EURECOM
Department of Mobile Communications
Campus SophiaTech
CS 50193
06904 Sophia Antipolis cedex
FRANCE

Research Report RR-16-326

On AoA Estimation in the Presence of Mutual Coupling: Algorithms and Performance Analysis

November 1st, 2016

Ahmad Bazzi and Dirk. T.M. Slock

Tel : (+33) 4 93 00 81 00
Fax : (+33) 4 93 00 82 00
Email : {bazzi, slock}@eurecom.fr

[1]EURECOM's research is partially supported by its industrial members: BMW Group Research and Technology, IABG, Monaco Telecom, Orange, Principaut de Monaco, SAP, ST Microelectronics, Symantec.

On AoA Estimation in the Presence of Mutual Coupling: Algorithms and Performance Analysis

Ahmad Bazzi and Dirk. T.M. Slock

Abstract

This paper presents two novel methods for Angle-of-Arrival (AoA) estimation in the presence of unknown mutual coupling. We mainly focus on the first method, where we show that this method could estimate AoAs of multiple sources when more mutual coupling parameters are included in the model, compared to existing methods. Furthermore, we derive a closed form Mean-Squared-Error (MSE) expression of the proposed method and compare it to the MSE of MUSIC with known coupling parameters. In addition, the derived MSE expression is also compared to the *Cramér-Rao* bound of joint AoA and mutual coupling estimation in the asymptotic regime, i.e. at high signal-to-Noise ratio or large number of antennas. The second method serves as a refinement of the first one. Simulation results have demonstrated the potential of the proposed methods as they enjoy better performance than existing methods. A better description of the paper could be found in the Conclusions section.

Index Terms

Angle-of-Arrival estimation, Mutual Coupling, Mean-Squared-Error analysis, MUSIC.

Contents

List of Figures

1 Introduction

1.1 Localization and Array Signal Processing

Localisation, also known as location estimation, of multiple emitting sources by observing measurements at the output of an array of antennas arises frequently in signal processing applications, such as sonar, seismology, radar, and astronomy [1, 2]. Indeed, locating multiple sources could be determined by processing the different signals at the output of the receiving array. This is so because the phase shifts across different antennas is a function of the angular position of the emitter [3, 4], known as the Angle-of-Arrival, or AoA for short. Once the AoA is known, say by two different array receivers, then one could apply techniques, like triangulation, to locate the source [5, 6]. Of course, there exists other signal parameters that could determine the position of a source, such as the Time-of-Arrival (ToA) and Recieved Signal Strength (RSS). For more information on this topic, we encourage the reader to refer to [7]. Our main focus in this paper is on Angle-of-Arrival estimation using an array of antennas.

Throughout several decades, the problem of Angle-of-Arrival estimation has been seen as a specific problem of array signal processing and parameter estimation [8]. The Maximum Likelihood (ML) approach was one of the first to be investigated [9]. The ML is optimal, in a sense that it minimizes the mean squared error, but it requires a multidimensional search over different AoAs. As a response, this issue attracted a number of researchers in order to find computationally efficient methods and, at the same time, optimise the ML cost function, such as the Iterative-Quadratic-ML (IQML) algorithm [10], the Alternating Projection (AP) method [11], the Expectation-Maximisation (EM) method [12], the Method-Of-Direction-Estimation (MODE) [13], and Space-Alternating-Generalised-EM (SAGE) [14]. These methods are "much faster" than a multidimensional search, but might suffer from convergence issues. For instance, see [15, 16] for convergence properties on IQML, MODE, EM and SAGE.

Subspace methods are another class of algorithms, such as MUltiple SIgnal Classification, MUSIC [17], and Estimation of Signal Parameters via Rotational Invariance Techniques, ESPRIT [18]. These class of algorithms have attracted a numerous amout of researchers because they are computationally more attractive than the algorithms listed above, however this comes at the price of performance. More specifically, MUSIC is a large sample realization of an ML estimator [19, 20]. The ESPRIT algorithm was also analyzed in [21, 22] and was reported to be statistically less efficient than MUSIC in [21]. Moreover, the performance of both MUSIC and ESPRIT deteriorate when the sources are highly correlated. As a result, spatial smoothing [23] was proposed as a remedy of highly correlated sources. In this paper, we focus on the MUSIC algorithm to solve the problem we are interested in.

1.2 Literature Review

In practical scenarios, an ideal model is, rarely, satisfied and modeling errors are the main reason behind this non-ideality. Moreover, it is well-known that different modeling errors could degrade the performance of direction-finding algorithms, when not taken into consideration [24, 25]. More specifically, these modeling errors induce a bias in the MUSIC estimator and could result in large mean-squared errors of AoA estimates, when the modeling errors are large enough [25]. These modeling errors include: antenna position uncertainty [26], unknown gain/phases between different antennas [27], and mutual coupling between antennas [28].

Mutual coupling is a popular problem in array signal processing. This phenomenon arises when antennas are close to each other [29], and thus the current developed in an antenna element depends on its own excitation and on the contributions from adjacent antennas. Methods that aim on solving the mutual coupling problem are sometimes referred to as calibration methods, which are of two types: Offline and Online. In an offline calibration approach, one estimates the mutual coupling parameters using known locations [30–34]. For the sake of fairness, we should distinguish the paper in [34], since this method could calibrate the antennas in the presence of multipath. Offline calibration methods could estimate all modeling errors, thanks to the aid of known positions. However, one might argue that this approach is time-consuming and, in some scenarios, impossible to implement. Other methods assume partly calibrated antennas [35–38] or partly known positions [39–41].

The other type of calibration methods, which is our main interest, are known as online (or auto-calibration) methods. One of the first papers to appear that deal with online calibration is [42]. This method is iterative and performs alternating minimisation steps between the AoAs and mutual coupling parameters to optimize the MUSIC cost function. However, this method [42] doesn't admit a unique solution (See [43] for the case of linear arrays and [44] for non-linear arrays). Moreover, an initial estimate of the coupling parameters are required. Other iterative methods we could mention are [45–48]. In [45], the convergence rate may be slow. Also, convergence is not always guaranteed. In [46], initial estimates are needed and this is not always available. The paper in [48] assume a diagonal source covariance matrix, i.e. the sources have no correlation. This is not always true. In addition, they treat the mutual coupling matrix and its conjugate-transpose as independent matrices. This might lead to some sub-optimality. Furthermore, they have reported, through simulations, that their algorithm needs 12 iterations to converge, and that global convergence is not guaranteed. Moreover, the methods in [49–52] add auxiliary, or "dummy" antennas, to solve the mutual coupling issue. For instance, in [52], antennas were added on the left and right edges of the array. Then, the main array, or the "middle sub-array" of the total array, was used to estimate the AoAs. However, adding antennas is not possible in some scenarios. Also, the algorithms in [53–55] focus on processing the "middle sub-array" of the main array, but this is sub-optimal since not all the antennas are being processed.

A RAnk-REduction estimator, also known as RARE, was first proposed in [56] in the context of partly calibrated arrays. The same idea was used for totally uncalibrated Uniform Circular Arrays (UCA) in [57, 58] and Uniform Linear Arrays (ULA) in [59, 60]. This method makes use of the MUSIC algorithm to estimate AoAs in the presence of mutual coupling via rank reduction of an appropriate matrix. The method in [60] is a Recursive-RARE (R-RARE), which was shown to achieve better performance than the traditional RARE. A similar rank-reduction approach was adopted in [61]. Moreover, the algorithm in [62] is based on a minimum eigenvalues instead. In addition, the method in [63] formulates the problem through a quadratic minimisation problem. We shall elaborate on those algorithms in Section III.

1.3 Contributions

Our main contributions are summarized as follows.

- We present some introductory algorithms that turn out to be essential for the bulk of this paper. The theorems lead to a consequence for ULAs that suffer from mutual coupling.
- Using these theorems, we derive an algorithm that is able to estimate AoAs when more mutual coupling parameters are present in the model.
- We derive an asymptotic Mean-Squared-Error expression and compare it to the appropriate *Cramér-Rao* bound.
- Another algorithm is introduced to further refine the AoA estimates by solving a multidimensional problem by alternating minimisation.
- Simulation results have demonstrated the potential of the proposed algorithms, which perform better than existing algorithms.

The remainder of the paper is divided as follows: Section 2 presents the system model used throughout the paper. Section 3 revises the MUSIC algorithm, with and without mutual coupling. Also, we summarize some related methods that could estimate AoAs in the presence of mutual coupling. We derive and state some properties of the proposed algorithm that could estimate the AoAs in the presence of mutual coupling in Section 4. The Mean-Squared-Error (MSE) of the proposed algorithm is derived in Section 5, where we compare the MSE expression to the MSE of MUSIC with *known* mutual coupling. We also compare the derived MSE expression to the *Cramér-Rao* bound in Section 6. Furthermore, Section 7 presents a refinement of the proposed method that could further enhance the AoA estimates. Some simulation results are demonstrated in Section 8. We conclude the paper in Section 9. The Appendix of the paper is given in Section 10.

Table 1: Nomenclature

$\mathbb{C}^{M\times N}$	Class of all $M \times N$ complex-valued matrices
$\mathbb{C}^{*}$	Class of all complex-valued numbers except zero
$\boldsymbol{A}^{\mathrm{T}}$	Transposition of matrix $\boldsymbol{A}$
$\boldsymbol{A}^{*}$	Conjugatation of matrix $\boldsymbol{A}$
$\boldsymbol{A}^{\mathrm{H}}$	Conjugate-transposition of matrix $\boldsymbol{A}$
$\boldsymbol{A}^{+}$	*Moore-Penrose* pseudo-inverse of matrix $\boldsymbol{A}$
$\mathrm{Re}\{\boldsymbol{A}\}$	Real part of matrix $\boldsymbol{A}$
$\mathrm{rank}\,\boldsymbol{A}$	Rank of matrix $\boldsymbol{A}$
$\|\boldsymbol{A}\|$	*Frobenius* norm of matrix $\boldsymbol{A}$
$(\boldsymbol{A})_{i,j}$	The $(i,j)^{th}$ element of matrix $\boldsymbol{A}$
$\boldsymbol{I}$	The identity matrix with appropriate dimensions
$\boldsymbol{J}_k$	An all-zero matrix in $\mathbb{C}^{k\times k}$ except ones at its anti-diagonal
$\boldsymbol{0}$	The zero matrix with appropriate dimensions
$\boldsymbol{A}\odot\boldsymbol{B}$	The *Hadamard* product of $\boldsymbol{A}$ and $\boldsymbol{B}$
$\mathcal{N}(\boldsymbol{A})$	The null-space of $\boldsymbol{A}$, i.e. $\boldsymbol{x}\in\mathcal{N}(\boldsymbol{A})$ if $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}$
$\boldsymbol{e}_i$	The i^{th} column of $\boldsymbol{I}$
$X \Longrightarrow Y$	If X is true, then Y is true
$X \Longleftrightarrow Y$	Statements X and Y are equivalent
$\mathbb{E}\{X\}$	Expectation of a random variable X
$\mathcal{R}(\boldsymbol{A},\boldsymbol{x})$	*Rayleigh quotient*, $\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}$
$\mathbb{1}_k$	$k\times 1$ vector of all-ones
$\delta_{i,j}$	The *Dirac* delta (If $i=j$, then $\delta_{i,j}=1$, else 0)
$\|z\|$	Magnitude of complex number z
$\mathcal{CN}(\boldsymbol{\nu},\boldsymbol{\Sigma})$	Complex Gaussian of mean $\boldsymbol{\nu}$ and covariance matrix $\boldsymbol{\Sigma}$

2 System Model

This section formulates the signal model and presents the assumptions used throughout the paper. Then, the problem of mutual coupling is addressed.

2.1 Problem Formulation

Consider a Uniform Linear Array (ULA) of N antennas. Furthermore, assume $q < N$ narrow-band sources, centered around a known frequency, say f_c, attacking the array from different angles, $\boldsymbol{\Theta} = [\theta_1 \dots \theta_q]$. Since narrow-bandedness in the context of array processing means that the propagation delays of the signals across the array are much smaller than the reciprocal of the bandwidth of the signals, it follows that these propagation delays translate into phase shifts that depend on the location.

Now following [1], the received analog signal across all antennas, in the absence of mutual coupling, could be written as

$$\boldsymbol{x}(t) = \sum_{i=1}^{q} \boldsymbol{a}(\theta_i) s_i(t) + \boldsymbol{w}(t) \tag{1}$$

where

$$\boldsymbol{x}(t) = \big[x_1(t) \dots x_N(t)\big]^{\mathrm{T}} \tag{2}$$

is the received vector across all antennas at time t. Moreover, the vector $\boldsymbol{a}(\theta)$ is referred to as the "steering vector" of the array towards angle θ. It is this vector that allows us to perform Angle-of-Arrival (AoA) estimation, and is given by

$$\boldsymbol{a}(\theta) = \big[1, z_\theta, \dots z_\theta^{N-1}\big]^{\mathrm{T}} \tag{3}$$

where $z_\theta = e^{-j2\pi \frac{d}{\lambda} \sin(\theta)}$, d is the inter-element spacing and λ is the signal's wavelength. Moreover, the signal $s_i(t)$ is the signal emitted by the i^{th} source at time t and

$$\boldsymbol{w}(t) = \big[w_1(t) \dots w_N(t)\big]^{\mathrm{T}} \tag{4}$$

is background noise across all antennas at time t. Equation (1) could be written in a more compact way as follows

$$\boldsymbol{x}(t) = \boldsymbol{A}(\boldsymbol{\Theta}) \boldsymbol{s}(t) + \boldsymbol{w}(t) \tag{5}$$

where $\boldsymbol{A}(\boldsymbol{\Theta}) \in \mathbb{C}^{N \times q}$ is referred to as "steering matrix" and is given as

$$\boldsymbol{A}(\boldsymbol{\Theta}) = \big[\boldsymbol{a}(\theta_1) \dots \boldsymbol{a}(\theta_q)\big] \tag{6}$$

and $\boldsymbol{s}(t) \in \mathbb{C}^{q \times 1}$ is the vector of transmitted signals, viz.

$$\boldsymbol{s}(t) = \big[s_1(t) \dots s_q(t)\big]^{\mathrm{T}} \tag{7}$$

Finally, sampling (5) at L time instances, say $t = \{0, T, \dots, (L-1)T\}$, where T is the sampling period, we get

$$\boldsymbol{X} = \boldsymbol{A}(\boldsymbol{\Theta}) \boldsymbol{S} + \boldsymbol{W} \tag{8}$$

where $\boldsymbol{X} = [\boldsymbol{x}(0), \boldsymbol{x}(T), \dots, \boldsymbol{x}\big((L-1)T\big)] \in \mathbb{C}^{N \times L}$ is the data collected over the observed interval of time. Matrices $\boldsymbol{S} \in \mathbb{C}^{q \times L}$ and $\boldsymbol{W} \in \mathbb{C}^{N \times L}$ are similarly defined.

Equation (8) assumes an ideal model, in the sense that each antenna acts independently of all the others. In reality, the current developed in an antenna element depends on its own excitation and on the contributions from adjacent antennas. As a consequence, an ideal model is no longer valid. This phenomenon is called "Mutual Coupling" between array elements, and it enters the model as follows [42]

$$\bar{\boldsymbol{X}} = \boldsymbol{T}(\boldsymbol{c}) \boldsymbol{A}(\boldsymbol{\Theta}) \boldsymbol{S} + \boldsymbol{W} \tag{9}$$

where $\boldsymbol{T}(\boldsymbol{c}) \in \mathbb{C}^{N \times N}$ captures the effect of mutual coupling, and is known as the "Mutual Coupling Matrix" (MCM). Due to the linear and uniform configuration of the different elements of the array, the MCM $\boldsymbol{T}(\boldsymbol{c})$ is given by a symmetric Toeplitz matrix. Let c_i be the coupling coefficient between two elements placed i inter-element spacings apart. Since the amplitude of the coupling parameters tend to decay as a function of increasing distance, namely

$$1 > |c_1| > \ldots > |c_{N-1}| \tag{10}$$

then a well-approximation of $\boldsymbol{T}(\boldsymbol{c})$ is a banded symmetric Toeplitz matrix [66] with bandwidth p, i.e.

$$\big(\boldsymbol{T}(\boldsymbol{c})\big)_{i,j} = \begin{cases} c_{|i-j|} & \text{if } |i-j| < p \\ 0 & \text{else} \end{cases} \tag{11}$$

In other words, antennas that are placed at least p inter-element spacings apart do not interfere, i.e. $c_i = 0$ for all $i \geq p$. In what follows, the MCM of a ULA configuration is modelled as banded symmetric Toeplitz matrix of bandwidth p and denoted as $\boldsymbol{T}(\boldsymbol{c})$, where $\boldsymbol{c} = [1, c_1 \ldots c_{p-1}]^{\mathrm{T}}$ is the vector of coupling parameters.

2.2 Assumptions

Before the problem is clearly addressed, some assumptions have to be made so as to proceed. They are given as follows:

- *Assumption 1*: The Angles-of-Arrival of the q sources are distinct, namely $\theta_i \neq \theta_j$ for $i \neq j$.
- *Assumption 2*: The processes $\boldsymbol{s}(t)$ and $\boldsymbol{w}(t)$ are ergodic and stationary over the observed interval of time.
- *Assumption 3*: The signals $\boldsymbol{s}(t)$ are not coherent, i.e. they are not fully correlated.
- *Assumption 4*: The number of signals q and the bandwidth of $\boldsymbol{T}(\boldsymbol{c})$, which is p, are known.
- *Assumption 5*: The noise $\boldsymbol{w}(lT)$ is modelled as a white circular complex Gaussian process of zero mean and covariance $\sigma^2\boldsymbol{I}$ and independent from the source signals.

2.3 Problem Statement

We are now ready to address our Mutual Coupling problem:

"Given $\bar{\boldsymbol{X}}$, q, and p, estimate the angles-of-arrival $\boldsymbol{\Theta}$ of the incoming signals in the presence of mutual coupling $\boldsymbol{T}(\boldsymbol{c})$."

3 The MUSIC Algorithm and Related Work

3.1 Preliminaries

This subsection serves as a review of the MUSIC algorithm, in the absence of mutual coupling. In other words, the model in equation (8) is assumed. The covariance matrix of the received data could be written as

$$\begin{aligned} \boldsymbol{R_{xx}} &\triangleq \mathbb{E}\{\boldsymbol{x}(t)\boldsymbol{x}^{\mathrm{H}}(t)\} \\ &= \boldsymbol{A}(\boldsymbol{\Theta})\boldsymbol{R_{ss}}\boldsymbol{A}^{\mathrm{H}}(\boldsymbol{\Theta}) + \sigma^2\boldsymbol{I} \end{aligned} \tag{12}$$

where the second equality is due to *Assumption 5* and

$$\boldsymbol{R_{ss}} \triangleq \mathbb{E}\{\boldsymbol{s}(t)\boldsymbol{s}^{\mathrm{H}}(t)\} \tag{13}$$

is the source covariance matrix. Using spectral decomposition, the matrix $\boldsymbol{R_{xx}}$ is expressed as

$$\begin{aligned} \boldsymbol{R_{xx}} &= \left[\begin{array}{c|c} \boldsymbol{U_s} & \boldsymbol{U_n} \end{array}\right] \left[\begin{array}{c|c} \boldsymbol{\Sigma_s} & \boldsymbol{0} \\ \hline \boldsymbol{0} & \boldsymbol{\Sigma_n} \end{array}\right] \left[\begin{array}{c|c} \boldsymbol{U_s} & \boldsymbol{U_n} \end{array}\right]^{\mathrm{H}} \\ &= \boldsymbol{U_s}\boldsymbol{\Sigma_s}\boldsymbol{U_s}^{\mathrm{H}} + \boldsymbol{U_n}\boldsymbol{\Sigma_n}\boldsymbol{U_n}^{\mathrm{H}} \end{aligned} \tag{14}$$

The partitioning in equation (14) is done because $\boldsymbol{R_{xx}}$ is composed of two major parts: Signal and Noise. The signal part $\boldsymbol{A}(\boldsymbol{\Theta})\boldsymbol{R_{ss}}\boldsymbol{A}^{\mathrm{H}}(\boldsymbol{\Theta})$ is rank q, under *Assumptions 1* and *3*. Therefore, due to the noise part $\sigma^2\boldsymbol{I}$, one can say that $\boldsymbol{\Sigma_s}$ is a $q \times q$ diagonal matrix composed of eigenvalues strictly greater than σ^2 and $\boldsymbol{\Sigma_n} = \sigma^2\boldsymbol{I}_{N-q}$. The eigenvectors $\boldsymbol{U_s}$ and $\boldsymbol{U_n}$ are, often, referred to as the *signal* and *noise subspaces*, respectively.
In the absence of mutual coupling, the key to MUSIC is the following observation:

$$\|\boldsymbol{U_n}^{\mathrm{H}}\boldsymbol{a}(\theta)\|^2 = 0 \Longrightarrow \theta \in \boldsymbol{\Theta} \tag{15}$$

In practice, one could estimate the covariance quantities through sample averaging, viz.

$$\widehat{\boldsymbol{R}}_{\boldsymbol{xx}} = \frac{1}{L}\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}} = \widehat{\boldsymbol{U}}_{\boldsymbol{s}}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{s}}\widehat{\boldsymbol{U}}_{\boldsymbol{s}}^{\mathrm{H}} + \widehat{\boldsymbol{U}}_{\boldsymbol{n}}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{n}}\widehat{\boldsymbol{U}}_{\boldsymbol{n}}^{\mathrm{H}} \tag{16}$$

MUSIC estimates the angles-of-arrival $\boldsymbol{\Theta}$ through peak finding, as follows

$$\{\widehat{\theta}_i\}_{i=1}^{q} = \arg\max_{\theta} \frac{1}{\boldsymbol{a}^{\mathrm{H}}(\theta)\widehat{\boldsymbol{U}}_{\boldsymbol{n}}\widehat{\boldsymbol{U}}_{\boldsymbol{n}}^{\mathrm{H}}\boldsymbol{a}(\theta)} \tag{17}$$

3.2 Mutual Coupling in the sense of MUSIC

The previous subsection tells us that one can estimate the angles-of-arrival in the absence of mutual coupling by performing a 1D-search according to equation (17). Now, for the ease of exposition, let $\bar{\boldsymbol{a}}(\theta)$ denote the steering vector in the presence of mutual coupling, i.e.

$$\bar{\boldsymbol{a}}(\theta) = \boldsymbol{T}(\boldsymbol{c})\boldsymbol{a}(\theta) \tag{18}$$

Similarly, define $\bar{\boldsymbol{A}}(\boldsymbol{\Theta})$ as follows

$$\bar{\boldsymbol{A}}(\boldsymbol{\Theta}) = \boldsymbol{T}(\boldsymbol{c})\boldsymbol{A}(\boldsymbol{\Theta}) = [\bar{\boldsymbol{a}}(\theta_1) \dots \bar{\boldsymbol{a}}(\theta_q)] \tag{19}$$

Taking into account mutual coupling, i.e. the model in equation (9), one could follow the same steps from equation (12) till (16) in order to say that the angles-of-arrival could be estimated as follows

$$\{\widehat{\theta}_i\}_{i=1}^{q} = \arg\max_{\theta} \frac{1}{\bar{\boldsymbol{a}}^{\mathrm{H}}(\theta)\widehat{\boldsymbol{U}}_{\boldsymbol{n}}\widehat{\boldsymbol{U}}_{\boldsymbol{n}}^{\mathrm{H}}\bar{\boldsymbol{a}}(\theta)} \tag{20}$$

where $\widehat{\boldsymbol{U}}_{\boldsymbol{n}}$ is the sample estimate of $\boldsymbol{U}_{\boldsymbol{n}}$. Throughout the rest of this paper, $\boldsymbol{U}_{\boldsymbol{n}}$ is the noise subspace, namely $\boldsymbol{U}_{\boldsymbol{n}}\boldsymbol{U}_{\boldsymbol{n}}^{\mathrm{H}} = \boldsymbol{P}_{\bar{\boldsymbol{A}}}^{\perp} = \boldsymbol{I} - \boldsymbol{P}_{\bar{\boldsymbol{A}}}$, where

$$\boldsymbol{P}_{\bar{\boldsymbol{A}}} = \bar{\boldsymbol{A}}(\bar{\boldsymbol{A}}^{\mathrm{H}}\bar{\boldsymbol{A}})^{-1}\bar{\boldsymbol{A}}^{\mathrm{H}} \tag{21}$$

However, applying MUSIC directly as in equation (20) to the problem in hand is not possible, since the functional form of the steering vector is not known. In other terms, we have partial knowledge of vector $\bar{\boldsymbol{a}}(\theta)$, which is that it is a known *Vandermonde* vector $\boldsymbol{a}(\theta)$ pre-multiplied by an unknown banded symmetric Toeplitz matrix $\boldsymbol{T}(\boldsymbol{c})$, as in equation (18). Nevertheless, MUSIC implies the following

$$\|\boldsymbol{U}_{\boldsymbol{n}}^{\mathrm{H}}\boldsymbol{T}(\boldsymbol{m})\boldsymbol{a}(\theta)\|^2 = 0 \Longrightarrow \{\theta \in \boldsymbol{\Theta} \text{ and } \boldsymbol{m} = \boldsymbol{c}\} \tag{22}$$

In order to proceed, we find the following theorem useful:

Theorem 1: *Let $\boldsymbol{\alpha} = [\alpha_0, \alpha_1 \dots \alpha_{p-1}]^T \in \mathbb{C}^{p\times 1}$ and $\boldsymbol{a} \in \mathbb{C}^{N\times 1}$. Define the corresponding matrix $\boldsymbol{T}(\boldsymbol{\alpha})$. Then the following is true for any $1 \leq p \leq N$*

$$\boldsymbol{T}(\boldsymbol{\alpha})\boldsymbol{a} = \boldsymbol{B}_p\boldsymbol{\alpha} \tag{23}$$

where

$$\boldsymbol{B}_p \triangleq \mathcal{G}_p(\boldsymbol{a}) = \left[\ \boldsymbol{a} \ \middle|\ \boldsymbol{S}_1\boldsymbol{a} \ \middle|\ \dots \ \middle|\ \boldsymbol{S}_{p-1}\boldsymbol{a}\ \right] \tag{24}$$

and $\boldsymbol{S}_k \in \mathbb{C}^{N\times N}$ is an all-zero matrix except at the k^{th} sub- and super-diagonals, which are set to 1.

Proof. See [42, 63]. □

Using this theorem, we can say that

$$\bar{\boldsymbol{a}}(\theta) = \boldsymbol{T}(\boldsymbol{c})\boldsymbol{a}(\theta) = \boldsymbol{B}(\theta)\boldsymbol{c} \tag{25}$$

where

$$\boldsymbol{B}(\theta) = \mathcal{G}_p(\boldsymbol{a}(\theta)) \tag{26}$$

Therefore, equation (22) could be re-written as

$$\|\boldsymbol{U}_{\boldsymbol{n}}^{\mathrm{H}}\boldsymbol{B}(\theta)\boldsymbol{m}\|^2 = 0 \Longrightarrow \{\theta \in \boldsymbol{\Theta} \text{ and } \boldsymbol{m} = \boldsymbol{c}\} \tag{27}$$

Said differently and in a more compact way, equation (27) also means

$$\left\| \begin{bmatrix} \boldsymbol{U_n}^{\mathrm{H}}\boldsymbol{B}(\theta_1) \\ \vdots \\ \boldsymbol{U_n}^{\mathrm{H}}\boldsymbol{B}(\theta_q) \end{bmatrix} \boldsymbol{m} \right\|^2 = 0 \Longrightarrow \boldsymbol{m} = \boldsymbol{c} \tag{28}$$

Therefore, one way to formulate the problem is

$$(\mathcal{P}_1): \min_{\boldsymbol{m},\bar{\theta}_1 \ldots \bar{\theta}_q} \boldsymbol{m}^{\mathrm{H}} \widehat{\boldsymbol{S}}(\bar{\theta}_1 \ldots \bar{\theta}_q)\boldsymbol{m} \tag{29}$$

where

$$\widehat{\boldsymbol{S}}(\theta_1 \ldots \theta_q) = \begin{bmatrix} \widehat{\boldsymbol{U}}_{\boldsymbol{n}}^{\mathrm{H}}\boldsymbol{B}(\theta_1) \\ \vdots \\ \widehat{\boldsymbol{U}}_{\boldsymbol{n}}^{\mathrm{H}}\boldsymbol{B}(\theta_q) \end{bmatrix}^{\mathrm{H}} \begin{bmatrix} \widehat{\boldsymbol{U}}_{\boldsymbol{n}}^{\mathrm{H}}\boldsymbol{B}(\theta_1) \\ \vdots \\ \widehat{\boldsymbol{U}}_{\boldsymbol{n}}^{\mathrm{H}}\boldsymbol{B}(\theta_q) \end{bmatrix} = \sum_{j=1}^{q} \widehat{\boldsymbol{K}}(\theta_j) \tag{30}$$

where

$$\widehat{\boldsymbol{K}}(\theta) = \boldsymbol{B}(\theta)\widehat{\boldsymbol{U}}_{\boldsymbol{n}}\widehat{\boldsymbol{U}}_{\boldsymbol{n}}^{\mathrm{H}}\boldsymbol{B}(\theta) \tag{31}$$

Assuming true subspaces (i.e. $\widehat{\boldsymbol{U}}_{\boldsymbol{n}} = \boldsymbol{U}_{\boldsymbol{n}}$) and excluding the trivial solution $\boldsymbol{m} = \boldsymbol{0}$, it is clear that one solution of problem $(\mathcal{P}_1)$ is attained when $\boldsymbol{m} = \boldsymbol{c}$ and $[\bar{\theta}_1 \ldots \bar{\theta}_q] = [\theta_1 \ldots \theta_q] = \boldsymbol{\Theta}$. Said differently, $\boldsymbol{S}(\boldsymbol{\Theta})$ admits a null space of dimension 1 spanned by the vector of coupling parameters, $\boldsymbol{c}$.
In any case, this is a multidimensional problem in the AoA parameters, and a number of papers have resorted to an alternative and sub-optimal problem, namely

$$(\mathcal{P}_2): \min_{\boldsymbol{m},\theta} \boldsymbol{m}^{\mathrm{H}} \widehat{\boldsymbol{K}}(\theta)\boldsymbol{m} \tag{32}$$

The sub-optimality here has a nice interpretation: It is "as if" the coupling parameters are treated to be angular-dependent and therefore, one does not acknowledge that the vector of coupling parameters $\boldsymbol{c}$ is fixed for any θ. Consequently, the objective function in $(\mathcal{P}_2)$ would have been a reasonable choice if the coupling parameters are a function of θ, i.e. $\boldsymbol{c} = \boldsymbol{c}(\theta)$. *Surprisingly, a problem involving angular-dependent coupling parameters suggests a computationally less optimisation problem in terms of the AoA parameters.* Indeed, this approach is sub-optimal when $\boldsymbol{c}$ is independent of θ.

3.3 Related Work

As mentioned in the introduction, some methods make use of the MUSIC algorithm to estimate the AoAs of multiple sources in the presence of mutual coupling. We provide a brief review of these algorithms in this subsection. The RARE algorithm in [56, 59] maximizes the following cost function through peak searching

$$f_{\mathrm{RARE}}(\theta) = \frac{1}{\det\{\widehat{\boldsymbol{K}}(\theta)\}} \tag{33}$$

The peaks of the RARE cost function are estimates of the AoAs. Indeed, this is a reasonable cost function, since equation (27) tells us that $\boldsymbol{K}(\theta)$ admits a null space when $\theta \in \boldsymbol{\Theta}$. The identifiability conditions of RARE is: $\{p+q \leq N \text{ and } p \leq \frac{N}{2}\}$ (See [60]). This condition should be distinguished from the one in [56], since the RARE method developed there was for partly calibrated arrays. The paper in [60] develop a Recursive-RARE algorithm to optimize problem $(\mathcal{P}_1)$ via multiple 1-D searches. They also assume $p \leq \frac{N}{2}$. Furthermore, the method in [61] is similar to the RARE cost function, in the sense that a determinant criterion is used, however it involves a different type of matrix. Nevertheless, the identifiability conditions of the method in [61] is: $\{2p+q \leq N+1\}$. It should be noted that this condition also implies that p could not exceed $\frac{N}{2}$. In addition, the method in [62] propose to maximize the following cost function

$$f_{\text{MinEig}}(\theta) = \frac{1}{\lambda_{\min}\{\widehat{\boldsymbol{K}}(\theta)\}} \tag{34}$$

where $\lambda_{\min}\{\}$ stands for the minimum eigenvalue. Whether stated in [62], or not, the above cost function is a solution of $(\mathcal{P}_2)$ under a norm constraint on $\boldsymbol{m}$. Moreover, a linear constraint (LC) was imposed on $\boldsymbol{m}$, i.e. $\boldsymbol{m}^{\text{H}}\boldsymbol{e_1} = 1$ to optimize $(\mathcal{P}_2)$ in [63], which gives the following cost function

$$f_{\text{LC}}(\theta) = \boldsymbol{e_1^{\text{H}}}\widehat{\boldsymbol{K}}^{-1}(\theta)\boldsymbol{e_1} \tag{35}$$

The identifiability conditions for this method is similar to the one in RARE. To the best of our knowledge, the reason of condition $p \leq \frac{N}{2}$ is not so clear through the literature. One aspect of this paper is to have a clear understanding on why this inequality has to be satisfied for all the stated algorithms (see ***Theorem 3*** and ***Consequence 1***). Furthermore, we propose an algorithm that is able to estimate the AoAs of multiple sources in the presence of mutual coupling, even when $p > \frac{N}{2}$.

4 The Proposed Mutual Coupling Agnostic Algorithm

This section presents the algorithm that could estimate the angles-of-arrival of multiple sources, in the presence of mutual coupling. The algorithm is based on optimising the cost function given in $(\mathcal{P}_2)$, under a suitable constraint. The constraint is based on the following theorems:

4.1 Introductory Theorems

Theorem 2: *Let $\boldsymbol{\alpha}_p = [\alpha_0, \alpha_1 \dots \alpha_{p-1}]^T$ and $\boldsymbol{a} = [1, z \dots z^{N-1}]^T$. Define the corresponding matrix $\boldsymbol{T}(\boldsymbol{\alpha}_p)$. Then for any $1 \leq p \leq N$, the following holds*

$$\boldsymbol{T}(\boldsymbol{\alpha}_p)\boldsymbol{a} = g(z, \boldsymbol{\alpha}_p)\boldsymbol{a} - \boldsymbol{M}_p\widetilde{\boldsymbol{\alpha}}_p \tag{36}$$

where the polynomial $g(z, \boldsymbol{\alpha})$ *is given by*

$$g(z, \boldsymbol{\alpha}_p) = \alpha_0 + \sum_{k=1}^{p-1} \alpha_k (z^k + z^{-k}) \tag{37}$$

The matrix $\boldsymbol{M}_p \in \mathbb{C}^{N \times (p-1)}$ *is defined as*

$$\boldsymbol{M}_p = \begin{bmatrix} \boldsymbol{U}_p \\ \hline \boldsymbol{0} \end{bmatrix} + \begin{bmatrix} \boldsymbol{0} \\ \hline z^{N-1} \boldsymbol{J}_{p-1} \boldsymbol{U}_p^* \end{bmatrix} \tag{38}$$

where $\boldsymbol{U}_p \in \mathbb{C}^{(p-1) \times (p-1)}$ *is written as*

$$\left(\boldsymbol{U}_p\right)_{i,j} = \begin{cases} z^{-(j-i+1)} & \text{if } j \geq i \\ 0 & \text{else} \end{cases} \tag{39}$$

and $\widetilde{\boldsymbol{\alpha}}_p = [\alpha_1, \alpha_2 \dots \alpha_{p-1}]^{\mathrm{T}}$.

Proof. See Appendix A. □

Theorem 2 is key to ***Theorem 3***, which comes next:

Theorem 3: *Let* $\boldsymbol{a} = [1, z \dots z^{N-1}]^T$ *and* $\boldsymbol{B}_p = \mathcal{G}_p(\boldsymbol{a})$. *Then,* $\boldsymbol{B}_p$ *has the following spectral characteristics:*

1. *If* $p \leq \frac{N+1}{2}$, *then* $\boldsymbol{B}_p$ *is full column rank.*
2. *If* $p = \frac{N+2}{2}$ *and* z *is an* N^{th} *unit root (i.e.* $z^N = 1$*) then* $\text{rank}(\boldsymbol{B}_p) = \frac{N}{2}$. *The null space is given in equation* (121). *Otherwise, it is full column rank.*
3. *We distinguish 2 cases when* $p > \frac{N+2}{2}$*:*
   (a) N is even:
      i. *If* $z^N \neq \pm 1$, *then* $\boldsymbol{B}_p$ *is full column rank.*
      ii. *If* $z^N = -1$, *then* $\text{rank}(\boldsymbol{B}_p) = \frac{N}{2} + 1$. *The null space is given in equation* (140).
      iii. *If* $z^N = 1$, *then* $\text{rank}(\boldsymbol{B}_p) = \frac{N}{2}$. *The null space is given in equation* (144).
   (b) N is odd:
      i. *If* $z^N = \pm 1$, *then* $\text{rank}(\boldsymbol{B}_p) = \frac{N+1}{2}$. The null space is given in equation (145).
      ii. *Otherwise,* $\boldsymbol{B}_p$ *is full column rank.*

Proof. See Appendix B. □

For ULA configurations, a direct consequence of ***Theorem 3*** is the following:

Consequence 1: *For ULA type configurations, i.e.* $\boldsymbol{a}(\theta) = [1, z_\theta, \dots z_\theta^{N-1}]^T$ *with* $z_\theta = e^{-j2\pi \frac{d}{\lambda} sin(\theta)}$. *Define the following sets*

$$\boldsymbol{\Theta_+} = \left\{ sin^{-1}(\frac{k\lambda}{Nd}), \quad k = -\frac{N}{2} \dots \frac{N}{2} \right\} \tag{40}$$

$$\boldsymbol{\Theta_-} = \left\{ sin^{-1}(\frac{(k+\frac{1}{2})\lambda}{Nd}), \quad k = -\frac{N}{2} \dots \frac{N}{2} \right\} \tag{41}$$

$$\boldsymbol{\Theta_\pm} = \left\{ \boldsymbol{\Theta_+} \cup \boldsymbol{\Theta_-} \right\} \tag{42}$$

The matrix $\boldsymbol{B}(\theta) = \mathcal{G}_p(\boldsymbol{a}(\theta))$ *has the following characteristics:*

- *If* $p < \frac{N+2}{2}$, *the matrix* $\boldsymbol{B}(\theta)$ *is full column rank.*
- *When* $p \geq \frac{N+2}{2}$, *we distinguish the following cases:*
  - *If* N *is even and* $\theta \in \boldsymbol{\Theta_+}$, *then* $\text{rank}(\boldsymbol{B}(\theta)) = \frac{N}{2}$.
  - *If* N *is even and* $\theta \in \boldsymbol{\Theta_-}$, *then* $\text{rank}(\boldsymbol{B}(\theta)) = \frac{N}{2} + 1$.
  - *If* N *is odd and* $\theta \in \boldsymbol{\Theta_\pm}$, *then* $\text{rank}(\boldsymbol{B}(\theta)) = \frac{N+1}{2}$.
  - *Else* $\boldsymbol{B}(\theta)$ *is full column rank.*

Proof. See Appendix C. □

One purpose of this paper is to understand the behaviour of matrix $\boldsymbol{B}(\theta)$ as function of θ. Let $\nu_1 \leq \nu_2 \leq \dots \leq \nu_p$ be the eigenvalues of $\boldsymbol{B}^{\text{H}}(\theta)\boldsymbol{B}(\theta)$. In order to partially verify ***Consequence 1***, we have depicted two figures where $p > \frac{N+2}{2}$. In Fig. 1a, we fix $N = 8$ (even) and $p = 7$. The red and green dashed vertical lines correspond to angles in $\boldsymbol{\Theta_+}$ and $\boldsymbol{\Theta_-}$, respectively. Observe that when θ approaches angles in $\boldsymbol{\Theta_+}$, we have three eigenvalues, i.e. ν_1, ν_2, and ν_3, dropping to zero. This implies that, when $\theta \in \boldsymbol{\Theta_+}$, the rank of $\boldsymbol{B}(\theta)$ is $p - 3 = 4 = \frac{N}{2}$. However, when when $\theta \in \boldsymbol{\Theta_-}$, only two eigenvalues, namely ν_1 and ν_2, go to zero. In this case, the rank of $\boldsymbol{B}(\theta)$ is $p-2 = 5 = \frac{N}{2}+1$. Also note that ν_4 is strictly positive. In Fig. 1b, we fix $N = 9$ (odd) and $p = 8$. Again, ν_4 is strictly positive. When $\theta \in \boldsymbol{\Theta_\pm}$, three eigenvalues go to zero, implying that the rank of $\boldsymbol{B}(\theta)$ is $p - 3 = 5 = \frac{N+1}{2}$.

4.2 Algorithm Derivation

The previous subsection reveals an important phenomenon of matrix $\boldsymbol{B}(\theta)$. According to ***Consequence 1***, if $\theta_k \in \boldsymbol{\Theta_\pm}$ and $p \geq \frac{N+2}{2}$, then $\boldsymbol{B}(\theta_k)$ admits a null-space. Therefore, optimising the cost function given in $(\mathcal{P}_2)$, without choosing an appropriate constraint, gives false AoAs. Mathematically speaking, the cost

function in $(\mathcal{P}_2)$ is exactly zero for all $\theta_k \in \boldsymbol{\Theta_{\pm}}$ when $p \geq \frac{N+2}{2}$. To circumvent this issue, we form the following optimisation problem

$$\begin{cases} \underset{\boldsymbol{m},\theta}{\text{minimize}} & \boldsymbol{m}^{\mathrm{H}}\widehat{\boldsymbol{K}}(\theta)\boldsymbol{m} \\ \text{subject to} & \boldsymbol{e}_1^{\mathrm{H}}\boldsymbol{B}(\theta)\boldsymbol{m} = 1 \end{cases} \tag{43}$$

It is easy to see that, for any θ, the trivial solution $\boldsymbol{m} = \boldsymbol{0}$ and the vectors that lie in the null space of $\boldsymbol{B}(\theta)$ (i.e. $\boldsymbol{B}(\theta)\boldsymbol{m} = 0$) are not feasible solutions because they do not satisfy the constraint. Therefore, optimising the above problem will exclude the latter false solutions.

The *Lagrangian* function corresponding to the optimisation problem in (43) is the following:

$$\mathcal{L}(\boldsymbol{m}, \nu) = \boldsymbol{m}^{\mathrm{H}}\widehat{\boldsymbol{K}}(\theta)\boldsymbol{m} - \nu\big(\boldsymbol{e}_1^{\mathrm{H}}\boldsymbol{B}(\theta)\boldsymbol{m} - 1\big) \tag{44}$$

Setting the derivative of $\mathcal{L}(\boldsymbol{m}, \nu)$ with respect to $\boldsymbol{m}$ to 0, we get

$$\frac{\partial}{\partial \boldsymbol{m}}\mathcal{L}(\boldsymbol{m}, \nu) = 2\widehat{\boldsymbol{K}}(\theta)\boldsymbol{m} - \nu\boldsymbol{B}^{\mathrm{H}}(\theta)\boldsymbol{e}_1 = 0 \tag{45}$$

Equation (45) gives the optimal coupling parameters, $\boldsymbol{m^o}$, for a given θ, in terms of the optimal Lagrangian multiplier ν^o as

$$\boldsymbol{m^o} = \frac{\nu^o}{2}\widehat{\boldsymbol{K}}^{-1}(\theta)\boldsymbol{B}^{\mathrm{H}}(\theta)\boldsymbol{e}_1 \tag{46}$$

It is easy to prove that

$$\boldsymbol{B}^{\mathrm{H}}(\theta)\boldsymbol{e}_1 = \boldsymbol{a}_p^*(\theta) \tag{47}$$

where $\boldsymbol{a}_p(\theta)$ is a $p \times 1$ vector defined as in equation (3). The expression of ν^o is obtained by plugging equations (46) and (47) in the constraint of the problem in (43), viz.

$$\nu^o = \frac{2}{\boldsymbol{a}_p^{\mathrm{T}}(\theta)\widehat{\boldsymbol{K}}^{-1}(\theta)\boldsymbol{a}_p^*(\theta)} \tag{48}$$

Therefore, $\boldsymbol{m^o}$ is now given as

$$\boldsymbol{m^o} = \frac{\widehat{\boldsymbol{K}}^{-1}(\theta)\boldsymbol{a}_p^*(\theta)}{\boldsymbol{a}_p^{\mathrm{T}}(\theta)\widehat{\boldsymbol{K}}^{-1}(\theta)\boldsymbol{a}_p^*(\theta)} \tag{49}$$

Substituting $\boldsymbol{m^o}$ in the objective function of (43), the q AoAs could be estimated as follows

$$\big\{\widehat{\theta}_k\big\}_{k=1}^{q} = \arg\min_{\theta} \frac{1}{f(\theta)} \tag{50a}$$

where

$$f(\theta) = \boldsymbol{a}_p^{\mathrm{T}}(\theta)\widehat{\boldsymbol{K}}^{-1}(\theta)\boldsymbol{a}_p^*(\theta) \tag{50b}$$

Note that $\widehat{\boldsymbol{K}}(\theta)$ is not invertible for the cases given in ***Consequence 1*** and when $\theta \in \boldsymbol{\Theta}$ at infinite SNR. For that, we adopt diagonal loading as done in [64], namely $\big(\boldsymbol{K}(\theta) + \epsilon \boldsymbol{I}\big)^{-1}$, where $\epsilon > 0$ is small. Additionally, it has been mentioned in [64] that there is generally no known method for determining the optimal value of ϵ, and it is usually determined experimentally. We have found that $\epsilon = 10^{-14}$ serves as a good value.

4.3 Properties of $f(\theta)$

For a better understanding of the behaviour of the cost function given in equation (50), we reveal some of its properties

Property 1*: For $p = 1$, i.e. no mutual coupling, the function $f(\theta)$ "boils down" to the traditional MUSIC cost function in equation* (17)*.*

Proof. Trivial. □

Property 2*: This property characterizes the null space of $\boldsymbol{K}(\theta)$ for $p + q \leq N$ as a function of θ*

$$\mathcal{N}(\boldsymbol{K}(\theta_i)) = \begin{cases} \{\boldsymbol{0}\}, & \text{if } \theta_i \notin \boldsymbol{\Theta} \cup \boldsymbol{\Theta}_{\pm} \\ \{\boldsymbol{0} \cup \boldsymbol{c}\}, & \text{if } \theta_i \in \boldsymbol{\Theta} \text{ and } \theta_i \notin \boldsymbol{\Theta}_{\pm} \\ \mathcal{N}\big(\boldsymbol{B}(\theta_i)\big), & \text{if } \theta_i \notin \boldsymbol{\Theta} \text{ and } \theta_i \in \boldsymbol{\Theta}_{\pm} \\ \mathcal{N}\big(\boldsymbol{B}(\theta_i)\big) \cup \{\boldsymbol{c}\}, & \text{if } \theta_i \in \boldsymbol{\Theta} \cap \boldsymbol{\Theta}_{\pm} \end{cases} \tag{51}$$

Note that if $p < \frac{N+2}{2}$, then $\mathcal{N}\big(\boldsymbol{B}(\theta_i)\big) = \{\boldsymbol{0}\}$. Also note that this property assumes true subspaces, i.e. $\widehat{\boldsymbol{K}}(\theta) = \boldsymbol{K}(\theta) = \boldsymbol{B}^{\boldsymbol{H}}(\theta)\boldsymbol{U_n}\boldsymbol{U_n^H}\boldsymbol{B}(\theta)$.

Proof.

- The first two cases are a direct consequence of equation (27).
- The third case is a result of ***Consequence 1*** and equation (27).
- As for the fourth case, assume that the sets $\boldsymbol{\Theta}$ and $\boldsymbol{\Theta}_{\pm}$ overlap and $\frac{N+2}{2} \leq p < N$. Let $\theta_i \in \boldsymbol{\Theta} \cap \boldsymbol{\Theta}_{\pm}$. Therefore, $\boldsymbol{K}(\theta_i)\boldsymbol{\alpha} = \boldsymbol{0}$ only when $\boldsymbol{\alpha} \in \mathcal{N}\big(\boldsymbol{B}(\theta_i)\big)$ or $\boldsymbol{\alpha} = \boldsymbol{c}$. It suffices to prove that the set $\boldsymbol{c}$ is linearly independent from the span of $\mathcal{N}\big(\boldsymbol{B}(\theta_i)\big)$.
  Let Δ be the dimension of $\mathcal{N}\big(\boldsymbol{B}(\theta_i)\big)$. Furthermore, let $\boldsymbol{\gamma} \in \mathbb{C}^{(\Delta+1)\times 1}$ be an arbitrary vector and $\boldsymbol{E} \in \mathbb{C}^{p\times(\Delta+1)}$ be a matrix where the first Δ columns span $\mathcal{N}\big(\boldsymbol{B}(\theta_i)\big)$ and the last column is the vector $\boldsymbol{c}$. It remains to show that $\boldsymbol{E}\boldsymbol{\gamma} = \boldsymbol{0} \Rightarrow \boldsymbol{\gamma} = \boldsymbol{0}$. Under the assumption that $p < N$ and using the structure

of the null space of $\boldsymbol{B}(\theta_i)$ given in equations (121), (140), (144), and (145), one could easily verify that the second row of $\boldsymbol{E}$ is given as

$$[\underbrace{0 \dots 0}_{\Delta}, c_1] \tag{52}$$

which implies that the last element of $\boldsymbol{\gamma}$ is 0, since by construction $c_1 \neq 0$, for $p \geq 2$. Hence, $\boldsymbol{\gamma} = \mathbf{0}$ because the first Δ columns of $\boldsymbol{E}$ are linearly independent.

□

***Property 3**: Assuming true subspaces (i.e. $\widehat{\boldsymbol{U}}_{\boldsymbol{n}} = \boldsymbol{U}_{\boldsymbol{n}}$), the function $f(\theta)$ is bounded when $\theta \notin \boldsymbol{\Theta}$ and unbounded when $\theta \in \boldsymbol{\Theta}$.*

Proof. Let the function $f_\epsilon(\theta)$ be defined as follows:

$$f_\epsilon(\theta) = \boldsymbol{a}_p^{\mathrm{T}}(\theta)\big(\boldsymbol{K}(\theta) + \epsilon \boldsymbol{I}\big)^{-1}\boldsymbol{a}_p^*(\theta) \tag{53}$$

and therefore

$$\lim_{\epsilon \to 0} f_\epsilon(\theta) = f(\theta) \tag{54}$$

By spectral decomposition,

$$\boldsymbol{K}(\theta) = \boldsymbol{V}\boldsymbol{\Phi}\boldsymbol{V}^{\mathrm{H}} \tag{55}$$

where the k^{th} column of $\boldsymbol{V}$ is the k^{th} normalized eigenvector[1] of $\boldsymbol{K}(\theta)$, denoted as $\boldsymbol{v}_k$ and its corresponding eigenvalue is the k^{th} smallest eigenvalue found in the k^{th} diagonal entry of $\boldsymbol{\Phi}$, denoted as λ_k. We could then express $f_\epsilon(\theta)$ as

$$f_\epsilon(\theta) = \boldsymbol{a}_p^{\mathrm{T}}(\theta)\boldsymbol{V}\big(\boldsymbol{\Phi} + \epsilon \boldsymbol{I}\big)^{-1}\boldsymbol{V}^{\mathrm{H}}\boldsymbol{a}_p^*(\theta) \tag{56}$$

- When $\theta \notin \boldsymbol{\Theta}$, we distinguish two sub-cases:
  - If $\theta \notin \boldsymbol{\Theta}_{\pm}$, then $\boldsymbol{K}(\theta)$ is full rank according to ***Property 2*** and hence $\lambda_k > 0$ for all k, so

$$f(\theta) = \|\boldsymbol{\Phi}^{-1/2}\boldsymbol{V}^{\mathrm{H}}\boldsymbol{a}_p^*(\theta)\|^2 < \infty \tag{57}$$

  - If $\theta \in \boldsymbol{\Theta}_{\boldsymbol{\pm}}$, then $\boldsymbol{K}(\theta)$ is full rank (if $p < \frac{N+2}{2}$) and the preceeding argument holds. However, if $p \geq \frac{N+2}{2}$, then $\boldsymbol{K}(\theta)$ admits the same null-space as that of $\boldsymbol{B}(\theta)$ according to ***Property 2***. As before, let Δ be the dimension of $\boldsymbol{K}(\theta)$, therefore $f_\epsilon(\theta)$ behaves as

$$f_\epsilon(\theta) \backsim \sum_{k=1}^{\Delta} \frac{1}{\lambda_k + \epsilon}\|\boldsymbol{a}_p^{\mathrm{T}}(\theta)\boldsymbol{v}_k\|^2 = \sum_{k=1}^{\Delta} \frac{1}{\epsilon}\|\boldsymbol{e}_1^{\mathrm{T}}\boldsymbol{B}(\theta)\boldsymbol{v}_k\|^2 \tag{58}$$

    Note that $\{\boldsymbol{v}_k\}_{k=1}^{\Delta}$ span the null space of $\boldsymbol{B}(\theta)$ and therefore $\boldsymbol{B}(\theta)\boldsymbol{v}_k = \mathbf{0}$. So, $f_\epsilon(\theta) = f(\theta) = 0 < \infty$.

[1] Indeed, $\boldsymbol{V}$ and $\boldsymbol{\Phi}$ are functions of θ. This is omitted for the sake of compact exposition.

- When $\theta \in \boldsymbol{\Theta}$, we also distinguish the same sub-cases:
  - If $\theta \notin \boldsymbol{\Theta}_{\pm}$, then there is only one singularity in $\boldsymbol{K}(\theta)$ according to ***Property 2***, i.e. $\lambda_1 = 0, \boldsymbol{v}_1 = \frac{\boldsymbol{c}}{\|\boldsymbol{c}\|}$, and $\lambda_k > 0$ for all $k \geq 2$. Hence

$$f_\epsilon(\theta) \backsim \frac{1}{\epsilon} \frac{\|\boldsymbol{a}_p^{\mathrm{T}}(\theta)\boldsymbol{c}\|^2}{\|\boldsymbol{c}\|^2} \tag{59}$$

    Notice that the term $\boldsymbol{a}_p^{\mathrm{T}}(\theta)\boldsymbol{c}$ is a polynomial of degree $p-1$ evaluated at the unit circle. For a polynomial with non-zero coefficients to have zeros on the unit-circle, the coefficient vector $\boldsymbol{c}$ must be conjugate-symmetric [68], which is not the case according to equation (10). Therefore, $\boldsymbol{a}_p^{\mathrm{T}}(\theta)\boldsymbol{c} \neq 0$ and thus

$$f(\theta) = \lim_{\epsilon \to 0} f_\epsilon(\theta) = \infty \tag{60}$$

  - If $\theta \in \boldsymbol{\Theta}_{\pm}$, then the null space of $\boldsymbol{K}(\theta)$ is spanned by $\Delta + 1$ vectors given in ***Property 2***, and we have

$$f_\epsilon(\theta) \backsim \sum_{k=1}^{\Delta} \frac{1}{\epsilon} \|\boldsymbol{e}_1^{\mathrm{T}} \boldsymbol{B}(\theta) \boldsymbol{v}_k\|^2 + \frac{1}{\epsilon} \frac{\|\boldsymbol{a}_p^{\mathrm{T}}(\theta)\boldsymbol{c}\|^2}{\|\boldsymbol{c}\|^2} \tag{61}$$

    Using the same argument as before, as ϵ goes to zero, the 1^{st} term of the above expression goes to zero, whereas the 2^{nd} term goes to ∞.

□

Property 4*: The condition so that $f(\theta)$ uniquely identifies the AoAs is that $p + q \leq N$.*

Proof. This is so because the cost function in equation (50) depends on the inversion of $\boldsymbol{K}(\theta)$. Hence, in the case where $\theta \notin \boldsymbol{\Theta} \cup \boldsymbol{\Theta}_{\pm}$, the matrix $\boldsymbol{U}_{\boldsymbol{n}}^{\mathrm{H}} \boldsymbol{B}(\theta)$ is full column rank when $p \leq N - q$. As for the case when $\theta \in \boldsymbol{\Theta} \cup \boldsymbol{\Theta}_{\pm}$, we have the argument in ***Property 3***. □

Remark*:* The existing methods in equations (33), (34), and (35) can not identify the true AoAs, when the number of coupling parameters $p > \frac{N}{2}$. According to ***Property 2***, the cost functions of these exsiting methods would yield peaks whenever $\theta \in \boldsymbol{\Theta}_{\pm}$ and $p > \frac{N}{2}$. One could not, simply, remove these peaks because they would affect the estimation, when the true AoAs are close to those in $\boldsymbol{\Theta}_{\pm}$.

5 MSE Analysis

It is well known that the noise subspace could be decomposed into two parts:

$$\widehat{\boldsymbol{U}}_{\boldsymbol{n}} = \boldsymbol{U}_{\boldsymbol{n}} + \widetilde{\boldsymbol{U}}_{\boldsymbol{n}} \tag{62}$$

where the first part, $\boldsymbol{U_n}$, is the true noise subspace and the second one, $\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}$, is the error term. Using this decomposition and other asymptotic properties (i.e. for large L or high SNR) which will appear in this section, we wish to derive an asymptotic MSE expression for the AoA estimates obtained from equation (50). In other words, we seek an asymptotic expression of $\mathbb{E}\{(\widetilde{\theta}_k)^2\}$, where $\widetilde{\theta}_k$ is the error part

$$\widehat{\theta}_k = \theta_k + \widetilde{\theta}_k \tag{63}$$

Since $\{\widehat{\theta}_k\}_{k=1}^{q}$ are minimum points of $f^{-1}(\theta)$, then

$$\frac{\partial f^{-1}(\widehat{\theta}_k)}{\partial \theta} \triangleq \frac{\partial f^{-1}(\theta)}{\partial \theta}\Big|_{\theta=\widehat{\theta}_k} = 0 \tag{64}$$

As done in [19], since $\widehat{\theta}_k$ is an estimate of θ_k, we could, asymptotically, expand the above derivative in the neighborhood of the true θ_k using *Taylor* series

$$\frac{\partial f^{-1}(\widehat{\theta}_k)}{\partial \theta} = \frac{\partial f^{-1}(\theta_k)}{\partial \theta} + \frac{\partial^2 f^{-1}(\theta_k)}{\partial \theta^2}(\widehat{\theta}_k - \theta_k) + \dots \tag{65}$$

which gives an approximate expression of the error $\widetilde{\theta}_k = \widehat{\theta}_k - \theta_k$

$$\widetilde{\theta}_k \backsimeq -\frac{\frac{\partial f^{-1}(\theta_k)}{\partial \theta}}{\frac{\partial^2 f^{-1}(\theta_k)}{\partial \theta^2}} = -\frac{f'(\theta_k)}{f''(\theta_k) - 2\frac{(f'(\theta_k))^2}{f(\theta_k)}} \tag{66}$$

where $f'(\theta_k)$ and $f''(\theta_k)$ are the 1^{st} and 2^{nd} order derivatives of $f(\theta)$ evaluated at point θ_k, respectively.

Property 5*: The derivatives $f'(\theta)$ and $f''(\theta)$ are given as*

$$f'(\theta) = g_1(\theta) + g_2(\theta) \tag{67}$$
$$f''(\theta) = h_1(\theta) + h_2(\theta) + h_3(\theta) \tag{68}$$

where $g_1(\theta)$ and $g_2(\theta)$ are given in equation (151) *and $h_1(\theta)$, $h_2(\theta)$, and $h_3(\theta)$ are given in equation* (152)*.*

Proof. See Appendix D. □

The expressions of $f'(\theta)$ and $f''(\theta)$ in equations (67) and (68), respectively, turn out to be too complicated to analyze the error in equation (66). However, some simplifications could be done, asymptotically, thanks to the following theorem

Theorem 4*: Let λ_j and $\boldsymbol{v}_j$ be the j^{th} smallest eigenvalue and its corresponding*

normalized eigenvector of $\boldsymbol{K}(\theta_k)$. Similarly, define $\widehat{\lambda}_j$ and $\widehat{\boldsymbol{v}}_j$ for $\widehat{\boldsymbol{K}}(\theta_k)$. The smallest eigenvalue $\widehat{\lambda}_1$ and its eigenvector $\widehat{\boldsymbol{v}}_1$ could be approximated as

$$\widehat{\lambda}_1 = \frac{1}{\|\boldsymbol{c}\|^2}\boldsymbol{c}^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}}(\theta_k)\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}\boldsymbol{P}_k^{\perp}\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}^{\mathrm{H}}\boldsymbol{B}(\theta_k)\boldsymbol{c} + \mathcal{O}(\|\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}\|^3) \tag{69}$$

$$\widehat{\boldsymbol{v}}_1 = \frac{1}{\|\boldsymbol{c}\|}\Big(\boldsymbol{c} - \sum_{i=\Delta+2}^{p}\frac{\boldsymbol{v}_i^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}}(\theta_k)\boldsymbol{U}_{\boldsymbol{n}}\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}^{\mathrm{H}}\boldsymbol{B}(\theta_k)\boldsymbol{c}}{\lambda_i}\boldsymbol{v}_i\Big) + \mathcal{O}(\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}^2) \tag{70}$$

where $\boldsymbol{P}_k^{\perp} = \boldsymbol{I} - \boldsymbol{P}_k$ and

$$\boldsymbol{P}_k = \boldsymbol{U}_{\boldsymbol{n}}^{\mathrm{H}}\boldsymbol{B}(\theta_k)\boldsymbol{K}^{+}(\theta_k)\boldsymbol{B}^{\mathrm{H}}(\theta_k)\boldsymbol{U}_{\boldsymbol{n}} \tag{71}$$

*and Δ is the dimension of $\mathcal{N}\big(\boldsymbol{B}(\theta_k)\big)$, which is 0 when $p \leq \frac{N+2}{2}$ or $\{p > \frac{N+2}{2}$ and $\theta_k \notin \boldsymbol{\Theta}_{\pm}\}$ and non-zero otherwise (according to **Consequence 1**). Note that $\mathcal{O}(\|\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}\|^k)$ and $\mathcal{O}(\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}^k)$ are scalar and vector terms, respectively, in which $\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}$ appears k times in each term.*

Proof. See Appendix E. □

This theorem reveals a behaviour of $\widehat{\lambda}_1$, i.e. it acts as $\mathcal{O}(\|\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}\|^2)$. Using ***Theorem 4***, and some straightforward algebra, we have the following asymptotic approximations of $f(\theta)$, $f'(\theta)$, and $f''(\theta)$

$$f(\theta) = \frac{1}{\widehat{\lambda}_1\|\boldsymbol{c}\|^2}\mu_k + \mathcal{O}\big(\|\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}\|^{-1}\big) \tag{72}$$

$$f'(\theta) = -\frac{2}{\widehat{\lambda}_1^2\|\boldsymbol{c}\|^4}\mu_k\mathrm{Re}\{\widetilde{\rho}_k\} + \mathcal{O}\big(\|\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}\|^{-2}\big) \tag{73}$$

$$f''(\theta) = \frac{2}{\widehat{\lambda}_1^3\|\boldsymbol{c}\|^4}\mu_k\big(\frac{4}{\|\boldsymbol{c}\|^2}(\mathrm{Re}\{\widetilde{\rho}_k\})^2 - \widehat{\lambda}_1\upsilon_k\Big) + \mathcal{O}\big(\|\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}\|^{-3}\big) \tag{74}$$

where

$$\mu_k = \|\boldsymbol{c}^{\mathrm{H}}\boldsymbol{a}_p^*(\theta_k)\|^2 \tag{75}$$

$$\widetilde{\rho}_k = \boldsymbol{c}^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}}(\theta_k)\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}\boldsymbol{P}_k^{\perp}\boldsymbol{U}_{\boldsymbol{n}}^{\mathrm{H}}\boldsymbol{D}(\theta_k)\boldsymbol{c} \tag{76}$$

$$\upsilon_k = \boldsymbol{c}^{\mathrm{H}}\boldsymbol{D}^{\mathrm{H}}(\theta_k)\boldsymbol{U}_{\boldsymbol{n}}\boldsymbol{P}_k^{\perp}\boldsymbol{U}_{\boldsymbol{n}}^{\mathrm{H}}\boldsymbol{D}(\theta_k)\boldsymbol{c} \tag{77}$$

Substituting these expressions in equation (66), we arrive at

$$\widetilde{\theta}_k \simeq \frac{\mathrm{Re}\{\widetilde{\rho}_k\}}{\upsilon_k} \tag{78}$$

In order to proceed, we use the following lemma, which gives the probabilistic distribution of the columns of $\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}$

Lemma 1: *Let $\widetilde{\boldsymbol{n}}_i$ be the i^{th} column of $\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}$. Asymptotically, the vectors $\boldsymbol{U_s}\boldsymbol{U_s}^H\widetilde{\boldsymbol{n}}_i$ are jointly Gaussian distributed with zero means and covariance matrices given by*

$$\mathbb{E}\Big\{\big(\boldsymbol{U_s}\boldsymbol{U_s}^{\mathrm{H}}\widetilde{\boldsymbol{n}}_i\big)\big(\boldsymbol{U_s}\boldsymbol{U_s}^{\mathrm{H}}\widetilde{\boldsymbol{n}}_j\big)^{\mathrm{H}}\Big\} = \frac{\sigma^2}{L}\boldsymbol{U}\delta_{i,j} \tag{79}$$

$$\mathbb{E}\Big\{\big(\boldsymbol{U_s}\boldsymbol{U_s}^{\mathrm{H}}\widetilde{\boldsymbol{n}}_i\big)\big(\boldsymbol{U_s}\boldsymbol{U_s}^{\mathrm{H}}\widetilde{\boldsymbol{n}}_j\big)^{\mathrm{T}}\Big\} = \boldsymbol{0} \tag{80}$$

where

$$\boldsymbol{U} = \boldsymbol{U_s}\boldsymbol{\Sigma_s}\big(\boldsymbol{\Sigma_s} - \sigma^2\boldsymbol{I}\big)^{-2}\boldsymbol{U_s}^{\mathrm{H}} \tag{81}$$

Proof. See [19]. □

This lemma is key to the following theorem, which gives the MSE expression $\mathbb{E}\big\{(\widetilde{\theta}_k)^2\big\}$

Theorem 5*: The estimates $\big\{\widehat{\theta}_k\big\}_{k=1}^{q}$ estimated through $f(\theta)$ by equation* (50) *are asymptotically unbiased. Furthermore, the MSE expression $\mathbb{E}\big\{(\widetilde{\theta}_k)^2\big\}$ is given as*

$$\mathbb{E}\big\{(\widetilde{\theta}_k)^2\big\} \triangleq \mathrm{var}_f^{(p)}(\widehat{\theta}_k) = \frac{\sigma^2}{2L}\frac{\bar{\boldsymbol{a}}^{\mathrm{H}}(\theta_k)\boldsymbol{U}\bar{\boldsymbol{a}}(\theta_k)}{\bar{\boldsymbol{d}}^{\mathrm{H}}(\theta_k)\boldsymbol{U_n}\boldsymbol{P}_k^{\perp}\boldsymbol{U_n}^{\mathrm{H}}\bar{\boldsymbol{d}}(\theta_k)} \tag{82}$$

where $\bar{\boldsymbol{a}}(\theta_k)$ and $\boldsymbol{U}$ are defined in equations (18) *and* (81)*, respectively. Also, $\bar{\boldsymbol{d}}(\theta_k) = \frac{\partial\bar{\boldsymbol{a}}(\theta)}{\partial\theta}\Big|_{\theta=\widehat{\theta}_k}$.*

Proof. See Appendix F. □

It is interesting and easy to see that when $p = 1$, the above MSE expression coincides with the MSE expression of MUSIC derived in [19]. In other words, if $p = 1$, we have $\bar{\boldsymbol{a}}(\theta_k) = \boldsymbol{a}(\theta_k)$, $\bar{\boldsymbol{d}}(\theta_k) = \boldsymbol{d}(\theta_k)$, and $\boldsymbol{P}_k^{\perp} = \boldsymbol{I}$, hence

$$\mathrm{var}_f^{(1)}(\widehat{\theta}_k) = \frac{\sigma^2}{2L}\frac{\boldsymbol{a}^{\mathrm{H}}(\theta_k)\boldsymbol{U}\boldsymbol{a}(\theta_k)}{\boldsymbol{d}^{\mathrm{H}}(\theta_k)\boldsymbol{U_n}\boldsymbol{U_n}^{\mathrm{H}}\boldsymbol{d}(\theta_k)} = \mathrm{var}_{\mathrm{MU}}(\widehat{\theta}_k;\boldsymbol{a}) \tag{83}$$

where $\mathrm{var}_{\mathrm{MU}}(\widehat{\theta}_k;\boldsymbol{a})$ is read as follows: The variance of $\widehat{\theta}_k$ obtained by MUSIC by utilising a steering vector $\boldsymbol{a}(\theta)$. We adopt this notation because the MSE expression, $\mathrm{var}_f^{(p)}(\widehat{\theta}_k)$, could also be expressed as

$$\mathrm{var}_f^{(p)}(\widehat{\theta}_k) = \Big(\frac{1}{1-\gamma_k}\Big)\mathrm{var}_{\mathrm{MU}}(\widehat{\theta}_k;\bar{\boldsymbol{a}}) \tag{84}$$

where

$$0 \leq \gamma_k = \mathcal{R}\Big(\boldsymbol{P}_k, \boldsymbol{U_n}^{\mathrm{H}}\bar{\boldsymbol{d}}(\theta_k)\Big) < 1 \tag{85}$$

where the bounds in equation (85) are due to the fact that γ_k is a *Rayleigh quotient*, which is always bounded between the minimum and maximum eigenvalues of $\boldsymbol{P}_k$. Since $\boldsymbol{P}_k$ is a projector matrix, then the eigenvalues are either 0 or 1. Note that

$\gamma_k = 1$ only when $N - q = \text{rank}(\boldsymbol{P}_k) = p - 1$, thus violating the identifiability condition given in ***Property 4***.

Observation: It is very important to observe that $\text{var}_{\text{MU}}(\widehat{\theta}_k; \bar{\boldsymbol{a}})$ appearing in equation (84) is, indeed, the MSE of $\widehat{\theta}_k$ estimated through MUSIC with *known mutual coupling parameters*. Therefore, the quantity $\frac{1}{1-\gamma_k}$ quantifies the loss of performance, or "gap" in terms of MSE, between the proposed method in equation (50) and the MUSIC algorithm with *known mutual coupling parameters*. Through exhaustive simulations, we have noticed that γ_k is increasing as a function of p, given that the coupling parameters decay as in equation (10) (See Section VIII).

6 Comparison with the Cramér-Rao Bound

The *Cramér-Rao* Bound (CRB) on the AoA estimates of a model that includes unknown mutual coupling, i.e. equation (9) was derived in [74]. The CRB is given as

$$\text{var}_{\text{CRB}}(\widehat{\theta}_k) = \frac{\sigma^2}{2L}\Big(\big[\bar{\boldsymbol{D}}^{\text{H}} \boldsymbol{P}_{\bar{\boldsymbol{A}}}^{\perp} \bar{\boldsymbol{D}} \odot \boldsymbol{R}_{\boldsymbol{ss}} \big]^{-1} \Big)_{k,k} \tag{86}$$

where $\boldsymbol{P}_{\bar{\boldsymbol{A}}}^{\perp} = \boldsymbol{I} - \boldsymbol{P}_{\bar{\boldsymbol{A}}}$ is given in equation (21) and $\bar{\boldsymbol{A}}$ is given in equation (19). Also

$$\bar{\boldsymbol{D}} = \left[\begin{array}{c|c|c} \frac{\partial \bar{\boldsymbol{a}}(\theta_1)}{\partial \theta_1} & \dots & \frac{\partial \bar{\boldsymbol{a}}(\theta_q)}{\partial \theta_q} \end{array} \right] \tag{87}$$

Following similar steps as in [19], we re-write the MSE equation, $\text{var}_f^{(p)}(\widehat{\theta}_k)$, in a way that turns out to be useful when comparing to the CRB

$$\text{var}_f^{(p)}(\widehat{\theta}_k) = \frac{\sigma^2}{2L} \frac{\big(\boldsymbol{R}_{\boldsymbol{ss}}^{-1}\big)_{k,k} + \sigma^2 \big(\boldsymbol{R}_{\boldsymbol{ss}}^{-1} (\bar{\boldsymbol{A}}^{\text{H}} \bar{\boldsymbol{A}})^{-1} \boldsymbol{R}_{\boldsymbol{ss}}^{-1} \big)_{k,k}}{\bar{\boldsymbol{d}}^{\text{H}}(\theta_k) \boldsymbol{U}_{\boldsymbol{n}} \boldsymbol{P}_k^{\perp} \boldsymbol{U}_{\boldsymbol{n}}^{\text{H}} \bar{\boldsymbol{d}}(\theta_k)} \tag{88}$$

6.1 Large Number of Antennas

We study the performance of the algorithm proposed in equation (50) in the asymptotic regime when $\frac{p}{N} \to 0$, i.e. $N \to \infty$ for fixed p. We have the following Theorem, which is a generalisation of the case with no mutual coupling in [19]

Theorem 6*: The limits of $var_{CRB}(\widehat{\theta}_k)$ and $var_f^{(p)}(\widehat{\theta}_k)$ are given as*

$$\text{var}_{\text{CRB}}(\widehat{\theta}_k) \xrightarrow[\frac{p}{N} \to 0]{} \frac{6\sigma^2}{N^3 L |\boldsymbol{h}_k^{\text{H}} \boldsymbol{c}|^2} \frac{1}{\big(\boldsymbol{R}_{\boldsymbol{ss}}\big)_{k,k}} \tag{89}$$

$$\text{var}_f^{(p)}(\widehat{\theta}_k) \xrightarrow[\frac{p}{N} \to 0]{} \frac{6\sigma^2}{N^3 L |\boldsymbol{h}_k^{\text{H}} \boldsymbol{c}|^2} \big(\boldsymbol{R}_{\boldsymbol{ss}}^{-1}\big)_{k,k} \tag{90}$$

$$\gamma_k \xrightarrow[\frac{p}{N} \to 0]{} 0 \tag{91}$$

where

$$\boldsymbol{h}_k = \boldsymbol{a}_p(\theta_k) + \boldsymbol{a}_p^*(\theta_k) - \boldsymbol{e}_1 \tag{92}$$

Proof. See Appendix G. □

Using this theorem, we have that

$$\frac{\mathrm{var}_f^{(p)}(\widehat{\theta}_k)}{\mathrm{var}_{\mathrm{CRB}}(\widehat{\theta}_k)} = \big(\boldsymbol{R_{ss}}\big)_{k,k}\big(\boldsymbol{R_{ss}^{-1}}\big)_{k,k} \tag{93}$$

and hence the CRB is attained for uncorrelated signals (i.e. $\boldsymbol{R_{ss}}$ is diagonal), when $\frac{p}{N} \to 0$.

6.2 High SNR

For high SNR and uncorrelated signals, one could show the following relation

$$\frac{\mathrm{var}_f^{(p)}(\widehat{\theta}_k)}{\mathrm{var}_{\mathrm{CRB}}(\widehat{\theta}_k)} = \Big(1 + \frac{\big((\bar{\boldsymbol{A}}^{\mathrm{H}}\bar{\boldsymbol{A}})^{-1}\big)_{k,k}}{\mathrm{SNR}_k}\Big)\Big(\frac{1}{1-\gamma_k}\Big) \tag{94}$$

where $\mathrm{SNR}_k = \frac{(\boldsymbol{R_{ss}})_{k,k}}{\sigma^2}$. For high SNR, the ratio in equation (94) is controlled by the factor $\frac{1}{1-\gamma_k}$, i.e. the "gap" between the MSE of the proposed algorithm and the CRB is $\frac{1}{1-\gamma_k}$.

7 Refining the AoA estimates by alternating minimisation

As explained in Section III.B, the optimisation problem formed in $(\mathcal{P}_2)$ is sub-optimal. This is due to the fact that it, implicitly, assumes that each AoA is exposed to different mutual coupling parameters, namely $\boldsymbol{c} = \boldsymbol{c}(\theta)$. Fortunately, problem $(\mathcal{P}_1)$ is optimal, since it forces the same coupling parameters on all the AoAs. In this section, we propose an efficient algorithm that aims at optimising problem $(\mathcal{P}_1)$.

Consider the following problem:

$$\begin{cases} \underset{\boldsymbol{m},\theta_1 \ldots \theta_q}{\text{minimize}} & \boldsymbol{m}^{\mathrm{H}}\widehat{\boldsymbol{S}}(\boldsymbol{\Theta})\boldsymbol{m} \\ \text{subject to} & \Big(\sum_{k=1}^{q} \boldsymbol{e}_1^{\mathrm{H}}\boldsymbol{B}(\theta_k)\Big)\boldsymbol{m} = 1 \end{cases} \tag{95}$$

The constraint here is a generalisation of that in problem (43) in a sense that it prevents the cost function to be zero when the AoA variables $\boldsymbol{\Theta}$ are "simultaneously" in the set $\boldsymbol{\Theta}_{\pm}$, i.e. when $\theta_1 \in \boldsymbol{\Theta}_{\pm} \ldots \theta_q \in \boldsymbol{\Theta}_{\pm}$. Following similar steps as in equations (44) till (49), the optimal coupling parameters are given as

$$\boldsymbol{m^o} = \frac{\widehat{\boldsymbol{S}}^{-1}(\boldsymbol{\Theta})\boldsymbol{A}_p^*(\boldsymbol{\Theta})\mathbb{1}_q}{\mathbb{1}_q^{\mathrm{T}}\boldsymbol{A}_p^{\mathrm{T}}(\boldsymbol{\Theta})\widehat{\boldsymbol{S}}^{-1}(\boldsymbol{\Theta})\boldsymbol{A}_p^*(\boldsymbol{\Theta})\mathbb{1}_q} \tag{96}$$

where $\boldsymbol{A}_p(\boldsymbol{\Theta})$ is similarly defined as $\boldsymbol{A}(\boldsymbol{\Theta})$ in equation (6) but of size $p \times q$. Plugging this expression of $\boldsymbol{m^o}$ in the objective function of (95), we get

$$\widehat{\boldsymbol{\Theta}} = \arg\max_{\boldsymbol{\Theta}} \left\{ \mathbb{1}_q^{\mathrm{T}} \boldsymbol{A}_p^{\mathrm{T}}(\boldsymbol{\Theta}) \widehat{\boldsymbol{S}}^{-1}(\boldsymbol{\Theta}) \boldsymbol{A}_p^{*}(\boldsymbol{\Theta}) \mathbb{1}_q \right\} \tag{97}$$

which involves a $q-$dimensional search in the AoA parameters. We, hereby, propose q "1$-$dimensional" searches done by alternating minimisations: At an iteration i, the following AoAs are estimated from previous iterations:

$$\widehat{\boldsymbol{\Theta}}_{\bar{i}} = [\widehat{\theta}_1 \dots \widehat{\theta}_{i-1}] \tag{98}$$

Estimate $\widehat{\theta}_i$ as

$$\widehat{\theta}_i = \arg\max_{\theta} \left\{ \mathbb{1}_i^{\mathrm{T}} \boldsymbol{A}_p^{\mathrm{T}}(\widehat{\boldsymbol{\Theta}}_{\bar{i}}, \theta) \widehat{\boldsymbol{S}}^{-1}(\widehat{\boldsymbol{\Theta}}_{\bar{i}}, \theta) \boldsymbol{A}_p^{*}(\widehat{\boldsymbol{\Theta}}_{\bar{i}}, \theta) \mathbb{1}_i \right\} \tag{99}$$

by picking $\widehat{\theta}_i \notin \widehat{\boldsymbol{\Theta}}_{\bar{i}}$ because values in $\widehat{\boldsymbol{\Theta}}_{\bar{i}}$ also maximize the above cost function. It is easy to see that the first iteration of this algorithm, i.e. $i = 1$, is equivalent to maximising $f(\theta)$. However, the difference is that, the first approach involves picking q peaks from $f(\theta)$, whereas, the alternating minimisation algorithm in equation (99) picks one peak at each iteration, and therefore refining the estimates of each AoA. Moreover, this approach could also estimate the coupling parameters. This is done by using all estimated AoAs, say $\widehat{\boldsymbol{\Theta}}$ and inserting them into equation (96), namely

$$\widehat{\boldsymbol{c}} = \frac{\widehat{\boldsymbol{S}}^{-1}(\widehat{\boldsymbol{\Theta}}) \boldsymbol{A}_p^{*}(\widehat{\boldsymbol{\Theta}}) \mathbb{1}_q}{\mathbb{1}_q^{\mathrm{T}} \boldsymbol{A}_p^{\mathrm{T}}(\widehat{\boldsymbol{\Theta}}) \widehat{\boldsymbol{S}}^{-1}(\widehat{\boldsymbol{\Theta}}) \boldsymbol{A}_p^{*}(\widehat{\boldsymbol{\Theta}}) \mathbb{1}_q} \tag{100}$$

8 Simulation Results

This section provides some computer simulations to validate some MSE properties and demonstrate the potential of the proposed algorithms.

In Fig. 2, we study the behaviour of γ_k given in equation (85) by fixing $p = 3$ and increasing N, i.e. $\frac{p}{N} \to 0$. Fig. 2a plots γ_1 for one source $q = 1$, but different AoAs. The coupling parameters are set to

$$\boldsymbol{c} = \left[1;\ -0.08 + 0.5j;\ -0.14 - 0.3j\right]^{\mathrm{T}} \tag{101}$$

In addition, Fig. 2b plots γ_1 and γ_2 when $q = 2$ sources are present. The coupling parameters are set to

$$\boldsymbol{c} = \left[1;\ 0.28 + 0.41j;\ 0.18 + 0.2j\right]^{\mathrm{T}} \tag{102}$$

We observe that in both cases $\gamma_k \to 0$ as $\frac{p}{N} \to 0$. Furthermore, the rate of decay depends on the AoA, number of sources, and the coupling parameters.

In Fig. 3, we study the behaviour of γ_k by fixing N and increasing p. We have simulated two different scenarios when $q = 1$ source is present. The coupling parameters are generated by first forming a vector $\overline{\boldsymbol{c}}$, such that $\left\{\overline{c}_k = \frac{1}{k+1} e^{j2\pi\phi_k}\right\}_{k=1}^{N}$, where ϕ_k is randomly chosen. Then, in order to compute γ_k, for $p = p_0$, we choose the first p_0 elements of $\overline{\boldsymbol{c}}$ to form the vector $\boldsymbol{c} \in \mathbb{C}^{p_0 \times 1}$. In Fig. 3a and Fig. 3b, we have set $N = 10$ and $N = 50$, respectively. We also observe that γ_k is increasing as p increases for fixed N. This results in an increase of the MSE given in equation (82), when p increases due to the factor $\left(\frac{1}{1-\gamma_k}\right)$, as we shall next.

The MSE of the proposed algorithm in equation (50), namely $\mathrm{var}_f^{(p)}(\widehat{\theta}_k)$, is simulated in Fig. 4. In Fig. 4a, we set $N = 6$, $q = 1$, and $\theta_1 = 50^\circ$. The number of snapshots is $L = 10^3$. The coupling parameters are chosen from vector

$$\overline{\boldsymbol{c}} = [1;\ -0.08 + 0.5j;\ -0.14 - 0.3j;\ -0.04 + 0.04j;\ 0.03 - 0.02j]^{\mathrm{T}} \tag{103}$$

as done in the case of Fig. 3. This figure tells us that a higher MSE is obtained for increasing p. In Fig. 4b, we quantify this loss of performance. We have $q = 1$ source impinging an array of $N = 6$ at $\theta_1 = 10^\circ$. The number of snapshots is $L = 10^2$. The number of coupling parameters is $p = 3$ with $\boldsymbol{c}$ equal to that in the scenario depicted in Fig. 2a. We have plotted the experimental and theoretical MSE of MUSIC with *known coupling parameters* and the proposed algorithm in equation (50). For the experimental MSE, we have averaged over 10^3 Monte-Carlo simulations. This figure validates the gap between the MSE of MUSIC and the proposed algorithm, which is about $\frac{1}{1-\gamma_1}$, for sufficiently high SNR. The value of γ_1 could be extracted from Fig. 2a, since we have used the same coupling parameters. We could see that $\gamma_1 \simeq 0.758$ for $\theta_1 = 10^\circ$, which gives $10\log_{10}\left(\frac{1}{1-\gamma_1}\right) \simeq 6$dB. This factor is the loss of performance compared to MUSIC with *known coupling parameters*. Furthermore, we could also observe that the experimental and theoretical MSE curves are in agreement for sufficiently high SNR.

In Fig. 5, different spectra of methods that estimate AoAs in the presence of mutual coupling are depicted for a particular scenario. There are two sources $\theta_1 = 2^\circ$ and $\theta_2 = 20^\circ$ attacking a ULA composed of $N = 8$ antennas. The ULA suffers from mutual coupling with $p = \frac{N+2}{2} = 5$ coupling parameters given as

$$\boldsymbol{c} = [1;\ -0.44 + 0.23j;\ 0.33 + 0.01j;\ -0.23 - 0.1j;\ 0.1 + 0.16j]^{\mathrm{T}} \tag{104}$$

The SNR is set to 10 dB and the collected number of snapshots is $L = 500$. We observe that the methods in Figures 5a, 5b, and 5d yield fake peaks when $\theta \in \boldsymbol{\Theta}_+$ according to ***Consequence 1***. In addition, there is no peaks corresponding to the true positions. This is so because fake peaks may overlap with the true ones, when the latter are sufficiently close to the former. Furthermore, the recursive RARE depicted in Fig. 5e is initialized by RARE, and therefore selecting a false peak in the first iteration may deteriorate the performance of recursive RARE in further iterations. As we can see, recursive RARE has not successfully identified the true

positions. Moreover, the method in [61] depicted in Fig. 5c does not perform well at all. As stated earlier, this is so because the method requires that $2p+q \leq N+1$. On the other hand, the iterative method in [48] gives broad and biased peaks away from the true positions. Moreover, the proposed method in equation (50) depicted in Fig. 5g gives peaks at the true positions. The ratio between the highest true peak and the highest fake peak is about 50dB. Additionally, the ratio between the 2^{nd} highest true peak and the highest fake peak is about 25dB. Indeed, there is a great improvement between the proposed method and the previously mentioned one. Finally, the refined method discussed in Section VII could further diminish the fake peaks as we can see in Fig. 5h. In addition, the refined method also exhibits better performance in terms of bias and MSE of AoAs and coupling parameters, when compared to all these methods.

We now conduct three experiments to compare the bias and MSE of estimated parameters. In all what follows, the experiments are conducted under 500 Monte-Carlo simulations. At a given SNR, let $\widehat{\theta}_k^{(m)}$ be the estimate of θ_k at the m^{th} Monte-Carlo simulation. Similarly, let $\widehat{\boldsymbol{c}}^{(m)}$ be the estimate of $\boldsymbol{c}$ at the m^{th} Monte-Carlo simulation. Then, we define the bias of AoA parameters at a given SNR as follows:

$$\text{Bias} = \frac{1}{500q} \sum_{m=1}^{500} \sum_{k=1}^{q} \theta_k - \widehat{\theta}_k^{(m)} \tag{105}$$

Also, at a given SNR, the MSE of AoA parameters is given as follows:

$$\text{MSE}_{\text{AoA}} = \frac{1}{500q} \sum_{m=1}^{500} \sum_{k=1}^{q} \left(\theta_k - \widehat{\theta}_k^{(m)}\right)^2 \tag{106}$$

Similarly, the MSE of the coupling parameters are computed as follows

$$\text{MSE}_{\boldsymbol{c}} = \frac{1}{500} \sum_{m=1}^{500} \frac{\left\|\boldsymbol{c} - \widehat{\boldsymbol{c}}^{(m)}\right\|}{\left\|\boldsymbol{c}\right\|} \tag{107}$$

In Experiment 1, we fix the following parameters: $N = 8$, $q = 2$ i.i.d. uncorrelated Gaussian sources impinge the array at $\theta_1 = 5^\circ$ and $\theta_2 = 20^\circ$. The collected number of snapshots is $L = 10^3$, and the number of coupling parameters is $p = 3$ with

$$\boldsymbol{c} = \left[1;\ 0.2 + 0.46j;\ 0.33 + 0.04j\right]^{\text{T}} \tag{108}$$

According to Fig. 6a, all methods, except for [48], show no bias when SNR$>$ 2dB. However, it is interesting to observe that the proposed method and its refinement are the least biased. In terms of the MSE of AoA estimates, which is depicted in Fig. 6b, we also observe that the proposed method and its refinement exhibit less MSE for any SNR. Interestingly, all algorithms (except for [61] and [48]), are exposed to the same MSE, when the SNR exceeds 2dB. In Fig. 9a, we have plotted the error on estimated coupling parameters, according to equation (107), for

methods that could estimate $\boldsymbol{c}$. Indeed, the AoA estimates are directly related to coupling parameter estimates, and this is why we see that the method in [48] shows the highest MSE error. In addition, we also observe that the error on coupling parameter estimates of the Recursive RARE [60] and the refined method show similar behaviour, when SNR exceeds 2dB.

In Experiment 2 (Fig. 7a, 7b, and 9b), we fix the same parameters as in Experiment 1, except now that the 2 sources are correlated. The sources are Gaussian with covariance matrix

$$\boldsymbol{R_{ss}} = \begin{bmatrix} 1 & \rho \\ \rho^* & 1 \end{bmatrix} \tag{109}$$

where the correlation coefficient is set to $|\rho| = 0.8$. Again, the method in [48] does not perform well at all (in terms of bias and MSE). This is so because the method was based on the assumption that $\boldsymbol{R_{ss}}$ is diagonal, and therefore correlation between sources is not allowed. On the other hand, all other methods require higher SNR when sources are correlated, since they are MUSIC-based methods. For example, the proposed method in equation (50) requires an SNR of -2 dB to achieve 0 dB MSE, when the sources are un-correlated (Experiment 1). On the other hand, and in order to achieve the same MSE for correlated sources with correlation coefficient $|\rho| = 0.8$, an SNR of 13 dB is needed. This is so because the MSE of this method depends on $\boldsymbol{R_{ss}}^{-1}$ (equation (88)), and hence a higher MSE is obtained as correlation between sources increase. According to Fig. 7a and 7b, we also observe that the proposed and refined methods are the least biased and enjoy better MSE performance than other methods. As for the MSE of coupling parameters, i.e. Fig. 9b, we see that the refined method outperforms Recursive RARE by an MSE of about 6 dB, for any SNR > 7 dB.

In Experiment 3 (Fig. 8a, 8b, and 9c), we fix the same parameters as in Experiment 1, except for $p = \frac{N+2}{2} = 5$, with

$$\boldsymbol{c} = \left[1;\ 0.2 + 0.46j;\ 0.33 + 0.04j;\ 0.12 + 0.01j;\ 0.01 + 0.03\right]^{\mathrm{T}} \tag{110}$$

According to Figures 8a, 8b, and 9c, we see that all algorithms, except for [48] and the proposed ones, do not operate properly in terms of bias and MSE. This is so since p was chosen to be $\frac{N+2}{2}$. Therefore, according to ***Consequence 1***, the matrix $\boldsymbol{B}(\theta)$, and consequently $\widehat{\boldsymbol{K}}(\theta)$ admits a null-space whenever $\theta \in \boldsymbol{\Theta_+}$, and therefore the mentioned methods will always choose peaks corresponding to angles in $\theta \in \boldsymbol{\Theta_+}$. At sufficiently high SNR, we see that the MSE of the proposed algorithm and the refined method coincide (Fig. 8b). Additionally, the refined method outperforms all other algorithms in terms of bias and MSE of AoAs and coupling parameters.

9 Conclusions

There are several new results in this paper that should be highlighted.

- We have first presented and proven two theorems, namely ***Theorem 2*** and ***Theorem 3***, that allowed us to characterize the spectral behaviour of an important matrix, i.e. $\boldsymbol{B}(\theta)$, through ***Consequence 1***. This consequence explains the reason why other algorithms, revised in Section III.C, suffer from *"non-identifiability"* (i.e. when $p > \frac{N}{2}$) through that particular matrix.

- In the light of these results, we propose a new algorithm in Section IV, that does not suffer from this *"non-identifiability"* issue, namely this algorithm could estimate the Angles-of-Arrival of q sources in the presence of p mutual coupling parameters, given that $p + q \leq N$. We have also proved some properties that are related to the cost function $f(\theta)$ to give a better insight on how the proposed method operates.

- We have derived a closed-form asymptotic MSE expression of the proposed algorithm with the help of the paper in [19] and some *Perturbation Theory* tools. Moreover, we have shown that the estimates of the Angles-of-Arrival through peak finding of $f(\theta)$ are asymptotically unbiased.

- We observed the "gap" between the MSE of the proposed method and the MSE of MUSIC with *known mutual coupling parameters*. This is given by equation (84). For the k^{th} source, this "gap" is given by $\left(\frac{1}{1-\gamma_k}\right)$.

- Furthermore, the derived MSE reveals that the proposed algorithm attains the *Cramér-Rao* bound of joint mutual coupling and Angle-of-Arrival estimation when $\frac{p}{N} \to 0$ for uncorrelated signals. However, for high SNR, this is not generally the case.

- Simulation results have demonstrated the potential of the proposed method and its refined version (which is derived in Section VII) for different scenarios, as they enjoy better performance than existing methods for any p (even when $p \leq \frac{N}{2}$).

10 Appendix

10.1 Appendix A: Proof of Theorem 2

We shall prove this theorem by mathematical recurrence. Clearly, the theorem is true when $p = 1$ for any $N \geq 1$. Assume equality (36) holds true for $p - 1$. Our

task is to prove the same equality for p. Using ***Theorem 1***, we can say

$$\begin{aligned}\boldsymbol{T}(\boldsymbol{\alpha}_p)\boldsymbol{a} &= \boldsymbol{B}_p\boldsymbol{\alpha}_p \\ &= \boldsymbol{B}_{p-1}\boldsymbol{\alpha}_{p-1} + \alpha_{p-1}\boldsymbol{S}_{p-1}\boldsymbol{a} \\ &= g(z,\boldsymbol{\alpha}_{p-1})\boldsymbol{a} - \boldsymbol{M}_{p-1}\widetilde{\boldsymbol{\alpha}}_{p-1} + \alpha_{p-1}\boldsymbol{S}_{p-1}\boldsymbol{a} \\ &= g(z,\boldsymbol{\alpha}_p)\boldsymbol{a} - \alpha_{p-1}(z^{p-1} + z^{-(p-1)})\boldsymbol{a} - \boldsymbol{M}_{p-1}\tilde{\boldsymbol{\alpha}}_{p-1} \\ &+ \alpha_{p-1}\boldsymbol{S}_{p-1}\boldsymbol{a} \\ &= g(z,\boldsymbol{\alpha}_p)\boldsymbol{a} - \left[\begin{array}{c|c} \boldsymbol{M}_{p-1} & \boldsymbol{m}_{p-1} \end{array}\right]\widetilde{\boldsymbol{\alpha}}_p\end{aligned} \tag{111}$$

where $\boldsymbol{B}_p = \mathcal{G}_p(\boldsymbol{\alpha}_p)$ and $\boldsymbol{S}_{p-1}$ are given in ***Definition 2***. The vector $\boldsymbol{m}_{p-1}$ is given as

$$\boldsymbol{m}_{p-1} = \Big((z^{p-1} + z^{-(p-1)})\boldsymbol{I}_N - \boldsymbol{S}_{p-1}\Big)\boldsymbol{a} \tag{112}$$

which is easily verified to be the last column of $\boldsymbol{M}_p$.

10.2 Appendix B: Proof of Theorem 3

Using ***Theorem 1*** and ***Theorem 2***, then for any $1 \leq p \leq N$ and $\boldsymbol{\alpha} = [\alpha_0, \alpha_1 \ldots \alpha_{p-1}]^{\mathrm{T}}$ we could say

$$\boldsymbol{B}_p\boldsymbol{\alpha} = \boldsymbol{T}(\boldsymbol{\alpha})\boldsymbol{a} = g(z,\boldsymbol{\alpha})\boldsymbol{a} - \boldsymbol{M}_p\widetilde{\boldsymbol{\alpha}} \tag{113}$$

where quantities have been previously defined in their corresponding theorems.

- ***Case 1***: Here, we should prove that $\boldsymbol{B}_p\boldsymbol{\alpha} = \boldsymbol{0}$ implies $\boldsymbol{\alpha} = \boldsymbol{0}$. For $p \leq \frac{N+1}{2}$, the matrix $\boldsymbol{M}_p$ could be expressed as

$$\boldsymbol{M}_p = \left[\begin{array}{c} \boldsymbol{U}_p \\ \hline \boldsymbol{0}_{[(N-2p+2)\times(p-1)]} \\ \hline z^{N-1}\boldsymbol{J}_{p-1}\boldsymbol{U}_p^* \end{array}\right] \tag{114}$$

  Note the "zero" gap in matrix $\boldsymbol{M}_p$, which gives $g(z,\boldsymbol{\alpha}) = 0$. Due to the upper-triangular nature of $\boldsymbol{U}_p$, we get $\widetilde{\boldsymbol{\alpha}} = \boldsymbol{0}$, which, in turn, by plugging in $g(z,\boldsymbol{\alpha}) = 0$ gives $\alpha_0 = 0$.

- ***Case 2***: Fix $p = \frac{N+2}{2}$ and N is even. Then

$$\boldsymbol{M}_p = \left[\begin{array}{c} \boldsymbol{U}_p \\ \hline z^{N-1}\boldsymbol{J}_{p-1}\boldsymbol{U}_p^* \end{array}\right] \tag{115}$$

  Note that the two block matrices $\boldsymbol{U}_p$ and $\boldsymbol{J}_{p-1}\boldsymbol{U}_p^*$ do not overlap. Assume $\boldsymbol{B}_p\boldsymbol{\alpha} = \boldsymbol{0}$. The $(p-1)^{th}$ row implies

$$z^{p-2}g(z,\boldsymbol{\alpha}) = z^{-1}\alpha_{p-1} \tag{116}$$

Plugging equation (116) in the equation given by row $p-2$ gives $\alpha_{p-1}=0$. By backward substitution from rows $p-3$ till 1, we get

$$\alpha_{p-2}=\ldots=\alpha_1=0 \tag{117}$$

Now, the polynomial $g(z,\boldsymbol{\alpha})$ is given as

$$g(z,\boldsymbol{\alpha})=\alpha_0+\alpha_{p-1}(z^{p-1}+z^{-(p-1)}) \tag{118}$$

Therefore, using equation (116), row $(p-1)$ gives

$$\alpha_0=-\alpha_{p-1}z^{p-1} \tag{119}$$

Similarly, the p^{th} row and using $z^{2(p-1)}=z^N$ since $p=\frac{N+2}{2}$, we get

$$\alpha_0=-\alpha_{p-1}z^{-(p-1)} \tag{120}$$

Equations (119) and (120) together give $z^N=1$ if $\alpha_{p-1}\neq 0$. Moreover, equations (119) and (120) give us the null space of $\boldsymbol{B}_p$, namely

$$\mathcal{N}(\boldsymbol{B}_p)=\{\boldsymbol{\beta}\in\mathbb{C}^{p\times 1}, z\in\mathbb{C}^*|\boldsymbol{\beta}=[1,0\ldots 0,-z^{p-1}]^{\mathrm{T}}\} \tag{121}$$

Therefore, the rank of $\boldsymbol{B}_p$ is $p-1$.

- ***Case 3(a)***: Here, N is even and $p>\frac{N+2}{2}$. Fix $k=p-\frac{N+2}{2}$. In this case, the two block matrices $\boldsymbol{U}_p$ and $\boldsymbol{J}_p\boldsymbol{U}_p^*$ overlap. Furthermore, the structure of $\boldsymbol{M}_p$ is given as follows

$$\boldsymbol{M}_p=\left[\begin{array}{c|c} \boldsymbol{U}_{p-k} & \\ \hline z^{N-1}\boldsymbol{J}_{p-k-1}\boldsymbol{U}_{p-k}^* & \boldsymbol{V}\end{array}\right] \tag{122}$$

The i^{th} column of $\boldsymbol{V}\in\mathbb{C}^{N\times k}$ is

$$\boldsymbol{v}_i=\left[\begin{array}{c|c|c}\boldsymbol{u}_i^{\mathrm{T}} & \boldsymbol{m}_i^{\mathrm{T}} & \boldsymbol{b}_i^{\mathrm{T}}\end{array}\right]^{\mathrm{T}} \tag{123}$$

where

$$\boldsymbol{u}_i=[z^{-(\frac{N}{2}+i)},z^{-(\frac{N}{2}+i)+1}\ldots z^{-(2i+1)}]^{\mathrm{T}} \tag{124}$$

$$\boldsymbol{m}_i=[z^{-2i}+z^N,z(z^{-2i}+z^N)\ldots z^{2i-1}(z^{-2i}+z^N)]^{\mathrm{T}} \tag{125}$$

$$\boldsymbol{b}_i=[z^{N+2i},z^{N+2i+1}\ldots z^{\frac{3N}{2}-1+i}]^{\mathrm{T}} \tag{126}$$

Realising the above equations, the system of equations $\boldsymbol{B}_p\boldsymbol{\alpha}=\boldsymbol{0}$ could be partitioned into 4 subsystems of equations given as follows:

Rows $1\ldots\frac{N}{2}-k$ of $\boldsymbol{B}_p\boldsymbol{\alpha}=\boldsymbol{0}$ are given by system $\mathcal{S}_1$

$$\mathcal{S}_1:\quad \boldsymbol{g}(z,\boldsymbol{\alpha})=\sum_{i=l}^{\frac{N}{2}+k}\alpha_i z^{-i},\quad l=1\ldots\frac{N}{2}-k \tag{127}$$

Rows $\frac{N}{2}-k+1\ldots\frac{N}{2}$ of $\boldsymbol{B}_p\boldsymbol{\alpha}=\mathbf{0}$ are given by system $\mathcal{S}_2$

$$\mathcal{S}_2: \quad \boldsymbol{g}(z,\boldsymbol{\alpha})=\sum_{i=\frac{N}{2}-k+l}^{\frac{N}{2}+k-l}\alpha_i z^{-i}+\sum_{i=\frac{N}{2}-k-l+1}^{\frac{N}{2}+k}\alpha_i(z^i+z^{-i}), \quad l=1\ldots k \tag{128}$$

Rows $\frac{N}{2}+1\ldots\frac{N}{2}+k$ of $\boldsymbol{B}_p\boldsymbol{\alpha}=\mathbf{0}$ are given by system $\mathcal{S}_3$

$$\mathcal{S}_3: \quad \boldsymbol{g}(z,\boldsymbol{\alpha})=\sum_{i=\frac{N}{2}-k+l}^{\frac{N}{2}+k-l}\alpha_i z^{i}+\sum_{i=\frac{N}{2}-k-l+1}^{\frac{N}{2}+k}\alpha_i(z^i+z^{-i}), \quad l=1\ldots k \tag{129}$$

Rows $\frac{N}{2}+k+1\ldots N$ of $\boldsymbol{B}_p\boldsymbol{\alpha}=\mathbf{0}$ are given by system $\mathcal{S}_4$

$$\mathcal{S}_4: \quad \boldsymbol{g}(z,\boldsymbol{\alpha})=\sum_{i=l}^{\frac{N}{2}+k}\alpha_i z^{i}, \quad l=1\ldots\frac{N}{2}-k \tag{130}$$

Now, system $\mathcal{S}_1$ (or equivalently $\mathcal{S}_4$) imply the following

$$\alpha_1=\ldots=\alpha_{\frac{N}{2}-k-1}=0 \tag{131}$$

which is carried on by backward substitution. Therefore, systems $\mathcal{S}_1$ and $\mathcal{S}_4$ each break down to one and only one equation (for $l=\frac{N}{2}-k$). Furthermore, for $l=k$, system $\mathcal{S}_2$ gives

$$\alpha_{\frac{N}{2}}z^{-\frac{N}{2}}(z^N-1)=0 \tag{132}$$

According to equation (132), two cases arise:

$$\boxed{\alpha_{\frac{N}{2}}=0 \text{ and } z^N\neq 1}$$

Using systems $\mathcal{S}_1$ and $\mathcal{S}_2$, we get

$$\sum_{i=\frac{N}{2}-k}^{\frac{N}{2}-k+l-1}\alpha_i z^{-i}=\sum_{i=\frac{N}{2}+k-l+1}^{\frac{N}{2}+k}\alpha_i z^{i}, \quad l=1\ldots k \tag{133}$$

Similarly, systems $\mathcal{S}_3$ and $\mathcal{S}_4$ give

$$\sum_{i=\frac{N}{2}-k}^{\frac{N}{2}-k+l-1}\alpha_i z^{i}=\sum_{i=\frac{N}{2}+k-l+1}^{\frac{N}{2}+k}\alpha_i z^{-i}, \quad l=1\ldots k \tag{134}$$

Equation (133) reads

$$\alpha_{\frac{N}{2}-k+l-1}=\alpha_{\frac{N}{2}+k-l+1}z^N, \qquad l=1\ldots k \tag{135}$$

Equation (134) gives

$$\alpha_{\frac{N}{2}-k+l-1} = \alpha_{\frac{N}{2}+k-l+1} z^{-N}, \qquad l = 1 \dots k \tag{136}$$

Equations (135) and (136) together give

$$\alpha_{\frac{N}{2}+k-l+1}(1 - z^N)(1 + z^{-N}) = 0 \tag{137}$$

Based on assumption $z^N \neq 1$, equation (137) gives two subcases:

- ***Case 3(a.i)***: $\{\alpha_{\frac{N}{2}+k-l+1} = 0\}_{l=1}^{k}$ and $z^N \neq -1$. In this case, one could easily verify that $\alpha_0 = \ldots = \alpha_{p-1} = 0$. Hence, iff $z^N \neq \pm 1$, the matrix $\boldsymbol{B}_p$ is full rank.

- ***Case 3(a.ii)***: $\{\alpha_{\frac{N}{2}+k-l+1} \neq 0\}_{l=1}^{k}$ and $z^N = -1$. In this case, equation (136) gives

$$\alpha_{\frac{N}{2}-k+l-1} = -\alpha_{\frac{N}{2}+k-l+1}, \qquad l = 1 \dots k \tag{138}$$

Systems $\mathcal{S}_1$ and $\mathcal{S}_4$ now give

$$\alpha_0 = -z^{\frac{N}{2}} \sum_{i=1}^{k} \alpha_{\frac{N}{2}+i}(z^i - z^{-i}) \tag{139}$$

The dimension of the corresponding null space is $(2k+1) - (k+1) = k$. Note that the quantity $(2k+1)$ is the number of non-zero variables and $(k+1)$ is the number of linearly independent equations. The null space of $\boldsymbol{B}_p$ when $z^N = -1$ is given by

$$\mathcal{N}(\boldsymbol{B}_p) = \left\{ \boldsymbol{b} \in \mathbb{C}^{p\times 1}, \boldsymbol{\beta} \neq \boldsymbol{0} \middle| \boldsymbol{b} = \begin{bmatrix} h_-(\boldsymbol{\beta}) \\ \boldsymbol{0} \\ -\boldsymbol{J}_k\boldsymbol{\beta} \\ 0 \\ \boldsymbol{\beta} \end{bmatrix} \right\} \tag{140a}$$

where $\boldsymbol{\beta} = [\beta_1 \dots \beta_k]^{\mathrm{T}}$ and

$$h_-(\boldsymbol{\beta}) = -z^{\frac{N}{2}} \sum_{i=1}^{k} \beta_i (z^i - z^{-i}) \tag{140b}$$

Therefore,

$$\text{rank}(\boldsymbol{B}_p) = p - k = \frac{N}{2} + 1 \tag{141}$$

$$\boxed{\alpha_{\frac{N}{2}} \neq 0 \text{ and } z^N = 1}$$

Following the same steps as Case 3(a.1), one reaches equation (135) and concludes

$$\alpha_{\frac{N}{2}-k+l-1} = \alpha_{\frac{N}{2}+k-l+1}, \qquad l = 1 \dots k \tag{142}$$

Also, as previously done, systems $\mathcal{S}_1$ and $\mathcal{S}_4$ imply

$$\alpha_0 = -z^{\frac{N}{2}} \Big(\alpha_{\frac{N}{2}} + \sum_{i=1}^{k} \alpha_{\frac{N}{2}+i} (z^i + z^{-i}) \Big) \tag{143}$$

The null space therefore spans $k+1$ dimensions, namely

$$\mathcal{N}(\boldsymbol{B}_p) = \Big\{ \boldsymbol{b} \in \mathbb{C}^{p\times 1}, [\boldsymbol{\beta}, \gamma]^{\mathrm{T}} \neq \mathbf{0} \Big| \boldsymbol{b} = \begin{bmatrix} h_+(\boldsymbol{\beta}, \gamma) \\ \mathbf{0} \\ \boldsymbol{J}_k \boldsymbol{\beta} \\ \gamma \\ \boldsymbol{\beta} \end{bmatrix} \Big\} \tag{144a}$$

$$h_+(\boldsymbol{\beta}, \gamma) = -z^{\frac{N}{2}} \Big(\gamma + \sum_{i=1}^{k} \beta_i (z^i + z^{-i}) \Big) \tag{144b}$$

Hence, we conclude that the rank of $\boldsymbol{B}_p$ is $\frac{N}{2}$. This completes the proof of ***Case 3(a.iii)***.

- ***Case 3(b)***: Here, N is odd and $p > \frac{N+2}{2}$. Fix $k = p - \frac{N+1}{2}$. The proof follows similar steps as ***Case 3(a)***. The null space is given as follows

$$\mathcal{N}(\boldsymbol{B}_p) = \Big\{ \boldsymbol{b} \in \mathbb{C}^{p\times 1}, \boldsymbol{\beta} \neq \mathbf{0} \Big| \boldsymbol{b} = \begin{bmatrix} y(\boldsymbol{\beta}) \\ \mathbf{0} \\ -z^N \boldsymbol{J}_k \boldsymbol{\beta} \\ \boldsymbol{\beta} \end{bmatrix} \Big\} \tag{145a}$$

$$y(\boldsymbol{\beta}) = -z^{\frac{N}{2}} \sum_{i=1}^{k} \beta_i \Big(z^{(i+\frac{1}{2})} - z^N z^{-(i+\frac{1}{2})} \Big) \tag{145b}$$

10.3 Appendix C: Proof of Consequence 1

Using the results of ***Theorem 3*** and restricting ourselves with $z = z_\theta = e^{-j2\pi \frac{d}{\lambda} \sin(\theta)}$, it suffices to derive the two sets, $\boldsymbol{\Theta_+}$ and $\boldsymbol{\Theta_-}$. The equation $z_\theta^N = 1$ reads the following

$$e^{-j2\pi \frac{d}{\lambda} N \sin(\theta)} = e^{j2k\pi}, \quad k = -\frac{N}{2} \dots \frac{N}{2} \tag{146}$$

With some straightforward algebra, equation (146) implies that $\theta \in \boldsymbol{\Theta_+}$. In a similar manner, $z_\theta^N = -1$ implies $\theta \in \boldsymbol{\Theta_-}$. Combining ***Theorem 3*** with the above completes the proof.

10.4 Appendix D: 1^{st} and 2^{nd} order derivatives of $f(\theta)$

The 1^{st} order derivative is computed as

$$f'(\theta) = 2\text{Re}\{\boldsymbol{a}_p^{\text{T}}(\theta)\widehat{\boldsymbol{K}}^{-1}(\theta)\frac{\partial \boldsymbol{a}_p^*(\theta)}{\partial \theta}\} + \boldsymbol{a}_p^{\text{T}}(\theta)\frac{\partial \widehat{\boldsymbol{K}}^{-1}(\theta)}{\partial \theta}\boldsymbol{a}_p^*(\theta) \tag{147}$$

Denoting

$$\boldsymbol{d}_p(\theta) = \frac{\partial \boldsymbol{a}_p(\theta)}{\partial \theta} \tag{148}$$

$$\boldsymbol{D}(\theta) = \frac{\partial \boldsymbol{B}(\theta)}{\partial \theta} = \mathcal{G}_p(\frac{\partial \boldsymbol{a}(\theta)}{\partial \theta}) \tag{149}$$

and using the following identity [73]

$$\frac{\partial \widehat{\boldsymbol{K}}^{-1}(\theta)}{\partial \theta} = -\widehat{\boldsymbol{K}}^{-1}(\theta)\frac{\partial \widehat{\boldsymbol{K}}(\theta)}{\partial \theta}\widehat{\boldsymbol{K}}^{-1}(\theta) \tag{150}$$

then $f'(\theta) = g_1(\theta) + g_2(\theta)$, where

$$g_1(\theta) = 2\text{Re}\{\boldsymbol{a}_p^{\text{T}}(\theta)\widehat{\boldsymbol{K}}^{-1}(\theta)\boldsymbol{d}_p^*(\theta)\} \tag{151a}$$

$$g_2(\theta) = -2\text{Re}\{\boldsymbol{a}_p^{\text{T}}(\theta)\widehat{\boldsymbol{K}}^{-1}(\theta)\widehat{\boldsymbol{G}}(\theta)\widehat{\boldsymbol{K}}^{-1}(\theta)\boldsymbol{a}_p^*(\theta)\} \tag{151b}$$

and

$$\widehat{\boldsymbol{G}}(\theta) = \boldsymbol{B}^{\text{H}}(\theta)\widehat{\boldsymbol{U}}_{\boldsymbol{n}}\widehat{\boldsymbol{U}}_{\boldsymbol{n}}^{\text{H}}\boldsymbol{D}(\theta) \tag{151c}$$

In a similar manner, after some straightforward, but lengthy, calculations, one could verify that $f''(\theta) = h_1(\theta) + h_2(\theta) + h_3(\theta)$, where $h_k(\theta)$ are given as

$$h_1(\theta) = 2\text{Re}\Big\{\boldsymbol{d}_p^{\text{T}}(\theta)\widehat{\boldsymbol{K}}^{-1}(\theta)\boldsymbol{d}_p^*(\theta) + \boldsymbol{a}_p^{\text{T}}(\theta)\widehat{\boldsymbol{K}}^{-1}(\theta)\frac{\partial \boldsymbol{d}_p^*(\theta)}{\partial \theta}\Big\} \tag{152a}$$

$$\begin{aligned} h_2(\theta) = &-4\text{Re}\Big\{\boldsymbol{a}_p^{\text{T}}(\theta)\widehat{\boldsymbol{K}}^{-1}(\theta)\Big(\widehat{\boldsymbol{G}}(\theta) + \widehat{\boldsymbol{G}}^{\text{H}}(\theta)\Big)\widehat{\boldsymbol{K}}^{-1}(\theta)\boldsymbol{d}_p^*(\theta)\Big\} \\ &-2\text{Re}\Big\{\boldsymbol{a}_p^{\text{T}}(\theta)\widehat{\boldsymbol{K}}^{-1}(\theta)\boldsymbol{B}^{\text{H}}(\theta)\widehat{\boldsymbol{U}}_{\boldsymbol{n}}\widehat{\boldsymbol{U}}_{\boldsymbol{n}}^{\text{H}}\frac{\partial \boldsymbol{D}(\theta)}{\partial \theta}\widehat{\boldsymbol{K}}^{-1}(\theta)\boldsymbol{a}_p^*(\theta)\Big\} \\ &-2\boldsymbol{a}_p^{\text{T}}(\theta)\widehat{\boldsymbol{K}}^{-1}(\theta)\boldsymbol{D}^{\text{H}}(\theta)\widehat{\boldsymbol{U}}_{\boldsymbol{n}}\widehat{\boldsymbol{U}}_{\boldsymbol{n}}^{\text{H}}\boldsymbol{D}(\theta)\widehat{\boldsymbol{K}}^{-1}(\theta)\boldsymbol{a}_p^*(\theta) \end{aligned} \tag{152b}$$

$$h_3(\theta) = 4\text{Re}\Big\{\boldsymbol{a}_p^{\text{T}}(\theta)\widehat{\boldsymbol{K}}^{-1}(\theta)\Big(\widehat{\boldsymbol{G}}(\theta)\widehat{\boldsymbol{K}}^{-1}(\theta)\widehat{\boldsymbol{G}}(\theta)+\widehat{\boldsymbol{G}}^{\text{H}}(\theta)\widehat{\boldsymbol{K}}^{-1}(\theta)\widehat{\boldsymbol{G}}(\theta)\Big)\widehat{\boldsymbol{K}}^{-1}(\theta)\boldsymbol{a}_p^*(\theta)\Big\} \tag{152c}$$

10.5 Appendix E: Perturbation of $\widehat{\lambda}_1$ and $\widehat{\boldsymbol{v}}_1$

$\boldsymbol{K}(\theta_k)$ could, also, be decomposed as follows

$$\widehat{\boldsymbol{K}}(\theta_k) = \boldsymbol{K}(\theta_k) + \widetilde{\boldsymbol{K}}(\theta_k) \tag{153}$$

where $\boldsymbol{K}(\theta_k) = \boldsymbol{B}^{\mathrm{H}}(\theta_k)\boldsymbol{U_n}\boldsymbol{U_n}^{\mathrm{H}}\boldsymbol{B}(\theta_k)$ and

$$\widetilde{\boldsymbol{K}}(\theta_k) = 2\mathrm{Re}\big\{\boldsymbol{B}^{\mathrm{H}}(\theta_k)\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}\boldsymbol{U_n}^{\mathrm{H}}\boldsymbol{B}(\theta_k)\big\} + \boldsymbol{B}^{\mathrm{H}}(\theta_k)\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}^{\mathrm{H}}\boldsymbol{B}(\theta_k) \tag{154}$$

Using well-known results in *Perturbation Theory* [69, 70], we seek to use the following methodology: Given two Hermitian positive semi-definite matrices $\boldsymbol{K}(\theta_k)$ and $\widetilde{\boldsymbol{K}}(\theta_k)$, where the latter perturbs the former, each $\widehat{\lambda}_j$ and $\widehat{\boldsymbol{v}}_j$ could be approximated by a linear combination as follows:

$$\widehat{\lambda}_j = \lambda_j + \boldsymbol{v}_j^{\mathrm{H}}\widetilde{\boldsymbol{K}}(\theta_k)\boldsymbol{v}_j + \sum_{i\neq j}\frac{|\boldsymbol{v}_i^{\mathrm{H}}\widetilde{\boldsymbol{K}}(\theta_k)\boldsymbol{v}_j|^2}{\lambda_j - \lambda_i} + \mathcal{O}(\|\widetilde{\boldsymbol{K}}\|^3) \tag{155}$$

and

$$\widehat{\boldsymbol{v}}_j = \boldsymbol{v}_j + \sum_{i\neq j}\frac{\boldsymbol{v}_i^{\mathrm{H}}\widetilde{\boldsymbol{K}}(\theta_k)\boldsymbol{v}_j}{\lambda_j - \lambda_i}\boldsymbol{v}_i + \mathcal{O}(\widetilde{\boldsymbol{K}}^2) \tag{156}$$

This approximation is valid if the the eigenvalue λ_j is non-degenerate. In our case, $\lambda_1 = 0$ is non-degenerate as long as $p \leq \frac{N+2}{2}$ or $\{p > \frac{N+2}{2}$ *and* $\theta_k \notin \boldsymbol{\Theta_{\pm}}\}$ according to ***Consequence 1***. In that case, applying equation (155) to $\widehat{\lambda}_1$ and denoting $\boldsymbol{B} = \boldsymbol{B}(\theta_k)$ and $\boldsymbol{K} = \boldsymbol{K}(\theta_k)$ for short, we get

$$\begin{aligned}
\widehat{\lambda}_1 &= \frac{1}{\|\boldsymbol{c}\|^2}\Big(\boldsymbol{c}^{\mathrm{H}}\widetilde{\boldsymbol{K}}\boldsymbol{c} - \sum_{i=2}^{p}\frac{|\boldsymbol{v}_i^{\mathrm{H}}\widetilde{\boldsymbol{K}}\boldsymbol{c}|^2}{\lambda_i}\Big) + \mathcal{O}(\|\widetilde{\boldsymbol{K}}\|^3) \\
&= \frac{1}{\|\boldsymbol{c}\|^2}\Big(\boldsymbol{c}^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}}\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{c} - \sum_{i=2}^{p}\frac{|\boldsymbol{v}_i^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{U_n}\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{c}|^2}{\lambda_i}\Big) + \dots \\
&= \frac{1}{\|\boldsymbol{c}\|^2}\boldsymbol{c}^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}}\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}\Big(\boldsymbol{I} - \boldsymbol{U_n}^{\mathrm{H}}\boldsymbol{B}\underbrace{(\sum_{i=2}^{p}\frac{\boldsymbol{v}_i\boldsymbol{v}_i^{\mathrm{H}}}{\lambda_i})}_{\boldsymbol{K}^{+}}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{U_n}\Big)\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{c} \\
&= \frac{1}{\|\boldsymbol{c}\|^2}\boldsymbol{c}^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}}\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}\Big(\boldsymbol{I} - \boldsymbol{P}_k\Big)\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{c} + \mathcal{O}(\|\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}\|^3) \\
&= \frac{1}{\|\boldsymbol{c}\|^2}\boldsymbol{c}^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}}\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}\boldsymbol{P}_k^{\perp}\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{c} + \mathcal{O}(\|\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}\|^3)
\end{aligned} \tag{157}$$

In a similar manner, using equation (156), $\widehat{\boldsymbol{v}}_{\boldsymbol{1}}$ could be written as

$$\begin{aligned}
\widehat{\boldsymbol{v}}_1 &= \frac{1}{\|\boldsymbol{c}\|}\Big(\boldsymbol{c} - \sum_{i=2}^{p}\frac{\boldsymbol{v}_i^{\mathrm{H}}\widetilde{\boldsymbol{K}}\boldsymbol{c}}{\lambda_i}\boldsymbol{v}_i\Big) + \mathcal{O}(\widetilde{\boldsymbol{K}}^2) \\
&= \frac{1}{\|\boldsymbol{c}\|}\Big(\boldsymbol{c} - \sum_{i=2}^{p}\frac{\boldsymbol{v}_i^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{U_n}\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{c}}{\lambda_i}\boldsymbol{v}_i\Big) + \mathcal{O}(\widetilde{\boldsymbol{U}}_{\boldsymbol{n}}^2)
\end{aligned} \tag{158}$$

As for the degenerate case, i.e. $\lambda_1 = \dots = \lambda_{\Delta+1} = 0$, which occurs when $p > \frac{N+2}{2}$ and $\theta_k \in \boldsymbol{\Theta_{\pm}}$; we follow similar steps and use the approximations given in [71].

10.6 Appendix F: MSE Expression

Let's call

$$\widetilde{\boldsymbol{\omega}}_k \triangleq \widetilde{\boldsymbol{U}}_{\boldsymbol{n}}^{\mathrm{H}} \boldsymbol{B}(\theta_k)\boldsymbol{c} = \widetilde{\boldsymbol{U}}_{\boldsymbol{n}}^{\mathrm{H}} \boldsymbol{U}_{\boldsymbol{s}} \boldsymbol{U}_{\boldsymbol{s}}^{\mathrm{H}} \boldsymbol{B}(\theta_k)\boldsymbol{c} \tag{159}$$

where the second equality is due to the fact that $\boldsymbol{B}(\theta_k)\boldsymbol{c} = \bar{\boldsymbol{a}}(\theta_k) = \boldsymbol{T}(\boldsymbol{c})\boldsymbol{a}(\theta_k)$. Using ***Lemma 1***, it is easy to see that $\widetilde{\boldsymbol{\omega}}_k$ is Gaussian distributed with zero mean and covariance matrix

$$\begin{aligned}
\Big(\mathbb{E}\{\widetilde{\boldsymbol{\omega}}_k \widetilde{\boldsymbol{\omega}}_k^{\mathrm{H}}\}\Big)_{i,j} &= \mathbb{E}\Big\{ \big(\boldsymbol{U}_{\boldsymbol{s}} \boldsymbol{U}_{\boldsymbol{s}}^{\mathrm{H}} \widetilde{\boldsymbol{n}}_i\big)^{\mathrm{H}} \bar{\boldsymbol{a}}(\theta_k) \bar{\boldsymbol{a}}^{\mathrm{H}}(\theta_k) \big(\boldsymbol{U}_{\boldsymbol{s}} \boldsymbol{U}_{\boldsymbol{s}}^{\mathrm{H}} \widetilde{\boldsymbol{n}}_j\big) \Big\} \\
&= \bar{\boldsymbol{a}}^{\mathrm{H}}(\theta_k) \mathbb{E}\Big\{ \big(\boldsymbol{U}_{\boldsymbol{s}} \boldsymbol{U}_{\boldsymbol{s}}^{\mathrm{H}} \widetilde{\boldsymbol{n}}_j\big) \big(\boldsymbol{U}_{\boldsymbol{s}} \boldsymbol{U}_{\boldsymbol{s}}^{\mathrm{H}} \widetilde{\boldsymbol{n}}_i\big)^{\mathrm{H}} \Big\} \bar{\boldsymbol{a}}(\theta_k) \\
&= \underbrace{\frac{\sigma^2}{L} \bar{\boldsymbol{a}}^{\mathrm{H}}(\theta_k) \boldsymbol{U} \bar{\boldsymbol{a}}(\theta_k)}_{\widetilde{\sigma}_k^2} \delta_{i,j}
\end{aligned} \tag{160}$$

where the last equality is a result of equation (79). Therefore, $\widetilde{\boldsymbol{\omega}}_k \sim \mathcal{CN}(\boldsymbol{0}, \widetilde{\sigma}_k^2 \boldsymbol{I})$. Similarly, $\mathbb{E}\{\widetilde{\boldsymbol{\omega}}_k \widetilde{\boldsymbol{\omega}}_k^{\mathrm{T}}\} = \boldsymbol{0}$. Using the moments of $\widetilde{\boldsymbol{\omega}}_k$, we have

$$\begin{aligned}
\mathbb{E}\big\{(\widetilde{\theta}_k)\big\} &= \mathbb{E}\big\{\frac{\mathrm{Re}\{\widetilde{\rho}_k\}}{\upsilon_k}\big\} \\
&= \frac{1}{2\upsilon_k}\Big(\bar{\boldsymbol{d}}^{\mathrm{H}}(\theta_k) \boldsymbol{U}_{\boldsymbol{n}} \boldsymbol{P}_k^{\perp} \mathbb{E}\{\widetilde{\boldsymbol{\omega}}_k\} + \mathbb{E}\{\widetilde{\boldsymbol{\omega}}_k^{\mathrm{H}}\} \boldsymbol{P}_k^{\perp} \boldsymbol{U}_{\boldsymbol{n}}^{\mathrm{H}} \bar{\boldsymbol{d}}(\theta_k)\Big) \\
&= 0
\end{aligned} \tag{161}$$

$$\begin{aligned}
\mathbb{E}\big\{(\widetilde{\theta}_k)^2\big\} &= \mathbb{E}\big\{\frac{\big(\mathrm{Re}\{\widetilde{\rho}_k\}\big)^2}{\upsilon_k^2}\big\} \\
&= \frac{1}{2\upsilon_k^2}\mathbb{E}\big\{|\widetilde{\rho}_k|^2\big\} \\
&= \frac{1}{2\upsilon_k^2} \bar{\boldsymbol{d}}^{\mathrm{H}}(\theta_k) \boldsymbol{U}_{\boldsymbol{n}} \boldsymbol{P}_k^{\perp} \mathbb{E}\{\widetilde{\boldsymbol{\omega}}_k \widetilde{\boldsymbol{\omega}}_k^{\mathrm{H}}\} \boldsymbol{P}_k^{\perp} \boldsymbol{U}_{\boldsymbol{n}}^{\mathrm{H}} \bar{\boldsymbol{d}}(\theta_k) \\
&= \frac{\widetilde{\sigma}_k^2}{2\upsilon_k^2} \bar{\boldsymbol{d}}^{\mathrm{H}}(\theta_k) \boldsymbol{U}_{\boldsymbol{n}} \boldsymbol{P}_k^{\perp} \boldsymbol{U}_{\boldsymbol{n}}^{\mathrm{H}} \bar{\boldsymbol{d}}(\theta_k) = \frac{\widetilde{\sigma}_k^2}{2\upsilon_k} \\
&= \frac{\sigma^2}{2L} \frac{\bar{\boldsymbol{a}}^{\mathrm{H}}(\theta_k) \boldsymbol{U} \bar{\boldsymbol{a}}(\theta_k)}{\upsilon_k}
\end{aligned} \tag{162}$$

where we have used

$$\big(\mathrm{Re}\{z\}\big)^2 = \frac{1}{2}\big(|z|^2 + \mathrm{Re}\{z^2\}\big) \tag{163}$$

and $\mathbb{E}\big\{\mathrm{Re}\{\widetilde{\rho}_k^2\}\big\} = 0$ since $\mathbb{E}\{\widetilde{\boldsymbol{\omega}}_k \widetilde{\boldsymbol{\omega}}_k^{\mathrm{T}}\} = \boldsymbol{0}$.

10.7 Appendix G: Proof of Theorem 6

The terms $\boldsymbol{B}^{\mathrm{H}}(\theta_k)\boldsymbol{B}(\theta_l)$, $\boldsymbol{B}^{\mathrm{H}}(\theta_k)\boldsymbol{D}(\theta_l)$, and $\boldsymbol{D}^{\mathrm{H}}(\theta_k)\boldsymbol{D}(\theta_l)$ appear in $\mathrm{var}_f^{(p)}(\widehat{\theta}_k)$ and $\mathrm{var}_{\mathrm{CRB}}(\widehat{\theta}_k)$. We first compute the limits of these three expressions as $\frac{p}{N} \to 0$.

With some straightforward calculations, one could verify the following equality

$$\Big(\boldsymbol{B}^{\mathrm{H}}(\theta_k)\boldsymbol{B}(\theta_l)\Big)_{m,n} = \begin{cases} b_{k,l}(m,n), & \text{if } m \geq n \\ b^*_{l,k}(n,m), & \text{else} \end{cases} \tag{164}$$

where

$$\begin{aligned} & b_{k,l}(m,n) \\ & = \Big(z_{\theta_k}^{-(m-1)} z_{\theta_l}^{(n-1)} + (1-\delta_{1,m}\delta_{1,n}) z_{\theta_l}^{(m-n)}\Big)\Big(\sum_{i=0}^{N-m} [z^*_{\theta_k} z_{\theta_l}]^i\Big) \\ & + (1-\delta_{1,n})\Big(z_{\theta_k}^{-(m+n-2)} + z_{\theta_l}^{(m+n-2)}\Big)\Big(\sum_{i=0}^{N-m-n+1} [z^*_{\theta_k} z_{\theta_l}]^i\Big) \end{aligned} \tag{165}$$

Using the following identity

$$\frac{1}{m^{k+1}} \sum_{t=1}^{m} t^k e^{jt(w_1-w_2)} \xrightarrow[m\to\infty]{} \frac{1}{k+1}\delta_{w_1,w_2} \tag{166}$$

and keeping in mind that p is fixed, we could complete the summation terms appearing in equation (165) by a "finite" amount of terms of order p so that the limits of the sum span all integers $i = 0 \dots N$, and therefore we have

$$\frac{\boldsymbol{B}^{\mathrm{H}}(\theta_k)\boldsymbol{B}(\theta_l)}{N} \xrightarrow[\frac{p}{N}\to 0]{} \boldsymbol{h}_k \boldsymbol{h}_k^{\mathrm{H}} \delta_{k,l} \tag{167a}$$

which is a rank-one contribution, and $\boldsymbol{h}_k$ is given in equation (92). In a very similar manner, we can prove

$$\frac{\boldsymbol{B}^{\mathrm{H}}(\theta_k)\boldsymbol{D}(\theta_l)}{N^2} \xrightarrow[\frac{p}{N}\to 0]{} \frac{j}{2} \boldsymbol{h}_k \boldsymbol{h}_k^{\mathrm{H}} \delta_{k,l} \tag{167b}$$

and

$$\frac{\boldsymbol{D}^{\mathrm{H}}(\theta_k)\boldsymbol{D}(\theta_l)}{N^3} \xrightarrow[\frac{p}{N}\to 0]{} \frac{1}{3} \boldsymbol{h}_k \boldsymbol{h}_k^{\mathrm{H}} \delta_{k,l} \tag{167c}$$

With those limits in hand, we could verify the following

$$\frac{1}{N^3} \bar{\boldsymbol{D}}^{\mathrm{H}} \boldsymbol{P}^{\perp}_{\bar{\boldsymbol{A}}} \bar{\boldsymbol{D}} \xrightarrow[\frac{p}{N}\to 0]{} \frac{1}{12} \begin{bmatrix} |\boldsymbol{h}_1^{\mathrm{H}}\boldsymbol{c}|^2 & 0 & \cdots & 0 \\ 0 & |\boldsymbol{h}_2^{\mathrm{H}}\boldsymbol{c}|^2 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & |\boldsymbol{h}_q^{\mathrm{H}}\boldsymbol{c}|^2 \end{bmatrix} \tag{168}$$

Note that, when $p = 1$, the above diagonal matrix is the identity matrix, which coincides with the result in [19]. Plugging the limit of equation (168) in the CRB

expression given in equation (86), we get equation (89). To verify equation (90), we expand the denominator of equation (88) as follow

$$\begin{aligned}&\bar{\boldsymbol{d}}^{\mathrm{H}}(\theta_k)\boldsymbol{U_n}\boldsymbol{P}_k^{\perp}\boldsymbol{U_n}^{\mathrm{H}}\bar{\boldsymbol{d}}(\theta_k)\\&=\boldsymbol{c}^{\mathrm{H}}\Big[\boldsymbol{D}^{\mathrm{H}}(\theta_k)\boldsymbol{U_n}\boldsymbol{U_n}^{\mathrm{H}}\boldsymbol{D}(\theta_k)-\big(\boldsymbol{D}^{\mathrm{H}}(\theta_k)\boldsymbol{U_n}\boldsymbol{U_n}^{\mathrm{H}}\boldsymbol{B}(\theta_k)\big)\\&\big(\boldsymbol{B}^{\mathrm{H}}(\theta_k)\boldsymbol{U_n}\boldsymbol{U_n}^{\mathrm{H}}\boldsymbol{B}(\theta_k)\big)^{+}\big(\boldsymbol{B}^{\mathrm{H}}(\theta_k)\boldsymbol{U_n}\boldsymbol{U_n}^{\mathrm{H}}\boldsymbol{D}(\theta_k)\big)\Big]\boldsymbol{c}\end{aligned}\tag{169}$$

By using the limits computed in (167), we have

$$\frac{\boldsymbol{B}^{\mathrm{H}}(\theta_k)\boldsymbol{U_n}\boldsymbol{U_n}^{\mathrm{H}}\boldsymbol{B}(\theta_k)}{N}\xrightarrow[\frac{p}{N}\to 0]{}\boldsymbol{0}\tag{170a}$$

$$\frac{\boldsymbol{B}^{\mathrm{H}}(\theta_k)\boldsymbol{U_n}\boldsymbol{U_n}^{\mathrm{H}}\boldsymbol{D}(\theta_k)}{N^2}\xrightarrow[\frac{p}{N}\to 0]{}\boldsymbol{0}\tag{170b}$$

$$\frac{\boldsymbol{D}^{\mathrm{H}}(\theta_k)\boldsymbol{U_n}\boldsymbol{U_n}^{\mathrm{H}}\boldsymbol{D}(\theta_k)}{N^3}\xrightarrow[\frac{p}{N}\to 0]{}\frac{1}{12}\boldsymbol{h}_k\boldsymbol{h}_k^{\mathrm{H}}\tag{170c}$$

and therefore

$$\frac{1}{N^3}\bar{\boldsymbol{d}}^{\mathrm{H}}(\theta_k)\boldsymbol{U_n}\boldsymbol{P}_k^{\perp}\boldsymbol{U_n}^{\mathrm{H}}\bar{\boldsymbol{d}}(\theta_k)\xrightarrow[\frac{p}{N}\to 0]{}\frac{1}{12}|\boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{c}|^2\tag{171}$$

Equations (170) directly imply that $\gamma_k \xrightarrow[\frac{p}{N}\to 0]{} 0$. Now, using equation (167a), we can verify that the second term in the numerator of equation (88) goes to zero, viz.

$$\big(\boldsymbol{R_{ss}}^{-1}(\bar{\boldsymbol{A}}^{\mathrm{H}}\bar{\boldsymbol{A}})^{-1}\boldsymbol{R_{ss}}^{-1}\big)_{k,k}\xrightarrow[\frac{p}{N}\to 0]{}0\tag{172}$$

Another proof could be done by using the asymptotic equivalence between *Toeplitz* and *Circulant* type matrices [75].

References

[1] H. L. Van Trees, *"Detection, Estimation, and Modulation Theory," Part IV, Optimum Array Processing,* New York: Wiley, 2002.

[2] T. Tuncer and B. Friedlander, *"Classical and Modern Direction-of-Arrival Estimation,"* New York: Academic, 2009.

[3] G. Sun, J. Chen, W. Guo, and K.J.R. Liu, "Signal processing techniques in network-aided positioning," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 12?23, 2005.

[4] A. H. Sayed, A.Taroghat, and N. Khajehnouri, "Network-based wireless location," *IEEE Signal Processing Magazine*, 22(4), pp. 24?40, 2005.

[5] F. Seco, A. R. Jimenez, C. Prieto, J. Roa, and K. Koutsou, "A survey of mathematical methods for indoor localization," in Proceedings of the 6th *IEEE International Symposium on Intelligent Signal Processing*, pp. 9?14, 2009.

[6] R. Zekavat and R. M. Buehrer, *"Handbook of Position Location,"* Hoboken, NJ: John Wiley and Sons, Inc., 2011.

[7] N. Patwari, J. Ash, S. Kyperountas, A. O. Hero, R. M. Moses, and N. S. Correal, "Locating the nodes: Cooperative localization in wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 54?69, 2005.

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

[9] F. C. Schweppe, "Sensor array data processing for multiple signal sources," *IEEE Transactions on Information Theory*, vol. IT, 14, pp. 294-305, 1968.

[10] Y. Bresler and A. Macovski, "Exact Maximum Likelihood Parameter Estimation of Superimposed Exponential Signals in Noise," *IEEE Transactions on Acoustics, Speech, Signal Processing*, vol. 35, no. 10, pp. 1081?1089, 1986.

[11] I. Ziskind and M. Wax, "Maximum likelihood localization of multiple sources by alternating projections," *IEEE Transactions on Acoustics, Speech, Signal Processing*, vol. 36, pp. 1553-1560, 1988.

[12] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm," *IEEE Transactions on Acoustics, Speech, Signal Processing*, vol. 36, pp. 477-489, 1988.

[13] P. Stoica and K. C. Sharman, ?Maximum likelihood methods for direction-of-arrival estimation,? *IEEE Transactions on Signal Processing*, vol. 38, no. 7, pp. 1132-1143, 1990.

[14] J. A. Fessler and A. O. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Transactions on Signal Processing*, vol. 42, no. 10, pp. 2664?2677, 1994.

[15] J. Li, P. Stoica, and Z.-S. Liu, "Comparative study of IQML and MODE direction-of-arrival estimators," *IEEE Transactions on Signal Processing*, vol. 46, no.1, pp. 149-160, 1998.

[16] P. J. Chung and J. F. Bohme. "Comparative convergence analysis of EM and SAGE algorithms in DOA estimation." *IEEE Transactions on Signal Processing*, vol. 49, no. 12, pp. 2940-2949, 2001.

[17] R. 0 . Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. AP-34, pp. 276-280, 1986.

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

[19] P. Stoica and A. Nehorai, "MUSIC, Maximum Likelihood and Cramér-Rao bound," *IEEE Transactions on Acoustics, Speech, Signal Processing*, vol. 37, pp. 720-741, 1989.

[20] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramér-Rao bound: further results and comparisons," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 12, pp. 2140-2150, 1990.

[21] P. Stoica and A. Nehorai, "Performance comparison of subspace rotation and MUSIC methods for direction estimation," *IEEE Transactions on Acoustics, Speech, Signal Processing*, vol. 39, no. 2, pp. 446-453, 1991.

[22] 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 Transactions on Acoustics, Speech, Signal Processing*, vol. 37, no. 12, pp. 1990-1995, 1989.

[23] T. J. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals." *IEEE Transactions on Acoustics, Speech, Signal Processing*, vol. 33, no. 4, pp. 806-811, 1985.

[24] A. L. Swindlehurst and T. Kailath. "A performance analysis of subspace-based methods in the presence of model errors. I. The MUSIC algorithm," *IEEE Transactions on Signal Processing*, vol. 40, no. 7, pp. 1758-1774, 1992.

[25] A. Ferreol, P. Larzabal, and M. Viberg, "On the asymptotic performance analysis of subspace DOA estimation in the presence of modeling errors: case of MUSIC," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 907-920, 2006.

[26] C-MS. See and B-K. Poh, "Parametric sensor array calibration using measured steering vectors of uncertain locations," *IEEE Transactions on Signal Processing*, vol. 47, no. 4, pp. 1133-1137, 1999.

[27] A. Liu, G. Liao, C. Zeng, Z. Yang, and Q. Xu, "An eigenstructure method for estimating DOA and sensor gain-phase errors," *IEEE Transactions on Acoustics, Speech, Signal Processing*, vol. 59, no. 12, pp. 5944-5956, 2011.

[28] T. Svantesson, "The effects of mutual coupling using a linear array of thin dipoles of finite length," *IEEE Statistical Signal and Array Processing Workshop*, 1998.

[29] W. L. Stutzman, and G. A. Thiele, *"Antenna theory and design,"* John Wiley and Sons, 2012.

[30] C-MS. See, "Method for array calibration in high-resolution sensor array processing," *IEE Proceedings-Radar, Sonar and Navigation*, vol. 142, no. 3, pp. 90-96, 1995.

[31] B. C. Ng, and C-MS. See, "Sensor-array calibration using a maximum-likelihood approach." *IEEE Transactions on Antennas and Propagation*, vol. 44, no. 6, pp. 827-835, 1996.

[32] E. K. L. Hung, "Matrix-construction calibration method for antenna arrays," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, no. 3, pp. 819-828, 2000.

[33] P. Cherntanomwong, J. I. Takada, H. Tsuji, and R. Miura, "Modified array calibration for precise angle-of-arrival estimation," *Proc. of 8th International Symposium on Wireless Personal Multimedia Communications*, Denmark, 2005.

[34] A. Leshem and M. Wax, "Array calibration in the presence of multipath," *IEEE Transactions on Signal Processing*, vol. 48, no. 1, pp. 53-59, 2000.

[35] P. Stoica, M. Viberg, K. M. Wong, and Q. Wu, "Maximum-likelihood bearing estimation with partly calibrated arrays in spatially correlated noise fields," *IEEE Transactions on Signal Processing*, vol. 44, no. 4, pp. 888-899, 1996.

[36] M. Pesavento, A. B. Gershman, and K. M. Wong, "Direction of arrival estimation in partly calibrated time-varying sensor arrays," ? *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Salt Lake City, UT, 2001.

[37] C-MS. See, and A. B. Gershman, "Direction-of-arrival estimation in partly calibrated subarray-based sensor arrays," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 329-338, 2004.

[38] B. Wang, Y. Wang, and Y. Guo, "Mutual coupling calibration with instrumental sensors," *Electronics Letters*, vol. 40, no. 7, pp. 406-408, 2004.

[39] R. Boyer and G. Bouleux, "Oblique projections for direction-of-arrival estimation with prior knowledge," *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1374-1387, 2008.

[40] G. Bouleux, P. Stoica, and R. Boyer, "An optimal prior-knowledge-based DOA estimation method," in *17th European Signal Processing Conference*, Glasgow, UK, 2009.

[41] V. Ollier, M. N. El Korso, R. Boyer, P. Larzabal, and M. Pesavento, "Joint ML calibration and DOA estimation with separated arrays," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Shanghai, China, 2016.

[42] B. Friedlander, and A. J. Weiss. "Direction finding in the presence of mutual coupling." *IEEE Transactions on Antennas and Propagation*, vol. 39, no. 3, pp. 273-284, 1991.

[43] J. Pierre and M. Kaveh. "Experimental performance of calibration and direction-finding algorithms." *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 1991.

[44] E. K. L. Hung, "A critical study of a self-calibrating direction-finding method for arrays," *IEEE Transactions on Signal Processing*, vol. 42, no. 2, pp. 471-474, 1994.

[45] I. S. D. Solomon, D. A. Gray, Y. I. Abramovich, and S. J. Anderson, "Receiver array calibration using disparate sources," *IEEE Transactions on Antennas and Propagation*, vol. 47, no. 3, pp. 496-505, 1991.

[46] Q. Bao, C. C. Ko, and W. Zhi, "DOA estimation under unknown mutual coupling and multipath," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 41, no. 2, pp. 565-573, 2005.

[47] A. C. Polycarpou and C. A. Balanis, "Finite-element domain decomposition using an iterative approach: Computation of mutual coupling," *Antennas and Propagation Society International Symposium*, Vol. 2. IEEE, 2000.

[48] F. Sellone and A. Serra, "A novel online mutual coupling compensation algorithm for uniform and linear arrays," *IEEE Transactions on Signal Processing*, vol. 55, no. 2, pp. 560-573, 2007.

[49] S. Lundgren, "A study of mutual coupling effects on the direction finding performance of ESPRIT with a linear microstrip patch array using the method of moments," *Antennas and Propagation Society International Symposium, AP-S, Digest*, Vol. 2. IEEE, 1996.

[50] J. E. F. del Rio, O. M. Conde-Portilla, and M. F. Catedra. "Estimating azimuth and elevation angles when mutual coupling is significant [direction finding]." *Antennas and Propagation Society International Symposium*, IEEE. Vol. 1, 1998.

[51] Y. C. Yu, M. Okada, and H. Yamamoto, "Dummy elements add on both sides of monopole-array assisted Doppler spread compensator for digital terrestrial television broadcasting receiver," in *Proc. IEEE Int. Workshop on Antenna Technology Small Antennas and Novel Metamaterials*, pp. 377?380, 2006.

[52] Z. Ye and C. Liu, "On the resiliency of MUSIC direction finding against antenna sensor coupling," *IEEE Transactions on Antennas and Propagation*, vol. 56, no. 2, pp. 371-380, 2008.

[53] Z. Ye, J. Dai, X. Xu, and X. Wu, "DOA estimation for uniform linear array with mutual coupling." *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 1, pp. 280-288, 2009.

[54] J. Dai, W. Xu, and D. Zhao, "Real-valued DOA estimation for uniform linear array with unknown mutual coupling," *Signal Processing*, vol. 92, no. 9, pp. 2056-2065, 2012.

[55] C. Liu, Z. Ye, and Y. Zhang, "Autocalibration algorithm for mutual coupling of planar array,"*Signal Processing*, vol. 90, no. 3, pp. 784-794, 2010.

[56] M. Pesavento, A. B. Gershman, and K. M. Wong, "Direction finding in partly calibrated sensor arrays composed of multiple subarrays." *IEEE Transactions on Signal Processing*, vol. 50, no. 9, pp. 2103-2115, 2002.

[57] C. Qi, Y. Wang, Y. Zhang, and H. Chen, "DOA estimation and self-calibration algorithm for uniform circular array," *Electronics Letters*, vol. 41, no. 20, pp. 1092-1094, 2005.

[58] R. Goossens and H. Rogier, "A hybrid UCA-RARE/Root-MUSIC approach for 2-D direction of arrival estimation in uniform circular arrays in the presence of mutual coupling," *IEEE Transactions on Antennas and Propagation*, vol. 55, no. 3, pp. 841-849, 2007.

[59] W. Bu-Hong, W. Yong-liang, and C. Hui, "A robust DOA estimation algorithm for uniform linear array in the presence of mutual coupling," *IEEE Antennas and Propagation Society International Symposium*, Vol. 3, pp. 924-927, 2003.

[60] J. Dai, X. Bao, N. Hu, C. Chang, and W. Xu "A recursive RARE algorithm for DOA estimation with unknown mutual coupling," *IEEE Antennas and Wireless Propagation Letters* 13, pp. 1593-1596, 2014.

[61] B. Liao, Z. G. Zhang, and S. C. Chan, "DOA estimation and tracking of ULAs with mutual coupling," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 1, pp. 891-905, 2012.

[62] E. W. J. Ding and B. Su, "A new method for DOA estimation in the presence of unknown mutual coupling of an antenna array," *48th Asilomar Conference on Signals, Systems and Computers,* IEEE, 2014.

[63] A. Bazzi, D. T.M. Slock, L. Meilhac, "Online Angle of Arrival Estimation in the Presence of Mutual Coupling," *IEEE International Workshop on Statistical Signal Processing (SSP)*, 2016.

[64] A. Jaffe and M. Wax, "Single-site localization via maximum discrimination multipath fingerprinting," *IEEE Transactions on Signal Processing*, vol. 62, no. 7, pp. 1718-172, 2014.

[65] T. Su, K. Dandekar, and H. Ling, "Simulation of mutual coupling effect in circular arrays for direction?finding applications," *Microwave and optical technology letters*, vol. 26, no. 5, pp. 331-336, 2000.

[66] H. T. Hui, "A new definition of mutual impedance for application in dipole receiving antenna arrays," *IEEE Antennas and Wireless Propagation Letters*, vol. 3, no. 1, 364-367, 2004.

[67] T. Svantesson, "Direction finding in the presence of mutual coupling," Chalmers University of Technology, Tech. Rep. 307L, 1999.

[68] T. Svantesson, "Modeling and estimation of mutual coupling in a uniform linear array of dipoles," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Phoenix, AZ, March 1999.

[69] T. Kato, "Perturbation theory for linear operators," *Springer Science and Business media*, Vol. 132, 2013.

[70] D. Bau and L. Trefethen, "Numerical linear algebra," *Society for Industrial and Applied Mathematics (SIAM)*, Philadelphia, PA, 1997.

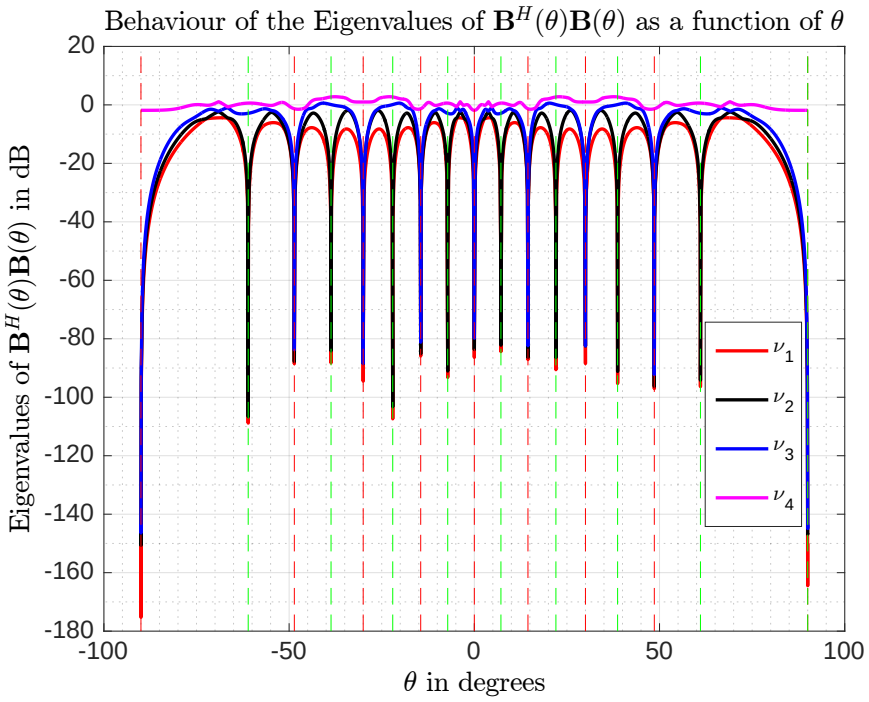
[71] J. J. Sakurai and J. Napolitano, "Modern quantum mechanics," *Addison-Wesley*, 2011.

[72] S. Boyd, L. Vandenberghe, "Convex Optimization," *Cambridge University Press*, 2009.

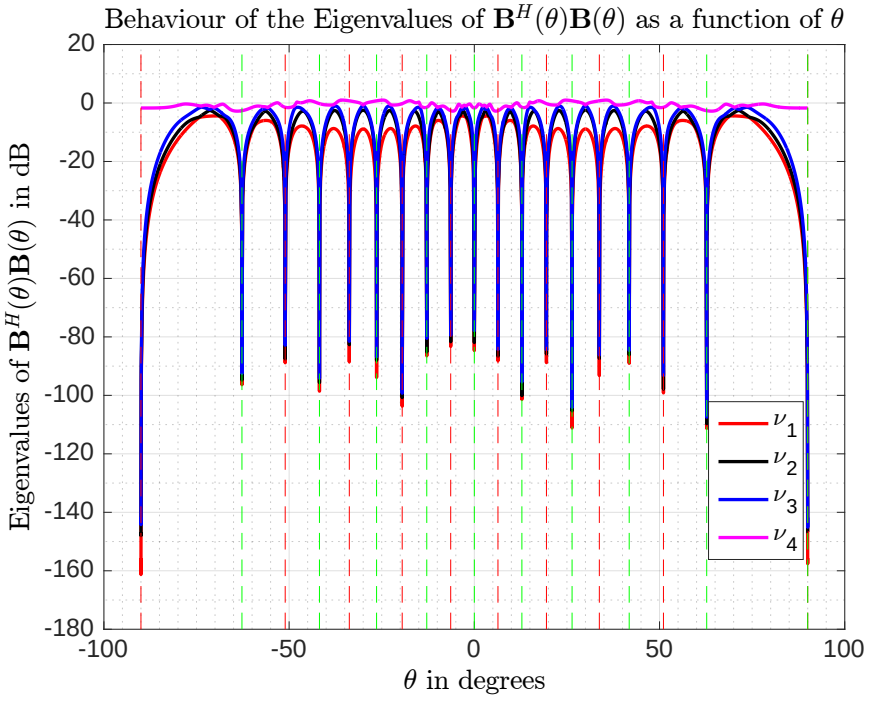
[73] K. B. Petersen and M. S. Pedersen, "The matrix cookbook,", Version 20070905. [Online]. Available: http://www2.imm.dtu.dk/pubdb/p.php?3274, Feb. 2008.

[74] T. Svantesson, "Antennas and Propagation from a Signal Processing Perspective," PhD thesis, Chalmers University of Technology, Gothenburg, Sweden, 2001.

[75] R. M. Gray, *"Toeplitz and Circulant Matrices : A Review,"* Available online at http://www-ee.stanford.edu/ gray/toeplitz.pdf, 2002.

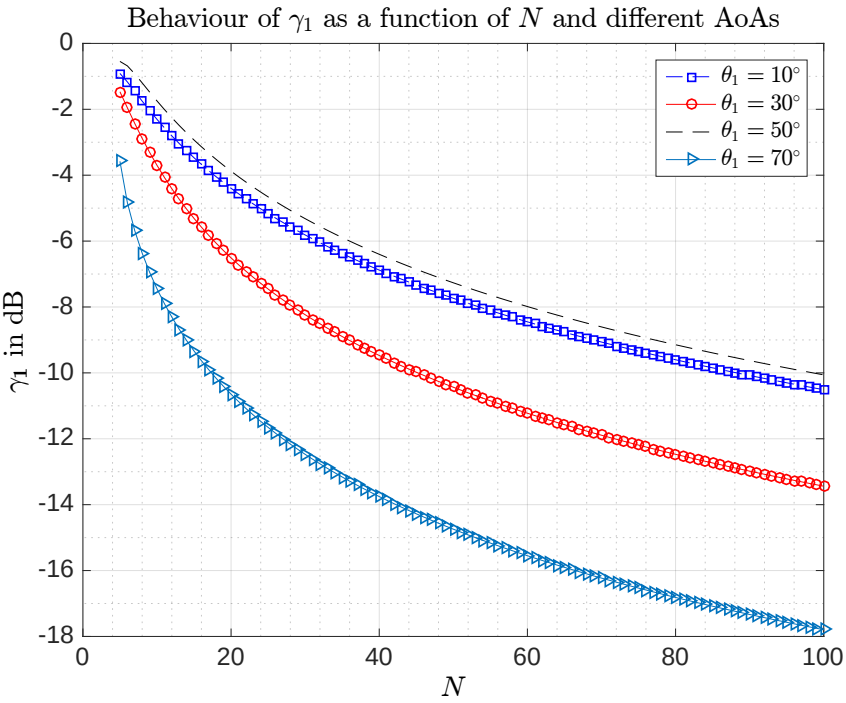


(a) $N = 8$ and $p = 7$

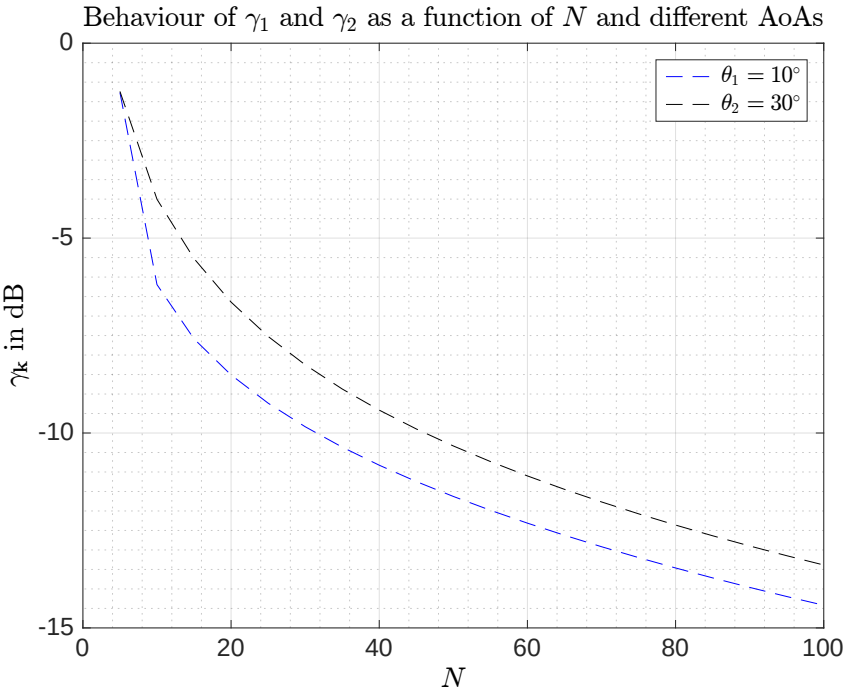


(b) $N = 9$ and $p = 8$

Figure 1: Eigenvalues of $\boldsymbol{B}^{\mathrm{H}}(\theta)\boldsymbol{B}(\theta)$ as a function of θ for different values of N and p.

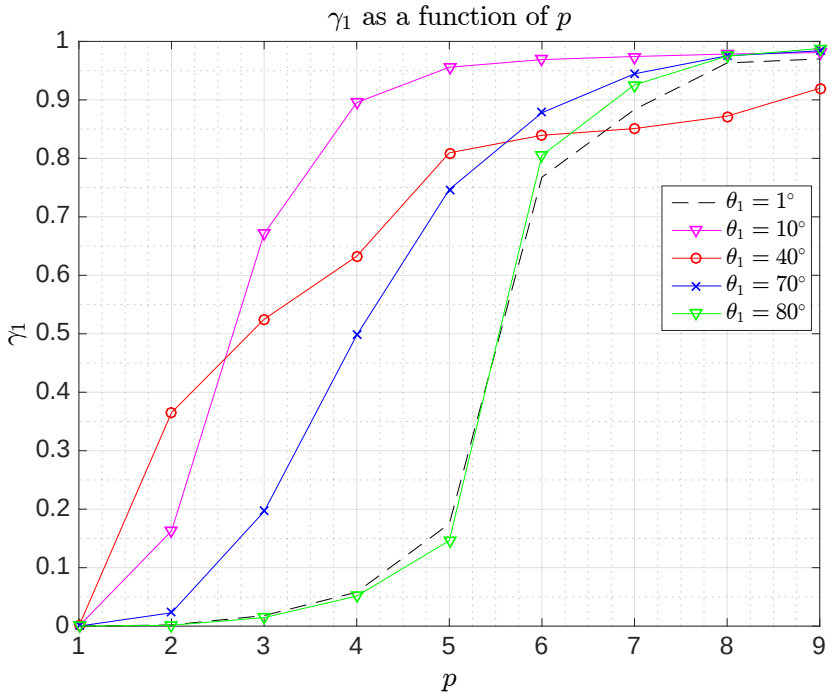


(a) $q = 1$

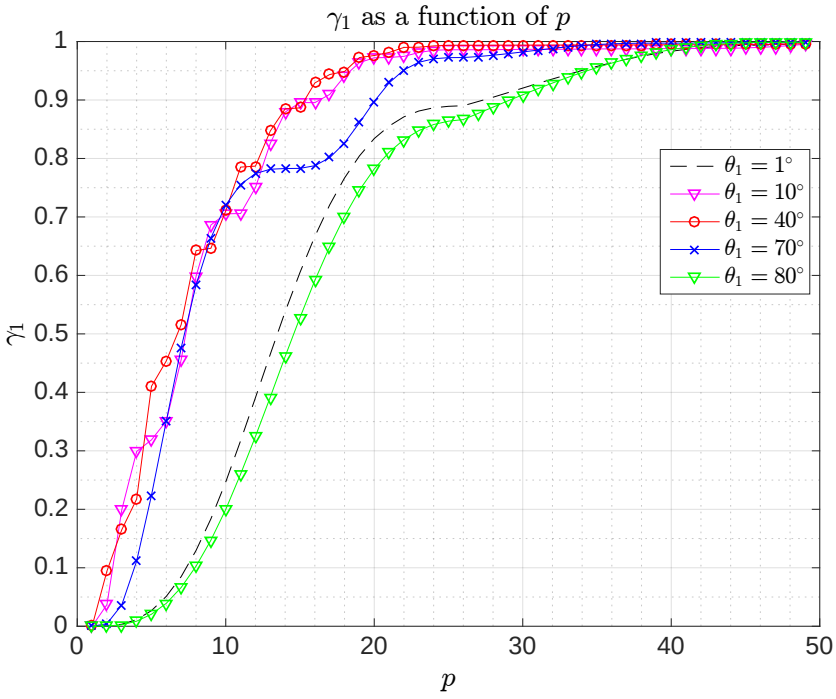


(b) $q = 2$

Figure 2: The behaviour of γ_k for fixed $p = 3$ as a function of N.

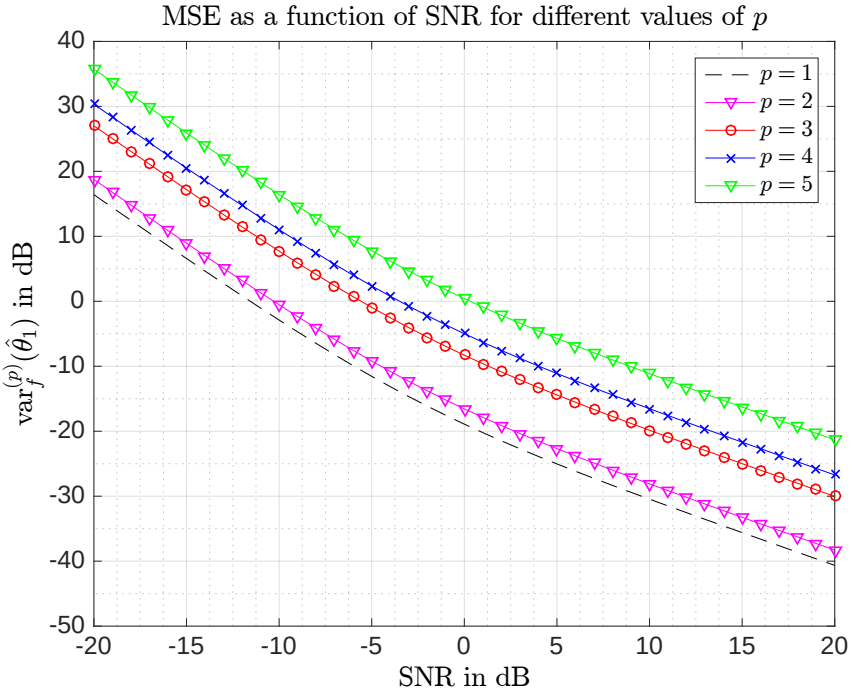


(a) $N = 10$

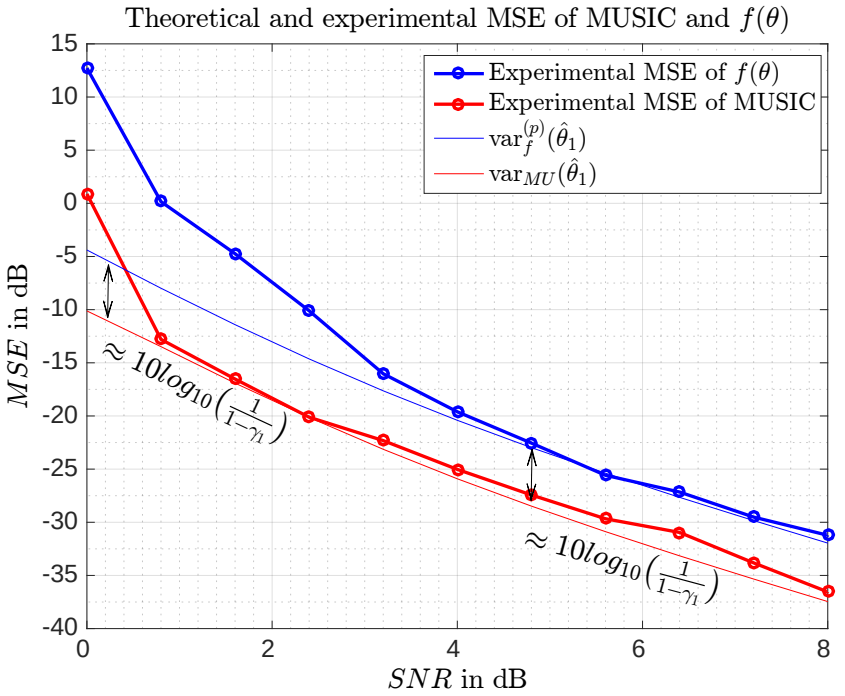


(b) $N = 50$

Figure 3: The behaviour of γ_k for fixed N as a function of p.

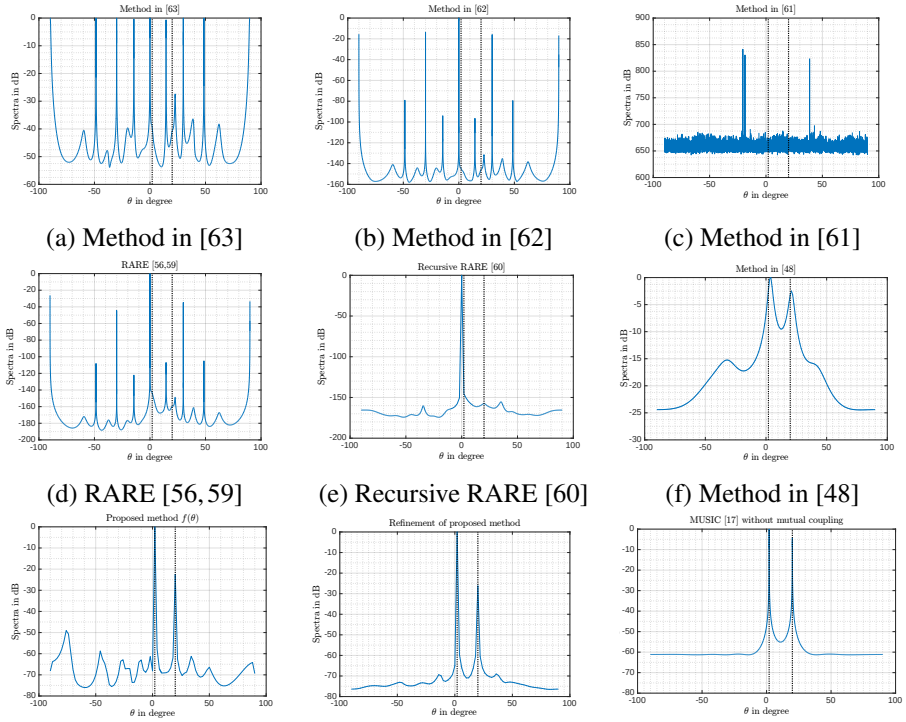


(a) $N = 6$, $q = 1$, and $\theta_1 = 50°$



(b) $N = 6$, $q = 1$, $\theta_1 = 10°$, and $p = 3$

Figure 4: MSE of the proposed method in equation (50) for different values of p and θ_1.

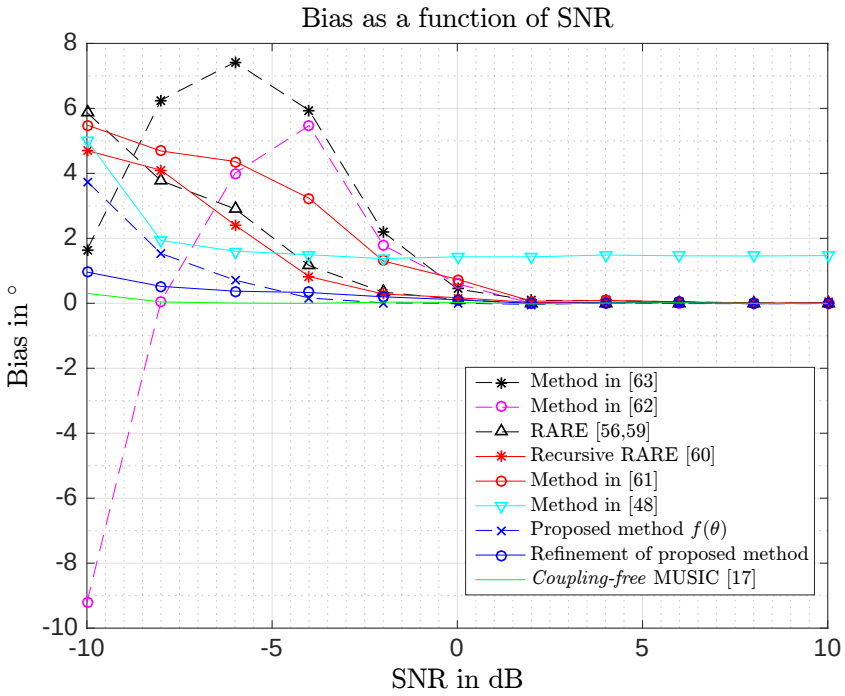


(a) Method in [63] (b) Method in [62] (c) Method in [61]

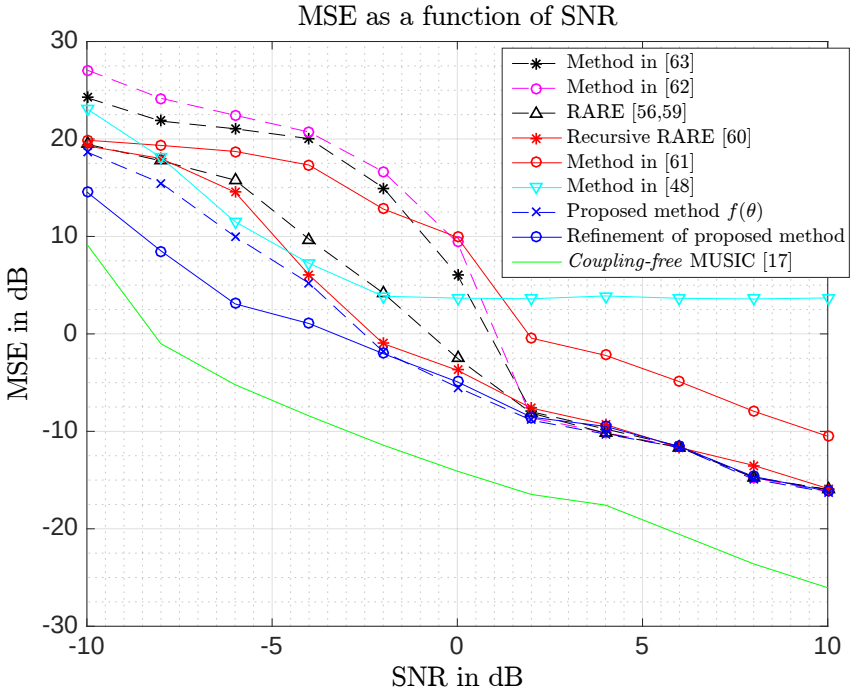
(d) RARE [56, 59] (e) Recursive RARE [60] (f) Method in [48]

(g) Proposed method $f(\theta)$ in equation (50) (h) Refinement of proposed method (i) *Coupling-free* MUSIC [17]

Figure 5: Different normalized spectra (in dB) of methods that estimate AoAs in the presence of mutual coupling.

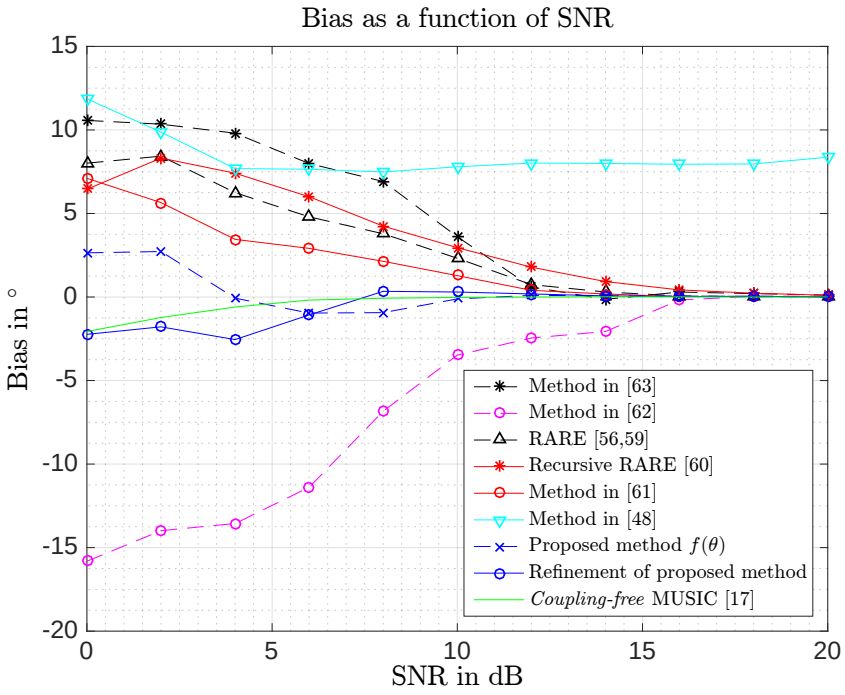


(a) Bias in ° of AoA estimates as a function of SNR.

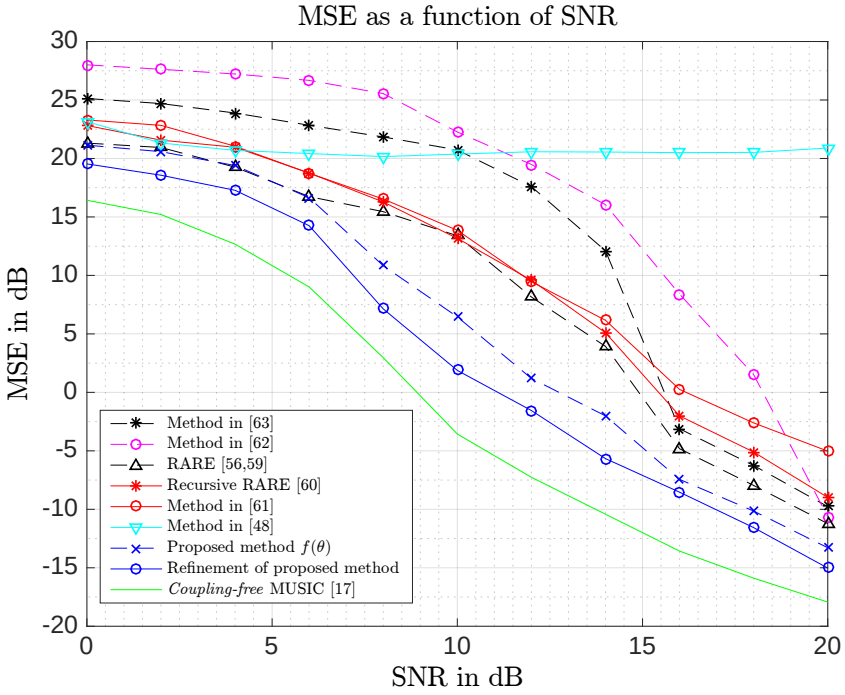


(b) MSE in dB of AoA estimates as a function of SNR.

Figure 6: Bias and MSE of the AoA estimates as a function of SNR for Experiment 1

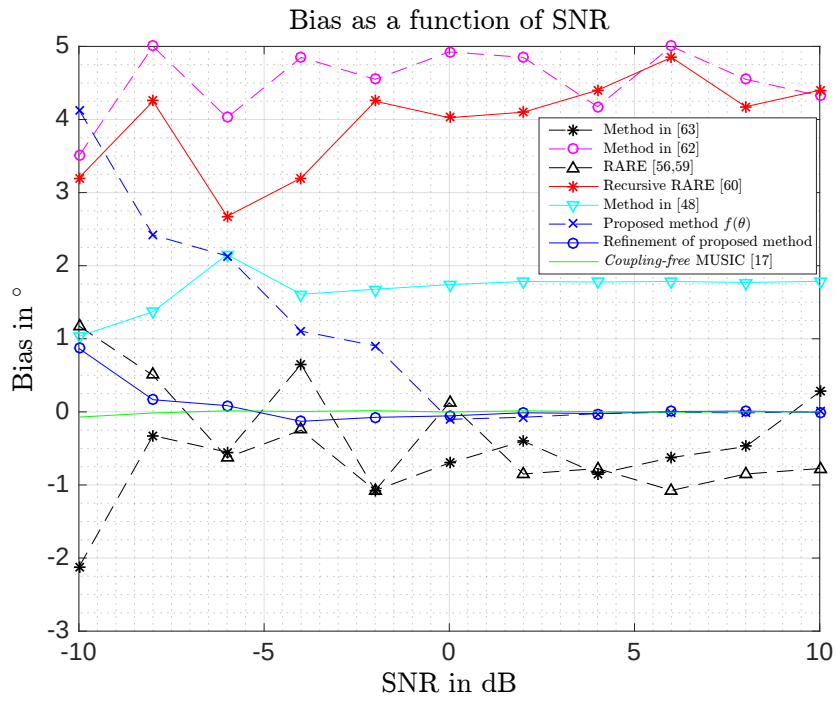


(a) Bias in ° of AoA estimates as a function of SNR.

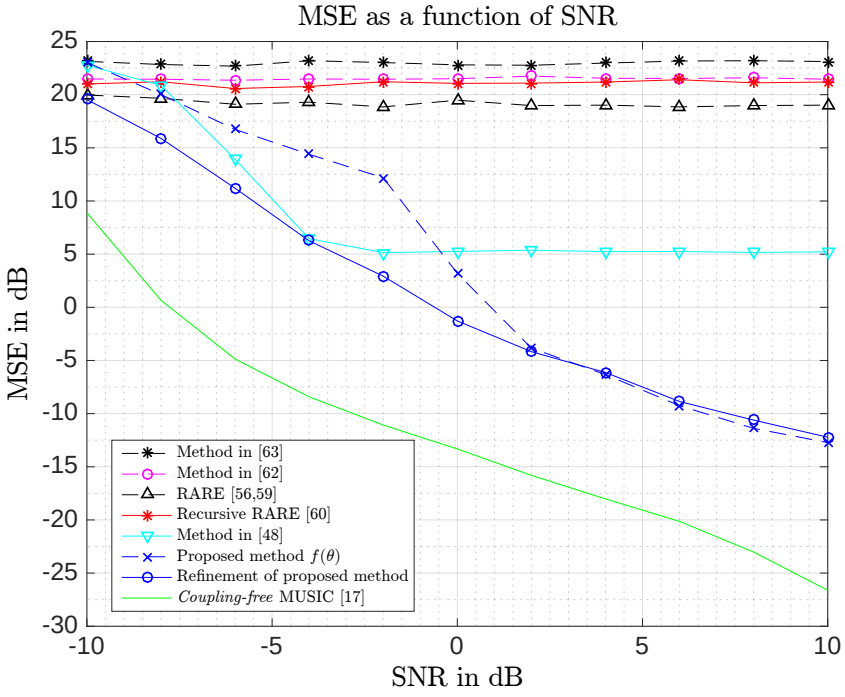


(b) MSE in dB of AoA estimates as a function of SNR.

Figure 7: Bias and MSE of the AoA estimates as a function of SNR for Experiment 2

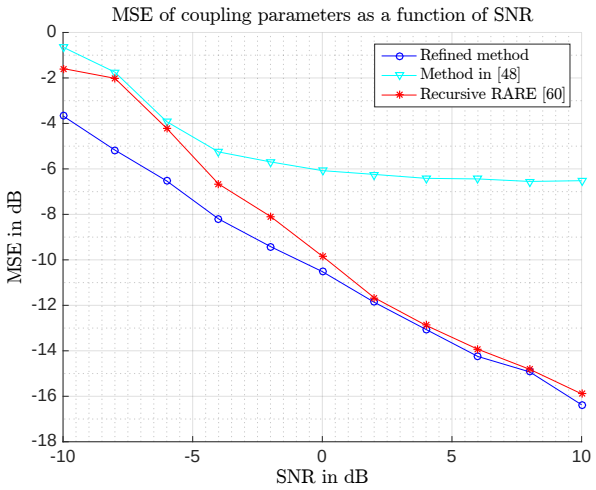


(a) Bias in ° of AoA estimates as a function of SNR.

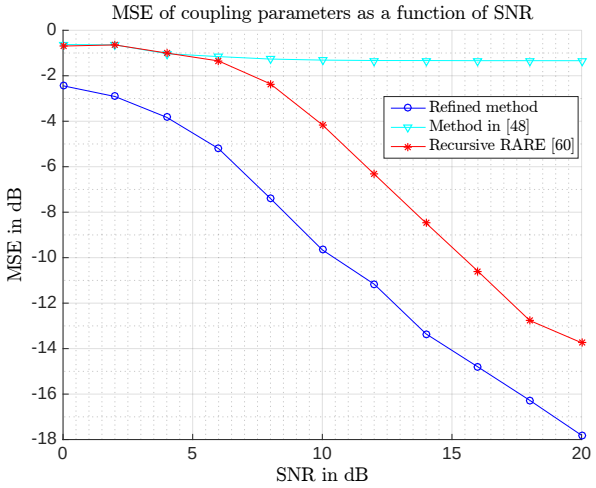


(b) MSE in dB of AoA estimates as a function of SNR.

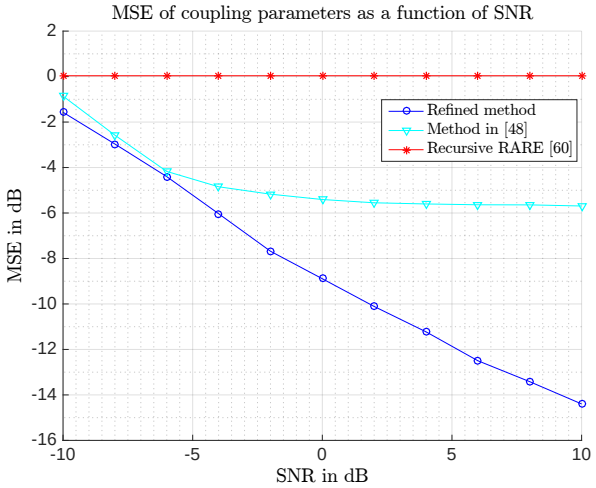
Figure 8: Bias and MSE of the AoA estimates as a function of SNR for Experiment 3



(a) Experiment 1



(b) Experiment 2



(c) Experiment 3

Figure 9: Error on coupling parameters as a function of SNR for the three experiments