\mathcal{D}} = \{\bar{d}_1, \bar{d}_2, \ldots, \bar{d}_M\}$, where $\bar{d}_i = d_i/d_0$ for $i = 1, 2, \ldots, M$. Without loss of generality, we assume that $d_1 = 0$.

We consider K narrow-band sources $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_K]^T$ impinging on the array from the far field. Denoting λ as the wavelength of the carrier frequency, we can express the steering

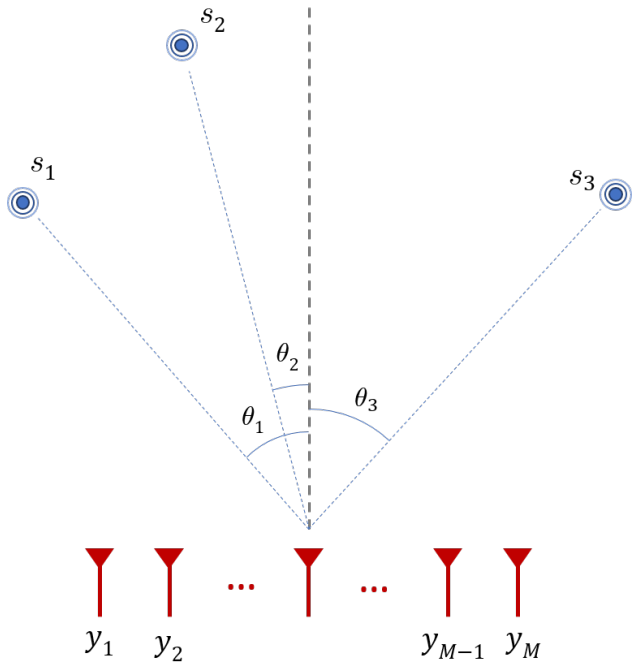


Figure 2.4: Illustration of our signal model. s_1, s_2, and s_3 denote three far-field narrow-band sources whose DOAs are given by θ_1, θ_2, and θ_3, respectively. $y_1, \ldots, y_M$ represent the sensor output of a linear array.

vector for the k-th source as

$$\boldsymbol{a}(\theta_k) = \begin{bmatrix} 1 & \exp\left(j\frac{2\pi d_2}{\lambda}\sin\theta_k\right) & \ldots & \exp\left(j\frac{2\pi d_M}{\lambda}\sin\theta_k\right) \end{bmatrix}^T, \tag{2.1}$$

or equivalently,

$$\boldsymbol{a}(\omega_k) = \begin{bmatrix} 1 & e^{j\bar{d}_2\omega_k} & \ldots & e^{j\bar{d}_M\omega_k} \end{bmatrix}^T, \tag{2.2}$$

where $\omega_k = (2\pi d_0 \sin\theta_k)/\lambda$. Because a one-to-one mapping exists between ω_k and θ_k for every $\theta_k \in (-\pi/2, \pi/2)$, there is no loss of information if we represent the DOAs using $\boldsymbol{\omega} = [\omega_1,\ \omega_2, \ldots, \omega_K]^T$ instead of $\boldsymbol{\theta}$. Typically, d_0 is chosen to be $\lambda/2$, and we have $\omega_k \in (-\pi, \pi)$.

The received signal vectors are given by [3]

$$\underbrace{\boldsymbol{y}(t)}_{M\times 1} = \underbrace{\boldsymbol{A}(\theta)}_{M\times K}\underbrace{\boldsymbol{x}(t)}_{K\times 1} + \underbrace{\boldsymbol{n}(t)}_{M\times 1}, t = 1, 2, \cdots, N, \tag{2.3}$$

where $\boldsymbol{A} = [\boldsymbol{a}(\theta_1)\, \boldsymbol{a}(\theta_2)\, \cdots\, \boldsymbol{a}(\theta_K)]$ denotes the array steering matrix, $\boldsymbol{x}(t)$ denotes the source signal vector, $\boldsymbol{n}(t)$ denotes additive noise, and N denotes the number of snapshots. Fig. 2.4 shows our signal model.

In the stochastic signal model, the source signals are assumed to be random and unknown [52]. In the following discussion, we make the following assumptions:

A1 The source signals are uncorrelated, and follow a zero-mean circularly-symmetric complex Gaussian distribution.

A2 The source DOAs are distinct (i.e., $\theta_k \neq \theta_l\ \forall k \neq l$).

A3 The additive noise is circularly-symmetric complex, white, and uncorrelated with the sources.

A4 The is no temporal correlation between snapshots.

Under **A1**–**A4**, the received snapshots follow a circularly-symmetric complex Gaussian distribution whose mean is zero and whose covariance matrix is given by

$$\boldsymbol{R} = \boldsymbol{A}\boldsymbol{P}\boldsymbol{A}^H + \sigma^2\boldsymbol{I}, \tag{2.4}$$

where $\boldsymbol{P} = \mathrm{diag}(p_1, p_2, \ldots, p_K)$ denotes the source covariance matrix, and σ^2 denotes the variance of the additive noise. In practice, only a finite number of snapshots are available,

and $\boldsymbol{R}$ is estimated via

$$\hat{\boldsymbol{R}} = \frac{1}{N}\sum_{t=1}^{N} \boldsymbol{y}(t)\boldsymbol{y}^H(t). \tag{2.5}$$

We can then apply various direction finding algorithms (e.g., MUSIC) to $\hat{\boldsymbol{R}}$ to obtain the DOAs.

2.2.2 The difference coarray model

Based on (2.4), we now introduce the difference coarray model. By vectorizing $\boldsymbol{R}$ in (2.4), we obtain that

$$\boldsymbol{r} := \mathrm{vec}(\boldsymbol{R}) = \boldsymbol{A}_{\mathrm{d}}\boldsymbol{p} + \sigma^2\boldsymbol{i}, \tag{2.6}$$

where $\boldsymbol{A}_{\mathrm{d}} = \boldsymbol{A}^* \odot \boldsymbol{A}$, $\boldsymbol{p} = [p_1, p_2, \ldots, p_K]^T$, and $\boldsymbol{i} = \mathrm{vec}(\boldsymbol{I})$. From the observation that

$$\boldsymbol{A}_{\mathrm{d}} = \begin{bmatrix} e^{j(\bar{d}_1-\bar{d}_1)\omega_1} & \ldots & e^{j(\bar{d}_1-\bar{d}_1)\omega_K} \\ \vdots & \ddots & \vdots \\ e^{j(\bar{d}_m-\bar{d}_n)\omega_1} & \ldots & e^{j(\bar{d}_m-\bar{d}_n)\omega_K} \\ \vdots & \ddots & \vdots \\ e^{j(\bar{d}_M-\bar{d}_M)\omega_1} & \ldots & e^{j(\bar{d}_M-\bar{d}_M)\omega_K} \end{bmatrix}_{M^2\times K}, \tag{2.7}$$

we know that $\boldsymbol{A}_{\mathrm{d}}$ corresponds to the steering matrix of the coarray whose sensor locations are given by $\mathcal{D}_{\mathrm{co}} = \{d_m - d_n | 1 \le m, n \le M\}$. By carefully combining repeated rows of $(\boldsymbol{A}^* \odot \boldsymbol{A})$, we can construct a new steering matrix representing a virtual ULA with enhanced degrees of freedom. Because $\mathcal{D}_{\mathrm{co}}$ is symmetric, this virtual ULA is centered at the origin. The sensor locations of the virtual ULA are given by $[-M_{\mathrm{co}}+1, -M_{\mathrm{co}}+2, \ldots, 0, \ldots, M_{\mathrm{co}}-1]d_0$, where M_{co} is defined so that $2M_{\mathrm{co}}-1$ is the size of the virtual ULA. Fig. 2.5 provides an illustrative example of the relationship between the physical array and the corresponding

virtual ULA. To capture this combination process, we need to introduce the definition of a *coarray selection matrix*.

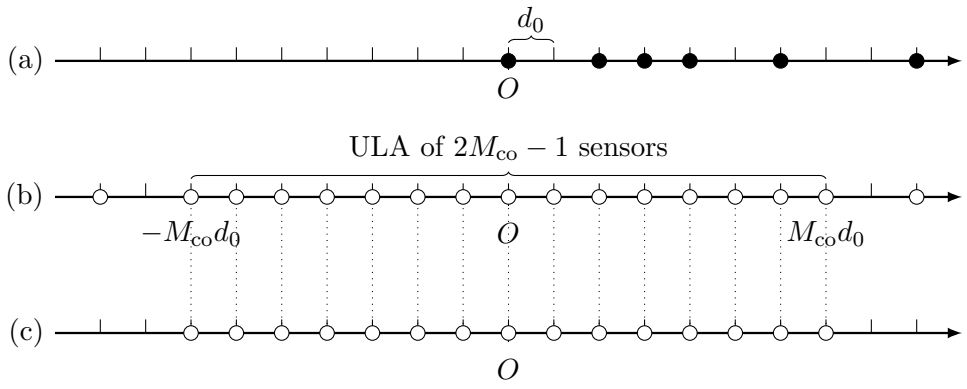


Figure 2.5: (a) A co-prime array with sensors located at $[0, 2, 3, 4, 6, 9]d_0$; (b) its difference coarray; (c) central ULA part of the difference coarray.

According to (2.4),

$$R_{mn} = \sum_{k=1}^{K} p_k \exp[j(\bar{d}_m - \bar{d}_n)\omega_k] + \delta_{mn}\sigma^2,$$

where δ_{mn} denotes Kronecker's delta. This equation implies that the (m, n)-th element of $\boldsymbol{R}$ is associated with the difference $(\bar{d}_m - \bar{d}_n)$. To capture this property, we introduce the difference matrix $\boldsymbol{\Delta}$, such that $\Delta_{mn} = \bar{d}_m - \bar{d}_n$. We also define the *weight function* as follows:

Definition 2.4. *The weight function $w(n) : \mathbb{Z} \mapsto \mathbb{Z}$ is defined as [38]*

$$w(l) = |\{(m, n)|\Delta_{mn} = l\}|,$$

where $|\mathcal{A}|$ denotes the cardinality of the set $\mathcal{A}$.

Intuitively, $w(l)$ counts the number of all possible pairs of $(\bar{d}_m, \bar{d}_n)$ such that $\bar{d}_m - \bar{d}_n = l$. Clearly, $w(l) = w(-l)$. With the definition of the weight function, we can formally introduce the definition of the coarray selection matrix as follows:

Definition 2.5. *The coarray selection matrix* $\boldsymbol{F}$ *is a* $(2M_{\text{co}}-1)\times M^2$ *matrix satisfying*

$$F_{m,p+(q-1)M}=\begin{cases}\frac{1}{w(m-M_{\text{co}})} & ,\Delta_{pq}=m-M_{\text{co}},\\ 0 & ,\text{otherwise},\end{cases} \tag{2.8}$$

for $m=1,2,\ldots,2M_{\text{co}}-1, p=1,2,\ldots,M, q=1,2,\ldots,M$.

To better illustrate the construction of $\boldsymbol{F}$, we consider a toy array whose sensor locations are given by $\{0,d_0,4d_0\}$. The corresponding difference matrix of this array is

$$\boldsymbol{\Delta}=\begin{bmatrix}0 & -1 & -4\\ 1 & 0 & -3\\ 4 & 3 & 0\end{bmatrix}.$$

The ULA part of the difference coarray consists of three sensors located at $-d_0$, 0, and d_0. The weight function satisfies $w(-1)=w(1)=1$, and $w(0)=3$, so $M_{\text{co}}=2$. We can write the coarray selection matrix as

$$\boldsymbol{F}=\begin{bmatrix}0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0\\ \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3}\\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\end{bmatrix}.$$

If we pre-multiply the vectorized sample covariance matrix $\boldsymbol{r}$ by $\boldsymbol{F}$, we obtain

$$\boldsymbol{z}=\begin{bmatrix}z_1\\ z_2\\ z_3\end{bmatrix}=\begin{bmatrix}R_{12}\\ \frac{1}{3}(R_{11}+R_{22}+R_{33})\\ R_{21}\end{bmatrix}.$$

It can be seen that z_m is obtained by averaging all the elements in $\boldsymbol{R}$ that correspond to the difference $m - M_{\text{co}}$, for $m = 1, 2, \ldots, 2M_{\text{co}} - 1$. This process is also referred to as redundancy averaging [57]. To provide a more intuitive understanding, we have illustrated this process in Fig. 2.6.

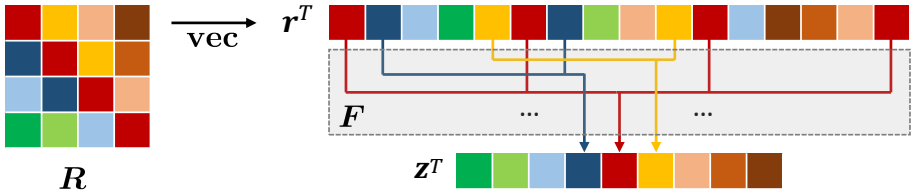


Figure 2.6: An illustration of the redundancy averaging process: (1) $\boldsymbol{R}$ is first vectorized into $\boldsymbol{r}$; (2) elements in $\boldsymbol{r}$ are grouped according to the difference matrix $\boldsymbol{\Delta}$; (3) $\boldsymbol{z}$ is constructed by averaging the elements in each group. Elements R_{mn} having the same Δ_{mn} share the same color. For example, the diagonal elements of $\boldsymbol{R}$ share the same red color because $\Delta_{mm} = 0$ for all m.

Based on Definition 2.5, we now derive several useful properties of $\boldsymbol{F}$.

Property 2.1. *$F_{m,p+(q-1)M} = F_{2M_{\text{co}}-m,q+(p-1)M}$ for $m = 1, 2, \ldots, 2M_{\text{co}}-1, p = 1, 2, \ldots, M, q = 1, 2, \ldots, M$.*

Proof. If $F_{m,p+(q-1)M} = 0$, then $\Delta_{pq} \neq m - M_{\text{co}}$. Because $\Delta_{qp} = -\Delta_{pq}$, $\Delta_{qp} \neq -(m - M_{\text{co}})$. Hence $(2M_{\text{co}} - m) - M_{\text{co}} = -(m - M_{\text{co}}) \neq \Delta_{qp}$, which implies that $F_{2M_{\text{co}}-m,q+(p-1)M}$ is also zero.

If $F_{m,p+(q-1)M} \neq 0$, then $\Delta_{pq} = m - M_{\text{co}}$ and $F_{m,p+(q-1)M} = 1/w(m - M_{\text{co}})$. Note that $(2M_{\text{co}} - m) - M_{\text{co}} = -(m - M_{\text{co}}) = -\Delta_{pq} = \Delta_{qp}$. We thus have $F_{2M_{\text{co}}-m,q+(p-1)M} = 1/w(-(m - M_{\text{co}})) = 1/w(m - M_{\text{co}}) = F_{m,p+(q-1)M}$. □

Property 2.2. *Let $\boldsymbol{R} \in \mathbb{C}^M$ be Hermitian symmetric. Then $\boldsymbol{z} = \boldsymbol{F}\,\text{vec}(\boldsymbol{R})$ is conjugate symmetric.*

Proof. By Property 2.1 and $\boldsymbol{R} = \boldsymbol{R}^H$,

$$z_m = \sum_{p=1}^{M}\sum_{q=1}^{M} F_{m,p+(q-1)M} R_{pq} = \sum_{q=1}^{M}\sum_{p=1}^{M} F_{2M_{\mathrm{co}}-m,q+(p-1)M} R_{qp}^* = z_{2M_{\mathrm{co}}-m}^*.$$

□

Property 2.3. *Let $\boldsymbol{z} \in \mathbb{C}^{2M_{\mathrm{co}}-1}$ be conjugate symmetric. Then $\mathrm{mat}_{M,M}(\boldsymbol{F}^T\boldsymbol{z})$ is Hermitian symmetric.*

Proof. Let $\boldsymbol{H} = \mathrm{mat}_{M,M}(\boldsymbol{F}^T\boldsymbol{z})$. Then

$$H_{pq} = \sum_{m=1}^{2M_{\mathrm{co}}-1} z_m F_{m,p+(q-1)M}. \tag{2.9}$$

We know that $\boldsymbol{z}$ is conjugate symmetric, so $z_m = z_{2M_{\mathrm{co}}-m}^*$. Therefore, by Lemma 2.1

$$H_{pq} = \sum_{m=1}^{2M_{\mathrm{co}}-1} z_{2M_{\mathrm{co}}-m}^* F_{2M_{\mathrm{co}}-m,q+(p-1)M} = \left[\sum_{m'=1}^{2M_{\mathrm{co}}-1} z_{m'} F_{m',q+(p-1)M}\right]^* = H_{qp}^*. \tag{2.10}$$

□

From definition 2.5, the observation vector of the virtual ULA is given by

$$\underbrace{\boldsymbol{z}}_{(2M_{\mathrm{co}}-1)\times 1} = \underbrace{\boldsymbol{F}}_{(2M_{\mathrm{co}}-1)\times M^2}\ \underbrace{\boldsymbol{r}}_{M^2\times 1} = \underbrace{\boldsymbol{A}_{\mathrm{c}}}_{(2M_{\mathrm{co}}-1)\times K}\ \underbrace{\boldsymbol{p}}_{K\times 1} + \underbrace{\sigma^2\boldsymbol{F}\boldsymbol{i}}_{(2M_{\mathrm{co}}-1)\times 1}, \tag{2.11}$$

where $\boldsymbol{A}_{\mathrm{c}}$ represents the steering matrix of the virtual ULA whose sensors are located at $[-M_{\mathrm{co}}+1, \ldots, 0, \ldots, M_{\mathrm{co}}-1]d_0$. We refer to (2.11) as the *difference coarray model*[2]. In

[2]For brevity, we will use the terms *difference coarray model* and *coarray model* interchangeably in the following discussion.

practice, only a finite number of snapshots are available, and $\boldsymbol{z}$ is replaced with its estimate $\hat{\boldsymbol{z}} = \boldsymbol{F}\hat{\boldsymbol{r}}$, where $\hat{\boldsymbol{r}} = \mathrm{vec}(\hat{\boldsymbol{r}})$.

2.3 Direction finding using MUSIC

MUSIC is a classical subspace-based DOA estimation algorithm first introduced by Schmidt [9]. In this section, we first provide a brief review of the MUSIC algorithm, and then extend it to the difference coarray model and introduce two commonly used coarray-based MUSIC algorithms.

2.3.1 Direct MUSIC

Recall that the covariance matrix of $\boldsymbol{y}(t)$ is given by

$$\boldsymbol{R} = \boldsymbol{A}\boldsymbol{P}\boldsymbol{A}^H + \sigma^2\boldsymbol{I}. \tag{2.12}$$

Assuming that both $\boldsymbol{A}$ and $\boldsymbol{P}$ are full column rank, then $\boldsymbol{A}\boldsymbol{P}\boldsymbol{A}^H$ will be a rank-K matrix. If we perform eigendecomposition over $\boldsymbol{R}$, the eigenvectors corresponding to the first K largest eigenvalues will span the same subspace as $\boldsymbol{A}$, which we call the *signal subspace*. The last $M-K$ eigenvalues will all equal to σ^2, and their corresponding eigenvectors span the *noise subspace*. In other words, we can rewrite the eigendecomposition of $\boldsymbol{R}$ as

$$\boldsymbol{R} = \boldsymbol{E}_{\mathrm{s}}\boldsymbol{\Lambda}_{\mathrm{s}}\boldsymbol{E}_{\mathrm{s}}^H + \sigma^2\boldsymbol{E}_{\mathrm{n}}\boldsymbol{E}_{\mathrm{n}}^H, \tag{2.13}$$

where $\boldsymbol{E}_{\mathrm{s}}$ and $\boldsymbol{\Lambda}_{\mathrm{s}}$ denote the eigenvectors and eigenvalues for the signal subspace, respectively. $\boldsymbol{\Lambda}_{\mathrm{n}}$ denotes the eigenvectors for the noise subspace. Because the two subspaces are orthogonal, if $\theta \in \{\theta_1, \theta_2, \ldots, \theta_K\}$, we must have $\boldsymbol{a}^H(\theta)\boldsymbol{E}_{\mathrm{n}}\boldsymbol{E}_{\mathrm{n}}^H = \boldsymbol{0}$. Therefore we can find the DOAs by searching for $\theta \in (-\pi/2, \pi/2)$ that minimizes $\boldsymbol{a}^H(\theta)\boldsymbol{E}_{\mathrm{n}}\boldsymbol{E}_{\mathrm{n}}^H\boldsymbol{a}(\theta)$.

Given the estimated covariance matrix $\hat{\boldsymbol{R}}$, its eigendecomposition can be expressed as

$$\hat{\boldsymbol{R}} = \hat{\boldsymbol{E}}_{\mathrm{s}}\hat{\boldsymbol{\Lambda}}_{\mathrm{s}}\hat{\boldsymbol{E}}_{\mathrm{s}}^H + \hat{\boldsymbol{E}}_{\mathrm{n}}\hat{\boldsymbol{\Lambda}}_{\mathrm{n}}\hat{\boldsymbol{E}}_{\mathrm{n}}^H. \tag{2.14}$$

Following the above reasoning, the MUSIC pseudo-spectrum is defined as

$$P_{\mathrm{MUSIC}}(\theta) = \frac{1}{\boldsymbol{a}(\theta)^H\hat{\boldsymbol{E}}_{\mathrm{n}}\hat{\boldsymbol{E}}_{\mathrm{n}}^H\boldsymbol{a}(\theta)}, \tag{2.15}$$

and the DOAs can be identified by performing a grid search and finding the peaks in the resulting pseudo-spectrum.

Remark. The MUSIC algorithm requires knowing the number of sources. In the literature of array signal processing, there are various source number detection methods, such as the Akaike information criterion (AIC) [58], Rissanen's minimum description length (MDL) [59], and the second order statistic of eigenvalues (SORTE) [60]. Therefore, we assume that the number of sources K is known when conducting the performance analysis of MUSIC-based direction finding algorithms.

2.3.2 Coarray-based MUSIC

Because $\boldsymbol{R}$ is an $M \times M$ matrix, we can resolve only up to $M-1$ sources using the MUSIC algorithm. In Fig. 2.5, we observe that the central ULA part has more virtual sensors than

physical sensors in the original array. If we can apply MUSIC to the difference coarray model, we can gain more degrees of freedom and resolve more DOAs. However, we effectively have only one snapshot from the difference coarray model: $\hat{\boldsymbol{z}}$. Therefore, $\hat{\boldsymbol{z}}\hat{\boldsymbol{z}}^H$ will be a rank-one matrix and MUSIC cannot be applied.

To apply MUSIC to the difference coarray model, we need to construct an augmented covariance matrix. We observe that the virtual ULA can be divided into M_{co} overlapping uniform subarrays of size M_{co}. The output of the i-th subarray is given by $\hat{\boldsymbol{z}}_i = \boldsymbol{\Gamma}_i\hat{\boldsymbol{z}}$ for $i = 1, 2, \ldots, M_{\text{co}}$, where $\boldsymbol{\Gamma}_i = [\boldsymbol{0}_{M_{\text{co}}\times(i-1)}\, \boldsymbol{I}_{M_{\text{co}}\times M_{\text{co}}}\, \boldsymbol{0}_{M_{\text{co}}\times(M_{\text{co}}-i)}]$ represents the selection matrix for the i-th subarray.

Given the outputs of the M_{co} subarrays, the augmented covariance matrix is commonly constructed via one of the following methods [38, 55]:

$$\hat{\boldsymbol{R}}_{\text{v1}} = [\hat{\boldsymbol{z}}_{M_{\text{co}}}\, \hat{\boldsymbol{z}}_{M_{\text{co}}-1}\, \cdots\, \hat{\boldsymbol{z}}_1], \tag{2.16a}$$

$$\hat{\boldsymbol{R}}_{\text{v2}} = \frac{1}{M_{\text{co}}}\sum_{i=1}^{M_{\text{co}}} \hat{\boldsymbol{z}}_i\hat{\boldsymbol{z}}_i^H, \tag{2.16b}$$

where method (2.16a) corresponds to DAA , while method (2.16b) corresponds to the spatial smoothing approach. Here $\hat{\boldsymbol{R}}_{\text{v1}}$ and $\hat{\boldsymbol{R}}_{\text{v2}}$ are estimates of their true versions, $\boldsymbol{R}_{\text{v1}}$ and $\boldsymbol{R}_{\text{v2}}$.

Following the results in [38] and [55], $\boldsymbol{R}_{\text{v1}}$ and $\boldsymbol{R}_{\text{v2}}$ are related via the following equality:

$$\boldsymbol{R}_{\text{v2}} = \frac{1}{M_{\text{co}}}\boldsymbol{R}_{\text{v1}}^2 = \frac{1}{M_{\text{co}}}(\boldsymbol{A}_{\text{co}}\boldsymbol{P}\boldsymbol{A}_{\text{co}}^H + \sigma^2\boldsymbol{I})^2, \tag{2.17}$$

where $\boldsymbol{A}_{\text{co}}$ corresponds to the steering matrix of a ULA whose sensors are located at $[0, 1, \ldots,\ M_{\text{co}} - 1]d_0$. If we design a sparse linear array such that $M_{\text{co}} > M$, we immediately gain enhanced degrees of freedom, because the rank of $\boldsymbol{R}_{\text{v1}}$ (or $\boldsymbol{R}_{\text{v2}}$) is greater than

that of $\boldsymbol{R}$. For example, in Fig. 2.5, we have a co-prime array with $M_{\text{co}} = 8 > 6 = M$. Because MUSIC is applicable only when the number of sources is less than the number of sensors, we assume that $K < M_{\text{co}}$ throughout this dissertation. This assumption, combined with **A2**, ensures that $\boldsymbol{A}_{\text{co}}$ is full column rank.

For brevity, we use the term direct augmentation based MUSIC (DA-MUSIC), and the term spatial smoothing based MUSIC (SS-MUSIC) to denote the MUSIC algorithm applied to $\boldsymbol{R}_{\text{v1}}$ and $\boldsymbol{R}_{\text{v2}}$, respectively. We will focus on analyzing the performance of these two coarray-based MUSIC algorithms in the following chapter.

2.4 Chapter summary

In this chapter, we presented the background of direction finding using sparse linear arrays. We reviewed the concepts of ULAs and sparse linear arrays, the stochastic signal model, and the MUSIC algorithm. We introduced the difference coarray model and the underlying mathematics. We showed the construction of the augmented covariance matrix based on the difference coarray model, and also reviewed two existing coarray-based MUSIC algorithms.

Chapter 3

Statistical Performance Analysis of the Coarray Model

Statistical performance analysis remains an important topic in array signal processing. The performance of the classical MUSIC estimator and its variants (e.g., root-MUSIC [22, 23]) was thoroughly analyzed by Stoica et al. in [61], [62] and [52], where the authors derived the asymptotic MSE expression of the MUSIC estimator and rigorously studied its statistical efficiency. In [53], Li et al. derived a unified MSE expression for common subspace-based estimators (e.g., MUSIC and ESPRIT [24]) via first-order perturbation analysis. However, the aforementioned results are based on the physical array model and make use of the statistical properties of the original sample covariance matrix, which cannot be applied when the difference coarray model is utilized. In [63], Gorokhov et al. first derived a general MSE expression for the MUSIC algorithm applied to matrix-valued transforms of the sample covariance matrix. While this expression is applicable to coarray-based MUSIC, its explicit form is rather complicated, making it difficult to conduct analytical performance studies. Therefore, a simpler and more revealing MSE expression is desired. The classical stochastic

CRB was derived and analyzed in [52]. However, it does not exist in the underdetermined case, when there are more sources than the number of sensors.

In this chapter, we consider the statistical performance analysis of sparse linear arrays[3]. We first derive the closed-form asymptotic MSE expression for DA-MUSIC and SS-MUSIC and investigate their properties. Our expression successfully explains the "saturation" behavior of SS-MUSIC in high SNR regions in previous studies. We next derive the CRB for sparse linear arrays. We analyze its behavior in high SNR regions, establish its connection with the classical stochastic CRB, and derive its approximated expression for co-prime and nested arrays with large numbers of sensors. Through these analyses, we theoretically show that co-prime and nested arrays can achieve much better performance than ULAs with the same number of sensors. Finally, we use numerical experiments to demonstrate the correctness of our theoretical results.

3.1 Asymptotic MSE of coarray-based MUSIC

Recall that in Chapter 2, we constructed the augmented covariance matrices $\hat{\boldsymbol{R}}_{\text{v1}}$ and $\hat{\boldsymbol{R}}_{\text{v2}}$ from the coarray measurement vector $\hat{\boldsymbol{z}}$, which is transformed from $\hat{\boldsymbol{R}}$. Because only a finite number of snapshots are available in practice, the estimation error $\Delta\boldsymbol{R} = \hat{\boldsymbol{R}} - \boldsymbol{R}$ is nonzero. Consequently, $\hat{\boldsymbol{R}}_{\text{v1}}$ and $\hat{\boldsymbol{R}}_{\text{v2}}$ will deviate from their true values, $\boldsymbol{R}_{\text{v1}}$ and $\boldsymbol{R}_{\text{v2}}$. When applying MUSIC, the estimated noise eigenvectors will also deviate from the true one, leading to DOA estimation errors.

[3]This chapter is based on 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, © IEEE 2017, and M. Wang, Z. Zhang, and A. Nehorai, "Further results on Coarrays, MUSIC, and the Cramér Rao bound," submitted to *IEEE Trans. Signal Process.*

In general, the eigenvectors of a perturbed matrix are not well-determined [64]. For instance, in the very low SNR scenario, $\Delta \boldsymbol{R}$ may cause a subspace swap, and the estimated noise eigenvectors will deviate drastically from the true ones [65, 66]. Nevertheless, as shown in [53, 63] and [67], given enough samples and sufficient SNR, it is possible to obtain a closed-form expressions for DOA estimation errors via first-order analysis. Following similar ideas, we are able to derive the closed-form error expression for DA-MUSIC and SS-MUSIC, as stated in Theorem 3.1.

Theorem 3.1. *Let $\hat{\theta}_k^{(1)}$ and $\hat{\theta}_k^{(2)}$ denote the estimated values of the k-th DOA by DA-MUSIC and SS-MUSIC, respectively. Let $\Delta \boldsymbol{r} = \mathrm{vec}(\hat{\boldsymbol{R}} - \boldsymbol{R})$. Assume the signal subspace and the noise subspace are well-separated, so that $\Delta \boldsymbol{r}$ does not cause a subspace swap. Then*

$$\hat{\theta}_k^{(1)} - \theta_k \doteq \hat{\theta}_k^{(2)} - \theta_k \doteq -(\gamma_k p_k)^{-1} \Re(\boldsymbol{\xi}_k^T \Delta \boldsymbol{r}), \tag{3.1}$$

where $\doteq$ denotes asymptotic equality, and

$$\boldsymbol{\xi}_k = \boldsymbol{F}^T \boldsymbol{\Gamma}^T (\boldsymbol{\beta}_k \otimes \boldsymbol{\alpha}_k), \tag{3.2a}$$

$$\boldsymbol{\alpha}_k^T = -\boldsymbol{e}_k^T \boldsymbol{A}_{\mathrm{co}}^{\dagger}, \tag{3.2b}$$

$$\boldsymbol{\beta}_k = \boldsymbol{\Pi}_{\boldsymbol{A}_{\mathrm{co}}}^{\perp} \dot{\boldsymbol{a}}_{\mathrm{co}}(\theta_k), \tag{3.2c}$$

$$\gamma_k = \dot{\boldsymbol{a}}_{\mathrm{co}}^H(\theta_k) \boldsymbol{\Pi}_{\boldsymbol{A}_{\mathrm{co}}}^{\perp} \dot{\boldsymbol{a}}_{\mathrm{co}}(\theta_k), \tag{3.2d}$$

$$\boldsymbol{\Gamma} = [\boldsymbol{\Gamma}_{M_{\mathrm{co}}}^T \, \boldsymbol{\Gamma}_{M_{\mathrm{co}}-1}^T \cdots \boldsymbol{\Gamma}_1^T]^T, \tag{3.2e}$$

$$\dot{\boldsymbol{a}}_{\mathrm{co}}(\theta_k) = \frac{\partial \boldsymbol{a}_{\mathrm{co}}(\theta_k)}{\partial \theta_k}. \tag{3.2f}$$

Proof. See Appendix A. □

Theorem 3.1 can be reinforced by Proposition 3.1. $\boldsymbol{\beta}_k \neq 0$ ensures that γ_k^{-1} exists and (3.1) is well-defined, while $\boldsymbol{\xi}_k \neq 0$ ensures that (3.1) depends on $\Delta \boldsymbol{r}$ and cannot be trivially zero.

Proposition 3.1. *$\boldsymbol{\beta}_k, \boldsymbol{\xi}_k \neq \mathbf{0}$ for $k = 1, 2, \ldots, K$.*

Proof. We first show that $\boldsymbol{\beta}_k \neq \mathbf{0}$ by contradiction. Assume $\boldsymbol{\beta}_k = \mathbf{0}$. Then $\boldsymbol{\Pi}^{\perp}_{\boldsymbol{A}_{\text{co}}} \boldsymbol{D}\boldsymbol{a}_{\text{co}}(\theta_k) = \mathbf{0}$, where $\boldsymbol{D} = \text{diag}(0, 1, \ldots, M_{\text{co}} - 1)$. This implies that $\boldsymbol{D}\boldsymbol{a}_{\text{co}}(\theta_k)$ lies in the column space of $\boldsymbol{A}_{\text{co}}$. Let $\boldsymbol{h} = \exp[-j(2\pi d_0 \sin\theta_k)/\lambda]\boldsymbol{D}\boldsymbol{a}_{\text{co}}(\theta_k)$. We immediately obtain that $[\boldsymbol{A}_{\text{co}}\ \boldsymbol{h}]$ is not full column rank. We now add $M_{\text{co}} - K - 1$ distinct DOAs in $(-\pi/2, \pi/2)$ that are different from $\theta_1, \ldots, \theta_K$, denoted by $\theta_{K+1}, \theta_{K+2}, \ldots, \theta_{M_{\text{co}}-1}$. Then we can construct an extended steering matrix $\bar{\boldsymbol{A}}_{\text{co}}$ of these $M_{\text{co}} - 1$ distinct DOAs. Let $\boldsymbol{B} = [\bar{\boldsymbol{A}}_{\text{co}}\ \boldsymbol{h}]$. It follows that $\boldsymbol{B}$ is also not full column rank. Because $\boldsymbol{B}$ is a square matrix, it is also not full row rank. Therefore there exists some non-zero $\boldsymbol{c} \in \mathbb{C}^M_{\text{co}}$ such that $\boldsymbol{c}^H\boldsymbol{B} = \mathbf{0}$. Let $t_l = \exp[j(2\pi d_0 \sin\theta_l)/\lambda]$ for $l = 1, 2, \ldots, M_{\text{co}} - 1$. We can express $\boldsymbol{B}$ as

$$\begin{bmatrix} 1 & 1 & \cdots & 1 & 0 \\ t_1 & t_2 & \cdots & t_{M_{\text{co}}-1} & 1 \\ t_1^2 & t_2^2 & \cdots & t_{M_{\text{co}}-1}^2 & 2t_k \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ t_1^{M_{\text{co}}-1} & t_2^{M_{\text{co}}-1} & \cdots & t_{M_{\text{co}}-1}^{M_{\text{co}}-1} & (M_{\text{co}}-1)t_k^{M_{\text{co}}-2} \end{bmatrix}.$$

We define the complex polynomial $f(x) = \sum_{d=1}^{M_{\text{co}}} c_d x^{d-1}$. It can be observed that $\boldsymbol{c}^T\boldsymbol{B} = \mathbf{0}$ is equivalent to $f(t_l) = 0$ for $l = 1, 2, \ldots, M_{\text{co}} - 1$, and $f'(t_k) = 0$. By construction, θ_l are distinct, so t_l are $M_{\text{co}} - 1$ different roots of $f(x)$. Because $\boldsymbol{c} \neq \mathbf{0}$, $f(x)$ is not a constant-zero polynomial, and has at most $M_{\text{co}} - 1$ roots. Therefore each root t_l has a multiplicity of at most one. However, $f'(t_k) = 0$ implies that t_k has a multiplicity of at least two, which contradicts the previous conclusion and completes the proof of $\boldsymbol{\beta}_k \neq \mathbf{0}$.

We now show that $\boldsymbol{\xi}_k \neq \mathbf{0}$. By the definition of $\boldsymbol{F}$ in Definition 2.5, each row of $\boldsymbol{F}$ has at least one non-zero element, and each column of $\boldsymbol{F}$ has at most one non-zero element. Hence $\boldsymbol{F}^T\boldsymbol{x} = \mathbf{0}$ for some $\boldsymbol{x} \in \mathbb{C}^{2M_{\mathrm{co}}-1}$ if and only of $\boldsymbol{x} = \mathbf{0}$. It suffices to show that $\boldsymbol{\Gamma}^T(\boldsymbol{\beta}_k \otimes \boldsymbol{\alpha}_k) \neq \mathbf{0}$. By the definition of $\boldsymbol{\Gamma}$, we can rewrite $\boldsymbol{\Gamma}^T(\boldsymbol{\beta}_k \otimes \boldsymbol{\alpha}_k)$ as $\tilde{\boldsymbol{B}}_k\boldsymbol{\alpha}_k$, where

$$\tilde{\boldsymbol{B}}_k = \begin{bmatrix} \beta_{kM_{\mathrm{co}}} & 0 & \cdots & 0 \\ \beta_{k(M_{\mathrm{co}}-1)} & \beta_{kM_{\mathrm{co}}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{k1} & \beta_{k2} & \cdots & \beta_{kM_{\mathrm{co}}} \\ 0 & \beta_{k1} & \cdots & \beta_{k(M_{\mathrm{co}}-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \beta_{k1} \end{bmatrix},$$

and β_{kl} is the l-th element of $\boldsymbol{\beta}_k$. Because $\boldsymbol{\beta}_k \neq \mathbf{0}$ and $K < M_{\mathrm{co}}$, $\tilde{\boldsymbol{B}}_k$ is full column rank. By the definition of pseudo inverse, we know that $\boldsymbol{\alpha}_k \neq \mathbf{0}$. Therefore $\tilde{\boldsymbol{B}}_k\boldsymbol{\alpha}_k \neq \mathbf{0}$, which completes the proof of $\boldsymbol{\xi}_k \neq \mathbf{0}$. □

One important implication of Theorem 3.1 is that DA-MUSIC and SS-MUSIC share the same first-order error expression, despite the fact that $\boldsymbol{R}_{\mathrm{v1}}$ is constructed from the second-order statistics, while $\boldsymbol{R}_{\mathrm{v2}}$ is constructed from the fourth-order statistics. Theorem 3.1 enables a unified analysis of the MSEs of DA-MUSIC and SS-MUSIC, which we present in Theorem 3.2.

Theorem 3.2. *Under the same assumptions as in Theorem 3.1, the asymptotic second-order statistics of the DOA estimation errors by DA-MUSIC and SS-MUSIC share the same form:*

$$\mathbb{E}[(\hat{\theta}_{k_1} - \theta_{k_1})(\hat{\theta}_{k_2} - \theta_{k_2})] \doteq \frac{\Re[\boldsymbol{\xi}_{k_1}^H(\boldsymbol{R} \otimes \boldsymbol{R}^T)\boldsymbol{\xi}_{k_2}]}{Np_{k_1}p_{k_2}\gamma_{k_1}\gamma_{k_2}}. \tag{3.3}$$

Proof. See Appendix B. □

By Theorem 3.2, it is straightforward to write the unified asymptotic MSE expression as

$$\epsilon(\theta_k) = \frac{\boldsymbol{\xi}_k^H(\boldsymbol{R} \otimes \boldsymbol{R}^T)\boldsymbol{\xi}_k}{Np_k^2\gamma_k^2}. \tag{3.4}$$

For brevity, when we use the acronym "MSE" in the following discussion in this chapter, we refer to the asymptotic MSE, $\epsilon(\theta_k)$, unless explicitly stated. We observe that the MSE depends on both the physical array geometry and the coarray geometry. The physical array geometry is captured by $\boldsymbol{A}$, which appears in $\boldsymbol{R} \otimes \boldsymbol{R}^T$. The coarray geometry is captured by $\boldsymbol{A}_{\text{co}}$, which appears in $\boldsymbol{\xi}_k$ and γ_k. Therefore, even if two arrays share the same coarray geometry, they may not share the same MSE because their physical array geometry may be different.

It can be easily observed from (3.4) that $\epsilon(\theta_k) \to 0$ as $N \to \infty$. However, because p_k appears in both the denominator and numerator in (3.4), it is not obvious how the MSE varies with respect to the source power p_k and noise power σ^2. Let $\bar{p}_k = p_k/\sigma^2$ denote the signal-to-noise ratio of the k-th source. Let $\bar{\boldsymbol{P}} = \text{diag}(\bar{p}_1, \bar{p}_2, \ldots, \bar{p}_K)$, and $\bar{\boldsymbol{R}} = \boldsymbol{A}\bar{\boldsymbol{P}}\boldsymbol{A}^H + \boldsymbol{I}$. We can then rewrite $\epsilon(\theta_k)$ as

$$\epsilon(\theta_k) = \frac{\boldsymbol{\xi}_k^H(\bar{\boldsymbol{R}} \otimes \bar{\boldsymbol{R}}^T)\boldsymbol{\xi}_k}{N\bar{p}_k^2\gamma_k^2}. \tag{3.5}$$

Hence the MSE depends on the SNRs instead of the absolute values of p_k or σ^2. To provide an intuitive understanding how SNR affects the MSE, we consider the case when all sources have the same power. In this case, we show in Corollary 3.1 that the MSE asymptotically decreases as the SNR increases.

Corollary 3.1. *Assume all sources have the same power p. Let $\bar{p} = p/\sigma^2$ denote the common SNR. Given sufficiently large N, the MSE $\epsilon(\theta_k)$ decreases monotonically as $\bar{p}$ increases, and*

$$\lim_{\bar{p}\to\infty} \epsilon(\theta_k) = \frac{1}{N\gamma_k^2} \|\boldsymbol{\xi}_k^H (\boldsymbol{A} \otimes \boldsymbol{A}^*)\|_2^2. \tag{3.6}$$

Proof. The limiting expression can be derived straightforwardly from (3.5). For monotonicity, without loss of generality, let $p = 1$, so $\bar{p} = 1/\sigma^2$. Because $f(x) = 1/x$ is monotonically decreasing on $(0, \infty)$, it suffices to show that $\epsilon(\theta_k)$ increases monotonically as σ^2 increases. Assume $0 < s_1 < s_2$, and we have

$$\epsilon(\theta_k)|_{\sigma^2=s_2} - \epsilon(\theta_k)|_{\sigma^2=s_1} = \frac{1}{N\gamma_k^2} \boldsymbol{\xi}_k^H \boldsymbol{Q} \boldsymbol{\xi}_k,$$

where $\boldsymbol{Q} = (s_2 - s_1)[(\boldsymbol{A}\boldsymbol{A}^H) \otimes \boldsymbol{I} + \boldsymbol{I} \otimes (\boldsymbol{A}\boldsymbol{A}^H)^T + (s_2 + s_1)\boldsymbol{I}]$. Because $\boldsymbol{A}\boldsymbol{A}^H$ is positive semidefinite, both $(\boldsymbol{A}\boldsymbol{A}^H) \otimes \boldsymbol{I}$ and $\boldsymbol{I} \otimes (\boldsymbol{A}\boldsymbol{A}^H)^T$ are positive semidefinite. Combined with our assumption that $0 < s_1 < s_2$, we conclude that $\boldsymbol{Q}$ is positive definite. By Proposition 3.1 we know that $\boldsymbol{\xi}_k \neq \boldsymbol{0}$. Therefore $\boldsymbol{\xi}_k^H \boldsymbol{Q} \boldsymbol{\xi}_k$ is strictly greater than zero, which implies the MSE monotonically increases as σ^2 increases. □

Because both DA-MUSIC and SS-MUSIC work also in cases when the number of sources exceeds the number of sensors, we are particularly interested in their limiting performance in such cases. As shown in Corollary 3.2, when $K \geq M$, the corresponding MSE is strictly greater than zero, even though the SNR approaches infinity. This corollary explains the "saturation" behavior of SS-MUSIC in the high SNR region as observed in [45] and [38]. Another interesting implication of Corollary 3.2 is that when $2 \leq K < M$, the limiting MSE is not necessarily zero. Recall that in [61], it was shown that the MSE of the traditional

MUSIC algorithm will converge to zero as SNR approaches infinity. We know that both DA-MUSIC and SS-MUSIC will be outperformed by traditional MUSIC in high SNR regions when $2 \leq K < M$. Therefore, we suggest using DA-MUSIC or SS-MUSIC only when $K \geq M$.

Corollary 3.2. *Following the same assumptions in Corollary 3.1,*

1. *When* $K = 1$, $\lim_{\bar{p}\to\infty} \epsilon(\theta_k) = 0$;
2. *When* $2 \leq K < M$, $\lim_{\bar{p}\to\infty} \epsilon(\theta_k) \geq 0$;
3. *When* $K \geq M$, $\lim_{\bar{p}\to\infty} \epsilon(\theta_k) > 0$.

Proof. The right-hand side of (3.6) can be expanded into

$$\frac{1}{N\gamma_k^2} \sum_{m=1}^{K} \sum_{n=1}^{K} \|\boldsymbol{\xi}_k^H [\boldsymbol{a}(\theta_m) \otimes \boldsymbol{a}^*(\theta_n)]\|_2^2.$$

By the definition of $\boldsymbol{F}$, $\boldsymbol{F}[\boldsymbol{a}(\theta_m) \otimes \boldsymbol{a}^*(\theta_m)]$ becomes

$$[e^{j(M_{\text{co}}-1)\omega_m}, e^{j(M_{\text{co}}-2)\omega_m}, \ldots, e^{-j(M_{\text{co}}-1)\omega_m}].$$

Hence $\boldsymbol{\Gamma}\boldsymbol{F}[\boldsymbol{a}(\theta_m) \otimes \boldsymbol{a}^*(\theta_m)] = \boldsymbol{a}_{\text{co}}(\theta_m) \otimes \boldsymbol{a}_{\text{co}}^*(\theta_m)$. Observe that

$$\begin{aligned}
\boldsymbol{\xi}_k^H [\boldsymbol{a}(\theta_m) \otimes \boldsymbol{a}^*(\theta_m)] =& (\boldsymbol{\beta}_k \otimes \boldsymbol{\alpha}_k)^H (\boldsymbol{a}_{\text{co}}(\theta_m) \otimes \boldsymbol{a}_{\text{co}}^*(\theta_m)) \\
=& (\boldsymbol{\beta}_k^H \boldsymbol{a}_{\text{co}}(\theta_m))(\boldsymbol{\alpha}_k^H \boldsymbol{a}_{\text{co}}^*(\theta_m)) \\
=& (\dot{\boldsymbol{a}}_{\text{co}}^H(\theta_k) \boldsymbol{\Pi}_{\boldsymbol{A}_{\text{co}}}^{\perp} \boldsymbol{a}_{\text{co}}(\theta_m))(\boldsymbol{\alpha}_k^H \boldsymbol{a}_{\text{co}}^*(\theta_m)) \\
=& 0.
\end{aligned}$$

We can reduce the right-hand side of (3.6) into

$$\frac{1}{N\gamma_k^2} \sum_{\substack{1\le m,n\le K \\ m\neq n}} \|\boldsymbol{\xi}_k^H[\boldsymbol{a}(\theta_m)\otimes \boldsymbol{a}^*(\theta_n)]\|_2^2.$$

Therefore when $K = 1$, the limiting expression is exactly zero. When $2 \le K < M$, the limiting expression is not necessary zero because when $m \neq n$, $\boldsymbol{\xi}_k^H[\boldsymbol{a}(\theta_m)\otimes \boldsymbol{a}^*(\theta_n)]$ is not necessarily zero.

When $K \ge M$, $\boldsymbol{A}$ is full row rank. Hence $\boldsymbol{A}\otimes\boldsymbol{A}^*$ is also full row rank. By Proposition 3.1 we know that $\boldsymbol{\xi}_k \neq 0$, which implies that $\epsilon(\theta_k)$ is strictly greater than zero. □

3.2 CRB for sparse linear arrays

In this section, we provide detailed analyses of the CRB for sparse linear arrays. The CRB gives the lower bound on the variance of any unbiased estimators under regularity conditions. Investigating the CRB for sparse linear arrays enables us to better understand the performance limits of these arrays, which will aid us in identifying optimal designs of sparse linear arrays.

3.2.1 Derivation

The CRB for the stochastic model (2.3) has been well studied in [52], but only when the number of sources is less than the number of sensors and no prior knowledge of $\boldsymbol{P}$ is given. For the coarray model, the number of sources can exceed the number of sensors, and $\boldsymbol{P}$ is assumed to be diagonal. Therefore, the CRB derived in [52] cannot be directly applied.

Based on [68, Appendix 15C], we provide an alternative CRB based on the signal model (2.3), under assumptions **A1**–**A4**.

For the signal model (2.3), the vector of unknown parameters is defined by

$$\boldsymbol{\eta} = [\theta_1, \ldots, \theta_K, p_1, \ldots, p_k, \sigma^2]^T, \tag{3.7}$$

and the (m, n)-th element of the Fisher information matrix (FIM) is given by [52, 68]

$$\mathrm{FIM}_{mn} = N \,\mathrm{tr}\left[\frac{\partial \boldsymbol{R}}{\partial \eta_m}\boldsymbol{R}^{-1}\frac{\partial \boldsymbol{R}}{\partial \eta_n}\boldsymbol{R}^{-1}\right]. \tag{3.8}$$

Observing that $\mathrm{tr}(\boldsymbol{AB}) = \mathrm{vec}(\boldsymbol{A}^T)^T \mathrm{vec}(\boldsymbol{B})$, and that $\mathrm{vec}(\boldsymbol{AXB}) = (\boldsymbol{B}^T \otimes \boldsymbol{A})\,\mathrm{vec}(\boldsymbol{X})$, we can rewrite (3.8) as

$$\mathrm{FIM}_{mn} = N\left[\frac{\partial \boldsymbol{r}}{\partial \eta_m}\right]^H (\boldsymbol{R}^T \otimes \boldsymbol{R})^{-1}\frac{\partial \boldsymbol{r}}{\partial \eta_n}.$$

Denote the derivatives of $\boldsymbol{r}$ with respect to $\boldsymbol{\eta}$ as

$$\frac{\partial \boldsymbol{r}}{\partial \boldsymbol{\eta}} = \left[\frac{\partial \boldsymbol{r}}{\partial \theta_1} \cdots \frac{\partial \boldsymbol{r}}{\partial \theta_K}\ \frac{\partial \boldsymbol{r}}{\partial p_1} \cdots \frac{\partial \boldsymbol{r}}{\partial p_K}\ \frac{\partial \boldsymbol{r}}{\partial \sigma^2}\right]. \tag{3.9}$$

The FIM can be compactly expressed by

$$\mathrm{FIM} = \left[\frac{\partial \boldsymbol{r}}{\partial \boldsymbol{\eta}}\right]^H (\boldsymbol{R}^T \otimes \boldsymbol{R})^{-1}\frac{\partial \boldsymbol{r}}{\partial \boldsymbol{\eta}}. \tag{3.10}$$

According to (2.6), we can compute the derivatives in (3.9) and obtain

$$\frac{\partial \boldsymbol{r}}{\partial \boldsymbol{\eta}} = \left[\dot{\boldsymbol{A}}_{\mathrm{d}}\boldsymbol{P} \quad \boldsymbol{A}_{\mathrm{d}} \quad \boldsymbol{i}\right], \tag{3.11}$$

where $\dot{\boldsymbol{A}}_{\mathrm{d}} = \dot{\boldsymbol{A}}^* \odot \boldsymbol{A} + \boldsymbol{A}^* \odot \dot{\boldsymbol{A}}$, $\boldsymbol{A}_{\mathrm{d}}$ and $\boldsymbol{i}$ follow the same definitions as in (2.6), and

$$\dot{\boldsymbol{A}} = \left[\frac{\partial \boldsymbol{a}(\theta_1)}{\partial \theta_1} \frac{\partial \boldsymbol{a}(\theta_2)}{\partial \theta_2} \cdots \frac{\partial \boldsymbol{a}(\theta_K)}{\partial \theta_K} \right].$$

Note that (3.11) can be partitioned into two parts, specifically, the part corresponding to DOAs and the part corresponding to the source and noise powers. We can also partition the FIM. Because $\boldsymbol{R}$ is positive definite, $(\boldsymbol{R}^T \otimes \boldsymbol{R})^{-1}$ is also positive definite, and its square root $(\boldsymbol{R}^T \otimes \boldsymbol{R})^{-1/2}$ also exists. Let

$$\begin{aligned} \boldsymbol{M}_{\boldsymbol{\theta}} &= (\boldsymbol{R}^T \otimes \boldsymbol{R})^{-1/2} \dot{\boldsymbol{A}}_{\mathrm{d}} \boldsymbol{P}, \\ \boldsymbol{M}_{\boldsymbol{s}} &= (\boldsymbol{R}^T \otimes \boldsymbol{R})^{-1/2} \big[\boldsymbol{A}_{\mathrm{d}} \, \boldsymbol{i} \big]. \end{aligned}$$

We can write the partitioned FIM as

$$\text{FIM} = N \begin{bmatrix} \boldsymbol{M}_{\boldsymbol{\theta}}^H \boldsymbol{M}_{\boldsymbol{\theta}} & \boldsymbol{M}_{\boldsymbol{\theta}}^H \boldsymbol{M}_{\boldsymbol{s}} \\ \boldsymbol{M}_{\boldsymbol{s}}^H \boldsymbol{M}_{\boldsymbol{\theta}} & \boldsymbol{M}_{\boldsymbol{s}}^H \boldsymbol{M}_{\boldsymbol{s}} \end{bmatrix}.$$

The CRB matrix for the DOAs is then obtained by block-wise inversion:

$$\boldsymbol{B}_{(\text{sto-uc})}(\boldsymbol{\theta}) = \frac{1}{N} (\boldsymbol{M}_{\boldsymbol{\theta}}^H \boldsymbol{\Pi}_{\boldsymbol{M}_{\boldsymbol{s}}}^{\perp} \boldsymbol{M}_{\boldsymbol{\theta}})^{-1}, \tag{3.12}$$

where $\boldsymbol{\Pi}_{\boldsymbol{M}_{\boldsymbol{s}}}^{\perp} = \boldsymbol{I} - \boldsymbol{M}_{\boldsymbol{s}} (\boldsymbol{M}_{\boldsymbol{s}}^H \boldsymbol{M}_{\boldsymbol{s}})^{-1} \boldsymbol{M}_{\boldsymbol{s}}^H$.

From (3.11), we observe that the invertibility of the FIM depends on the coarray structure. In the noisy case, $(\boldsymbol{R}^T \otimes \boldsymbol{R})^{-1}$ is always full rank, so the FIM is invertible if and only if $\partial \boldsymbol{r} / \partial \boldsymbol{\eta}$ is full column rank. By (3.11) we know that the rank of $\partial \boldsymbol{r} / \partial \boldsymbol{\eta}$ is closely related to $\boldsymbol{A}_{\mathrm{d}}$, which encodes the coarray structure. Therefore, unlike the classical stochastic CRB for the stochastic model introduced in [52, Remark 1], $\boldsymbol{B}_{(\text{sto-uc})}(\boldsymbol{\theta})$ is applicable even if the number

of sources exceeds the number of sensors. However, $\boldsymbol{B}_{(\text{sto-uc})}(\boldsymbol{\theta})$ is not valid for an arbitrary number of sources, because $\boldsymbol{A}_{\text{d}}$ may not be full column rank when too many sources are present. A more detailed identifiability analysis can be found in [69].

3.2.2 Behavior in high SNR regions

In Section 3.1, we showed that the asymptotic MSE of DA-MUSIC and SS-MUSIC depends on the SNRs instead of the absolute values of p_k or σ^2. We now show that $\boldsymbol{B}_{(\text{sto-uc})}$ exhibits a similar behavior. Let $\bar{p}_k = p_k/\sigma^2$, and $\bar{\boldsymbol{P}} = \text{diag}(\bar{p}_1, \bar{p}_2, \ldots, \bar{p}_K)$. We have

$$\boldsymbol{M}_{\boldsymbol{\theta}} = (\bar{\boldsymbol{R}}^T \otimes \bar{\boldsymbol{R}})^{-1/2} \dot{\boldsymbol{A}}_{\text{d}} \bar{\boldsymbol{P}}, \tag{3.13}$$

$$\boldsymbol{M}_{\boldsymbol{s}} = \sigma^{-2} (\bar{\boldsymbol{R}}^T \otimes \bar{\boldsymbol{R}})^{-1/2} \big[\boldsymbol{A}_{\text{d}}\, \boldsymbol{i}\big]. \tag{3.14}$$

Substituting (3.13) and (3.14) into (3.12), the term σ^2 gets canceled, and the resulting $\boldsymbol{B}_{(\text{sto-uc})}(\boldsymbol{\theta})$ depends on the ratios $\bar{p}_k$ instead of absolute values of p_k or σ^2.

We now analyze the behavior of $\boldsymbol{B}_{(\text{sto-uc})}$ in high SNR regions. The results are summarized in the following proposition.

Proposition 3.2. *Assume all sources have the same power p, and $\partial \boldsymbol{r}/\partial \boldsymbol{\eta}$ is full column rank. Let $\bar{p} = p/\sigma^2$.*

(1) If $K < M$, and $\lim_{\bar{p}\to\infty} \boldsymbol{B}_{(\text{sto-uc})}(\boldsymbol{\theta})$ exists, it is zero under mild conditions.

(2) If $K \geq M$, and $\lim_{\bar{p}\to\infty} \boldsymbol{B}_{(\text{sto-uc})}(\boldsymbol{\theta})$ exists, it is positive definite.

Proof. See Appendix C. □

While infinite SNR is unachievable from a practical standpoint, Proposition 3.2 gives some useful theoretical implications. When $K < M$, the limiting MSE (13) in Corollary 3.1 is not necessarily zero. However, Proposition 3.2 reveals that the CRB generally approaches zero when SNR goes to infinity. This observation implies that both DA-MUSIC and SS-MUSIC may have poor statistical efficiency in high SNR regions. When $K \geq M$, Proposition 3.2 implies that the CRB of each DOA will converge to a positive constant as the SNR approaches infinity. This unusual behavior puts a strictly positive lower bound on unbiased DOA estimators in the underdetermined case. It is also consistent with the behavior of the asymptotic MSE described in Corollary 3.2.

3.2.3 Connection to the classical stochastic CRB

In this section, we establish the connection between $\boldsymbol{B}_{(\text{sto-uc})}$ and the classical stochastic CRB derived in [52]. In this section and the next section, to avoid complications in the derivatives, we use $\boldsymbol{\omega}$ instead of $\boldsymbol{\theta}$ to represent the DOAs. Recall that in Section 2.2.1, we showed that there exists a one-to-one mapping between $\boldsymbol{\omega}$ and $\boldsymbol{\theta}$.

The classical stochastic CRB, which we denote as $\boldsymbol{B}_{(\text{sto})}$, is derived without prior knowledge that the sources are uncorrelated. The unknown parameters consist of the DOAs, $\boldsymbol{\omega}$, the real and imaginary parts of $\boldsymbol{P}$, and the noise variance σ^2. Because $\boldsymbol{P}$ is Hermitian, there are $K^2 + K + 1$ unknown parameters. In this case, we have [52, 56]:

$$\boldsymbol{B}_{(\text{sto})}(\boldsymbol{\omega}) = \frac{\sigma^2}{2N}\left\{\Re[(\dot{\boldsymbol{A}}^H \boldsymbol{\Pi}_{\boldsymbol{A}}^{\perp} \dot{\boldsymbol{A}}) \circ (\boldsymbol{P}\boldsymbol{A}^H \boldsymbol{R}^{-1} \boldsymbol{A}\boldsymbol{P})^T]\right\}^{-1}, \tag{3.15}$$

where

$$\dot{\boldsymbol{A}} = \left[\frac{\partial \boldsymbol{a}(\omega_1)}{\partial \omega_1} \quad \frac{\partial \boldsymbol{a}(\omega_1)}{\partial \omega_2} \quad \dots \quad \frac{\partial \boldsymbol{a}(\omega_1)}{\partial \omega_K} \right].$$

Our CRB $\boldsymbol{B}_{\text{(sto-uc)}}$ is derived with the prior knowledge that $\boldsymbol{P}$ is a diagonal matrix. Recall that it can be compactly expressed as

$$\boldsymbol{B}_{\text{(sto-uc)}}(\boldsymbol{\omega}) = \frac{1}{N}(\boldsymbol{M}_{\boldsymbol{\omega}}^H \boldsymbol{\Pi}_{\boldsymbol{M_s}}^{\perp} \boldsymbol{M}_{\boldsymbol{\omega}})^{-1}, \tag{3.16}$$

where

$$\boldsymbol{M}_{\boldsymbol{\omega}} = (\boldsymbol{R}^T \otimes \boldsymbol{R})^{-1/2} \dot{\boldsymbol{A}}_{\text{d}} \boldsymbol{P}, \tag{3.17}$$

$$\boldsymbol{M}_{\boldsymbol{s}} = (\boldsymbol{R}^T \otimes \boldsymbol{R})^{-1/2} [\boldsymbol{A}_{\text{d}} \ \boldsymbol{i}], \tag{3.18}$$

$$\dot{\boldsymbol{A}}_{\text{d}} = \dot{\boldsymbol{A}}^* \odot \boldsymbol{A} + \boldsymbol{A}^* \odot \dot{\boldsymbol{A}}, \tag{3.19}$$

$$\boldsymbol{A}_{\text{d}} = \boldsymbol{A}^* \odot \boldsymbol{A}. \tag{3.20}$$

While the compact form (3.16) of $\boldsymbol{B}_{\text{(sto-uc)}}$ provides great convenience when analyzing the maximum number of resolvable sources [69], it is not well-suited for our asymptotic analysis in the following discussion. Therefore, we provide a brief review of its more traditional form, obtained by block-wise computation of the FIM. Under the assumption that the sources are uncorrelated, the FIM of the stochastic model is given by [3]

$$\boldsymbol{J}_{\text{(sto-uc)}} = N \begin{bmatrix} \boldsymbol{J}_{\boldsymbol{\omega}\boldsymbol{\omega}} & \boldsymbol{J}_{\boldsymbol{\omega}\boldsymbol{p}} & \boldsymbol{J}_{\boldsymbol{\omega}\sigma^2} \\ \boldsymbol{J}_{\boldsymbol{p}\boldsymbol{\omega}} & \boldsymbol{J}_{\boldsymbol{p}\boldsymbol{p}} & \boldsymbol{J}_{\boldsymbol{p}\sigma^2} \\ \boldsymbol{J}_{\sigma^2\boldsymbol{\omega}} & \boldsymbol{J}_{\sigma^2\boldsymbol{p}} & \boldsymbol{J}_{\sigma^2\sigma^2} \end{bmatrix}, \tag{3.21}$$

where

$$\begin{aligned}
\boldsymbol{J}_{\boldsymbol{\omega\omega}} =& 2\Re[(\dot{\boldsymbol{A}}^H\boldsymbol{R}^{-1}\dot{\boldsymbol{A}})^* \circ (\boldsymbol{P}\boldsymbol{A}^H\boldsymbol{R}^{-1}\boldsymbol{A}\boldsymbol{P}) + (\dot{\boldsymbol{A}}^H\boldsymbol{R}^{-1}\boldsymbol{A})^* \circ (\boldsymbol{P}\boldsymbol{A}^H\boldsymbol{R}^{-1}\dot{\boldsymbol{A}}\boldsymbol{P})], \\
\boldsymbol{J}_{\boldsymbol{pp}} =& (\boldsymbol{A}^H\boldsymbol{R}^{-1}\boldsymbol{A})^* \circ (\boldsymbol{A}^H\boldsymbol{R}^{-1}\boldsymbol{A}), \\
\boldsymbol{J}_{\sigma^2\sigma^2} =& \operatorname{tr}(\boldsymbol{R}^{-2}), \\
\boldsymbol{J}_{\boldsymbol{\omega p}} =& 2\Re[(\dot{\boldsymbol{A}}^H\boldsymbol{R}^{-1}\boldsymbol{A})^* \circ (\boldsymbol{P}\boldsymbol{A}^H\boldsymbol{R}^{-1}\boldsymbol{A})], \\
\boldsymbol{J}_{\boldsymbol{\omega}\sigma^2} =& 2\Re[\operatorname{diag}(\boldsymbol{P}\dot{\boldsymbol{A}}^H\boldsymbol{R}^{-2}\boldsymbol{A})], \\
\boldsymbol{J}_{\boldsymbol{p}\sigma^2} =& \operatorname{diag}(\boldsymbol{A}^H\boldsymbol{R}^{-2}\boldsymbol{A}),
\end{aligned}$$

and $\boldsymbol{J}_{\boldsymbol{p\omega}} = \boldsymbol{J}_{\boldsymbol{\omega p}}^H$, $\boldsymbol{J}_{\sigma^2\boldsymbol{\omega}} = \boldsymbol{J}_{\boldsymbol{\omega}\sigma^2}^H$, $\boldsymbol{J}_{\sigma^2\boldsymbol{p}} = \boldsymbol{J}_{\boldsymbol{p}\sigma^2}^H$.

By inverting $\boldsymbol{J}_{(\text{sto-uc})}$, we obtain the alternative expression of $\boldsymbol{B}_{(\text{sto-uc})}$. While this expression seems much more complicated than the one in (3.16), it can be shown that they are equivalent via Lemma 3.1 listed below. In the following derivations, we make extensive use of (3.21) instead of (3.16).

Lemma 3.1. *Let $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$, $\boldsymbol{E}$, and $\boldsymbol{F}$ be compatible matrices. Then*

$$(\boldsymbol{A} \odot \boldsymbol{B})^H(\boldsymbol{C} \otimes \boldsymbol{D})(\boldsymbol{E} \odot \boldsymbol{F}) = (\boldsymbol{A}^H\boldsymbol{C}\boldsymbol{E}) \circ (\boldsymbol{B}^H\boldsymbol{D}\boldsymbol{F}). \tag{3.22}$$

Proof. The left-hand side of (3.22) can be expanded as

$$\begin{bmatrix} \boldsymbol{a}_1^H \otimes \boldsymbol{b}_1^H \\ \vdots \\ \boldsymbol{a}_M^H \otimes \boldsymbol{b}_M^H \end{bmatrix} (\boldsymbol{C} \otimes \boldsymbol{D}) \begin{bmatrix} \boldsymbol{e}_1 \otimes \boldsymbol{f}_1 & \cdots & \boldsymbol{e}_N \otimes \boldsymbol{f}_N \end{bmatrix}, \tag{3.23}$$

whose (i,j)-th element is given by

$$(\boldsymbol{a}_i^H \otimes \boldsymbol{b}_i^H)(\boldsymbol{C} \otimes \boldsymbol{D})(\boldsymbol{e}_j \otimes \boldsymbol{f}_j) = (\boldsymbol{a}_i^H \boldsymbol{C} \boldsymbol{e}_j)(\boldsymbol{b}_i^H \boldsymbol{D} \boldsymbol{f}_j).$$

Observing that $\boldsymbol{a}_i^H \boldsymbol{C} \boldsymbol{e}_j$ is the (i,j)-th element of $\boldsymbol{A}^H \boldsymbol{C} \boldsymbol{E}$, and that $\boldsymbol{b}_i^H \boldsymbol{D} \boldsymbol{f}_j$ is the (i,j)-th element of $\boldsymbol{B}^H \boldsymbol{D} \boldsymbol{F}$, we immediately conclude that the left-hand side is equal to the right-hand side in (3.22). □

When $\boldsymbol{P}$ is diagonal, there is a subtle distinction between $\boldsymbol{B}_{(\text{sto})}$ and $\boldsymbol{B}_{(\text{sto-uc})}$. $\boldsymbol{B}_{(\text{sto})}$ gives the CRB when the sources are uncorrelated and this knowledge is not known a priori. $\boldsymbol{B}_{(\text{sto-uc})}$ gives the CRB when the sources are uncorrelated and this knowledge is known a priori. This subtle distinction implies that $\boldsymbol{B}_{(\text{sto})}$ and $\boldsymbol{B}_{(\text{sto-uc})}$ are not equal. In fact, it is straightforward to show that $\boldsymbol{B}_{(\text{sto-uc})} \preceq \boldsymbol{B}_{(\text{sto})}$, implying that incorporating the prior knowledge reduces uncertainties in estimating the DOAs. If we compare (3.15) with (3.21), we can observe that the term $\boldsymbol{P} \boldsymbol{A}^H \boldsymbol{R}^{-1} \boldsymbol{A} \boldsymbol{P}$ appears in both expressions, suggesting a potential connection between $\boldsymbol{B}_{(\text{sto})}$ and $\boldsymbol{B}_{(\text{sto-uc})}$. To establish such a connection, we first introduce the following three lemmas.

Lemma 3.2 (Woodbury matrix inversion lemma [70])**.**

$$(\boldsymbol{A} + \boldsymbol{U}\boldsymbol{C}\boldsymbol{V})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{U}(\boldsymbol{C}^{-1} + \boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{U})^{-1}\boldsymbol{V}\boldsymbol{A}^{-1}.$$

Lemma 3.3. *Let $\boldsymbol{A}$ be nonsingular and $\boldsymbol{B}$ have a sufficiently small norm. Then*

$$(\boldsymbol{A} + \boldsymbol{B})^{-1} \approx \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}. \tag{3.24}$$

Proof. For $\boldsymbol{B}$ with a sufficiently small norm, the spectral radius of $\boldsymbol{A}^{-1}\boldsymbol{B}$ will be less than one, and the Taylor series expansion of $(\boldsymbol{A}+\boldsymbol{B})^{-1}$ converges [70, P. 421]. Therefore, (3.24) is just the first-order Taylor approximation. □

Lemma 3.4. *For sufficiently small σ^2, $\sigma^2\boldsymbol{R}^{-1} = \boldsymbol{\Pi}_{\boldsymbol{A}}^{\perp} + O(\sigma^2)$, where $O(\sigma^2)$ denotes terms with the same order of σ^2.*

Proof. By Lemma 3.2, we have

$$\sigma^2\boldsymbol{R}^{-1} = \boldsymbol{I} - \boldsymbol{A}(\sigma^2\boldsymbol{P}^{-1} + \boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{A}^H. \tag{3.25}$$

Because $\boldsymbol{A}^H\boldsymbol{A}$ is full rank, by Lemma 3.3, $(\sigma^2\boldsymbol{P}^{-1} + \boldsymbol{A}^H\boldsymbol{A})^{-1} = (\boldsymbol{A}^H\boldsymbol{A})^{-1} + O(\sigma^2)$. □

We now reveal this connection in Theorem 3.3.

Theorem 3.3. *Assume that the sources are uncorrelated. If we fix the diagonal matrix $\boldsymbol{P} \succ 0$, $\boldsymbol{B}_{(\text{sto})} \doteq \boldsymbol{B}_{(\text{sto-uc})}$ as $\sigma^2 \to 0$, where $\doteq$ denotes that the equality is up to the first order with respect to σ^2.*

Proof. Without loss of generality, we assume that $N = 1$. We already know that when $\boldsymbol{P}$ is diagonal, the following inequalities hold:

$$\boldsymbol{J}_{\boldsymbol{\omega}\boldsymbol{\omega}}^{-1} \preceq \boldsymbol{B}_{(\text{sto-uc})} \preceq \boldsymbol{B}_{(\text{sto})}. \tag{3.26}$$

It suffices to show that $\boldsymbol{J}_{\boldsymbol{\omega}\boldsymbol{\omega}}^{-1} \doteq \boldsymbol{B}_{(\text{sto})}$. Using the above Lemma 3.4, we observe that

$$\begin{aligned}\sigma^2\Re[(\dot{\boldsymbol{A}}^H\boldsymbol{R}^{-1}\dot{\boldsymbol{A}})^* \circ (\boldsymbol{P}\boldsymbol{A}^H\boldsymbol{R}^{-1}\boldsymbol{A}\boldsymbol{P})] =& \Re[(\dot{\boldsymbol{A}}^H(\sigma^2\boldsymbol{R}^{-1})\dot{\boldsymbol{A}})^* \circ (\boldsymbol{P}\boldsymbol{A}^H\boldsymbol{R}^{-1}\boldsymbol{A}\boldsymbol{P})] \\ =& \Re[(\dot{\boldsymbol{A}}^H\boldsymbol{\Pi}_{\boldsymbol{A}}^{\perp}\dot{\boldsymbol{A}})^* \circ (\boldsymbol{P}\boldsymbol{A}^H\boldsymbol{R}^{-1}\boldsymbol{A}\boldsymbol{P}) + O(\sigma^2)].\end{aligned}$$

Because that $\boldsymbol{\Pi}_{\boldsymbol{A}}^{\perp}\boldsymbol{A}=\boldsymbol{0}$, we have

$$\sigma^2\Re[(\dot{\boldsymbol{A}}^H\boldsymbol{R}^{-1}\boldsymbol{A})^*\circ(\boldsymbol{P}\boldsymbol{A}^H\boldsymbol{R}^{-1}\dot{\boldsymbol{A}}\boldsymbol{P})]=\Re[(\dot{\boldsymbol{A}}^H(\sigma^2\boldsymbol{R}^{-1})\boldsymbol{A})^*\circ(\boldsymbol{P}\boldsymbol{A}^H\boldsymbol{R}^{-1}\dot{\boldsymbol{A}}\boldsymbol{P})]=O(\sigma^2).$$

Combined with the fact that $\Re(\boldsymbol{X})=\Re(\boldsymbol{X}^*)$, we have

$$\begin{aligned}\boldsymbol{J}_{\boldsymbol{\omega}\boldsymbol{\omega}}^{-1}&=\frac{\sigma^2}{2}\big\{\sigma^2\Re[(\dot{\boldsymbol{A}}^H\boldsymbol{R}^{-1}\dot{\boldsymbol{A}})^*\circ(\boldsymbol{P}\boldsymbol{A}^H\boldsymbol{R}^{-1}\boldsymbol{A}\boldsymbol{P})]\\&\quad+\sigma^2\Re[(\dot{\boldsymbol{A}}^H\boldsymbol{R}^{-1}\boldsymbol{A})^*\circ(\boldsymbol{P}\boldsymbol{A}^H\boldsymbol{R}^{-1}\dot{\boldsymbol{A}}\boldsymbol{P})]\big\}^{-1}\\&=\frac{\sigma^2}{2}\big\{\Re[(\dot{\boldsymbol{A}}^H\boldsymbol{\Pi}_{\boldsymbol{A}}^{\perp}\dot{\boldsymbol{A}})^*\circ(\boldsymbol{P}\boldsymbol{A}^H\boldsymbol{R}^{-1}\boldsymbol{A}\boldsymbol{P})+O(\sigma^2)]\big\}^{-1}\\&=\frac{\sigma^2}{2}\big\{\Re[(\dot{\boldsymbol{A}}^H\boldsymbol{\Pi}_{\boldsymbol{A}}^{\perp}\dot{\boldsymbol{A}})\circ(\boldsymbol{P}\boldsymbol{A}^H\boldsymbol{R}^{-1}\boldsymbol{A}\boldsymbol{P})^T]+\Re[O(\sigma^2)]\big\}^{-1}.\end{aligned}$$

By Lemma 3.3, we obtain that $\boldsymbol{J}_{\boldsymbol{\omega}\boldsymbol{\omega}}^{-1}\doteq\boldsymbol{B}_{(\text{sto})}$, which immediately leads to $\boldsymbol{B}_{(\text{sto})}\doteq\boldsymbol{B}_{(\text{sto-uc})}$.

□

Theorem 3.3 shows that when the sources are uncorrelated and the number of sources is less than the number of sensors, $\boldsymbol{B}_{(\text{sto})}$ and $\boldsymbol{B}_{(\text{sto-uc})}$ are approximately equal when the SNR is large. This result is in agreement with our intuition. When the SNR is larger, we can clearly identify the signals, and incorporating the prior knowledge will not give much improvement in estimation performance. When the SNR is low, the signals cannot be clearly distinguished from the noise, and we are more uncertain about whether the sources are correlated. In this case, incorporating the prior knowledge will help improve the estimation performance.

3.2.4 Analysis for co-prime and nested arrays with large number of sensors

In this section, we analyze the behavior of $\boldsymbol{B}_{(\text{sto-uc})}$ for co-prime and nested arrays with large numbers of sensors. We will focus on co-prime and nested arrays, whose array configurations have closed-form solutions. It is possible to extend our analysis to other variants, such as generalized co-prime arrays [45]. While numerical simulations show that MRAs share behaviors similar to co-prime and nested arrays [71, 72], we cannot not obtain similar analytical results because MRA configurations do not have closed-form solutions.

For reference, we will provide the results for the ULA case first. In [61], the authors showed that for an M-sensor ULA, the CRB of the deterministic signal model decreases at a rate of $O(M^{-3})$ for large M. In the following proposition, we show that $\boldsymbol{B}_{(\text{sto-uc})}$ has the same behavior.

Proposition 3.3. *Assume that* $\text{SNR}_i^{-1} := \sigma^2/p_i \ll M$ *for all* $i = 1, 2, \ldots, K$ *and that* $K < M$. *Then for ULAs, as* $M \to \infty$,

$$\boldsymbol{B}_{(\text{sto-uc})}(\boldsymbol{\omega}) \approx \frac{6}{M^3 N}\sigma^2 \boldsymbol{P}^{-1}. \tag{3.27}$$

Proof. See Appendix D. □

Unlike ULAs, the physical array geometries of sparse linear arrays can be drastically different, even if they share the same number of sensors[4]. To avoid complications, we will consider nearly optimal configurations in the following discussion. For co-prime arrays, we

[4]For example, the nested arrays generated by $(8, 2)$ and $(5, 5)$ both have 10 sensors. However, the latter can achieve 30 degrees of freedom, while the former can achieve only 18 degrees of freedom.

will consider configurations generated by co-prime pairs $(Q, Q+1)$. For nested arrays, we will consider the configurations generated by (Q, Q). Co-prime and nested arrays generated by these configurations are nearly optimal in terms of maximum achievable degrees of freedom.

Because co-prime and nested arrays are non-uniform and $\boldsymbol{B}_{(\text{sto-uc})}$ is used instead of $\boldsymbol{B}_{(\text{sto})}$, the overall derivation is much more involved than for the ULA case. We will start from the one source case.

Theorem 3.4 (One source case)**.** *Let $K = 1$ and assume that $\text{SNR}^{-1} := \sigma^2/p \ll Q$. Then as $Q \to \infty$,*

1. *For a co-prime array generated with the co-prime pair $(Q, Q+1)$,*

$$\boldsymbol{B}_{(\text{sto-uc})}(\omega) \approx \frac{6}{11}\frac{1}{N}\frac{1}{Q^5}\frac{1}{\text{SNR}}. \tag{3.28}$$

2. *For a nested array generated with the parameter pair (Q, Q),*

$$\boldsymbol{B}_{(\text{sto-uc})}(\omega) \approx \frac{12}{5}\frac{1}{N}\frac{1}{Q^5}\frac{1}{\text{SNR}}. \tag{3.29}$$

Proof. See Appendix E. □

We observe that, similar to the ULA case, the CRB is inversely proportional to the number of samples, N, and the SNR. The interesting term here is $1/Q^5$. According to Definition 2.2, a co-prime array generated with the co-prime pair $(Q, Q+1)$ consists of $M = 3Q$ sensors. Similarly, a nested array generated with the parameter pair (Q, Q) consists of $M = 2Q$ sensors. Theorem 3.4 shows that, in the one source case, $\boldsymbol{B}_{(\text{sto-uc})}$ of co-prime and nested arrays can indeed decrease at a rate of $O(M^{-5})$. This finding implies that the resolution

limit of such co-prime and nested arrays is inversely proportional to M^5, as opposed to M^3 in the ULA case. In other words, such co-prime and nested arrays have much better resolution than ULAs with the same number of sensors. This behavior can be explained by the fact that they have much larger apertures than ULAs with the same number of sensors.

Remark. The constant term for the co-prime arrays in (3.28) is smaller than that of the nested arrays because for a fixed Q, the co-prime array generated by the co-prime pair $(Q, Q+1)$ has a larger aperture than that of the nested array generated by the parameter pair (Q, Q).

Next, we generalize the results in Theorem 3.4 to the multiple source case. Unlike ULAs, this generalization is not straightforward because both co-prime and nested arrays contain subarrays with inter-element spacing greater than d_0. Such subarrays have grating lobes in their beam patterns [73]. Therefore, one of the two subarrays of a co-prime (or nested) array will not be able to identify certain source placements, leading to degenerated estimation performance. We call such source placements degenerative placements. To illustrate this behavior, we plotted $\boldsymbol{B}_{(\text{sto-uc})}$ for a co-prime array and a ULA for the two-source case in Fig. 3.1. We can observe that, unlike the ULA, there are multiple off-diagonal black bands, implying that the values of $\boldsymbol{B}_{(\text{sto-uc})}$ can be significantly larger for certain placements of ω_1, ω_2. While such a degenerative behavior is interesting, we want to focus on approximating the fastest rate $\boldsymbol{B}_{(\text{sto-uc})}$ can decrease with respect to the number of sensors, M, in the following discussion. Therefore, we need to exclude such degenerative source placements from our analysis. Hence, we introduce the following definition:

Definition 3.1. *Let L be a positive integer and $0 < \delta < 1$. Define the set Ω_L^δ as follows:*

$$\Omega_L^\delta = \{\omega | \omega L/2 \in [k\pi + \arcsin\delta, (k+1)\pi - \arcsin\delta], k \in \mathbb{Z}\},$$

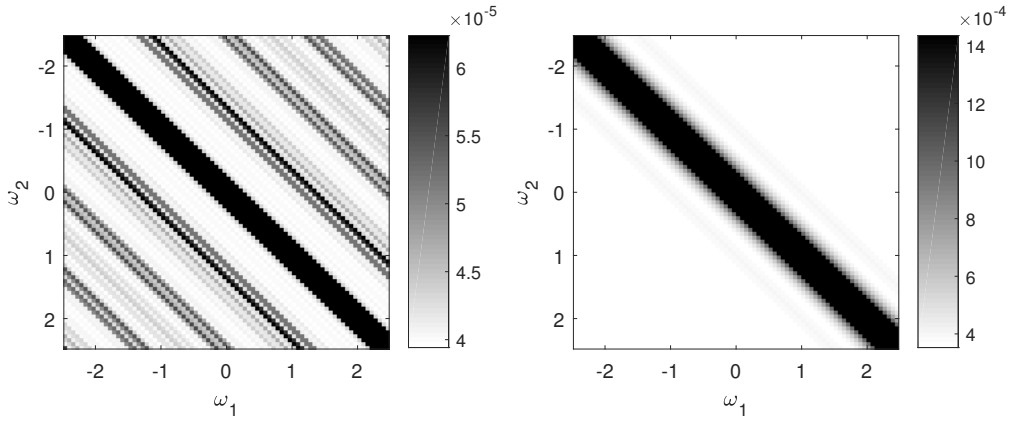


Figure 3.1: $\boldsymbol{B}_{(\text{sto-uc})}(\omega_1, \omega_2)$ computed from difference combinations of (ω_1, ω_2) for (a) a co-prime array generated by the co-prime pair $(4, 5)$ and (b) a 12-element ULA.

where $\mathbb{Z}$ denote the set of integers.

The intuition behind this definition is explained in Appendix F. Using Definition 3.1, the result for the multiple source case is summarized in Theorem 3.5:

Theorem 3.5 (Multiple source case)**.** *Let $K < Q$ and assume that $\text{SNR}_i^{-1} := \sigma^2/p_i \ll Q$. Choose $\delta = 0.5$. Then for large values of Q,*

1. *For a co-prime array generated with co-prime pair $(Q, Q+1)$, if $\omega_m - \omega_n \in \Omega_Q^\delta \cap \Omega_{Q+1}^\delta$, $\forall m \neq n$, $m, n \in \{1, 2, \ldots, K\}$, then*

$$\boldsymbol{B}_{(\text{sto-uc})}(\omega) \approx \frac{6}{11}\frac{1}{N}\frac{1}{Q^5}\sigma^2 \boldsymbol{P}^{-1}. \tag{3.30}$$

2. *For a nested array generated with the parameter pair (Q, Q), if $\omega_m - \omega_n \in \Omega_{Q+1}^\delta$, $\forall m \neq n$, $m, n \in \{1, 2, \ldots, K\}$, then*

$$\boldsymbol{B}_{(\text{sto-uc})}(\omega) \approx \frac{12}{5}\frac{1}{N}\frac{1}{Q^5}\sigma^2 \boldsymbol{P}^{-1}. \tag{3.31}$$

Proof. See Appendix F. □

Theorem 3.4 and Theorem 3.5 lead to the following two important implications:

1. Given the same number of sensors, co-prime and nested arrays can achieve a much better estimation performance than ULAs.

2. Given the same aperture, co-prime and nested arrays need many more snapshots to achieve the same estimation performance.

The first implication, which comes directly from Theorem 3.4 and Theorem 3.5, shows a great advantage of co-prime and nested arrays, in addition to their attractive ability to identify more uncorrelated sources than the number of sensors.

To understand the second implication, we consider a ULA with M^2 sensors. From Proposition 3.3, we know that the CRB of this ULA is $O(M^{-6})$. To achieve the same aperture, we need a co-prime (or nested) array with only $O(M)$ sensors. However, according Theorem 3.4 and Theorem 3.5, the resulting CRB of this co-prime (or nested) array will be only $O(M^{-5})$. Therefore, we need $O(M)$ times more snapshots to achieve the same estimation performance as the ULA. By thinning a ULA into a co-prime (or nested) array, we can reduce the number of sensors from $O(M^2)$ to $O(M)$, while keeping the array's ability to identify up to $O(M^2)$ uncorrelated sources. However, this thinning operation indeed comes with a cost: the variance of the estimated DOAs can be M times larger. The second implication shows the trade-off between the number of spatial samples and the number of temporal samples.

Remark. In the above analysis, the number of sources, K, is assumed to be smaller than the number of sensors, M. Because co-prime and nested arrays can identify more sources than

the number of sensors, it would be interesting to conduct a similar analysis for the $K \geq M$ case. However, when M is very large and $K \geq M$ holds, the sources become densely located within $(-\pi/2, \pi/2)$. In this case, $\omega_i - \omega_j$ is close to zero for any two different sources i and j, rendering the approximations in Appendix F invalid. Therefore, the results in Theorem 3.5 cannot be directly extended to the cases when $K \geq M$.

3.3 Numerical results

In this section, we use numerical experiments to demonstrate our analytical results. We first verify the MSE expression (3.3) introduced in Theorem 3.2 through Monte Carlo simulations. We then examine the application of (3.1) in predicting the resolvability of two closely placed sources, and analyze the asymptotic efficiency of both estimators from various aspects. Finally, we numerically verify our analytical results on the CRB in Theorem 3.3–3.5. In all experiments, we define the signal-to-noise ratio (SNR) as

$$\text{SNR} = 10 \log_{10} \frac{\min_{k=1,2,\ldots,K} p_k}{\sigma^2},$$

where K is the number of sources.

Throughout Section 3.3.1, 3.3.2 and 3.3.3, we consider the following three different types of linear arrays with the following sensor configurations:

- Co-prime Array [39]: $[0, 3, 5, 6, 9, 10, 12, 15, 20, 25]\lambda/2$
- Nested Array [38]: $[1, 2, 3, 4, 5, 10, 15, 20, 25, 30]\lambda/2$
- MRA [37]: $[0, 1, 4, 10, 16, 22, 28, 30, 33, 35]\lambda/2$

All three arrays share the same number of sensors, but difference apertures.

3.3.1 Numerical verification of Theorem 3.2

We first verify (3.4) via numerical simulations. We consider 11 sources with equal power, evenly placed between $-67.50°$ and $56.25°$, which is more than the number of sensors. We compare the difference between the analytical MSE and the empirical MSE under different combinations of SNR and snapshot numbers. The analytical MSE is defined by

$$\mathrm{MSE}_{\mathrm{an}} = \frac{1}{K}\sum_{k=1}^{K}\epsilon(\theta_k),$$

and the empirical MSE is defined by

$$\mathrm{MSE}_{\mathrm{em}} = \frac{1}{KL}\sum_{l=1}^{L}\sum_{k=1}^{K}\left(\hat{\theta}_k^{(l)} - \theta_k^{(l)}\right)^2,$$

where $\theta_k^{(l)}$ is the k-th DOA in the l-th trial, and $\hat{\theta}_k^{(l)}$ is the corresponding estimate.

Fig. 3.2 illustrates the relative errors between $\mathrm{MSE}_{\mathrm{an}}$ and $\mathrm{MSE}_{\mathrm{em}}$ obtained from 10,000 trials under various scenarios. It can be observed that $\mathrm{MSE}_{\mathrm{em}}$ and $\mathrm{MSE}_{\mathrm{an}}$ agree very well given enough snapshots and a sufficiently high SNR. It should be noted that at 0 dB SNR, (3.1) is quite accurate when 250 snapshots are available. In addition. there is no significant difference between the relative errors obtained from DA-MUSIC and those from SS-MUSIC. These observations are consistent with our assumptions, and verify Theorem 3.1 and Theorem 3.2.

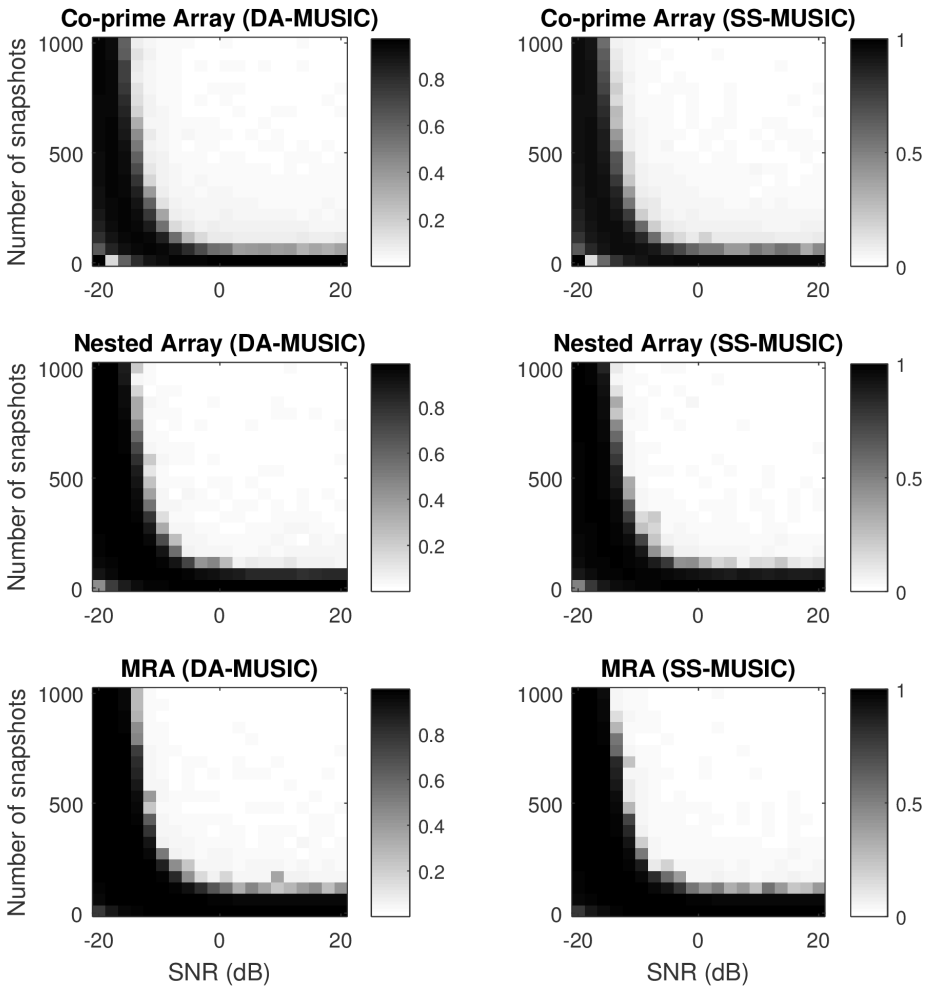


Figure 3.2: $|\mathrm{MSE}_{\mathrm{an}} - \mathrm{MSE}_{\mathrm{em}}|/\mathrm{MSE}_{\mathrm{em}}$ for different types of arrays under different numbers of snapshots and different SNRs.

We observe that in some of the low SNR regions, $|\mathrm{MSE}_{\mathrm{an}} - \mathrm{MSE}_{\mathrm{em}}|/\mathrm{MSE}_{\mathrm{em}}$ appears to be smaller even if the number of snapshots is limited. In such regions, $\mathrm{MSE}_{\mathrm{em}}$ actually "saturates", and $\mathrm{MSE}_{\mathrm{an}}$ happens to be close to the saturated value. Therefore, this observation does not imply that (3.4) is valid in such regions.

3.3.2 Prediction of resolvability

One direct application of Theorem 3.2 is predicting the resolvability of two closely located sources. We consider two sources with equal power, located at $\theta_1 = 30^\circ - \Delta\theta/2$, and $\theta_2 = 30^\circ + \Delta\theta/2$, where $\Delta\theta$ varies from 0.3° to 3.0°. We say the two sources are correctly resolved if the MUSIC algorithm is able to identify two sources, and the two estimated DOAs satisfy $|\hat{\theta}_i - \theta_i| < \Delta\theta/2$, for $i \in \{1, 2\}$. The probability of resolution is computed from 500 trials. For all trials, the number of snapshots is fixed at 500, the SNR is set to 0 dB, and SS-MUSIC is used.

For illustration purpose, we analytically predict the resolvability of the two sources via the following simple criterion:

$$\epsilon(\theta_1) + \epsilon(\theta_2) \underset{\text{Resolvable}}{\overset{\text{Unresovalble}}{\gtrless}} \Delta\theta. \tag{3.32}$$

Readers are directed to [74] for a more comprehensive criterion.

Fig. 3.3 illustrates the resolution performance of the three arrays under different $\Delta\theta$, as well as the thresholds predicted by (3.32). The MRA shows best resolution performance of the three arrays, which can be explained by the fact that the MRA has the largest aperture. The co-prime array, with the smallest aperture, shows the worst resolution performance. Despite the differences in resolution performance, the probability of resolution of each array drops to nearly zero at the predicted thresholds. This confirms that (3.4) provides a convenient way of predicting the resolvability of two close sources.

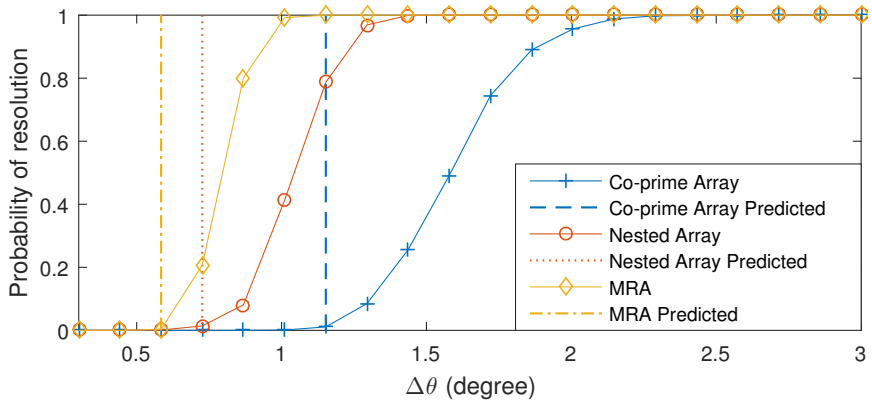


Figure 3.3: Probability of resolution vs. source separation, obtained from 500 trials. The number of snapshots is fixed at 500, and the SNR is set to 0 dB.

3.3.3 Asymptotic efficiency study

In this section, we utilize (3.4) and (3.12) to study the asymptotic statistical efficiency of DA-MUSIC and SS-MUSIC under different array geometries and parameter settings. We define their average efficiency as

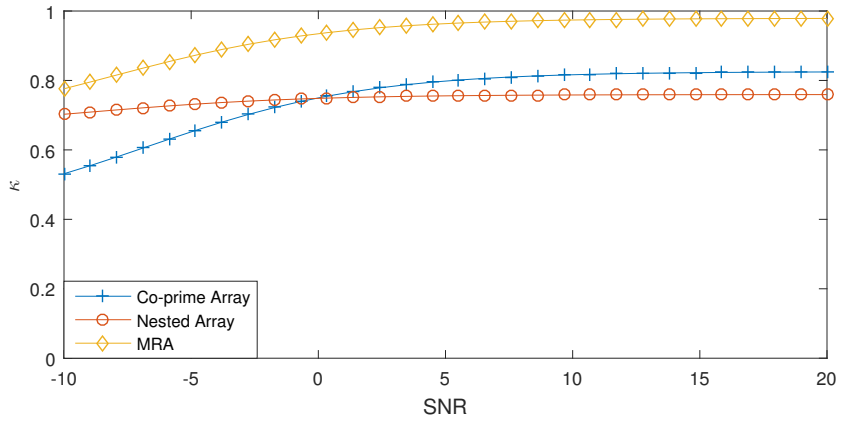
$$\kappa = \frac{\operatorname{tr} \boldsymbol{B}_{(\text{sto-uc})}(\boldsymbol{\theta})}{\sum_{k=1}^{K} \epsilon(\theta_k)}. \tag{3.33}$$

For efficient estimators we expect $\kappa = 1$, while for inefficient estimators we expect $0 \leq \kappa < 1$.

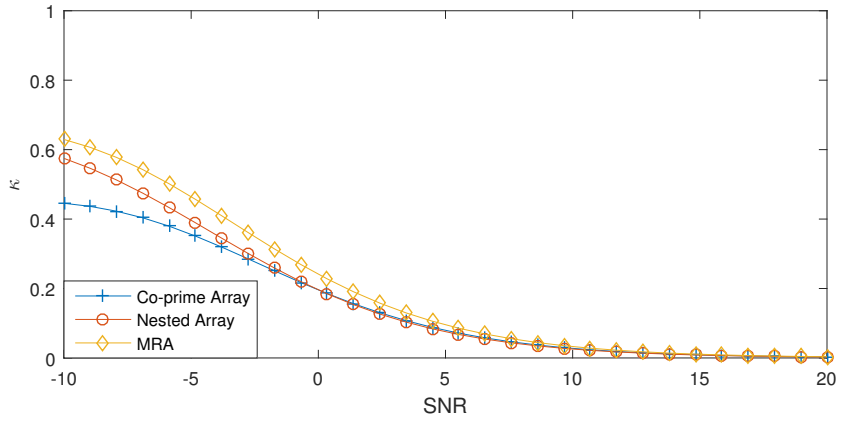
We first compare the κ value under different SNRs for the three different arrays. We consider three cases: $K = 1$, $K = 6$, and $K = 12$. The K sources are located at $\{-60^\circ + [120(k-1)/(K-1)]^\circ | k = 1, 2, \ldots, K\}$, and all sources have the same power. As shown in Fig. 3.4(a), when only one source is present, κ increases as the SNR increases for all three arrays. However, none of the arrays leads to efficient DOA estimation. Interestingly, despite being the least efficient geometry in the low SNR region, the co-prime array achieves higher efficiency than the nested array in the high SNR region. When $K = 6$, we can observe

in Fig. 3.4(b) that κ decreases to zero as SNR increases. This rather surprising behavior suggests that both DA-MUSIC and SS-MUSIC are not statistically efficient methods for DOA estimation when the number of sources is greater than one and less than the number of sensors. It is consistent with the implication of Proposition 3.2 when $K < M$. When $K = 12$, the number of sources exceeds the number of sensors. We can observe in Fig. 3.4(c) that κ also decreases as SNR increases. However, unlike the case when $K = 6$, κ converges to a positive value instead of zero. The above observations imply that DA-MUSIC and SS-MUSIC achieve higher degrees of freedom at the cost of decreased statistical efficiency. When statistical efficiency is concerned and the number of sources is less than the number of sensors, one might consider applying MUSIC directly to the original sample covariance $\boldsymbol{R}$ defined in (2.4) [75].

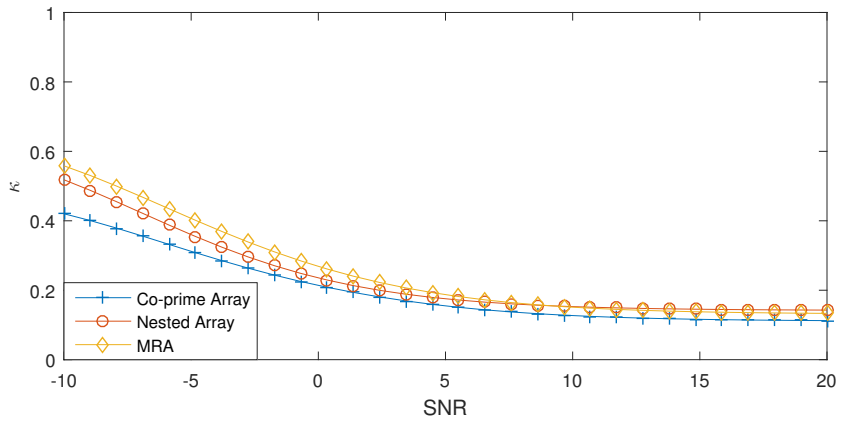
Next, we then analyze how κ is affected by angular separation. Two sources located at $-\Delta\theta$ and $\Delta\theta$ are considered. We compute the κ values under different choices of $\Delta\theta$ for all three arrays. For reference, we also include the empirical results obtained from 1000 trials. To satisfy the asymptotic assumption, the number of snapshots is fixed at 1000 for each trial. As shown in Fig. 3.5(a)–(c), the overall statistical efficiency decreases as the SNR increases from 0 dB to 10 dB for all three arrays, which is consistent with our previous observation in Fig. 3.4(b). We can also observe that the relationship between κ and the normalized angular separation $\Delta\theta/\pi$ is rather complex, as opposed to the traditional MUSIC algorithm (c.f. [61]). The statistical efficiency of DA-MUSIC and SS-MUSIC is highly dependent on array geometry and angular separation.



(a)

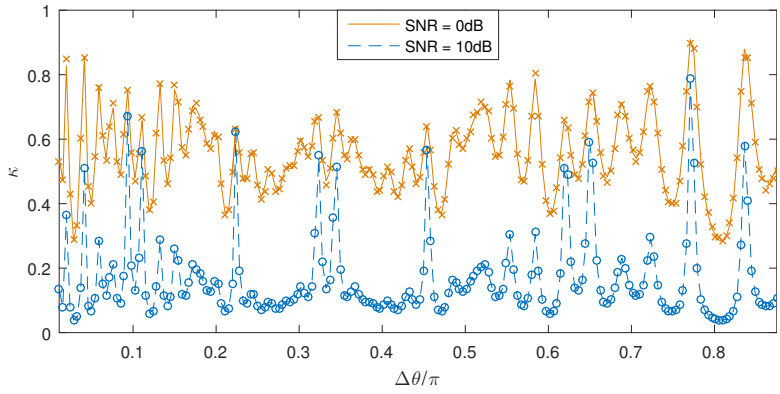


(b)

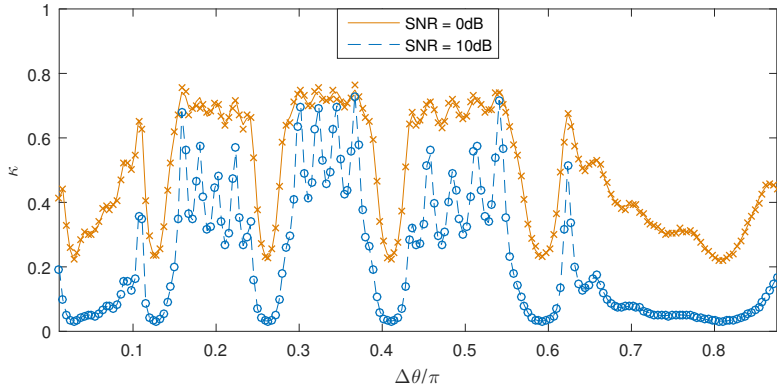


(c)

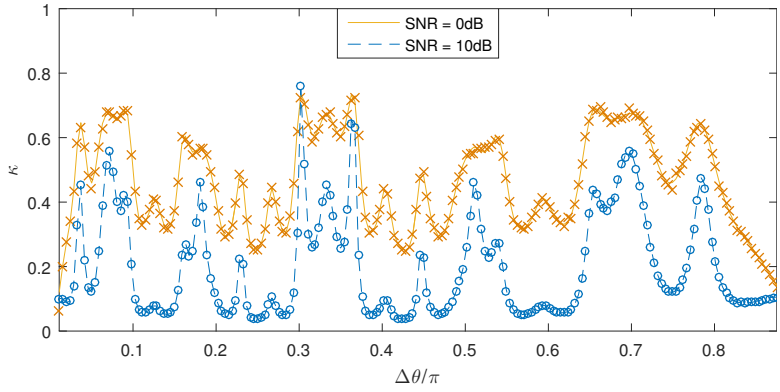
Figure 3.4: Average efficiency vs. SNR: (a) $K = 1$, (b) $K = 6$, (c) $K = 12$.



(a)



(b)



(c)

Figure 3.5: Average efficiency vs. angular separation for the co-prime array: (a) MRA, (b) nested array, (c) co-prime array. The solid lines and dashed lines are analytical values obtained from (3.33). The circles and crosses are empirical results averaged from 1000 trials.

3.3.4 Classical stochastic CRB vs. our CRB

In this section, we demonstrate Theorem 3.3 using numerical experiments. We consider the following four different sparse linear arrays:

- Co-prime (3,5): $[0, 3, 5, 6, 9, 10, 12, 15, 20, 25]d_0$;
- MRA 10[37]: $[0, 1, 4, 10, 16, 22, 28, 30, 33, 35]d_0$;
- Nested (4,6): $[0, 1, 2, 3, 4, 9, 14, 19, 24, 29]d_0$;
- Nested (5,5): $[0, 1, 2, 3, 4, 5, 11, 17, 23, 29]d_0$.

We consider six sources with equal power, whose the DOAs, θ_k, are given by $\theta_k = -\pi/3 + 2(k-1)/15\pi, k = 1, 2, \ldots, 6$. We vary the SNR from -20 dB to 20 dB and plot the relative difference between $\boldsymbol{B}_{(\text{sto})}$ and $\boldsymbol{B}_{(\text{sto-uc})}$ in Fig. 3.6. It can be observed that when the SNR is above 0 dB, the relative difference between $\boldsymbol{B}_{(\text{sto})}$ and $\boldsymbol{B}_{(\text{sto-uc})}$ for all four sparse linear arrays drastically decreases to zero as SNR increases. When the SNR is below 0 dB, $\boldsymbol{B}_{(\text{sto-uc})}$ becomes more optimistic and deviates from $\boldsymbol{B}_{(\text{sto})}$. These observations agree with our theoretical results in Theorem 3.3.

3.3.5 CRB vs. number of sensors

We next verify Theorem 3.4 and Theorem 3.5 via numerical experiments. We consider co-prime arrays generated by the co-prime pair $(Q, Q+1)$, and nested arrays generated by the parameter pair (Q, Q), where we vary Q between 3 and 20. We consider four different SNR settings: -20 dB, -10 dB, 0 dB, and 10 dB.

The results for the one source case are plotted in Fig. 3.7, where the only source is placed at the the origin. We can observe that, give large enough Q values and sufficient SNR, the simple approximation given in Theorem 3.4 is very close to the accurate value of $\boldsymbol{B}_{(\text{sto-uc})}$ for both co-prime and nested arrays. When the SNR is low, the noise variance term can no longer be neglected and our approximation deviates from the true values. When the value of Q is small, the contribution of the terms with lower degrees with respect to Q is no longer negligible, and our approximation is no longer accurate.

The results for the multiple sources case are plotted in Fig. 3.8, where we consider five sources with equal power, whose DOAs, ω_k, are give by $\omega_k = -\pi/3 + (k-1)/6\pi, k = 1, 2, \ldots, 5$. Similar to the results in Fig. 3.7, the CRBs of both the co-prime array and the nested arrays follow the trend predicted by Theorem 3.5 as Q increases. However, unlike the one-source case, the CRBs do not monotonically decrease. At some particular Q values, the CRBs deviate from the prediction by Theorem 3.5, regardless of the SNR. This is because we fix the source placement for all experiments. For some particular Q values, this placement may be close to a degenerative placement of the array generated by parameter Q, leading to larger CRB values than our approximations.

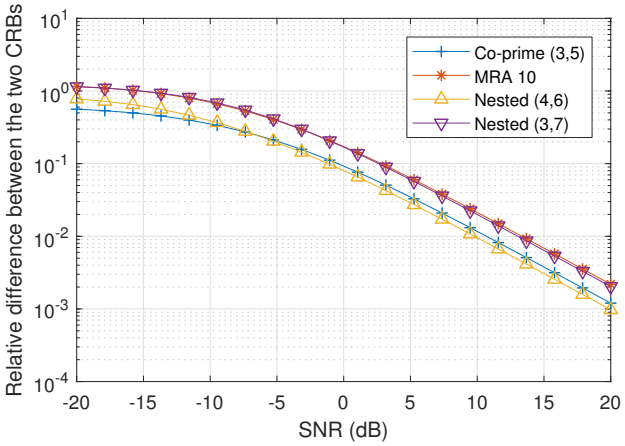
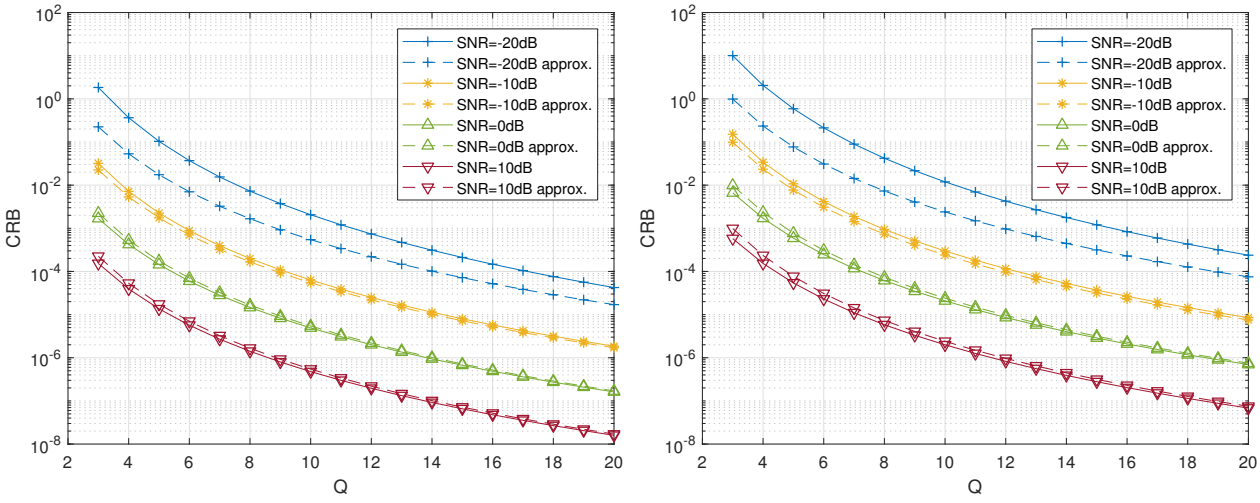
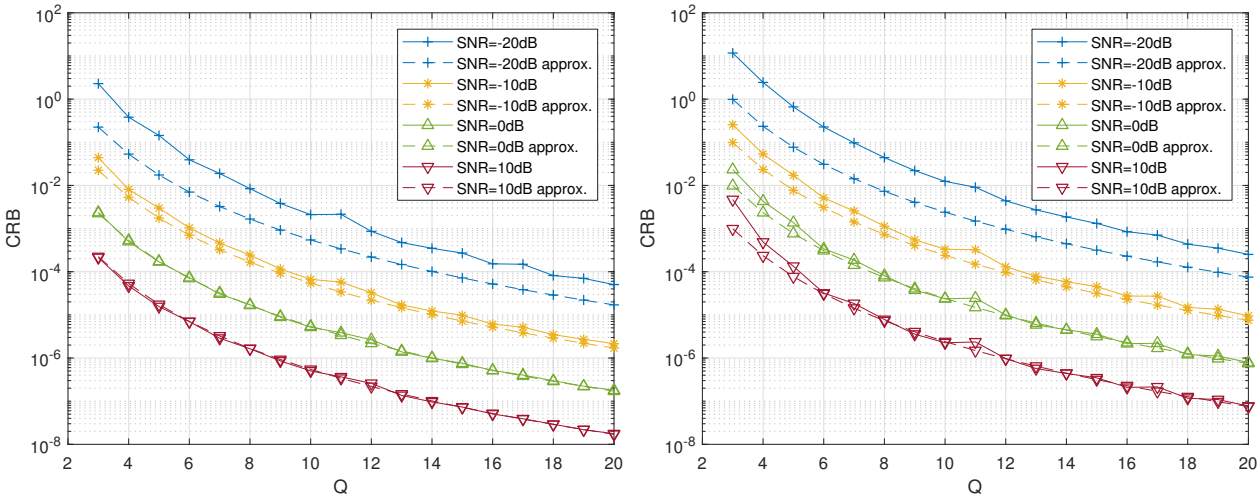


Figure 3.6: $|\operatorname{tr}(\boldsymbol{B}_{(\text{sto})} - \boldsymbol{B}_{(\text{sto-uc})})| / \operatorname{tr}(\boldsymbol{B}_{(\text{sto-uc})})$ for the four arrays under different SNRs.



(a) (b)

Figure 3.7: $\boldsymbol{B}_{(\text{sto-uc})}$ vs. Q for (a) co-prime arrays; (b) nested arrays. One source case. The solid lines represent accurate values computed using (3.16), while the dashed lines represent approximations given by Theorem 3.4.



(a) (b)

Figure 3.8: $\boldsymbol{B}_{(\text{sto-uc})}$ vs. Q for (a) co-prime arrays; (b) nested arrays. Multiple source case ($K = 5$). The solid lines represent accurate values computed using (3.16), while the dashed lines represent approximations given by Theorem 3.4.

In the previous experiments, we assume that the sources have equal power. However, Theorem 3.5 does not require all sources share the same power. Therefore, we conduct addition

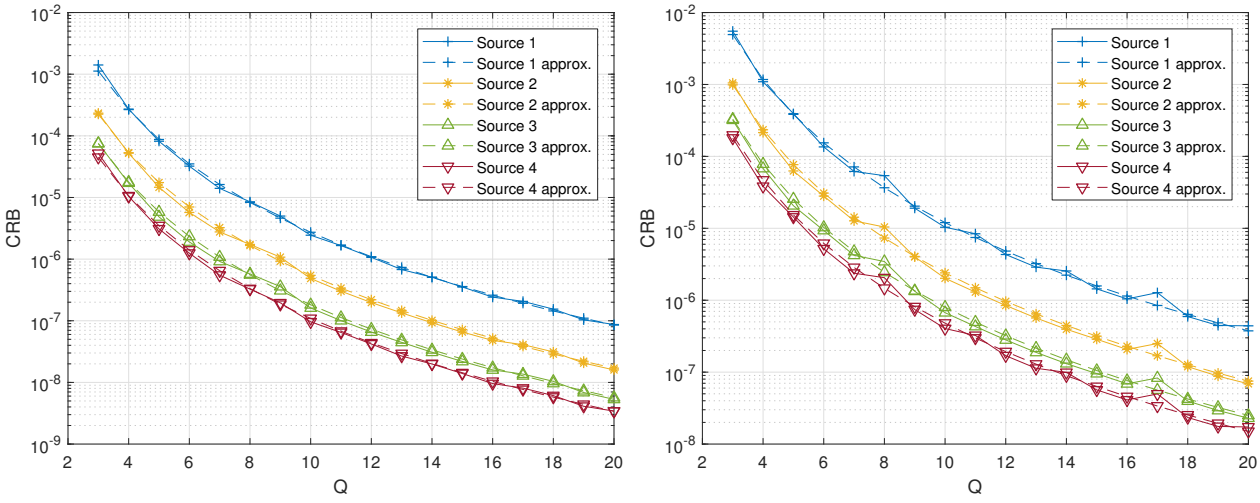


Figure 3.9: $\boldsymbol{B}_{(\text{sto-uc})}$ of individual sources vs. Q for (a) co-prime arrays and (b) nested arrays. Four sources with different powers are considered. The solid lines represent accurate values computed using (3.16), while the dashed lines represent approximations given by Theorem 3.4.

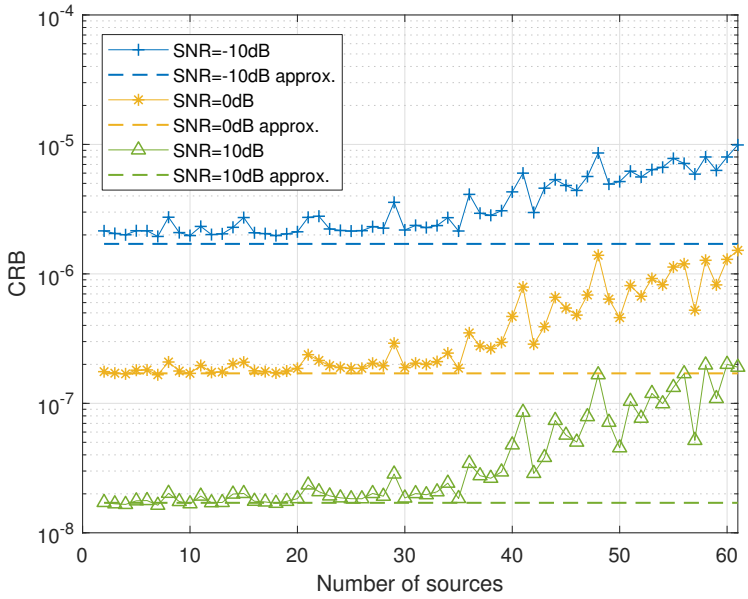


Figure 3.10: $\boldsymbol{B}_{(\text{sto-uc})}$ and the approximation given by (3.30) versus the number of sources. The co-prime array has 60 sensors. The solid lines represent accurate values computed using (3.16), while the dashed lines represent approximations given by (3.30).

experiments for the multiple source case when the source powers are not equal. We consider four sources with $\boldsymbol{p} = [2, 10, 30, 50]\sigma^2$. The results are plotted in Fig. 3.9. We observe that the actual CRBs closely follow the approximations given by Theorem 3.5 for all four sources. Because there is more than one source, we observe that the actual CRBs do not monotonically decrease as Q increase, similar to our observations in Fig. 3.8.

We close this section by addressing the comments in the last remark in Section 3.2.4 by using numerical experiments. We consider a co-prime array generated by the co-prime pair $(Q, Q+1)$ where $Q = 20$ is fixed. The resulting co-prime array has 60 sensors. We evenly place the DOAs, ω_k, at $\omega_k = -\pi/3 + 2(k-1)/(3K-3)\pi, k = 1, 2, \ldots, K$. We vary the number of sources, K, from 2 to 61. We plot the actual CRB, $\boldsymbol{B}_{\text{(sto-uc)}}$, together with the approximation given by (3.30) in Fig. 3.10. The real CRB values are denoted by solid lines, and the approximations given by (3.30) are denoted by dashed lines. We can observe that, when the number of sources is small, the actual CRB values are very close to our approximations, despite some fluctuations. However, as the number of sources increases, the actual CRB values begin to deviate from our approximations. In such cases, these sources become very close to each other, and the assumption that $\omega_m - \omega_n \in \Omega_Q^\delta \cap \Omega_{Q+1}^\delta$, $\forall m \neq n$, $m, n \in \{1, 2, \ldots, K\}$ becomes difficult to satisfy. Consequently, our approximation (3.30) is no longer accurate and the actual CRB values start to deviate from our approximation.

3.4 Chapter summary

In this chapter, we presented our key results from statistical performance analyses of sparse linear arrays. We theoretically proved that DA-MUSIC and SS-MUSIC share the same asymptotic MSE error expression, and then we derived this analytical MSE expression,

which can be applied to various types of sparse linear arrays. Our expression successfully explained the "saturation" behavior of SS-MUSIC observed in previous work. We derived and analyzed the CRB of sparse linear arrays under the assumption that the sources are uncorrelated, denoted by $\boldsymbol{B}_{(\text{sto-uc})}$. We showed that, when the SNR is high, $\boldsymbol{B}_{(\text{sto-uc})}$ coincides with the classical stochastic CRB, $\boldsymbol{B}_{(\text{sto})}$. We analyzed the behavior of $\boldsymbol{B}_{(\text{sto-uc})}$ for co-prime and nested arrays with a large number of sensors. We showed that, given a fixed number of sensors, M, $\boldsymbol{B}_{(\text{sto-uc})}$ for co-prime and nested arrays can decrease at a rate of $O(M^{-5})$, while $\boldsymbol{B}_{(\text{sto-uc})}$ for an M-sensor ULA decreases at a rate of only $O(M^{-3})$. We also showed that, when the aperture is fixed, co-prime and nested arrays need many more snapshots to achieve the same performance as ULAs, demonstrating the trade-off between the number of spatial samples and the number of temporal samples. Our results show both the pros and cons of sparse linear arrays, and will aid in choosing between sparse linear arrays and ULAs in practical problems.

Chapter 4

Perturbation Analysis of the Coarray Model

In the previous chapters, we assume that the arrays are perfectly calibrated[5]. However, this assumption may not hold in real-world applications. Various array imperfections exist, such as mutual coupling [76, 77], gain and phase errors [78, 79], and sensor location errors [80–82]. These array imperfections will generally degrade the DOA estimation performance [67, 83]. Various works consider the sensitivity of direction finding algorithms and the achievable bounds in the presence of array imperfections. In [80], the authors derived a hybrid Cramér-Rao bound on calibration and source localization for two-dimensional arrays in the presence of sensor location errors. Based on the derived CRB, the authors showed the conditions under which the CRB goes to zero as the SNR approaches infinity. In [67] and [84], the authors conducted a thorough performance analysis of subspace-based DOA estimators in the presence of model errors. In [85], the authors analyzed the resolution probability of the MUSIC algorithm, while taking into account model errors. However, the aforementioned

[5]This chapter is based on M. Wang, Z. Zhang, and A. Nehorai, "Performance analysis of coarray-based MUSIC in the presence of sensor location errors," *IEEE Trans. Signal Process.*, vol. 66, pp. 3074-3085, June 2018

analyses are based on the physical array model, and the number of sources is usually fewer than the number of sensors. The performance of direction finding algorithms based on the difference coarray model in the presence of array imperfections has not been widely analyzed. Recently, in [86], the authors conducted a performance analysis of uniform and nonuniform samplers based on the CRB of a grid-based model in the presence of model errors. These results can be applied to grid-based direction finding algorithms based on the difference coarray model. However, their analysis assumes one-dimensional perturbations along the array and that the DOAs lie on a predefined grid. Here, we neither restrict our analysis to one-dimensional perturbations, nor we assume a grid-based model.

In this chapter, we analyze the effect of sensor location errors on the difference coarray model. Unlike gain and phase errors, perturbed array manifolds are nonlinear with respect to sensor location errors [87]. This nonlinearity makes it more challenging to analyze the impact of sensor location errors. We first introduce a signal model for deterministic sensor location errors. We consider the commonly used SS-MUSIC [38] algorithm and derive a closed-form expression of its asymptotic MSE in the presence of small sensor location errors. Next, we present an brief extension of our analysis to incorporate stochastic (or time-variant) sensor location errors. We also derive the CRB on joint estimation of the DOAs and sensor location errors. Our CRB is applicable even if the number of sources exceeds the number of sensors. Finally, we use extensive numerical experiments to demonstration our analytical results. While our analyses are focused on sensor location errors, they can be readily extended to incorporate other array imperfections.

4.1 The deterministic error model

In this section, we consider the *deterministic error model*, where the sensor location errors are assumed deterministic and unknown. We derive the closed-form asymptotic MSE expression for SS-MUSIC under the deterministic error model, as well as the CRB for joint estimation of DOA parameters and sensor location errors.

4.1.1 Asymptotic MSE of SS-MUSIC

To obtain a more general perturbation model, we consider sensor location errors along both the x-axis and the y-axis[6]. We use $\boldsymbol{u} = [u_1, u_2, \ldots, u_M]^T$ to denote the sensor locations errors along the x-axis, and $\boldsymbol{v} = [v_1, v_2, \ldots, v_M]^T$ to denote the sensor location errors along the y-axis. The perturbed sensor locations are then given by $\tilde{\mathcal{D}} = \{(d_1 + u_1, v_1), (d_2 + u_2, v_2), \ldots, (d_M + u_M, v_M)\}$. When the sensor location errors are large, the linear array structure will be completely destroyed, resulting large DOA estimation errors that are difficult to characterize. Therefore, our performance analysis will focus on cases when the sensor location errors are small. In this chapter, in addition to assumptions **A1**–**A4**, we make the following additional assumption:

A5 The sensor location errors are small compared with d_0.

[6]We do not need to consider the perturbations along the z-axis under the far-field and co-planar assumption of the source signals.

Let $\boldsymbol{\delta} = [\boldsymbol{u}^T \ \boldsymbol{v}^T]^T$ denote the collection of sensor location error parameters. Under assumption **A1**–**A5**, the N snapshots received by the perturbed array can be expressed as

$$\tilde{\boldsymbol{y}}(t) = \tilde{\boldsymbol{A}}(\boldsymbol{\theta}, \boldsymbol{\delta})\boldsymbol{x}(t) + \boldsymbol{n}(t), t = 1, 2, \ldots, N, \tag{4.1}$$

where $\tilde{\boldsymbol{A}}(\boldsymbol{\theta}, \boldsymbol{\delta})$ denotes the perturbed steering matrix, and

$$\tilde{A}_{ik}(\boldsymbol{\theta}, \boldsymbol{\delta}) = \exp\left[j\frac{2\pi}{\lambda}(d_i \sin\theta_k + u_i \sin\theta_k + v_i \cos\theta_k)\right].$$

The perturbed covariance matrix is then given by

$$\tilde{\boldsymbol{R}} = \tilde{\boldsymbol{A}}(\boldsymbol{\theta}, \boldsymbol{\delta})\boldsymbol{P}\tilde{\boldsymbol{A}}^H(\boldsymbol{\theta}, \boldsymbol{\delta}) + \sigma^2\boldsymbol{I}. \tag{4.2}$$

The corresponding observation model of the difference coarray is then given by

$$\tilde{\boldsymbol{r}} = (\tilde{\boldsymbol{A}}^* \odot \tilde{\boldsymbol{A}})\boldsymbol{p} + \sigma^2 \operatorname{vec}(\boldsymbol{I}). \tag{4.3}$$

Here we drop the explicit dependencies on $\boldsymbol{\theta}, \boldsymbol{\delta}$ for notational simplicity. The matrix $(\tilde{\boldsymbol{A}}^* \odot \tilde{\boldsymbol{A}})$ now resembles a steering matrix of the perturbed difference coarray, whose sensor locations are given by $\tilde{\mathcal{D}}_{\text{co}} = \{(d_m - d_n + u_m - u_n, v_m - v_n) | m, n = 1, 2, \ldots, M\}$. As illustrated in Fig. 4.1, the perturbed difference coarray no longer embeds a ULA, and can no longer be divided into multiple overlapping subarrays of the same shape. Consequently, applying SS-MUSIC to the perturbed difference coarray model without error compensations will lead to degraded DOA estimation performance.

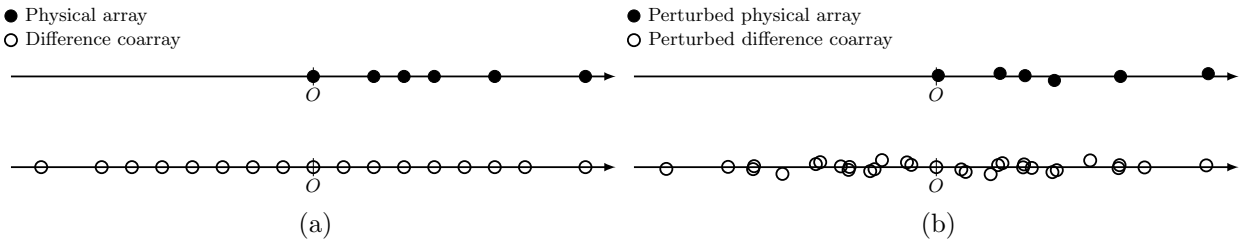


Figure 4.1: Illustration of a perturbed difference coarray: (a) a co-prime array and its difference coarray; (b) a perturbed co-prime array and its perturbed difference coarray.

To establish the link between the coarray perturbation and the DOA estimation errors, we start with the perturbed steering matrix $\tilde{\boldsymbol{A}}$. Because $\tilde{\boldsymbol{A}}$ is analytic in the neighborhood of $\boldsymbol{\delta} = \boldsymbol{0}$, we can linearize $\tilde{\boldsymbol{A}}$ around $\boldsymbol{\delta} = \boldsymbol{0}$ via the first-order Taylor expansion under assumption **A5**:

$$\tilde{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{U}\tilde{\boldsymbol{A}}_u + \boldsymbol{V}\tilde{\boldsymbol{A}}_v + o(\boldsymbol{\delta}), \tag{4.4}$$

where

$$\boldsymbol{U} = \mathrm{diag}(u_1, u_2, \ldots, u_M), \tag{4.5a}$$

$$\boldsymbol{V} = \mathrm{diag}(v_1, v_2, \ldots, v_M), \tag{4.5b}$$

$$\tilde{\boldsymbol{A}}_u = j\frac{2\pi}{\lambda}\boldsymbol{A}\boldsymbol{D}_{\mathrm{s}}, \tag{4.5c}$$

$$\tilde{\boldsymbol{A}}_v = j\frac{2\pi}{\lambda}\boldsymbol{A}\boldsymbol{D}_{\mathrm{c}}, \tag{4.5d}$$

$$\boldsymbol{D}_{\mathrm{s}} = \mathrm{diag}(\sin\theta_1, \sin\theta_2, \ldots, \sin\theta_K), \tag{4.5e}$$

$$\boldsymbol{D}_{\mathrm{c}} = \mathrm{diag}(\cos\theta_1, \cos\theta_2, \ldots, \cos\theta_K), \tag{4.5f}$$

and $o(\boldsymbol{\delta})$ denotes the higher order terms with respect to $\boldsymbol{\delta}$. The perturbed covariance matrix $\tilde{\boldsymbol{R}}$ can then be approximated as

$$\tilde{\boldsymbol{R}} = \boldsymbol{R} + \boldsymbol{U}\tilde{\boldsymbol{A}}_u\boldsymbol{P}\boldsymbol{A}^H + \boldsymbol{A}\boldsymbol{P}\tilde{\boldsymbol{A}}_u^H\boldsymbol{U} + \boldsymbol{V}\tilde{\boldsymbol{A}}_v\boldsymbol{P}\boldsymbol{A}^H + \boldsymbol{A}\boldsymbol{P}\tilde{\boldsymbol{A}}_v^H\boldsymbol{V} + o(\boldsymbol{\delta}). \tag{4.6}$$

In practice, the true covariance matrix is unknown, and we obtain only the estimate of $\tilde{\boldsymbol{R}}$ with $\hat{\boldsymbol{R}} = \frac{1}{N}\sum_{t=1}^{N} \boldsymbol{y}(t)\boldsymbol{y}(t)^H$. Hence, the discrepancy between the estimate, $\hat{\boldsymbol{R}}$, and nominal covariance matrix, $\boldsymbol{R}$, can be decomposed into two parts:

$$\Delta\boldsymbol{R} = \hat{\boldsymbol{R}} - \boldsymbol{R} = \underbrace{(\hat{\boldsymbol{R}} - \tilde{\boldsymbol{R}})}_{\boldsymbol{E}} + \underbrace{(\tilde{\boldsymbol{R}} - \boldsymbol{R})}_{\boldsymbol{G}}, \tag{4.7}$$

where $\boldsymbol{E}$ denotes the estimation errors resulting from finite snapshots, and $\boldsymbol{G}$ denotes the estimation errors resulting from sensor location errors. To derive the asymptotic MSE expression of SS-MUSIC in the presence of sensor location errors, we make use of Theorem 3.1.

It is straightforward to verify that $\hat{\boldsymbol{R}}$ is still Hermitian in the presence of sensor location errors. Combining (4.7) and Theorem 3.1 and neglecting all the high order terms, we obtain

$$\Delta\theta_k \doteq -(\gamma_k p_k)^{-1}\Re[\boldsymbol{\xi}_k^T(\boldsymbol{e}+\boldsymbol{g})], \tag{4.8}$$

where $\doteq$ denotes equality up to the first order, $\boldsymbol{e} = \mathrm{vec}(\boldsymbol{E})$, and $\boldsymbol{g} = \mathrm{vec}(\boldsymbol{G})$. Hence, for a large number of snapshots, the asymptotic MSE can be evaluated as

$$\mathbb{E}[\Delta\theta_k^2] \doteq \frac{\mathbb{E}\{[\Re(\boldsymbol{\xi}_k^T(\boldsymbol{e}+\boldsymbol{g}))]^2\}}{\gamma_k^2 p_k^2}. \tag{4.9}$$

Using the fact that $\Re(\boldsymbol{AB}) = \Re(\boldsymbol{A})\Re(\boldsymbol{B}) - \Im(\boldsymbol{A})\Im(\boldsymbol{B})$, we can expand the numerator in (4.9) as follows:

$$\begin{aligned}
&\mathbb{E}\{[\Re(\boldsymbol{\xi}_k^T(\boldsymbol{e}+\boldsymbol{g}))]^2\} \\
=&\Re(\boldsymbol{\xi}_k)^T\mathbb{E}[\Re(\boldsymbol{e}+\boldsymbol{g})\Re(\boldsymbol{e}+\boldsymbol{g})^T]\Re(\boldsymbol{\xi}_k) + \Im(\boldsymbol{\xi}_k)^T\mathbb{E}[\Im(\boldsymbol{e}+\boldsymbol{g})\Im(\boldsymbol{e}+\boldsymbol{g})^T]\Im(\boldsymbol{\xi}_k) \\
&- 2\Re(\boldsymbol{\xi}_k)^T\mathbb{E}[\Re(\boldsymbol{e}+\boldsymbol{g})\Im(\boldsymbol{e}+\boldsymbol{g})^T]\Im(\boldsymbol{\xi}_k).
\end{aligned} \tag{4.10}$$

Because $\mathbb{E}[\boldsymbol{e}] = \mathbf{0}$, we have

$$\begin{aligned}
\mathbb{E}[\Re(\boldsymbol{e}+\boldsymbol{g})\Re(\boldsymbol{e}+\boldsymbol{g})^T] &= \mathbb{E}[\Re(\boldsymbol{e})\Re(\boldsymbol{e})^T] + \Re(\boldsymbol{g})\Re(\boldsymbol{g})^T, \\
\mathbb{E}[\Im(\boldsymbol{e}+\boldsymbol{g})\Im(\boldsymbol{e}+\boldsymbol{g})^T] &= \mathbb{E}[\Im(\boldsymbol{e})\Im(\boldsymbol{e})^T] + \Im(\boldsymbol{g})\Im(\boldsymbol{g})^T, \\
\mathbb{E}[\Re(\boldsymbol{e}+\boldsymbol{g})\Im(\boldsymbol{e}+\boldsymbol{g})^T] &= \mathbb{E}[\Re(\boldsymbol{e})\Im(\boldsymbol{e})^T] + \Re(\boldsymbol{g})\Im(\boldsymbol{g})^T.
\end{aligned}$$

Hence we can expand (4.10) as

$$\begin{aligned}
\mathbb{E}\{[\Re(\boldsymbol{\xi}_k^T(\boldsymbol{e}+\boldsymbol{g}))]^2\} =& \Re(\boldsymbol{\xi}_k)^T\mathbb{E}[\Re(\boldsymbol{e})\Re(\boldsymbol{e})^T]\Re(\boldsymbol{\xi}_k) + \Im(\boldsymbol{\xi}_k)^T\mathbb{E}[\Im(\boldsymbol{e})\Im(\boldsymbol{e})^T]\Im(\boldsymbol{\xi}_k) \\
&- 2\Re(\boldsymbol{\xi}_k)^T\mathbb{E}[\Re(\boldsymbol{e})\Im(\boldsymbol{e})^T]\Im(\boldsymbol{\xi}_k) + \Re(\boldsymbol{\xi}_k)^T\Re(\boldsymbol{g})\Re(\boldsymbol{g})^T\Re(\boldsymbol{\xi}_k) \\
&+ \Im(\boldsymbol{\xi}_k)^T\Im(\boldsymbol{g})\Im(\boldsymbol{g})^T\Im(\boldsymbol{\xi}_k) - 2\Re(\boldsymbol{\xi}_k)^T\Re(\boldsymbol{g})\Im(\boldsymbol{g})^T\Im(\boldsymbol{\xi}_k) \\
=& \Re[\boldsymbol{\xi}_k^H(\tilde{\boldsymbol{R}} \otimes \tilde{\boldsymbol{R}}^T)\boldsymbol{\xi}_k]/N + \Re(\boldsymbol{g}^T\boldsymbol{\xi}_k)^T\Re(\boldsymbol{g}^T\boldsymbol{\xi}_k).
\end{aligned} \tag{4.11}$$

The first three terms evaluate into $\Re[\boldsymbol{\xi}_k^H(\tilde{\boldsymbol{R}} \otimes \tilde{\boldsymbol{R}}^T)\boldsymbol{\xi}_k]/N$. The derivation follows the same idea as in [71, Appendix C], but with $\boldsymbol{R}$ replaced with $\tilde{\boldsymbol{R}}$. The second three terms can be combined into $\Re(\boldsymbol{g}^T\boldsymbol{\xi}_k)^T\Re(\boldsymbol{g}^T\boldsymbol{\xi}_k)$. To obtain the final MSE expression, we still need to expand $\boldsymbol{g}$ in terms of $\boldsymbol{\delta}$, which requires Lemma 4.1 below.

Lemma 4.1. *Let* $\boldsymbol{D} = \operatorname{diag}(\boldsymbol{d})$ *be a diagonal matrix. Then* $\operatorname{vec}(\boldsymbol{DX}) = (\boldsymbol{X}^T \odot \boldsymbol{I})\boldsymbol{d}$ *and* $\operatorname{vec}(\boldsymbol{XD}) = (\boldsymbol{I} \odot \boldsymbol{X})\boldsymbol{d}$ *for any matrix* $\boldsymbol{X}$ *with a proper shape.*

Proof. The two equalities follow immediately from the following fact [70]: for any diagonal matrix $\boldsymbol{X}$ and any two matrices $\boldsymbol{A}$, $\boldsymbol{B}$ with proper shapes,

$$\operatorname{vec}(\boldsymbol{AXB}) = (\boldsymbol{B}^T \odot \boldsymbol{A})\operatorname{diag}(\boldsymbol{X}). \tag{4.12}$$

□

Using Lemma 4.1 and (4.6), we can rewrite $\boldsymbol{g}$ as $\boldsymbol{B\delta} + o(\boldsymbol{\delta})$, where $\boldsymbol{B} = [\boldsymbol{B_u}\ \boldsymbol{B_v}]$ and

$$\boldsymbol{B_u} = \boldsymbol{I} \odot (\boldsymbol{AP}\tilde{\boldsymbol{A}}_u^H) + (\boldsymbol{AP}\tilde{\boldsymbol{A}}_u^H)^* \odot \boldsymbol{I}, \tag{4.13a}$$

$$\boldsymbol{B_v} = \boldsymbol{I} \odot (\boldsymbol{AP}\tilde{\boldsymbol{A}}_v^H) + (\boldsymbol{AP}\tilde{\boldsymbol{A}}_v^H)^* \odot \boldsymbol{I}. \tag{4.13b}$$

Substituting the expression for $\boldsymbol{g}$ back into (4.11), we obtain the following result.

Corollary 4.1. *Under the deterministic error model, the asymptotic MSE of SS-MUSIC for the k-th DOA in the presence of small sensor location errors is given by*

$$\frac{1}{p_k^2\gamma_k^2}\left\{\frac{1}{N}\Re[\boldsymbol{\xi}_k^H(\tilde{\boldsymbol{R}}\otimes\tilde{\boldsymbol{R}}^T)\boldsymbol{\xi}_k] + [\boldsymbol{\delta}^T\Re(\boldsymbol{B}^T\boldsymbol{\xi}_k)]^2\right\}, \tag{4.14}$$

where $\boldsymbol{\xi}_k$ and $\boldsymbol{B}$ follow the same definition in Theorem 3.1 and (4.13a)–(4.13b).

The asymptotic MSE (4.14) consists of two terms. The first term results from the estimation errors of the covariance matrix, which will vanish as the number of snapshots goes to infinity. It should be also noted that this term is also affected by the sensor location errors, because $\tilde{\boldsymbol{R}}$ depends on $\boldsymbol{\delta}$. However, given a sufficient number of snapshots N, such an effect is negligible after being divided by N. The second term is the result from sensor location errors, which will not vanish as the number of snapshots goes to infinity, leading to a constant bias among the DOA estimates.

Corollary 4.1 gives the asymptotic MSE for a particular realization of the sensor locations errors, $\boldsymbol{\delta}$. We are also interested in the ensemble behavior of (4.14) under different realizations of sensor location errors. Following the idea of the hybrid CRB, we assume that the sensor location errors $\boldsymbol{\delta}$ follows a Gaussian prior $\mathcal{N}(\boldsymbol{0}, \boldsymbol{C})$ [3], and evaluate the average asymptotic MSE under this Gaussian prior. The results are summarized in Corollary 4.2.

Corollary 4.2. *Let $\boldsymbol{\delta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{C})$, where $\|\boldsymbol{C}\|$ is sufficiently small such that the high order moments of $\boldsymbol{\delta}/d_0$ are $o(\|\boldsymbol{C}\|)$. Then then the average asymptotic MSE (AAMSE) of SS-MUSIC in the presence of sensor location errors is given by*

$$\frac{1}{p_k^2 \gamma_k^2} \left\{ \frac{1}{N} \Re[\boldsymbol{\xi}_k^H (\boldsymbol{R} \otimes \boldsymbol{R}^T) \boldsymbol{\xi}_k] + \Re(\boldsymbol{B}^T \boldsymbol{\xi}_k)^T \boldsymbol{C} \Re(\boldsymbol{B}^T \boldsymbol{\xi}_k) \right\}, \tag{4.15}$$

Proof. Let $\boldsymbol{\Delta} = \boldsymbol{U} \tilde{\boldsymbol{A}}_u \boldsymbol{P} \boldsymbol{A}^H + \boldsymbol{A} \boldsymbol{P} \tilde{\boldsymbol{A}}_u^H \boldsymbol{U} + \boldsymbol{V} \tilde{\boldsymbol{A}}_v \boldsymbol{P} \boldsymbol{A}^H + \boldsymbol{A} \boldsymbol{P} \tilde{\boldsymbol{A}}_v^H \boldsymbol{V}$. Using (4.6), we have

$$\tilde{\boldsymbol{R}} \otimes \tilde{\boldsymbol{R}}^T = \boldsymbol{R} \otimes \boldsymbol{R}^T + \boldsymbol{R} \otimes \boldsymbol{\Delta}^T + \boldsymbol{\Delta} \otimes \boldsymbol{R}^T + o(\|\boldsymbol{C}\|).$$

Because $\mathbb{E}_{\boldsymbol{\delta}}[\boldsymbol{\Delta}] = \mathbf{0}$, using the assumption that the high order moments of $\boldsymbol{\delta}/d_0$ are $o(\|\boldsymbol{C}\|)$, we obtain $\mathbb{E}_{\boldsymbol{\delta}}[\tilde{\boldsymbol{R}} \otimes \tilde{\boldsymbol{R}}^T] \doteq \boldsymbol{R} \otimes \boldsymbol{R}^T$. This leads to the first term in (4.15). The second term in (4.15) is due to the fact that $\mathbb{E}_{\boldsymbol{\delta}}[\boldsymbol{\delta}\boldsymbol{\delta}^T] = \boldsymbol{C}$. The remaining high order terms are still $o(\|\boldsymbol{C}\|)$ under the assumption that that the high order moments of $\boldsymbol{\delta}/d_0$ are $o(\|\boldsymbol{C}\|)$. □

Because the second error term in (4.15) is linear in $\boldsymbol{C}$, we can use $\Re(\boldsymbol{B}^T \boldsymbol{\xi}_k)^T \Re(\boldsymbol{B}^T \boldsymbol{\xi}_k)$ as a sensitivity metric of the robustness of SS-MUSIC against the sensor location errors for the k-th DOA. It can be observed that this term is affected by both the physical array geometry and the coarray geometry. The physical array geometry is encoded in the matrix $\boldsymbol{B}$, which depends on the nominal physical array steering matrix $\boldsymbol{A}$. The coarray geometry is encoded in the vector $\boldsymbol{\xi}_k$, which depends on the coarray steering matrix $\boldsymbol{A}_{\text{co}}$ as well as the transform matrix $\boldsymbol{F}$. This observation implies that even if two sparse linear arrays share the same coarray structure, their sensitivities against model errors may not be the same.

Corollary 4.3. *Assume all sources share the same power p. Let $\varepsilon(\theta_k)$ denote the AAMSE of the k-th DOA in Corollary 4.2. Fixing σ^2, we have*

$$\lim_{p\to\infty} \varepsilon(\theta_k) = \frac{1}{\gamma_k^2}\left\{ \frac{1}{N}\|\boldsymbol{\xi}_k^H(\boldsymbol{A}\otimes\boldsymbol{A}^*)\|_2^2 + \Re(\bar{\boldsymbol{B}}^T\boldsymbol{\xi}_k)^T\boldsymbol{C}\Re(\bar{\boldsymbol{B}}^T\boldsymbol{\xi}_k)\right\}, \tag{4.16}$$

where $\bar{\boldsymbol{B}} = [\bar{\boldsymbol{B}}_{\boldsymbol{u}}\ \bar{\boldsymbol{B}}_{\boldsymbol{v}}]$, and

$$\bar{\boldsymbol{B}}_{\boldsymbol{u}} = \boldsymbol{I}\odot(\boldsymbol{A}\tilde{\boldsymbol{A}}_u^H) + (\boldsymbol{A}\tilde{\boldsymbol{A}}_u^H)^*\odot\boldsymbol{I},$$
$$\bar{\boldsymbol{B}}_{\boldsymbol{v}} = \boldsymbol{I}\odot(\boldsymbol{A}\tilde{\boldsymbol{A}}_v^H) + (\boldsymbol{A}\tilde{\boldsymbol{A}}_v^H)^*\odot\boldsymbol{I}.$$

Proof. The result follows directly from Corollary 4.2 and [71, Corollary 1]. □

The first term in (4.16) is the limiting expression of the asymptotic MSE of SS-MUSIC in the absence of sensor location errors as the SNR approaches infinity, which is generally non-zero when multiple sources are present [71]. The second term in (4.16) is the result from sensor location errors. Because $\bar{\boldsymbol{B}}$ is independent of the source power p, we conclude that the DOA estimation bias of SS-MUSIC introduced by the sensor location errors cannot be mitigated by increasing the SNR alone.

4.1.2 CRB for joint estimation of DOA and location error parameters

In this section, we derive the CRB for general sparse linear arrays under the deterministic error model. In addition to the DOAs, source powers, and noise power, we also treat sensor location errors as unknown parameters. To obtain a more general expression of the FIM, we assume that the precise sensor locations are partially known. This assumption includes the

case when sensor location errors among all the sensors are unknown. Let $\{i_1, i_2, \ldots, i_{M_1}\} \subseteq \{1, 2, \ldots, M\}$ denote the indices of sensors with unknown location errors along the x-axis, and $\{l_1, l_2, \ldots, l_{M_2}\} \subseteq \{1, 2, \ldots, M\}$ denote the indices of sensors with unknown location errors along the y-axis. The collection of unknown parameters is given by the $(2K+M_1+M_2+1)\times 1$ real vector:

$$\boldsymbol{\eta} = [\boldsymbol{\theta}^T, \boldsymbol{p}^T, u_{i_1}, \cdots, u_{i_{M_1}}, v_{l_1}, \cdots, v_{l_{M_2}}, \sigma^2]^T. \tag{4.17}$$

The FIM is then given by:

Proposition 4.1. *Under assumptions **A1**–**A3**, the FIM of the deterministic error model is give by*

$$\boldsymbol{J} = N\boldsymbol{M}^H(\tilde{\boldsymbol{R}}^T \otimes \tilde{\boldsymbol{R}})^{-1}\boldsymbol{M}. \tag{4.18}$$

Here,

$$\boldsymbol{M} = \left[\frac{\partial \tilde{\boldsymbol{r}}}{\partial \boldsymbol{\theta}} \; \frac{\partial \tilde{\boldsymbol{r}}}{\partial \boldsymbol{p}} \; \frac{\partial \tilde{\boldsymbol{r}}}{\partial \boldsymbol{u}} \; \frac{\partial \tilde{\boldsymbol{r}}}{\partial \boldsymbol{v}} \; \frac{\partial \tilde{\boldsymbol{r}}}{\partial \sigma^2}\right], \tag{4.19}$$

where

$$\frac{\partial \tilde{\boldsymbol{r}}}{\partial \boldsymbol{\theta}} = (\tilde{\boldsymbol{A}}_{\boldsymbol{\theta}}^* \odot \tilde{\boldsymbol{A}} + \tilde{\boldsymbol{A}}^* \odot \tilde{\boldsymbol{A}}_{\boldsymbol{\theta}})\boldsymbol{P}, \tag{4.20a}$$

$$\frac{\partial \tilde{\boldsymbol{r}}}{\partial \boldsymbol{p}} = \tilde{\boldsymbol{A}}^* \odot \tilde{\boldsymbol{A}}, \tag{4.20b}$$

$$\frac{\partial \tilde{\boldsymbol{r}}}{\partial \boldsymbol{u}} = [(\tilde{\boldsymbol{A}}\boldsymbol{P}\tilde{\boldsymbol{A}}_u^H)^*\boldsymbol{L}_1] \odot \boldsymbol{L}_1 + \boldsymbol{L}_1 \odot (\tilde{\boldsymbol{A}}\boldsymbol{P}\tilde{\boldsymbol{A}}_u^H\boldsymbol{L}_1), \tag{4.20c}$$

$$\frac{\partial \tilde{\boldsymbol{r}}}{\partial \boldsymbol{v}} = [(\tilde{\boldsymbol{A}}\boldsymbol{P}\tilde{\boldsymbol{A}}_v^H)^*\boldsymbol{L}_2] \odot \boldsymbol{L}_2 + \boldsymbol{L}_2 \odot (\tilde{\boldsymbol{A}}\boldsymbol{P}\tilde{\boldsymbol{A}}_v^H\boldsymbol{L}_2), \tag{4.20d}$$

$$\frac{\partial \tilde{\boldsymbol{r}}}{\partial \sigma^2} = \mathrm{vec}(\boldsymbol{I}_M), \tag{4.20e}$$

and

$$\begin{aligned}
\boldsymbol{L}_1 &= [\boldsymbol{e}_M^{(i_1)} \ \boldsymbol{e}_M^{(i_2)} \ \cdots \ \boldsymbol{e}_M^{(i_{M_1})}], \\
\boldsymbol{L}_2 &= [\boldsymbol{e}_M^{(l_1)} \ \boldsymbol{e}_M^{(l_2)} \ \cdots \ \boldsymbol{e}_M^{(l_{M_2})}], \\
\tilde{\boldsymbol{A}}_{\boldsymbol{\theta}} &= \left[\frac{\partial \tilde{\boldsymbol{a}}(\theta_1)}{\partial \theta_1} \ \frac{\partial \tilde{\boldsymbol{a}}(\theta_2)}{\partial \theta_2} \ \cdots \ \frac{\partial \tilde{\boldsymbol{a}}(\theta_K)}{\partial \theta_K} \right].
\end{aligned}$$

Proof. The (m,n)-th element of the single snapshot FIM for the observation model (4.1) is given by [52, 68]

$$\boldsymbol{J}_{mn} = \operatorname{tr}\left[\frac{\partial \tilde{\boldsymbol{R}}}{\partial \eta_m} \tilde{\boldsymbol{R}}^{-1} \frac{\partial \tilde{\boldsymbol{R}}}{\partial \eta_n} \tilde{\boldsymbol{R}}^{-1} \right].$$

Using the properties that $\operatorname{tr}(\boldsymbol{AB}) = \operatorname{vec}(\boldsymbol{A}^T)^T \operatorname{vec}(\boldsymbol{B})$, and that $\operatorname{vec}(\boldsymbol{AXB}) = (\boldsymbol{B}^T \otimes \boldsymbol{A}) \operatorname{vec}(\boldsymbol{X})$ [70], we can express the FIM as (4.18).

To obtain the FIM, we need to evaluate the partial derivatives in (4.19). The partial derivatives of $\tilde{\boldsymbol{r}}$ with respect to $\boldsymbol{\theta}$, $\boldsymbol{p}$, and σ^2 have been derived in [69, 71, 88]. We will focus on deriving the partial derivatives of $\tilde{\boldsymbol{r}}$ with respect to the sensor location errors, making use of the following lemma:

Lemma 4.2. *Let* $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{C}^{M \times K}$, $\boldsymbol{e} \in \mathbb{C}^M$, *and* $\boldsymbol{p} \in \mathbb{C}^K$. *Then*

$$\begin{aligned}
(\boldsymbol{A} \odot \boldsymbol{e}\boldsymbol{e}^T\boldsymbol{B})\boldsymbol{p} &= (\boldsymbol{A}\boldsymbol{P}\boldsymbol{B}^T\boldsymbol{e}) \otimes \boldsymbol{e}, \\
(\boldsymbol{e}\boldsymbol{e}^T\boldsymbol{B} \odot \boldsymbol{A})\boldsymbol{p} &= \boldsymbol{e} \otimes (\boldsymbol{A}\boldsymbol{P}\boldsymbol{B}^T\boldsymbol{e}),
\end{aligned}$$

where $\boldsymbol{P} = \operatorname{diag}(\boldsymbol{p})$.

Proof. For brevity, we show only the proof of the first equality. The proof of the second equality follows the same idea. By the definition of the Khatri-Rao product and the fact

that $\boldsymbol{a} \otimes \boldsymbol{b} = \mathrm{vec}(\boldsymbol{b}\boldsymbol{a}^T)$, the left hand side can be expressed as

$$\sum_i p_i(\boldsymbol{a}_i \otimes \boldsymbol{e}\boldsymbol{e}^T\boldsymbol{b}_i) = \sum_i p_i \,\mathrm{vec}(\boldsymbol{e}\boldsymbol{e}^T\boldsymbol{b}_i\boldsymbol{a}_i^T). \tag{4.21}$$

Because the Kronecker product follows the distributive rule, the right hand side is given by

$$\Big(\sum_i p_i\boldsymbol{a}_i\boldsymbol{b}_i^T\boldsymbol{e}\Big) \otimes \boldsymbol{e} = \sum_i p_i(\boldsymbol{a}_i\boldsymbol{b}_i^T\boldsymbol{e} \otimes \boldsymbol{e}) = \sum_i p_i \,\mathrm{vec}(\boldsymbol{e}\boldsymbol{e}^T\boldsymbol{b}_i\boldsymbol{a}_i^T), \tag{4.22}$$

which is equal to the left hand side. □

Because the partial derivative of Khatri-Rao products follows the Leibniz rule, we have

$$\begin{aligned}
\frac{\partial\tilde{\boldsymbol{r}}}{\partial u_i} &= \frac{\partial}{\partial u_i}[(\tilde{\boldsymbol{A}}^* \odot \tilde{\boldsymbol{A}})\boldsymbol{p} + \sigma^2 \,\mathrm{vec}(\boldsymbol{I}_M)] \\
&= \left(\frac{\partial\tilde{\boldsymbol{A}}^*}{\partial u_i} \odot \tilde{\boldsymbol{A}} + \tilde{\boldsymbol{A}}^* \odot \frac{\partial\tilde{\boldsymbol{A}}}{\partial u_i}\right)\boldsymbol{p} \\
&= \left\{[\boldsymbol{e}_M^{(i)}(\boldsymbol{e}_M^{(i)})^T\tilde{\boldsymbol{A}}_u^*] \odot \tilde{\boldsymbol{A}} + \tilde{\boldsymbol{A}}^* \odot [\boldsymbol{e}_M^{(i)}(\boldsymbol{e}_M^{(i)})^T\tilde{\boldsymbol{A}}_u]\right\}\boldsymbol{p}
\end{aligned} \tag{4.23}$$

By Lemma 4.2, we immediately obtain that

$$\frac{\partial\tilde{\boldsymbol{r}}}{\partial u_i} = (\tilde{\boldsymbol{A}}^*\boldsymbol{P}\tilde{\boldsymbol{A}}_u^T\boldsymbol{e}_M^{(i)}) \otimes \boldsymbol{e}_M^{(i)} + \boldsymbol{e}_M^{(i)} \otimes (\tilde{\boldsymbol{A}}\boldsymbol{P}\tilde{\boldsymbol{A}}_u^H\boldsymbol{e}_M^{(i)}). \tag{4.24}$$

Combining (4.24) with the definition of the Khatri-Rao product leads to (4.20c). The derivation of (4.20d) follows the same idea. □

If the FIM is nonsingular, the CRB for the DOAs can be readily obtained by inverting the FIM. However, this CRB does not always exist, due to the potential ambiguities introduced by sensor location errors. In the presence of sensor location errors, it is possible that certain

combinations of DOAs, $\boldsymbol{\theta}$, and sensor location errors, $\boldsymbol{\delta}$, lead to the same perturbed steering matrix and same observations. Consequently, it is impossible to distinguished between these combinations from the observations. For a perturbed steering matrix, we formally define the local ambiguity as follows:

Definition 4.1. *An perturbed steering matrix* $\boldsymbol{A}(\boldsymbol{\theta}, \boldsymbol{\delta})$ *is called* locally ambiguous *if for any* $(\boldsymbol{\theta}, \boldsymbol{\delta}) \in \boldsymbol{\Theta} \times \boldsymbol{\Delta}$*, there exists a non-empty neighborhood* $U \subset \boldsymbol{\Theta} \times \boldsymbol{\Delta}$*, such that for any* $(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\delta}}) \in U$*,* $\boldsymbol{A}(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\delta}}) = \boldsymbol{A}(\boldsymbol{\theta}, \boldsymbol{\delta})$*.*

In practice, the first sensor is usually chosen as the reference sensor, whose location is assumed known. However, this is not sufficient to eliminate the local ambiguity, because the perturbed steering matrix remains the same if we rotate the array by a small angle and shift all the DOAs by the same amount. Even if we restrict the perturbation along the x-axis only, the local ambiguity still exists because we can obtain the same steering matrix by expanding or shrinking the whole array along the x-axis by a small amount and adjusting the DOAs accordingly. When such local ambiguities exist, the set of unknown parameters will be locally unidentifiable, leading to a singular FIM [89]. In the following discussion, we assume that the FIM is nonsingular.

Unlike the CRB derived in [3, Ch. 8], our CRB utilizes the assumption that the sources are uncorrelated. Observing that $(\tilde{\boldsymbol{R}}^T \otimes \tilde{\boldsymbol{R}})^{-1}$ is always full rank in the noisy case, the FIM is non-singular if and only if $\boldsymbol{M}$ is full rank. Because $\boldsymbol{M}$ is a matrix of dimension $M^2 \times (2K + M_1 + M_2 + 1)$, the FIM (4.18) can remain nonsingular for up to $O(M^2)$ sources. Therefore our CRB can work in the underdetermined case when $K > M$, while the CRB in [3] cannot. Our derivation is also different from that in [87]. In [87], the FIM is evaluated partition by partition under the assumption that both the source powers and the noise power are known. In our derivation, the FIM is derived in a "factorized" form, which is more concise

than that in [87]. In addition, using our derivation, we conclude that the FIM can remain nonsingular for up to $O(M^2)$ sources. This conclusion is not easily seen from the derivation in [87].

Because the FIM (4.18) shares a form similar to the location error free FIM in [71], it is straightforward to show that the corresponding CRB depends on the SNRs instead of absolute values of p_k or σ^2. For sparse linear arrays, we are particularly interested in the underdetermined case when $K \geq M$. In [71], we have shown that the location error free CRB remains positive definite even if the SNR approaches infinity. This unusual behavior still exists in the presence of sensor location errors. If both $\tilde{\boldsymbol{A}}$ and $\boldsymbol{M}$ are full rank, $\tilde{\boldsymbol{R}}^T \otimes \tilde{\boldsymbol{R}}$ remains full rank as σ^2 approaches 0, and the resulting FIM remains positive definite. Hence the Schur complement corresponding to the DOAs is also positive definite, leading to a positive definite CRB matrix. This behavior puts a strictly positive lower bound on the MSE of all unbiased estimators when $K \geq M$.

4.2 The stochastic error model

One extension to the deterministic error model is the *stochastic error model*, where the sensor location errors are time-dependent. Such a model is applicable when the array is mounted on a non-stationary surface (e.g. [44, 90]), and the sensor location errors cannot be assumed constant during the N snapshots. By replacing $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{\delta}$ with their time-dependent counterparts, we can express the N snapshots received by the perturbed array as

$$\tilde{\boldsymbol{y}}(t) = \tilde{\boldsymbol{A}}(\boldsymbol{\theta}, \boldsymbol{\delta}(t))\boldsymbol{x}(t) + \boldsymbol{n}(t), t = 1, 2, \ldots, N. \tag{4.25}$$

in which $\boldsymbol{\delta}(t)$ follows some stochastic model. To avoid complications and obtain a general idea of the impact of stochastic sensor location errors, we make the following additional assumption:

A6 The sensor location errors $\boldsymbol{\delta}(t)$ are i.i.d. and are uncorrelated from both the source signals $\boldsymbol{s}(t)$ and the additive noise $\boldsymbol{n}(t)$.

Because $\tilde{\boldsymbol{A}}(\boldsymbol{\delta}(t))$ is nonlinear in the random variable $\boldsymbol{\delta}(t)$, $\tilde{\boldsymbol{y}}$ no longer follows the complex circularly-symmetric Gaussian distribution as in the deterministic error model. Consequently, it is rather difficult to derive the distribution of $\hat{\boldsymbol{R}}$ for the stochastic error model in the case of a finite number of snapshots. On the other hand, as implied by (4.7), the effect of sensor location errors dominates only when the number of snapshots is sufficiently large. Hence for the stochastic error model, we will analyze how the sensor location errors affect the estimation performance when an infinite number of snapshots is available.

Under assumption **A1**–**A5**, the perturbed covariance matrix can be evaluated as

$$\begin{aligned}
\tilde{\boldsymbol{R}} =& \mathbb{E}[\boldsymbol{y}(t)\boldsymbol{y}^H(t)] \\
=& \mathbb{E}[\tilde{\boldsymbol{A}}(\boldsymbol{\delta}(t))\boldsymbol{s}(t)\boldsymbol{s}^H(t)\tilde{\boldsymbol{A}}^H(\boldsymbol{\delta}(t))] + \mathbb{E}[\tilde{\boldsymbol{A}}(\boldsymbol{\delta}(t))\boldsymbol{s}(t)\boldsymbol{n}^H(t)] \\
&+ \mathbb{E}[\boldsymbol{n}(t)\boldsymbol{s}^H(t)\tilde{\boldsymbol{A}}^H(\boldsymbol{\delta}(t))] + \mathbb{E}[\boldsymbol{n}(t)\boldsymbol{n}^H(t)]. \\
=& \underbrace{\mathbb{E}[\tilde{\boldsymbol{A}}(\boldsymbol{\delta}(t))\boldsymbol{s}(t)\boldsymbol{s}^H(t)\tilde{\boldsymbol{A}}^H(\boldsymbol{\delta}(t))]}_{\boldsymbol{S}} + \sigma^2\boldsymbol{I},
\end{aligned}$$

where the cross terms vanish, because the sources and the additive noise have zero means and are uncorrelated. The first term $\boldsymbol{S}$ can be expressed as

$$\boldsymbol{S} = \sum_{i=1}^{K}\sum_{l=1}^{K} \mathbb{E}[\tilde{\boldsymbol{a}}(\theta_i, \boldsymbol{\delta}(t))s_i(t)s_l^*(t)\tilde{\boldsymbol{a}}^H(\theta_l, \boldsymbol{\delta}(t))],$$

whose (m,n)-th element is given by

$$S_{mn} = \sum_{i=1}^{K}\sum_{l=1}^{K} \mathbb{E}[\tilde{a}_m(\theta_i, \boldsymbol{\delta}(t))\tilde{a}_n^*(\theta_l, \boldsymbol{\delta}(t))s_i(t)s_l^*(t)]. \tag{4.26}$$

Using assumption **A6**, we can decouple the expectation evaluations with respect to $\boldsymbol{\delta}(t)$ and $\boldsymbol{s}(t)$. Noting that $\mathbb{E}[s_i(t)s_l^*(t)] = p_l$ only if $i = l$, and is otherwise 0, we need to consider only the terms where $i = l$. We can then rewrite (4.26) as

$$\begin{aligned} S_{mn} &= \sum_{k=1}^{K} p_i \mathbb{E}[\tilde{a}_m(\theta_k, \boldsymbol{\delta}(t))\tilde{a}_n^*(\theta_k, \boldsymbol{\delta}(t))] \\ &= \sum_{k=1}^{K} p_i a_m(\theta_k) a_n^*(\theta_k) \mathbb{E}\Big\{ e^{j(\boldsymbol{t}_{k,m} - \boldsymbol{t}_{k,n})^T \boldsymbol{\delta}} \Big\} \\ &= \sum_{k=1}^{K} p_i a_m(\theta_k) a_n^*(\theta_k) \phi_{\boldsymbol{\delta}}(\boldsymbol{t}_{k,m} - \boldsymbol{t}_{k,n}), \end{aligned} \tag{4.27}$$

where $\phi_{\boldsymbol{\delta}}(\boldsymbol{t})$ is the characteristic function of $\boldsymbol{\delta}(t)$, $\boldsymbol{t}_{k,n} = \frac{2\pi}{\lambda}\begin{bmatrix} \boldsymbol{e}_M^{(n)} \sin\theta_k \\ \boldsymbol{e}_M^{(n)} \cos\theta_k \end{bmatrix}$, and $\boldsymbol{e}_M^{(n)}$ is an M-dimensional vector with only the n-th element being one and other elements being zero. Let $\boldsymbol{\Phi}_k$ be an $M \times M$ matrix whose (m,n)-th element is given by $\phi_{\boldsymbol{\delta}}(\boldsymbol{t}_{k,m} - \boldsymbol{t}_{k,n})$. We can then express $\tilde{\boldsymbol{R}}$ as

$$\tilde{\boldsymbol{R}} = \sum_{k=1}^{K} p_k [\boldsymbol{a}(\theta_k)\boldsymbol{a}^H(\theta_k)] \circ \boldsymbol{\Phi}_k + \sigma^2 \boldsymbol{I}. \tag{4.28}$$

Here, the effect of the sensor location errors is encoded in matrices $\boldsymbol{\Phi}_k$. Because $\boldsymbol{t}_{k,m}$ depends on the k-th DOA, the effect of sensor location errors is generally DOA dependent and cannot be treated as colored Gaussian noise.

Vectorizing the (4.28) leads to

$$\tilde{\boldsymbol{r}} = [(\boldsymbol{A}^* \odot \boldsymbol{A}) \circ \boldsymbol{\Phi}]\boldsymbol{p} + \sigma^2 \operatorname{vec}(\boldsymbol{I}), \tag{4.29}$$

where $\boldsymbol{\Phi} = [\text{vec}(\boldsymbol{\Phi}_1)\ \text{vec}(\boldsymbol{\Phi}_2)\ \cdots\ \text{vec}(\boldsymbol{\Phi}_K)]$. Comparing (4.29) with (2.6), we observe that, under the stochastic error model, the coarray steering matrix $(\boldsymbol{A}^* \odot \boldsymbol{A})$ is modulated by $\boldsymbol{\Phi}$. Because characteristic functions usually do not evaluate to one outside the origin, $\boldsymbol{\Phi}$ will not be a matrix of ones and the corresponding difference coarray model will be perturbed.

To give a better idea of (4.28) and (4.29), we consider the case when $\boldsymbol{\delta}(t)$ follows a zero-mean Gaussian distribution with the covariance matrix denoted by $\boldsymbol{C}$. We partition $\boldsymbol{C}$ as $\left[\begin{smallmatrix} \boldsymbol{C}_{uu} & \boldsymbol{C}_{uv} \\ \boldsymbol{C}_{vu} & \boldsymbol{C}_{vv} \end{smallmatrix}\right]$, where $\boldsymbol{C}_{uu}$ and $\boldsymbol{C}_{vv}$ are the covariance of the location errors along the x-axis and y-axis, respectively, and $\boldsymbol{C}_{uv}$ denotes the corresponding cross covariance. The corresponding characteristic function of $\boldsymbol{\delta}(t)$ is then given by $\phi_{\boldsymbol{\delta}}(\boldsymbol{t}) = \exp(-1/2\boldsymbol{t}^T\boldsymbol{C}\boldsymbol{t})$. Substituting $\boldsymbol{t}_{k,n}$ into $\phi_{\boldsymbol{\delta}}(\boldsymbol{t})$ and expanding the terms in the exponent, we obtain that in the Gaussian case

$$\Phi_k(m,n) = \exp\left\{-\frac{2\pi^2}{\lambda^2}[\mu_1(m,n)\sin^2\theta_k + \mu_2(m,n)\cos^2\theta_k + 2\mu_3(m,n)\sin\theta_k\cos\theta_k]\right\}, \quad (4.30)$$

where

$$\begin{aligned}
\mu_1(m,n) &= C_{uu}(m,m) + C_{uu}(n,n) - 2C_{uu}(m,n), \\
\mu_2(m,n) &= C_{vv}(m,m) + C_{vv}(n,n) - 2C_{vv}(m,n), \\
\mu_3(m,n) &= C_{uv}(m,m) + C_{uv}(n,n) - C_{uv}(m,n) - C_{uv}(n,m).
\end{aligned}$$

We also observe that $\Phi_k(m,n)$ is still dependent on the k-th DOA. Hence for a general covariance matrix, the effect of the random sensor location errors is still DOA dependent. However, as shown in the following proposition, for certain covariance matrices, $\Phi_k(m,n)$ is independent of k.

Proposition 4.2. *Let $\boldsymbol{\delta}(t) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{C})$. Then $\boldsymbol{\Phi}_k$ $(k = 1, 2, \ldots, K)$ are independent of the DOAs if and only if $\mu_1(m,n) = \mu_2(m,n)$ and $\mu_3(m,n) = 0$ holds for every $m, n = 1, 2, \ldots, M$.*

Proof. Let $a, b, c \in \mathbb{C}$. Define $f(\theta) = a\sin^2\theta + b\cos^2\theta + c\sin\theta\cos\theta$. It suffices to show that $f(\theta)$ is a constant for all $\theta \in (-\pi/2, \pi/2)$ if and only if $a = b$ and $c = 0$.

The sufficiency is trivial and we need to show only necessity. Suppose $f(\theta) = d, \forall\theta \in (-\pi/2, \pi/2)$. Choose $\theta = \pi/4$ and we obtain $a + b + c = 2d$. Choose $\theta = -\pi/4$ and we obtain $a + b - c = 2d$. Therefore c must be 0. Choose $\theta = 0$ and we obtain $b = d$, which implies that $(a - b)\sin^2\theta = 0$ must hold for every $\theta \in (-\pi/2, \pi/2)$. Therefore we must have $a = b$. □

One special case that satisfies the conditions given in Proposition 4.2 is when $\boldsymbol{C} = \sigma_{\mathrm{p}}^2\boldsymbol{I}$, which leads to the following corollary.

Corollary 4.4. *Let* $\boldsymbol{\delta}_t \sim \mathcal{N}(\boldsymbol{0}, \sigma_{\mathrm{p}}^2\boldsymbol{I})$. *Then*

$$\tilde{\boldsymbol{R}} = C_1\Big\{\boldsymbol{A}\boldsymbol{P}\boldsymbol{A}^H + \frac{1}{C_1}\Big[\sigma^2 + (1 - C_1)\sum_{k=1}^{K} p_k\Big]\boldsymbol{I}\Big\}, \tag{4.31}$$

where $C_1 = \exp(-4\pi^2\sigma_{\mathrm{p}}^2/\lambda^2)$.

Proof. The expression (4.31) can be obtained by substituting $\boldsymbol{C} = \sigma_{\mathrm{p}}^2\boldsymbol{I}$ into (4.30) and simplifying the resulting $\tilde{\boldsymbol{R}}$ according to (4.28). □

We observe that if the sensor location perturbations are i.i.d. zero-mean Gaussian with the same variance, the effect of the sensor location errors can be indeed modeled as additive white noise as the number of snapshots goes to infinity. The signal subspace remains unchanged. However, the effective SNR is decreased because $0 < C_1 < 1$. In this special case, we can approximate the asymptotic MSE of SS-MUSIC for the k-th DOA with $\Re[\boldsymbol{\xi}_k^H(\boldsymbol{R} \otimes \boldsymbol{R}^T)\boldsymbol{\xi}_k]/(N\gamma_k^2 p_k^2)$, but with the original noise variance σ^2 replaced with the "effective noise

variance"

$$\frac{1}{C_1}\Big[\sigma^2 + (1 - C_1)\sum_{k=1}^{K} p_k\Big].$$

4.3 Numerical results

In this section, we use numerical simulations to demonstrate how sensor location errors affect the DOA estimation performance for sparse linear arrays. We consider both the deterministic error model and the stochastic error model. Unlike ULAs, sparse linear arrays sharing the same number of sensors can have different structures. For a comprehensive comparison, we consider two sets of sparse linear arrays throughout the simulations. The first set consists of four different sparse linear arrays sharing the same number of sensors:

- Co-prime (3,5): $[0, 3, 5, 6, 9, 10, 12, 15, 20, 25]d_0$;
- MRA 10[37]: $[0, 1, 4, 10, 16, 22, 28, 30, 33, 35]d_0$;
- Nested (4,6): $[0, 1, 2, 3, 4, 9, 14, 19, 24, 29]d_0$;
- Nested (5,5): $[0, 1, 2, 3, 4, 5, 11, 17, 23, 29]d_0$.

The second set consists of four different sparse linear arrays sharing the same aperture:

- Co-prime (2,3): $[0, 2, 3, 4, 6, 9]d_0$;
- MRA 5[37]: $[0, 1, 2, 6, 9]d_0$;
- Nested (1,5): $[0, 1, 3, 5, 7, 9]d_0$;
- Nested (4,2): $[0, 1, 2, 3, 4, 9]d_0$.

Throughout all experiments, we define the SNR as follows:

$$\mathrm{SNR} = 10 \log_{10} \frac{\min_{k=1,2,\ldots,K} p_k}{\sigma^2}.$$

Given the results from L trials, we compute the empirical MSE with

$$\mathrm{MSE}_{\mathrm{em}} = \frac{1}{KL} \sum_{l=1}^{L} \sum_{k=1}^{K} \left(\hat{\theta}_k^{(l)} - \theta_k^{(l)} \right)^2,$$

where $\theta_k^{(l)}$ is the k-th DOA in the l-th trial, and $\hat{\theta}_k^{(l)}$ is the estimate of $\theta_k^{(l)}$.

4.3.1 Numerical analysis of the deterministic error model

We begin by verifying our closed-form asymptotic MSE expression (4.15) for the deterministic error model via numerical simulations. We consider 11 sources, which is more than the number of sensors, uniformly distributed between $-\pi/3$ and $\pi/3$ with equal power. We set the SNR to 0 dB. We generate the sensor location errors from a zero-mean Gaussian distribution with covariance matrix $\sigma_{\mathrm{p}}^2 \boldsymbol{I}$. The magnitude of sensor location errors can then be tuned with σ_{p}^2. We consider the first set of sparse linear arrays. We compute the difference between the AAMSE given by (4.15) and the empirical MSE under different combinations of snapshot numbers and magnitudes of perturbations. The results are summarized in Fig. 4.2. It can be observed that the empirical results agree very well with our analytical results when the number of snapshots is above 200 and the perturbation level is below 0.05. When the number of snapshots is small, the asymptotic assumption no longer holds, and the discrepancy between our analytical results and the empirical results becomes evident. When the magnitude of the sensor location errors is large, the high order terms with respect to the

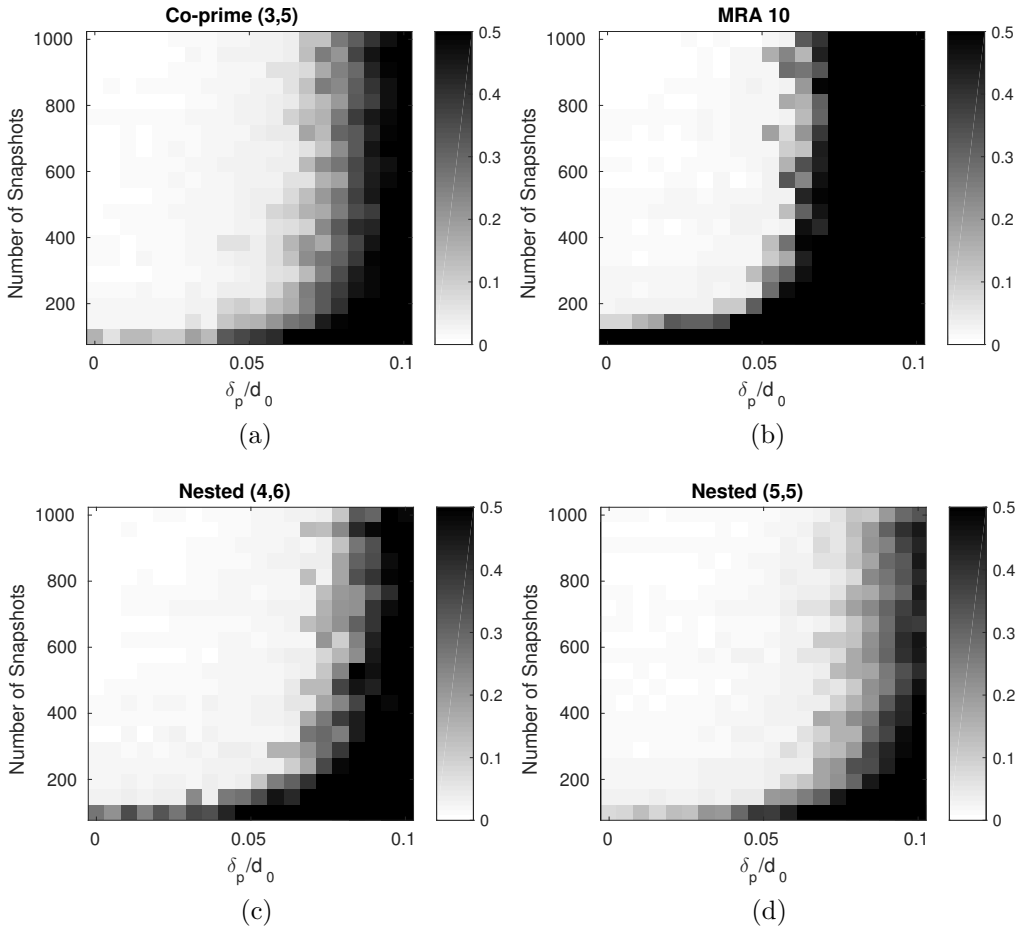


Figure 4.2: $|\mathrm{MSE}_{\mathrm{an}} - \mathrm{MSE}_{\mathrm{em}}|/\mathrm{MSE}_{\mathrm{em}}$ for different types of arrays under different numbers of snapshots and different magnitudes of perturbations. The results are averaged from 3000 trials.

sensor location errors are no longer negligible, leading to discrepancies between our analytical results and the empirical results.

We next demonstrate how the DOA estimation errors vary with respect to sensor location errors for different types of sparse linear arrays. The results are plotted in Fig. 4.3 and Fig. 4.4. In Fig. 4.3, we plot the RMSE vs. σ_{p}/d_0 for four different sparse linear arrays with the same number of sensors. We observe that the MRA achieves the lowest RMSE, the co-prime array achieves the highest RMSE, and the two nested arrays sit in the middle.

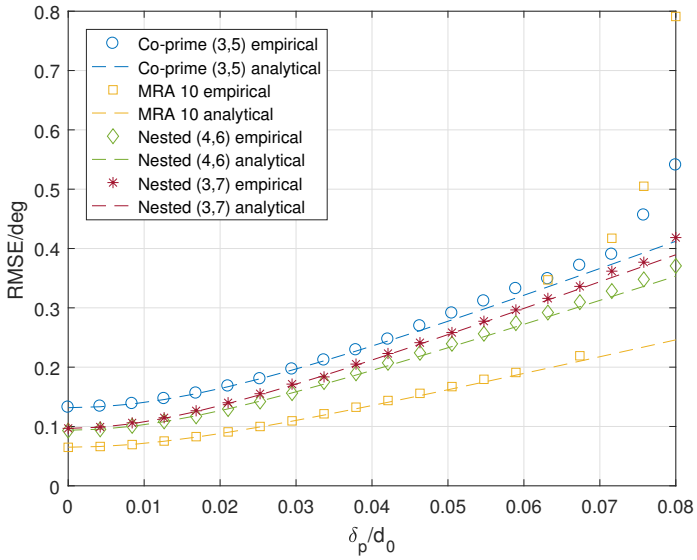


Figure 4.3: RMSE vs. perturbation level for four different sparse linear arrays with the same number of sensors. The empirical results are averaged from 1000 trials.

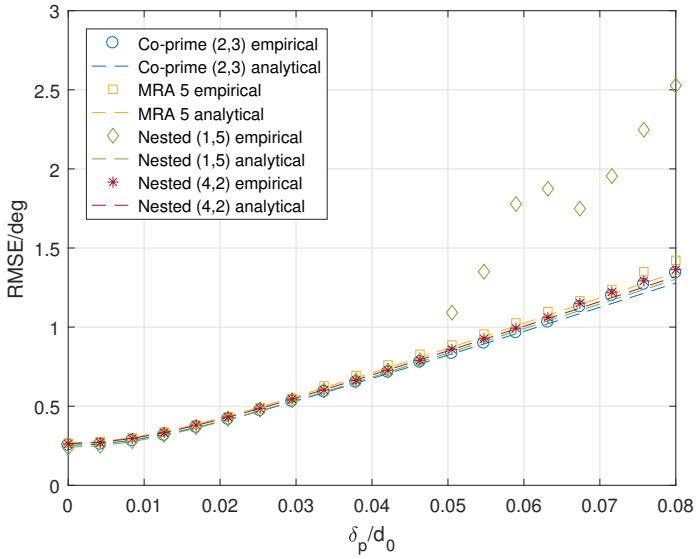


Figure 4.4: RMSE vs. perturbation level for four different sparse linear arrays with the same aperture. The empirical results are averaged from 1000 trials.

This observation reflects the fact that the MRA has the largest aperture among the four arrays, while the co-prime array has the smallest. In Fig. 4.4, we plot the RMSE vs. $\sigma_{\rm p}/d_0$

for four different sparse linear arrays with the same aperture. We observe that while all four arrays show similar performance, MRA 5 is least sensitive to sensor location errors. Another interesting observation is that, Nested (1,5), Nested (4,2), and MRA 5, despite sharing the same central ULA part in their difference coarrays, show different sensitivities with respect the sensor location errors. This observation agrees with our analysis of (4.15).

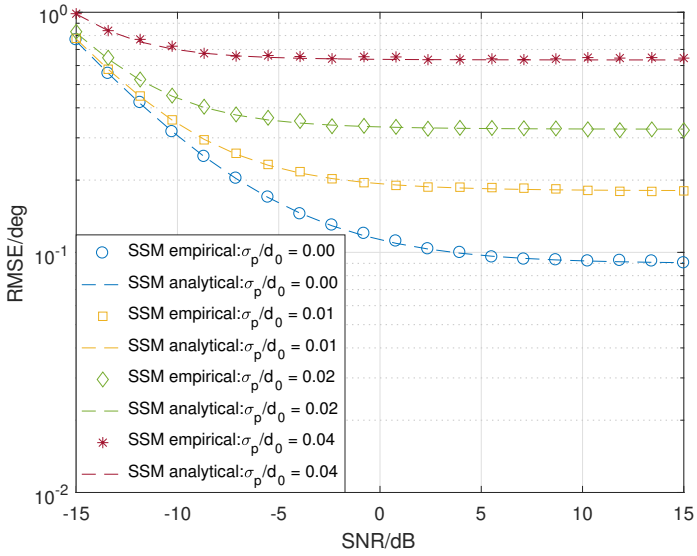


Figure 4.5: RMSE vs. SNR for Co-prime (2,3) under different perturbation levels. The empirical results are averaged from 1000 trials.

Finally, we show how the variance of sensor location errors, σ_{p}, affects the MSE of SS-MUSIC in high SNR regions. We consider 6 sources evenly placed between $-\pi/3$ and $\pi/3$, and fix the number of snapshots to 5000. Fig. 4.5 plots the results for Co-prime (3,5). We observe that the empirical MSEs well agree with our theoretical results. In the absence of sensor location errors, the MSE of SS-MUSIC converges to a positive constant as the SNR approaches infinity, which agrees with our analysis of Corollary 4.3. As the variance of sensor location errors increase, this positive constant also increases, because the bias resulting from sensor location errors grows larger. Additionally, we observe that the gap between the MSE

values when the sensor location errors are present and when they are not present does not decrease as the SNR increases. This observation confirms our analysis of (4.15) that the bias cannot be mitigated by increasing only the SNR.

4.3.2 Numerical analysis of the stochastic error model

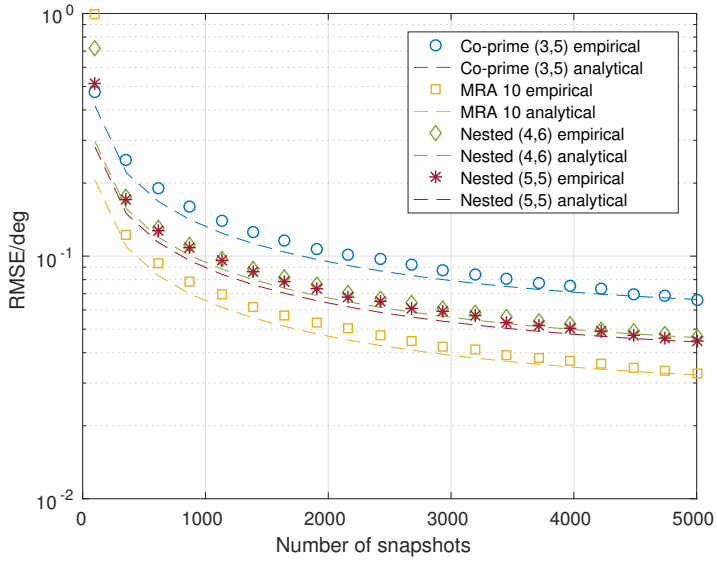


Figure 4.6: Empirical RMSEs vs. analytical approximations under different numbers of snapshots for four different sparse linear arrays with the same number of sensors, based on the stochastic error model. The empirical results are averaged from 5000 trials.

In this subsection, we verify our derivations in Section 4.2 via numerical simulations. For the first set of sparse linear arrays, we consider 11 sources evenly distributed between $-\pi/3$ and $\pi/3$. For the second set of sparse linear arrays, we consider 6 sources evenly distributed between $-\pi/3$ and $\pi/3$. For both sets of sparse linear arrays, the number of sources is chosen to be larger than or equal to the number of sensors. We sample the sensor location errors $\boldsymbol{\delta}(t)$ from a zero-mean Gaussian distribution with covariance matrix $\sigma_{\mathrm{p}}^2\boldsymbol{I}$, and the standard deviation of sensor location errors, σ_{p}, is fixed to $0.1d_0$. Because the sensor location errors are i.i.d. zero-mean Gaussian, we approximate the analytical MSE by evaluating the

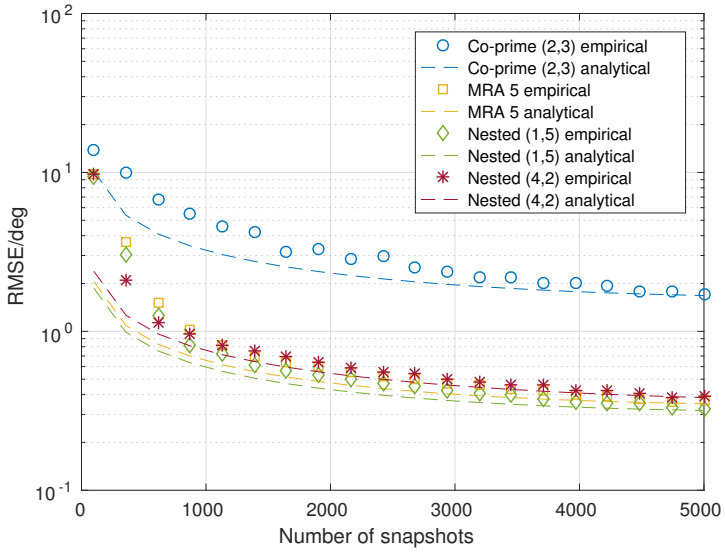


Figure 4.7: Empirical RMSEs vs. analytical approximations under different numbers of snapshots for four different sparse linear arrays with the same aperture, based on the stochastic error model. The empirical results are averaged from 5000 trials.

location error free asymptotic MSE of SS-MUSIC [71] with the noise power replaced with the "effective noise power" given by Corollary 4.4. We fix the SNR to 0 dB and vary the number of snapshots.

The results are plotted in Fig. 4.6 and Fig. 4.7. We observe that, when the number of snapshots is small, the empirical MSE deviates from the analytical MSE. As the number of snapshots increases, the empirical MSE approaches the our analytical approximation. This is because our analytical approximation is based on the assumption of infinite number of snapshots. In Fig. 4.6, we observe that the MRA, which has the largest aperture, achieves the lowest MSE. The co-prime array, which has the smallest aperture, has higher MSE than the MRA and two nested arrays. In Fig. 4.7, we observe that the MSE of the co-prime array is significantly higher than the other three arrays. This is because the co-prime array is the only array among the four arrays whose difference coarray is not a full ULA. Consequently,

the central ULA part of the co-prime array is the smallest among the four, resulting a significantly higher MSE.

4.3.3 Numerical results of the CRB

We close this section with numerical results of the CRB we derived in Section 4.1.2. We demonstrate that the CRB obtained from Proposition 4.1 is indeed achievable in cases when the number of sources is greater than the number of sensors. We consider 11 sources evenly distributed between $-\pi/3$ and $\pi/3$ and fix the number of snapshots to 5000. We consider the first set of sparse linear arrays with the same number of sensors. We compare the CRB and the empirical MSE obtained by solving the following stochastic maximum likelihood problem using the optimization toolbox in MATLAB:

$$\min_{\boldsymbol{\theta},\boldsymbol{p},\sigma^2,\boldsymbol{\delta}} \log\det(\tilde{\boldsymbol{R}}(\boldsymbol{\theta},\boldsymbol{p},\sigma^2,\boldsymbol{\delta})) + \mathrm{tr}(\tilde{\boldsymbol{R}}^{-1}(\boldsymbol{\theta},\boldsymbol{p},\sigma^2,\boldsymbol{\delta})\hat{\boldsymbol{R}}),$$

where $\tilde{\boldsymbol{R}}(\boldsymbol{\theta},\boldsymbol{p},\sigma^2,\boldsymbol{\delta})$ follows the definition in (4.2).

The results are plotted in Fig. 4.8. For comparison, we also include the CRB without considering sensor location errors[71,72]. We first notice that the CRB converges to a positive constant as SNR increases, which agrees with our analysis of the CRB in the underdetermined case in Section 4.1.2. We then observe that, given sufficient SNR, the MSE of the MLE indeed achieves the CRB for all four arrays. Additionally, there is a significant gap between the values of the CRB when the sensor location errors are considered and when they are not. This gap shows that unknown sensor location errors have a drastic impact on the achievable DOA estimation performance of sparse linear arrays.

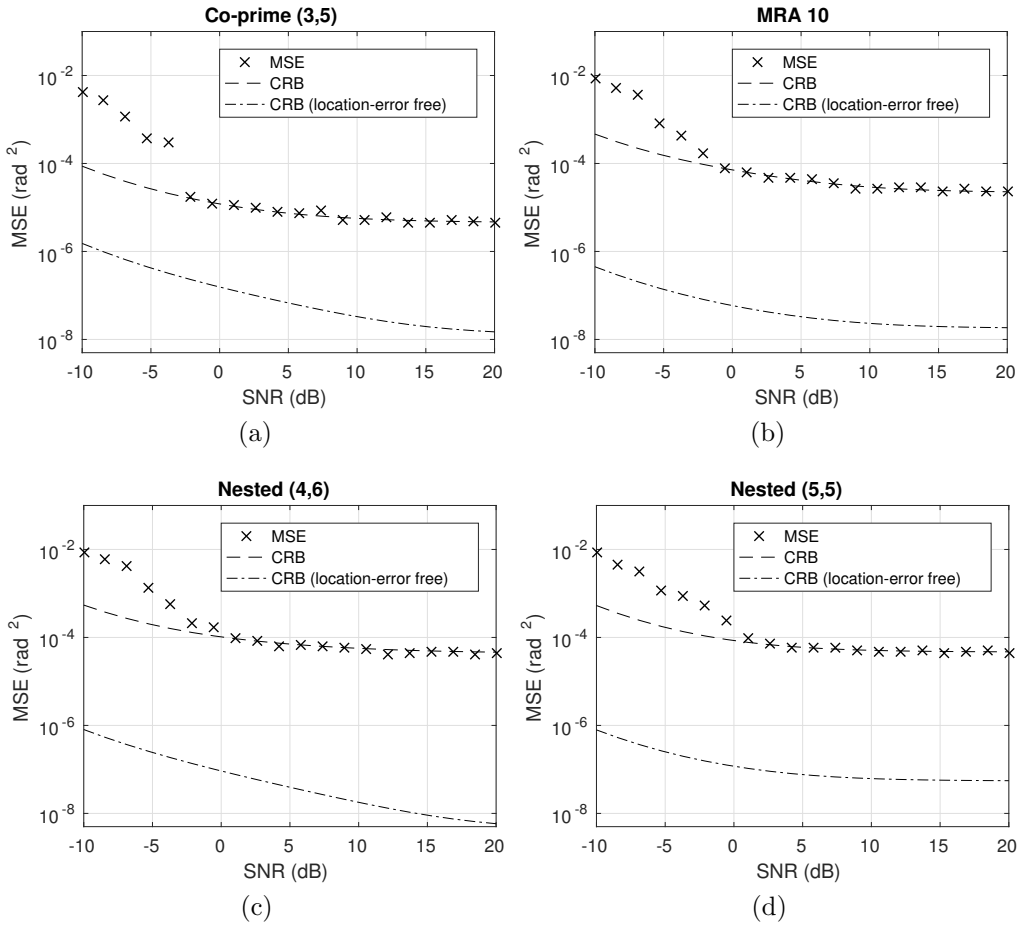


Figure 4.8: CRB versus the empirical MSE of the maximum likelihood estimator for four different sparse linear arrays under different SNRs. The empirical MSEs are averaged from 500 trials.

4.4 Chapter summary

In this chapter, we analyzed the performance of sparse linear arrays in the presence of sensor location errors. We derived a closed-form asymptotic MSE expression for SS-MUSIC in the presence of small sensor location errors. Under the deterministic error model, the sensor location errors introduce a constant bias that cannot be mitigated by only increasing the SNR. To extend our analysis, we introduced the stochastic error model and discussed the

Gaussian case. Our results will benefit future research on the development of robust DOA estimators using sparse linear arrays and the optimal design of sparse linear arrays.

Chapter 5

Robust Direction Finding in Cases of Missing Data

In the previous chapters, we focused on statistically analyzing the performance of sparse linear arrays. In this chapter, we introduce robust DOA estimators that handle missing data for sparse linear arrays[7]. Missing data problems arise when one or more sensors malfunction and fail to provide correct data during the measurement period. Because sparse linear arrays depend on their difference coarray model to resolve more sources than the number of sensors, they are more susceptible to sensor failures. If the measurements from one or more sensors are missing, the coarray structure will be partially destroyed, leading to performance degradation and loss of degrees of freedom.

Many previous works have addressed the problem of direction finding in cases of missing data. In [91], Larsson et al. proposed a Cholesky parameterization based maximum likelihood estimator and analyzed its asymptotic performance. However, their model is based on ULAs, and requires a sequential failure pattern. In practice, any sensor may fail, so

[7]This chapter is based on M. Wang, Z. Zhang, and A. Nehorai, "Direction finding using sparse linear arrays with missing data," *Proc. 42nd IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, New Orleans, LA, Mar. 5–9, 2017.

the sequential assumption may not be true. Recent advances in matrix completion [92, 93] and atomic norm minimization [34, 94] also bring new methods to tackle the missing data problem. By exploiting the low-rank property of the signal subspace, it is possible to extrapolate the missing data via semidefinite programming (SDP). However, when the number of measurements is large, the resulting SDP will be computationally expensive to solve.

In this chapter, we consider the direction finding problem with general sparse linear arrays and incomplete measurements, but without assuming a sequential failure pattern. Base on the maximum-likelihood approach, we focus on deriving algorithms that utilizes the information in both complete measurements and incomplete measurements. We first estimate the augmented covariance matrix of the difference coarray model by exploiting its Toeplitz structure, and then apply the MUSIC algorithm [9] to obtain the DOA estimates. We derive the CRB and confirm the efficacy of our algorithms via numerical examples.

5.1 Measurement model

Recall that in Section 2.2, we knew that an M-sensor sparse linear array could be viewed as a thinned ULA of $M_0 = \bar{d}_M + 1$ sensors. For example, a co-prime array whose sensors are located at $[0, 2, 3, 4, 6, 9]d_0$ can be viewed as a 10-sensor ULA with the 2nd, 6th, 8th, and 9th sensors removed. Therefore, we can rewrite (2.3) as

$$\boldsymbol{y}(t) = \boldsymbol{S}\boldsymbol{A}_{\mathrm{ULA}}(\boldsymbol{\theta})\boldsymbol{x}(t) + \boldsymbol{n}(t),\ t = 1, 2, \ldots, N, \tag{5.1}$$

where $\boldsymbol{A}_{\text{ULA}}(\boldsymbol{\theta}) = [\boldsymbol{a}_{\text{ULA}}(\theta_1), \boldsymbol{a}_{\text{ULA}}(\theta_2), \ldots, \boldsymbol{a}_{\text{ULA}}(\theta_K)]$ is the steering matrix of a M_0-sensor ULA [52]. $\boldsymbol{S}$ is a $M \times M_0$ selection matrix, where S_{mn} is one if and only if the m-th sensor in the sparse linear array corresponds to the n-th sensor in the ULA, and otherwise zero.

We can then rewrite the covariance matrix $\boldsymbol{R}$ as

$$\boldsymbol{R} = \boldsymbol{S}\boldsymbol{R}_{\text{ULA}}\boldsymbol{S}^T, \tag{5.2}$$

where $\boldsymbol{R}_{\text{ULA}} = \boldsymbol{A}_{\text{ULA}}\boldsymbol{P}\boldsymbol{A}_{\text{ULA}}^H + \sigma^2\boldsymbol{I}$. Therefore the covariance matrix of a sparse linear array is a compressed version of the covariance matrix of a ULA.

By vectorizing $\boldsymbol{R}$, we obtain

$$\boldsymbol{r} = (\boldsymbol{S} \otimes \boldsymbol{S})(\boldsymbol{A}_{\text{ULA}}^* \odot \boldsymbol{A}_{\text{ULA}})\boldsymbol{p} + \sigma^2\boldsymbol{i}, \tag{5.3}$$

where $\boldsymbol{r} = \text{vec}(\boldsymbol{R})$, $\boldsymbol{p} = [p_1, p_2, \cdots, p_K]^T$, and $\boldsymbol{i} = \text{vec}(\boldsymbol{I})$. As discussed in Section 2.2.2, model (5.3) resembles a measurement model with deterministic sources and noise, and $(\boldsymbol{S} \otimes \boldsymbol{S})(\boldsymbol{A}_{\text{ULA}}^* \odot \boldsymbol{A}_{\text{ULA}})$ embeds a steering matrix of a virtual array with enhanced degrees of freedom, whose sensor locations are given by $\bar{\mathcal{D}}_{\text{co}} = \{(\bar{d}_m - \bar{d}_n) | \bar{d}_m, \bar{d}_n \in \bar{\mathcal{D}}\}$. If $\bar{\mathcal{D}}_{\text{co}}$ consists of consecutive integers from $-M_0 + 1$ to $M_0 - 1$, we call the sparse linear array *complete*. If a sparse linear array is complete (e.g., minimum redundancy arrays and nested arrays), it is possible to estimate the elements in $\boldsymbol{R}_{\text{ULA}}$ using rank enhanced spatial smoothing [38] or more sophisticated methods [54]. We are then able to identify more sources than the number of sensors through $\boldsymbol{R}_{\text{ULA}}$. On the other hand, if a sparse linear array is incomplete (e.g., co-prime arrays), we define $\tilde{M}_0$ as the largest M such that $\{-M+1, \ldots, 0, \ldots, M-1\} \subset \bar{\mathcal{D}}_{\text{co}}$. In this case, we can recover only a $\tilde{M}_0 \times \tilde{M}_0$ submatrix of $\boldsymbol{R}_{\text{ULA}}$ using similar methods. If $\tilde{M}_0 > M$, we again are able to identify more sources than the number of sensors.

We now consider the signal model with missing data. Without loss of generality, we consider L sampling periods. During the first period, we assume all the sensors are functioning normally. This assumption is reasonable because if some sensors fail from the beginning, we can simply remove them and form a new sparse linear array whose sensors are all functional during the first period. During the l-th ($2 \leq l \leq L$) period, some sensors fail and the measurement data from these sensors are missing. Let M_l be the number of working sensors during the l-th period. Let $\boldsymbol{T}_l$ be a selection matrix of size $M_l \times M$ such that the (i,j)-th element $\boldsymbol{T}_l$ is one if and only if the j-th sensor in the sparse linear array is the i-th working sensor during the l-th period, and otherwise zero. For notational simplicity, we define $\boldsymbol{T}_1 = \boldsymbol{I}_M$. After discarding the measurements from the malfunctioning sensors, the snapshots taken during the l-th period are given by

$$\boldsymbol{y}_l(t) = \boldsymbol{T}_l[\boldsymbol{S}\boldsymbol{A}_{\mathrm{ULA}}(\boldsymbol{\theta})\boldsymbol{x}(t) + \boldsymbol{n}(t)], \tag{5.4}$$

for $t = N_1 + \cdots + N_{l-1} + 1, \ldots, N_1 + \cdots + N_{l-1} + N_l$, where N_l is the number of snapshots collected during the l-th period. The total number of snapshots is denoted by $N = \sum_{l=1}^{L} N_l$. Correspondingly, we can collect L different sample covariance matrices $\hat{\boldsymbol{R}}_l = 1/N_l \sum_{t=N_1+\cdots+N_{l-1}+1}^{N_1+\cdots+N_{l-1}+N_l} \boldsymbol{y}_l(t)\boldsymbol{y}_l^H(t)$, $l = 1, 2, \ldots, L$. We also define their expectations as

$$\boldsymbol{R}_l = \mathbb{E}[\hat{\boldsymbol{R}}_l] = \boldsymbol{T}_l\boldsymbol{S}\boldsymbol{R}_{\mathrm{ULA}}\boldsymbol{S}^T\boldsymbol{T}_l^T + \sigma^2\boldsymbol{I}, \tag{5.5}$$

whose vectorized versions are given by

$$\boldsymbol{r}_l = \mathrm{vec}(\boldsymbol{R}_l) = (\boldsymbol{T}_l\boldsymbol{S} \otimes \boldsymbol{T}_l\boldsymbol{S})(\boldsymbol{A}_{\mathrm{ULA}}^* \odot \boldsymbol{A}_{\mathrm{ULA}})\boldsymbol{p} + \sigma^2\boldsymbol{i}. \tag{5.6}$$

Because of the missing data, $(\boldsymbol{T}_l\boldsymbol{S} \otimes \boldsymbol{T}_l\boldsymbol{S})(\boldsymbol{A}^*_{\text{ULA}} \odot \boldsymbol{A}_{\text{ULA}})$ no longer embeds a desired virtual array steering matrix and existing methods cannot be directly applied. If we use only $\hat{\boldsymbol{R}}_1$ for estimation, we lose much information provided in $\hat{\boldsymbol{R}}_l$ $(2 \leq l \leq L)$. Therefore an estimator that utilizes all the information in $\hat{\boldsymbol{R}}_l$ $(1 \leq l \leq L)$ is desired.

5.2 Estimation in the presence of missing data

In this section, by exploiting the Toeplitz structure of $\boldsymbol{R}_{\text{ULA}}$, we introduce three DOA estimators that utilize all the information from $\hat{\boldsymbol{R}}_1, \hat{\boldsymbol{R}}_2, \ldots, \hat{\boldsymbol{R}}_L$.

5.2.1 Ad-hoc estimator

The ad-hoc estimator for our signal model is inspired by redundancy averaging [57, 63], and is an extension of the ad-hoc estimator in [91]. Let $\mathcal{V}_k = \{(m,n) | \bar{d}_m - \bar{d}_n = k, \bar{d}_m, \bar{d}_n \in \bar{\mathcal{D}}\}$. Let $\mathcal{A}_{m,n}$ denote the set of snapshot indices when both the m-th and the n-th sensor are working. We define

$$u_k = \frac{\sum_{(m,n)\in\mathcal{V}_k} \sum_{t\in\mathcal{A}_{m,n}} y_m(t) y_n^*(t)}{\sum_{(m,n)\in\mathcal{V}_k} |\mathcal{A}_{m,n}|}, \tag{5.7}$$

where $\boldsymbol{y}(t) = [y_1(t), \cdots, y_M(t)]$ is the full measurement vector before discarding invalid data, $y_m(t)$ is the output of the m-th sensor, and $|\mathcal{A}|$ denotes the cardinality of $\mathcal{A}$. For complete arrays, we can obtain u_k for $k = -M_0+1, -M_0+2, \ldots, M_0-1$, and the ad-hoc estimate of

$\boldsymbol{R}_{\mathrm{ULA}}$ is given by

$$\hat{\boldsymbol{R}}_{\mathrm{U}}^{(\mathrm{ad-hoc})} = \begin{bmatrix} u_0 & u_{-1} & \cdots & u_{-M_0+1} \\ u_1 & u_0 & \cdots & u_{-M_0+2} \\ \vdots & \vdots & \ddots & \vdots \\ u_{M_0} & u_{M_0-1} & \cdots & u_0 \end{bmatrix}. \tag{5.8}$$

We can then apply MUSIC or other DOA estimation methods to $\hat{\boldsymbol{R}}_{\mathrm{U}}^{(\mathrm{ad-hoc})}$ to obtain the DOA estimates.

For incomplete arrays, we can use a similar construction to obtain a $\tilde{M}_0 \times \tilde{M}_0$ matrix from u_k, $k = -\tilde{M}_0 + 1, \tilde{M}_0 + 2, \ldots, \tilde{M}_0 - 1$, which is the estimate of a submatrix of $\boldsymbol{R}_{\mathrm{ULA}}$.

It should be noted that while (5.7) and (5.8) are computationally efficient to evaluate, the resulting $\hat{\boldsymbol{R}}_{\mathrm{U}}^{(\mathrm{ad-hoc})}$ is not guaranteed to be positive definite, which may be undesired in some applications.

5.2.2 Maximum-likelihood based estimators

Based on the results in [3], the negative log-likelihood function of our model is given by

$$L(\boldsymbol{R}_1, \ldots, \boldsymbol{R}_L) = \sum_{l=1}^{L} N_l[\log |\boldsymbol{R}_l| + \mathrm{tr}(\boldsymbol{R}_l^{-1} \hat{\boldsymbol{R}}_l)], \tag{5.9}$$

where we have omitted the constants.

Observe that $\boldsymbol{R}_{\mathrm{ULA}}$ is Hermitian Toeplitz. It is possible to reparameterize $\boldsymbol{R}_{\mathrm{ULA}}$ by exploiting the Toeplitz structure, and the estimation of $\boldsymbol{R}_{\mathrm{ULA}}$ becomes a structure covariance estimation problem. In the following discussion, we consider only complete arrays. Extension to non-restricted arrays will be discussed in the remarks.

Following the idea in [95], we construct the structured matrices as follows. Let $\boldsymbol{I}_M^{(i)}$ denotes the $M \times M$ matrix whose elements are zero except for the i-th upper diagonal (i.e., $\boldsymbol{I}_M^{(i)}(m,n) = \delta(n-m-i)$, where $\delta(n)$ is the Kronecker delta). For a given positive integer M, we define the matrices $\{\boldsymbol{Q}_M^{(i)}\}_{i=1}^{2M-1}$ as

$$\boldsymbol{Q}_M^{(i)} = \begin{cases} \boldsymbol{I}_M, & i = 1, \\ \boldsymbol{I}_M^{(i-1)} + (\boldsymbol{I}_M^{(i-1)})^T, & 2 \le i \le M, \\ -j\boldsymbol{I}_M^{(i-M)} + j(\boldsymbol{I}_M^{(i-M)})^T, & M+1 \le i \le 2M-1. \end{cases} \tag{5.10}$$

Then we are able to express $\boldsymbol{R}_{\text{ULA}}$ as

$$\boldsymbol{R}_{\text{ULA}} = \sum_{i=1}^{2M_0-1} c_i \boldsymbol{Q}_{M_0}^{(i)}, \tag{5.11}$$

where $\boldsymbol{c} = [c_1, c_2, \cdots, c_{2M_0-1}]^T \in \mathbb{R}^{2M_0-1}$ is the Hermitian Toeplitz parameterization of $\boldsymbol{R}_{\text{ULA}}$. After obtaining its estimate, we can reconstruct $\boldsymbol{R}_{\text{ULA}}$ from (5.11) and then perform DOA estimation. Substituting (5.11) into (5.9) and taking the derivative with respect to c_i, we obtain

$$\frac{\partial L(\boldsymbol{c})}{\partial c_i} = \sum_{l=1}^{L} N_l \operatorname{tr}\left[\boldsymbol{T}_l \boldsymbol{S} \boldsymbol{Q}_{M_0}^{(i)} \boldsymbol{S}^T \boldsymbol{T}_l^T \boldsymbol{R}_l^{-1} (\boldsymbol{R}_l - \hat{\boldsymbol{R}}_l) \boldsymbol{R}_l^{-1}\right]$$

for $i = 1, 2, \ldots, 2M_0 - 1$. Because $\operatorname{vec}(\boldsymbol{AXB}) = (\boldsymbol{B}^T \otimes \boldsymbol{A}) \operatorname{vec}(\boldsymbol{X})$, and because $(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}$ for nonsingular $\boldsymbol{A}$, $\boldsymbol{B}$ [70], we have

$$\operatorname{vec}(\boldsymbol{T}_l \boldsymbol{S} \boldsymbol{Q}_{M_0}^{(i)} \boldsymbol{S}^T \boldsymbol{T}_l^T) = \boldsymbol{\Phi}_l \boldsymbol{q}_{M_0}^{(i)}, \tag{5.12}$$

where $\boldsymbol{q}_{M_0}^{(i)} = \mathrm{vec}(\boldsymbol{Q}_{M_0}^{(i)})$, and $\boldsymbol{\Phi}_l = \boldsymbol{T}_l\boldsymbol{S} \otimes \boldsymbol{T}_l\boldsymbol{S}$. We also have

$$\mathrm{vec}(\boldsymbol{R}_l^{-1}(\boldsymbol{R}_l - \hat{\boldsymbol{R}}_l)\boldsymbol{R}_l^{-1}) = \boldsymbol{W}_l^{-1}(\boldsymbol{\Phi}_l\boldsymbol{Q}_{M_0}\boldsymbol{c} - \hat{\boldsymbol{r}}_l), \tag{5.13}$$

where $\boldsymbol{W}_l = \boldsymbol{R}_l^T \otimes \boldsymbol{R}_l$, $\boldsymbol{Q}_{M_0} = [\boldsymbol{q}_{M_0}^{(1)}, \boldsymbol{q}_{M_0}^{(2)}, \cdots, \boldsymbol{q}_{M_0}^{(2M_0-1)}]$, and $\hat{\boldsymbol{r}}_l = \mathrm{vec}(\hat{\boldsymbol{R}}_l)$. Let all the partial derivatives with respect to c_i be zero. Then, we utilize (5.12) and (5.13) to obtain

$$\Big(\sum_{l=1}^{L} N_l\boldsymbol{G}_l\Big)\boldsymbol{c} = \sum_{l=1}^{L} N_l\boldsymbol{h}_l \tag{5.14}$$

where $\boldsymbol{G}_l = \boldsymbol{Q}_{M_0}^T\boldsymbol{\Phi}_l^T\boldsymbol{W}_l^{-1}\boldsymbol{\Phi}_l\boldsymbol{Q}_{M_0}$, and $\boldsymbol{h}_l = \boldsymbol{Q}_{M_0}^T\boldsymbol{\Phi}_l^T\boldsymbol{W}_l^{-1}\hat{\boldsymbol{r}}_l$. Note that if we have sufficient snapshots in each period, $\hat{\boldsymbol{R}}_l$ will be very close to $\boldsymbol{R}_l$, and we can replace $\boldsymbol{W}_l$ with its estimate $\hat{\boldsymbol{W}}_l = \hat{\boldsymbol{R}}_l^T \otimes \hat{\boldsymbol{R}}_l$. In this case the only unknown in (5.14) will be $\boldsymbol{c}$, whose estimate can be readily given by

$$\hat{\boldsymbol{c}}_{\mathrm{WLS}} = \Big[\sum_{l=1}^{L} N_l\hat{\boldsymbol{G}}_l\Big]^{-1}\Big[\sum_{l=1}^{L} N_l\hat{\boldsymbol{h}}_l\Big]. \tag{5.15}$$

where $\hat{\boldsymbol{G}}_l$ denotes $\boldsymbol{G}_l$ with $\boldsymbol{W}_l$ replaced by $\hat{\boldsymbol{W}}_l$, and $\hat{\boldsymbol{h}}_l$ denotes $\boldsymbol{h}_l$ with $\boldsymbol{W}_l$ replaced by $\hat{\boldsymbol{W}}_l$. Lemma 5.2 ensures that (5.15) produces real results.

Lemma 5.1. *Let $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ be Hermitian symmetric. Then* $\mathrm{tr}(\boldsymbol{ABAC})$ *is real.*

Proof. This can be shown by the fact that

$$\mathrm{tr}(\boldsymbol{ABAC})^* = \mathrm{tr}[(\boldsymbol{ABAC})^H] = \mathrm{tr}(\boldsymbol{CABA}) = \mathrm{tr}(\boldsymbol{ABAC}).$$

□

Lemma 5.2. *Both $\hat{\boldsymbol{G}}_l$ and $\hat{\boldsymbol{h}}_l$ are real.*

Proof. Through algebraic manipulations, the (m,n)-th element of $\hat{\boldsymbol{G}}_l$ can be rewritten as

$$\mathrm{tr}[\hat{\boldsymbol{R}}_l^{-1}\boldsymbol{T}_l\boldsymbol{S}\boldsymbol{Q}_{M_0}^{(m)}\boldsymbol{S}^T\boldsymbol{T}_l^T\hat{\boldsymbol{R}}_l^{-1}\boldsymbol{T}_l\boldsymbol{S}\boldsymbol{Q}_{M_0}^{(n)}\boldsymbol{S}^T\boldsymbol{T}_l^T].$$

By the definition of $\boldsymbol{Q}_{M_0}^{(m)}$ in (5.10), we know that $\boldsymbol{T}_l\boldsymbol{S}\boldsymbol{Q}_{M_0}^{(m)}\boldsymbol{S}^T\boldsymbol{T}_l^T$ is Hermitian symmetric. Because $\hat{\boldsymbol{R}}^{-1}$ is also Hermitian symmetric, we know that each element of $\hat{\boldsymbol{G}}_l$ is real by Lemma 5.2. The proof for the second claim follows the same idea. □

We call (5.15) the "weighted least squares" (WLS) estimate, because (5.15) is the solution to the weighted least squares problem: $\min_{\boldsymbol{c}}\sum_{l=1}^{L}N_l\|\boldsymbol{\Phi}_l\boldsymbol{Q}_{M_0}\boldsymbol{c}-\hat{\boldsymbol{r}}_l\|^2_{\hat{\boldsymbol{W}}_l^{-1}}$, where $\|\boldsymbol{x}\|_{\boldsymbol{W}}=\sqrt{\boldsymbol{x}^H\boldsymbol{W}\boldsymbol{x}}$.

We can also observe that (5.14) leads to the following fixed-point type iteration:

$$\hat{\boldsymbol{c}}_{\mathrm{FP}}^{(k)}=\Big[\sum_{l=1}^{L}N_l\boldsymbol{G}_l\big(\hat{\boldsymbol{c}}_{\mathrm{FP}}^{(k-1)}\big)\Big]^{-1}\Big[\sum_{l=1}^{L}N_l\boldsymbol{h}_l\big(\hat{\boldsymbol{c}}_{\mathrm{FP}}^{(k-1)}\big)\Big], \tag{5.16}$$

where $\boldsymbol{G}_l\big(\hat{\boldsymbol{c}}_{\mathrm{FP}}^{(k-1)}\big)$ and $\boldsymbol{h}_l\big(\hat{\boldsymbol{c}}_{\mathrm{FP}}^{(k-1)}\big)$ are constructed from $\hat{\boldsymbol{c}}_{\mathrm{FP}}^{(k-1)}$.

Remark. In practice, the computation of $\hat{\boldsymbol{G}}_l$ and $\hat{\boldsymbol{h}}_l$ can be efficiently implemented by exploiting the properties of Kronecker product and the fact that $\boldsymbol{\Phi}_l$ are Kronecker products of simple selection matrices. In our experiments, by setting the initial value as $\hat{\boldsymbol{c}}_{\mathrm{WLS}}$, $\{\hat{\boldsymbol{c}}_{\mathrm{FP}}^{(k)}\}$ showed good convergence in a few iterations.

Remark. When the signal-to-noise ratio (SNR) is very high, the conditional number of $\boldsymbol{R}_{\mathrm{ULA}}$ will be large, and the reconstructed $\boldsymbol{R}_{\mathrm{ULA}}hat$ becomes indefinite. In this case, we project $\boldsymbol{R}_{\mathrm{ULA}}hat$ onto the intersection of the positive semidefinite cone $\mathbb{PSD}$ and the Toeplitz subspace $\mathbb{T}$. This can be achieved via the alternating projections method. Because both $\mathbb{PSD}$ and $\mathbb{T}$ are convex and their $\mathbb{PSD}\cap\mathbb{T}\neq\emptyset$, the alternating projections method converges [96].

Remark. For incomplete arrays, not all elements in $\boldsymbol{R}_{\text{ULA}}$ are present in $\boldsymbol{R}_l$. Therefore $\boldsymbol{Q}_{M_0}^T \boldsymbol{\Phi}_l$ is no longer full rank, and we cannot perform the matrix inversion in (5.15) or (5.16). In this case, we first delete the elements we cannot estimate from $\boldsymbol{c}$ and their corresponding basis matrices from $\{\boldsymbol{Q}_{M_0}^i\}_{i=1}^{2M_0-1}$ to form $\tilde{\boldsymbol{c}}$ and $\tilde{\boldsymbol{Q}}_{M_0}$. We then estimate $\tilde{\boldsymbol{c}}$ using (5.15) or (5.16), with $\boldsymbol{Q}_{M_0}$ replaced by $\tilde{\boldsymbol{Q}}_{M_0}$. Finally, we construct a submatrix of $\boldsymbol{R}_{\text{ULA}}$ from the estimated $\tilde{\boldsymbol{c}}$.

5.3 Performance bounds

Because the measurements are assumed independent, the (m,n)-th element of the FIM for our signal model is given by [3, 68]:

$$\text{FIM}_{mn} = \sum_{L=1}^{L} N_l \,\text{tr}\left[\frac{\partial \boldsymbol{R}_l}{\partial \eta_m}\boldsymbol{R}_l^{-1}\frac{\partial \boldsymbol{R}_l}{\partial \eta_n}\boldsymbol{R}_l^{-1}\right].$$

Using the properties of the Kronecker product, we can express the FIM as

$$\text{FIM}_{mn} = \sum_{L=1}^{L} N_l \left[\frac{\partial \boldsymbol{r}_{\text{U}}}{\partial \eta_m}\right]^H \boldsymbol{\Phi}_l^H (\boldsymbol{R}_l^T \otimes \boldsymbol{R}_l)^{-1} \boldsymbol{\Phi}_l \frac{\partial \boldsymbol{r}_{\text{U}}}{\partial \eta_n},$$

where $\boldsymbol{r}_{\text{U}} = \text{vec}(\boldsymbol{R}_{\text{U}})$. Therefore, for complete arrays, the FIM for the Toeplitz parametrization is given by

$$\text{FIM}_{\boldsymbol{c}} = \sum_{l=1}^{L} N_l \boldsymbol{Q}_{M_0}^H \boldsymbol{\Phi}_l^H (\boldsymbol{R}_l^T \otimes \boldsymbol{R}_l)^{-1} \boldsymbol{\Phi}_l \boldsymbol{Q}_{M_0}. \tag{5.17}$$

For incomplete arrays, as stated in Remark 5.2.2, not all elements in $\boldsymbol{c}$ is estimable. To compute the FIM of the estimable elements in $\boldsymbol{c}$, we need to replace $\boldsymbol{Q}_{M_0}$ by $\tilde{\boldsymbol{Q}}_{M_0}$ in a similar fashion.

For parameters $\boldsymbol{\eta} = [\boldsymbol{\theta}, \boldsymbol{p}, \sigma^2]^T$, the FIM is given by

$$\text{FIM}_{\boldsymbol{\eta}} = \sum_{l=1}^{L} N_l \boldsymbol{D}^H \boldsymbol{\Phi}_l^H (\boldsymbol{R}_l^T \otimes \boldsymbol{R}_l)^{-1} \boldsymbol{\Phi}_l \boldsymbol{D}, \tag{5.18}$$

where $\boldsymbol{D} = [\dot{\boldsymbol{A}}_{\text{d}} \boldsymbol{P} \ \boldsymbol{A}_{\text{d}} \ \boldsymbol{i}]$, and $\dot{\boldsymbol{A}}_{\text{d}} = \dot{\boldsymbol{A}}_{\text{U}}^* \odot \boldsymbol{A}_{\text{ULA}} + \boldsymbol{A}_{\text{ULA}}^* \odot \dot{\boldsymbol{A}}_{\text{U}}$, $\boldsymbol{A}_{\text{d}} = \boldsymbol{A}_{\text{ULA}}^* \odot \boldsymbol{A}_{\text{ULA}}$, $\boldsymbol{i} = \text{vec}(\boldsymbol{I}_{M_0})$, and

$$\dot{\boldsymbol{A}}_{\text{U}} = \left[\frac{\partial \boldsymbol{a}_{\text{ULA}}(\theta_1)}{\partial \theta_1} \quad \dots \quad \frac{\partial \boldsymbol{a}_{\text{ULA}}(\theta_K)}{\partial \theta_K} \right]. \tag{5.19}$$

The corresponding CRBs can be obtained by inverting the FIMs in (5.17) and (5.18).

5.4 Numerical examples

We consider the following two sparse linear array configurations in the numerical examples:

- Nested array: $[0, 1, 2, 3, 7, 11, 15, 19]d_0$;
- Coprime array: $[0, 3, 5, 6, 9, 10, 12, 15, 20, 25]d_0$.

In all the experiments, we consider 12 sources uniformly distributed between $-\pi/3$ and $\pi/3$. The number of sources is more than the number of sensors of either array. We set L to be 3. When $L = 2$ the last sensor of each array fails, and when $L = 3$, the last two sensors of each array fail. We set $N_1 = 50\mu$, $N_2 = 100\mu$, and $N_3 = 150\mu$, where μ is a tunable parameter. Hence we have more snapshots with missing data than those with complete data. When making comparisons under different numbers of snapshots, we fixed SNR = 0dB and varied μ from 1 to 20. When making comparisons under different SNRs, we fixed $\mu = 1$ and varied SNR from -20 dB to 20 dB. The root-mean-square errors (RMSEs) were obtained from 500 trials, and the DOAs were estimated by MUSIC.

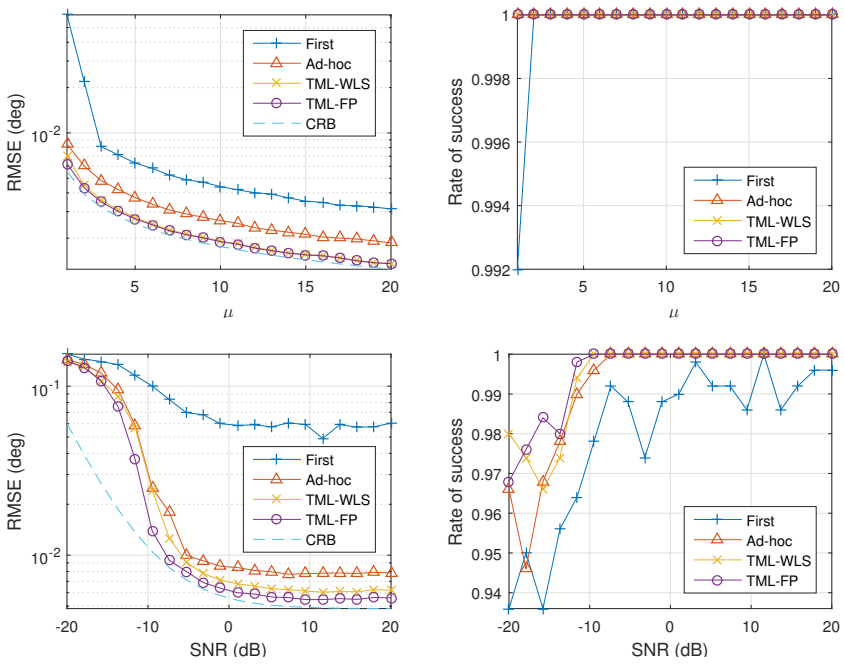


Figure 5.1: Performance of different algorithms for the nested array configuration.

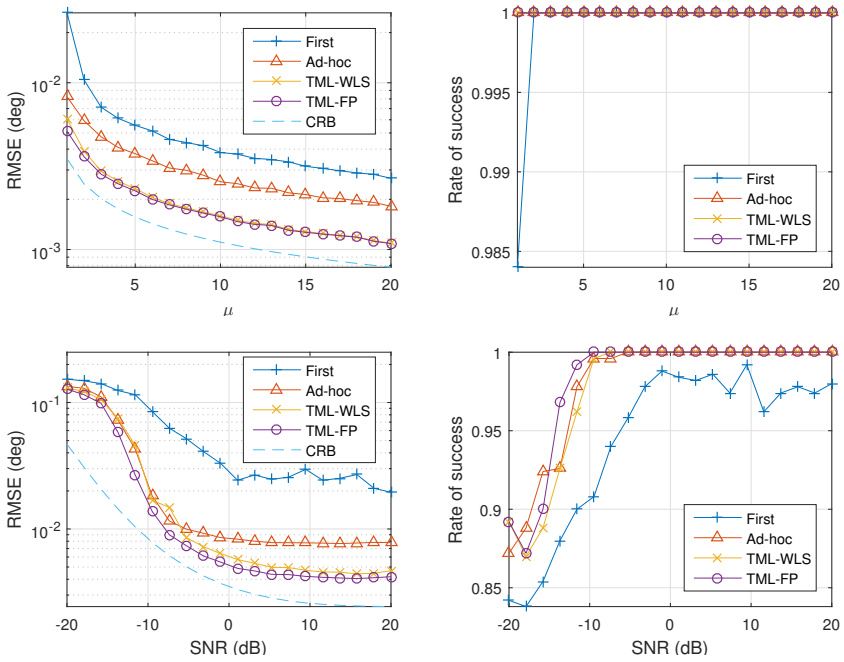


Figure 5.2: Performance of different algorithms for the co-prime array configuration.

In all the figures, "First" denotes the results obtained using only $\hat{\boldsymbol{R}}_1$, while Ad-hoc, TML-WLS, and TML-FP denote the results obtained from (5.8), (5.15), and (5.16), respectively. We also include the CRB obtained from (5.18) for comparison.

Fig. 5.1 illustrates the performance of different algorithms for the nested array configuration. We observe that TML-FP achieves the best performance, and is very close to the CRB, while "First" results in the worst performance because it cannot utilize the information in $\hat{\boldsymbol{R}}_l$ ($l \geq 2$). We observe similar results for the co-prime configuration in Fig. 5.2. However, a gap exists between the RMSE of TML-FP and the CRB, which may be attributed to the fact that the co-prime array is incomplete.

5.5 Chapter summary

In this chapter, we addressed the problem of direction finding using sparse linear arrays with incomplete measurements. By exploiting the difference coarray of a sparse linear array, we proposed to first reconstruct an augmented covariance matrix with enhanced degrees of freedom using the Toeplitz parameterization, and then to apply MUSIC to obtain the DOAs. Specifically, we showed that, by applying our method to co-prime and nested arrays, we can resolve more sources than the number of sensors, even in the missing data case. Through numerical experiments, we demonstrated that our methods achieve better estimation accuracy than the traditional method that uses only the complete measurements.

Chapter 6

Conclusions and Future Work

6.1 Summary and conclusions

In this dissertation, we focused on analyzing the the performance of sparse linear arrays, with and without sensor location errors. We also introduced robust DOA estimation algorithms to tackle the missing data problem caused by sensor failures.

We began by revisiting the background of direction finding using sparse linear arrays. We introduced a mathematical formulation of the difference coarray model, and reviewed two commonly used coarray-based MUSIC algorithms, DA-MUSIC and SS-MUSIC. We proved that, although they are based on different augmented covariance matrices, DA-MUSIC and SS-MUSIC share the same asymptotic estimator error. With this finding, we derived a closed-form asymptotic MSE expression for both DA-MUSIC and SS-MUSIC. We showed that this expression is strictly non-zero in the underdetermined case even if the SNR approaches infinity. This finding analytically explained the "saturation" behavior of SS-MUSIC observed in various numerical simulations in previous studies.

We next derived and analyzed the CRB for general sparse linear arrays. First, we observed that, unlike the classical stochastic CRB, our CRB is applicable even if the number of sources is greater than the number of sensors. Combining our CRB with our closed-form MSE expression of DA-MUSIC and SS-MUSIC, we studied their statistical efficiency. DA-MUSIC and SS-MUSIC excel in the underdetermined case, while the classical MUSIC algorithm is preferred when the number of sources is less than the number of sensors. We investigated the behavior of our CRB in high SNR regions, and showed that, in the underdetermined case, the CRB remains positive definite even if the SNR approaches infinity. Then we established the connection between our CRB and the classical stochastic CRB for uncorrelated sources, and showed that they are asymptotically equal in high SNR regions. Next, we further analyzed the behavior of our CRB for co-prime and nested arrays with large numbers of sensors. We showed that the CRB can decrease at a rate of $O(M^{-5})$ for co-prime and nested arrays with M sensors. This rate is much faster than that of an M-sensor ULA, which is only $O(M^{-3})$. On the other hand, for a fixed aperture, co-prime and nested array require many more snapshots to attain the same performance as a ULA. This finding demonstrates the trade-off between the number of spatial samples and the number of temporal samples.

We next investigated the impact of sensor location errors on the difference coarray model, considering two error models: the deterministic error model and the stochastic error model. For the deterministic error model, we derived a closed-form asymptotic MSE expression for SS-MUSIC, which can be utilized to analyze the sensitivity of SS-MUSIC against the sensor location errors. We showed that the sensor location errors lead to a constant bias in the DOA estimates, which cannot be eliminated by only increasing the SNR. We also derived the CRB for joint estimation of DOA parameters and sensor location errors., and showed that this CRB is indeed applicable, even if the number of sources exceeds the number of sensors. For the stochastic error model, we considered the case when the sensor location

errors follow a Gaussian distribution. When the number of snapshots is large and the sensor locations errors follow a white Gaussian distribution, the effect of the sensor location errors can be indeed modeled as additive white noise. Numerical experiments verified our analytical expressions.

Finally, we proposed new DOA estimators for sparse linear arrays in cases of missing data, considering a more general sensor failure model that does not assume a sequential failure pattern. By exploiting the difference coarray structure, we introduced three algorithms to construct an augmented covariance matrix with enhanced degrees of freedom by combining all the information from the snapshots. Then, we applied MUSIC to this augmented covariance matrix to estimate the DOAs. We also computed the corresponding CRB in the missing data case, and used numerical experiments to demonstrate the efficacy of our algorithms.

6.2 Future directions

In the future, we can potentially extend our research to the following directions:

Partially correlated sources: Throughout this dissertation, we assumed that the sources are uncorrelated. In practice, this assumption may not hold in every environment, due to multi-path effects [97, 98]. Consequently, the sources will be partially correlated. Recently, using sparse linear arrays, new algorithms have been proposed to resolve more correlated sources than the number of sensors [99, 100]. It would be interesting to extend our analysis in Chapter 3 to cases of partially correlated sources and to analyze the maximum number of identifiable partially correlated sources.

Optimal sparse linear array design: In this dissertation, we derived the closed-form asymptotic MSE expression of DA-MUSIC and SS-MUSIC, as well as the CRB. We also analyzed the performance of sparse linear arrays in the presence of sensor location errors. These results enabled us to formulate optimal array design problems. Instead of using pre-configured array geometries, it would be interesting to be able to set constraints on metrics such as the MSE and sensitivity to model errors, and to obtain the optimal array geometries for specific application scenarios by solving the resulting optimization problems.

Extension of the stochastic error model: In our analysis of the stochastic error model, we simply assumed that the time-variant sensor location errors are i.i.d. This assumption, while convenient in statistical analysis, may not hold in practice. In the future, our analysis of the stochastic error model could be extended by introducing motion models for the sensor location errors. Then robust DOA estimation algorithms could be developed based on such models.

Bibliography

[1] S. S. Haykin and J. H. Justice, Eds., *Array signal processing*, ser. Prentice-Hall signal processing series. Englewood Cliffs, N.J: Prentice-Hall, 1985.

[2] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, Jul. 1996.

[3] H. L. Van Trees, *Optimum Array Processing*, ser. Detection, Estimation, and Modulation Theory / Harry L. Van Trees. New York: Wiley, 2002, no. 4.

[4] P. Stoica and R. L. Moses, *Spectral analysis of signals*. Upper Saddle River, N.J: Pearson/Prentice Hall, 2005.

[5] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.

[6] B. Van Veen and K. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, Apr. 1988.

[7] C. Vaidyanathan and K. Buckley, "Performance analysis of the MVDR spatial spectrum estimator," *IEEE Trans. Signal Process.*, vol. 43, no. 6, pp. 1427–1437, Jun. 1995.

[8] J. Li and P. Stoica, Eds., *Robust adaptive beamforming*. Hoboken, NJ: John Wiley, 2006.

[9] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[10] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 5, pp. 1081–1089, Oct. 1986.

[11] R. Kumaresan, L. Scharf, and A. Shaw, "An algorithm for pole-zero modeling and spectral analysis," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 3, pp. 637–640, Jun. 1986.

[12] I. Ziskind and M. Wax, "Maximum likelihood localization of multiple sources by alternating projection," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 10, pp. 1553–1560, Oct. 1988.

[13] P. Stoica, R. L. Moses, B. Friedlander, and T. Soderstrom, "Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 3, pp. 378–392, Mar. 1989.

[14] P. Stoica and K. C. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 7, pp. 1132–1143, Jul. 1990.

[15] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai, "Exact and large sample maximum likelihood techniques for parameter estimation and detection in array processing," in *Radar Array Processing*, T. S. Huang, T. Kohonen, M. R. Schroeder, H. K. V. Lotsch, S. Haykin, J. Litva, and T. J. Shepherd, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1993, vol. 25, pp. 99–151.

[16] P. Stoica and A. Nehorai, "On the concentrated stochastic likelihood function in array signal processing," *Circuits, Systems and Signal Processing*, vol. 14, no. 5, pp. 669–674, Sep. 1995.

[17] P. Stoica, B. Ottersten, M. Viberg, and R. L. Moses, "Maximum likelihood array processing for stochastic coherent sources," *IEEE Trans. Signal Process.*, vol. 44, no. 1, pp. 96–105, Jan. 1996.

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

[19] M. Pesavento and A. B. Gershman, "Maximum-likelihood direction-of-arrival estimation in the presence of unknown nonuniform noise," *IEEE Trans. Signal Process.*, vol. 49, no. 7, pp. 1310–1324, Jul. 2001.

[20] C. E. Chen, F. Lorenzelli, R. E. Hudson, and K. Yao, "Stochastic maximum-likelihood DOA estimation in the presence of unknown nonuniform noise," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3038–3044, Jul. 2008.

[21] B. D. Rao and K. V. S. Hari, "Performance analysis of Root-Music," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 12, pp. 1939–1949, Dec. 1989.

[22] B. Friedlander, "The root-MUSIC algorithm for direction finding with interpolated arrays," *Signal Processing*, vol. 30, no. 1, pp. 15–29, Jan. 1993.

[23] M. Pesavento, A. Gershman, and M. Haardt, "Unitary root-MUSIC with a real-valued eigendecomposition: a theoretical and experimental performance study," *IEEE Trans. Signal Process.*, vol. 48, no. 5, pp. 1306–1314, May 2000.

[24] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul. 1989.

[25] B. D. Rao and K. V. S. Hari, "Performance analysis of ESPRIT and TAM in determining the direction of arrival of plane waves in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 12, pp. 1990–1995, Dec. 1989.

[26] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[27] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9-10, pp. 589–592, May 2008.

[28] D. Malioutov, M. Cetin, and A. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[29] J. Yin and T. Chen, "Direction-of-arrival estimation using a sparse representation of array covariance vectors," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4489–4493, Sep. 2011.

[30] Y. Zhang, M. Amin, and B. Himed, "Sparsity-based DOA estimation using co-prime arrays," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2013, pp. 3967–3971.

[31] Z. Tan and A. Nehorai, "Sparse direction of arrival estimation using co-prime arrays with off-grid targets," *IEEE Signal Process. Lett.*, vol. 21, no. 1, pp. 26–29, Jan. 2014.

[32] Z.-M. Liu, Z.-T. Huang, and Y.-Y. Zhou, "An efficient maximum likelihood method for direction-of-arrival estimation via sparse Bayesian learning," *IEEE Transactions on Wireless Communications*, vol. 11, no. 10, pp. 1–11, Oct. 2012.

[33] P. Gerstoft, C. F. Mecklenbrauker, A. Xenaki, and S. Nannuru, "Multisnapshot sparse bayesian learning for DOA," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1469–1473, Oct. 2016.

[34] G. Tang, B. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7465–7490, Nov. 2013.

[35] Z. Tan, Y. C. Eldar, and A. Nehorai, "Direction of arrival estimation using co-prime arrays: A super resolution viewpoint," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5565–5576, Nov. 2014.

[36] A. Moffet, "Minimum-redundancy linear arrays," *IEEE Trans. Antennas Propag.*, vol. 16, no. 2, pp. 172–175, Mar. 1968.

[37] M. Ishiguro, "Minimum redundancy linear arrays for a large number of antennas," *Radio Science*, vol. 15, no. 6, pp. 1163–1170, Nov. 1980.

[38] P. Pal and P. Vaidyanathan, "Nested arrays: A novel approach to array processing with enhanced degrees of freedom," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4167–4181, Aug. 2010.

[39] P. Pal and P. P. Vaidyanathan, "Coprime sampling and the MUSIC algorithm," in *2011 IEEE Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop (DSP/SPE)*, Jan. 2011, pp. 289–294.

[40] K. Adhikari, J. R. Buck, and K. E. Wage, "Beamforming with extended co-prime sensor arrays." IEEE, May 2013, pp. 4183–4186. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6638447

[41] K. Han and A. Nehorai, "Improved source number detection and direction estimation with nested arrays and ulas using jackknifing," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 6118–6128, Dec. 2013.

[42] ——, "Nested array processing for distributed sources," *IEEE Signal Process. Lett.*, vol. 21, no. 9, pp. 1111–1114, Sep. 2014.

[43] P. Pakrooh, L. L. Scharf, and A. Pezeshki, "Modal analysis using co-prime arrays," *IEEE Trans. Signal Process.*, vol. 64, no. 9, pp. 2429–2442, May 2016.

[44] J. Ramirez and J. L. Krolik, "Synthetic aperture processing for passive co-prime linear sensor arrays," *Digital Signal Processing*, vol. 61, pp. 62–75, Feb. 2017.

[45] S. Qin, Y. Zhang, and M. Amin, "Generalized coprime array configurations for direction-of-arrival estimation," *IEEE Trans. Signal Process.*, vol. 63, no. 6, pp. 1377–1390, Mar. 2015.

[46] C. L. Liu and P. P. Vaidyanathan, "Super nested arrays: Linear sparse arrays with reduced mutual coupling – Part I: Fundamentals," *IEEE Trans. Signal Process.*, vol. 64, no. 15, pp. 3997–4012, Aug. 2016.

[47] ——, "Super nested arrays: Linear sparse arrays with reduced mutual coupling – Part II: High-order extensions," *IEEE Trans. Signal Process.*, vol. 64, no. 16, pp. 4203–4217, Aug. 2016.

[48] P. Pal and P. P. Vaidyanathan, "Nested arrays in two dimensions, Part I: Geometrical considerations," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4694–4705, Sep. 2012.

[49] ——, "Nested arrays in two dimensions, Part II: Application in two dimensional array processing," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4706–4718, Sep. 2012.

[50] K. Han and A. Nehorai, "Nested vector-sensor array processing via tensor modeling," *IEEE Trans. Signal Process.*, vol. 62, no. 10, pp. 2542–2553, May 2014.

[51] ——, "Direction of arrival estimation using nested vector-sensor arrays via tensor modeling," in *2014 IEEE 8th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Jun. 2014, pp. 429–432.

[52] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1783–1795, Oct. 1990.

[53] F. Li, H. Liu, and R. 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.

[54] Y. Abramovich, N. Spencer, and A. Gorokhov, "Detection-estimation of more uncorrelated Gaussian sources than sensors in nonuniform linear antenna arrays .I. Fully augmentable arrays," *IEEE Trans. Signal Process.*, vol. 49, no. 5, pp. 959–971, May 2001.

[55] C.-L. Liu and P. Vaidyanathan, "Remarks on the spatial smoothing step in coarray MUSIC," *IEEE Signal Process. Lett.*, vol. 22, no. 9, Sep. 2015.

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

[57] M. A. Doron and A. J. Weiss, "Performance analysis of direction finding using lag redundancy averaging," *IEEE Trans. Signal Process.*, vol. 41, no. 3, pp. 1386–1391, Mar. 1993.

[58] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, no. 6, pp. 716–723, Dec. 1974.

[59] M. Wax and I. Ziskind, "Detection of the number of coherent signals by the MDL principle," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 8, pp. 1190–1196, Aug. 1989.

[60] Zhaoshui He, A. Cichocki, Shengli Xie, and Kyuwan Choi, "Detecting the number of clusters in n-way probabilistic clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 2006–2021, Nov. 2010.

[61] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 5, pp. 720–741, May 1989.

[62] ——, "MUSIC, maximum likelihood, and Cramer-Rao bound: further results and comparisons," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 12, pp. 2140–2150, Dec. 1990.

[63] A. Gorokhov, Y. Abramovich, and J. F. Bohme, "Unified analysis of DOA estimation algorithms for covariance matrix transforms," *Signal Processing*, vol. 55, no. 1, pp. 107–115, Nov. 1996.

[64] G. W. Stewart, "Error and perturbation bounds for subspaces associated with certain eigenvalue problems," *SIAM Review*, vol. 15, no. 4, pp. 727–764, Oct. 1973.

[65] M. Hawkes, A. Nehorai, and P. Stoica, "Performance breakdown of subspace-based methods: prediction and cure," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01)*, vol. 6, 2001, pp. 4005–4008 vol.6.

[66] B. Johnson, Y. Abramovich, and X. Mestre, "The role of subspace swap in MUSIC performance breakdown," in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008*, Mar. 2008, pp. 2473–2476.

[67] A. Swindlehurst and T. Kailath, "A performance analysis of subspace-based methods in the presence of model errors. I. The MUSIC algorithm," *IEEE Trans. Signal Process.*, vol. 40, no. 7, pp. 1758–1774, Jul. 1992.

[68] S. M. Kay, *Fundamentals of statistical signal processing*, ser. Prentice Hall signal processing series. Englewood Cliffs, N.J: Prentice-Hall PTR, 1993.

[69] C.-L. Liu and P. Vaidyanathan, "Cramér-Rao bounds for coprime and other sparse arrays, which find more sources than sensors," *Digital Signal Processing*, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1051200416300264

[70] G. A. F. Seber, *A Matrix Handbook for Statisticians*. Hoboken, N.J.: Wiley-Interscience, 2008.

[71] 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.

[72] M. Wang, Z. Zhang, and A. Nehorai, "Performance analysis of coarray-based MUSIC and the Cramér-Rao bound," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2017, pp. 3061–3065.

[73] K. Adhikari and J. R. Buck, "Spatial spectral estimation with product processing of a pair of colinear arrays," *IEEE Trans. Signal Process.*, vol. 65, no. 9, pp. 2389–2401, May 2017.

[74] Z. Liu and A. Nehorai, "Statistical angular resolution limit for point sources," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5521–5527, Nov. 2007.

[75] P. P. Vaidyanathan and P. Pal, "Direct-MUSIC on sparse arrays," in *2012 International Conference on Signal Processing and Communications (SPCOM)*, Jul. 2012, pp. 1–5.

[76] B. Friedlander and A. J. Weiss, "Direction finding in the presence of mutual coupling," *IEEE Trans. Antennas Propag.*, vol. 39, no. 3, pp. 273–284, Mar. 1991.

[77] E. BouDaher, F. Ahmad, M. G. Amin, and A. Hoorfar, "Mutual coupling effect and compensation in non-uniform arrays for direction-of-arrival estimation," *Digital Signal Processing*, Jun. 2016.

[78] A. J. Weiss and B. Friedlander, "Eigenstructure methods for direction finding with sensor gain and phase uncertainties," *Circuits, Systems and Signal Processing*, vol. 9, no. 3, pp. 271–300, Sep. 1990.

[79] A. Liu, G. Liao, C. Zeng, Z. Yang, and Q. Xu, "An eigenstructure method for estimating DOA and sensor gain-phase errors," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 5944–5956, Dec. 2011.

[80] Y. Rockah and P. Schultheiss, "Array shape calibration using sources in unknown locations–Part I: Far-field sources," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 3, pp. 286–299, Mar. 1987.

[81] M. Aktas and T. Tuncer, "Iterative HOS-SOS (IHOSS) algorithm for direction-of-arrival estimation and sensor localization," *IEEE Trans. Signal Process.*, vol. 58, no. 12, pp. 6181–6194, Dec. 2010.

[82] A. Koochakzadeh and P. Pal, "Sparse source localization using perturbed arrays via bi-affine modeling," *Digital Signal Processing*, vol. 61, pp. 15–25, 2017.

[83] K. Han, P. Yang, and A. Nehorai, "Calibrating nested sensor arrays with model errors," *IEEE Trans. Antennas Propag.*, vol. 63, no. 11, pp. 4739–4748, Nov. 2015.

[84] A. Swindlehurst and T. Kailath, "A performance analysis of subspace-based methods in the presence of model error. II. Multidimensional algorithms," *IEEE Trans. Signal Process.*, vol. 41, no. 9, pp. 2882–2890, Sep. 1993.

[85] A. Ferreol, P. Larzabal, and M. Viberg, "Statistical analysis of the MUSIC algorithm in the presence of modeling errors, taking into account the resolution probability," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4156–4166, Aug. 2010.

[86] A. Koochakzadeh and P. Pal, "Performance of uniform and sparse non-uniform samplers in presence of modeling errors: A Cramér-Rao bound based study," *IEEE Trans. Signal Process.*, vol. 65, no. 6, pp. 1607–1621, Mar. 2017.

[87] A. Weiss and B. Friedlander, "Array shape calibration using sources in unknown locations-a maximum likelihood approach," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 12, pp. 1958–1966, Dec. 1989.

[88] A. Koochakzadeh and P. Pal, "Cramér Rao bounds for underdetermined source localization," *IEEE Signal Process. Lett.*, vol. 23, no. 7, pp. 919–923, Jul. 2016.

[89] T. J. Rothenberg, "Identification in parametric models," *Econometrica*, vol. 39, no. 3, pp. 577–591, 1971.

[90] M. Hawkes and A. Nehorai, "Acoustic vector-sensor processing in the presence of a reflecting boundary," *IEEE Trans. Signal Process.*, vol. 48, no. 11, pp. 2981–2993, Nov. 2000.

[91] E. G. Larsson and P. Stoica, "High-resolution direction finding: the missing data case," *IEEE Trans. Signal Process.*, vol. 49, no. 5, pp. 950–958, May 2001.

[92] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, 2009.

[93] E. J. Candes and Y. Plan, "Matrix completion with noise," *Proc. IEEE*, vol. 98, no. 6, pp. 925–936, June 2010.

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

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

[96] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK ; New York: Cambridge University Press, 2004.

[97] P. Stoica, B. Ottersten, and M. Viberg, "Optimal array signal processing in the presence of coherent wavefronts," in *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*, vol. 5, May 1996, pp. 2904–2907 vol. 5.

[98] J. Wang, H. Xu, G. J. T. Leus, and G. A. E. Vandenbosch, "Experimental assessment of the co-array concept for DoA estimation in wireless communications," *IEEE Trans. Antennas Propag.*, pp. 1–1, 2018.

[99] D. D. Arianandа and G. Leus, "Direction of arrival estimation for more correlated sources than active sensors," *Signal Processing*, vol. 93, no. 12, pp. 3435–3448, Dec. 2013.

[100] S. Qin, Y. D. Zhang, and M. G. Amin, "DOA estimation of mixed coherent and uncorrelated targets exploiting coprime MIMO radar," *Digital Signal Processing*, vol. 61, pp. 26–34, Feb. 2017.

Appendix A

Proof of Theorem 3.1

We first derive the first-order expression of DA-MUSIC. Denote the eigendecomposition of $\boldsymbol{R}_{\mathrm{v1}}$ by

$$\boldsymbol{R}_{\mathrm{v1}} = \boldsymbol{E}_{\mathrm{s}}\boldsymbol{\Lambda}_{\mathrm{s1}}\boldsymbol{E}_{\mathrm{s}}^{H} + \boldsymbol{E}_{\mathrm{n}}\boldsymbol{\Lambda}_{\mathrm{n1}}\boldsymbol{E}_{\mathrm{n}}^{H},$$

where $\boldsymbol{E}_{\mathrm{s}}$ and $\boldsymbol{E}_{\mathrm{n}}$ are eigenvectors of the signal subspace and noise subspace, respectively, and $\boldsymbol{\Lambda}_{\mathrm{s1}}, \boldsymbol{\Lambda}_{\mathrm{n1}}$ are the corresponding eigenvalues. Specifically, we have $\boldsymbol{\Lambda}_{\mathrm{n1}} = \sigma^2\boldsymbol{I}$.

Let $\tilde{\boldsymbol{R}}_{\mathrm{v1}} = \boldsymbol{R}_{\mathrm{v1}} + \Delta\boldsymbol{R}_{\mathrm{v1}}$, $\tilde{\boldsymbol{E}}_{\mathrm{n1}} = \boldsymbol{E}_{\mathrm{n}} + \Delta\boldsymbol{E}_{\mathrm{n1}}$, and $\tilde{\boldsymbol{\Lambda}}_{\mathrm{n1}} = \boldsymbol{\Lambda}_{\mathrm{n1}} + \Delta\boldsymbol{\Lambda}_{\mathrm{n1}}$ be the perturbed versions of $\boldsymbol{R}_{\mathrm{v1}}$, $\boldsymbol{E}_{\mathrm{n}}$, and $\boldsymbol{\Lambda}_{\mathrm{n1}}$. The following equality holds:

$$(\boldsymbol{R}_{\mathrm{v1}} + \Delta\boldsymbol{R}_{\mathrm{v1}})(\boldsymbol{E}_{\mathrm{n}} + \Delta\boldsymbol{E}_{\mathrm{n1}}) = (\boldsymbol{E}_{\mathrm{n}} + \Delta\boldsymbol{E}_{\mathrm{n1}})(\boldsymbol{\Lambda}_{\mathrm{n1}} + \Delta\boldsymbol{\Lambda}_{\mathrm{n1}}).$$

If the perturbation is small and the SNR is high, we can omit high-order terms and obtain [53, 64, 67]

$$\boldsymbol{A}_{\mathrm{co}}^{H}\Delta\boldsymbol{E}_{\mathrm{n1}} \doteq -\boldsymbol{P}^{-1}\boldsymbol{A}_{\mathrm{co}}^{\dagger}\Delta\boldsymbol{R}_{\mathrm{v1}}\boldsymbol{E}_{\mathrm{n}}. \tag{A.1}$$

Because $\boldsymbol{P}$ is diagonal, for a specific θ_k, we have

$$\boldsymbol{a}^H(\theta_k)\Delta\boldsymbol{E}_{\mathrm{n1}} \doteq -p_k^{-1}\boldsymbol{e}_k^T\boldsymbol{A}_{\mathrm{co}}^{\dagger}\Delta\boldsymbol{R}_{\mathrm{v1}}\boldsymbol{E}_{\mathrm{n}}, \tag{A.2}$$

where $\boldsymbol{e}_k$ is the k-th column of the identity matrix $\boldsymbol{I}_{K\times K}$. Based on the conclusion in Appendix B of [61], under sufficiently small perturbations, the error expression of DA-MUSIC for the k-th DOA is given by

$$\hat{\theta}_k^{(1)} - \theta_k \doteq -\frac{\Re[\boldsymbol{a}_{\mathrm{co}}^H(\theta_k)\Delta\boldsymbol{E}_{\mathrm{n1}}\boldsymbol{E}_{\mathrm{n}}^H\dot{\boldsymbol{a}}_{\mathrm{co}}(\theta_k))]}{\dot{\boldsymbol{a}}_{\mathrm{co}}^H(\theta_k)\boldsymbol{E}_{\mathrm{n}}\boldsymbol{E}_{\mathrm{n}}^H\dot{\boldsymbol{a}}_{\mathrm{co}}(\theta_k)}, \tag{A.3}$$

where $\dot{\boldsymbol{a}}_{\mathrm{co}}(\theta_k) = \partial\boldsymbol{a}_{\mathrm{co}}(\theta_k)/\partial\theta_k$.

Substituting (A.2) into (A.3) gives

$$\hat{\theta}_k^{(1)} - \theta_k \doteq \frac{\Re[\boldsymbol{e}_k^T\boldsymbol{A}_{\mathrm{co}}^{\dagger}\Delta\boldsymbol{R}_{\mathrm{v1}}\boldsymbol{E}_{\mathrm{n}}\boldsymbol{E}_{\mathrm{n}}^H\dot{\boldsymbol{a}}_{\mathrm{co}}(\theta_k)]}{p_k\dot{\boldsymbol{a}}_{\mathrm{co}}^H(\theta_k)\boldsymbol{E}_{\mathrm{n}}\boldsymbol{E}_{\mathrm{n}}^H\dot{\boldsymbol{a}}_{\mathrm{co}}(\theta_k)}. \tag{A.4}$$

Because $\mathrm{vec}(\boldsymbol{AXB}) = (\boldsymbol{B}^T\otimes\boldsymbol{A})\,\mathrm{vec}(\boldsymbol{X})$ and $\boldsymbol{E}_{\mathrm{n}}\boldsymbol{E}_{\mathrm{n}}^H = \boldsymbol{\Pi}_{\boldsymbol{A}_{\mathrm{co}}}^{\perp}$, we can use the notations introduced in (3.2b)–(3.2d) to express (A.4) as

$$\hat{\theta}_k^{(1)} - \theta_k \doteq -(\gamma_k p_k)^{-1}\Re[(\boldsymbol{\beta}_k\otimes\boldsymbol{\alpha}_k)^T\Delta\boldsymbol{r}_{\mathrm{v1}}], \tag{A.5}$$

where $\Delta\boldsymbol{r}_{\mathrm{v1}} = \mathrm{vec}(\Delta\boldsymbol{R}_{\mathrm{v1}})$.

Note that $\tilde{\boldsymbol{R}}_{\mathrm{v1}}$ is constructed from $\tilde{\boldsymbol{R}}$. It follows that $\Delta\boldsymbol{R}_{\mathrm{v1}}$ actually depends on $\Delta\boldsymbol{R}$, which is the perturbation part of the covariance matrix $\boldsymbol{R}$. By the definition of $\boldsymbol{R}_{\mathrm{v1}}$,

$$\Delta\boldsymbol{r}_{\mathrm{v1}} = \mathrm{vec}(\begin{bmatrix}\boldsymbol{\Gamma}_{M_{\mathrm{co}}}\Delta\boldsymbol{z} & \cdots & \boldsymbol{\Gamma}_2\Delta\boldsymbol{z} & \boldsymbol{\Gamma}_1\Delta\boldsymbol{z}\end{bmatrix}) = \boldsymbol{\Gamma}\boldsymbol{F}\Delta\boldsymbol{r},$$

where $\boldsymbol{\Gamma} = [\boldsymbol{\Gamma}_{M_{\rm co}}^T \, \boldsymbol{\Gamma}_{M_{\rm co}-1}^T \cdots \boldsymbol{\Gamma}_1^T]^T$ and $\Delta\boldsymbol{r} = \mathrm{vec}(\Delta\boldsymbol{R})$.

Let $\boldsymbol{\xi}_k = \boldsymbol{F}^T\boldsymbol{\Gamma}^T(\boldsymbol{\beta}_k \otimes \boldsymbol{\alpha}_k)$. We can now express (A.5) in terms of $\Delta\boldsymbol{r}$ as

$$\hat{\theta}_k^{(1)} - \theta_k \doteq -(\gamma_k p_k)^{-1}\Re(\boldsymbol{\xi}_k^T\Delta\boldsymbol{r}), \tag{A.6}$$

which completes the first part of the proof.

We next consider the first-order error expression of SS-MUSIC. From (2.17) we know that $\boldsymbol{R}_{\rm v2}$ shares the same eigenvectors as $\boldsymbol{R}_{\rm v1}$. Hence the eigendecomposition of $\boldsymbol{R}_{\rm v2}$ can be expressed by

$$\boldsymbol{R}_{\rm v2} = \boldsymbol{E}_{\rm s}\boldsymbol{\Lambda}_{\rm s2}\boldsymbol{E}_{\rm s}^H + \boldsymbol{E}_{\rm n}\boldsymbol{\Lambda}_{\rm n2}\boldsymbol{E}_{\rm n}^H,$$

where $\boldsymbol{\Lambda}_{\rm s2}$ and $\boldsymbol{\Lambda}_{\rm n2}$ are the eigenvalues of the signal subspace and noise subspace. Specifically, we have $\boldsymbol{\Lambda}_{\rm n2} = \sigma^4/M_{\rm co}\boldsymbol{I}$. Note that $\boldsymbol{R}_{\rm v2} = (\boldsymbol{A}_{\rm co}\boldsymbol{P}\boldsymbol{A}_{\rm co}^H + \sigma^2\boldsymbol{I})^2/M_{\rm co}$. Following a similar approach to the one we used to obtain (A.1), we get

$$\boldsymbol{A}_{\rm co}^H\Delta\boldsymbol{E}_{\rm n2} \doteq -M_{\rm co}\boldsymbol{P}^{-1}(\boldsymbol{P}\boldsymbol{A}_{\rm co}^H\boldsymbol{A}_{\rm co} + 2\sigma^2\boldsymbol{I})^{-1}\boldsymbol{A}_{\rm co}^\dagger\Delta\boldsymbol{R}_{\rm v2}\boldsymbol{E}_{\rm n},$$

where $\Delta\boldsymbol{E}_{\rm n2}$ is the perturbation of the noise eigenvectors produced by $\Delta\boldsymbol{R}_{\rm v2}$. After omitting high-order terms, $\Delta\boldsymbol{R}_{\rm v2}$ is given by

$$\Delta\boldsymbol{R}_{\rm v2} \doteq \frac{1}{M_{\rm co}}\sum_{k=1}^{M_{\rm co}}(\boldsymbol{z}_k\Delta\boldsymbol{z}_k^H + \Delta\boldsymbol{z}_k\boldsymbol{z}_k^H).$$

According to [38], each subarray observation vector $\boldsymbol{z}_k$ can be expressed by

$$\boldsymbol{z}_k = \boldsymbol{A}_{\rm co}\boldsymbol{\Psi}^{M_{\rm co}-k}\boldsymbol{p} + \sigma^2\boldsymbol{i}_{M_{\rm co}-k+1}, \tag{A.7}$$

for $k = 1, 2, \ldots, M_{\text{co}}$, where $\boldsymbol{i}_l$ is a vector of length M_{co} whose elements are zero except for the l-th element being one, and

$$\boldsymbol{\Psi} = \text{diag}(e^{-j\omega_1}, e^{-j\omega_2}, \ldots, e^{-j\omega_K}).$$

Here $\omega_k = (2\pi d_0 \sin\theta_k)/\lambda$. We can further obtain that

$$\sum_{k=1}^{M_{\text{co}}} \sigma^2 \boldsymbol{i}_{M_{\text{co}}-k+1} \Delta \boldsymbol{z}_k^H = \sigma^2 \Delta \boldsymbol{R}_{\text{v1}}^H,$$

and that

$$\begin{aligned}
&\sum_{k=1}^{M_{\text{co}}} \boldsymbol{A}_{\text{co}} \boldsymbol{\Psi}^{M_{\text{co}}-k} \boldsymbol{p} \Delta \boldsymbol{z}_k^H \\
=&\boldsymbol{A}_{\text{co}} \boldsymbol{P} \begin{bmatrix} e^{-j(M_{\text{co}}-1)\omega_1} & e^{-j(M_{\text{co}}-2)\omega_1} & \cdots & 1 \\ e^{-j(M_{\text{co}}-1)\omega_2} & e^{-j(M_{\text{co}}-2)\omega_2} & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ e^{-j(M_{\text{co}}-1)\omega_K} & e^{-j(M_{\text{co}}-2)\omega_K} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \Delta \boldsymbol{z}_1^H \\ \Delta \boldsymbol{z}_2^H \\ \vdots \\ \Delta \boldsymbol{z}_{M_{\text{co}}}^H \end{bmatrix} \\
=&\boldsymbol{A}_{\text{co}} \boldsymbol{P} (\boldsymbol{T}_{M_{\text{co}}} \boldsymbol{A}_{\text{co}})^H \boldsymbol{T}_{M_{\text{co}}} \Delta \boldsymbol{R}_{\text{v1}}^H \\
=&\boldsymbol{A}_{\text{co}} \boldsymbol{P} \boldsymbol{A}_{\text{co}}^H \Delta \boldsymbol{R}_{\text{v1}}^H,
\end{aligned}$$

where $\boldsymbol{T}_{M_{\text{co}}}$ is a $M_{\text{co}} \times M_{\text{co}}$ permutation matrix whose anti-diagonal elements are one, and whose remaining elements are zero. Because $\Delta \boldsymbol{R} = \Delta \boldsymbol{R}^H$, by Lemma 2.2 we know that $\Delta \boldsymbol{z}$ is conjugate symmetric. According to the definition of $\boldsymbol{R}_{\text{v1}}$, it is straightforward to show that $\Delta \boldsymbol{R}_{\text{v1}} = \Delta \boldsymbol{R}_{\text{v1}}^H$ also holds. Hence

$$\Delta \boldsymbol{R}_{\text{v2}} \doteq \frac{1}{M_{\text{co}}} [(\boldsymbol{A}_{\text{co}} \boldsymbol{P} \boldsymbol{A}_{\text{co}}^H + 2\sigma^2 \boldsymbol{I}) \Delta \boldsymbol{R}_{\text{v1}} + \Delta \boldsymbol{R}_{\text{v1}} \boldsymbol{A}_{\text{co}} \boldsymbol{P} \boldsymbol{A}_{\text{co}}^H].$$

Substituting $\Delta \boldsymbol{R}_{\mathrm{v}2}$ into the expression of $\boldsymbol{A}_{\mathrm{co}}^H \Delta \boldsymbol{E}_{\mathrm{n}2}$, and utilizing the property that $\boldsymbol{A}_{\mathrm{co}}^H \boldsymbol{E}_{\mathrm{n}} = \boldsymbol{0}$, we can express $\boldsymbol{A}_{\mathrm{co}}^H \Delta \boldsymbol{E}_{\mathrm{n}2}$ as

$$-\boldsymbol{P}^{-1}(\boldsymbol{P}\boldsymbol{A}_{\mathrm{co}}^H \boldsymbol{A}_{\mathrm{co}} + 2\sigma^2 \boldsymbol{I})^{-1} \boldsymbol{A}_{\mathrm{co}}^{\dagger}(\boldsymbol{A}_{\mathrm{co}} \boldsymbol{P}\boldsymbol{A}_{\mathrm{co}}^H + 2\sigma^2 \boldsymbol{I}) \Delta \boldsymbol{R}_{\mathrm{v}1} \boldsymbol{E}_{\mathrm{n}}.$$

Observe that

$$\begin{aligned}
\boldsymbol{A}_{\mathrm{co}}^{\dagger}(\boldsymbol{A}_{\mathrm{co}} \boldsymbol{P}\boldsymbol{A}_{\mathrm{co}}^H + 2\sigma^2 \boldsymbol{I}) =& (\boldsymbol{A}_{\mathrm{co}}^H \boldsymbol{A}_{\mathrm{co}})^{-1} \boldsymbol{A}_{\mathrm{co}}^H (\boldsymbol{A}_{\mathrm{co}} \boldsymbol{P}\boldsymbol{A}_{\mathrm{co}}^H + 2\sigma^2 \boldsymbol{I}) \\
=& [\boldsymbol{P}\boldsymbol{A}_{\mathrm{co}}^H + 2\sigma^2 (\boldsymbol{A}_{\mathrm{co}}^H \boldsymbol{A}_{\mathrm{co}})^{-1} \boldsymbol{A}_{\mathrm{co}}^H] \\
=& (\boldsymbol{P}\boldsymbol{A}_{\mathrm{co}}^H \boldsymbol{A}_{\mathrm{co}} + 2\sigma^2 \boldsymbol{I}) \boldsymbol{A}_{\mathrm{co}}^{\dagger}.
\end{aligned}$$

Hence the term $(\boldsymbol{P}\boldsymbol{A}_{\mathrm{co}}^H \boldsymbol{A}_{\mathrm{co}} + 2\sigma^2 \boldsymbol{I})$ gets canceled and we obtain

$$\boldsymbol{A}_{\mathrm{co}}^H \Delta \boldsymbol{E}_{\mathrm{n}2} \doteq -\boldsymbol{P}^{-1} \boldsymbol{A}_{\mathrm{co}}^{\dagger} \Delta \boldsymbol{R}_{\mathrm{v}1} \boldsymbol{E}_{\mathrm{n}}, \tag{A.8}$$

which coincides with the first-order error expression of $\boldsymbol{A}_{\mathrm{co}}^H \Delta \boldsymbol{E}_{\mathrm{n}1}$.

Appendix B

Proof of Theorem 3.2

Before proceeding to the main proof, we introduce the following definition.

Definition B.1. *Let* $\boldsymbol{A} = [\boldsymbol{a}_1\, \boldsymbol{a}_2\, \ldots \boldsymbol{a}_N] \in \mathbb{R}^{N\times N}$*, and* $\boldsymbol{B} = [\boldsymbol{b}_1\, \boldsymbol{b}_2\, \ldots \boldsymbol{b}_N] \in \mathbb{R}^{N\times N}$*. The structured matrix* $\boldsymbol{C_{AB}} \in \mathbb{R}^{N^2\times N^2}$ *is defined as*

$$\boldsymbol{C_{AB}} = \begin{bmatrix} \boldsymbol{a}_1\boldsymbol{b}_1^T & \boldsymbol{a}_2\boldsymbol{b}_1^T & \ldots & \boldsymbol{a}_N\boldsymbol{b}_1^T \\ \boldsymbol{a}_1\boldsymbol{b}_2^T & \boldsymbol{a}_2\boldsymbol{b}_2^T & \ldots & \boldsymbol{a}_N\boldsymbol{b}_2^T \\ \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{a}_1\boldsymbol{b}_N^T & \boldsymbol{a}_2\boldsymbol{b}_N^T & \ldots & \boldsymbol{a}_N\boldsymbol{b}_N^T \end{bmatrix}.$$

We now start deriving the explicit MSE expression. According to (A.6),

$$\begin{aligned}
&\mathbb{E}[(\hat{\theta}_{k_1}-\theta_{k_1})(\hat{\theta}_{k_2}-\theta_{k_2})]\\
\doteq&(\gamma_{k_1}p_{k_1})^{-1}(\gamma_{k_2}p_{k_2})^{-1}\mathbb{E}[\Re(\boldsymbol{\xi}_{k_1}^T\Delta\boldsymbol{r})\Re(\boldsymbol{\xi}_{k_2}^T\Delta\boldsymbol{r})]\\
=&(\gamma_{k_1}p_{k_1})^{-1}(\gamma_{k_2}p_{k_2})^{-1}\big\{\Re(\boldsymbol{\xi}_{k_1})^T\mathbb{E}[\Re(\Delta\boldsymbol{r})\Re(\Delta\boldsymbol{r})^T]\Re(\boldsymbol{\xi}_{k_2})\\
&+\Im(\boldsymbol{\xi}_{k_1})^T\mathbb{E}[\Im(\Delta\boldsymbol{r})\Im(\Delta\boldsymbol{r})^T]\Im(\boldsymbol{\xi}_{k_2})\\
&-\Re(\boldsymbol{\xi}_{k_1})^T\mathbb{E}[\Re(\Delta\boldsymbol{r})\Im(\Delta\boldsymbol{r})^T]\Im(\boldsymbol{\xi}_{k_2})\\
&-\Re(\boldsymbol{\xi}_{k_2})^T\mathbb{E}[\Re(\Delta\boldsymbol{r})\Im(\Delta\boldsymbol{r})^T]\Im(\boldsymbol{\xi}_{k_1})\big\},
\end{aligned}\tag{B.1}$$

where we used the property that $\Re(\boldsymbol{AB})=\Re(\boldsymbol{A})\Re(\boldsymbol{B})-\Im(\boldsymbol{A})\Im(\boldsymbol{B})$ for two complex matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ with proper dimensions.

To obtain the closed-form expression for (B.1), we need to compute the four expectations. It should be noted that in the case of finite snapshots, $\Delta\boldsymbol{r}$ does not follow a circularly-symmetric complex Gaussian distribution. Therefore we cannot directly use the properties of the circularly-symmetric complex Gaussian distribution to evaluate the expectations. For brevity, we demonstrate the computation of only the first expectation in (B.1). The computation of the remaining three expectations follows the same idea.

Let $\boldsymbol{r}_i$ denote the i-th column of $\boldsymbol{R}$ in (2.4). Its estimate, $\hat{\boldsymbol{r}}_i$, is given by $\sum_{t=1}^{N}\boldsymbol{y}(t)y_i^*(t)$, where $y_i(t)$ is the i-th element of $\boldsymbol{y}(t)$. Because $\mathbb{E}[\hat{\boldsymbol{r}}_i]=\boldsymbol{r}_i$,

$$\mathbb{E}[\Re(\Delta\boldsymbol{r}_i)\Re(\Delta\boldsymbol{r}_l)^T]=\mathbb{E}[\Re(\hat{\boldsymbol{r}}_i)\Re(\hat{\boldsymbol{r}}_l)^T]-\Re(\boldsymbol{r}_i)\Re(\boldsymbol{r}_l)^T.\tag{B.2}$$

The second term in (B.2) is deterministic, and the first term in (B.2) can be expanded into

$$
\begin{aligned}
&\frac{1}{N^2}\mathbb{E}\left[\Re\Big(\sum_{s=1}^{N}\boldsymbol{y}(s)y_i^*(s)\Big)\Re\Big(\sum_{t=1}^{N}\boldsymbol{y}(t)y_l^*(t)\Big)^T\right]\\
=&\frac{1}{N^2}\mathbb{E}\left[\sum_{s=1}^{N}\sum_{t=1}^{N}\Re(\boldsymbol{y}(s)y_i^*(s))\Re(\boldsymbol{y}(t)y_l^*(t))^T\right]\\
=&\frac{1}{N^2}\sum_{s=1}^{N}\sum_{t=1}^{N}\mathbb{E}\Big\{\big[\Re(\boldsymbol{y}(s))\Re(y_i^*(s))-\Im(\boldsymbol{y}(s))\Im(y_i^*(s))\big]\\
&\quad\big[\Re(\boldsymbol{y}(t))^T\Re(y_l^*(t))-\Im(\boldsymbol{y}(t))^T\Im(y_l^*(t))\big]\Big\} \qquad \text{(B.3)}\\
=&\frac{1}{N^2}\sum_{s=1}^{N}\sum_{t=1}^{N}\Big\{\mathbb{E}[\Re(\boldsymbol{y}(s))\Re(y_i(s))\Re(\boldsymbol{y}(t))^T\Re(y_l(t))]\\
&+\mathbb{E}[\Re(\boldsymbol{y}(s))\Re(y_i(s))\Im(\boldsymbol{y}(t))^T\Im(y_l(t))]\\
&+\mathbb{E}[\Im(\boldsymbol{y}(s))\Im(y_i(s))\Re(\boldsymbol{y}(t))^T\Re(y_l(t))]\\
&+\mathbb{E}[\Im(\boldsymbol{y}(s))\Im(y_i(s))\Im(\boldsymbol{y}(t))^T\Im(y_l(t))]\Big\}.
\end{aligned}
$$

We first consider the partial sum of the cases when $s \neq t$. By **A4**, $\boldsymbol{y}(s)$ and $\boldsymbol{y}(t)$ are uncorrelated Gaussians. Recall that for $\boldsymbol{x} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$,

$$
\begin{aligned}
&\mathbb{E}[\Re(\boldsymbol{x})\Re(\boldsymbol{x})^T] = \frac{1}{2}\Re(\boldsymbol{\Sigma}),\ \mathbb{E}[\Re(\boldsymbol{x})\Im(\boldsymbol{x})^T] = -\frac{1}{2}\Im(\boldsymbol{\Sigma})\\
&\mathbb{E}[\Im(\boldsymbol{x})\Re(\boldsymbol{x})^T] = \frac{1}{2}\Im(\boldsymbol{\Sigma}),\ \mathbb{E}[\Im(\boldsymbol{x})\Im(\boldsymbol{x})^T] = \frac{1}{2}\Re(\boldsymbol{\Sigma}).
\end{aligned}
$$

We have

$$
\mathbb{E}[\Re(\boldsymbol{y}(s))\Re(y_i(s))\Re(\boldsymbol{y}(t))^T\Re(y_l(t))] = \mathbb{E}[\Re(\boldsymbol{y}(s))\Re(y_i(s))]\mathbb{E}[\Re(\boldsymbol{y}(t))^T\Re(y_l(t))] = \frac{1}{4}\Re(\boldsymbol{r}_i)\Re(\boldsymbol{r}_l)^T.
$$

Similarly, we can obtain that when $s \neq t$,

$$\begin{aligned}
\mathbb{E}[\Re(\boldsymbol{y}(s))\Re(y_i(s))\Im(\boldsymbol{y}(t))^T\Im(y_l(t))] &= \frac{1}{4}\Re(\boldsymbol{r}_i)\Re(\boldsymbol{r}_l)^T, \\
\mathbb{E}[\Im(\boldsymbol{y}(s))\Im(y_i(s))\Re(\boldsymbol{y}(t))^T\Re(y_l(t))] &= \frac{1}{4}\Re(\boldsymbol{r}_i)\Re(\boldsymbol{r}_l)^T, \\
\mathbb{E}[\Im(\boldsymbol{y}(s))\Im(y_i(s))\Im(\boldsymbol{y}(t))^T\Im(y_l(t))] &= \frac{1}{4}\Re(\boldsymbol{r}_i)\Re(\boldsymbol{r}_l)^T.
\end{aligned} \tag{B.4}$$

Therefore the partial sum of the cases when $s \neq t$ is given by $(1 - 1/N)\Re(\boldsymbol{r}_i)\Re(\boldsymbol{r}_l)^T$.

We now consider the partial sum of the cases when $s = t$. We first consider the first expectation inside the double summation in (B.3).

Recall that for $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$, $\mathbb{E}[x_i x_l x_p x_q] = \sigma_{il}\sigma_{pq} + \sigma_{ip}\sigma_{lq} + \sigma_{iq}\sigma_{lp}$. We can express the (m, n)-th element of the matrix $\mathbb{E}[\Re(\boldsymbol{y}(t))\Re(y_i(t))\Re(\boldsymbol{y}(t))^T\Re(y_l(t))]$ as

$$\begin{aligned}
&\mathbb{E}[\Re(y_m(t))\Re(y_i(t))\Re(y_n(t))\Re(y_l(t))] \\
=&\mathbb{E}[\Re(y_m(t))\Re(y_i(t))\Re(y_l(t))\Re(y_n(t))] \\
=&\mathbb{E}[\Re(y_m(t))\Re(y_i(t))]\mathbb{E}[\Re(y_l(t))\Re(y_n(t))] \\
&+ \mathbb{E}[\Re(y_m(t))\Re(y_l(t))]\mathbb{E}[\Re(y_i(t))\Re(y_n(t))] \\
&+ \mathbb{E}[\Re(y_m(t))\Re(y_n(t))]\mathbb{E}[\Re(y_i(t))\Re(y_l(t))] \\
=&\frac{1}{4}[\Re(R_{mi})\Re(R_{ln}) + \Re(R_{ml})\Re(R_{in}) + \Re(R_{mn})\Re(R_{il})].
\end{aligned}$$

Hence

$$\mathbb{E}[\Re(\boldsymbol{y}(t))\Re(y_i(t))\Re(\boldsymbol{y}(t))^T\Re(y_l(t))] = \frac{1}{4}[\Re(\boldsymbol{r}_i)\Re(\boldsymbol{r}_l)^T + \Re(\boldsymbol{r}_l)\Re(\boldsymbol{r}_i)^T + \Re(\boldsymbol{R})\Re(R_{il})].$$

Similarly, we obtain that

$$\mathbb{E}[\Im(\boldsymbol{y}(t))\Im(y_i(t))\Im(\boldsymbol{y}(t))^T\Im(y_l(t))] = \frac{1}{4}[\Re(\boldsymbol{r}_i)\Re(\boldsymbol{r}_l)^T + \Re(\boldsymbol{r}_l)\Re(\boldsymbol{r}_i)^T + \Re(\boldsymbol{R})\Re(R_{il})],$$
$$\mathbb{E}[\Re(\boldsymbol{y}(t))\Re(y_i(t))\Im(\boldsymbol{y}(t))^T\Im(y_l(t))] = \frac{1}{4}[\Re(\boldsymbol{r}_i)\Re(\boldsymbol{r}_l)^T - \Im(\boldsymbol{r}_l)\Im(\boldsymbol{r}_i)^T + \Im(\boldsymbol{R})\Im(R_{il})],$$
$$\mathbb{E}[\Im(\boldsymbol{y}(t))\Im(y_i(t))\Re(\boldsymbol{y}(t))^T\Re(y_l(t))] = \frac{1}{4}[\Re(\boldsymbol{r}_i)\Re(\boldsymbol{r}_l)^T - \Im(\boldsymbol{r}_l)\Im(\boldsymbol{r}_i)^T + \Im(\boldsymbol{R})\Im(R_{il})].$$

Therefore the partial sum of the cases when $s = t$ is given by

$$\frac{1}{N}\Re(\boldsymbol{r}_i)\Re(\boldsymbol{r}_l)^T + \frac{1}{2N}[\Re(\boldsymbol{R})\Re(R_{il}) + \Im(\boldsymbol{R})\Im(R_{il}) + \Re(\boldsymbol{r}_l)\Re(\boldsymbol{r}_i)^T - \Im(\boldsymbol{r}_l)\Im(\boldsymbol{r}_i)^T].$$

Combined with the previous partial sum of the cases when $s \neq t$, we obtain that

$$\mathbb{E}[\Re(\Delta\boldsymbol{r}_i)\Re(\Delta\boldsymbol{r}_l)^T] = \frac{1}{2N}[\Re(\boldsymbol{R})\Re(R_{il}) + \Im(\boldsymbol{R})\Im(R_{il}) + \Re(\boldsymbol{r}_l)\Re(\boldsymbol{r}_i)^T - \Im(\boldsymbol{r}_l)\Im(\boldsymbol{r}_i)^T]. \quad \text{(B.5)}$$

Therefore

$$\mathbb{E}[\Re(\Delta\boldsymbol{r})\Re(\Delta\boldsymbol{r})^T] = \frac{1}{2N}[\Re(\boldsymbol{R}) \otimes \Re(\boldsymbol{R}) + \Im(\boldsymbol{R}) \otimes \Im(\boldsymbol{R}) + \boldsymbol{C}_{\Re(\boldsymbol{R})\Re(\boldsymbol{R})} - \boldsymbol{C}_{\Im(\boldsymbol{R})\Im(\boldsymbol{R})}], \quad \text{(B.6)}$$

which completes the computation of first expectation in (B.1). Utilizing the same technique, we obtain that

$$\mathbb{E}[\Im(\Delta\boldsymbol{r})\Im(\Delta\boldsymbol{r})^T] = \frac{1}{2N}[\Re(\boldsymbol{R}) \otimes \Re(\boldsymbol{R}) + \Im(\boldsymbol{R}) \otimes \Im(\boldsymbol{R}) + \boldsymbol{C}_{\Im(\boldsymbol{R})\Im(\boldsymbol{R})} - \boldsymbol{C}_{\Re(\boldsymbol{R})\Re(\boldsymbol{R})}], \quad \text{(B.7)}$$

and

$$\mathbb{E}[\Re(\Delta\boldsymbol{r})\Im(\Delta\boldsymbol{r})^T] = \frac{1}{2N}[\Im(\boldsymbol{R}) \otimes \Re(\boldsymbol{R}) - \Re(\boldsymbol{R}) \otimes \Im(\boldsymbol{R}) + \boldsymbol{C}_{\Re(\boldsymbol{R})\Im(\boldsymbol{R})} + \boldsymbol{C}_{\Im(\boldsymbol{R})\Re(\boldsymbol{R})}]. \quad \text{(B.8)}$$

Substituting (B.6)–(B.8) into (B.1) gives a closed-form MSE expression. However, this expression is too complicated for analytical study. In the following steps, we make use of the properties of $\boldsymbol{\xi}_k$ to simply the MSE expression.

Lemma B.1. *Let* $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{N\times N}$ *satisfying* $\boldsymbol{X}^T = (-1)^{n_x}\boldsymbol{X}$, $\boldsymbol{A}^T = (-1)^{n_a}\boldsymbol{A}$, *and* $\boldsymbol{B}^T = (-1)^{n_b}\boldsymbol{B}$, *where* $n_x, n_a, n_b \in \{0, 1\}$. *Then*

$$\begin{aligned}
\operatorname{vec}(\boldsymbol{X})^T(\boldsymbol{A}\otimes\boldsymbol{B})\operatorname{vec}(\boldsymbol{Y}) &= (-1)^{n_x+n_b}\operatorname{vec}(\boldsymbol{X})^T\boldsymbol{C}_{\boldsymbol{AB}}\operatorname{vec}(\boldsymbol{Y}),\\
\operatorname{vec}(\boldsymbol{X})^T(\boldsymbol{B}\otimes\boldsymbol{A})\operatorname{vec}(\boldsymbol{Y}) &= (-1)^{n_x+n_a}\operatorname{vec}(\boldsymbol{X})^T\boldsymbol{C}_{\boldsymbol{BA}}\operatorname{vec}(\boldsymbol{Y}).
\end{aligned}$$

Proof. By Definition B.1,

$$\begin{aligned}
&\operatorname{vec}(\boldsymbol{X})^T\boldsymbol{C}_{\boldsymbol{AB}}\operatorname{vec}(\boldsymbol{Y})\\
=&\sum_{m=1}^{N}\sum_{n=1}^{N}\boldsymbol{x}_m^T\boldsymbol{a}_n\boldsymbol{b}_m^T\boldsymbol{y}_n\\
=&\sum_{m=1}^{N}\sum_{n=1}^{N}\Big(\sum_{p=1}^{N}A_{pn}X_{pm}\Big)\Big(\sum_{p=1}^{N}B_{qm}Y_{qn}\Big)\\
=&\sum_{m=1}^{N}\sum_{n=1}^{N}\sum_{p=1}^{N}\sum_{q=1}^{N}A_{pn}X_{pm}B_{qm}Y_{qn}\\
=&(-1)^{n_x+n_b}\sum_{p=1}^{N}\sum_{n=1}^{N}\sum_{m=1}^{N}\sum_{q=1}^{N}(X_{mp}B_{mq}Y_{qn})A_{pn}\\
=&(-1)^{n_x+n_b}\sum_{p=1}^{N}\sum_{n=1}^{N}\boldsymbol{x}_p^TA_{pn}\boldsymbol{B}\boldsymbol{y}_n\\
=&(-1)^{n_x+n_b}\operatorname{vec}(\boldsymbol{X})^T(\boldsymbol{A}\otimes\boldsymbol{B})\operatorname{vec}(\boldsymbol{Y}).
\end{aligned}$$

The proof of the second equality follows the same idea. □

Lemma B.2. $\boldsymbol{T}_{M_{\text{co}}}\boldsymbol{\Pi}^{\perp}_{\boldsymbol{A}_{\text{co}}}\boldsymbol{T}_{M_{\text{co}}} = (\boldsymbol{\Pi}^{\perp}_{\boldsymbol{A}_{\text{co}}})^*$.

Proof. Since $\boldsymbol{\Pi}^{\perp}_{\boldsymbol{A}_{\text{co}}} = \boldsymbol{I} - \boldsymbol{A}_{\text{co}}(\boldsymbol{A}^H_{\text{co}}\boldsymbol{A}_{\text{co}})^{-1}\boldsymbol{A}^H_{\text{co}}$, it suffices to show that $\boldsymbol{T}_{M_{\text{co}}}\boldsymbol{A}_{\text{co}}(\boldsymbol{A}^H_{\text{co}}\boldsymbol{A}_{\text{co}})^{-1}\boldsymbol{A}^H_{\text{co}}\boldsymbol{T}_{M_{\text{co}}} = (\boldsymbol{A}_{\text{co}}(\boldsymbol{A}^H_{\text{co}}\boldsymbol{A}_{\text{co}})^{-1}\boldsymbol{A}^H_{\text{co}})^*$. Because $\boldsymbol{A}_{\text{co}}$ is the steering matrix of a ULA with M_{co} sensors, it is straightforward to show that $\boldsymbol{T}_{M_{\text{co}}}\boldsymbol{A}_{\text{co}} = (\boldsymbol{A}_{\text{co}}\boldsymbol{\Phi})^*$, where

$$\boldsymbol{\Phi} = \text{diag}(e^{-j(M_{\text{co}}-1)\omega_1}, e^{-j(M_{\text{co}}-1)\omega_2}, \ldots, e^{-j(M_{\text{co}}-1)\omega_K}).$$

Because $\boldsymbol{T}_{M_{\text{co}}}\boldsymbol{T}_{M_{\text{co}}} = \boldsymbol{I}, \boldsymbol{T}^H_{M_{\text{co}}} = \boldsymbol{T}_{M_{\text{co}}}$,

$$\begin{aligned}
&\boldsymbol{T}_{M_{\text{co}}}\boldsymbol{A}_{\text{co}}(\boldsymbol{A}^H_{\text{co}}\boldsymbol{A}_{\text{co}})^{-1}\boldsymbol{A}^H_{\text{co}}\boldsymbol{T}_{M_{\text{co}}} \\
=&\boldsymbol{T}_{M_{\text{co}}}\boldsymbol{A}_{\text{co}}(\boldsymbol{A}^H_{\text{co}}\boldsymbol{T}^H_{M_{\text{co}}}\boldsymbol{T}_{M_{\text{co}}}\boldsymbol{A}_{\text{co}})^{-1}\boldsymbol{A}^H_{\text{co}}\boldsymbol{T}^H_{M_{\text{co}}} \\
=&(\boldsymbol{A}_{\text{co}}\boldsymbol{\Phi})^*((\boldsymbol{A}_{\text{co}}\boldsymbol{\Phi})^T(\boldsymbol{A}_{\text{co}}\boldsymbol{\Phi})^*)^{-1}(\boldsymbol{A}_{\text{co}}\boldsymbol{\Phi})^T \\
=&(\boldsymbol{A}_{\text{co}}(\boldsymbol{A}^H_{\text{co}}\boldsymbol{A}_{\text{co}})^{-1}\boldsymbol{A}^H_{\text{co}})^*.
\end{aligned}$$

□

Lemma B.3. *Let* $\boldsymbol{\Xi}_k = \text{mat}_{M,M}(\boldsymbol{\xi}_k)$. *Then* $\boldsymbol{\Xi}^H_k = \boldsymbol{\Xi}_k$ *for* $k = 1, 2, \ldots, K$.

Proof. Note that $\boldsymbol{\xi}_k = \boldsymbol{F}^T\boldsymbol{\Gamma}^T(\boldsymbol{\beta}_k \otimes \boldsymbol{\alpha}_k)$. We first prove that $\boldsymbol{\beta}_k \otimes \boldsymbol{\alpha}_k$ is conjugate symmetric, or that $(\boldsymbol{T}_{M_{\text{co}}} \otimes \boldsymbol{T}_{M_{\text{co}}})(\boldsymbol{\beta}_k \otimes \boldsymbol{\alpha}_k) = (\boldsymbol{\beta}_k \otimes \boldsymbol{\alpha}_k)^*$. Similar to the proof of Lemma B.2, we utilize the properties that $\boldsymbol{T}_{M_{\text{co}}}\boldsymbol{A}_{\text{co}} = (\boldsymbol{A}_{\text{co}}\boldsymbol{\Phi})^*$ and that $\boldsymbol{T}_{M_{\text{co}}}\boldsymbol{a}_{\text{co}}(\theta_k) = (\boldsymbol{a}_{\text{co}}(\theta_k)e^{-j(M_{\text{co}}-1)\omega_k})^*$ to show that

$$\boldsymbol{T}_{M_{\text{co}}}(\boldsymbol{A}^{\dagger}_{\text{co}})^H\boldsymbol{e}_k\boldsymbol{a}^H_{\text{co}}(\theta_k)\boldsymbol{T}_{M_{\text{co}}} = [(\boldsymbol{A}^{\dagger}_{\text{co}})^H\boldsymbol{e}_k\boldsymbol{a}^H_{\text{co}}(\theta_k)]^*. \tag{B.9}$$

Observe that $\dot{\boldsymbol{a}}_{\mathrm{co}}(\theta_k) = j\dot{\omega}_k \boldsymbol{D}\boldsymbol{a}_{\mathrm{co}}(\theta_k)$, where $\dot{\omega}_k = (2\pi d_0 \cos\theta_k)/\lambda$ and $\boldsymbol{D} = \mathrm{diag}(0, 1, \ldots, M_{\mathrm{co}}-1)$. We have

$$\begin{aligned}
&(\boldsymbol{T}_{M_{\mathrm{co}}} \otimes \boldsymbol{T}_{M_{\mathrm{co}}})(\boldsymbol{\beta}_k \otimes \boldsymbol{\alpha}_k) = (\boldsymbol{\beta}_k \otimes \boldsymbol{\alpha}_k)^* \\
\Longleftrightarrow\ &\boldsymbol{T}_{M_{\mathrm{co}}} \boldsymbol{\alpha}_k \boldsymbol{\beta}_k^T \boldsymbol{T}_{M_{\mathrm{co}}} = (\boldsymbol{\alpha}_k \boldsymbol{\beta}_k^T)^* \\
\Longleftrightarrow\ &\boldsymbol{T}_{M_{\mathrm{co}}}[(\boldsymbol{A}_{\mathrm{co}}^\dagger)^H \boldsymbol{e}_k \boldsymbol{a}_{\mathrm{co}}^H(\theta_k) \boldsymbol{D}\boldsymbol{\Pi}_{\boldsymbol{A}_{\mathrm{co}}}^\perp]^* \boldsymbol{T}_{M_{\mathrm{co}}} = -(\boldsymbol{A}_{\mathrm{co}}^\dagger)^H \boldsymbol{e}_k \boldsymbol{a}_{\mathrm{co}}^H(\theta_k) \boldsymbol{D}\boldsymbol{\Pi}_{\boldsymbol{A}_{\mathrm{co}}}^\perp.
\end{aligned}$$

Since $\boldsymbol{D} = \boldsymbol{T}_{M_{\mathrm{co}}} \boldsymbol{T}_{M_{\mathrm{co}}} \boldsymbol{D} \boldsymbol{T}_{M_{\mathrm{co}}} \boldsymbol{T}_{M_{\mathrm{co}}}$, combining with Lemma B.2 and (B.9), it suffices to show that

$$(\boldsymbol{A}_{\mathrm{co}}^\dagger)^H \boldsymbol{e}_k \boldsymbol{a}_{\mathrm{co}}^H(\theta_k) \boldsymbol{T}_{M_{\mathrm{co}}} \boldsymbol{D} \boldsymbol{T}_{M_{\mathrm{co}}} \boldsymbol{\Pi}_{\boldsymbol{A}_{\mathrm{co}}}^\perp = -(\boldsymbol{A}_{\mathrm{co}}^\dagger)^H \boldsymbol{e}_k \boldsymbol{a}_{\mathrm{co}}^H(\theta_k) \boldsymbol{D}\boldsymbol{\Pi}_{\boldsymbol{A}_{\mathrm{co}}}^\perp. \tag{B.10}$$

Observe that $\boldsymbol{T}_{M_{\mathrm{co}}} \boldsymbol{D} \boldsymbol{T}_{M_{\mathrm{co}}} + \boldsymbol{D} = (M_{\mathrm{co}} - 1)\boldsymbol{I}$. We have

$$\boldsymbol{\Pi}_{\boldsymbol{A}_{\mathrm{co}}}^\perp (\boldsymbol{T}_{M_{\mathrm{co}}} \boldsymbol{D} \boldsymbol{T}_{M_{\mathrm{co}}} + \boldsymbol{D}) \boldsymbol{a}_{\mathrm{co}}(\theta_k) = \boldsymbol{0},$$

or equivalently

$$\boldsymbol{a}_{\mathrm{co}}^H(\theta_k) \boldsymbol{T}_{M_{\mathrm{co}}} \boldsymbol{D} \boldsymbol{T}_{M_{\mathrm{co}}} \boldsymbol{\Pi}_{\boldsymbol{A}_{\mathrm{co}}}^\perp = -\boldsymbol{a}_{\mathrm{co}}^H(\theta_k) \boldsymbol{D}\boldsymbol{\Pi}_{\boldsymbol{A}_{\mathrm{co}}}^\perp. \tag{B.11}$$

Pre-multiplying both sides of (B.11) with $(\boldsymbol{A}_{\mathrm{co}}^\dagger)^H \boldsymbol{e}_k$ leads to (B.10), which completes the proof that $\boldsymbol{\beta}_k \otimes \boldsymbol{\alpha}_k$ is conjugate symmetric. According to the definition of $\boldsymbol{\Gamma}$ in (3.2e), it is straightforward to show that $\boldsymbol{\Gamma}^T(\boldsymbol{\beta}_k \otimes \boldsymbol{\alpha}_k)$ is also conjugate symmetric. Combined with Property 2.3, we conclude that $\mathrm{mat}_{M,M}(\boldsymbol{F}^T \boldsymbol{\Gamma}^T (\boldsymbol{\beta}_k \otimes \boldsymbol{\alpha}_k))$ is Hermitian symmetric, or that $\boldsymbol{\Xi}_k = \boldsymbol{\Xi}_k^H$. □

Given Lemma B.1–B.3, we are able continue the simplification. We first consider the term $\Re(\boldsymbol{\xi}_{k_1})^T \mathbb{E}[\Re(\Delta\boldsymbol{r})\Re(\Delta\boldsymbol{r})^T]\Re(\boldsymbol{\xi}_{k_2})$ in (B.1). Let $\boldsymbol{\Xi}_{k_1} = \mathrm{mat}_{M,M}(\boldsymbol{\xi}_{k_1})$, and $\boldsymbol{\Xi}_{k_2} = \mathrm{mat}_{M,M}(\boldsymbol{\xi}_{k_2})$. By Lemma B.3, we have $\boldsymbol{\Xi}_{k_1} = \boldsymbol{\Xi}_{k_1}^H$, and $\boldsymbol{\Xi}_{k_2} = \boldsymbol{\Xi}_{k_2}^H$. Observe that $\Re(\boldsymbol{R})^T = \Re(\boldsymbol{R})$, and that

$\Im(\boldsymbol{R})^T = \Im(\boldsymbol{R})$. By Lemma B.1 we immediately obtain the following equalities:

$$\begin{aligned}
\Re(\boldsymbol{\xi}_{k_1})^T(\Re(\boldsymbol{R}) \otimes \Re(\boldsymbol{R}))\Re(\boldsymbol{\xi}_{k_2}) &= \Re(\boldsymbol{\xi}_{k_1})^T \boldsymbol{C}_{\Re(\boldsymbol{R})\Re(\boldsymbol{R})}\Re(\boldsymbol{\xi}_{k_2}), \\
\Re(\boldsymbol{\xi}_{k_1})^T(\Im(\boldsymbol{R}) \otimes \Im(\boldsymbol{R}))\Re(\boldsymbol{\xi}_{k_2}) &= -\Re(\boldsymbol{\xi}_{k_1})^T \boldsymbol{C}_{\Im(\boldsymbol{R})\Im(\boldsymbol{R})}\Re(\boldsymbol{\xi}_{k_2}).
\end{aligned}$$

Therefore $\Re(\boldsymbol{\xi}_{k_1})^T\mathbb{E}[\Re(\Delta\boldsymbol{r})\Re(\Delta\boldsymbol{r})^T]\Re(\boldsymbol{\xi}_{k_2})$ can be compactly expressed as

$$\begin{aligned}
&\Re(\boldsymbol{\xi}_{k_1})^T\mathbb{E}[\Re(\Delta\boldsymbol{r})\Re(\Delta\boldsymbol{r})^T]\Re(\boldsymbol{\xi}_{k_2}) \\
=&\frac{1}{N}\Re(\boldsymbol{\xi}_{k_1})^T[\Re(\boldsymbol{R}) \otimes \Re(\boldsymbol{R}) + \Im(\boldsymbol{R}) \otimes \Im(\boldsymbol{R})]\Re(\boldsymbol{\xi}_{k_2}) \\
=&\frac{1}{N}\Re(\boldsymbol{\xi}_{k_1})^T\Re(\boldsymbol{R}^T \otimes \boldsymbol{R})\Re(\boldsymbol{\xi}_{k_2}),
\end{aligned} \tag{B.12}$$

where we make use of the properties that $\boldsymbol{R}^T = \boldsymbol{R}^*$, and $\Re(\boldsymbol{R}^* \otimes \boldsymbol{R}) = \Re(\boldsymbol{R}) \otimes \Re(\boldsymbol{R}) + \Im(\boldsymbol{R}) \otimes \Im(\boldsymbol{R})$. Similarly, we can obtain that

$$\Im(\boldsymbol{\xi}_{k_1})^T\mathbb{E}[\Im(\Delta\boldsymbol{r})\Im(\Delta\boldsymbol{r})^T]\Im(\boldsymbol{\xi}_{k_2}) = \frac{1}{N}\Im(\boldsymbol{\xi}_{k_1})^T\Re(\boldsymbol{R}^T \otimes \boldsymbol{R})\Im(\boldsymbol{\xi}_{k_2}), \tag{B.13}$$

$$\Re(\boldsymbol{\xi}_{k_1})^T\mathbb{E}[\Re(\Delta\boldsymbol{r})\Im(\Delta\boldsymbol{r})^T]\Im(\boldsymbol{\xi}_{k_2}) = -\frac{1}{N}\Re(\boldsymbol{\xi}_{k_1})^T\Im(\boldsymbol{R}^T \otimes \boldsymbol{R})\Im(\boldsymbol{\xi}_{k_2}), \tag{B.14}$$

$$\Re(\boldsymbol{\xi}_{k_2})^T\mathbb{E}[\Re(\Delta\boldsymbol{r})\Im(\Delta\boldsymbol{r})^T]\Im(\boldsymbol{\xi}_{k_1}) = -\frac{1}{N}\Re(\boldsymbol{\xi}_{k_2})^T\Im(\boldsymbol{R}^T \otimes \boldsymbol{R})\Im(\boldsymbol{\xi}_{k_1}). \tag{B.15}$$

Substituting (B.12)–(B.15) into (B.1) completes the proof.

Appendix C

Proof of Proposition 3.2

Without loss of generality, let $p = 1$ and $\sigma^2 \to 0$. For brevity, we denote $\boldsymbol{R}^T \otimes \boldsymbol{R}$ by $\boldsymbol{W}$. We first consider the case when $K < M$. Denote the eigendecomposition of $\boldsymbol{R}^{-1}$ by $\boldsymbol{E}_{\mathrm{s}}\boldsymbol{\Lambda}_{\mathrm{s}}^{-1}\boldsymbol{E}_{\mathrm{s}}^H + \sigma^{-2}\boldsymbol{E}_{\mathrm{n}}\boldsymbol{E}_{\mathrm{n}}^H$. We have

$$\boldsymbol{W}^{-1} = \sigma^{-4}\boldsymbol{K}_1 + \sigma^{-2}\boldsymbol{K}_2 + \boldsymbol{K}_3,$$

where

$$\begin{aligned}
\boldsymbol{K}_1 &= \boldsymbol{E}_{\mathrm{n}}^*\boldsymbol{E}_{\mathrm{n}}^T \otimes \boldsymbol{E}_{\mathrm{n}}\boldsymbol{E}_{\mathrm{n}}^H, \\
\boldsymbol{K}_2 &= \boldsymbol{E}_{\mathrm{s}}^*\boldsymbol{\Lambda}_{\mathrm{s}}^{-1}\boldsymbol{E}_{\mathrm{s}}^T \otimes \boldsymbol{E}_{\mathrm{n}}\boldsymbol{E}_{\mathrm{n}}^H + \boldsymbol{E}_{\mathrm{n}}^*\boldsymbol{E}_{\mathrm{n}}^T \otimes \boldsymbol{E}_{\mathrm{s}}\boldsymbol{\Lambda}_{\mathrm{s}}^{-1}\boldsymbol{E}_{\mathrm{s}}^H, \\
\boldsymbol{K}_3 &= \boldsymbol{E}_{\mathrm{s}}^*\boldsymbol{\Lambda}_{\mathrm{s}}^{-1}\boldsymbol{E}_{\mathrm{s}}^T \otimes \boldsymbol{E}_{\mathrm{s}}\boldsymbol{\Lambda}_{\mathrm{s}}^{-1}\boldsymbol{E}_{\mathrm{s}}^H.
\end{aligned}$$

Recall that $\boldsymbol{A}^H\boldsymbol{E}_{\mathrm{n}} = \boldsymbol{0}$. We have

$$\begin{aligned}
\boldsymbol{K}_1\dot{\boldsymbol{A}}_{\mathrm{d}} &= (\boldsymbol{E}_{\mathrm{n}}^*\boldsymbol{E}_{\mathrm{n}}^T \otimes \boldsymbol{E}_{\mathrm{n}}\boldsymbol{E}_{\mathrm{n}}^H)(\dot{\boldsymbol{A}}^* \odot \boldsymbol{A} + \boldsymbol{A}^* \odot \dot{\boldsymbol{A}}) \\
&= \boldsymbol{E}_{\mathrm{n}}^*\boldsymbol{E}_{\mathrm{n}}^T\dot{\boldsymbol{A}}^* \odot \boldsymbol{E}_{\mathrm{n}}\boldsymbol{E}_{\mathrm{n}}^H\boldsymbol{A} + \boldsymbol{E}_{\mathrm{n}}^*\boldsymbol{E}_{\mathrm{n}}^T\boldsymbol{A}^* \odot \boldsymbol{E}_{\mathrm{n}}\boldsymbol{E}_{\mathrm{n}}^H\dot{\boldsymbol{A}} \\
&= \boldsymbol{0}. \qquad \text{(C.1)}
\end{aligned}$$

Therefore

$$\boldsymbol{M}_{\boldsymbol{\theta}}^H \boldsymbol{M}_{\boldsymbol{\theta}} = \dot{\boldsymbol{A}}_{\mathrm{d}}^H \boldsymbol{W}^{-1} \dot{\boldsymbol{A}}_{\mathrm{d}} = \sigma^{-2} \dot{\boldsymbol{A}}_{\mathrm{d}}^H (\boldsymbol{K}_2 + \sigma^2 \boldsymbol{K}_3) \dot{\boldsymbol{A}}_{\mathrm{d}}. \tag{C.2}$$

Similar to $\boldsymbol{W}^{-1}$, we denote $\boldsymbol{W}^{-\frac{1}{2}} = \sigma^{-2}\boldsymbol{K}_1 + \sigma^{-1}\boldsymbol{K}_4 + \boldsymbol{K}_5$, where

$$\begin{aligned}
\boldsymbol{K}_4 &= \boldsymbol{E}_{\mathrm{s}}^* \boldsymbol{\Lambda}_{\mathrm{s}}^{-\frac{1}{2}} \boldsymbol{E}_{\mathrm{s}}^T \otimes \boldsymbol{E}_{\mathrm{n}} \boldsymbol{E}_{\mathrm{n}}^H + \boldsymbol{E}_{\mathrm{n}}^* \boldsymbol{E}_{\mathrm{n}}^T \otimes \boldsymbol{E}_{\mathrm{s}} \boldsymbol{\Lambda}_{\mathrm{s}}^{-\frac{1}{2}} \boldsymbol{E}_{\mathrm{s}}^H, \\
\boldsymbol{K}_5 &= \boldsymbol{E}_{\mathrm{s}}^* \boldsymbol{\Lambda}_{\mathrm{s}}^{-\frac{1}{2}} \boldsymbol{E}_{\mathrm{s}}^T \otimes \boldsymbol{E}_{\mathrm{s}} \boldsymbol{\Lambda}_{\mathrm{s}}^{-\frac{1}{2}} \boldsymbol{E}_{\mathrm{s}}^H.
\end{aligned}$$

Therefore

$$\boldsymbol{M}_{\boldsymbol{\theta}}^H \boldsymbol{\Pi}_{\boldsymbol{M_s}}^{\perp} \boldsymbol{M}_{\boldsymbol{\theta}} = \dot{\boldsymbol{A}}_{\mathrm{d}}^H \boldsymbol{W}^{-\frac{1}{2}} \boldsymbol{\Pi}_{\boldsymbol{M_s}}^{\perp} \boldsymbol{W}^{-\frac{1}{2}} \dot{\boldsymbol{A}}_{\mathrm{d}} = \sigma^{-2} \dot{\boldsymbol{A}}_{\mathrm{d}}^H (\sigma \boldsymbol{K}_5 + \boldsymbol{K}_4) \boldsymbol{\Pi}_{\boldsymbol{M_s}}^{\perp} (\sigma \boldsymbol{K}_5 + \boldsymbol{K}_4) \dot{\boldsymbol{A}}_{\mathrm{d}},$$

where $\boldsymbol{\Pi}_{\boldsymbol{M_s}}^{\perp} = \boldsymbol{M_s} \boldsymbol{M_s}^{\dagger}$. We can then express the CRB as

$$\mathrm{CRB}_{\boldsymbol{\theta}} = \sigma^2 (\boldsymbol{Q}_1 + \sigma \boldsymbol{Q}_2 + \sigma^2 \boldsymbol{Q}_3)^{-1}, \tag{C.3}$$

where

$$\begin{aligned}
\boldsymbol{Q}_1 &= \dot{\boldsymbol{A}}_{\mathrm{d}}^H (\boldsymbol{K}_2 - \boldsymbol{K}_4 \boldsymbol{\Pi}_{\boldsymbol{M_s}}^{\perp} \boldsymbol{K}_4) \dot{\boldsymbol{A}}_{\mathrm{d}}, \\
\boldsymbol{Q}_2 &= -\dot{\boldsymbol{A}}_{\mathrm{d}}^H (\boldsymbol{K}_4 \boldsymbol{\Pi}_{\boldsymbol{M_s}}^{\perp} \boldsymbol{K}_5 + \boldsymbol{K}_5 \boldsymbol{\Pi}_{\boldsymbol{M_s}}^{\perp} \boldsymbol{K}_4) \dot{\boldsymbol{A}}_{\mathrm{d}}, \\
\boldsymbol{Q}_3 &= \dot{\boldsymbol{A}}_{\mathrm{d}}^H (\boldsymbol{K}_3 - \boldsymbol{K}_5 \boldsymbol{\Pi}_{\boldsymbol{M_s}}^{\perp} \boldsymbol{K}_5) \dot{\boldsymbol{A}}_{\mathrm{d}}.
\end{aligned}$$

When $\sigma^2 = 0$, $\boldsymbol{R}$ reduces to $\boldsymbol{A}\boldsymbol{A}^H$. Observe that the eigendecomposition of $\boldsymbol{R}$ always exists for $\sigma^2 \geq 0$. We use $\boldsymbol{K}_1^\star$–$\boldsymbol{K}_5^\star$ to denote the corresponding $\boldsymbol{K}_1$–$\boldsymbol{K}_5$ when $\sigma^2 \to 0$.

Lemma C.1. *Let $K < M$. Assume $\partial \boldsymbol{r} / \partial \boldsymbol{\eta}$ is full column rank. Then $\lim_{\sigma^2 \to 0^+} \boldsymbol{\Pi}_{\boldsymbol{M_s}}^{\perp}$ exists.*

Proof. Because $\boldsymbol{A}^H \boldsymbol{E}_{\mathrm{n}} = \boldsymbol{0}$,

$$\begin{aligned}\boldsymbol{K}_2 \boldsymbol{A}_{\mathrm{d}} =&(\boldsymbol{E}_{\mathrm{s}}^* \boldsymbol{\Lambda}_{\mathrm{s}}^{-1} \boldsymbol{E}_{\mathrm{s}}^T \otimes \boldsymbol{E}_{\mathrm{n}} \boldsymbol{E}_{\mathrm{n}}^H)(\boldsymbol{A}^* \odot \boldsymbol{A}) + (\boldsymbol{E}_{\mathrm{n}}^* \boldsymbol{E}_{\mathrm{n}}^T \otimes \boldsymbol{E}_{\mathrm{s}} \boldsymbol{\Lambda}_{\mathrm{s}}^{-1} \boldsymbol{E}_{\mathrm{s}}^H)(\boldsymbol{A}^* \odot \boldsymbol{A}) \\ =&\boldsymbol{E}_{\mathrm{s}}^* \boldsymbol{\Lambda}_{\mathrm{s}}^{-1} \boldsymbol{E}_{\mathrm{s}}^T \boldsymbol{A}^* \odot \boldsymbol{E}_{\mathrm{n}} \boldsymbol{E}_{\mathrm{n}}^H \boldsymbol{A} + \boldsymbol{E}_{\mathrm{n}}^* \boldsymbol{E}_{\mathrm{n}}^T \boldsymbol{A}^* \odot \boldsymbol{E}_{\mathrm{s}} \boldsymbol{\Lambda}_{\mathrm{s}}^{-1} \boldsymbol{E}_{\mathrm{s}}^H \boldsymbol{A} \\ =&\boldsymbol{0}\end{aligned}$$

Similarly, we can show that $\boldsymbol{K}_4 \boldsymbol{A}_{\mathrm{d}} = \boldsymbol{0}$, $\boldsymbol{i}^H \boldsymbol{K}_2 \boldsymbol{i} = \boldsymbol{i}^H \boldsymbol{K}_4 \boldsymbol{i} = 0$, and $\boldsymbol{i}^H \boldsymbol{K}_1 \boldsymbol{i} = \mathrm{rank}(\boldsymbol{E}_{\mathrm{n}}) = M - K$. Hence

$$\boldsymbol{M}_{\boldsymbol{s}}^H \boldsymbol{M}_{\boldsymbol{s}} = \begin{bmatrix} \boldsymbol{A}_{\mathrm{d}}^H \boldsymbol{K}_3 \boldsymbol{A}_{\mathrm{d}} & \boldsymbol{A}_{\mathrm{d}}^H \boldsymbol{K}_3 \boldsymbol{i} \\ \boldsymbol{i}^H \boldsymbol{K}_3 \boldsymbol{A}_{\mathrm{d}} & \boldsymbol{i}^H \boldsymbol{W}^{-1} \boldsymbol{i} \end{bmatrix}.$$

Because $\partial \boldsymbol{r} / \partial \boldsymbol{\eta}$ is full column rank, $\boldsymbol{M}_{\boldsymbol{s}}^H \boldsymbol{M}_{\boldsymbol{s}}$ is full rank and positive definite. Therefore the Schur complements exist, and we can inverse $\boldsymbol{M}_{\boldsymbol{s}}^H \boldsymbol{M}_{\boldsymbol{s}}$ block-wisely. Let $\boldsymbol{V} = \boldsymbol{A}_{\mathrm{d}}^H \boldsymbol{K}_3 \boldsymbol{A}_{\mathrm{d}}$ and $v = \boldsymbol{i}^H \boldsymbol{W}^{-1} \boldsymbol{i}$. After tedious but straightforward computation, we obtain

$$\begin{aligned}\boldsymbol{\Pi}_{\boldsymbol{M}_{\boldsymbol{s}}}^{\perp} =&\boldsymbol{K}_5 \boldsymbol{A}_{\mathrm{d}} \boldsymbol{S}^{-1} \boldsymbol{A}_{\mathrm{d}}^H \boldsymbol{K}_5 \\ &- s^{-1} \boldsymbol{K}_5 \boldsymbol{A}_{\mathrm{d}} \boldsymbol{V}^{-1} \boldsymbol{A}_{\mathrm{d}}^H \boldsymbol{K}_3 \boldsymbol{i} \boldsymbol{i}^H (\boldsymbol{K}_5 + \sigma^{-2} \boldsymbol{K}_1) \\ &- v^{-1} (\boldsymbol{K}_5 + \sigma^{-2} \boldsymbol{K}_1) \boldsymbol{i} \boldsymbol{i}^H \boldsymbol{K}_3 \boldsymbol{A}_{\mathrm{d}} \boldsymbol{S}^{-1} \boldsymbol{A}_{\mathrm{d}}^H \boldsymbol{K}_5 \\ &+ s^{-1} (\boldsymbol{K}_5 + \sigma^{-2} \boldsymbol{K}_1) \boldsymbol{i} \boldsymbol{i}^H (\boldsymbol{K}_5 + \sigma^{-2} \boldsymbol{K}_1),\end{aligned}$$

where $\boldsymbol{S}$ and s are Schur complements given by

$$\begin{aligned}\boldsymbol{S} &= \boldsymbol{V} - v^{-1} \boldsymbol{A}_{\mathrm{d}}^H \boldsymbol{K}_3 \boldsymbol{i} \boldsymbol{i}^H \boldsymbol{K}_3 \boldsymbol{A}_{\mathrm{d}}, \\ s &= v - \boldsymbol{i}^H \boldsymbol{K}_3 \boldsymbol{A}_{\mathrm{d}} \boldsymbol{V}^{-1} \boldsymbol{A}_{\mathrm{d}}^H \boldsymbol{K}_5 \boldsymbol{i}.\end{aligned}$$

Observe that

$$v = \boldsymbol{i}^H \boldsymbol{W}^{-1} \boldsymbol{i} = \sigma^{-4}(M-K) + \boldsymbol{i}^H \boldsymbol{K}_3 \boldsymbol{i}.$$

We know that both v^{-1} and s^{-1} decrease at the rate of σ^4. As $\sigma^2 \to 0$, we have

$$\begin{aligned}
&\boldsymbol{S} \to \boldsymbol{A}_{\mathrm{d}}^H \boldsymbol{K}_3^\star \boldsymbol{A}_{\mathrm{d}}, \\
&s^{-1}(\boldsymbol{K}_5 + \sigma^{-2}\boldsymbol{K}_1) \to \boldsymbol{0}, \\
&v^{-1}(\boldsymbol{K}_5 + \sigma^{-2}\boldsymbol{K}_1) \to \boldsymbol{0}, \\
&s^{-1}(\boldsymbol{K}_5 + \sigma^{-2}\boldsymbol{K}_1)\boldsymbol{i}\boldsymbol{i}^H(\boldsymbol{K}_5 + \sigma^{-2}\boldsymbol{K}_1) \to \frac{\boldsymbol{K}_1^\star \boldsymbol{i}\boldsymbol{i}^H \boldsymbol{K}_1^\star}{M-K}.
\end{aligned}$$

We now show that $\boldsymbol{A}_{\mathrm{d}}^H \boldsymbol{K}_3^\star \boldsymbol{A}_{\mathrm{d}}$ is nonsingular. Denote the eigendecomposition of $\boldsymbol{A}\boldsymbol{A}^H$ by $\boldsymbol{E}_{\mathrm{s}}^\star \boldsymbol{\Lambda}_{\mathrm{s}}^\star (\boldsymbol{E}_{\mathrm{s}}^\star)^H$. Recall that for matrices with proper dimensions, $(\boldsymbol{A} \odot \boldsymbol{B})^H (\boldsymbol{C} \odot \boldsymbol{D}) = (\boldsymbol{A}^H \boldsymbol{C}) \circ (\boldsymbol{B}^H \boldsymbol{D})$, where $\circ$ denotes the Hadamard product. We can expand $\boldsymbol{A}_{\mathrm{d}}^H \boldsymbol{K}_3^\star \boldsymbol{A}_{\mathrm{d}}$ into

$$[\boldsymbol{A}^H \boldsymbol{E}_{\mathrm{s}}^\star (\boldsymbol{\Lambda}_{\mathrm{s}}^\star)^{-1} (\boldsymbol{E}_{\mathrm{s}}^\star)^H \boldsymbol{A}]^* \circ [\boldsymbol{A}^H \boldsymbol{E}_{\mathrm{s}}^\star (\boldsymbol{\Lambda}_{\mathrm{s}}^\star)^{-1} (\boldsymbol{E}_{\mathrm{s}}^\star)^H \boldsymbol{A}].$$

Note that $\boldsymbol{A}\boldsymbol{A}^H \boldsymbol{E}_{\mathrm{s}}^\star (\boldsymbol{\Lambda}_{\mathrm{s}}^\star)^{-1} (\boldsymbol{E}_{\mathrm{s}}^\star)^H \boldsymbol{A} = \boldsymbol{E}_{\mathrm{s}}^\star (\boldsymbol{E}_{\mathrm{s}}^\star)^H \boldsymbol{A} = \boldsymbol{A}$, and that $\boldsymbol{A}$ is full column rank when $K < M$. We thus have $\boldsymbol{A}^H \boldsymbol{E}_{\mathrm{s}}^\star (\boldsymbol{\Lambda}_{\mathrm{s}}^\star)^{-1} (\boldsymbol{E}_{\mathrm{s}}^\star)^H \boldsymbol{A} = \boldsymbol{I}$. Therefore $\boldsymbol{A}_{\mathrm{d}}^H \boldsymbol{K}_3^\star \boldsymbol{A}_{\mathrm{d}} = \boldsymbol{I}$, which is nonsingular.

Combining the above results, we obtain that when $\sigma^2 \to 0$,

$$\boldsymbol{\Pi}_{\boldsymbol{M_s}}^\perp \to \boldsymbol{K}_5^\star \boldsymbol{A}_{\mathrm{d}} \boldsymbol{A}_{\mathrm{d}}^H \boldsymbol{K}_5^\star + \frac{\boldsymbol{K}_1^\star \boldsymbol{i}\boldsymbol{i}^H \boldsymbol{K}_1^\star}{M-K}.$$

□

For sufficiently small $\sigma^2 > 0$, it is easy to show that $\boldsymbol{K}_1$–$\boldsymbol{K}_5$ are bounded in the sense of Frobenius norm (i.e., $\|\boldsymbol{K}_i\|_F \leq C$ for some $C > 0$, for $i \in \{1, 2, 3, 4, 5\}$). Because $\partial \boldsymbol{r}/\partial \boldsymbol{\eta}$ is full rank, $\boldsymbol{M_s}$ is also full rank for any $\sigma^2 > 0$, which implies that $\boldsymbol{\Pi}^{\perp}_{\boldsymbol{M_s}}$ is well-defined for any $\sigma^2 > 0$. Observe that $\boldsymbol{\Pi}^{\perp}_{\boldsymbol{M_s}}$ is positive semidefinite, and that $\mathrm{tr}(\boldsymbol{\Pi}^{\perp}_{\boldsymbol{M_s}}) = \mathrm{rank}(\boldsymbol{M_s})$. We know that $\boldsymbol{\Pi}^{\perp}_{\boldsymbol{M_s}}$ is bounded for any $\sigma^2 > 0$. Therefore $\boldsymbol{Q}_2$ and $\boldsymbol{Q}_3$ are also bounded for sufficiently small σ^2, which implies that $\sigma \boldsymbol{Q}_2 + \sigma^2 \boldsymbol{Q}_3 \to \boldsymbol{0}$ as $\sigma^2 \to 0$.

By Lemma C.1, we know that $\boldsymbol{Q}_1 \to \boldsymbol{Q}_1^{\star}$ as $\sigma^2 \to 0$, where

$$\boldsymbol{Q}_1^{\star} = \dot{\boldsymbol{A}}_{\mathrm{d}}^H (\boldsymbol{K}_2^{\star} - \boldsymbol{K}_4^{\star} \boldsymbol{\Pi}^{\perp}_{\boldsymbol{M_s}}{}^{\star} \boldsymbol{K}_4^{\star}) \dot{\boldsymbol{A}}_{\mathrm{d}},$$

and $\boldsymbol{M_s^{\star}} = \lim_{\sigma^2 \to 0^+} \boldsymbol{\Pi}^{\perp}_{\boldsymbol{M_s}}$ as derived in Lemma C.1. Assume $\boldsymbol{Q}_1^{\star}$ is nonsingular[8]. By (C.3) we immediately obtain that $\mathrm{CRB}_{\boldsymbol{\theta}} \to \boldsymbol{0}$ as $\sigma^2 \to 0$.

When $K \geq M$, $\boldsymbol{R}$ is full rank regardless of the choice of σ^2. Hence $(\boldsymbol{R}^T \otimes \boldsymbol{R})^{-1}$ is always full rank. Because $\partial \boldsymbol{r}/\partial \boldsymbol{\eta}$ is full column rank, the FIM is positive definite, which implies its Schur complements are also positive definite. Therefore $\mathrm{CRB}_{\boldsymbol{\theta}}$ is positive definite.

[8]The condition when $\boldsymbol{Q}_1^{\star}$ is singular is difficult to obtain analytically. In numerical simulations, we have verified that it remains nonsingular for various parameter settings.

Appendix D

Proof of Proposition 3.3

Following [61, Appendix G], for ULAs with a large number of sensors, M, we have

$$\frac{1}{M}\boldsymbol{A}^H\boldsymbol{A} \approx \boldsymbol{I},\ \frac{1}{M^2}\boldsymbol{A}^H\dot{\boldsymbol{A}} \approx \frac{j}{2}\boldsymbol{I},\ \frac{1}{M^3}\dot{\boldsymbol{A}}^H\dot{\boldsymbol{A}} \approx \frac{1}{3}\boldsymbol{I}. \tag{D.1}$$

Applying Lemma 3.2, the inverse of $\boldsymbol{R}$ can be rewritten as

$$\boldsymbol{R}^{-1} = \sigma^{-2}[\boldsymbol{I} - \boldsymbol{A}(\sigma^2\boldsymbol{P}^{-1} + \boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{A}^H]. \tag{D.2}$$

Combined with the assumption that $\mathrm{SNR}_i^{-1} = \sigma^2/p_i \ll M$, we have

$$\begin{aligned}
\boldsymbol{A}^H\boldsymbol{R}^{-1}\boldsymbol{A} =& \sigma^{-2}\boldsymbol{A}^H\boldsymbol{A}[\boldsymbol{I} - (\sigma^2\boldsymbol{P}^{-1} + \boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{A}^H\boldsymbol{A}] \\
=& \sigma^{-2}\boldsymbol{A}^H\boldsymbol{A}(\sigma^2\boldsymbol{P}^{-1} + \boldsymbol{A}^H\boldsymbol{A})^{-1}[\sigma^2\boldsymbol{P}^{-1} + \boldsymbol{A}^H\boldsymbol{A} - \boldsymbol{A}^H\boldsymbol{A}] \\
=& \boldsymbol{A}^H\boldsymbol{A}(\sigma^2\boldsymbol{P}^{-1} + \boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{P}^{-1} \\
\approx& \boldsymbol{P}^{-1},
\end{aligned} \tag{D.3}$$

$$\begin{aligned}\dot{\boldsymbol{A}}^H\boldsymbol{R}^{-1}\boldsymbol{A} =&\sigma^{-2}[\dot{\boldsymbol{A}}^H\boldsymbol{A}-\dot{\boldsymbol{A}}^H\boldsymbol{A}(\sigma^2\boldsymbol{P}^{-1}+\boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{A}^H\boldsymbol{A}]\\ =&\sigma^{-2}\dot{\boldsymbol{A}}^H\boldsymbol{A}(\sigma^2\boldsymbol{P}^{-1}+\boldsymbol{A}^H\boldsymbol{A})^{-1}(\sigma^2\boldsymbol{P}^{-1}+\boldsymbol{A}^H\boldsymbol{A}-\boldsymbol{A}^H\boldsymbol{A})\\ \approx&\dot{\boldsymbol{A}}^H\boldsymbol{A}(\boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{P}^{-1}\\ \approx&-j\frac{M}{2}\boldsymbol{P}^{-1},\end{aligned} \tag{D.4}$$

and

$$\dot{\boldsymbol{A}}^H\boldsymbol{R}^{-1}\dot{\boldsymbol{A}}=\sigma^{-2}[\dot{\boldsymbol{A}}^H\dot{\boldsymbol{A}}-\dot{\boldsymbol{A}}^H\boldsymbol{A}(\sigma^2\boldsymbol{P}^{-1}+\boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{A}^H\dot{\boldsymbol{A}}]\approx\frac{M^3}{12}\sigma^{-2}\boldsymbol{I}. \tag{D.5}$$

Substituting (D.3)–(D.5) into the expression of $\boldsymbol{J}_{\omega\omega}$, we obtain

$$\boldsymbol{J}_{\omega\omega}=\frac{M^3}{6}\sigma^{-2}\boldsymbol{P}.$$

Using similar tricks, we can obtain the following:

$$\mathrm{tr}(\boldsymbol{R}^{-2})\approx\sigma^{-4}(M-K), \tag{D.6}$$

$$\boldsymbol{A}^H\boldsymbol{R}^{-2}\boldsymbol{A}=\boldsymbol{A}^H\boldsymbol{A}[(\sigma^2\boldsymbol{P}^{-1}+\boldsymbol{A}^H\boldsymbol{A})\boldsymbol{P}]^{-2}\approx\boldsymbol{0}, \tag{D.7}$$

$$\dot{\boldsymbol{A}}^H\boldsymbol{R}^{-2}\boldsymbol{A}=\dot{\boldsymbol{A}}^H\boldsymbol{A}[(\sigma^2\boldsymbol{P}^{-1}+\boldsymbol{A}^H\boldsymbol{A})\boldsymbol{P}]^{-2}\approx-\frac{j}{2}\boldsymbol{P}^{-2}. \tag{D.8}$$

The detailed derivation of (D.7) is summarized in (D.9) below. The derivation of (D.8) follows the same idea.

$$
\begin{aligned}
&\boldsymbol{A}^H\boldsymbol{R}^{-2}\boldsymbol{A}\\
=&\sigma^{-4}\boldsymbol{A}^H\boldsymbol{A}(\sigma^2\boldsymbol{P}^{-1}+\boldsymbol{A}^H\boldsymbol{A})^{-1}[\sigma^2\boldsymbol{P}^{-1}+\boldsymbol{A}^H\boldsymbol{A}-2\boldsymbol{A}^H\boldsymbol{A}+\boldsymbol{A}^H\boldsymbol{A}(\sigma^2\boldsymbol{P}^{-1}+\boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{A}^H\boldsymbol{A}]\\
=&\sigma^{-4}\boldsymbol{A}^H\boldsymbol{A}(\sigma^2\boldsymbol{P}^{-1}+\boldsymbol{A}^H\boldsymbol{A})^{-1}[\sigma^2\boldsymbol{P}^{-1}-\boldsymbol{A}^H\boldsymbol{A}(\sigma^2\boldsymbol{P}^{-1}+\boldsymbol{A}^H\boldsymbol{A})^{-1}(\sigma^2\boldsymbol{P}^{-1}+\boldsymbol{A}^H\boldsymbol{A}-\boldsymbol{A}^H\boldsymbol{A})]\\
=&\sigma^{-4}\boldsymbol{A}^H\boldsymbol{A}(\sigma^2\boldsymbol{P}^{-1}+\boldsymbol{A}^H\boldsymbol{A})^{-1}(\sigma^2\boldsymbol{P}^{-1}+\boldsymbol{A}^H\boldsymbol{A}-\boldsymbol{A}^H\boldsymbol{A})(\sigma^2\boldsymbol{P}^{-1}+\boldsymbol{A}^H\boldsymbol{A})^{-1}\sigma^2\boldsymbol{P}^{-1}\\
=&\boldsymbol{A}^H\boldsymbol{A}(\sigma^2\boldsymbol{P}^{-1}+\boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{P}^{-1}(\sigma^2\boldsymbol{P}^{-1}+\boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{P}^{-1}.
\end{aligned}
\tag{D.9}
$$

By (D.3) and (D.6), we obtain that $\boldsymbol{J}_{\boldsymbol{pp}} \approx \boldsymbol{P}^{-2}$ and that $\boldsymbol{J}_{\sigma^2\sigma^2} \approx \sigma^{-4}(M-K)$, both of which will not vanish as M grows. According to (D.4) and (D.8), in the expressions of $\boldsymbol{J}_{\boldsymbol{\omega p}}$ and $\boldsymbol{J}_{\boldsymbol{\omega}\sigma^2}$, the terms inside the $\Re(\cdot)$ operator will be almost imaginary. Therefore, both $\boldsymbol{J}_{\boldsymbol{\omega p}}$ and $\boldsymbol{J}_{\boldsymbol{\omega}\sigma^2}$ will be approximately zeros for large values of M. Consequently, the FIM will be block diagonal and we only need to evaluate $\boldsymbol{J}_{\boldsymbol{\omega\omega}}^{-1}$ to obtain $\boldsymbol{B}_{(\text{sto-uc})}$, which leads to

$$
\boldsymbol{B}_{(\text{sto-uc})}(\boldsymbol{\omega}) \approx \frac{1}{N}\boldsymbol{J}_{\boldsymbol{\omega\omega}}^{-1} = \frac{6}{M^3N}\sigma^2\boldsymbol{P}^{-1}.
$$

Appendix E

Proof of Theorem 3.4

We first prove the result for co-prime arrays. In the one source case, the steering matrix $\boldsymbol{A}$ reduces to a vector $\boldsymbol{a} = [\boldsymbol{a}_1^T \ \boldsymbol{a}_2^T]^T \in \mathbb{C}^{3Q\times 1}$, where

$$\boldsymbol{a}_1^T = \begin{bmatrix} 1 & e^{jQ\omega} & \dots & e^{jQ^2\omega} \end{bmatrix}, \tag{E.1}$$

$$\boldsymbol{a}_2^T = \begin{bmatrix} e^{j(Q+1)\omega} & e^{j2(Q+1)\omega} & \dots & e^{j(2Q-1)(Q+1)\omega} \end{bmatrix}. \tag{E.2}$$

With respect to $\boldsymbol{\omega}$, the derivative vector $\dot{\boldsymbol{a}}$ is given by $\dot{\boldsymbol{a}} = j\boldsymbol{D}\boldsymbol{a}$, where $\boldsymbol{D} = \text{diag}(\boldsymbol{D}_1, \boldsymbol{D}_2)$, and

$$\begin{aligned} \boldsymbol{D}_1 &= \text{diag}(0, Q, \dots, Q^2), \\ \boldsymbol{D}_2 &= \text{diag}(Q+1, 2(Q+1), \dots, (2Q-1)(Q+1)). \end{aligned} \tag{E.3}$$

Therefore, we have

$$\begin{aligned}
&\boldsymbol{a}^H\boldsymbol{a} = 3Q,\\
&\dot{\boldsymbol{a}}^H\boldsymbol{a} = -j\Big[\sum_{q=1}^{Q} qQ + \sum_{q=1}^{2Q-1} q(Q+1)\Big] \approx -j\frac{5}{2}Q^3,\\
&\dot{\boldsymbol{a}}^H\dot{\boldsymbol{a}} = \sum_{q=1}^{Q} q^2Q^2 + \sum_{q=1}^{2Q-1} q^2(Q+1)^2 \approx 3Q^5,
\end{aligned}$$

where the approximations are obtained by removing terms that are one-order smaller than the highest order terms. Following the proof of Lemma 3.4, we have

$$\boldsymbol{R}^{-1} = \sigma^{-2}[\boldsymbol{I} - \boldsymbol{a}(\sigma^2 p^{-1} + \boldsymbol{a}^H\boldsymbol{a})^{-1})^{-1}\boldsymbol{a}^H] = \sigma^{-2}\Big(\boldsymbol{I} - \frac{\boldsymbol{a}\boldsymbol{a}^H}{\sigma^2 p^{-1} + 3Q}\Big).$$

Therefore, when $\mathrm{SNR}^{-1} \ll Q$,

$$\boldsymbol{a}^H\boldsymbol{R}^{-1}\boldsymbol{a} = \sigma^{-2}\Big(\boldsymbol{a}^H\boldsymbol{a} - \frac{\boldsymbol{a}^H\boldsymbol{a}\boldsymbol{a}^H\boldsymbol{a}}{\sigma^2 p^{-1} + 3Q}\Big) = \frac{3Qp^{-1}}{\sigma^2 p^{-1} + 3Q} \approx p^{-1}.$$

Similarly, we can show that

$$\dot{\boldsymbol{a}}^H\boldsymbol{R}^{-1}\boldsymbol{a} = \sigma^{-2}\Big(\dot{\boldsymbol{a}}^H\boldsymbol{a} - \frac{\dot{\boldsymbol{a}}^H\boldsymbol{a}\boldsymbol{a}^H\boldsymbol{a}}{\sigma^2 p^{-1} + 3Q}\Big) \approx -j\frac{5}{6}Q^2p^{-1},$$

and that

$$\dot{\boldsymbol{a}}^H\boldsymbol{R}^{-1}\dot{\boldsymbol{a}} = \sigma^{-2}\Big(\dot{\boldsymbol{a}}^H\dot{\boldsymbol{a}} - \frac{\dot{\boldsymbol{a}}^H\boldsymbol{a}\boldsymbol{a}^H\dot{\boldsymbol{a}}}{\sigma^2 p^{-1} + 3Q}\Big) \approx \frac{11}{12}Q^5\sigma^{-2}.$$

Observing that $\dot{\boldsymbol{a}}^H \boldsymbol{R}^{-1} \boldsymbol{a}$ is purely imaginary, we immediately know that $J_{\omega p}$ and $J_{\omega \sigma^2}$ are exactly zero. Hence, the FIM takes the following form:

$$\boldsymbol{J} = N \begin{bmatrix} J_{\omega\omega} & 0 & 0 \\ 0 & * & * \\ 0 & * & * \end{bmatrix}. \tag{E.4}$$

Therefore, to obtain $\boldsymbol{B}_{(\text{sto-uc})}(\omega)$, we need to evaluate only $J_{\omega\omega}$, which is given by

$$\begin{aligned} J_{\omega\omega} &= 2\Re[(\dot{\boldsymbol{a}}^H \boldsymbol{R}^{-1} \dot{\boldsymbol{a}})^* \circ (p^2 \boldsymbol{a}^H \boldsymbol{R}^{-1} \boldsymbol{a}) + (\dot{\boldsymbol{a}}^H \boldsymbol{R}^{-1} \boldsymbol{a})^* \circ (p^2 \boldsymbol{a}^H \boldsymbol{R}^{-1} \dot{\boldsymbol{a}})] \\ &= 2\Re\Big[\frac{11}{12} Q^5 p \sigma^{-2} + \frac{25}{36} Q^4\Big] \\ &\approx \frac{11}{6} Q^5 p \sigma^{-2}. \end{aligned}$$

We finally obtain that

$$\boldsymbol{B}_{(\text{sto-uc})}(\omega) = \frac{1}{N} J_{\omega\omega}^{-1} \approx \frac{6}{11} \frac{1}{N} \frac{1}{Q^5} \frac{1}{\text{SNR}}. \tag{E.5}$$

Given a nested array configured with the parameter pair (Q, Q), its steering vector for the one source case is given by $\boldsymbol{a} = [\boldsymbol{a}_1^T \boldsymbol{a}_2^T]^T$, where

$$\begin{aligned} \boldsymbol{a}_1^T &= \begin{bmatrix} 1 & e^{j\omega} & \dots & e^{j(Q-1)\omega} \end{bmatrix}, \\ \boldsymbol{a}_2^T &= \begin{bmatrix} e^{jQ\omega} & e^{j[Q+(Q+1)]\omega} & \dots & e^{j[Q+(Q-1)(Q+1)]\omega} \end{bmatrix}. \end{aligned}$$

With respect to $\boldsymbol{\omega}$, the derivative vector $\dot{\boldsymbol{a}}$ is given by $\dot{\boldsymbol{a}} = j\boldsymbol{D}\boldsymbol{a}$, where $\boldsymbol{D} = \text{diag}(\boldsymbol{D}_1, \boldsymbol{D}_2)$, and

$$\begin{aligned} \boldsymbol{D}_1 &= \text{diag}(0, 1, \ldots, Q-1), \\ \boldsymbol{D}_2 &= \text{diag}(Q, Q+(Q+1), \ldots, Q+(Q-1)(Q+1)). \end{aligned} \tag{E.6}$$

Similar to the co-prime array case, we can calculate the following terms as

$$\begin{aligned} \boldsymbol{a}^H\boldsymbol{a} &= 2Q, \\ \dot{\boldsymbol{a}}^H\boldsymbol{a} &= -j\sum_{q=0}^{Q-1}[q+Q+q(Q+1)] \approx -j\frac{1}{2}Q^3, \\ \dot{\boldsymbol{a}}^H\dot{\boldsymbol{a}} &= \sum_{q=0}^{Q-1}[q^2+(Q+q(Q+1))^2] \approx \frac{1}{3}Q^5. \end{aligned}$$

We can calculate the inverse of $\boldsymbol{R}$ as

$$\boldsymbol{R}^{-1} = \sigma^{-2}\Big[\boldsymbol{I} - \frac{\boldsymbol{a}\boldsymbol{a}^H}{\sigma^2 p^{-1} + 2Q}\Big]. \tag{E.7}$$

Hence,

$$\boldsymbol{a}^H\boldsymbol{R}^{-1}\boldsymbol{a} = \sigma^{-2}\Big[2Q - \frac{4Q^2}{\sigma^2 p^{-1}+2Q}\Big] \approx p^{-1}.$$

$$\dot{\boldsymbol{a}}^H\boldsymbol{R}^{-1}\boldsymbol{a} \approx -j\sigma^{-2}\Big[\frac{1}{2}Q^3 - \frac{Q^4}{\sigma^2 p^{-1}+2Q}\Big] \approx -j\frac{1}{4}Q^2 p^{-1}.$$

$$\dot{\boldsymbol{a}}^H\boldsymbol{R}^{-1}\dot{\boldsymbol{a}} \approx \sigma^{-2}\Big[\frac{1}{3}Q^5 - \frac{1}{4}\frac{Q^6}{\sigma^2 p^{-1}+2Q}\Big] \approx \frac{5}{24}Q^5\sigma^{-2}.$$

Similar to the co-prime case, the FIM is block diagonal, and we need to evaluate only $J_{\omega\omega}$. Combining the above results, we obtain that

$$J_{\omega\omega} \approx \frac{5}{12}Q^5 p\sigma^{-2}. \tag{E.8}$$

Therefore,

$$\boldsymbol{B}_{(\text{sto-uc})}(\omega) = \frac{1}{N} J^{-1}_{\omega\omega} \approx \frac{12}{5} \frac{1}{N} \frac{1}{Q^5} \frac{1}{\text{SNR}}. \tag{E.9}$$

Appendix F

Proof of Theorem 3.5

Lemma F.1. *Sum of trigonometric series:*

$$\sum_{t=0}^{n-1} \sin(\phi + tld) = \frac{\sin\frac{nld}{2}}{\sin\frac{ld}{2}} \sin(\phi + \frac{n-1}{2} ld), \tag{F.1}$$

$$\sum_{t=0}^{n-1} \cos(\phi + tld) = \frac{\sin\frac{nld}{2}}{\sin\frac{ld}{2}} \cos(\phi + \frac{n-1}{2} ld). \tag{F.2}$$

In the proof of Theorem 3.4, because the steering matrix reduces to a vector, the resulting inner product is easy to compute. However, in the case of multiple sources, $\boldsymbol{A}^H\boldsymbol{A}$ will be a full matrix whose off-diagonal elements are generated by the inner products between $\boldsymbol{a}(\omega_i)$ and $\boldsymbol{a}(\omega_j)$, $i \neq j$. These elements are generally not zero. We can follow similar steps as we did in the proof of Proposition 3.3 if we can show that these off-diagonal elements are much smaller than the main diagonal elements under certain conditions and that $\boldsymbol{A}^H\boldsymbol{A}$ can be expressed as

$$\boldsymbol{A}^H\boldsymbol{A} = P_n(Q)\boldsymbol{I} + o(Q^n),$$

where $P_n(Q)$ is a polynomial of Q with degree n.

However, as will be shown later, the above approximation may not always be possible because one of the subarrays in co-prime/nested arrays has an inter-element spacing that is greater than d_0. Degenerative cases occur under some specific DOA configurations. Nevertheless, we shall show that the CRB can indeed decrease at a rate of $O(Q^{-5})$.

For brevity, we only show the detailed derivations for the co-prime array case. The derivation for the nested array case is actually simpler because the first subarray of a nested array is a ULA with an inter-element spacing d_0. In the multiple source case, the steering matrix of the co-prime array generated with the co-prime pair $(Q, Q+1)$ can be expressed as $\boldsymbol{A} = [\boldsymbol{A}_1^T \ \boldsymbol{A}_2^T]^T$, where

$$\boldsymbol{A}_1 = \begin{bmatrix} \boldsymbol{a}_1(\omega_1) & \boldsymbol{a}_1(\omega_2) & \cdots & \boldsymbol{a}_1(\omega_K) \end{bmatrix},$$
$$\boldsymbol{A}_2 = \begin{bmatrix} \boldsymbol{a}_2(\omega_1) & \boldsymbol{a}_2(\omega_2) & \cdots & \boldsymbol{a}_2(\omega_K) \end{bmatrix},$$

and $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ follow the same definitions as those in (E.1), (E.2). We also have $\dot{\boldsymbol{A}} = j\boldsymbol{D}\boldsymbol{A}$, where $\boldsymbol{D}$ follows the same definition as that in (E.3).

Therefore,

$$\left[\boldsymbol{A}^H\boldsymbol{A}\right]_{m,n} = [\boldsymbol{A}_1^H\boldsymbol{A}_1]_{m,n} + [\boldsymbol{A}_2^H\boldsymbol{A}_2]_{m,n} = \sum_{q=0}^{Q} e^{jqQ(\omega_m-\omega_n)} + \sum_{q=1}^{2Q-1} e^{jq(Q+1)(\omega_m-\omega_n)}. \qquad \text{(F.3)}$$

Here $[\cdot]_{m,n}$ denotes the (m,n)-th element.

When $m = n$, the sum reduces to $3Q$, as computed in Theorem 3.4. When $m \neq n$, by Lemma F.1, we have

$$\begin{aligned}[\boldsymbol{A}_1^H \boldsymbol{A}_1]_{m,n} =& \frac{\sin[(Q+1)Q(\omega_m - \omega_n)/2]}{\sin[Q(\omega_m - \omega_n)/2]} \cos\left[\frac{1}{2}Q^2(\omega_m - \omega_n)\right] \\ &+ j\frac{\sin[(Q+1)Q(\omega_m - \omega_n)/2]}{\sin[Q(\omega_m - \omega_n)/2]} \sin\left[\frac{1}{2}Q^2(\omega_m - \omega_n)\right].\end{aligned}$$

Note that the absolute values of the numerators are bounded above by one. $[\boldsymbol{A}_1^H \boldsymbol{A}_1]_{m,n}$ will become large when $\sin[Q(\omega_m - \omega_n)/2]$ is close to zero (the actual limit is $Q+1$ by L'Hospital's rule). Therefore, if we restrict the range of $\omega_m - \omega_n$, we can bound $[\boldsymbol{A}_1^H \boldsymbol{A}_1]_{m,n}$ from above by a constant that does not grow with Q. This the reason why we introduce Definition 3.1.

By Definition 3.1, we immediately know that $|\sin(\omega L/2)|^{-1}$ is bounded above by δ^{-1}, $\forall \omega \in \Omega_L^\delta$. For a fixed δ, if we restrict $\omega_m - \omega_n$ within Ω_Q^δ, $\forall m \neq n$, then $|[\boldsymbol{A}_1^H \boldsymbol{A}_1]_{m,n}| \leq \sqrt{2}\delta^{-1}$, $\forall m \neq n$, which leads to

$$\frac{1}{Q+1}\boldsymbol{A}_1^H \boldsymbol{A}_1 \approx \boldsymbol{I}. \tag{F.4}$$

Similarly, we have

$$\frac{1}{2Q-1}\boldsymbol{A}_2^H \boldsymbol{A}_2 \approx \boldsymbol{I} \tag{F.5}$$

if we restrict $\omega_m - \omega_n$ within Ω_{Q+1}^δ, $\forall m \neq n$.

Lemma F.2. *As long as δ is not very close to 1, $\Omega_Q^\delta \cap \Omega_{Q+1}^\delta \neq \emptyset$ for $Q \geq 2$.*

Proof. Let $\phi = \arcsin\delta$. Consider the interval $\frac{2}{Q}[\phi, \pi - \phi]$ in Ω_Q^δ and the interval $\frac{2}{Q+1}[\phi, \pi - \phi]$ in Ω_{Q+1}^δ. The condition of overlapping is given by

$$\frac{2}{Q}\phi < \frac{2}{Q+1}(\pi - \phi),$$

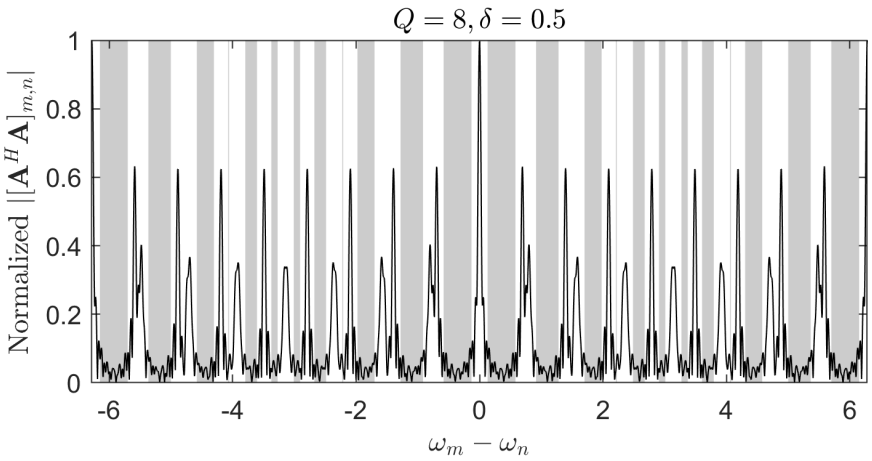


Figure F.1: $|[\boldsymbol{A}^H\boldsymbol{A}]_{m,n}|$ v.s. $\omega_m - \omega_n$ for $Q = 8$ and $\delta = 0.5$. The shaded regions are defined by $\Omega_Q^\delta \cap \Omega_{Q+1}^\delta$. It can be observed that $|[\boldsymbol{A}^H\boldsymbol{A}]_{m,n}|$ is very small in the shaded regions.

which is equivalent to

$$\phi < \frac{\pi}{2 + 1/Q}.$$

When $Q \geq 2$, we only need to choose $\phi < 2\pi/5$ and the above condition will hold, which corresponds to choosing $\delta < 0.95$. □

Therefore, for a reasonable choice of δ (e.g., 0.5), if $\omega_m - \omega_n \in \Omega_Q^\delta \cap \Omega_{Q+1}^\delta$, $\forall m \neq n$, then $\boldsymbol{A}^H\boldsymbol{A}/(3Q) \approx \boldsymbol{I}$, which is very similar to the result we obtained in Theorem 3.4. To demonstrate this, we plot $|[\boldsymbol{A}^H\boldsymbol{A}]_{m,n}|$ as a function of $\omega_m - \omega_n$ in Fig. F.1. We can observe that for certain values of $\omega_m - \omega_n$, the summation of the trigonometric series is indeed large and cannot be neglected. However, in the shaded areas defined by $\Omega_Q^\delta \cap \Omega_{Q+1}^\delta$, $|[\boldsymbol{A}^H\boldsymbol{A}]_{m,n}|$ is negligibly small.

Following the same reasoning as in Appendix D, we can obtain the following approximations:

$$\frac{1}{Q}\boldsymbol{A}^H\boldsymbol{A} \approx 3\boldsymbol{I}, \tag{F.6}$$

$$\frac{1}{Q^3}\dot{\boldsymbol{A}}^H\boldsymbol{A} \approx -j\frac{5}{2}\boldsymbol{I}, \tag{F.7}$$

$$\frac{1}{Q^5}\dot{\boldsymbol{A}}^H\dot{\boldsymbol{A}} \approx 3\boldsymbol{I}. \tag{F.8}$$

We can now substitute these terms back into the expression of $\boldsymbol{B}_{(\text{sto-uc})}$. Following the same approach as in the proof of Proposition 3.3, we obtain

$$\boldsymbol{B}_{(\text{sto-uc})}(\omega) \approx \frac{6}{11}\frac{1}{N}\frac{1}{Q^5}\sigma^2\boldsymbol{P}^{-1}, \tag{F.9}$$

if $\omega_m - \omega_n \in \Omega_Q^\delta \cap \Omega_{Q+1}^\delta$, $\forall m \neq n$, $m, n \in \{1, 2, \ldots, K\}$, and some reasonable choice of δ.

Following the same idea, we can obtain a similar result for nested arrays generated by the parameter pair (Q, Q):

$$\boldsymbol{B}_{(\text{sto-uc})}(\omega) \approx \frac{12}{5}\frac{1}{N}\frac{1}{Q^5}\sigma^2\boldsymbol{P}^{-1}, \tag{F.10}$$

if $\omega_m - \omega_n \in \Omega_{Q+1}^\delta$, $\forall m \neq n$, $m, n \in \{1, 2, \ldots, K\}$, and for a reasonable choice of δ.

Vita

Mianzhi Wang

Degrees

Ph.D., Electrical Engineering, Washington University in St. Louis, Missouri, USA, August 2018
M.S., Electrical Engineering, Washington University in St. Louis, Missouri, USA, May 2018
B.S., Electronic Engineering, Fudan University, Shanghai, China, June 2013

Professional Memberships

The Institute of Electrical and Electronics Engineers (IEEE)
IEEE Signal Processing Society

Publications

Journal Publications:

M. Wang, Z. Zhang, and A. Nehorai, "Further results on Coarrays, MUSIC, and the Cramér Rao bound," submitted to *IEEE Trans. Signal Process.*

Z. Zhang, **M. Wang**, and A. Nehorai, "Optimal transport in reproducing kernel Hilbert spaces: theory and applications," in revision for *IEEE Trans. Pattern Anal. Mach. Intell.*

M. Wang, Z. Zhang, and A. Nehorai, "Performance analysis of coarray-based MUSIC in the presence of sensor location errors," *IEEE Trans. Signal Process.*, vol. 66, pp. 3074-3085, June 2018.

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.

Conference Publications:

Z. Zhang, **M. Wang**, Y. Huang, and A. Nehorai, "Aligning infinite-dimensional covariance matrices in reproducing kernel Hilbert spaces for domain adaptation," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, USA, June 22–28, 2018.

G. Jiang, X. Mao, **M. Wang**, Y. Liu and A. Nehorai, "Underdetermined DOA estimation with unknown source number in nonuniform noise," *2018 IEEE Radar Conference (RadarConf18)*, Oklahoma City, OK, Apr. 23–27, 2018.

M. Wang, Z. Zhang, and A. Nehorai, "Performance analysis of coarray-based MUSIC and the Cramér-Rao bound," *Proc. 42nd IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, New Orleans, LA, Mar. 5–9, 2017.

M. Wang, Z. Zhang, and A. Nehorai, "Direction finding using sparse linear arrays with missing data," *Proc. 42nd IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, New Orleans, LA, Mar. 5–9, 2017.

Z. Zhang, **M. Wang**, Y. Xiang, A. Nehorai, "Geometry-adapted Gaussian random field regression," *Proc. 42nd IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, New Orleans, LA, Mar. 5-9, 2017.

August 2018