Source Detection and Parameter Estimation in Array Processing in the Presence of Nonuniform Noise

Vom Fachbereich 18
Elektrotechnik und Informationstechnik
der Technischen Universität Darmstadt
zur Erlangung der Würde
eines Doktor-Ingenieurs (Dr.-Ing.)
genehmigte Dissertation

von

Saïd Aouada

Referent:	Prof. Dr.-Ing. Abdelhak M. Zoubir
Korreferent:	Prof. Dr. Moeness G. Amin
Tag der Einreichung:	01.02.2006
Tag der mündlichen Prüfung :	12.05.2006

D 17
Darmstädter Dissertation

Zusammenfassung

Behandelt wird das Problem der Quellendetektion mit Hilfe von Sensorgruppen bei ungleichförmiger Verteilung der Rauschleistung auf die Sensoren. Derartiges kann bei verschiedenen Anwendungen wie Funkübertragungen, Radar, Sonar oder in der biomedizinischen Technik beobachtet werden. Das dargestellte Problem ist in der Theorie eindeutig lösbar, abgesehen von dem Fall eines gekoppelten Arrays. Dieser Fall ist in der Praxis jedoch unbedeutend. Die Schätzung der Anzahl der Quellensignale ist äquivalent zur Bestimmung der Dimension des Signal-unterraumes. Unter Betrachtung des idealen Falles mit gleichmäßiger Verteilung der Rauschleistung über alle Sensoren wird der Effekt der Modellstörung auf die Qualität der Signalraumschätzung untersucht und es werden praktische Limi-tierungen in der Trennbarkeit zwischen Signal- und Rauschunterraum identifiziert. Basierend auf dieser Analyse und unter Annahme einer normalförmigen Rauschdichte wird ein sequentieller Hypothesentest vorgeschlagen, wobei die a-symptotische Wahrscheinlichkeitsdichte der Teststatistik hergeleitet wird. Die Teststatistik resultiert aus einer Transformation des Arrays und Anwendung des Gerschgorin-Theorems. Im Falle von nicht-normalförmigem Rauschen oder bei ungenügenden Kenntnissen über die statistische Verteilung der Daten wird das Bootstrap-Verfahren zur empirischen Schätzung der Verteilung verwendet. Dieser Ansatz wurde erfolgreich auf Messdaten aus einem Kernkraftwerk angewandt. Basierend auf der Güte der Schätzung des Signalunterraumes werden verschiedene Informationskriterien vorgeschlagen und deren asymptotische Leistung sowohl analytisch als auch durch Simulationen untersucht. Unter der Annahme, dass ein Kriterium zur Quellendetektion zur Verfügung steht, wird das Problem der Parameterschätzung in ungleichförmigem Rauschen untersucht und ein modi-fizierter approximativer Maximum-Likelihood-Schätzer für gaußverteilte Signale angewandt. Simulationsbeispiele illustrieren die Charakteris-

tiken, Vorteile und Limitierungen der vorgestellten Methoden und vergleichen sie zu bereits exis-tierenden.

Abstract

We address the problem of source detection in array signal processing when the noise over the sensors does not have a uniform power. Such spatial nonuniformity is observed in several applications including communications, radar, sonar and biomedical engineering. The problem of interest is theoretically non-identifiable, however, the only case of non-identifiability is very unrealistic in practical coupled arrays. Estimation of the number of sources is equivalent to the determination of the dimension of the signal subspace. Considering the ideal uniform scenario, we evaluate the effect of noise-power perturbation on the quality of the estimated signal subspace and identify practical limits to the separability between the noise and signal subspaces. Based on this analysis, under the Gaussian-data scenario, we propose a sequential hypothesis test for source detection, deriving an expression for the asymptotic distribution of the proposed test statistics. The latter follows from a transformation of the array and Gerschgorin's theorem. When the data are assumed non-Gaussian, or when no sufficient prior knowledge of the distribution of the data is available, we propose to employ the bootstrap to empirically estimate the distributions of interest. To support the approach, we use a set of real data from a nuclear power plant. Based on the same measure of quality of the estimated signal subspace, we also propose a set of information theoretic criteria and analyze their asymptotic performance both analytically and through simulation. Given a criterion for source detection is available, we investigate the problem of parameter estimation in nonuniform noise and apply a modified approximate maximum likelihood estimator in the stochastic Gaussian case. Simulation examples are provided to illustrate the characteristics, advantages and limitations of the different methods, and compare them to existing methods.

Contents

Abbreviations **x**

Publications **xii**

Acknowledgments **xv**

1 Introduction **1**

1.1 Aims and Objectives 3

1.2 Contributions 3

1.3 Outline 5

2 Data Model and Problem Formulation **6**

2.1 Sensor Array Model 6

2.2 Assumptions 9

2.3 Signal and Noise Subspaces 10

2.4 Overview of Source Detection in Uniform Noise 11

2.4.1 Hypothesis Tests 13

2.4.2 Information Theoretic Criteria 15

2.5 Overview of DOA Estimation in Uniform Noise 18

2.5.1 Deterministic Maximum Likelihood (DML) 18

2.5.2 Stochastic Maximum Likelihood (SML) 19

2.5.3 Multiple Signal Classification (MUSIC) 20

3 Effect of Spatially Nonuniform Noise **22**

3.1 Introduction 22

3.2 Limits of Eigen-based Subspace Separation in Nonuniform Noise . . . 24

3.2.1 Effect of Poor Clustering of the Noise Eigenvalues on Classical Detection Criteria . . . 25
3.2.2 Identifiability of the Dimension of the Signal Subspace . . . 26
3.3 Subspace Estimation Quality . . . 27
3.4 Numerical Examples . . . 31
3.5 Conclusion . . . 33

4 Estimated Subspace Quality (ESQ) Test for Source Detection 35
4.1 Introduction . . . 35
4.2 Subspace Separation . . . 36
4.2.1 Array Transformation . . . 36
4.2.2 Geometric Interpretation . . . 38
4.3 Sequential Hypothesis Test for Source Detection . . . 38
4.4 Numerical Examples . . . 43
4.5 Conclusion . . . 48

5 Nonuniform Information Theoretic Criteria 49
5.1 Introduction . . . 49
5.2 Information Theoretic Criteria for Source Detection . . . 49
5.2.1 Asymptotic Performance . . . 51
5.2.2 Comparison with the Gerschgorin Likelihood Estimator . . . 55
5.3 Numerical Examples . . . 57
5.4 Conclusion . . . 62

6 Application of the Bootstrap to Source Detection 63
6.1 Introduction . . . 63
6.2 Source Detection . . . 64
6.2.1 Sequential Hypothesis Test . . . 64
6.2.2 Multiple Hypothesis Test . . . 64
6.3 Bootstrap-based Detector . . . 65
6.3.1 Estimation of the Null Distribution . . . 66
6.3.2 Studentizing and Bias Correction . . . 67
6.3.3 Bootstrap detection procedure . . . 69
6.4 Numerical Examples . . . 71

6.5 Application to Power Plant Process Monitoring 80
6.5.1 Data model 82
6.5.2 Data Sample and Pre-processing 83
6.5.3 Experimental Results 84
6.6 Conclusion 87

7 Direction of Arrival Estimation in Spatially Nonuniform Noise 88
7.1 Introduction 88
7.2 Existing DOA Estimators for Nonuniform Noise 89
7.2.1 Nonuniform Maximum Likelihood Estimator (MLDOA) 89
7.2.2 Weighted Least Squares Approach 91
7.2.3 Power Domain Solution 92
7.3 DOA Estimation using Approximate Subspace Separation 93
7.4 Nonuniform Approximate Maximum Likelihood 95
7.4.1 Noise Modeling 96
7.4.2 Parameter Space Mapping 97
7.4.3 Parameter Estimation 98
7.5 Identifiability and the Cramér-Rao Bound 101
7.6 Numerical Examples 104
7.7 Conclusion 108

8 Conclusions 110

A Joint Estimation of Angles and Frequencies in Partially Unknown Noise 114
A.1 Approximate Maximum Likelihood (AML) 114
A.2 Extension to 2D: Application to Low-Flying Target Localization . . . 115
A.3 Data Model 117
A.3.1 Model Transformation 118
A.4 Joint Maximum Likelihood Estimation 121
A.5 Cramér-Rao Bound 122
A.6 Numerical Examples 123

B Proofs of Lemmas and Results 128
B.1 Proof of Lemma 3.3.1 128

B.2 Proof of Corollary 3.3.2 131
B.3 Proof of Lemma 5.2.1 132
B.4 Proof of Result 5.2.2 133
B.5 Proof of Result 5.2.3 136
B.6 Proof of Result 5.2.4 136
B.7 Proof of Lemma 5.2.5 137
B.8 Proof of Result 5.2.6 137
B.9 Proof of Lemma A.5.1 138

List of Tables

6.1 Application of source detection to the process data. 85

List of Figures

1.1 Diagram of the research project. 4

2.1 Example of an arbitrary array receiving direct waves from distinct sources. 9

2.2 Geometric representation of the signal subspace. 12

3.1 Effect of noise nonuniformity on the performance of MUSIC. 23

3.2 Example of applications with nonuniform noise. 24

3.3 Approximate asymptotic distribution of $\cos(\gamma_3(2))$. 31

3.4 Signal subspace estimation quality vs SNR and WNPR. 32

3.5 Signal subspace estimation quality vs sample size and WNPR. 33

4.1 Transformed Gerschgorin disks on the complex plane. 39

4.2 Performance of the ESQ hypothesis test vs SNR. 43

4.3 Performance of the ESQ hypothesis test vs sample size. 44

4.4 Performance of the ESQ hypothesis test vs angular separation. 45

4.5 Performance of the ESQ hypothesis test vs WNPR. 46

4.6 ESQ test level in a signal-free environment vs sample size. 47

4.7 ESQ test level in a signal-free environment vs WNPR. 48

5.1 Performance of the nonuniform criteria vs SNR. 57

5.2 Performance of the nonuniform criteria vs sample size. 58

5.3 Performance of the nonuniform criteria vs angular separation. 59

5.4 Performance of the nonuniform criteria vs WNPR. 60

5.5 Performance of the nonuniform criteria vs sample size in noise only. . 61

5.6 Performance of the nonuniform criteria vs WNPR in noise only. . . . 61

6.1 Principle of the bootstrap. 66
6.2 Estimated noise Gerschgorin radii with and without bias correction. . 68
6.3 Performance of the bootstrap tests vs SNR (Gaussian sources and noise). 71
6.4 Performance of the bootstrap tests vs SNR (Laplacian sources, Gaussian noise). 72
6.5 Performance of the bootstrap tests vs SNR (Gaussian sources, Laplacian noise). 73
6.6 Performance of the bootstrap tests vs SNR (Laplacian sources and noise). 74
6.7 Performance of the bootstrap tests vs sample size. 75
6.8 Performance of the bootstrap tests vs angular separation. 76
6.9 Performance of the bootstrap tests vs WNPR. 77
6.10 Level of the bootstrap tests in noise only vs sample size and WNPR. 78
6.11 Performance of the bootstrap tests vs array and sample size. 80
6.12 Diagram of a PWR power plant. 81
6.13 Example of an output signal wave-form. 84
6.14 Sample eigenvalues and Gerschgorin radii. 84
6.15 Estimated process sources. 86

7.1 Variation of the CRB with the nuisance parameters. 103
7.2 Performance of the DOA estimators vs SNR. 105
7.3 Performance of the DOA estimators vs sample size. 106
7.4 Performance of the DOA estimators vs array size. 107
7.5 Performance of the DOA estimators vs WNPR. 108

A.1 Reflection model for a spherical smooth earth. 117
A.2 Data processing over three dimensions. 119
A.3 Performance of the joint estimators (DOA) vs SNR. 123
A.4 Performance of the joint estimators (Doppler) vs SNR. 124
A.5 Performance of the joint estimators (DOA) vs correlation coefficient. . 125
A.6 Performance of the joint estimators (Doppler) vs correlation coefficient.126
A.7 Tracking of the direct and reflected DOAs for the different methods. . 127

Abbreviations

AIC	Akaike Information Criterion
AML	Approximate Maximum Likelihood
AML-OP	Oblique Projection Approximate Maximum Likelihood
AP	Alternating Projections
COMET	Covariance Matching Estimation Techniques
CRB	Cramér-Rao Bound
DML	Deterministic Maximum Likelihood
DOA	Direction of Arrival
EDC	Efficient Detection Criterion
EM	Expectation Maximization
ESQ	Estimated Subspace Quality
FIM	Fisher Information Matrix
GA	Genetic Algorithms
GAIC	Gerschgorin Akaike Information Criterion
GLE	Gerschgorin Likelihood Estimator
GMDL	Gerschgorin Minimum Description Length
iid	Independent and Identically Distributed
IQML	Iterative Quadratic Maximum Likelihood
LL	Log-Likelihood
LR	Likelihood Ratio
MDL	Minimum Description Length
MHT	Multiple Hypotheses Test

ML	Maximum Likelihood
MLDOA	Maximum Likelihood DOA Estimator in Nonuniform Noise
MN	Minimum Norm
MODE	Method of Direction Estimation
MSE	Mean Square Error
MUSIC	Multiple Signal Classification
MVD	Minimum Variance Distortionless Beamformer
NB	Narrowband
NSF	Noise Subspace Fitting
NU-AIC	Nonuniform Akaike Information Criterion
NU-MDL	Nonuniform Minimum Description Length
NU-MUSIC	Nonuniform Multiple Signal Classification
OSAIC	Order Statistics Akaike Information Criterion
OSMDL	Order Statistics Minimum Description Length
OSMLE	Order Statistics Maximum Likelihood Estimator
pdf	Probability Density Function
PDS	Power Domain Solution
PRI	Pulse Repetition Interval
RCS	Radar Cross Section
SHT	Sequential Hypothesis test
SML	Stochastic Maximum Likelihood
SNR	Signal-to-Noise Ratio
SPH	Sphericity Test
SRB	Holm's Sequential Rejective Bonferroni Procedure
SSF	Signal Subspace Fitting
TDOA	Time Difference of Arrival
WLS	Weighted Least Squares
WNPR	Worst Noise Power Ratio
WSSF	Weighted Signal Subspace Fitting
2D	Two-dimensional

Publications

The following publications arose as a result of this research work:

Journal papers

- S. Aouada and A. M. Zoubir, "Source Detection using Signal-Subspace Quality Assessment and Gerschgorin Radii, Part I - Hypothesis Tests for Gaussian Data", *IEEE Trans. Signal Processing*, (under review).

- S. Aouada and A. M. Zoubir, "Source Detection using Signal-Subspace Quality Assessment and Gerschgorin Radii, Part II - Information Theoretic Criteria", *IEEE Trans. Signal Processing*, (under review).

- S. Aouada and A. M. Zoubir, "Source Detection using Signal-Subspace Quality Assessment and Gerschgorin Radii, Part III - Data of Unknown Distribution", *IEEE Trans. Signal Processing*, (under review).

- M. Djeddou, A. Belouchrani and S. Aouada, "Maximum Likelihood Angle-Frequency Estimation in Partially Known Correlated Noise for Low-Elevation Targets", *IEEE Trans. Signal Processing*, Vol. 53, No. 8, pp. 3057-3064, Aug. 2005.

Conference papers

- R. Aouada, S. Aouada, G. d'Urso, and A. M. Zoubir. "Source Detection and Separation in Power Plant Process Monitoring: Application of the Bootstrap". *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toulouse, France, May 2006.

- S. Aouada, D. Traskov, N. d'Heureuse, and A. M. Zoubir, "Application of the Bootstrap to Source Detection in Nonuniform Noise", *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Philadelphia, USA, Mar. 2005.

- S. Aouada, M. Bekara, A. M. Zoubir, G. Fleury and C. M. S. See, "A Gerschgorin-Kullback Criterion for Source Number Detection in Nonuniform Noise and Small Samples", *Proc. IEEE International Workshop on Sensor Array and Multichannel Signal Processing (SAM)*, Sitges-Barcelona, Spain, Jul. 2004.

- S. Aouada, A. M. Zoubir and C.M.S. See, "Source Detection in the Presence of Nonuniform Noise", *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Montreal, Canada, May 2004.

- A. M. Zoubir and S. Aouada, "High Resolution Estimation of Directions of Arrival in Nonuniform Noise", *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Montreal, Canada, May 2004.

- M. Djeddou, A. Belouchrani and S. Aouada, "Approximate ML Direction Finding in Spatially Correlated Noise Using Oblique Projections", *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Montreal, Canada, May 2004.

- A. Belouchrani and S. Aouada, "Maximum Likelihood Joint Angle and Delay Estimation in Unknown Noise Fields", *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Hong Kong, China, Apr. 2003.

- S. Aouada, A. M. Zoubir and C. M. S. See, "A Comparative Study on Source Number Detection", *Proc. International Symposium on Signal Processing and its Applications (ISSPA)*, Paris, France, Jul. 2003.

- M. Djeddou, S. Aouada and A. M. Zoubir, "Maximum Likelihood Angle-Frequency Parameter Estimation in Unknown Noise Fields for Low Elevation Target Tracking", *Proc. International Symposium on Signal Processing and its Applications (ISSPA)*, Paris, France, Jul. 2003.

- S. Aouada and A. M. Zoubir, "Maximum Likelihood DOA Estimation in Nonuniform Noise", *Proc. IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, Marrakech, Morocco, Dec. 2002.

Acknowledgments

I would like to thank Prof. Abdelahk M. Zoubir for his advice and support. I would also like to thank all people who, at different levels (colleagues, technical and administrative staff, etc), made the completion of this work easier. Special thanks to EDF in Chatou, France, and in particular Rostom, for providing the real data set and assisting with the experiments.

Above all, I am grateful to my family for their support and patience.

Chapter 1

Introduction

Source localization using an array of sensors has always been at the core of signal processing research [51, 91, 92, 99]. Localization often translates to the estimation of directions of arrival (DOA) and/or other related quantities characterizing the source signals from a parametric model [12, 107]. A prerequisite condition for correct localization is that the number of sources is known with enough accuracy prior to parameter estimation [99, 106, 108]. Applications of both source number estimation or source detection and DOA estimation include wireless communications, radar, sonar, biomedical engineering and seismic exploration [29, 47, 99, 108, 121].

Source detection employs model order selection techniques. Several detection schemes with a number of variants have been proposed and analyzed in the literature, ranging from hypothesis testing [17, 27, 37, 112, 120] to information theoretic criteria [1, 60, 80, 106, 113].

DOA estimation algorithms can be classified into two families. The first family corresponds to spectral-based methods, including beamforming [51, 99] and subspace separation techniques [24, 83, 104]. The second family encompasses parametric methods, i.e., maximum likelihood (ML) techniques [13, 42, 53, 65, 91].

The major classic detection and estimation techniques are reviewed in more detail in the sequel.

Fundamental detection and estimation methods rely on the key assumption that the background noise is spatially uniform, i.e., the noise powers are assumed to be identical over all array sensors. However, in many applications, where the array structure and the noise environment remain unknown or change slowly with time

[55, 74], the assumption of spatial uniformity of the background noise is not necessarily valid. Typical practical cases where spatial *nonuniformity* is observed include magneto-encephalography systems with a large number of sensors [47], sparse arrays with prevailing external noise and hardware non-idealities in receiving channels [34, 69, 97, 103].

If the noise powers are different from one sensor to another, performance of the standard detection and estimation methods can degrade dramatically due to the induced model mismatch. Similarly, methods designed for colored noise cannot be expected to yield satisfactory performance because they ignore important *a priori* information on the structure of the noise samples. In addition, these methods often deal with the detection or estimation problem following a case-by-case approach and result in a considerable computation effort, restricting them to pure theoretical analysis.

Several attempts were made to generalize the detection schemes to cases where the noise covariance matrix is not an identity matrix. Hypothesis tests were proposed in [37, 89] for the case of spatially correlated noise under the assumption that the noise covariance matrix has a band structure. The approach in [37] can be extended to nonuniform noise but can use half of the available sensors at most, thus making it a very restrictive approach. An information criterion was proposed based on Gerschgorin's theorem to increase the robustness against model mismatch. However, similarly to the other classical approaches, the method fails to correctly detect the sources in nonuniform noise because it incorporates the erroneously ordered eigenvalues of the data covariance matrix.

For estimation under the uniform noise scenario, the latter key assumption is inherent to many useful simplifications, leading to a reduction of the dimension of the unknown parameter space, as well as the resulting computational load [36, 76]. Few methods have been proposed to specifically deal with the spatial nonuniformity of the noise. A DOA estimator based on the deterministic ML was proposed in [76], leading to a non-convex optimization problem. It is important to mention that while the nonuniform ML of [76] provides accurate estimates at a relatively high computational cost, simplifications based on subspace methods, or specific high-resolution methods are not available, as no suitable subspace criterion has been proposed for nonuniform noise.

1.1 Aims and Objectives

Taking into account the limitations of classical detectors in nonuniform noise, we address in this work the problem of estimating the dimension and parameters of the signal subspace without using the erroneously ordered eigenvalues of the covariance matrix of the data. It is important to stress that the assumption of spatially white noise is maintained.

Throughout the analysis, we emphasize the applicability of the proposed solutions. We therefore favor computation-efficient and implementable approaches over iterative and case-specific strategies inspired from scenarios with colored noise. We also support the most general proposed detection scheme by a test involving real data.

More particularly, we propose to perform the following:

- Analyze the effect of noise-power perturbation on the quality of the estimated signal subspace,
- develop simple detection algorithms to estimate the dimension of the signal subspace, with a practical accuracy,
- analyze the performance of the proposed detectors and extend the approach to wider distributions,
- upon correctly detecting the number of sources, investigate parameter estimation in nonuniform noise.

The steps followed in our work can be summarized by the diagram of Figure 1.1.

1.2 Contributions

The major contributions of this work are:

- A hypothesis test relying on a measure of the *estimated subspace quality* (ESQ) for source detection, based on the discriminating property of the transformed Gerschgorin radii. This detector follows from an array transformation and is considerably more robust to noise-power variation than classical detectors, while being simple to implement.

- A nonuniform version of information theoretic criteria for source detection, following the same Gerschgorin radii approach and an exclusion of the explicit contribution of the sample eigenvalues.

- Robust bootstrap versions of the above ESQ hypothesis test, suitable for a wide range of data distributions.

- An approximate stochastic ML estimator of the DOAs in nonuniform noise, for a specific array configuration.

- An extension of the estimation approach to the two-dimensional (2D) case, through the example of joint DOA-Doppler frequency estimation.

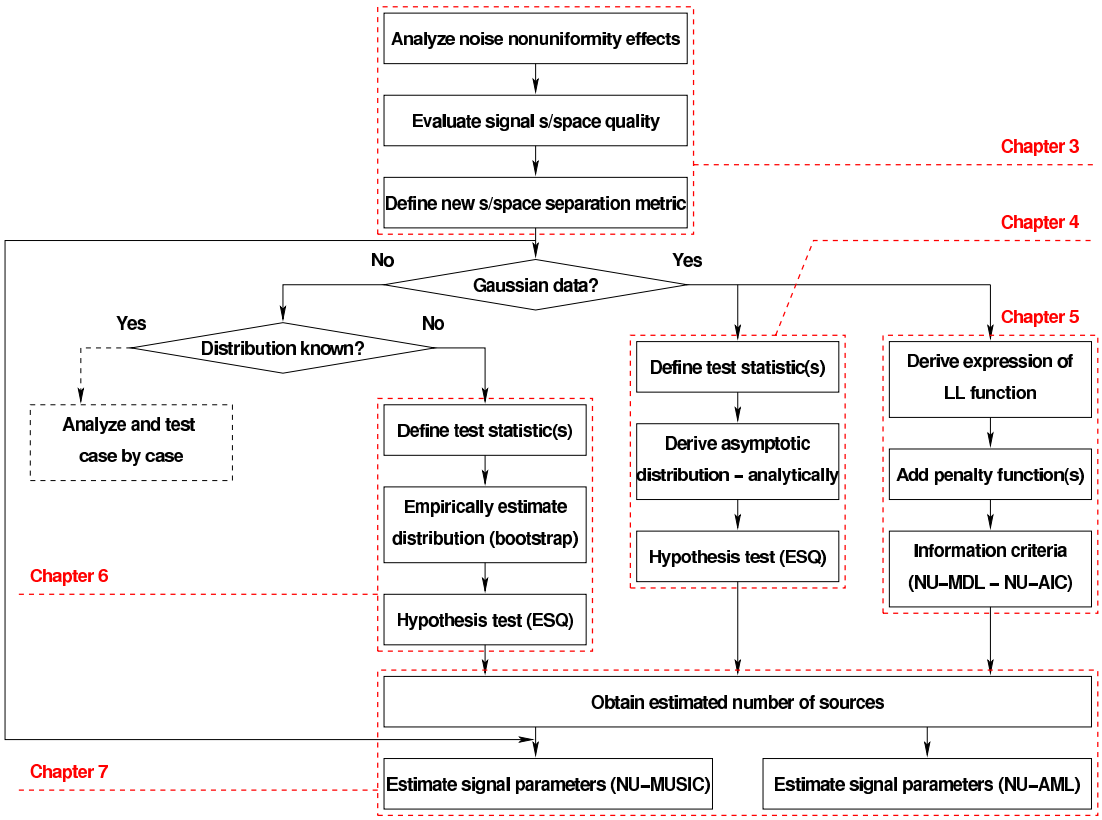

Figure 1.1: Summarized steps of the undertaken research work.

1.3 Outline

The dissertation is organized as follows:

Chapter 2: The basic assumptions that frame the considered data model, and an overview of the principal detection and estimation techniques in uniform noise are presented.

Chapter 3: The effect of spatial nonuniformity of the noise is described and the problem of separating the signal and noise subspaces is formulated. In particular, the statistical properties of the different metrics are outlined and the limitations of the subspace separation approach are taken into account.

Chapter 4: Following the analysis of the previous chapter, the ESQ hypothesis test is derived and its performance is assessed through simulation examples.

Chapter 5: Alternatively, a set of nonuniform detection criteria are derived and their performance is analyzed both analytically and through simulation examples, highlighting their superiority over classical approaches.

Chapter 6: An alternative formulation of the ESQ test using the bootstrap is presented. In particular a multiple hypotheses test (MHT) is applied to real data to illustrate the validity of the approach in practice, irrespective of the particular data distribution.

Chapter 7: DOA estimation in nonuniform noise is investigated in this chapter and the proposed estimator is compared to existing techniques through simulations. The principle behind the approach, namely the linear expansion of the noise covariance matrix, is also employed in the 2D case illustrated in **Appendix A**.

While concluding remarks and perspectives are summarized in **Chapter 8**, the various proofs and detailed derivations are left to **Appendix B**.

Chapter 2

Data Model and Problem Formulation

2.1 Sensor Array Model

In what follows, we focus on the sensor array at the receiver, regardless of the nature of the emitter such as its technology, configuration or physical nature.

We restrict ourselves to the case of a finite number of point sources emitting signals which are either physically independent in space, or a result of the propagation via distinct paths of the signal emanating from a single source. The sources are assumed to be located in the far field. This assumption permits a reasonable modeling of source propagation by plane waves, therefore simplifying its parameterization. Additionally, for simplicity and without loss of generality, we assume that the sources are coplanar. This assumption simplifies the array spatial response to a parameterization in the 2D case at most.

We also restrict our analysis to the narrowband (NB) assumption on the signals. Equivalently, we assume that the array aperture (i.e., its physical size measured in signal wavelengths) is significantly smaller than the inverse relative signal bandwidth. The assumption of NB signals essentially implies that the array spatial response to an impinging signal is constant and does not vary with time. From a system representation point of view, the array response is modeled as a linear instantaneous mixture of the received signals.

Without noise, for a source signal $s(t)$ impinging on an array composed of M

sensors arranged in a given geometry, the M-dimensional array output vector is given by

$$\mathbf{x}(t) := \mathbf{a}s(t) \tag{2.1}$$

where $\mathbf{a} := [a_1, \ldots, a_M]^T$ is the array steering vector, and $(\cdot)^T$ denotes matrix transpose. If p signals impinge on an M-sensor array, the noise-free output vector at instant t follows from superposition as

$$\mathbf{x}(t) := \sum_{q=1}^{p} \mathbf{a}_q s_q(t) \tag{2.2}$$

where $\mathbf{a}_q$ denotes the array response vector to the q-th baseband[1] signal waveform $s_q(t)$, for $q = 1, \ldots, p$.

In matrix form, the array output of (2.2) can be rewritten as

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \tag{2.3}$$

where the array steering matrix and the vector of signal waveforms are defined as

$$\mathbf{A} \;:=\; [\mathbf{a}_1, \ldots, \mathbf{a}_p] \tag{2.4}$$

and

$$\mathbf{s}(t) \;:=\; [s_1(t), \ldots, s_p(t)]^T \tag{2.5}$$

respectively.

In the presence of additive noise, represented by vector $\mathbf{n}(t)$, the model commonly used in array processing is obtained as

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t). \tag{2.6}$$

If the number of sources p is unknown *a priori*, we propose to estimate it, i.e., detect the sources. For the problem of detection, the functional form of the array response vector $\mathbf{a}$ is not of importance. However, the response to distinct sources must materialize in linearly independent steering vectors $\mathbf{a}_q$, $q = 1, \ldots, p$, resulting in a full column-rank steering matrix $\mathbf{A}$.

[1]In practice, the carrier frequency of the signal is normally dropped before further processing.

With a known number of sources p, we are interested in localizing the different sources in space. This localization is usually performed parametrically by identifying the DOA of each signal. DOA estimation requires an explicit parameterization of the data model. Thus, the model (2.6) becomes

$$\mathbf{x}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}(t) + \mathbf{n}(t) \tag{2.7}$$

where

$$\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_p)] \tag{2.8}$$

and the vector of DOAs of each received signal is given by

$$\boldsymbol{\theta} = [\theta_1, \ldots, \theta_p]^T. \tag{2.9}$$

The uniform linear array (ULA) configuration is an example widely employed in array signal processing [51, 99]. When the uniform separation between the array elements equals half the received wavelength, the functional form of the M-element ULA steering vector in terms of a DOA θ, satisfies a Vandermonde structure, and is simplified to

$$\mathbf{a}(\theta) = \left[1, e^{-j\pi\cos(\theta)}, \ldots, e^{-j(M-1)\pi\cos(\theta)}\right]. \tag{2.10}$$

Note that a source signal can be associated with a number of characteristic parameters, such as a time-difference of arrival (TDOA), or a DOA in elevation and in azimuth for general 2-D configurations. For simplicity, the analysis to follow is restricted to the case of a single parameter θ, although extension to multiple parameters per source is straightforward [51] (see Appendix A).

The knowledge of the array manifold is capital in evaluating the array's operation limits, both for source detection and DOA estimation [49]. A quantitative assessment of the shape and orientation of the array manifold provides a measure of its effect on the array's operation. This assessment, which is part of the array calibration in practice, can be accomplished effectively through a study of the manifold's differential geometry [64]. For instance, it was shown in [64] that the detection and resolution capabilities of a ULA are a function of the rate of change of the *arc length* and the *first curvature* of the array manifold curve (given by the functional form of the spatial response vector $\mathbf{a}(\boldsymbol{\theta})$). It is worth mentioning that array calibration

is an important and tedious practical operation. In our work, we assume that the array manifold is known with enough accuracy and that the limits on the spatial resolution of the array are known.

In practice, the sensor output is appropriately pre-processed and sampled at a given rate. The signal samples are collected at time instants $t = 1, \ldots, L$. For conciseness, we use the same notation, $\mathbf{x}(t)$, $t = 1, \ldots, L$, to denote a sample from a multi-channel random process, and for the generating model (2.7).

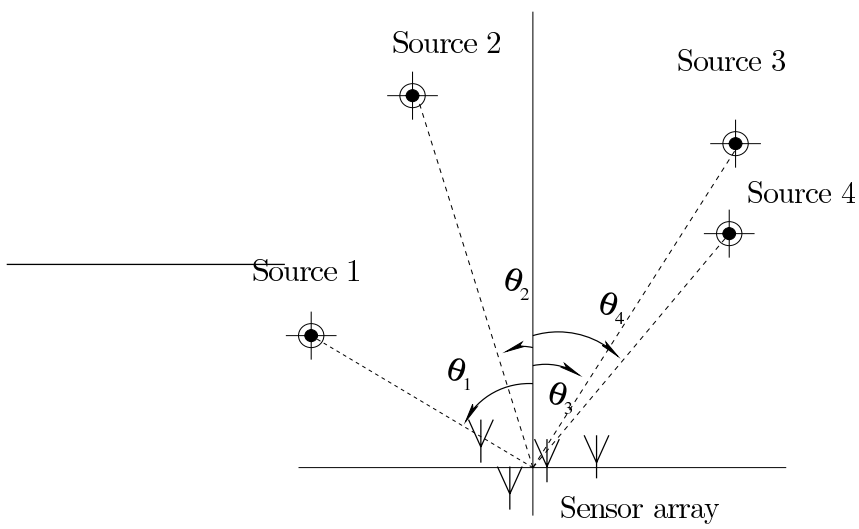

Figure 2.1: Example of an arbitrary array receiving direct waves from distinct sources, located in the far-field at DOAs θ_q, $q = 1, \ldots, p$.

2.2 Assumptions

For both source detection and DOA estimation, the linear system described by (2.7) must be over-determined, i.e., the condition on the number of sources is $p < M$.

The signal parameters of interest are spatial in nature and thus require the cross-covariance information among the various sensors, i.e., the spatial covariance matrix, which is given by

$$\mathbf{R} = E\left\{\mathbf{x}(t)\mathbf{x}^H(t)\right\} \tag{2.11}$$

where $E(\cdot)$ denotes expectation and $(\cdot)^H$ stands for Hermitian transpose.

The source signals are assumed to be spatially independent. Unless otherwise stated, the source samples are modeled in the time domain as unknown random processes. They are independent and identically distributed (iid), and have covariance matrix $\mathbf{R_s}$ with the following diagonal structure:

$$\begin{aligned} \mathbf{R_s} &= E\left\{\mathbf{s}(t)\mathbf{s}^H(t)\right\} \\ &= \operatorname{diag}\left\{\sigma_{s_1}^2, \ldots, \sigma_{s_p}^2\right\} \end{aligned} \tag{2.12}$$

with $\sigma_{s_q}^2$ being the power of the q-th source signal.

The additive noise is modeled as discrete interference sources over each sensor, and is independent of the sources. Its samples are independent from one sensor to another, and temporally white, with the same distribution, i.e., iid. The noise is assumed to satisfy the following:

$$E\left\{\mathbf{n}(t_1)\mathbf{n}^H(t_2)\right\} = \mathbf{Q}\delta_{t_1 t_2}; \qquad E\left\{\mathbf{n}^T(t_1)\mathbf{n}(t_2)\right\} = 0. \tag{2.13}$$

In the general case of spatial independence, the noise covariance matrix $\mathbf{Q}$ has a diagonal structure, as follows:

$$\begin{aligned} \mathbf{Q} &= E\left\{\mathbf{n}(t)\mathbf{n}^H(t)\right\} \\ &= \operatorname{diag}(\mathbf{q}) \end{aligned} \tag{2.14}$$

with

$$\mathbf{q} := \left[\sigma_1^2, \ldots, \sigma_M^2\right]^T \tag{2.15}$$

where σ_m^2 is the noise power at the m-th sensor, $m = 1, \ldots, M$.

Taking into account (2.13) and (2.14), the data covariance matrix (2.11) can rewritten as

$$\mathbf{R} = \mathbf{A}(\boldsymbol{\theta})\mathbf{R_s}\mathbf{A}^H(\boldsymbol{\theta}) + \mathbf{Q}. \tag{2.16}$$

2.3 Signal and Noise Subspaces

The eigen-decomposition of matrix $\mathbf{R}$ is given by

$$\begin{aligned} \mathbf{R} &= \sum_{m=1}^{M} \lambda_m \mathbf{e}_m \mathbf{e}_m^H \\ &= \mathbf{E}\boldsymbol{\Lambda}\mathbf{E}^H \end{aligned} \tag{2.17}$$

with

$$\mathbf{\Lambda} := \operatorname{diag}\{\lambda_1, \ldots, \lambda_M\} \tag{2.18}$$

and

$$\mathbf{E} := [\mathbf{e}_1, \ldots, \mathbf{e}_M] \tag{2.19}$$

where λ_m and $\mathbf{e}_m$, $m = 1, \ldots, M$, are the eigenvalues of $\mathbf{R}$ and their corresponding eigenvectors, respectively.

Examining (2.16), one can see that in the noise-free case, the array output is fully confined to a p-dimensional subspace of the complex M-dimensional space, $\mathbb{C}^M$, which is spanned by the steering vectors, i.e., the columns of $\mathbf{A}(\boldsymbol{\theta})$. In such a case, the $M - p$ smallest eigenvalues are all equal to zero. We then have [44]

$$\operatorname{span}\{\mathbf{A}(\boldsymbol{\theta})\} = \operatorname{span}\{\mathbf{E_s}\} \tag{2.20}$$

where $\operatorname{span}\{\mathbf{A}\}$ is the range space of the columns of $\mathbf{A}$, and $\mathbf{E_s}$ denotes the matrix formed by the column concatenation of the signal subspace eigenvectors, corresponding to the non-zero eigenvalues. Determining the basis $\mathbf{E_s}$ with enough precision constitutes a solution to most detection and parametric estimation problems in array processing.

If the power of the noise over all the sensors is the same, the noise covariance matrix defined by (2.14) reduces to an identity matrix (up to a scalar), i.e., $\mathbf{Q} = \sigma^2\mathbf{I}$, where $\sigma^2 = \sigma_1^2 = \ldots = \sigma_M^2$. The set of eigenvalues (ordered by magnitude), or simply eigenspectrum, are thus scaled up by the addition of the noise floor σ^2. This assumption of *uniform* noise provides a convenient means to distinguish between the signal and noise subspaces. If it is not verified, all classical detection and estimation algorithms fail to perform satisfactorily .

2.4 Overview of Source Detection in Uniform Noise

Source detection is a fundamental problem in many areas of signal processing. It is a prerequisite step for other processing such as source estimation.

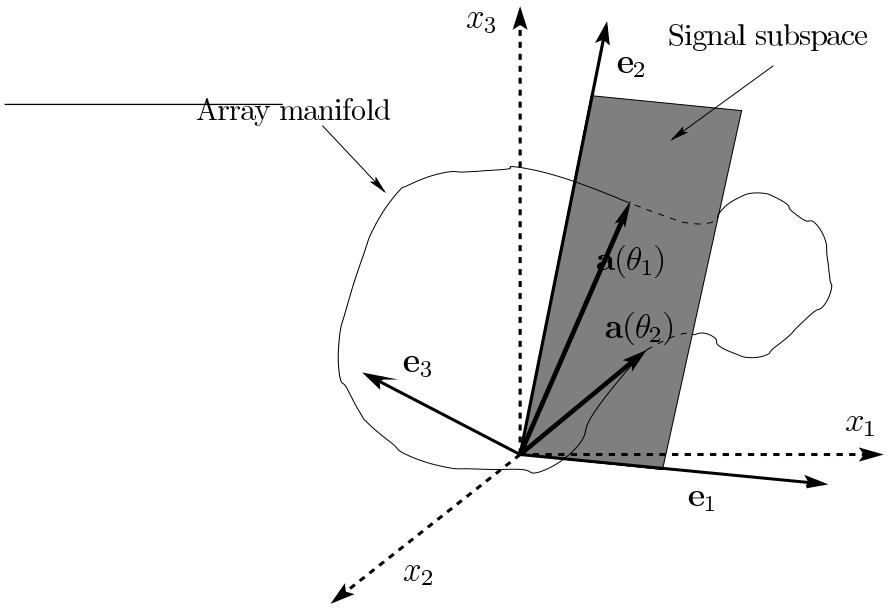

Figure 2.2: Geometric representation of the signal subspace in the case of an array of $M = 3$ sensors receiving $p = 2$ sources from directions. Note here that the signal subspace is spanned by eigenvectors $\mathbf{e}_1$ and $\mathbf{e}_2$.

Under the assumptions of Section 2.2 and considering that the noise is spatially uniform, the data covariance matrix $\mathbf{R}$ can be rewritten as

$$\mathbf{R} = \mathbf{A}\mathbf{R}_\mathbf{s}\mathbf{A}^H + \sigma^2\mathbf{I}. \tag{2.21}$$

From (2.14)-(2.16) and (2.21), it follows that

$$\lambda_1 > \ldots > \lambda_p > \lambda_{p+1} = \ldots = \lambda_M = \sigma^2 \tag{2.22}$$

and

$$\text{span}\{\mathbf{A}\} = \text{span}\{\mathbf{E}_\mathbf{s}\} = \text{span}\{\mathbf{e}_1, \ldots, \mathbf{e}_p\}. \tag{2.23}$$

Equation (2.22), indicates that the number of sources p can be determined from the multiplicity of the smallest eigenvalue of $\mathbf{R}$.

In practice, due to the finite number of samples L, we use $\hat{\mathbf{R}}$, the sample covariance matrix of the data $\mathbf{x}(t)$, $t = 1, \ldots, L$, which is defined by

$$\hat{\mathbf{R}} := \frac{1}{L}\sum_{t=1}^{L}\mathbf{x}(t)\mathbf{x}^H(t). \tag{2.24}$$

and is a consistent estimator for $\mathbf{R}$.

The $M-p$ smallest eigenvalues of $\hat{\mathbf{R}}$ which are denoted $\hat{\lambda}_m$, $m = p+1 \ldots, M$, are with probability one, not equal. However, it is shown that these $M-p$ eigenvalues concentrate around σ^2 with deviation $O(1/\sqrt{L})$ [114].

The covariance matrix $\mathbf{R}$ is parameterized by the vector of unknowns, $\boldsymbol{\eta}_p$, defined as

$$\boldsymbol{\eta}_p := \left[\lambda_1, \ldots, \lambda_p, \sigma^2, \mathbf{e}_1^T, \ldots, \mathbf{e}_p^T\right]^T. \tag{2.25}$$

The problem of model selection can be formulated as follows: Given the parameter space $\boldsymbol{H}$, consider a partition of $\boldsymbol{H}$ corresponding to mutually disjoint p-order model subsets. The problem is to find a criterion to decide, based only on the observations $\mathbf{x}(t)$, $t = 1, \ldots, L$, to which subset $\boldsymbol{H}_p$ the parameter $\boldsymbol{\eta}_p$ belongs [33, 114].

With the data samples rewritten in matrix form as

$$\mathbf{X} = [\mathbf{x}(1), \ldots, \mathbf{x}(L)] \tag{2.26}$$

the ML estimate of the parameter $\boldsymbol{\eta}_p$ is defined as

$$\hat{\boldsymbol{\eta}}_p := \arg \max_{\boldsymbol{\eta}_p \in \boldsymbol{H}_p} \; f(\mathbf{X}|\boldsymbol{\eta}_p) \tag{2.27}$$

where

$$f(\mathbf{X}|\boldsymbol{\eta}_p) = \prod_{t=1}^{L} f(\mathbf{x}(t)|\boldsymbol{\eta}_p) \tag{2.28}$$

is the probability density function (pdf) of the data samples $\mathbf{X}$, given the parameters $\boldsymbol{\eta}_p$.

Source enumeration rests on model selection techniques [99, 120]. These detection methods range from hypothesis testing [17, 27, 112, 113] to information theoretic criteria [1, 33, 80, 106, 116].

2.4.1 Hypothesis Tests

Eigenvalue-Equality and the Sphericity Test

In [7] and [54], the problem of source detection is tackled using a sequential hypothesis test (SHT). For this sequence of nested hypotheses, test statistics must be

adequately defined. The usual assumption of Gaussian data leads to variants of a statistic $\mathcal{T}_q$, based on the ratio of the geometric mean, $\mathcal{G}_q(\hat{\lambda})$, to the arithmetic mean, $\mathcal{M}_q(\hat{\lambda})$, of the $M-q$ smallest sample eigenvalues (corresponding to the candidate noise subspace), i.e.,

$$\mathcal{T}_q := \frac{\mathcal{G}_q(\hat{\lambda})}{\mathcal{M}_q(\hat{\lambda})} = \frac{\left(\prod_{m=q+1}^{M} \hat{\lambda}_m\right)^{1/(M-q)}}{\frac{1}{M-q}\sum_{m=q+1}^{M} \hat{\lambda}_m}, \quad q = 1, \ldots, M-1. \tag{2.29}$$

The test's threshold can be set according to the Neyman-Pearson criterion [99] given a significance level α. This procedure requires knowledge of the distribution of the test statistics under the global null. The exact form of the distribution of $\mathcal{T}_q$ is not available in a useful form since it is generally written as an infinite expansion in terms of basic distributions [66]. It is shown in [16, 112], from the general theory of likelihood ratio (LR) tests, that it is possible to employ asymptotic approximations leading to a χ^2 distribution with M^2-1 degrees of freedom. This however, is only valid for a large L and if all of the eigenvalues of $\mathbf{R}$ are tested for equality simultaneously. As the number of eigenvalues is reduced gradually through the sequences of the SHT, the distribution of the statistics $\mathcal{T}_q$ complies with the approximations only at a very high SNR [112]. Further investigation led to Liggett's variant of the test statistic [61], defined as

$$T_q := -2\left((L-1) - q - \frac{2(M-q)^2+1}{6(M-q)}\right)\ln(\mathcal{T}_q), \qquad q = 1, \ldots M-1 \tag{2.30}$$

For a large L, the statistic T_q is asymptotically χ^2-distributed with $(M-q)^2-1$ degrees of freedom. Other variants of test statistics based on $\mathcal{T}_q$ have been investigated to improve the performance of the tests, especially by taking into account the influence of the q larger eigenvalues (corresponding to the candidate signal subspace) on the distribution of the smaller eigenvalues being tested for equality [112, 117].

Predicted Eigen-Threshold

Recalling that the sample eigenvalues, $\hat{\lambda}_1, \ldots, \hat{\lambda}_M$, are all different with probability one, contrary to the true eigenvalues of $\mathbf{R}$, Chen *et al.* [27] proposed a modified hypothesis test. The method addresses the behavior of the sample eigenvalues, which can be described by the asymptotic joint distribution of the normalized eigenvalues,

as well as by their marginal distributions [27]. These marginal distributions are shown to be concentrated over certain regions, suggesting the definition of an upper and a lower threshold. The latter is defined as the *predicted eigen-threshold* and is distinct from previously used thresholds in that it is not fixed but is a function of the eigenvalue under consideration.

Bootstrap-based Detection

The sphericity test and the predicted eigen-threshold test rely on the assumption of Gaussian data and large samples. If Gaussianity cannot be justified or if only a small sample is available, the pdf of $\mathcal{T}_q$ in (2.29) cannot be determined exactly. The structure of $\mathcal{T}_q$ is well motivated and relevant in non-Gaussian scenarios. However, the performance of the SHT can be erratic [17]. To tackle the problem of an unknown data distribution, alternative test statistics made of all possible pairwise differences among the sample eigenvalues were introduced in [17] as follows:

$$T_{(i,j)_q} = \hat{\lambda}_i - \hat{\lambda}_j, \qquad i = q+1, \ldots, M-1 \qquad , j = i+1, \ldots, M. \tag{2.31}$$

Testing simultaneously all pairs results in a MHT which relies on Holm's sequential rejective Bonferroni (SRB) procedure [45].

Since the distribution of the test statistics is not always analytically tractable, the bootstrap [31, 121] is used successfully to approximate it [17].

A possible bias in the multiple sample eigenvalues can alter the assumption that the noise eigenvalues have equal means. To correct the bias in the estimated eigenvalues, a number of methods were introduced in [17]. Of special interest are the distribution-free techniques based on the bootstrap or the Jackknife [31, 120].

2.4.2 Information Theoretic Criteria

AIC and MDL

The general idea of information theoretic criteria is based on two major principles. First, following the LR theory, the criterion seeks to minimize the distance (or a measure of resemblance) between two probability functions, i.e., between two models. A representative criterion of this principle is Akaike's Information Criterion (AIC) [1]. Second, the criterion minimizes the length of coding of a data sample

generated by a candidate model, and at the same time, introduces a penalty for possibly using a more complex model [59]. The Minimum Description Length (MDL) [80] is a well known criterion of this type.

Formulation of information criteria can be summarized as follows: Given a family of pdfs $f(\mathbf{X}|\boldsymbol{\eta}_q)$, with $\boldsymbol{\eta}_q \in \boldsymbol{H}_q$, for $0 \leq q < M$, we want to determine the value $\hat{q}$ satisfying

$$\hat{q} = \arg\min_{q} \left\{ -\mathcal{L}(\hat{\boldsymbol{\eta}}_q) + \mathcal{P}(q) \right\} \tag{2.32}$$

where $\mathcal{L}(\boldsymbol{\eta}_q) = \ln(f(\mathbf{X}|\boldsymbol{\eta}_q))$, is the log-likelihood (LL) function of the collected data, and $\mathcal{P}(q)$ is a penalty function associated with the q-th family of models. $\hat{\boldsymbol{\eta}}_q$ is the ML estimate of the unknown parameters given the q-th family of distributions, i.e.,

$$\hat{\boldsymbol{\eta}}_q = \arg\max_{\boldsymbol{\eta}_q \in \boldsymbol{H}_q} f(\mathbf{X}|\boldsymbol{\eta}_q). \tag{2.33}$$

Denoting the number of free parameters in $\boldsymbol{H}_q$ by $\nu(\boldsymbol{H}_q)$, AIC and the MDL criterion are defined as follows:

$$\mathrm{AIC}(q) := \arg\min_{0 \leq q < M} \left(-2\mathcal{L}(\hat{\boldsymbol{\eta}}_q) + 2\nu(\boldsymbol{H}_q) \right) \tag{2.34}$$

and

$$\mathrm{MDL}(q) := \arg\min_{0 \leq q < M} \left(-\mathcal{L}(\hat{\boldsymbol{\eta}}_q) + \frac{\nu(\boldsymbol{H}_q)}{2} \ln(L) \right). \tag{2.35}$$

The second term of the right-hand side in (2.34) and (2.35) is the penalty function $\mathcal{P}(q)$. AIC and the MDL criterion essentially differ by their penalty functions. The penalty function $\mathcal{P}(q)$ is monotonically increasing with $\nu(\boldsymbol{H}_q)$, and aims at penalizing the use of increasingly complex models. It is different from one criterion to another.

In the particular case of Gaussian sources and noise, the above criteria become

$$\mathrm{AIC}(q) = \arg\min_{0 \leq q < M} \left(-2L(M-q) \ln(\mathcal{T}_q) + 2q(2M-q) \right) \tag{2.36}$$

and

$$\mathrm{MDL}(q) = \arg\min_{0 \leq q < M} \left(-L(M-q) \ln(\mathcal{T}_q) + \frac{q(2M-q)}{2} \ln(L) \right) \tag{2.37}$$

where $\mathcal{T}_q$ is the statistic defined in (2.29).

The asymptotic performance of the detection criteria is established based on the functional form of $\mathcal{P}(q)$ [33, 106]. More specifically, it is shown that AIC tends to asymptotically overestimate the number of sources, whereas the MDL criterion is strongly consistent[2] for iid data. For small samples, MDL often results in an underestimated model. In a sense, the information criteria can be viewed as eigenvalue-equality tests with an adaptive significance level.

Efficient Detection Criteria (EDC)

Based on the general theory of LR, when further analyzing the asymptotic performance of AIC and MDL, Zhao *et al.* [115] outlined a general class of strongly consistent information theoretic criteria. The proposed class is called Efficient Detection Criteria (EDC). The particularity of EDC arises with the general condition on the penalty function for a criterion to be consistent. Indeed, it is shown that a penalty function of the form

$$\mathcal{P}(q) = \nu(\boldsymbol{H}_q) \cdot \mathcal{C}(L) \tag{2.38}$$

ensures that the information criterion is strongly consistent, provided that the function $\mathcal{C}(L)$ satisfies the following:

$$\lim_{L\to\infty} \frac{\mathcal{C}(L)}{L} = 0 \qquad \text{and} \qquad \lim_{L\to\infty} \frac{\mathcal{C}(L)}{\ln(\ln(L))} = \infty. \tag{2.39}$$

When examining (2.35) and (2.38), it is clear that MDL is a particular case of EDC.

Order Statistics

In the previously described techniques, the expression for the ML estimates of the unknown parameters $\boldsymbol{\eta}_p$ is based on the assumption that the order of the sample eigenvalues and the actual eigenvalues is the same. While asymptotically this assumption is true, it is not verified for a finite L. Considering this limitation, the pdf

[2]The problem of source detection in array signal processing relies on the assumption that the generating model is of a finite order (specular multipath for example). Hence, an information theoretic criterion is said to be *consistent* if , for a large L, the estimated number of sources tends asymptotically to the true model order with probability one.

of the ordered sample eigenvalues is incorporated in the derivation of the goodness-of-fit part of the information theoretic criteria [33], leading to the Order Statistics ML Estimator (OSMLE) of the unknown parameters, and thus to the Order Statistics AIC (OSAIC) and Order Statistics MDL (OSMDL). These detectors provide an increased detection accuracy and retain the same consistency properties as the conventional AIC and MDL, but involve an increased computational load.

Gerschgorin Radii

In addition to the ordered eigenvalues of $\mathbf{R}$, the Gerschgorin-AIC (GAIC) and Gerschgorin-MDL (GMDL) [113] employ the ordered Gerschgorin radii [111] of a unitarily transformed version of $\mathbf{R}$. This additional information provides increased detection accuracy. While GMDL benefits from a strengthened consistency, GAIC is, unlike most AIC-based detectors, a consistent estimator of the number of sources. We show this fact analytically in Section 5.2.2, as we extend the principle of subspace separation using Gerschgorin radii further to partially nonuniform noise.

2.5 Overview of DOA Estimation in Uniform Noise

Estimation of the DOAs $\boldsymbol{\theta}$, from the data samples $\mathbf{x}(t)$, $t = 1, \ldots, L$, requires that the number of sources, p, is known *a priori*. Two major classes of estimation techniques are known in the literature. These are *parametric* methods, based on ML estimation, and *spectral-based* methods, where the DOAs constitute the highest peaks of a defined spectrum-like function of the unknown parameters [51]. These spectral methods are mostly based on an efficient separation between the signal and noise subspaces.

2.5.1 Deterministic Maximum Likelihood (DML)

The signal waveforms are considered to be deterministic and unknown, so that the measurements $\mathbf{x}(t)$ are modeled as iid Gaussian processes with mean $\mathbf{A}(\boldsymbol{\theta})\mathbf{s}(t)$ and covariance $\sigma^2\mathbf{I}$.

The negative LL function of the data is then given by [22, 81]

$$\mathcal{L}(\boldsymbol{\theta}, \mathbf{s}(t), \sigma^2) = M \log\left(\sigma^2\right) + \frac{1}{\sigma^2 L} \sum_{t=1}^{L} \|\mathbf{x}(t) - \mathbf{A}(\boldsymbol{\theta})\mathbf{s}(t)\|_2^2. \tag{2.40}$$

where $\|\cdot\|_N$ denotes the N-th norm operator. For simplicity, in the sequel we use only $N = 2$ and denote $\|\cdot\|_2$ by $\|\cdot\|$.

Since the parameters of interest are the DOAs $\boldsymbol{\theta}$, concentrating the LL function with respect to the other unknowns is performed by obtaining closed-form expressions of the DML estimates of σ^2 and $\mathbf{s}(t)$ [13, 105].

The concentrated cost function whose minima are the DML estimates of the DOAs is given by

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left(\text{trace}\left\{\boldsymbol{\Pi}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})\hat{\mathbf{R}}\right\}\right) \tag{2.41}$$

where $\boldsymbol{\Pi}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta}) = \mathbf{I} - \boldsymbol{\Pi}_{\mathbf{A}}(\boldsymbol{\theta})$ is the orthogonal complement of the projector onto the nullspace of $\mathbf{A}^H(\boldsymbol{\theta})$, denoted $\boldsymbol{\Pi}_{\mathbf{A}}(\boldsymbol{\theta}) = \mathbf{A}(\boldsymbol{\theta})\mathbf{A}^{\#}(\boldsymbol{\theta})$, with the Moor-Penrose pseudo-inverse of matrix $\mathbf{A}$ given by $\mathbf{A}^{\#} = \left(\mathbf{A}^H\mathbf{A}\right)^{-1}\mathbf{A}^H$.

The DML estimates require a numerical solution to the non-linear p-dimensional optimization problem of (2.41). With a good initialization, a Gauss-Newton technique (see [24, 102]) usually converges rapidly to the minimum of (2.41) [51]. Obtaining sufficiently accurate initial estimates, however, can be a computationally expensive task. Subspace separation methods (see Section 2.5.3) are a natural choice for an initial estimator, provided that all the sources can be resolved. Other possibilities are analyzed and commented on in [51, 71, 119].

It is worth mentioning that the DML estimates of $\boldsymbol{\theta}$ attain the corresponding deterministic, or conditional, Cramér-Rao bound (CRB), only when both M and L tend to infinity [91]. This results from the fact that in the deterministic model the number of signal waveform parameters $\mathbf{s}(t)$ grows without bound as the number of samples increases, implying that they cannot be consistently estimated [51].

2.5.2 Stochastic Maximum Likelihood (SML)

The sources are modeled as Gaussian random processes with zero mean and covariance $\mathbf{R}_{\mathbf{s}}$ according to the assumptions of Section 2.2. The observation vector $\mathbf{x}(t)$ is therefore a white zero-mean Gaussian random vector with a covariance matrix given

by (2.21). In this case, the likelihood function depends on $\boldsymbol{\theta}$, $\mathbf{R_s}$ and the power of the uniform noise, σ^2. Ignoring constant terms, the negative LL function can be reduced to

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}, \mathbf{R_s}, \sigma^2) &= \frac{1}{L}\sum_{t=1}^{L} \|\boldsymbol{\Pi}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})\mathbf{x}(t)\|^2 \\ &= \text{trace}\left\{\boldsymbol{\Pi}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})\hat{\mathbf{R}}\right\}. \end{aligned} \tag{2.42}$$

The above criterion is highly non-linear. However, it allows an explicit separation of the parameters. Hence, for fixed DOAs $\boldsymbol{\theta}$, concentrating the LL function (2.42) with respect to σ^2 and $\mathbf{R_s}$ leads to the following solution:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left[\log\left(\det\left(\mathbf{A}(\boldsymbol{\theta})\hat{\mathbf{R}}_{\mathbf{s}}\mathbf{A}^H(\boldsymbol{\theta}) + \hat{\sigma}^2\mathbf{I}\right)\right)\right]. \tag{2.43}$$

where the concentrated ML estimates are $\hat{\sigma}^2 = (M-p)^{-1}\text{trace}\left\{\boldsymbol{\Pi}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})\right\}$ and $\hat{\mathbf{R}}_{\mathbf{s}} = \mathbf{A}^{\#}(\boldsymbol{\theta})\left(\hat{\mathbf{R}} - \hat{\sigma}^2\mathbf{I}\right)\mathbf{A}^{\#H}(\boldsymbol{\theta})$.

The criterion of (2.43) is a non-linear function of its argument $\boldsymbol{\theta}$. A Newton-type technique can be employed for the numerical search [23, 71, 88, 101], as well as Alternating Projections (AP) [118] or Expected Maximization (EM) [32, 65] techniques. Upon achieving the global minimum, the SML method produces excellent statistical properties (consistency and efficiency). The DOA estimates obtained by SML are shown to have a better accuracy than the DML estimates [72, 92]. This difference, however, is only important for the non-asymptotic case. For Gaussian signals, the SML estimates attain the unconditional CRB. This follows from the general theory of ML estimation (see e.g. [99]), since all the unknowns in the stochastic model are estimated consistently.

2.5.3 Multiple Signal Classification (MUSIC)

In practice, the sample covariance matrix $\hat{\mathbf{R}}$ defined in (2.24) is used as a consistent estimate of $\mathbf{R}$, and its eigenvectors are separated into signal and noise eigenvector sets, $\hat{\mathbf{E}}_{\mathbf{s}}$ and $\hat{\mathbf{E}}_{\mathbf{n}}$, respectively, from (2.23). Ideally, since the noise eigenvectors in $\mathbf{E_n}$ are orthogonal to $\mathbf{A}(\boldsymbol{\theta})$, it is straightforward to see that

$$\mathbf{E}_{\mathbf{n}}^H\mathbf{a}(\theta) = \mathbf{0}, \qquad \text{for } \theta \in \{\theta_1, \ldots, \theta_p\}. \tag{2.44}$$

Also, since $\mathbf{A}(\boldsymbol{\theta})\mathbf{R_s}\mathbf{A}^H(\boldsymbol{\theta})$ is of rank p, it follows that θ_q, $q = 1, \ldots, p$, are the only possible solutions to (2.44), i.e., to uniquely resolve the DOA estimates, the array is usually assumed to be unambiguous.

The MUSIC spatial spectrum is defined as [83]

$$\mathcal{S}(\theta) := \frac{1}{\mathbf{a}^H(\theta)\hat{\mathbf{E}}_\mathbf{n}\hat{\mathbf{E}}_\mathbf{n}^H\mathbf{a}(\theta)}. \tag{2.45}$$

From (2.44), it is clear that the functional $\mathcal{S}(\theta)$ has peaks in the vicinity of the true DOAs. If the data sample and the SNR are sufficiently large, and the signal model is sufficiently accurate, the MUSIC algorithm can provide statistically consistent estimates [91].

For independent sources, MUSIC is a large sample realization of DML [91, 92]. However, the important limitation of the MUSIC functional of (2.45) is its failure to resolve closely spaced signals in small sample and low SNR scenarios. This loss of resolution is more pronounced for highly correlated signals.

In addition to its asymptotic equivalence to DML, MUSIC is related to other popular spatial-spectrum methods, over which it has resolution advantages, such as the Minimum Variance Distortionless (MVD) beamformer [6], and the Minimum Norm (MN) algorithm [68]. It was demonstrated in [68] that MUSIC is equivalent to the limit of the MVD beamformer when the SNR tends to infinity. Also, following the development of [22, 68], MUSIC is interpreted as a smoothed version of MN. Moreover, MUSIC is related to the subspace-separation approximations of ML methods. Indeed, when $\|\mathbf{a}(\theta)\|$ is independent of θ, MUSIC and the Noise Subspace Fitting (NSF) algorithm coincide if the weighting matrix of NSF reduces to the identity matrix, $\mathbf{I}$ [51]. By extension through NSF, it is also related to the Signal Subspace Fitting (SSF) and Weighted Signal Subspace Fitting (WSSF) techniques [100].

In the particular case of a ULA configuration, the polynomial-rooting version of MUSIC (Root-MUSIC) allows a closed-form solution for the estimated DOAs [6].

Given the above link to a variety of other methods and being a computationally-motivated sub-optimal high-resolution alternative to ML, MUSIC is a benchmark method that is central to our analysis.

Chapter 3

Effect of Spatially Nonuniform Noise

3.1 Introduction

In order to estimate the number of sources p from the model in (2.6), conventional detection methods rely on the ordered eigenvalues of $\mathbf{R}$ and essentially count the multiplicity of the smallest eigenvalue which ideally is equal to σ^2. In practice, we have a finite sample size L and thus the sample covariance matrix $\hat{\mathbf{R}}$ is used instead of $\mathbf{R}$.

As explained in Chapter 2, contrary to the eigenvalues of $\mathbf{R}$, with probability one, the smallest $M - p$ sample eigenvalues are not equal [33]. In addition, when the SNR is low or when the noise power is not ideally uniform over the sensors due to perturbations, it can happen that the magnitude-order of the sample eigenvalues does not coincide with the order of the true eigenvalues [33]. It results that the strict criterion of *eigenvalue-equality* described by the statistic of (2.29), on which most conventional detectors are based, may be broken, and the sample eigenvectors may not be separated into signal and noise eigenvectors in an optimal way.

To picture this mismodeling problem, consider for example a ULA with $M = 6$ elements receiving $p = 2$ sources from DOAs $\boldsymbol{\theta} = [0°, 13°]^T$. Assume that the two first sensors out of M are perturbed such that their individual noise-powers are respectively 5 and 4 times larger than the remaining noise-powers, which are equal and represent the reference of the original SNR. If this SNR is 0 dB, then

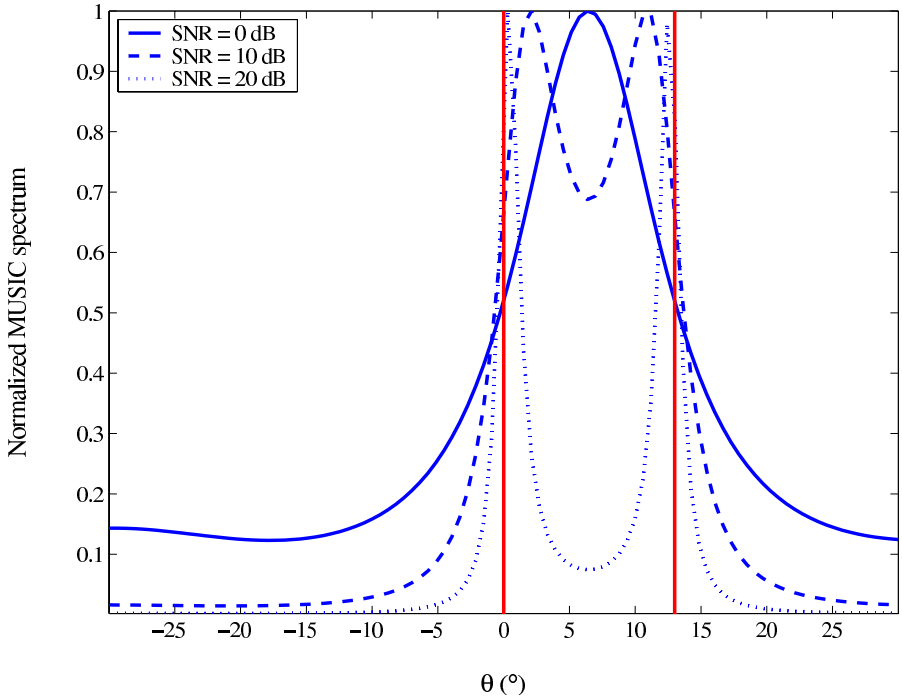

Figure 3.1: Mismodeling effect due to nonuniform noise on the MUSIC spectrum.

the ordered eigenvalues of the resulting data covariance matrix are $\lambda_1 = 14.41$, $\lambda_2 = 9.72$, $\lambda_3 = 7.10$, $\lambda_4 = 2.77$, $\lambda_5 = 2.00$ and $\lambda_6 = 2.00$, whereas the projection of their corresponding eigenvectors onto the columns of the array steering matrix $\mathbf{A}$, results in values 1.79, 1.08, 1.17, 0.50, 0.00 and 0.00, respectively. It is therefore clear that not only the true number of sources is difficult to deduce from the equality of the smallest eigenvalues, but also, if the true number of sources were known *a priori*, the second largest eigenvalue would not lead to the closest signal-subspace eigenvector.

Figure 3.1 illustrates the same mismodeling effect on DOA estimation using the classical MUSIC algorithm, where the same settings as above were employed. Note that even when the SNR is increased to 10 dB, the DOAs are not properly recovered.

In the following, we examine the causes and effect of noise-power perturbations on the problem of source detection and parameter estimation. The analysis essentially addresses the feasibility of subspace separation, and more practically, assesses the quality of an approximate separation. Based on the results of this chapter, we will propose a set of algorithms to better tackle the problem of non-uniform noise powers.

3.2 Limits of Eigen-based Subspace Separation in Nonuniform Noise

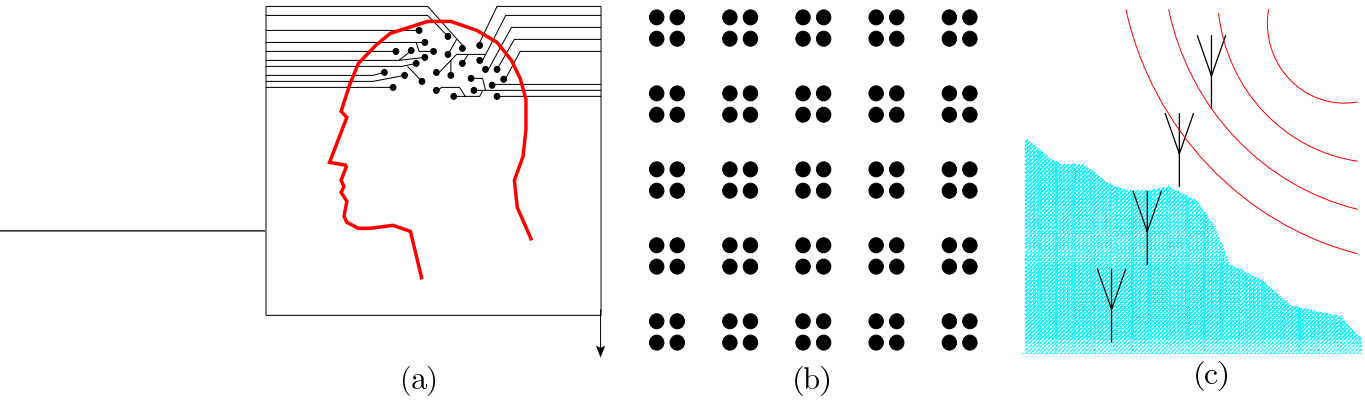

Figure 3.2: Example of applications where spatially nonuniform noise is observed: (a) magneto-encephalography, (b) sparse arrays, and (c) SONAR systems in inhomogeneous media.

Noise-power perturbation, or spatial *nonuniformity* of the noise can also be thought of when, for example, the inter-element spacing is increased to the limit of element-coupling, yet making each array element undergo perturbations separately. A typical practical case where spatial nonuniformity is observed includes magneto-encephalography (MEG), where an array of sensors is placed over the patient's head to record the brain's activity [47]. As the system contains a very large number of different sensors whose distances from the patient's head are slightly different, the noise varies considerably from one sensor to another, and no *a priori* knowledge is available on the array response for a reliable calibration [34]. Another application concerns sparse arrays with prevailing external noise and hardware non-idealities in receiving channels [68, 97, 76]. Other examples of nonuniform noise include RADAR applications with shortcomings in the array calibration, or SONAR systems, where the inhomogeneous propagation medium generates hydroacoustic reverberations of different intensities at the hydrophones, provided that the separation between the latter is large enough to ensure decorrelation of the noise from one sensor to another [38]. In Section 6.5, we also describe a practical scenario where noise uniformity cannot be guaranteed. Indeed, in a process monitoring system (nuclear power-plant),

the number of sensors is very high and their pattern around the process of interest is either random or too complex to be tractable. Furthermore, the observation time is very long (typically several months) which makes a continuous and reliable calibration of the sensors very costly and impractical, and makes the whole measurement process vulnerable to unpredictable small external perturbations at some sensors [4].

With the requirement that the noise is spatially independent, nonuniform noise is modeled by its covariance matrix $\mathbf{Q}$ as defined by (2.14), i.e.,

$$\mathbf{Q} = \operatorname{diag}\left\{\sigma_1^2, \ldots, \sigma_M^2\right\}$$

with different noise powers, σ_m^2, $m = 1, \ldots, M$.

With the above noise model, performance of the standard detection and estimation methods of Sections 2.4 and 2.5 can degrade dramatically due to the induced mismodeling of the data. In addition, because the distribution of the noise powers $\sigma_1^2, \ldots, \sigma_M^2$, over the array sensors is unknown, direct estimation of the noise powers and data prewhitening is problematic.

Denoting the eigenvalues of $\mathbf{AR_sA}^H$ by $\mathbf{\Lambda}_s := \operatorname{diag}\{\lambda_{1_s}, \ldots, \lambda_{p_s}, 0, \ldots, 0\}$, it is clear that contrary to the ideal uniform noise case where the noise covariance $\mathbf{Q}$ is equal to an identity matrix $\mathbf{I}$ (up to a positive scalar σ^2), in the nonuniform noise case, matrices $\mathbf{AR_sA}^H$ and $\mathbf{Q}$ are not commutative under multiplication. Therefore the equality $\lambda_m = \lambda_{m_s} + \sigma_m^2$, for $m = 1, \ldots, M$, does not hold, and it is not possible to directly count the multiplicity of the smallest eigenvalues of $\mathbf{R}$ to determine the number of sources p.

3.2.1 Effect of Poor Clustering of the Noise Eigenvalues on Classical Detection Criteria

Direct application of the classical detectors of Section 2.4 to nonuniform noise scenarios results in erroneous results. The relation between the quality of the estimated eigenvalues and the performances of AIC and the MDL criterion for instance, were analyzed in [60].

If the smallest eigenvalues are not clustered sufficiently closely, then the above criteria will very likely bypass a resulting gap between the candidate signal and noise

eigenvalues, thus leading to overmodeling. Moreover, for a fixed number of sensors M, this overmodeling becomes more probable for an increasing number of samples L. Following the arguments of [60], it can be deduced that in nonuniform noise the estimated noise eigenvalues will not be clustered sufficiently closely if the SNR is not sufficiently high, and the above criteria will result in severe overmodeling. Note that this property is also at the root of the breakdown of AIC and MDL in the presence of colored noise [60].

If the candidate signal and noise eigenvalues are not well separated in magnitude, and if the candidate noise eigenvalues are sufficiently close to each other, then undermodeling will very likely occur. Such a case often happens when the SNR is low. Note however that in nonuniform noise, the smallest eigenvalues are far from being equal if the SNR is not high enough.

3.2.2 Identifiability of the Dimension of the Signal Subspace

Another obstacle to correct source enumeration using ordered eigenvalues, is the ambiguity related to noise nonuniformity. This ambiguity is inherent to the data model (2.16).

For nonuniform noise, the data covariance matrix $\mathbf{R}$ is parameterized by the vector of unknowns $\boldsymbol{\eta}_p = \left[p, \mathbf{p}^T, \mathbf{q}^T\right]^T$, where $\mathbf{p} = \left[\mathrm{vec}^T\left(\Re\left(\mathbf{R_s}\right)\right), \mathrm{vec}^T\left(\Im\left(\mathbf{R_s}\right)\right)\right]^T$, with $\mathrm{vec}(\cdot)$ being the operator that stacks matrix columns into a single vector, and $\mathbf{q}$ is defined by (2.15). Note that it is always possible to find $p_1 \neq p_2$, and thus $\boldsymbol{\eta}_{p_1} \neq \boldsymbol{\eta}_{p_2}$, such that $\mathbf{R}(\boldsymbol{\eta}_{p_1}) = \mathbf{R}(\boldsymbol{\eta}_{p_2})$. This ambiguity can be easily illustrated by realizing that the uniform noise is a special case where all noise powers are equal, making it possible to rearrange the hypothetical number of sources and their respective powers. More specifically, it is shown in [33] that the problem of source detection in nonuniform noise becomes non-identifiable in the event where at least one of the p sources is received by only one of the M sensors. Such a source cannot be distinguished from the noise power over that same sensor.

In the sequel, we exclude cases falling in the above non-identifiable scenario. In practice, such a scenario is unlikely for arrays of coupled elements, and is rare in applications of non-destructive process control (see Section 6.5).

3.3 Subspace Estimation Quality

Ideally, the eigenvalues of $\mathbf{R}$ are used to indicate which subspace their corresponding eigenvectors are likely to span. When the noise covariance matrix is an identity (up to a scalar), the separation between the signal and noise subspaces is exact, as a result of the orthogonality between the respective eigenvectors. In practice, this orthogonality is lost and the separation between the subspaces is only approximate, leading to an *ambiguous* intersection between the signal and noise subspaces. Clearly, if the signals and noise are assumed spatially and temporally white, the quality of the estimated subspaces degrades with a decreasing sample size, a low SNR, or variable perturbations of the noise power over the sensors.

In [104], the quality of subspace estimation is measured by the angle between the estimated and the ideal signal subspaces, for a given dimension ($p = 2$). Specifically, the asymptotic expression of the Mean Square Error (MSE) of the cosine of the angle of interest is provided. Based on the development in [104], we derive in what follows the asymptotic distribution of the cosine of the angle between the estimated signal eigenvectors and the ideal signal subspace on one hand, and on the other hand, the distribution of the cosine of the angle between the estimated noise eigenvectors and the same signal subspace reference.

Consider the ideal uniform noise case. Let $\mathbf{E} = [\mathbf{e}_1, \ldots, \mathbf{e}_M]$ be the eigenvectors corresponding to the ordered eigenvalues $\lambda_1 \geq \ldots \geq \lambda_M$ of $\mathbf{R}$ as defined by (2.19) and (2.18), respectively. With the true number of sources p, we ideally have $\lambda_{p+1} = \ldots = \lambda_M = \sigma^2$. Also, consider $\hat{\mathbf{E}} = [\hat{\mathbf{e}}_1, \ldots, \hat{\mathbf{e}}_M]$, the eigenvectors of the sample covariance matrix $\hat{\mathbf{R}}$ corresponding to eigenvalues $\hat{\lambda}_1, \ldots, \hat{\lambda}_M$. Note that the order of the latter sample eigenvalues is not guaranteed to match the order of the true eigenvalues [33]. Define $\phi_m(p)$ as the angle between the ideal and the sample eigenvectors $\mathbf{e}_m$ and $\hat{\mathbf{e}}_m$, for $m = 1, \ldots, M$, given the hypothesis that there are p sources (H_p).

Lemma 3.3.1. *Given the assumptions of Section 2.2 and* H_p*, with Gaussian source signals and noise, for* $L \to \infty$ *and* $M \gg p$*, we have*

1. $\underline{m = 1, \ldots, p:}$

   For $m = 1, \ldots, p$*, the angles* $\phi_m(p)$*, corresponding to the candidate signal*

subspace, are asymptotically independent and satisfy the following:

$$\cos(\phi_m(p)) \xrightarrow{\mathcal{D}} X_{s_m} \tag{3.1}$$

where $\xrightarrow{\mathcal{D}}$ *denotes convergence in distribution, and the random variable* X_s *satisfies the following:*

$$X_{s_m} \sim \mathcal{N}\left(\mu_{s_m}(p), \varsigma^2_{s_m}(p)\right) \tag{3.2}$$

with

$$\mu_{s_m}(p) = 1 - \sum_{\substack{q=1 \\ q\neq m}}^{p} \frac{\lambda_m \lambda_q}{L(\lambda_m - \lambda_q)^2} - \frac{M-p}{L} \frac{\lambda_m \sigma^2}{(\lambda_m - \sigma^2)^2} \tag{3.3}$$

$$\varsigma^2_{s_m}(p) = \sum_{\substack{q=1 \\ q\neq m}}^{p} \left(\frac{\lambda_m \lambda_q}{L(\lambda_m - \lambda_q)^2}\right)^2 + \frac{M-p}{L^2}\left(\frac{\lambda_m \sigma^2}{(\lambda_m - \sigma^2)^2}\right)^2. \tag{3.4}$$

2. $\underline{m = p+1, \ldots, M:}$

For $m = p+1, \ldots, M$*, asymptotically, for* $L \to \infty$ *and* $M \gg p$*, the angles* $\phi_m(p)$ *corresponding to the complement of the candidate signal subspace, satisfy marginally the following:*

$$\cos(\phi_m(p)) \xrightarrow{\mathcal{D}} X_n \tag{3.5}$$

where

$$X_n \sim \mathcal{N}\left(\mu_n(p), \varsigma^2_n(p)\right) \tag{3.6}$$

with

$$\mu_n(p) = 1 - \sum_{q=1}^{p} \frac{\lambda_q \sigma^2}{L(\lambda_q - \sigma^2)^2} \tag{3.7}$$

$$\varsigma^2_n(p) = \sum_{q=1}^{p} \left(\frac{\lambda_q \sigma^2}{L(\lambda_q - \sigma^2)^2}\right)^2. \tag{3.8}$$

Proof. See Appendix B.1. □

From Lemma 3.3.1, one can visualize a threshold effect on the accuracy with which the signal subspace is estimated, as the SNR, L and M vary. To better illustrate this effect, consider for the noise-free case the expression of the expected value of $\cos(\phi_m(p))$, for $m = 1, \ldots, p$, denoted $\mu_{s_m}^{(0)}(p)$, i.e.,

$$\mu_{s_m}^{(0)}(p) = 1 - \sum_{\substack{q=1 \\ q \neq m}}^{p} \frac{\lambda_{m_s}\lambda_{q_s}}{L(\lambda_{m_s} - \lambda_{q_s})^2} \tag{3.9}$$

where λ_{m_s} are the eigenvalues of $\mathbf{A}\mathbf{R_s}\mathbf{A}^H$, i.e., pure signal eigenvalues. For a fixed sample size, accuracy of the estimated angle $\phi_m(p)$ depends on the spatial resolution of the array, i.e., two closely spaced sources can lead to very close eigenvalues λ_{m_s} and λ_{q_s} in the denominator of (3.9). For a large L, when the sources are well resolved by the array, $\mu_{s_m}^{(0)}$ tends asymptotically to its ideal value, 1.

In the presence of noise, using (3.3) and (3.9), $\mu_{s_m}(p)$ can be rewritten as

$$\begin{aligned} \mu_{s_m}(p) = \mu_{s_m}^{(0)}(p) \quad &+ \sum_{\substack{q=1 \\ q \neq m}}^{p} \frac{\sigma^2(\lambda_{m_s} + \lambda_{q_s} + \sigma^2)}{L(\lambda_{m_s} - \lambda_{q_s})^2} \\ &+ \frac{M-p}{L} \frac{1 + \lambda_{m_s}/\sigma^2}{(\lambda_{m_s}/\sigma^2)^2}. \end{aligned} \tag{3.10}$$

When L is large and $M \gg p$, the second term on the right-hand side of (3.10) can be neglected. The effect of noise is therefore described by the third term, i.e.,

$$\Delta\mu_{s_m}(p) \approx \frac{M-p}{L} \frac{1 + \lambda_{m_s}/\sigma^2}{(\lambda_{m_s}/\sigma^2)^2}. \tag{3.11}$$

The variation of $\Delta\mu_{s_m}(p)$ follows two patterns. First, when $\sigma^2 \gg \lambda_{m_s}$, (3.11) can be rewritten as

$$\Delta\mu_{s_m}(p) \approx \frac{M-p}{L} \frac{1}{(\lambda_{m_s}/\sigma^2)^2} \tag{3.12}$$

and second, when $\sigma^2 \ll \lambda_{m_s}$, we obtain

$$\Delta\mu_{s_m}(p) \approx \frac{M-p}{L} \frac{1}{\lambda_{m_s}/\sigma^2}. \tag{3.13}$$

The SNR is generally defined with respect to the average noise power. Hence, when it is high, the SNR is proportional to L, whereas for low values, it is proportional to $L^{1/2}$. This illustrates that due to additive noise, under the typical 0 dB

SNR threshold, the sample size L needs to be increased quadratically, rather than linearly, to compensate for the decrease of the SNR.

The same threshold effect can be deduced by examining the expression of $\varsigma^2_{m_s}(p)$ or by evaluating the MSE of $\cos(\phi_m(p))$, taking into account that its ideal value is 1. Using the same approximations as with $\mu_{m_s}(p)$ in (3.10), it is straightforward to see that this MSE reduces to

$$\mathrm{MSE} = \left(\frac{M-p}{L}\,\frac{\lambda_m\sigma^2}{(\lambda_m-\sigma^2)^2}\right)^2. \tag{3.14}$$

When $p=2$ and $m=2$, the above MSE is the same as the result of [104].

Let $\gamma_m(p)$ denote the angle between the true signal subspace and sample eigenvector $\hat{\mathbf{e}}_m$, for $m=p+1,\ldots,M$, corresponding to the candidate noise subspace. Then, using Lemma 3.3.1, we can derive the following result:

Corollary 3.3.2. *Given the assumptions of Lemma 3.3.1 and* H_p*, asymptotically, for* $L\to\infty$ *and* $M\gg p$*, the angles* $\gamma_m(p)$*, for* $m=p+1,\ldots,M$*, satisfy marginally the following:*

$$\cos(\gamma_m(p)) \xrightarrow{\mathcal{D}} X_\gamma \tag{3.15}$$

where

$$X_\gamma \sim \mathcal{N}\left(\mu_\gamma(p), \varsigma^2_\gamma(p)\right) \tag{3.16}$$

with

$$\mu_\gamma(p) = 1-\mu_n(p) \tag{3.17}$$

$$\varsigma^2_\gamma(p) = \varsigma^2_n(p) \tag{3.18}$$

where $\mu_n(p)$ *and* $\varsigma^2_n(p)$ *are defined by (3.7) and (3.8), respectively.*

Proof. See Appendix B.2. □

The above asymptotic results are an indication of the *separability* between the estimated signal and noise subspaces. In the sequel we will use these results to estimate the dimension of the approximately separated signal subspace, for a sufficiently high SNR, i.e., in the case where the SNR is roughly linearly proportional to the sample size L, as described by (3.13). Beyond this scenario, i.e., under (3.12), the separation between the signal and noise subspaces is theoretically impossible and any approximation is problematic.

3.4 Numerical Examples

In Figure 3.3, we illustrate the validity of the approximate asymptotic distribution given by (3.16). An ULA of $M = 12$ sensors in uniform sensor noise is considered, receiving $p = 2$ signals from DOAs given by $\boldsymbol{\theta} = [3°, 23°]^T$. In order to emphasize the quality of the asymptotic approximation, we use a large sample size $L = 1000$ and SNR=15 dB. The figure shows both the histogram of $\cos(\gamma_3(2))$ after 1000 Monte Carlo runs, and the Normal distribution with mean and variance given by (3.17) and (3.18), respectively. In this example, the theoretical expressions of $\mu_\gamma(2)$ and $\sigma_\gamma^2(2)$ as defined in Corollary 3.3.2 are 1.9294×10^{-4} and 1.8518×10^{-8}, respectively, whereas the expressions of $E\{\cos(\gamma_3(2))\}$ and $\text{var}\{\cos(\gamma_3(2))\}$ as obtained through Monte Carlo runs are 2.3942×10^{-4} and 1.8385×10^{-8}, respectively.

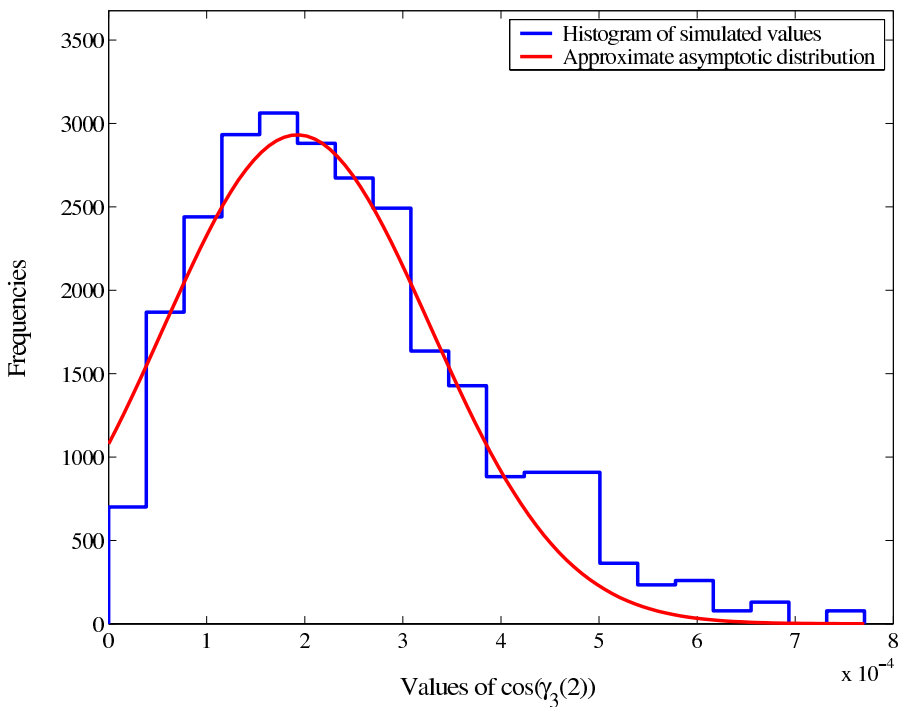

Figure 3.3: Example of a histogram of $\cos(\gamma_3(2))$ and the approximate asymptotic distribution of Corollary 3.3.2.

The examples of Figures 3.4 and 3.5 show the threshold effect due to noise on the quality of the estimated signal subspace. In the simulations, the noise power is perturbed from one array sensor to another. The noise powers are randomly

generated from a uniform distribution over the interval $[\sigma^2_{\min}\sigma^2_{\max}]$. The MSE of $\cos(\phi_2(2))$ as defined by (3.2), is shown varying with the Worst Noise Power Ratio (WNPR) which is defined as [76]

$$\text{WNPR} := \frac{\sigma^2_{\max}}{\sigma^2_{\min}} \tag{3.19}$$

and at the same time with the SNR on one hand, and the sample size L on the other hand. The SNR is evaluated with respect to the average noise power. In both examples, a ULA is assumed with $M = 10$ sensors receiving $p = 2$ sources from DOAs $\boldsymbol{\theta} = [3°, 23°]^T$. In Figure 3.4, L is fixed to 50 samples, whereas in Figure 3.5, the SNR is fixed to 5 dB.

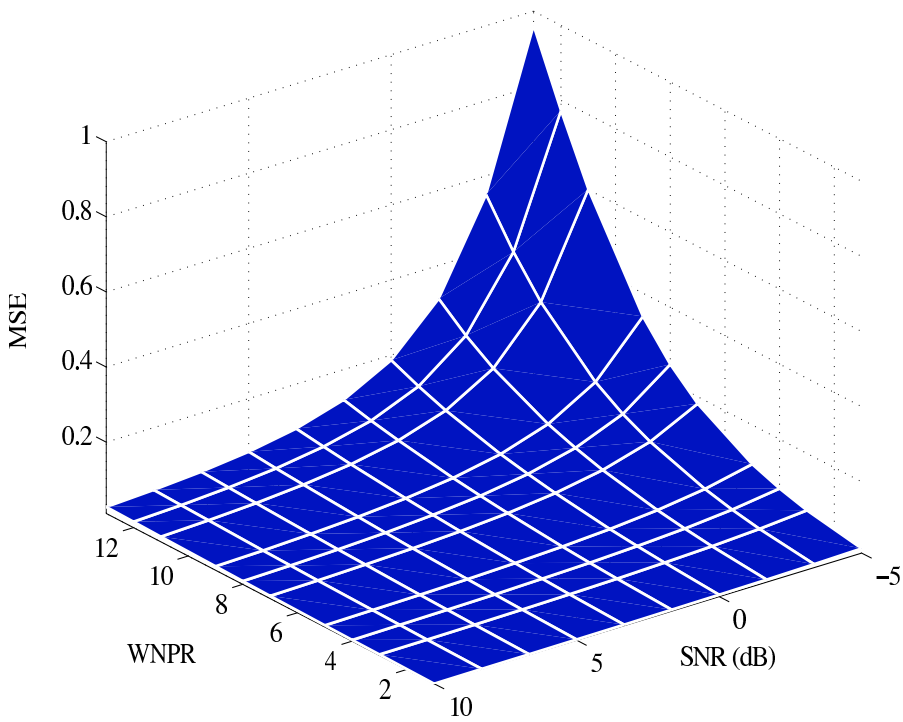

Figure 3.4: Joint effect of SNR and WNPR on the signal subspace estimation quality, measured by MSE of $\cos(\phi_2(2))$.

Both examples show similar results and give an idea on the relation between the SNR and the sample size, as well as the range of noise perturbations over which the quality of the estimated signal subspace remains *exploitable*. In all the examples, WNPR is variable so that the best case corresponding to uniform noise (WNPR=1)

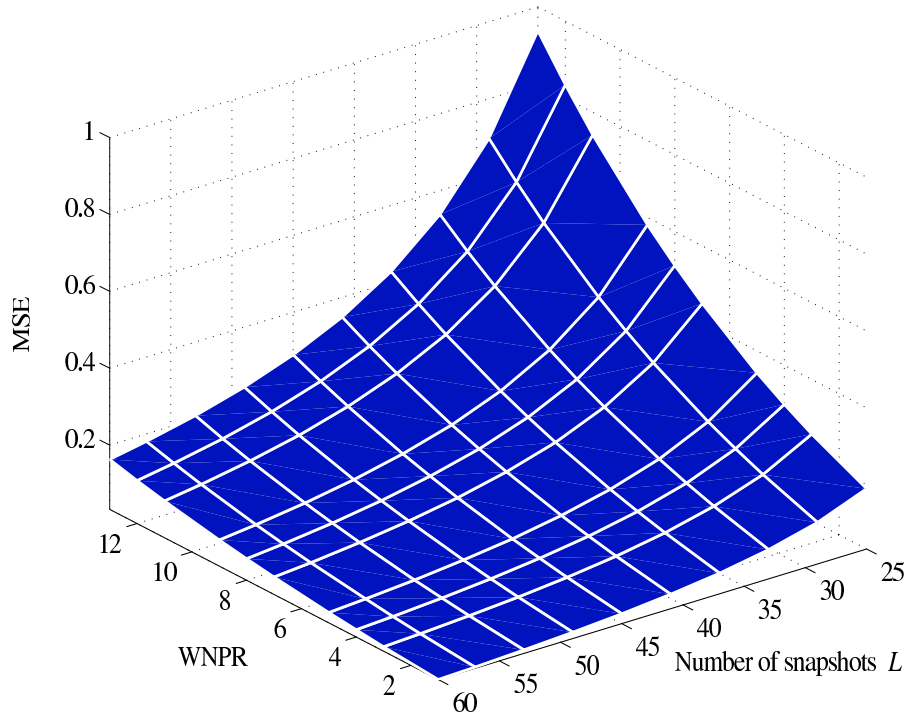

Figure 3.5: Joint effect of L and WNPR on the signal subspace estimation quality, measured by MSE of $\cos(\phi_2(2))$.

is well highlighted. Note that the WNPR does not provide an indication on the distribution of the noise powers.

Observe that the variation of the different parameters and the WNPR have a threshold effect of the MSE, indicating that beyond certain values of these parameters, the degradation in the quality of estimation grows very quickly. This threshold effect implies that the compensation for noise nonuniformity by improving the other parameters can be very costly. Beyond the indicated values of WNPR, it is not possible to rely on subspace separation techniques to correctly infer the unknown signal parameters.

3.5 Conclusion

In the very general nonuniform noise case, the problem of estimating the number of sources is non-identifiable. It is always possible to find two parameter sets $\boldsymbol{\eta}_{p_1}$ and $\boldsymbol{\eta}_{p_2}$ such that the resulting LL functions are equal. This ambiguity is inherent to

the data model and can be easily pictured by realizing that the ideal uniform noise is a special case of spatially nonuniform noise. However the theoretical framework that provides the only typical example of non-identifiability is extremely unlikely in practice.

Estimation of the number of sources is related to the estimation of the signal subspace. Under the assumptions of ideal uniform noise, and both Gaussian signals and noise, the asymptotic behavior of the angle between the estimated signal subspace and its theoretical reference is shown to satisfy normality conditions. More importantly, the MSE on the cosine of this angle shows a threshold effect which indicates the range of SNR, L, and noise-power perturbation that allows an exploitable subspace separation. It is important to stress that beyond this threshold, subspace separation is theoretically impossible.

Chapter 4

Estimated Subspace Quality (ESQ) Test for Source Detection

4.1 Introduction

We are interested in the problem of estimating the dimension of the signal subspace without relying on the order of the eigenvalues of the covariance matrix of the data, and determine to what extent such an estimation is reliable.

Based on the analysis of Chapter 3, we exploit the results of [104] and [113] in a more useful way. More specifically, we concentrate on a measure of the *closeness* of the estimated signal subspace to its ideal reference. We use this information instead of the ordered eigenvalues of the covariance matrix of the data to infer the dimension of the estimated signal subspace. As compared to the stricter eigenvalue-equality criterion, the proposed approach offers an increased robustness to small perturbations of the noise-power.

Our analysis leads to a metric for approximate subspace separation, which results from the array transformation introduced in [113]. Based on the proposed metric, we define a set of test statistics and formulate a new sequential hypothesis test (SHT) for source detection. We use several simulation examples to show superiority of the proposed hypothesis test over a number of classical approaches, including the ones proposed by [113].

4.2 Subspace Separation

4.2.1 Array Transformation

Consider first the ideal model of (2.21). In order to determine the number of sources, we propose to simultaneously use a second array, which is excited by the same wavefront from the same sources, and combine the information from both arrays. In practice, this approach is not always realistic, especially for complicated array geometries. It is also interesting to remark that for sensors undergoing different noise-power perturbations (due to a finite sample size for example), increasing the number of sensors does not necessarily improve the *accessibility* to the signal subspace, as it introduces more unknown parameters. To partially circumvent this effect, it is possible to *isolate* the location of the perturbations, and create a second array with a different array manifold, by discarding one or more sensors from the original array, similarly to [113]. Obviously, this procedure is conducted at a cost of a reduced spatial diversity.

In what follows, for simplicity and without loss of generality, we discard the last element of the array, corresponding to the M-th row of the original array steering matrix $\mathbf{A}$, by applying a selection matrix $\mathbf{U}^{(M)} = [\mathbf{I}_{M-1}|\mathbf{0}_{M-1}]$, where the subscripts indicate the dimensions of the identity matrix and the vector of zeros, and superscript (M) refers to the discarded array element[1].

Assuming that $p < M-1$, the resulting $((M-1)\times p)$-dimensional steering matrix is given as

$$\begin{aligned} \mathbf{A}^{(M)} &:= \mathbf{U}^{(M)}\mathbf{A} \\ &= [\mathbf{a}_1^{(M)}, \mathbf{a}_2^{(M)}, \ldots, \mathbf{a}_p^{(M)}] \end{aligned} \tag{4.1}$$

where the vectors $\mathbf{a}_q^{(M)}$, $q = 1, \ldots, p$, are the same as in (2.4) with the M-th element being removed.

The covariance matrix of the collected data over the reduced $(M-1)$-element array has a structure similar to $\mathbf{R}$ in (2.21) and is related to the latter through

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}^{(M)} & \mathbf{r}^{(M)} \\ \mathbf{r}^{(M)H} & r_{(M,M)} \end{bmatrix} \tag{4.2}$$

[1]Discarding an arbitrary array element is straightforward by appropriately positioning the $\mathbf{0}_{M-1}$ column in the selection matrix.

where $r_{(M,M)}$ is the (M, M)-th element of $\mathbf{R}$ and

$$\mathbf{r}^{(M)} = \mathbf{A}^{(M)}\mathbf{R_s}\mathbf{b}^{(M)H} \tag{4.3}$$

with $\mathbf{b}^{(M)}$ being the removed M-th row of $\mathbf{A}$.

The reduced covariance matrix $\mathbf{R}^{(M)}$ has the following eigen-decomposition:

$$\mathbf{R}^{(M)} = \mathbf{E}^{(M)}\mathbf{\Lambda}^{(M)}\mathbf{E}^{(M)H} \tag{4.4}$$

with $\mathbf{E}^{(M)} := \left[\mathbf{e}_1^{(M)}, \ldots, \mathbf{e}_{M-1}^{(M)}\right]$ and $\mathbf{\Lambda}^{(M)} := \text{diag}\left\{\lambda_1^{(M)}, \ldots, \lambda_{M-1}^{(M)}\right\}$, where $\lambda_m^{(M)}$ and $\mathbf{e}_m^{(M)}$, $m = 1, \ldots, M-1$, are the eigenvalues and their corresponding eigenvectors, respectively.

In order to exploit the fact that the two arrays, with covariance matrices $\mathbf{R}$ and $\mathbf{R}^{(M)}$, respectively, are excited by the same signal wavefront, a unitary transformation matrix denoted $\mathbf{T}^{(M)}$ (with $\mathbf{T}^{(M)}\mathbf{T}^{(M)H} = \mathbf{I}_M$), can be introduced similarly to [113], as follows:

$$\mathbf{T}^{(M)} := \begin{bmatrix} \mathbf{E}^{(M)} & \mathbf{0}_{M-1} \\ \mathbf{0}_{M-1}^T & 1 \end{bmatrix}. \tag{4.5}$$

Applying transformation $\mathbf{T}^{(M)}$ to the covariance matrix $\mathbf{R}$ leads to

$$\begin{aligned} \mathcal{R}^{(M)} &= \mathbf{T}^{(M)H}\mathbf{R}\mathbf{T}^{(M)} && (4.6) \\ &= \begin{bmatrix} \mathbf{\Lambda}^{(M)} & \mathbf{c}^{(M)} \\ \mathbf{c}^{(M)H} & r_{(M,M)} \end{bmatrix} && (4.7) \end{aligned}$$

where the m-th element $c_m^{(M)}$, $m = 1, \ldots, M-1$, of vector $\mathbf{c}^{(M)}$ in (4.7) has the following structure:

$$\begin{aligned} c_m^{(M)} &= \mathbf{e}_m^{(M)H}\mathbf{A}^{(M)}\mathbf{R_s}\mathbf{b}^{(M)H} \\ &= \mathbf{e}_m^{(M)H}\mathbf{r}^{(M)}. && (4.8) \end{aligned}$$

From (4.3), it is important to note that $\mathbf{r}^{(M)}$ is not corrupted by noise and is totally defined by the signal parameters. Moreover, due to the fact that the noise subspace is ideally orthogonal to matrix $\mathbf{A}^{(M)}$, the elements $c_m^{(M)}$ satisfy the following:

$$c_m^{(M)} \begin{cases} = 0, & \text{if } \mathbf{e}_m^{(M)} \text{ is a noise eigenvector,} \\ \neq 0, & \text{if } \mathbf{e}_m^{(M)} \text{ is a signal eigenvector.} \end{cases} \tag{4.9}$$

Based on the information contained in the elements $c_m^{(M)}$, $m = 1, \ldots, M-1$, it is possible to separate the noise and the signal subspaces by ordering the magnitudes $\rho_m^{(M)} = |c_m^{(M)}|$, $m = 1, \ldots, M-1$, such that

$$\rho_1^{(M)} \geq \ldots \geq \rho_p^{(M)} \geq \rho_{p+1}^{(M)} = \ldots = \rho_{M-1}^{(M)} = 0. \tag{4.10}$$

4.2.2 Geometric Interpretation

The elements $c_m^{(M)}$ of (4.9) can be interpreted as the projection of the noise-free component of the M-th column of $\mathbf{R}$ onto the m-th eigenvector, $\mathbf{e}_m^{(M)}$, of $\mathbf{R}^{(M)}$. This clearly illustrates the isolation of the M-th sensor noise power. Since the noise and signal subspaces are disjoint, it follows that the projection $c_m^{(M)}$ provides a means to separate the two subspaces (orthogonality) and thus determine their dimensions.

Alternatively, from Gerschgorin's theorem [111], the first $M-1$ eigenvalues of the transformed covariance matrix $\boldsymbol{\mathcal{R}}^{(M)}$ are identical to those of $\mathbf{R}$ and are located in the union of Gerschgorin disks whose centers are given by the diagonal elements of $\boldsymbol{\mathcal{R}}^{(M)}$. In other words, the first $M-1$ centers are the elements of $\boldsymbol{\Lambda}^{(M)}$ and their corresponding radii are $\rho_m^{(M)}$, $m = 1, \ldots, M-1$. The value of these radii indicates the multiplicity of the eigenvalues and the subspaces that their eigenvectors span [111, 113]. It follows from (4.10) that two distinct subsets of disks are identifiable, representing the signal subspace for the p largest radii, $\rho_1^{(M)}, \ldots, \rho_p^{(M)}$, and the noise subspace for the $M-1-p$ smallest radii, $\rho_{p+1}^{(M)}, \ldots, \rho_{M-1}^{(M)}$.

4.3 Sequential Hypothesis Test for Source Detection

We use the ordered Gerschgorin radii from (4.10) rather than the ordered eigenvalues, to estimate the number of sources. The proposed detection approach measures how close is the estimated signal subspace to its theoretical reference. As it can be inferred from the simulations of Chapter 3, this closeness criterion is looser than the eigenvalue-equality criterion, which is used by most detectors in the literature, and thus is expected to offer more flexibility and robustness within the case described by (3.13). Recall that beyond this scenario, i.e., under (3.12), the separation between

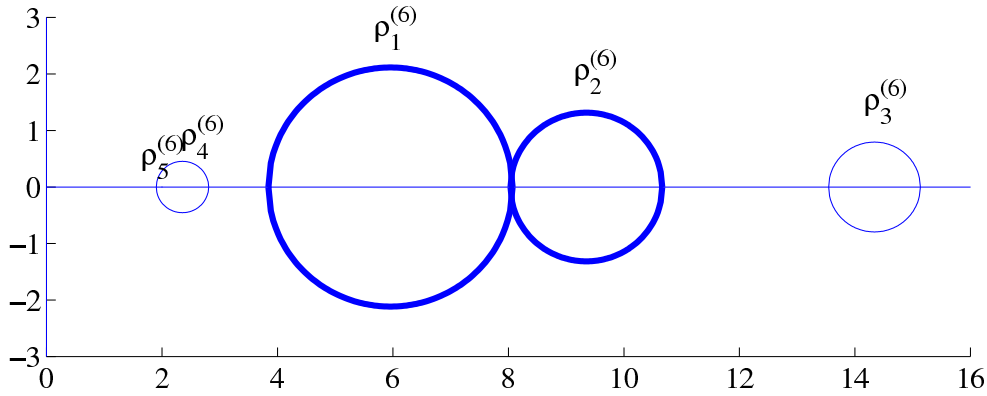

Figure 4.1: Transformed Gerschgorin disks corresponding to the example of Section 3.1. The largest two discs are connected and correspond to the *noisy* signal subspace. Note that the smallest disk is actually a point, i.e., $\rho_5^{(6)} = 0$.

the signal and noise subspaces is theoretically impossible and any approximation is problematic.

From (4.10) we can define a SHT to estimate the dimension of the signal subspace as follows:

$$\begin{array}{lcrcl} \mathsf{H}_0 & : & \rho_1^{(M)} = \ldots = \rho_{M-1}^{(M)} & = & 0 \\ \vdots & & & \vdots & \\ \mathsf{H}_q & : & \rho_{q+1}^{(M)} = \ldots = \rho_{M-1}^{(M)} & = & 0 \\ \vdots & & & \vdots & \\ \mathsf{H}_{M-2} & : & \rho_{M-1}^{(M)} & = & 0. \end{array} \tag{4.11}$$

Similarly to the sphericity test [112] (see Section 2.4.1), the above hypotheses H_q, $q = 0, \ldots, M-2$, are tested independently. The test (4.11) starts by checking that the global null, H_0, is verified, i.e., that no source is present. If H_0 is accepted, then the estimated number of sources will be $\hat{p} = 0$. If H_0 is rejected, then it can be deduced that at least one source is present. By stepping through the hypotheses H_q, $q = 0, \ldots, M-2$, the contribution of the largest Gerschgorin radius $\rho_{q+1}^{(M)}$ is eliminated sequentially and the dimension of the candidate noise subspace over which the null is tested, is reduced. The test stops when a hypothesis is accepted, or when it reaches H_{M-2}, indicating that there are $M-2$ sources.

Each hypothesis of the SHT in (4.11) is tested at the level of significance α separately. The test is performed according to the Neyman-Pearson criterion [81].

Thus, when no sources are present, the probability of correctly choosing $\hat{p} = 0$ must be controlled at $1 - \alpha$.

From (4.8) and (4.10), it is easy to verify that under H_p ($p \neq 0$), for $m = p+1, \ldots, M$, we have

$$\rho_m^{(M)} = \|\mathbf{e}_m^{(M)}\| \|\mathbf{r}^{(M)}\| \cos(\gamma_m(p)). \tag{4.12}$$

Hence, given that eigenvectors $\mathbf{e}_m^{(M)}$, $m = 1, \ldots, M-1$, are of unit norm, it follows readily that

$$\lim_{L \to \infty} \frac{\hat{\rho}_m^{(M)}}{\|\hat{\mathbf{r}}^{(M)}\|} = \cos(\gamma_m(p)) \tag{4.13}$$

provided that $\|\hat{\mathbf{r}}^{(M)}\| \neq 0$. In (4.13), $\hat{\rho}_m^{(M)}$ and $\hat{\mathbf{r}}$ are obtained from the sample covariance matrix $\hat{\mathbf{R}}$ and $\cos(\gamma_m(p))$ is defined by Corollary 3.3.2.

With the Gerschgorin radii ordered by magnitude, define the following statistic:

$$T_q := \sum_{m=q+1}^{M-1} \frac{\hat{\rho}_m^{(M)}}{\|\hat{\mathbf{r}}^{(M)}\|} \tag{4.14}$$

and observe that under H_p, for $q = p, \ldots, M-2$, the above statistic must asymptotically approach zero, for a large L. It is possible to employ the knowledge on the distribution of $\cos(\gamma_m(p))$, given by (3.16), to estimate the dimensions of the signal subspace and its complement. In particular, when $\|\hat{\mathbf{r}}^{(M)}\| \neq 0$, from Corollary 3.3.2, it follows that under H_p, when $L \to \infty$, we have asymptotically

$$T_p \xrightarrow{\mathcal{D}} X_{T_p} \tag{4.15}$$

where

$$X_{T_p} \sim \mathcal{N}\left((M-p-1)\mu_\gamma(p),\ (M-p-1)\varsigma_\gamma^2(p)\right). \tag{4.16}$$

However, the above result makes sense only for $\|\hat{\mathbf{r}}^{(M)}\| \neq 0$, which translates to the case of $p \neq 0$.

In the particular case of $p = 0$, we have $\lim_{L \to \infty} \hat{\mathbf{R}} = \sigma^2 \mathbf{I}_M$. Assuming for simplicity that the noise is real, the m-th element of the M-th column of $\hat{\mathbf{R}}$, denoted $\hat{r}_{(m,M)}$, for $m = 1, \ldots, M-1$, is defined as

$$\hat{r}_{(m,M)} := \frac{1}{L} \sum_{t=1}^{L} n_m(t) n_M(t). \tag{4.17}$$

Moreover, under H_0, the random variable $n_m(t)n_M(t)$ satisfies the normal-product distribution, given by

$$D_{n_m n_M}(u) = \frac{1}{\pi\sigma^2} K_0\left(\frac{|u|}{\sigma^2}\right) \tag{4.18}$$

where $K_\nu(\cdot)$ is a modified Bessel function of the second kind [85]. Also, since $\text{cov}\{n_m(t), n_M(t)\} = 0$, the variable $n_m(t)n_M(t)$ has a zero mean and variance σ^4. Consequently, if we consider the sum of elements of $\hat{\mathbf{r}}^{(M)}$, which is $\|\hat{\mathbf{r}}^{(M)}\|_1$, taking into account the reproductive property of the normal-product distribution, using the central limit theorem, we can state that under H_0, we have asymptotically, for $L \to \infty$

$$\|\hat{\mathbf{r}}^{(M)}\|_1 \xrightarrow{\mathcal{D}} X_{\mathbf{r}} \tag{4.19}$$

where

$$X_{\mathbf{r}} \sim \mathcal{N}\left(0, \frac{M-1}{L}\sigma^4\right). \tag{4.20}$$

Note that the statistics T_p and $\|\hat{\mathbf{r}}^{(M)}\|_1$ are completely confined to $\hat{\boldsymbol{\mathcal{R}}}^{(M)}$, which is defined as $\hat{\boldsymbol{\mathcal{R}}}^{(M)} := \mathbf{T}^{(M)H}\hat{\mathbf{R}}\mathbf{T}^{(M)}$. Since $\boldsymbol{\mathcal{R}}^{(M)}$ is a unitary transformation of $\mathbf{R}$, and $\hat{\mathbf{R}}$ is a sufficient statistic of $\mathbf{R}$, it is straightforward to verify that T_p and $\|\hat{\mathbf{r}}^{(M)}\|_1$ are also sufficient statistics, as the joint pdf of the data under the assumptions of Section 2.1, given by

$$f\left(\mathbf{X}|\boldsymbol{\eta}_p\right) = \frac{1}{[(2\pi)^M \det(\mathbf{R})]^{L/2}} \exp\left(-\frac{L}{2}\text{trace}\left\{\mathbf{R}^{-1}\hat{\mathbf{R}}\right\}\right) \tag{4.21}$$

satisfies Fisher-Neyman's factorization theorem [81].

For a candidate number of sources $q = 0, \ldots, M-2$, define $\hat{\sigma}^2 := \sum_{m=q+1}^{M-1} \frac{\hat{\lambda}_q^{(M)}}{(M-q-1)}$, and let $\hat{\mu}_\gamma(p)$ and $\hat{\varsigma}_\gamma^2(p)$ be defined by (3.5) and (3.8) respectively, with the sample eigenvalues of $\hat{\mathbf{R}}^{(M)}$ replacing the true eigenvalues of $\mathbf{R}^{(M)}$. It is important to mention here that the sample eigenvalues are ordered according to the magnitudes of the corresponding sample Gerschgorin radii following (4.10). Thus, the test (4.11), designated Estimated Subspace Quality (ESQ) test, can be summarized as follows:

1. Let $q = 0$. If $\|\hat{\mathbf{r}}^{(M)}\|_1 \leq \mathcal{E}_0(\alpha)$ is satisfied, where, given the asymptotic distribution (4.20), the threshold $\mathcal{E}_0(\alpha)$ is given by[2]

$$\mathcal{E}_0(\alpha) = \left(\hat{\sigma}^2\sqrt{(M-1)/L}\right)\operatorname{erfc}^{-1}(\alpha)$$

then accept H_0, set $\hat{p} = q = 0$ and stop the test.

2. Let $q \leftarrow q+1$. If $T_q \leq \mathcal{E}_q(\alpha)$ is satisfied, where, given the asymptotic distribution (4.16), we have

$$\mathcal{E}_q(\alpha) = \left(\hat{\varsigma}_\gamma(q)\sqrt{M-q-1}\right)\operatorname{erfc}^{-1}(\alpha) + (M-q-1)\hat{\mu}_\gamma(q)$$

then accept H_q, set $\hat{p} = q$ and stop the test.

3. If $q < M-2$, go to step 2.

4. set $\hat{p} = M-2$ and stop the test.

When the structure of the actual noise covariance matrix is known with enough accuracy, it is possible to make the above hypothesis test increasingly conservative, in the same way as the sphericity test [112]. When no accurate knowledge on the data structure is available beforehand, as it is mostly the case in practice, the test needs to be considerably more liberal to induce a significant result. As an indication, it was suggested in [112] to safely use α within the range 10% and 25% for the sphericity test. In our simulation examples, we restricted our analysis to levels not exceeding 5%. The global performance of our proposed hypothesis test varies with the actual structure of the data. Given the structure of the employed test statistics, the performance is expected to yield slightly different results when checking for $p = 0$ sources, or for any $p > 0$. In order to reduce the dependence on manually preset significance levels, one can resort to the use of penalty functions and information theoretic criteria, as explained in the next chapter.

[2] $\operatorname{erfc}^{-1}(\cdot)$ is the inverse of the complementary error function $\operatorname{erfc}(\cdot)$, which is defined as $\operatorname{erfc}(u) := 1 - \operatorname{erf}(u)$. The error function $\operatorname{erf}(\cdot)$, which will be used in the sequel, is defined as $\operatorname{erf}(u) := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{u}\exp(-v^2/2)dv$.

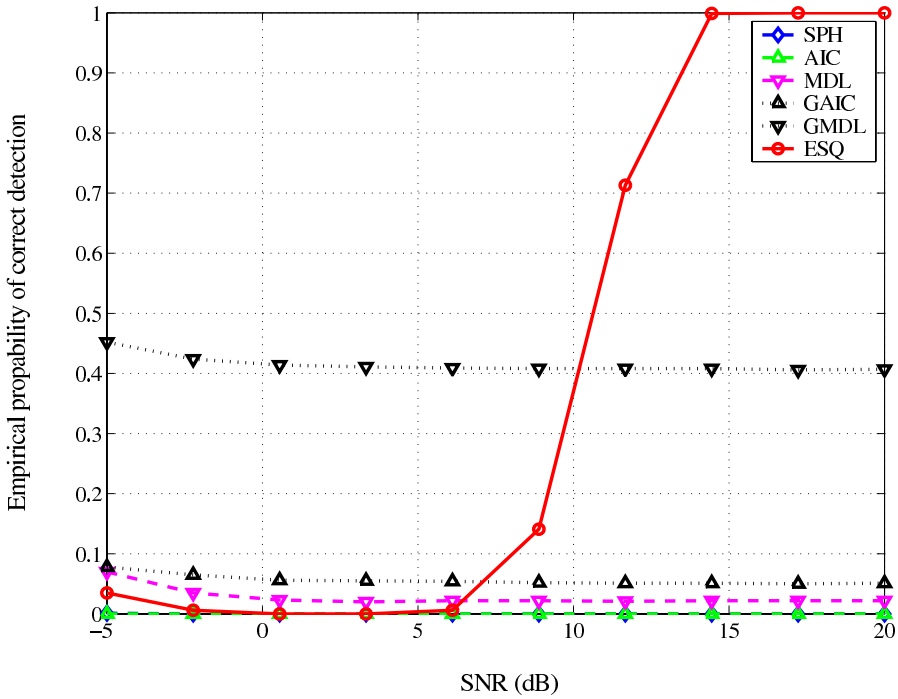

Figure 4.2: Performance of the detectors vs SNR.

4.4 Numerical Examples

We show the performance of the proposed ESQ test and compare it to the sphericity test (SPH), AIC, MDL, and GAIC, GMDL [113]. In the examples, a ULA is assumed with $M = 6$ sensors. The true number of sources is $p = 2$. We simulated different noise perturbations over the sensors. The employed noise powers are given by the vector

$$\mathbf{q} = [0.50, 0.81, 0.64, 0.51, 0.53, 1.00]^T$$

therefore we have WNPR=2. All the examples illustrate the empirical probability of correct detection resulting from 1000 Monte Carlo runs. In the examples, we illustrate both cases of $p > 0$ and $p = 0$. For the first case, we have a true number of sources $p = 2$ as using other values produces very similar results. The performance is illustrated with respect to a number of parameters as follows:

- Figure 4.2 describes a comparison of performance with respect to the SNR, which is defined with respect to the average noise-power. The fixed parameters are the number of snapshots $L = 100$ and the DOAs $\boldsymbol{\theta} = [0°, -23°]^T$, whereas

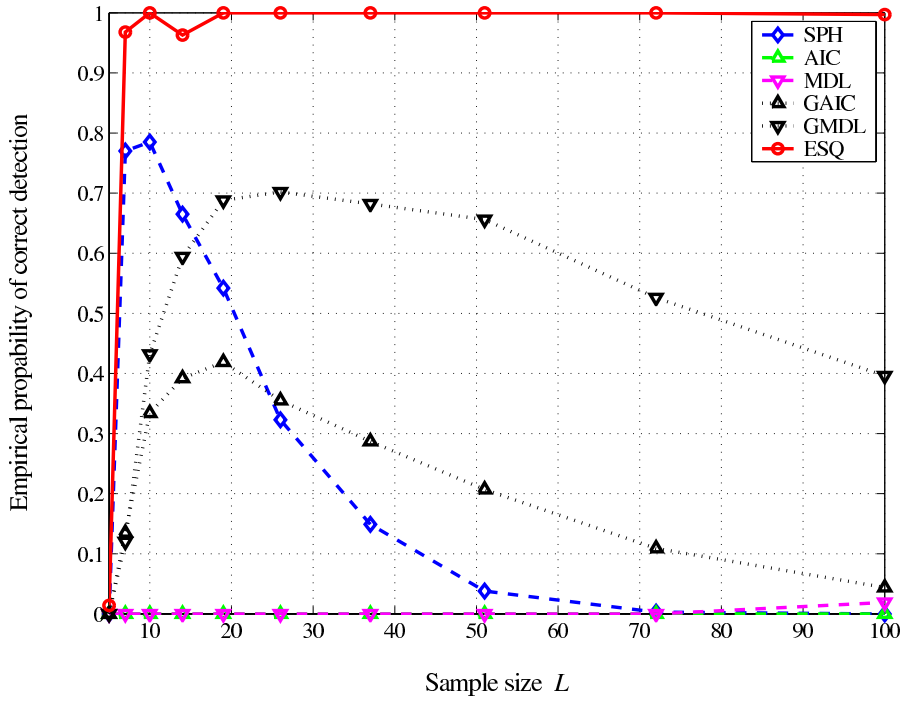

Figure 4.3: Performance of the detectors vs L.

the SNR varies from -5 dB to 20 dB. The proposed ESQ test is applied with a significance level $\alpha = 2\%$. The SPH test is applied with the same significance level. As expected, with perturbed noise powers, the proposed nonuniform detector outperforms the eigenvalue-based detectors. With a level $\alpha = 2\%$, the ESQ test is relatively sensitive to the variation of the SNR.

- Figure 4.3 illustrates the performance with respect to the number of snapshots L which varies from 5 to 100. The fixed parameters are SNR=15 dB and the DOAs $\boldsymbol{\theta} = [0°, -23°]^T$. The performance of the eigenvalue-based detectors does not improve within the illustrated range of the sample size. This shows the threshold effect on the quality of the estimated signal subspace and the relative robustness of the proposed ESQ test against the variation of noise powers.
- Figure 4.4 illustrates the performance with respect to the angular resolution $\Delta\theta$. The first DOA is fixed at $\theta_1 = 0°$ whereas the second one, θ_2, varies from $-26°$ to $0°$, while the other parameters are fixed as in the previous examples. Overall, the same relative performance of the applied detectors can

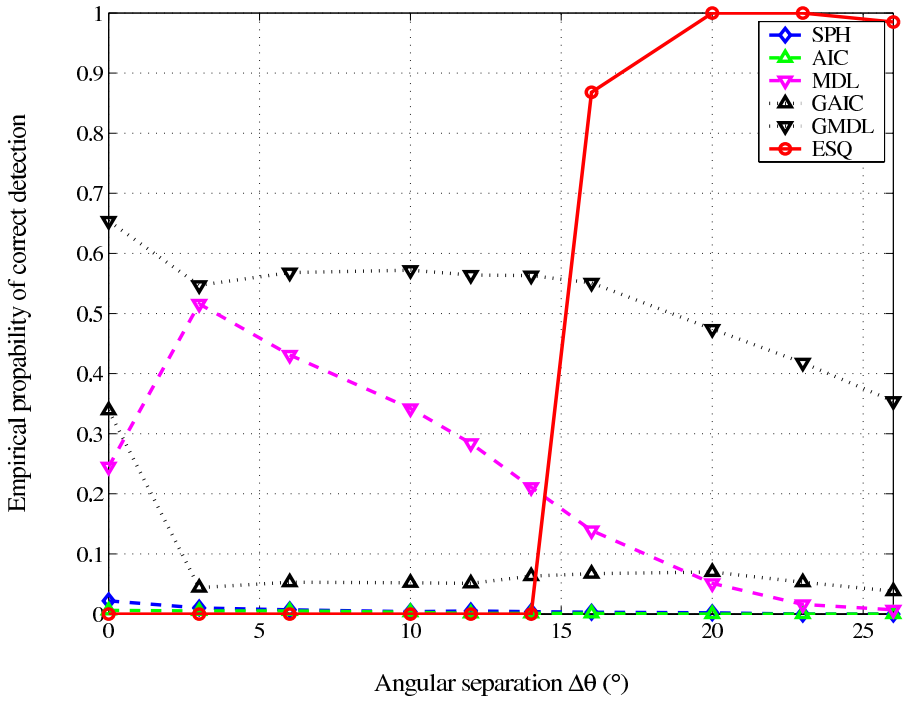

Figure 4.4: Performance of the detectors vs $\Delta\theta$.

be observed. Similarly to the SNR example, the ESQ test is relatively sensitive to angular separation. Note that the degradation in performance of the eigenvalue-based detectors appears to be due merely to the spatial nonuniformity of the noise rather than the spatial resolution of the DOAs, as seen in the following example.

- The effect of noise nonuniformity is illustrated in Figure 4.5. In this example, the WNPR varies from 1 to 10, and the noise powers are generated from a uniform distribution over $[1, \text{WNPR}]$ and normalized by their mean. The other parameters are fixed to SNR=10 dB, $L = 100$ and $\boldsymbol{\theta} = [0°, -23°]^T$. When the WNPR approaches 1, the scenario is close to the ideal uniform noise case. Recall that the general case, the variation of the WNPR does not provide an indication on the distribution of the noise powers. Observe that all the detectors undergo the effect of noise nonuniformity in a similar way, i.e., the performance decreases with increasing WNPR. However, the performance of the eigenvalue-based detectors drops abruptly as the criterion of equality of the noise eigenvalues is violated very quickly. Note the relative

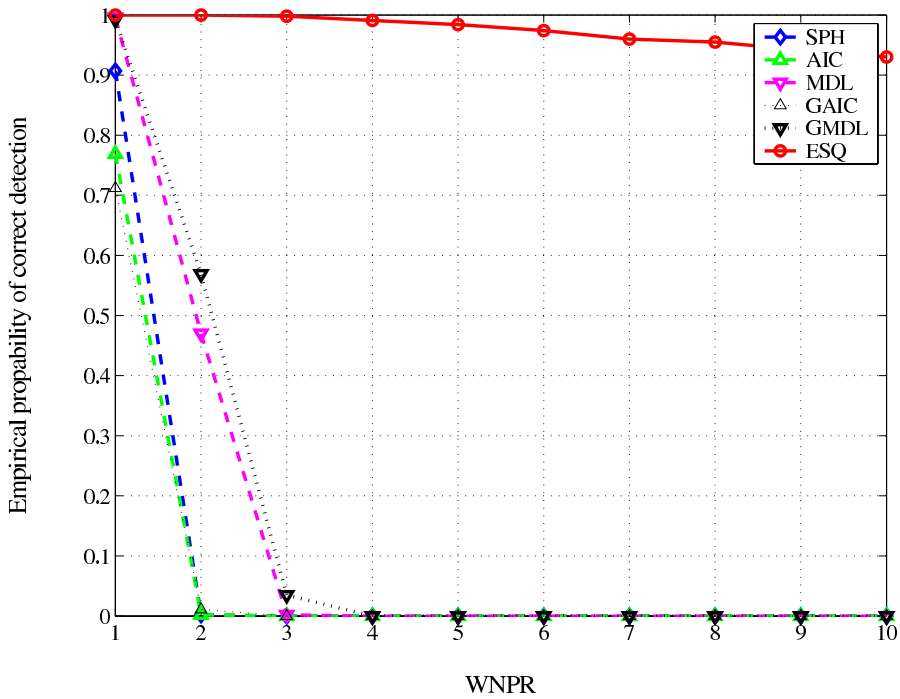

Figure 4.5: Performance of the detectors vs WNPR.

robustness of the ESQ test to the variation of the WNPR within the scenario of (3.13). Recalling the threshold effect on the estimation quality of the signal subspace, when the WNPR is further increased, the detectors' performance will degrade significantly, as the effect of additive noise will induce a strong model mismatch preventing subspace separation. Observe also that the ESQ test is perfectly suitable for uniform noise (corresponding to WNPR=1) as it performs similarly to the standard detectors. However, in this case, the ESQ suffers from a limited resolution as it requires an array transformation through the removal of one sensor.

- In Figures 4.6 and 4.7, we illustrate the ability of the detectors, and especially the hypothesis tests, to correctly decide that no sources are present. The data is thus limited to the noise. The performance is shown in terms of the variation of the sample size L and the WNPR. Here, the significance level of the ESQ and SPH tests is set to $\alpha = 2\%$.

  In the example of Figure 4.6, the WNPR is fixed to 3 and L varies from 5 to 140. The figure shows that the ESQ test performs at a level of 5%. This

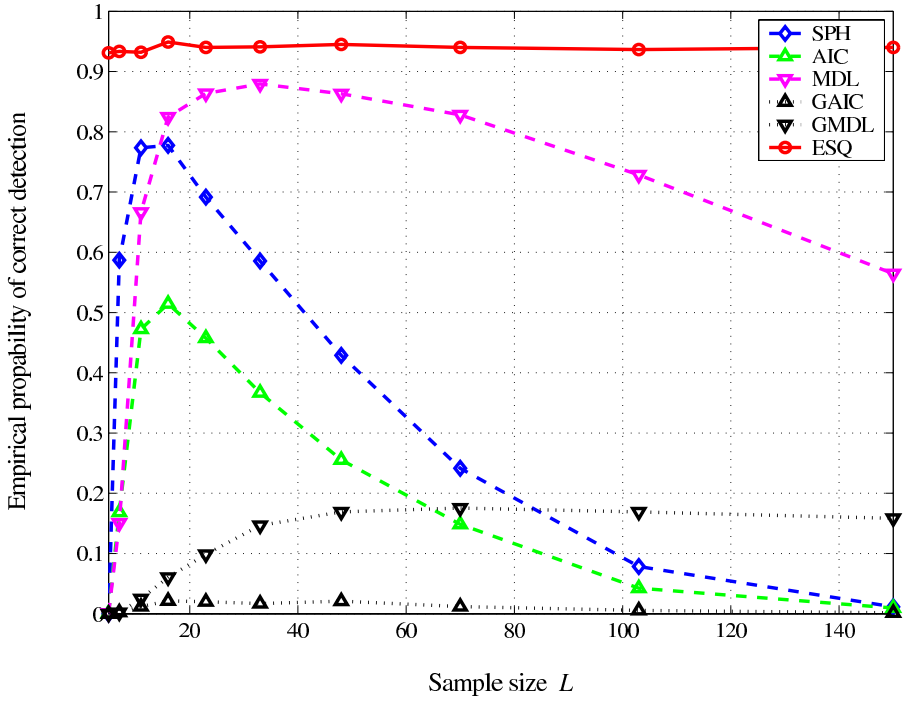

Figure 4.6: Probability of correctly detecting zero sources vs L.

fact is a result of a poor estimation of the statistics pdf tails. In addition to noise-power perturbations, one source of error that may account for this result is the finite sample size which deviates from the asymptotic conditions, i.e., the data samples may become increasingly correlated and far from the assumption of Gaussianity. For the same settings, when the significance level is made less conservative with $\alpha = 5\%$, the ESQ test tends to maintain the assigned level. The result appears to be independent of the sample size. The SPH test, as expected, rejects every hypothesis of equal eigenvalues, therefore failing to correctly assess that there are no sources present.

The effect of the noise nonuniformity through the variation of the WNPR is stronger, as illustrated by Figure 4.7, where the sample size is fixed to $L = 100$. If the WNPR increases, the noise sources with the largest powers will erroneously be assimilated to dominating signal sources, and eventually, the global significance level of the test will not be maintained.

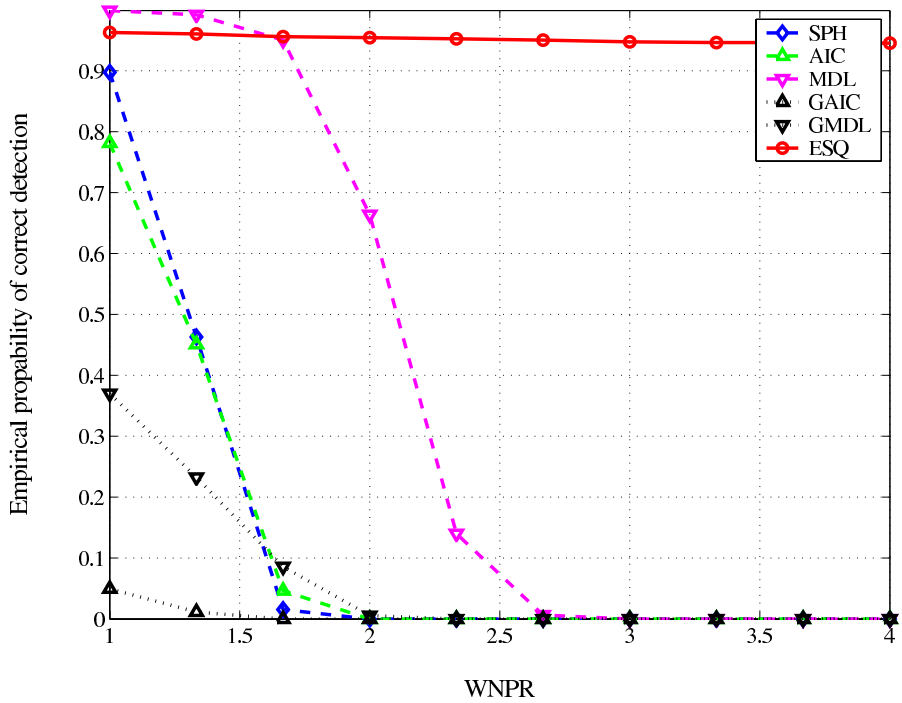

Figure 4.7: Probability of correctly detecting zero sources vs WNPR.

4.5 Conclusion

The proposed ESQ test indirectly measures the quality of the estimated signal subspace. It is more relaxed than the eigenvalue-equality criterion that is used by classical detectors. The offered flexibility can be exploited up to a certain threshold on the variation of the subspace estimation error. Beyond such a threshold, the effect of additive noise prevents any subspace separation.

The proposed detectors follow from a transformation of the covariance matrix of the data, resulting from array element suppression to cope with possible noise-power discrepancies. Using Gerschgorin's theorem, a metric different from the ordered eigenvalues is used for subspace separation.

Simulation results show the power of the detection schemes and their robustness to small noise-power perturbations. However, a superior performance in spatially nonuniform noise implies a more restrictive identifiability condition on the number of array sensors, i.e., $p < M - 1$ and not $p < M$.

Chapter 5

Nonuniform Information Theoretic Criteria

5.1 Introduction

In what follows, we propose an alternative detection criterion by deriving a new stochastic likelihood function. The method partially copes with noise-power perturbation by the same approach of successively eliminating the contribution of single elements from the array. The resulting LL function demonstrates robustness to the variation of the noise powers.

Even-though the proposed information criteria are based on large-sample approximations, they fill a gap between simple but sensitive detection approaches, and robust but more complicated methods requiring prior knowledge on the data.

5.2 Information Theoretic Criteria for Source Detection

Under the Gaussianity assumption, the stochastic negative LL function of the observed data is given by [81]

$$\mathcal{L}(\boldsymbol{\eta}_p) = L \ln \left\{ \det \left[\mathbf{R}(\boldsymbol{\eta}_p) \right] \right\} + \operatorname{trace} \left\{ \mathbf{R}^{-1}(\boldsymbol{\eta}_p) \hat{\mathbf{R}} \right\} \tag{5.1}$$

where we introduce the dependence of $\mathbf{R}$ on the vector of unknown parameters $\boldsymbol{\eta}_p$.

Similarly to the ESQ hypothesis test of Chapter 4, in what follows we do not use any *a priori* knowledge on the order of the eigenvalues of $\mathbf{R}$ and discard it completely from the LL function. Instead, we use the information provided by the ordered Gerschgorin radii $\rho_m^{(M)}$, $m = 1, \ldots, M-1$, as defined in (4.10).

Recalling that the transformation matrix $\mathbf{T}^{(M)}$ of (4.5) is unitary, matrix $\mathbf{R}$ can be replaced by the transformed covariance matrix $\boldsymbol{\mathcal{R}}^{(M)}$ defined in (4.6)[1]. Leaving an explicit dependence of the LL function only on p, the following modified negative LL function is obtained:

$$\mathcal{L}^{(M)}(p) = L \ln \left\{ \det \left(\boldsymbol{\mathcal{R}}^{(M)}(p) \right) \right\} + \operatorname{trace} \left\{ \left(\boldsymbol{\mathcal{R}}^{(M)}(p) \right)^{-1} \hat{\boldsymbol{\mathcal{R}}}^{(M)} \right\}. \tag{5.2}$$

Exploiting the fact that matrix $\boldsymbol{\Lambda}^{(M)}$ in (4.7) is diagonal, relation (4.10) and properties of (2×2)-block matrices, we have

$$\begin{aligned} \det \left(\boldsymbol{\mathcal{R}}^{(M)}(p) \right) &= \det \left(\boldsymbol{\Lambda}^{(M)} \right) \det \left(r_{(M,M)} - \mathbf{c}^{(M)H} \left(\boldsymbol{\Lambda}^{(M)} \right)^{-1} \mathbf{c}^{(M)} \right) \\ &= \left(\prod_{m=1}^{M-1} \lambda_m^{(M)} \right) \left(r_{(M,M)} - \sum_{m=1}^{p} \frac{\left(\rho_m^{(M)} \right)^2}{\lambda_m^{(M)}} \right). \end{aligned} \tag{5.3}$$

From (4.6) we also have

$$\operatorname{trace} \left\{ \left(\boldsymbol{\mathcal{R}}^{(M)}(p) \right)^{-1} \hat{\boldsymbol{\mathcal{R}}}^{(M)} \right\} = \operatorname{trace} \left\{ \mathbf{R}^{-1}(p) \hat{\mathbf{R}} \right\} \approx M. \tag{5.4}$$

Using (5.3) and (5.4), and omitting terms independent of p, the negative LL function (5.2) reduces to

$$\mathcal{L}^{(M)}(p) = L \ln \left(r_{(M,M)} - \sum_{m=1}^{p} \frac{\left(\rho_m^{(M)} \right)^2}{\lambda_m^{(M)}} \right). \tag{5.5}$$

The obtained LL function is monotonic with respect to the squared values of $\rho_m^{(M)}$, $m = 1, \ldots, p$. Note that only these elements provide the necessary information for the estimation of the number of sources p. The corresponding eigenvalues $\lambda_m^{(M)}$,

[1] The effect of the unitary transformation $\mathbf{T}^{(M)}$ can be seen as a re-parameterization of $\mathbf{R}(\boldsymbol{\eta}_p)$ into $\boldsymbol{\mathcal{R}}^{(M)}(p, \mathbf{c}^{(M)})$.

$m = 1, \ldots, p$, are ordered accordingly and are only scaling factors. Following Section 2.4.2, this LL function describes the goodness-of-fit part for a detection criterion [106]. The number of free parameters in the negative LL function (5.5) is clearly p for the elements $\rho_m^{(M)}$, and p^2 for the signal subspace which translates to the rank condition on $\mathbf{R}$ or $\mathbf{R}^{(M)}$ [108, 113]. Following the development in [106], a penalty function, denoted $\mathcal{P}(p)$, can be used, leading to an information theoretic criterion, $\mathcal{K}^{(M)}(p)$, for automatic source number estimation as follows [1, 80, 106, 116]:

$$\mathcal{K}^{(M)}(p) := \mathcal{L}^{(M)}(p) + \mathcal{P}(p) \tag{5.6}$$

where the estimated number of sources, $\hat{p}$, is obtained by minimizing the criterion (5.6). In practice, due to the finite data length, the sample Gerschgorin radii and the sample eigenvalues are used. The above criterion does not measure the equality of the smallest eigenvalues of $\mathbf{R}^{(M)}$ and is therefore less sensitive than the conventional criteria for cases where the noise-power undergoes small perturbations.

Bearing in mind this robustness to some degree of noise nonuniformity, using the expressions of the penalty functions in (2.36) and (2.37) respectively, we define the Nonuniform Akaike Information Criterion (NU-AIC) and the Nonuniform Minimum Description Length (NU-MDL) as

$$\text{NU-AIC}^{(M)}(p) := \arg \min_{0 \le p < M-1} \left\{ \mathcal{L}^{(M)}(p) + (p^2 + p) \right\} \tag{5.7}$$

and

$$\text{NU-MDL}^{(M)}(p) := \arg \min_{0 \le p < M-1} \left\{ \mathcal{L}^{(M)}(p) + \frac{1}{2}(p^2 + p)\ln(L) \right\} \tag{5.8}$$

or a more general version of the latter, the Nonuniform Efficient Detection Criterion (NU-EDC) as [115]

$$\text{NU-EDC}^{(M)}(p) = \arg \min_{0 \le p < M-1} \left\{ \mathcal{L}^{(M)}(p) + (p^2 + p) \cdot \mathcal{C}(L) \right\} \tag{5.9}$$

where $\mathcal{C}(L)$ is a function satisfying $\lim_{L \to \infty} \mathcal{C}(L)/L = 0$ and $\lim_{L \to \infty} \mathcal{C}(L)/\ln(\ln(L)) = \infty$, and superscript (M) stands for the removed M-th element of the array.

5.2.1 Asymptotic Performance

Following the derivation in [106], consistency of the above detectors can be established by showing that in the large-sample limit, the criteria defined in (5.6) are minimized for the true source number p.

Lemma 5.2.1. *Under the assumptions of Section 2.2, with Gaussian source samples and noise and L large, the NU-MDL criterion defined by (5.8) is a consistent source number estimator, whereas the NU-AIC defined by (5.7) is not consistent.*

Proof. See Appendix B.3. □

Consistency of the proposed detectors can also be verified through the evaluation of their probability of error. The latter comprises the probability of missing a source, or probability of under-modeling, denoted $P_M(p)$, and the probability of false alarm, or probability of over-modeling, denoted $P_F(p)$, for a given number of sources p. These probabilities are defined as

$$P_M(p) := P\left\{\hat{p} < p | \mathsf{H}_p\right\} \tag{5.10}$$

$$P_F(p) := P\left\{\hat{p} > p | \mathsf{H}_p\right\}. \tag{5.11}$$

Similarly to [104] and [33], for simplicity, we assume that for $L \to \infty$, we have

$$P\left\{\hat{p} = p-1 | \mathsf{H}_p\right\} \gg P\left\{\hat{p} < p-1 | \mathsf{H}_p\right\} \tag{5.12}$$

$$P\left\{\hat{p} = p+1 | \mathsf{H}_p\right\} \gg P\left\{\hat{p} > p+1 | \mathsf{H}_p\right\} \tag{5.13}$$

hence simplifying expressons (5.10) and (5.11) to

$$P_M(p) \approx P\left\{\mathcal{K}^{(M)}(\hat{p} = p-1) < \mathcal{K}^{(M)}(\hat{p} = p) | \mathsf{H}_p\right\} \tag{5.14}$$

$$P_F(p) \approx P\left\{\mathcal{K}^{(M)}(\hat{p} = p+1) < \mathcal{K}^{(M)}(\hat{p} = p) | \mathsf{H}_p\right\}. \tag{5.15}$$

The evaluation of the asymptotic probability of error requires at least a sufficiently accurate approximation of the distribution of the eigenvalues of $\hat{\mathbf{R}}$. This issue was addressed in [104, 50] for the case of classical information criteria. For $m = 1, \ldots, p$ the signal sample eigenvalues are known to be asymptotically independent and as $L \to \infty$, they satisfy the following [2]:

$$\hat{\lambda}_m \sim \mathcal{N}\left(\lambda_m, \lambda_m^2/L\right). \tag{5.16}$$

Besides, the empirical results of [50] show that as $L \to \infty$, the largest noise sample eigenvalue can be approximated with great accuracy by a Gaussian random variable X_σ, such that

$$X_\sigma \sim \mathcal{N}\left(\sigma^2\left(1 + \frac{\zeta(M,p,L)}{\sqrt{L}}\right), \sigma^4 \frac{\xi(M,p,L)}{L}\right) \tag{5.17}$$

with,

$$\zeta(M,p,L) = O(2\sqrt{M-p}) \tag{5.18}$$

$$0 < \xi(M,p,L) < 1 \quad . \tag{5.19}$$

Moreover, a less strict approximation resulting from setting $\zeta(M,p,L)/\sqrt{L} = 0$ and $\xi(M,p,L) = 1$ when $L \to \infty$, does not lead to a significant change in the asymptotic performance of the information criteria of interest [50]. Therefore, for simplicity, in order to easily compare the asymptotic performance of several detectors, we shall resort to the latter approximate asymptotic distribution of the sample eigenvalues, as was initially suggested by [104]. Hence, in what follows, as $L \to \infty$, for $m = 1,\ldots,p+1$, the sample eigenvalues $\hat{\lambda}_m$ are considered to asymptotically satisfy the distribution (5.16).

Result 5.2.2. *From Corollary 3.3.2, using the expression of the asymptotic distribution of* $\cos(\gamma_m(p))$ *provided by (3.16), and the expression of the LL function of (5.5), for* $p = 1,\ldots,M-2$ *and* $L \to \infty$*, probability (5.14) is given by*

$$P_M(p) = erf\left(\frac{\|\mathbf{r}^{(M)}\|^{-1}\sqrt{W(p)\Delta\mathcal{P}(p)/L} - \mu_\gamma(p)}{\varsigma_\gamma(p)}\right). \tag{5.20}$$

For $p = 1,\ldots,M-2$*, when* $L \to \infty$*, probability (5.15) is given by*

$$P_F(p) = 1 - erf\left(\frac{\|\mathbf{r}^{(M)}\|^{-1}\sqrt{W(p+1)\Delta\mathcal{P}(p+1)/L} - \mu_\gamma(p)}{\varsigma_\gamma(p)}\right) \tag{5.21}$$

and for $p = 0$ *and* $L \to \infty$*, probability (5.15) converges asymptotically to*

$$P_F(0) = 1 - erf\left(\frac{\sqrt{W(1)\Delta\mathcal{P}(1)/2}}{\sigma^2}\right) \tag{5.22}$$

where $\Delta\mathcal{P}(p) = \mathcal{P}(p) - \mathcal{P}(p-1)$*, i.e., a penalty-function difference,* $\mu_\gamma(p)$ *and* $\varsigma_\gamma^2(p)$ *are defined by (3.17) and (3.18), respectively, and*

$$W(p) = \lambda_p^{(M)}\left(r_{(M,M)} - \sum_{m=1}^{p-1}\frac{\left(\rho_m^{(M)}\right)^2}{\lambda_m^{(M)}}\right) \tag{5.23}$$

with $W(1) = \lambda_1^{(M)} r_{(M,M)}$.

Proof. See Appendix B.4. □

In [104], a similar result was established for the conventional eigenvalue-based information criteria (such as AIC, MDL), for the special cases of $p = 0, 1$ and 2. For instance, the expressions of $P_{M_{\mathbf{\Lambda}}}(1)$ and $P_{F_{\mathbf{\Lambda}}}(1)$ of [104] are given by

$$P_{M_{\mathbf{\Lambda}}}(1) = \operatorname{erf}\left(\frac{\sqrt{2\Delta\mathcal{P}(1)} - \sqrt{L}M\mathrm{SNR}}{1 + M\mathrm{SNR}}\right) \tag{5.24}$$

$$P_{F_{\mathbf{\Lambda}}}(1) = 1 - \operatorname{erf}\left(\sqrt{2\Delta\mathcal{P}(2)}\right) \tag{5.25}$$

where subscript $\mathbf{\Lambda}$ refers to the ordered eigenvalues of $\mathbf{R}$, and under H_1, we have $\lambda_1 = (1 + M\mathrm{SNR})\sigma^2$.

A simple comparison of the probabilities of false alarm $P_F(1)$ and $P_{F_{\mathbf{\Lambda}}}(1)$ indicates the following:

Result 5.2.3. *Using Result 5.2.2, under* H_1*, the proposed nonuniform information criteria result in an asymptotically smaller probability of false alarm than the conventional eigenvalue-based information criteria, i.e., for* $L \to \infty$*, we have*

$$P_F(1) \leq P_{F_{\mathbf{\Lambda}}}(1). \tag{5.26}$$

Proof. See Appendix B.5. □

However, it is also readily verifiable that the total probability of error is almost entirely determined by the probability of missing a source, as the probability of false alarm is negligible in comparison. Moreover, comparing $P_M(1)$ and $P_{M_{\mathbf{\Lambda}}}(1)$ shows a different result, i.e., asymptotically, for a large L, we have $P_{M_{\mathbf{\Lambda}}}(1) \leq P_M(1)$, as a result of a higher spatial resolution due to an extra active sensor and a better conditioning of the data covariance matrix. As L becomes infinite, both probabilities coincide.

An exception to this result occurs when L is sufficiently large to justify the asymptotic approximation (5.20) but not enough to *compensate* for a strong decrease of the SNR.

Result 5.2.4. *Using Result 5.2.2, under* H_1*, for large* L *and* M*, with* $M\mathrm{SNR} \ll 1$ *and* $L \ll 1/(M\mathrm{SNR})^2$*, the proposed nonuniform information criteria result in*

an asymptotically smaller probability of missing a source than the conventional eigenvalue-based information criteria, i.e.,

$$P_M(1) \;\leq\; P_{M_\Lambda}(1). \tag{5.27}$$

Proof. See Appendix B.5. □

Hence, for a sufficiently large observation time, under the above particular scenario, it appears that the transformed Gerschgorin radii lead to a better source detection than the ordered sample eigenvalues. This fact is accentuated when the noise is spatially nonuniform. Indeed, the effect of noise-power perturbations at the sensors can be seen as an increase in the overall SNR, or MSNR in (5.24). This follows from the observation that the closest matrix (in Frobenius norm) to $\mathbf{R} = \sum_{m=1}^{M} \lambda_m \mathbf{e}_m \mathbf{e}_m^H$ such that *the overall noise power* is of multiplicity $M-p$, is given by $\tilde{\mathbf{R}} = \sum_{m=1}^{p} \lambda_m \mathbf{e}_m \mathbf{e}_m^H + \sum_{m=p+1}^{M} \left(\sum_{m=p+1}^{M} \lambda_m/(M-p) \right) \mathbf{e}_m \mathbf{e}_m^H$, both matrices being of rank p [33]. In this case, nonuniform noise would result in a larger overall SNR. The dependence between the SNR and L and their effect on the quality of the estimated signal subspace, for both uniform and nonuniform noise, was illustrated through simulation examples in Chapter 3.

Under the assumptions of Chapter 2, a similar analysis can be extended to cases other than $p = 1$ (typically $P_F(0)$, $P_M(2)$ and $P_F(2)$). The above results essentially follow from the common penalty function $\mathcal{P}(p)$ and the asymptotic properties of the angles between the estimated noise eigenvectors and the ideal signal subspace.

5.2.2 Comparison with the Gerschgorin Likelihood Estimator

A similar formulation, namely the Gerschgorin Likelihood Estimator (GLE), was presented in [113]. The resulting GAIC and GMDL criteria (Section 2.4.2) apply the same penalty function as in (5.7) and (5.8), respectively, while their goodness-of-fit part, or LL function, has the following expression:

$$\mathcal{L}^{(M)}(p) = -L(M-1-p) \ln \left\{ \frac{\mathcal{G}_p(\lambda^{(M)})}{\mathcal{M}_p(\lambda^{(M)})} \right\} + L \ln \left(r_{(M,M)} - \sum_{m=1}^{p} \frac{\left(\rho_m^{(M)}\right)^2}{\lambda_m^{(M)}} \right) \tag{5.28}$$

given that the discarded sensor is of index M.

As compared to (5.5), the above LL function contains an extra part in addition to the contribution of the elements of $\mathbf{c}^{(M)}$ as seen previously. This extra part is related to the statistics of (2.29), as it specifically incorporates *a priori* knowledge on the order of the eigenvalues corresponding to the noise subspace. It measures the equality of these eigenvalues, like other conventional detection criteria.

Another important difference concerns consistency. It can be easily shown that GMDL is consistent, in the same way as MDL and NU-MDL. However, while most AIC-based criteria are not consistent, the GAIC is shown in [113] to be consistent by graphical proof [114].

For a more general framework, this consistency can be easily confirmed analytically.

Lemma 5.2.5. *Under the conditions of Section 2.2, for a large L, the GAIC criterion defined by the LL function of (5.5), and penalty function $\mathcal{P}(p) = p^2 + p$, is a consistent source number estimator.*

Proof. See Appendix B.7. □

The above Lemma indicates that asymptotically, as $L \to \infty$, the combination of the *correct order of the sample eigenvalues* with the transformed Gerschgorin radii on which the GLE criteria are based, provides an improved performance over both classical and our nonuniform criteria.

Result 5.2.6. *Under H_1, the GLE criteria of [113] result in an asymptotically smaller probability of error than both the conventional eigenvalue-based and the proposed nonuniform criteria, i.e., when $L \to \infty$, we have*

$$\begin{cases} P_{M_{GLE}}(1) & \leq & P_{M_{\mathbf{\Lambda}}}(1) \\ P_{F_{GLE}}(1) & \leq & P_{F_{\mathbf{\Lambda}}}(1) \end{cases} \tag{5.29}$$

$$\begin{cases} P_{M_{GLE}}(1) & \leq & P_{M}(1) \\ P_{F_{GLE}}(1) & \leq & P_{F}(1) \end{cases} \tag{5.30}$$

where $P_{M_{GLE}}(p)$ and $P_{F_{GLE}}(p)$ are the probabilities of missing a source and of false-alarm of the GLE criteria, respectively.

Proof. See Appendix B.8. □

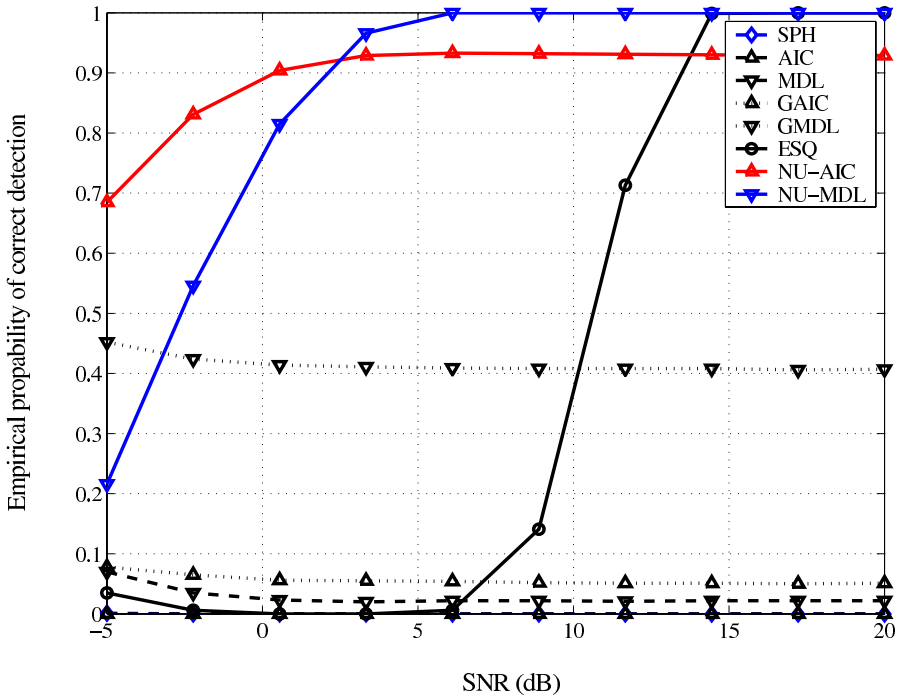

Figure 5.1: Performance of the detectors vs SNR.

However, it is worth mentioning that the advantage of the GLE criteria is based on the exact ordering of the sample eigenvalues. Similarly to other classical eigenvalue-based criteria, the GLE criteria are sensitive to noise-power perturbations and low SNR. In such a scenario, our proposed nonuniform criteria provide a better performance, as will be illustrated by the simulations of the next section.

Note also that although our proposed nonuniform criteria use the ordered Gerschgorin radii of $\boldsymbol{\mathcal{R}}^{(M)}$, they are not a generalization of the the GLE criteria of [113], as they result from a different LL function.

5.3 Numerical Examples

In this section we illustrate the performance of the proposed criteria for the same settings as in Chapter 4, i.e., we employ a ULA with $M = 6$ sensors for $p = 2$ sources, and the noise perturbations over the sensors are given by the vector

$$\mathbf{q} = [0.50, 0.81, 0.64, 0.51, 0.53, 1.00]^T$$

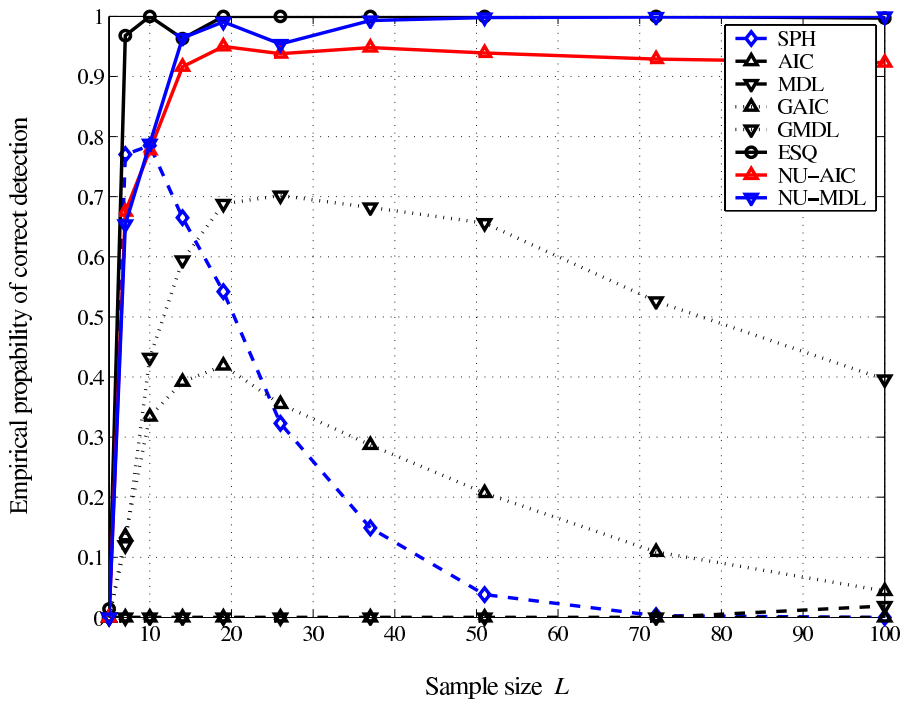

Figure 5.2: Performance of the detectors vs L.

therefore we have WNPR=2. Also, the examples illustrate the empirical probability of correct detection resulting from 1000 Monte Carlo runs. Similarly to the previous chapter, we illustrate both cases of $p > 0$ and $p = 0$. For the first case, we use a true number of sources $p = 2$ and emphasize that using other values produces similar results. It is important to mention that although the proposed nonuniform information criteria often outperform the ESQ test, the latter provides a probability measure of the correctness of the result, a feature important in practice.

- Figure 5.1 shows the performance with respect to the SNR, which is defined with respect to the average noise power. The ESQ hypothesis test of Chapter 4, as well as the SPH test, are applied with a significance level $\alpha = 2\%$. Observe that NU-AIC and NU-MDL retain the general relative performance of AIC and MDL in the uniform noise case [33, 104, 106]. Observe also that the eigenvalue-based information criteria consistently over-model the data. This observation conforms to the analysis of [60], illustrating the mismatch due to a poor clustering of the noise eigenvalues. Note the relative robustness to the SNR of the nonuniform information criteria as opposed to the ESQ test.

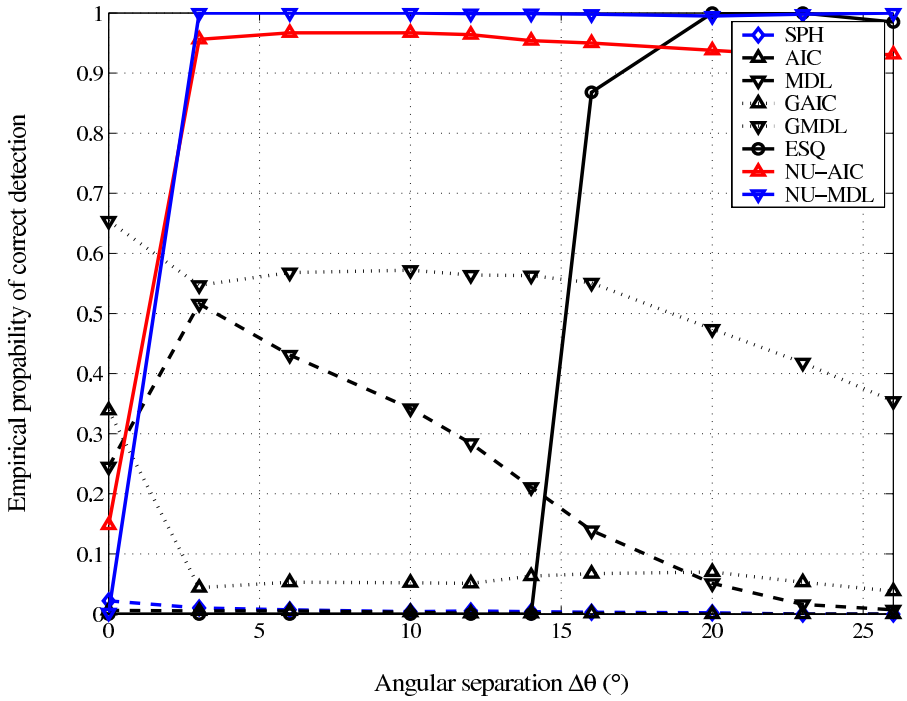

Figure 5.3: Performance of the detectors vs $\Delta\theta$.

- Figure 5.2 illustrates the performance with respect to the number of snapshots L. We can see the threshold effect on the quality of the estimated signal subspace and the relative robustness of the proposed nonuniform detectors against the variation of the noise powers. As expected, the nonuniform criteria consistently outperform the conventional detectors that undergo model mismatch.

- Figure 5.3 illustrates the performance with respect to the angular resolution $\Delta\theta$. Overall, the same relative performance of the applied detectors can be observed. Here again, the ESQ test is more sensitive to the angular separation than the nonuniform information criteria, which are known for their consistency. From the expressions of the first- and second-order moments of the cosine of the angle between the ideal signal subspace and the available estimate, it is clear that close sources may induce a significant bias in the sample Gerschgorin radii, which impacts on the ESQ test if it is too conservative. The nonuniform criteria use a fixed penalty function.

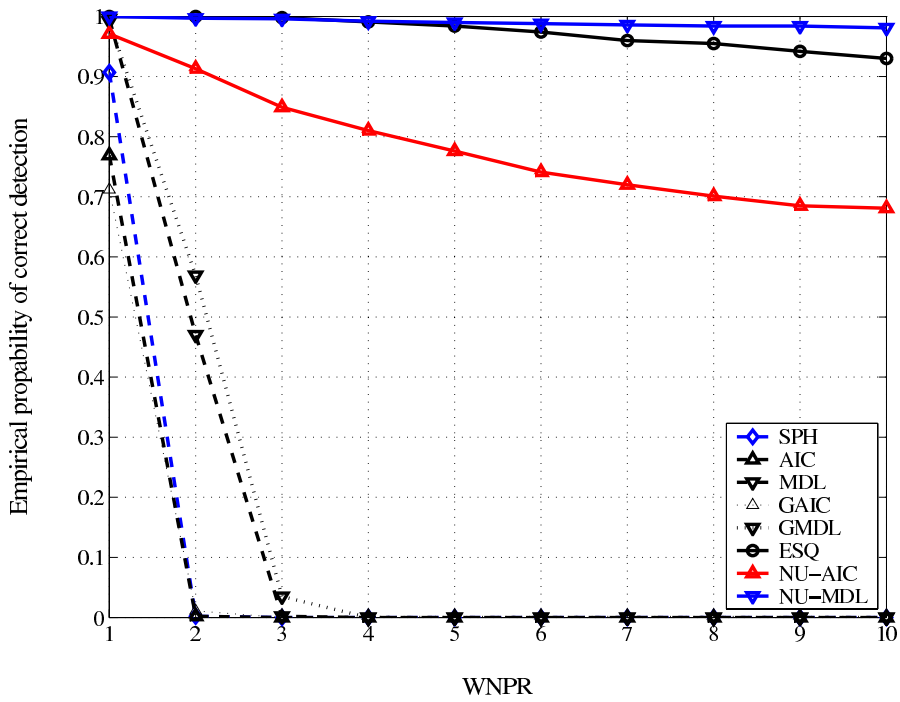

Figure 5.4: Performance of the detectors vs WNPR.

- The effect of noise nonuniformity is illustrated in Figure 5.4. Recalling the threshold effect on the estimation quality of the signal subspace, when the WNPR is further increased, the detectors' performance will degrade significantly, as the effect of additive noise will induce a strong model mismatch. Recall that in the general case, the WNPR does not provide an information on the distribution of the noise powers.

- In Figures 5.5 and 5.6, we illustrate the ability of the detectors, especially the hypothesis tests, to correctly decide that no sources are present. The data is thus limited to the noise. All the settings are the same as for the examples of the previous chapter. Note that while often outperforming the eigenvalue-based detectors for an increasing L, the proposed information criteria, i.e., NU-AIC and NU-MDL, perform poorly. In particular, NU-AIC provides too small a detection rate to be useful. The reason can be seen from the asymptotic expression of $P_F(0)$ in (5.22). Clearly, the probability of error depends only on the functional form of the particular penalty function and its variation with L. The sensitivity of NU-AIC then becomes apparent.

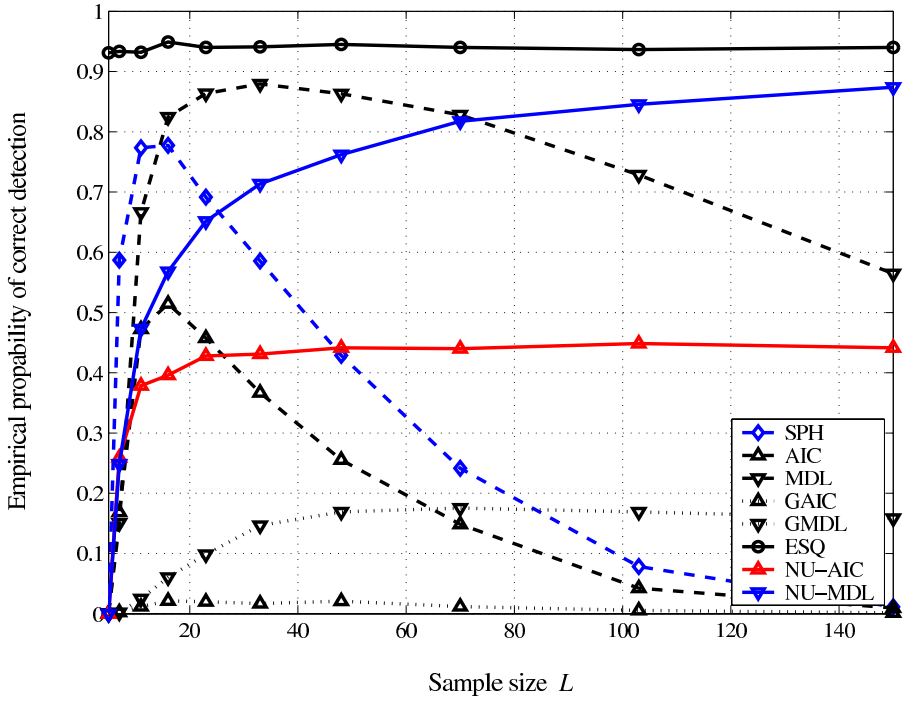

Figure 5.5: Probability of correctly detecting zero sources vs L.

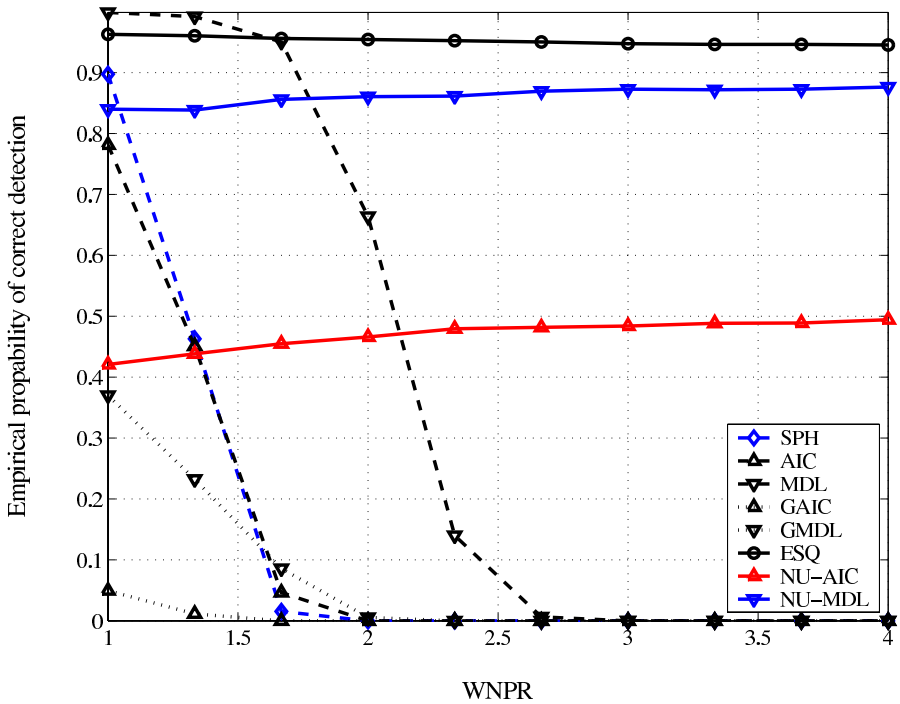

Figure 5.6: Probability of correctly detecting zero sources vs WNPR.

5.4 Conclusion

Similarly to the ESQ test, the proposed criteria, i.e., NU-AIC and NU-MDL, follow from array transformation and Gerschgorin's theorem. They are based on a criterion measuring the quality of the estimated signal subspace and do not include any *a priori* knowledge on the ordered eigenvalues of the data covariance matrix.

The performance of the criteria of interest is asymptotically analyzed and compared to the classic detectors. This analysis, supported by simulation examples, showed the superiority of the nonuniform detectors over classical methods of the same type.

Due to the required array transformation, this improvement in the global performance comes at a cost of reduced spatial diversity and stricter identifiability conditions on the number of resolvable sources with respect to array elements.

Chapter 6

Application of the Bootstrap to Source Detection

6.1 Introduction

The main drawback of the nonuniform information criteria of Chapter 5 is that their LL function is specifically derived for Gaussian signals. If the assumption of Gaussianity is not verified, especially for a low SNR and spatially close sources, the behavior of the aforementioned detectors can degrade significantly.

In what follows, we propose an alternative detection scheme based on hypothesis testing, without making prior assumptions on the distribution of the data. Similarly to the ESQ test, the proposed test statistics are based on the discriminating property of the Gerschgorin radii [113]. To fully take into account the information provided by the subspace projectors, or equivalently the transformed Gerschgorin radii, a formulation of a conventional SHT and a MHT is considered. To keep the experiment as wide as possible (wide variety of data distributions), the null distributions of the employed test statistics are empirically inferred using the bootstrap [31, 121].

6.2 Source Detection

6.2.1 Sequential Hypothesis Test

The SHT of (4.11) is applied in the same way as the sphericity test [112], i.e., it starts by testing the global null H_0 (no sources), and only upon rejection of H_0, hypothesis H_1 is tested. The same procedure is repeated until a hypothesis H_q is accepted where the test stops, indicating that $\hat{p} = q$ sources are present.

Beyond the stochastic Gaussian case, the separation of the signal and noise subspaces based on the ordered Gerschgorin radii remains valid, and the cumulated squared radii can still be used to indicate the dimension of the signal and noise subspaces. This suggests the following test statistic:

$$T_{\Sigma_q} := \sum_{m=q+1}^{M-1} \left(\hat{\rho}_m^{(M)}\right)^2, \qquad q = 0, \ldots, M-2. \tag{6.1}$$

In practice, the values of the statistic T_{Σ_q} will be close to zero if all the radii that it encloses correspond to the noise subspace, and significantly greater than zero otherwise.

Alternatively, in order to more easily check for zeros, a test statistic based on the difference between two metrics can be formulated for $q = 0, \ldots, M-2$, as follows:

$$T_{\Delta_q} := \left(\frac{1}{M-1-q} \sum_{m=q+1}^{M-1} \hat{\rho}_m^{(M)}\right) - \left(\prod_{m=q+1}^{M-1} \left(\hat{\rho}_m^{(M)}\right)^{\frac{1}{M-1-q}}\right). \tag{6.2}$$

The statistic T_{Δ_q} compares the arithmetic and geometric means of the radii. Each of these two metrics will be close to zero when the considered radii correspond to the noise subspace, i.e., when the null is in force, leading to a value of T_{Δ_q} that can be closer to zero, as compared to that of T_{Σ_q}.

As explained in Chapter 4, the global null corresponds to the case where all the Gerschgorin radii are equal to zero. The threshold for the test of hypotheses in (4.11) can be set according to the desired significance level α (Neyman-Pearson criterion).

6.2.2 Multiple Hypothesis Test

Following the same approach, to further emphasize the difference between a number of metrics when checking for zeros, one can consider all possible pairwise differences

between the radii. A test statistic corresponding to such a comparison is defined as

$$T_{(ij)_q} := \hat{\rho}_i^{(M)} - \hat{\rho}_j^{(M)}, \quad i = q+1\ldots, M-2, \;\; j = i+1, \ldots, M-1. \tag{6.3}$$

Obviously, the difference $T_{(ij)_q}$ will be closer to zero only when both radii $\rho_i^{(M)}$ and $\rho_j^{(M)}$ correspond to the noise subspace. A similar formulation based on the equality of the noise eigenvalues for the uniform noise case is reported in [17] and is summarized in Section 2.4.1 through the statistics (2.31).

To integrate the simultaneous test for zero of all possible pairs into the test of (4.11), every hypothesis H_q, $q = 0, \ldots, M-2$, can be formulated as the intersection between hypotheses $\mathsf{H}_{(ij)_q}$ as follows:

$$\mathsf{H}_q := \bigcap_{i,j} \mathsf{H}_{(ij)_q}, \qquad \mathsf{H}_{(ij)_q} : \rho_i^{(M)} = \rho_j^{(M)}, \quad i = q+1\ldots, M-2, \;\; j = i+1, \ldots, M-1. \tag{6.4}$$

with alternative

$$\mathsf{K}_q : \text{not } \mathsf{H}_q. \tag{6.5}$$

The MHT of H_q, $q = 0, \ldots, M-2$, in (6.4) is also conducted with the requirement of maintaining a global significance level α. In order to take into account the logical implications between the significance values $P_{(ij)_q}$, corresponding to $\mathsf{H}_{(ij)_q}$, we use Holm's SRB procedure [17]. For every combined test H_q, this procedure tests the single hypothesis $\mathsf{H}_{(ij)_q}$ corresponding to the smallest significance value, at a level $\alpha/\text{card}(\mathsf{H}_q)$, where $\text{card}(\mathsf{H}_q)$ denotes the number of hypotheses comprising H_q. The SRB allows a strong control of the global level of the test, ensuring that all single hypotheses comprising the global null are verified [17, 45]. Otherwise, if at least one of the single hypotheses $\mathsf{H}_{(ij)_q}$, $i = q+1, \ldots, M-2$, $j = i+1, \ldots, M-1$, is rejected, hypothesis H_q is systematically rejected.

6.3 Bootstrap-based Detector

Thresholding the significance values P_q (specifically P_{Σ_q}, P_{Δ_q} or $P_{(ij)_q}$) requires that the distribution of the corresponding test statistics T_q (specifically T_{Σ_q}, T_{Δ_q} or $T_{(ij)_q}$)

is known under the null. In Section 4.3, it was deduced that for the unconditional Gaussian scenario, under H_p, the Gerschgorin radii, $\hat{\rho}_m^{(M)}$, $m = p+1, \ldots, M-1$, corresponding to the candidate noise subspace, are asymptotically normal random variables. However, if either the distribution of the data is unknown, or the data sample size is too limited to permit useful asymptotic approximations, the distribution of the test statistic is consequently unknown or intractable.

Since no assumption is made on the distribution of the data, an empirical estimation of the required distributions can be thought of, using repeated experiments and Monte Carlo methods. Unfortunately, reproducing the exact experiment conditions is prohibitive in cost, if not impossible. Here, we employ the bootstrap as a means to empirically estimate the distribution of the statistics T_q from the available data sample, without *a priori* knowledge of its distribution.

6.3.1 Estimation of the Null Distribution

The bootstrap is based on the paradigm that the sample $\mathbf{x}(t)$, $t = 1, \ldots, L$, is an empirical estimate of the true distribution. By sampling with replacement from $\mathbf{x}(t)$, $t = 1, \ldots, L$, many times, bootstrap data sets are created. Computing for each of these resamples the statistics T_q, it is possible to infer their bootstrap distribution. More detailed analysis and different applications of the bootstrap for hypothesis testing can be found in [109, 121].

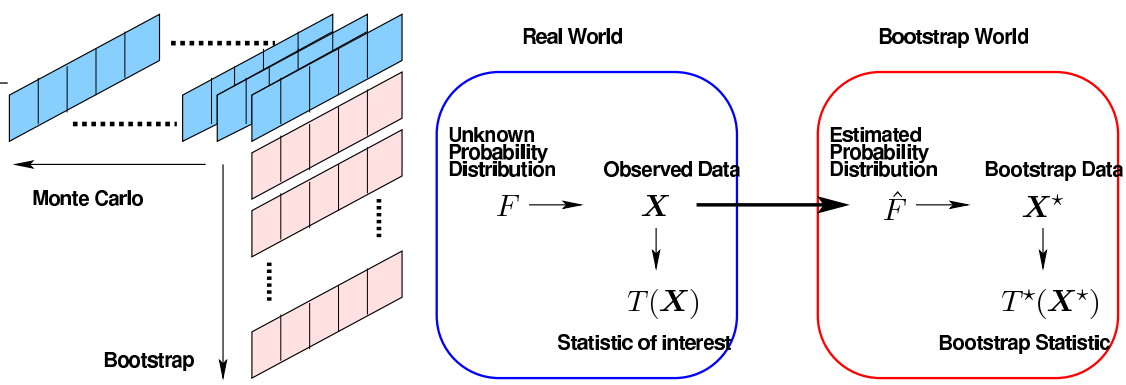

Figure 6.1: Principle of the bootstrap.

In [17] sub-sampling was investigated as a means to ensure a weak convergence of the noise eigenvalues to their asymptotic distributions. Nevertheless, it was experimentally shown that the conventional bootstrap provided a sufficiently accurate estimate. Here, this observation was also made for the estimated distribution of the smallest Gerschgorin radii, essentially under the assumption of iid data samples [78]. As a remark, when the data samples are not iid, *moving block bootstrap* variants have been investigated to create resamples [70], provided the number of blocks is much smaller than the number of data samples.

Let the b-th conventional bootstrap resample of the data be denoted $\mathbf{x}_b^\star(t)$, for $t = 1, \ldots, L$, and $b = 1, \ldots, B$. From B bootstrap resamples, the estimate of the empirical distribution of the test statistics under the null can be obtained as [31]

$$\hat{T}_q^{\mathsf{H}}(b) = T_q^\star(b) - T_q, \quad b = 1, \ldots, B \tag{6.6}$$

where T_q is the test statistic evaluated from the data $\mathbf{x}(t)$, while $T_q^\star(b)$ is the test statistic evaluated from the resample $\mathbf{x}_b^\star(t)$. The significance values for the hypothesis tests of (4.11) are given by [31]

$$P_q := \frac{1}{B} \sum_{b=1}^{B} I\left(|T_q| \leq |\hat{T}_q^{\mathsf{H}}(b)|\right) \tag{6.7}$$

with $I(\cdot)$ being the indicator function.

With a preset significance level α, starting from $q = 0$, if $P_q \geq \alpha$ then H_q is accepted, otherwise set $q \leftarrow q + 1$ and repeat the test.

6.3.2 Studentizing and Bias Correction

Deviation from asymptotic conditions usually results in an exceeded level of the test. This excess is, in many cases, inversely proportional to the data sample size [70]. To better control the global significance level, it is desirable to reduce the distribution's dependence on unknown parameters. To this end, one resorts to studentizing the test statistic by empirically estimating its sample variance [31, 121]. This estimation is also conducted using the bootstrap through a nested routine [121].

In our case though, under the iid assumption of the data, experimental results did not justify the computational cost associated with the nested bootstrap procedure[1],

[1]The shape of the distribution of the studentized test statistics is often different from that of

thus the Gerschgorin radii are not studentized. It is of interest to mention that for the eigenvalue-based case, the bootstrap was also applied without studentizing the sample eigenvalues [17].

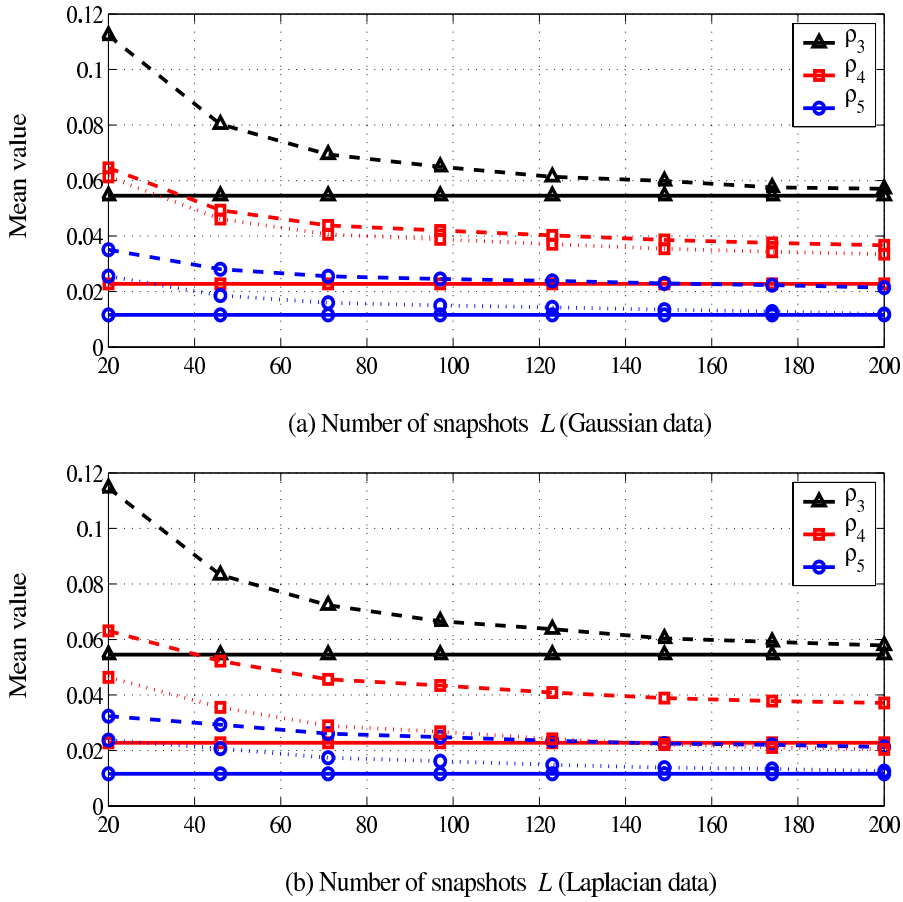

Figure 6.2: Mean value of the noise radii computed from $\mathbf{R}$ $(-)$, and from $\hat{\mathbf{R}}$, first without bias correction $(--)$, and then using a bootstrap bias estimator $(\ldots)$.

The significance level of the test can also be exceeded if the test statistic T_q is not estimated correctly. The test of hypotheses in (4.11) is sensitive to any statistically significant difference between the sample radii $\hat{\rho}_m^{(M)}$, $m = 1\ldots, M-1$. If these estimated radii contain an error, or bias, the resulting test statistics become inaccurate. It was shown in [17] that the sample eigenvalues corresponding to the

the original statistics. However, for relatively large data samples, the resulting improvement in the detectors' performance is imperceptible. On the other hand, for small samples, while the detection rate improves, it remains bounded at a relatively low level.

noise subspace are asymptotically biased. Several techniques were then proposed to estimate the bias and alleviate its effect, including robust estimators based on the bootstrap as well as the jackknife [31]. In our case, the sample radii $\hat{\rho}_m^{(M)}$, $m = p+1\ldots, M-1$, corresponding to the candidate noise subspace are very close to zero. It is expected therefore that estimating such small values may induce some bias. A bias in the estimated radii can be geometrically interpreted as an excess in the size of the disks corresponding to the disjoint signal and noise subspaces [111, 113]. If the bias is significant, the two sets of disks can expand to the point of intersecting, thus preventing an accurate separation of the two subspaces.

Figure 6.2 illustrates the mean of the sample radii corresponding to the candidate noise subspace as a function of the number of snapshots L, computed from both $\mathbf{R}$ and $\hat{\mathbf{R}}$. The examples employ an array of $M = 6$ sensors, with perturbed noise powers given by the vector $\mathbf{q} = [1.0, 0.5, 2.8, 1.7, 4.1, 5.0]^T$. The true number of sources is $p = 2$, and the average SNR is set to 10 dB. In example (a), the data samples are Gaussian, whereas in (b) they are Laplacian. The mean of the radii is averaged over 1000 Monte Carlo runs. In both examples, a bootstrap-based bias estimator is used to improve the quality of the estimated sample radii.

Observe that the individual bias of the radii is of the same order of magnitude. This suggests that for a test statistic based on the difference of two metrics, and more particularly the difference of two radii such as $T_{(ij)_q}$ in (6.3), the effect of this bias on the test statistic is significantly reduced by the algebraic structure of the latter. Indeed, from preliminary experimental results, the effect of the bias in the estimated radii on the detectors' performance is observed to be negligible, for different values of SNR. In the scenario considered here, including a nested bootstrap or jackknife routine in the evaluation of the test statistics does not appear to be necessary.

6.3.3 Bootstrap detection procedure

Direct application of the bootstrap to source detection using the SHT of (4.11) can be summarized as follows:

1. Let $q = 0$.
2. For $b = 1, \ldots, B$, using (6.6) and (6.1), calculate $\hat{T}^{\mathsf{H}}_{\Sigma_q}(b)$ (respectively $\hat{T}^{\mathsf{H}}_{\Delta_q}(b)$, using (6.6) and (6.2)).
3. Using (6.7), calculate the significance value P_q, with T_q replaced by T_{Σ_q} (respectively T_{Δ_q}).
4. If $P_q \geq \alpha$ is satisfied, then accept H_q, set $\hat{p} = q$ and stop the test.
5. If $q < M - 2$, set $q \leftarrow q + 1$ and go to step 2.
6. Set $\hat{p} = M - 2$ and stop the test.

The incorporation of the MHT of (6.4) in the bootstrap-based detection procedure yields the following variation:

1. Let $q = 0$.
2. For $i = q + 1, \ldots, M - 2$ and $j = i + 1, \ldots, M - 1$, calculate $\hat{T}^{\mathsf{H}}_{(ij)_q}(b)$ using (6.6) and (6.3), for $b = 1, \ldots, B$.
3. For $i = q + 1, \ldots, M - 2$ and $j = i + 1, \ldots, M - 1$, calculate the significance values $P_{(ij)_q}$ using (6.7), with T_q replaced by $T_{(ij)_q}$.
4. If $\min_{i,j} \left\{P_{(ij)_q}\right\} \geq \alpha/h$ is satisfied, where $h = M - 2 - q$ is the total number of individual hypotheses $\mathsf{H}_{(ij)_q}$ comprising H_q, then accept H_q, set $\hat{p} = q$ and stop the test.
5. If $q < M - 2$, set $q \leftarrow q + 1$ and go to step 2.
6. Set $\hat{p} = M - 2$ and stop the test.

Using the conventional bootstrap on the described test statistics, the performance of the resulting source detectors as a function of the data parameters, is illustrated in the following section.

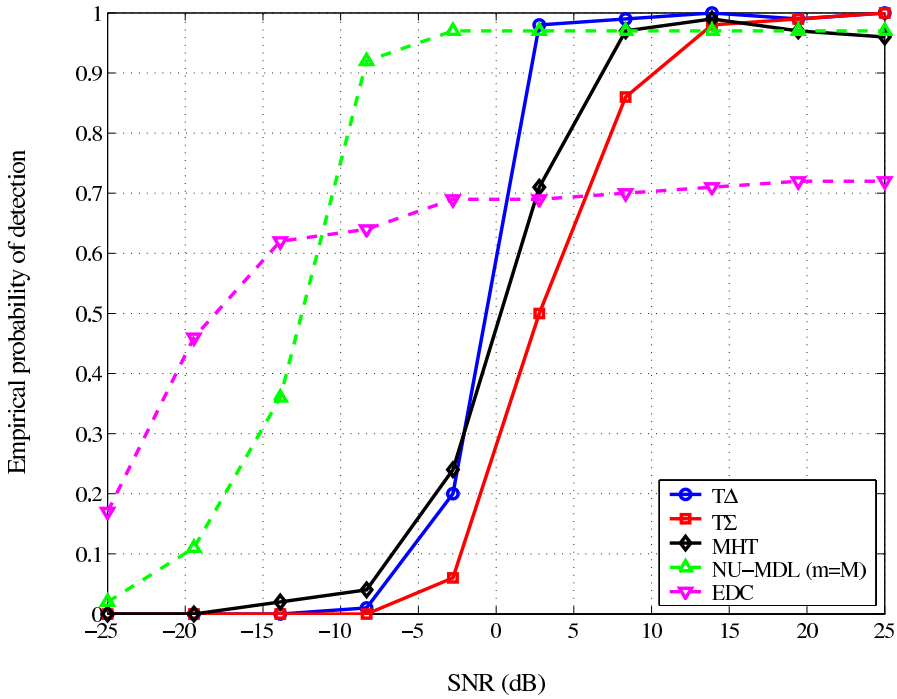

Figure 6.3: Performance of the detectors vs SNR (Gaussian sources and noise).

6.4 Numerical Examples

In what follows, the detectors corresponding to the test statistics (6.1), (6.2) and (6.3), are denoted T_Σ, T_Δ and MHT, respectively, and are compared to the NU-MDL of Chapter 5, and the EDC of [116] (see Section 2.4.2). The first four detectors are based on the transformed sample Gerschgorin radii. Because of their robustness to noise-power perturbation, we refer to them as nonuniform detectors as indicated in the previous chapter. In the simulations, the discarded sensor is of index M, i.e., the last array sensor. The EDC is designed for an arbitrary noise covariance matrix. Its use is best known for spatially correlated noise fields.

Although a specification of the functional form of the array manifold is not necessary for the detectors to work, the examples are restricted to a ULA with $M = 6$ sensors. The true number of sources is $p = 2$. The employed noise powers are given by the vector $\mathbf{q} = [1.0, 0.5, 2.8, 1.7, 4.1, 5.0]^T$, and WNPR $= 10$. The performance is illustrated through the empirical probability of correct detection resulting from 500 Monte Carlo runs.

Figures 6.3, 6.4, 6.5 and 6.6 describe a comparison of performance with respect

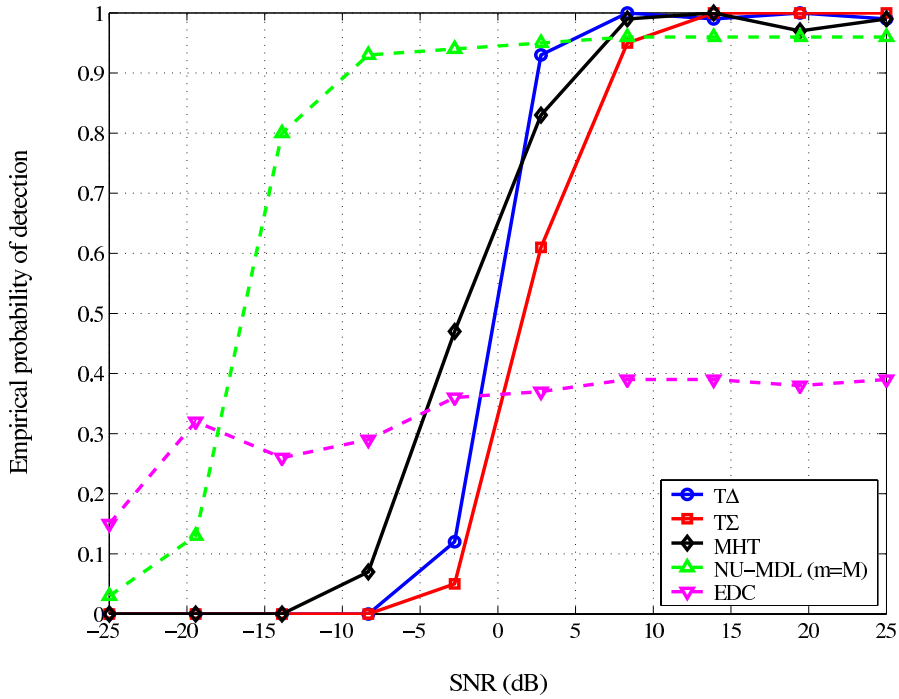

Figure 6.4: Performance of the detectors vs SNR (Laplacian sources, Gaussian noise).

to the SNR, for different combinations of source and noise distributions. The fixed parameters are the number of snapshots $L = 100$ and the DOAs $\boldsymbol{\theta} = [0°, 25°]^T$, necessary for the ULA parameterization. The SNR varies from -25 dB to 25 dB. The number of bootstrap resamples is the same for T_Σ, T_Δ and MHT, and is fixed at $B = 500$. The significance level of the tests is set to $\alpha = 2\%$.

- In the example of Figure 6.3, the sources and the noise are both Gaussian. In this scenario, all the nonuniform detectors outperform the EDC. The latter consistently overmodels the data, as it assumes that the main diagonal of the noise covariance matrix is equipotent. When the SNR is low, the poor clustering of the noise eigenvalues reduces the detection rate of EDC [60, 117]. When the SNR increases, this mismodeling still bounds the detection rate at a useless level. Note that because it is specifically derived for the unconditional Gaussian case, the NU-MDL provides better results than the hypothesis tests for the smaller values of SNR. In addition, the bootstrap detectors are different from the NU-MDL information criterion, which is known for its consistency properties [80, 106]. Observe the relative improvement in performance of T_Δ

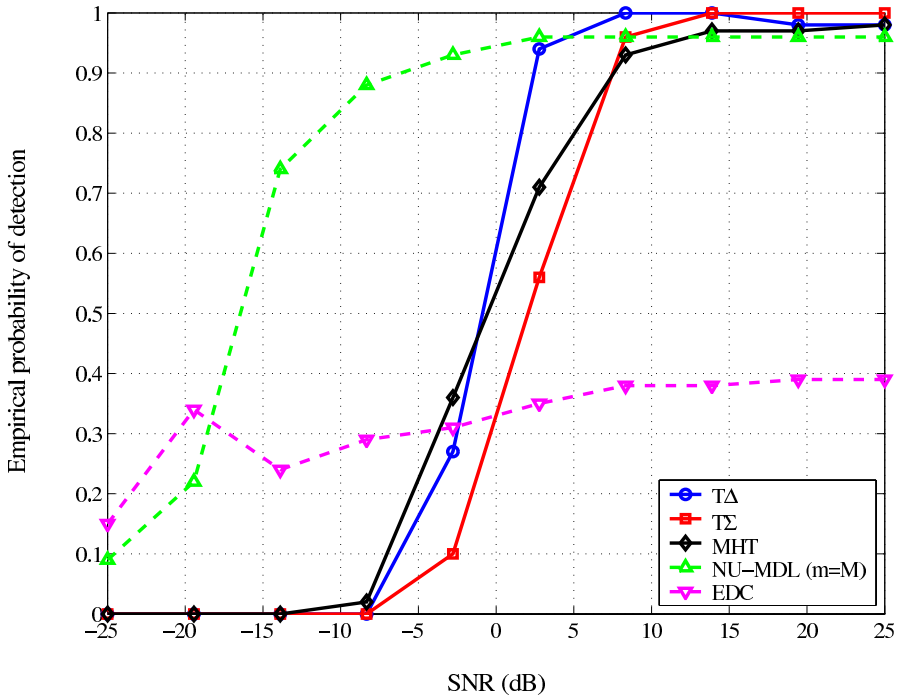

Figure 6.5: Performance of the detectors vs SNR (Gaussian sources, Laplacian noise).

as opposed to T_Σ. This can be explained by the fact that when testing for the null, the structure of T_Δ (difference between two metrics) allows it to be closer to zero "more often" than T_Σ. For the same reason, the MHT performs slightly better than T_Σ. It should be remembered however, that the MHT is more demanding in terms of bootstrap resamples. Indeed, the MHT requires an accurate estimation of the null distributions for a larger number of single statistics.

- In the examples of Figures 6.4 and 6.5, the sources and the noise are distributed differently as Gaussian or Laplacian, and vice versa. The global performance of all the previous detectors is similar to the fully Gaussian case. First, EDC overestimates the number of sources due to the same mismodeling problem. Second, NU-MDL similarly to all variants of MDL in uniform noise, is less sensitive to the distribution of the data, provided that the spatial resolution of the sources and the sample size are sufficiently high. Equivalently, NU-MDL is more sensitive to the correlation among the sources, and thus to the rank condition on the covariance matrix of the data. The settings of the examples,

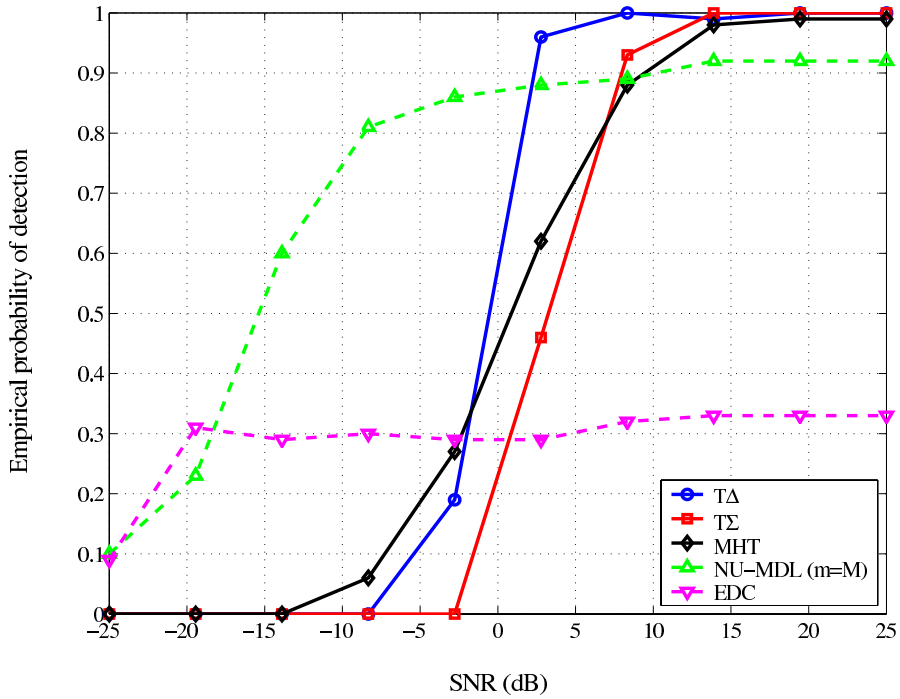

Figure 6.6: Performance of the detectors vs SNR (Laplacian sources and noise).

i.e., spatially well separated sources, do not seriously alter a correct detection. Finally, the bootstrap detectors are relatively robust and insensitive to the distribution of the data. This robustness is indicated by a performance similar to the fully Gaussian case. Moreover, as the SNR increases, the bootstrap detectors consistently outperform NU-MDL.

- Figure 6.6 shows the detectors' performance when both the sources and the noise are Laplacian. Overall, the same observations can be made as for the previous examples, noting by the way a further decline in the performance of NU-MDL as compared to the bootstrap detectors, which do not require any *a priori* knowledge of the distribution of the data.

- Figure 6.7 illustrates the performance with respect to the number of snapshots L which varies from 5 to 145. The fixed parameters are SNR=10 dB, WNPR=10, and the DOAs $\boldsymbol{\theta} = [0°, 45°]^T$. The number of bootstrap resamples is $B = 500$ and the level of the test is fixed at $\alpha = 2\%$. Here the data samples are Gaussian. As expected, the performance of the nonuniform detectors

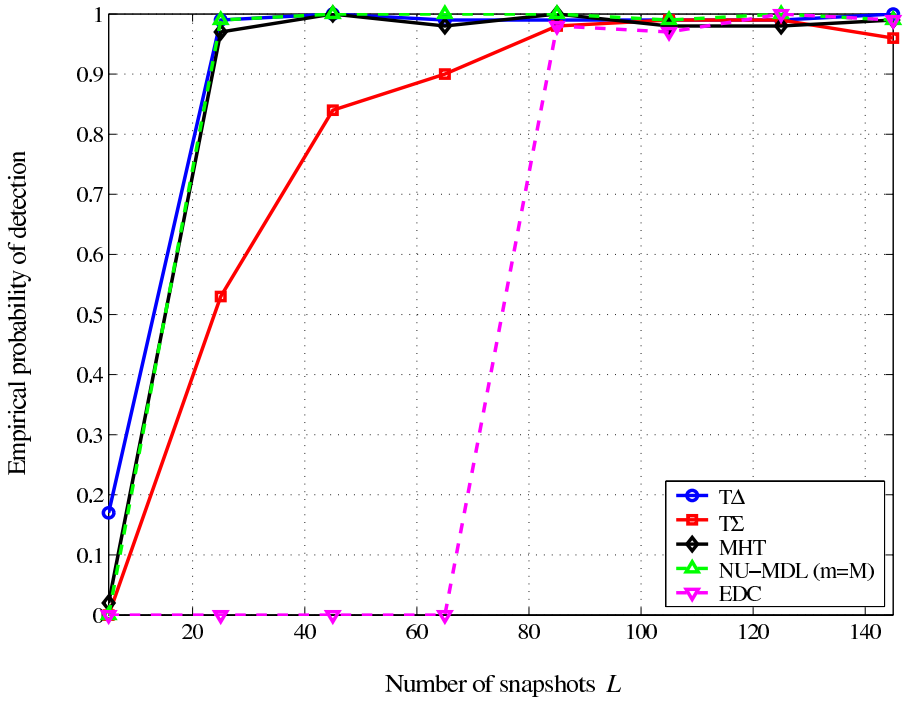

Figure 6.7: Performance of the detectors vs the sample size L (Gaussian sources and noise).

improves with an increasing L, as the discriminating capability of the transformed sample Gerschgorin radii on which the detectors are based, sharpens. The relative performance between the three bootstrap detectors is the same as for the previous settings. Also, as L increases, EDC abruptly ceases the overestimation of the source number. The increasing range of the sample size enhances the effect of the SNR and thus, the threshold between the signal and noise eigenvalues becomes more clear. This results in a performance of EDC, which is similar to other consistent information criteria like MDL. It is worth mentioning that EDC exploits a larger spatial diversity due to a larger number of sensors, M, as compared to the nonuniform detectors $(M-1)$.

- Figure 6.8 illustrates the performance with respect to the angular resolution $\Delta\theta$. The first DOA is fixed at $\theta_1 = 0°$ whereas the second one, θ_2, varies from $1°$ to $10°$, while SNR=10 dB, WNPR= 10, and $L = 100$. Here also, $B = 500$, $\alpha = 2\%$, and the data samples are Gaussian. Note that the small angular separation between the two sources has an effect on the rank condition of the

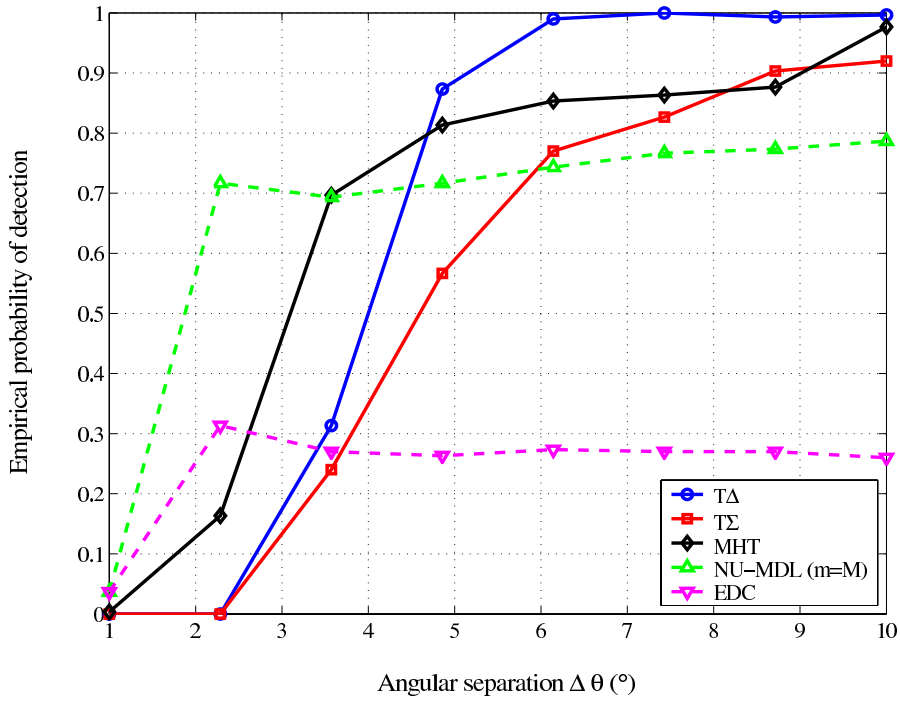

Figure 6.8: Performance of the detectors vs the source spatial resolution $\Delta\theta$ (Gaussian sources and noise).

data covariance matrix, which is accentuated with a discarded array element (for the nonuniform detectors). As the rank drops because of too closely separated sources, the detection rate falls considerably. The relative performance between the bootstrap detectors is similar to the previous examples, while EDC undergoes data mismodeling.

- The effect of noise-power perturbation is illustrated in Figure 6.9. In this example, the WNPR varies from 1 to 10, and the noise powers are generated from a uniform distribution over $[1, \text{WNPR}]$ and normalized by their mean. Here SNR=10 dB, $L = 80$ and $\boldsymbol{\theta} = [0°, 15°]^T$. The number of resamples $B = 500$ and $\alpha = 2\%$. The data are Gaussian. Note that when the WNPR approaches 1, the scenario is close to the ideal uniform noise case. As the noise nonuniformity increases, the performance of the detectors decreases. Observe that NU-MDL is less sensitive to the values of WNPR than the bootstrap detectors (for the settings of the example). As the WNPR increases, the contribution of the largest noise power is such that it contaminates the signal

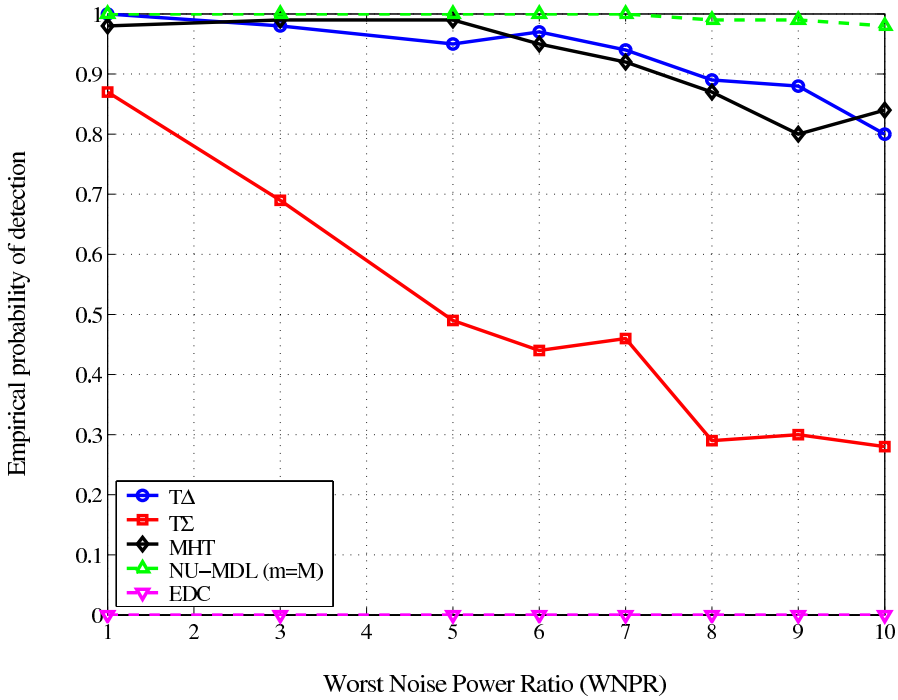

Figure 6.9: Performance of the detectors vs WNPR (Gaussian sources and noise).

subspace, therefore increasing its dimension and leading to an erroneous detection. Equivalently, estimation of the correct significance values for the test statistics becomes less accurate as the WNPR increases. The NU-MDL copes with this problem through a fixed penalty function (Gaussian case). Observe also that EDC systematically fails to correctly detect the number of sources due to a strong mismodeling of the noise. Because of the threshold effect on the estimation quality of the signal subspace, when the WNPR is further increased, the detectors' performance will degrade significantly, as the additive noise will induce a strong model mismatch preventing subspace separation.

- In Figure 6.10, we illustrate the ability of the bootstrap detectors to correctly assess that no sources are present. In the example, the data samples are limited to the Gaussian noise. The performance is shown in terms of the variation of the sample size L and the WNPR. Here, like for the previous examples, the significance level of the tests is set to $\alpha = 2\%$, and the number of bootstrap resamples is $B = 500$. For the settings of the example, Figure 6.10 (a) shows a strongly controlled significance level, independently of the sample size. This

indicates that the bootstrap permits a sufficiently precise estimation of the null distribution of the test statistics. It is worth mentioning that the quality of the estimation of the significance values P_q partially depends on the number of bootstrap resamples B. The number of resamples is in turn affected by the data size, i.e., the combination of the array size M, and the number of collected snapshots L. Hence, if B is not chosen appropriately, the global significance level of the test cannot be maintained.

- The effect of noise nonuniformity through the variation of the WNPR appears to be stronger, as illustrated by Figure 6.10 (b). However, for a sufficiently high B, the significance level is maintained. Note that in both cases, MHT is more demanding that T_Σ and T_Δ and normally requires a higher number of resamples B, due to the logical implications between the larger number of inherent single hypotheses. Yet, increasing B beyond a certain level does not necessarily improve the performance of the detectors, while inducing a much higher computational cost. For reference, it was shown in [43] that for univariate data for example, for a desired confidence interval $1 - \alpha$, it is

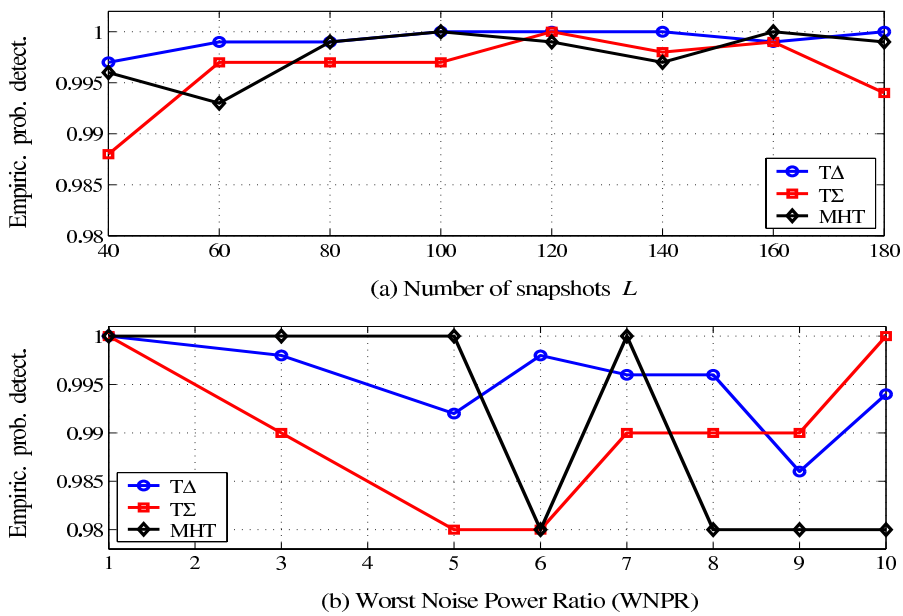

Figure 6.10: Probability of correctly detecting zero sources vs (a): L, and (b): WNPR (Gaussian sources and noise).

sufficient to choose an integer number of resamples B, such that $B/(B+1) = 1-\alpha$. This property[2] follows from the fact that a test statistic resulting from B resamples divides the real line into $B+1$ parts. Considering the multivariate case that we are interested in, the number of resamples B must be adjusted accordingly, as a smaller B can result in longer confidence intervals, and consequently, the confidence level of the test will be exceeded. For a given B, the effect of the total collected data is illustrated in the following example. As indicated previously, if the WNPR increases, the noise sources with the largest powers will erroneously be assimilated to dominating signal sources, and eventually, the global significance level of the test will not be maintained.

- The effect of the data size, through the combination of the sample size L and the number of sensors M, on the performance of the bootstrap detectors is shown in Figure 6.11. For simplicity, only the SHT based on T_Σ and T_Δ are illustrated. The fixed parameters are the DOAs $\boldsymbol{\theta} = [25°, 35°]^T$, SNR=10 dB, and WNPR=5. The number of snapshots is set to vary from 20 to 120. Here, $B = 100$ and $\alpha = 2\%$. The data samples are Gaussian. The performance is evaluated for different values of M. In general, as expected, the performance improves as L increases, as previously indicated in Figure 6.7. Note however that for a fixed significance level α, even asymptotically, the hypothesis tests stay below 100% detection rate. As the number of sensors M increases (increasing degrees of freedom of the data vector $\mathbf{x}(t)$), the bootstrap requires a larger "minimal" number of snapshots to faithfully estimate the empirical distribution of the different test statistics under the null. For instance, with T_Σ, when $L \leq 70$, the detection rate is lower with $M = 7$ than with $M = 4$. On the other hand, above a certain value ($L = 80$), the bootstrap gives better results, and the joint effect of increasing M and L improves the detection rate significantly. Because of its structure, T_Δ shows a better performance than T_Σ.

[2]This theoretical result is drawn for pivotal statistics. It is valid for a studentized test statistic. In the experiments however, for a given number of resamples B, the difference in detection performance with or without studentizing remains negligible.

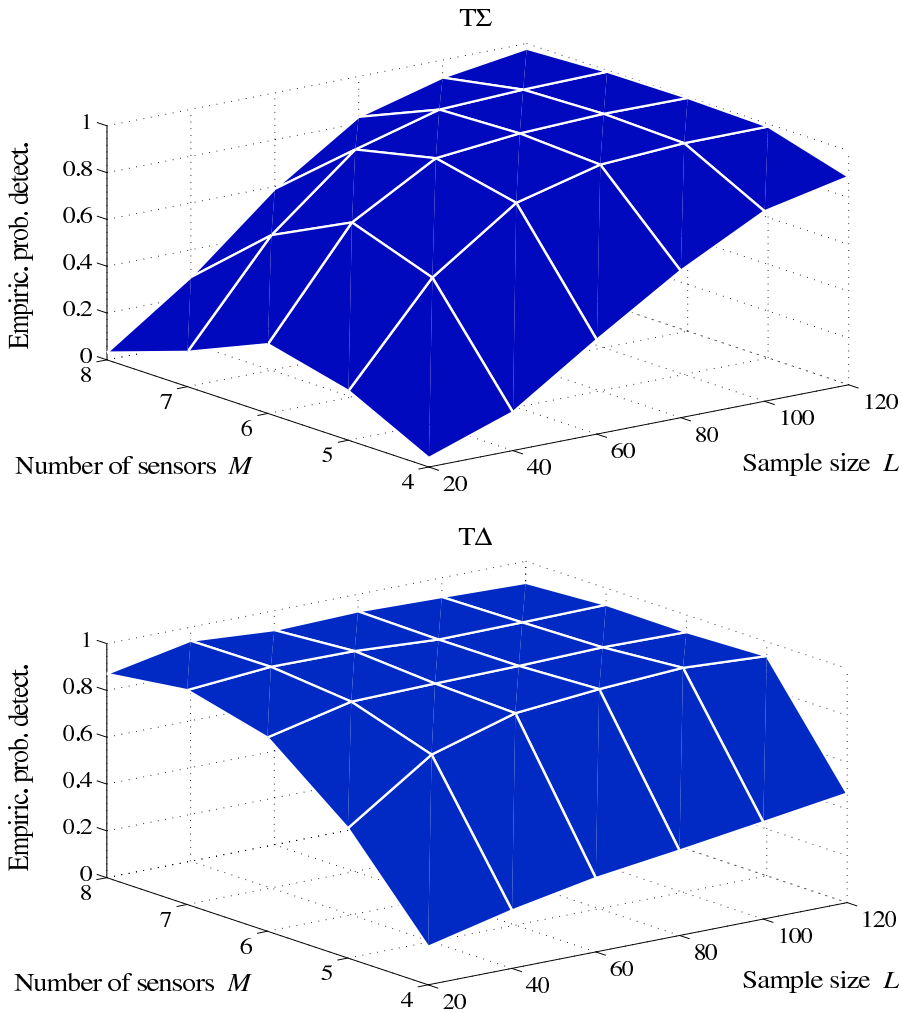

Figure 6.11: Joint effect of the number of sensors M and sample size L on the global performance of the detectors (Gaussian sources and noise).

6.5 Application to Power Plant Process Monitoring

In a nuclear power plant based on a pressurized water reactor (PWR), ordinary light water is used as coolant. The water in the primary circuit of a PWR is normally pressurized but is not directly used to drive a turbine. Instead, steam generators are employed to transfer heat to a secondary circuit, where water reaches boiling state, producing steam for electricity generation. The water in the cooling circuit is employed to condensate the steam in the secondary circuit (Figure 6.12).

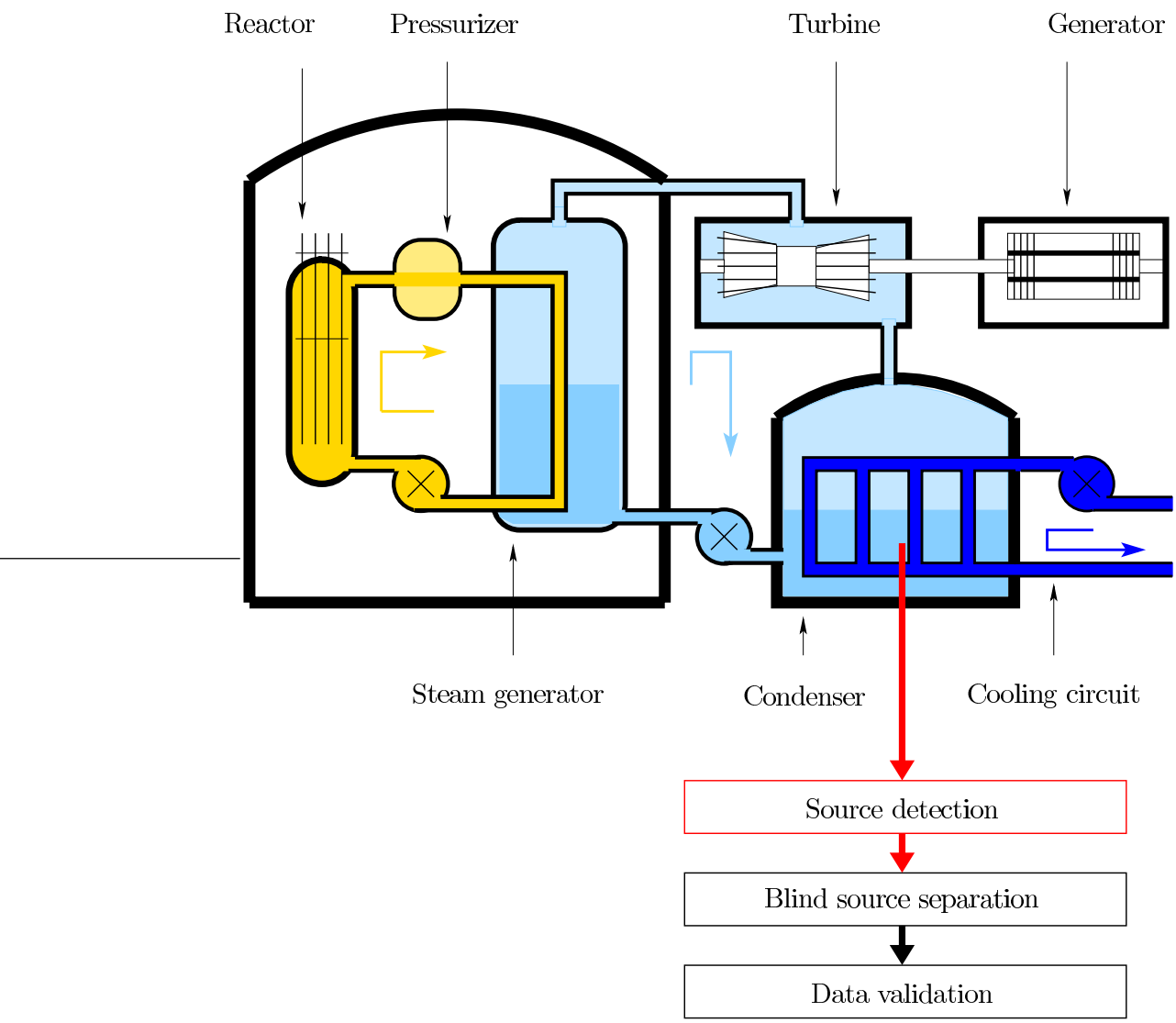

Figure 6.12: Diagram of a PWR power plant and application of source detection.

Electricité de France (EDF) is the European leader of electricity production. About 75% of its total production is of nuclear origin and most EDF's nuclear plants are based on PWRs. EDF regularly conducts non-destructive assessments of different parts of its plants in order to monitor and improve its industrial processes. One such assessment involves the evaluation of several physical variables and their effect on the behavior of the cooling process in a typical PWR. As part of this

assessment, many measurements are conducted around the cooling circuit, without direct access to the coolant water. The measurements are a combination of the effects of a number of physical quantities. Attempting to identify the individual influence of each major physical quatity from the data mixture is one step of the analysis. When the data are appropriately modeld, the problem translates to blind source separation, or more generally independent component analysis (ICA) [10, 26, 48].

Like for most array signal processing problems involving the model of (2.6), source detection is prerequisite for blind source separation as indicated in Chapters 1 and 2. In addition, as the distribution of the measured data is unknown, the use of a robust source detector becomes well motivated. Therefore, in what follows, we propose to use our bootstrap detectors on a set of real measurements from the cooling circuit of a nuclear reactor[3].

It is important to mention that the purpose of our application is the validation of the proposed detection approach on real data. Consequently, the physical interpretation of the whole assessment process is beyond the scope of this work.

6.5.1 Data model

Although the most precise modeling of the cooling circuit (as most processes in a nuclear plant) introduces convolutive mixtures, we approximate the behavior of our system by an instantaneous-mixture linear model following (2.6), as a trade-off between precision and computational complexity [30]. The size of the measured data sample is therefore defined to partially corroborate this assumption, i.e., a large M and a large number of collected snapshots, L.

We have p latent signals measured simultaneously by a set of sensors yielding M outputs. The source signals correspond to the physical quantities to be monitored, which are mainly the *temperature*, *pressure* and *flow* of the water. However, the operator is initially unable to observe a single characteristic signal for each quantity, but receives a mixture of measurements emanating from different and arbitrary locations in the process. The operator predicts 2 to 3 independent major sources in the system, given that the water temperature and pressure can be strongly correlated. Yet, for our blind analysis, the number of sources p is assumed unknown and is to be

[3]The tests and data formatting of this section were conducted within the scope of [4].

estimated and validated. The number of sensors M is much larger than the number of sources p. The modulation and conditioning of the sources are conducted in a NB framework. The sources in the process are known to be mostly non-Gaussian [4].

Depending on the combination of measured physical quantities, the technology of the particular sensors is variable. In addition, they are not necessarily located in a homogeneous pattern around the process, so that the most representative information is collected. Moreover, given the relatively large values of M and L, it appears that a uniform calibration of all the sensors over the total duration of the measurement is very difficult. Although the measurements are pre-processed to level the power of the noise, some residual perturbation may appear from one sensor to another.

6.5.2 Data Sample and Pre-processing

The data sample of interest is collected and formatted by EDF, from its Nogent-sur-Seine site. The measurements correspond to the operational phase of the PWR, as opposed to the trial phase, and the collected data sample is of size $M = 142$ by $L = 3 \times 10^5$. The data are sampled above the Nyquist rate, at a period of 1 min[4]. The sensor outputs are correlated and have the same general shape over time. As an example, one such output wave-form is illustrated in Fig. 6.13.

Two different operation modes of the power plant can be easily identified. First, between $t = 0$ and $t = 10^5$, a *steady-state* mode, where the measurements are mostly stationary around their nominal value, and second, between $t = 1.5 \times 10^5$ and $t = 3 \times 10^5$, a *network-steered* mode, where the measurements are non-stationary and strongly vary with the load on the network. In this analysis, we are interested only in the steady-state operation, and thus limit the sample size to $L = 10^5$ (10^5 minutes).

Note that other segments are visible. These are also discarded from our analysis. They correspond to irrelevant situations, such as an interrupted measurement due to the opening of the reactor's heart, or constant measured values due to a likely dysfunction of the sensors.

[4]The typical time constant of such a process is about 7 min [30]. The employed sample rate therefore emphasizes a linear quasi-static model which can be accurately described by (2.6).

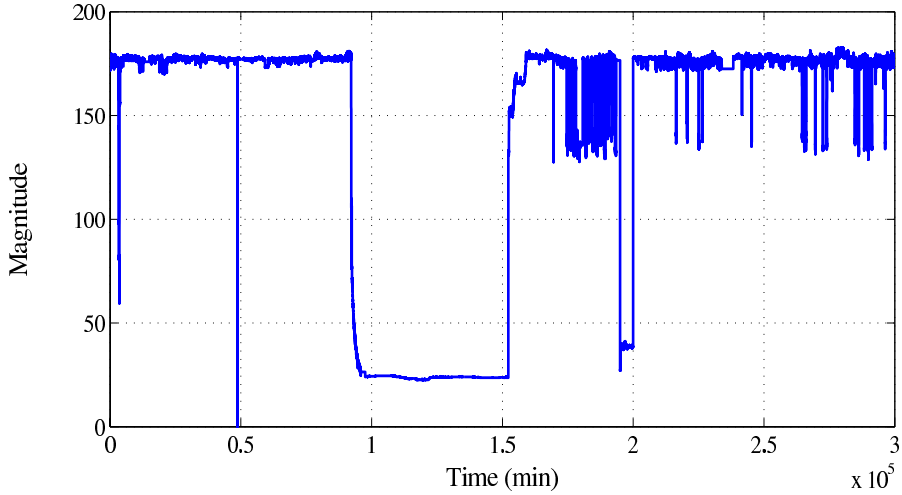

Figure 6.13: Example of an output signal wave-form.

The collected data of interest are further cleaned from short constant segments and outlier points, and then tested for stationarity. Depending on the application, the data can also be whitened and centered (see [4] for details).

6.5.3 Experimental Results

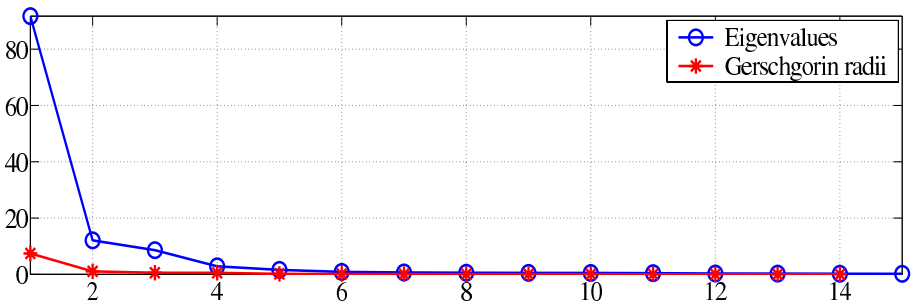

Figure 6.14: Sample eigenvalues and Gerschgorin radii.

The number of sources is estimated using MDL, NU-MDL, the bootstrap SHT, both with T_Σ and T_Δ of (6.1) and (6.2), and the MHT. In our example, the number of output signals used for this estimation is limited to 15 to reduce the computational load (recall that the complexity associated with the eigen-decomposition of $\hat{\mathbf{R}}$ is of the order $O(M^3)$). The test significance level is $\alpha = 5\%$, whereas the number of

	L ($\times$1000)								
	6	9	12	15	18	21	24	27	31
MDL	15	15	15	15	15	15	15	15	15
NU-MDL	4	4	4	4	7	7	7	7	7
T_Σ	2	2	2	2	3	3	3	3	3
T_Δ	2	2	2	3	3	3	3	3	3
MHT	2	2	2	2	3	3	3	3	3

Table 6.1: Estimated number of sources.

bootstrap resamples is set to $B = 3000$. The number of resamples B is chosen to take into account the dimension of the data and the total number of hypotheses in the MHT.

Fig. 6.14 shows the ordered eigenvalues and Gerschgorin radii. Note that we have a strongly dominating eigenvalue, which translates to a large Gerschgorin radius, although the transformation reduces the magnitude dispersion. Hence, in our relatively high SNR case, we expect the detectors to yield a small number of sources. It should be kept in mind however, that the dominating source can possibly mask the smaller sources. At the same time, because of the poor clustering of the smallest eigenvalues, the information criteria can typically overmodel the system [60].

Application of the detectors (25 independent trials) results in the estimates shown in Table 6.1. The accuracy of the detection improves with an increasing sample size L. The Gerschgorin-based criteria outperform the conventional MDL. Despite the high SNR, the main reason for the failure of MDL is the non-Gaussianity of the embedded sources, which are thus strongly mismodeled by the informative part of MDL. The non-Gaussianity of the sources affects the NU-MDL in the same negative way. The bootstrap detectors on the other hand are more suitable for unknown distributions of the data. In addition, determination of the signal and noise subspace dimensions is carried-out through the ordered Gerschgorin radii rather than the ordered eigenvalues, thus reducing the possibility of masking smaller sources. The result is more sensible, irrespective of the sample size L. Despite the increased computational complexity, this result validates the relevance of the bootstrap in practice, especially for off-line analysis.

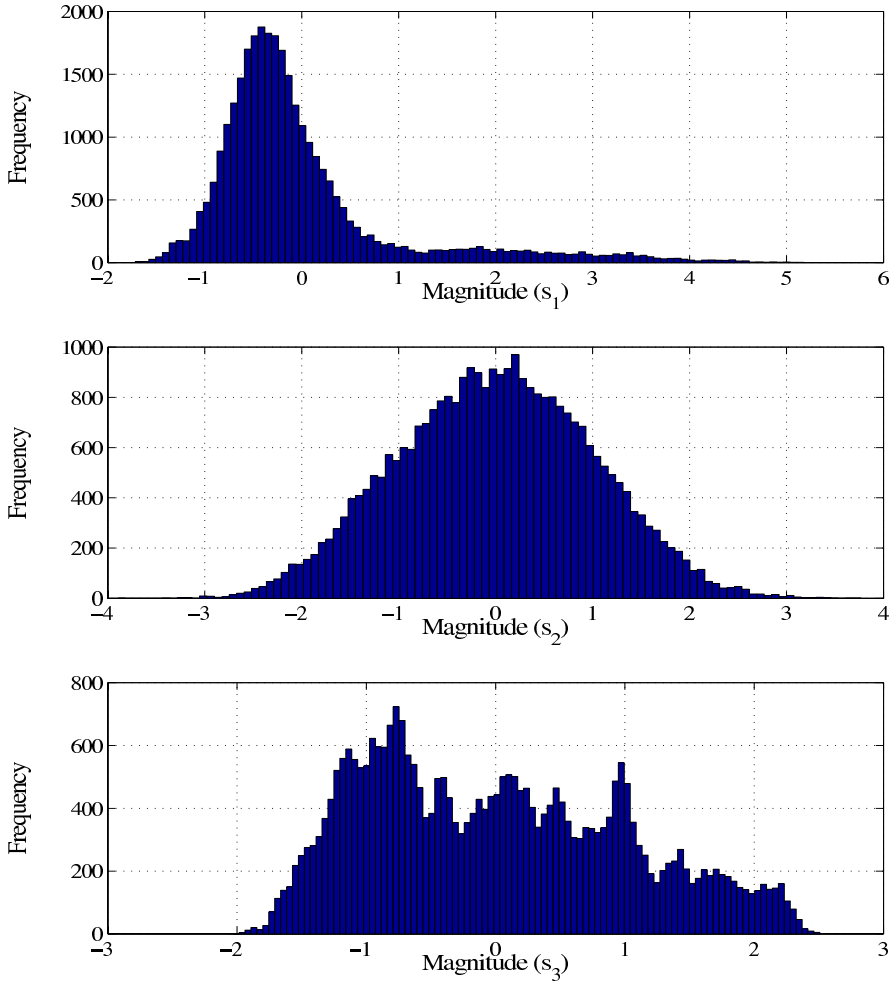

Figure 6.15: Histogram of the magnitudes of the estimated sources.

After application of the ICA algorithms to the data mixture, the obtained separated sources are mostly non-Gaussian[5], in conformity with the initial assumptions. The first source shows small fluctuations and some impulsive bursts. It is somehow heavytailed. In the steady-state operation mode, it is likely to correspond to the variation of the temperature/pressure of the cooling water. The second source is nearly Gaussian. It is likely to correspond to an internally-generated perturbation, which is rather independent of the variation of the physical quantities of interest. The third source is multi-modal and shows progressive variations, which likely cor-

[5]Source separation was applied to the data mixture using three different ICA methods, i.e., SOBI [10], JADE [26], and FastICA [48], with quasi-identical results [4].

respond to the flow of the cooling water.

6.6 Conclusion

A hypothesis test has been proposed for source detection when no *a priori* knowledge is available, neither on the distribution of the data, nor on the order of the eigenvalues of the sample covariance matrix.

When the assumptions on the distribution of the data are minimal, no information is available on the null distribution of the proposed test statistics. This distribution can be inferred empirically using the nonparametric bootstrap.

Simulation results show that the bootstrap detectors are relatively robust to the particular distribution of the data, and in cases where Gaussianity or asymptotic conditions are not verified, they outperform standard information criteria. Nevertheless, a good performance in spatially perturbed noise requires a more restrictive identifiability condition on the number of array sensors.

Also, a linear instantaneous-mixture has been successfully employed to model the cooling process in one of EDF's power plants. The bootstrap, through Gerschgorin-based detection criteria, is applied to the measured data and yields conclusive results. These experimental results show the power of the bootstrap-based detector and its advantage over other methods in a real case where the distribution of the data samples in unknown.

Chapter 7

Direction of Arrival Estimation in Spatially Nonuniform Noise

7.1 Introduction

Most standard DOA estimation techniques, as seen in Section 2.5, are based on the assumption that the additive noise is spatially uniform. When this assumption is not verified, the above techniques cannot be applied directly, due to the severe model mismatch. If the individual noise powers are known in advance, then it is possible to scale the data and transform the model to the standard uniform case. However, these noise powers are often unknown in practice.

In what follows, we summarize the few existing approaches that specifically take noise nonuniformity into account, and discuss their properties. We also suggest alternatives such as a modified MUSIC estimator, and a stochastic ML approach for the particular case of a ULA configuration. The first method is based on the approximated subspace separation principle discussed in the previous chapters, while the proposed SML variant attempts to estimate the noise covariance matrix as a first step. The estimate of noise is then parameterized in terms of the unknown signal attributes and is used in the LL function.

7.2 Existing DOA Estimators for Nonuniform Noise

In this section, we summarize existing DOA estimation algorithms which are specifically dedicated to the case of spatially nonuniform noise. These methods are mostly parametric methods related to the ML approach.

7.2.1 Nonuniform Maximum Likelihood Estimator (MLDOA)

ML methods play an important role in array signal processing for their excellent asymptotic and threshold performances [76, 94, 103]. However, total ML solutions are generally computationally very demanding [38]. It is possible to partially reduce this computational load if prior information about the structure of the covariance matrix of the data is available. In addition to a less complicated implementation, this information leads to a significant improvement of the ML estimator performance [76, 103].

Standard uniform ML estimators, as well as ML techniques for colored noise, ignore important knowledge on the noise and cannot be expected to yield good results when applied in nonuniform noise.

Many noise models have been addressed where the spectral distribution of the noise has been considered [41, 42, 76], leading to a linear parameterization of the noise covariance matrix. Despite this parameterization, the colored noise assumption remains very general. In many real applications, the general assumption of colored noise can be simplified to a spatially nonuniform white noise.

In [103], a ML estimator of DOAs in block-correlated noise has been proposed, with a variant called MLDOA, for the special case of nonuniform white noise [76].

Under the assumption that the signal waveforms are deterministic unknowns, the LL function of the data can be written as

$$\begin{aligned}\mathcal{L}(\boldsymbol{\eta}_p) &= -L\ln\left(\det(\mathbf{Q})\right) - \sum_{t=1}^{L}\left[\mathbf{x}(t)-\mathbf{A}(\boldsymbol{\theta})\mathbf{s}(t)\right]^H\mathbf{Q}^{-1}\left[\mathbf{x}(t)-\mathbf{A}(\boldsymbol{\theta})\mathbf{s}(t)\right] \\ &= -L\sum_{m=q}^{M}\ln(\sigma_m^2) - \sum_{t=1}^{L}\left\|\tilde{\mathbf{x}}(t)-\tilde{\mathbf{A}}(\boldsymbol{\theta})\mathbf{s}(t)\right\|^2 \qquad (7.1)\end{aligned}$$

where

$$\tilde{\mathbf{x}}(t) := \mathbf{Q}^{-1/2}\mathbf{x}(t) \tag{7.2}$$

$$\tilde{\mathbf{A}}(\boldsymbol{\theta}) := \mathbf{Q}^{-1/2}\mathbf{A}(\boldsymbol{\theta}) \tag{7.3}$$

for $t = 1, \ldots, L$, and $\mathbf{Q}$ defined by (2.14).

From (7.1), keeping the irrelevant parameters fixed and by nulling the derivative of the LL function with respect to the unknown matrix of signal samples, its ML estimate is obtained as follows:

$$\hat{\mathbf{S}} := \tilde{\mathbf{A}}^{\#}(\boldsymbol{\theta})\tilde{\mathbf{X}} \tag{7.4}$$

where

$$\tilde{\mathbf{X}} = \mathbf{Q}^{-1/2}\mathbf{X} \tag{7.5}$$

and matrix $\mathbf{X}$ is defined in (2.26).

Define the following $M \times L$ noise-sample matrix:

$$\hat{\mathbf{N}} := \mathbf{X} - \mathbf{A}(\boldsymbol{\theta})\hat{\mathbf{S}}. \tag{7.6}$$

Similarly, if the remaining parameters are considered to be fixed, the ML estimate of the noise covariance matrix can be obtained as

$$\hat{\mathbf{Q}} = \frac{1}{L}\mathrm{diag}\left(\hat{\mathbf{N}}\hat{\mathbf{N}}^H\right). \tag{7.7}$$

Hence, replacing (7.7) into (7.1) and omitting constant terms, we obtain the following expression of the LL function:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) &= \mathrm{trace}\left\{\ln\left(\frac{1}{L}\hat{\mathbf{N}}\hat{\mathbf{N}}^H\right)\right\} \\
&= \mathrm{trace}\left\{\ln\left(\frac{1}{L}\boldsymbol{\Pi}^{\perp}_{\tilde{\mathbf{A}}}(\boldsymbol{\theta})\tilde{\mathbf{X}}\tilde{\mathbf{X}}^H\boldsymbol{\Pi}^{\perp}_{\tilde{\mathbf{A}}}(\boldsymbol{\theta})\right)\right\} \\
&= \mathrm{trace}\left\{\ln\left(\boldsymbol{\Pi}^{\perp}_{\tilde{\mathbf{A}}}(\boldsymbol{\theta})\hat{\tilde{\mathbf{R}}}\boldsymbol{\Pi}^{\perp}_{\tilde{\mathbf{A}}}(\boldsymbol{\theta})\right)\right\}
\end{aligned} \tag{7.8}$$

where

$$\hat{\tilde{\mathbf{R}}} := \frac{1}{L}\tilde{\mathbf{X}}\tilde{\mathbf{X}}^H. \tag{7.9}$$

It is important to mention that in (7.8), the $\ln(\cdot)$ function is applied elementwise. The cost function (7.8) shows that the main difference between MLDOA and the uniform DML of Section 2.5.1 lies in the fact that the latter cost function is made independent of the common noise power by factoring it out.

Because of the mutual dependence between the unknown noise covariance matrix $\mathbf{Q}$ and the unknown signal parameters $\mathbf{S}$, it is not possible to analytically concentrate the LL function with respect to either one of these parameters. Instead, [76, 103] suggests an interactive numerical concentration by first initializing the estimated noise covariance matrix to $\hat{\mathbf{Q}} = \mathbf{I}$ and obtaining initial estimates of the DOAs, $\hat{\boldsymbol{\theta}}$. These estimates are in turn used to improve the noise estimate and the whole procedure is repeated a few times until a convergence criterion is satisfied. The optimization typically involves two iterations. This stepwise numerical concentration of the LL function is inspired from AP [119], IQML [18] and MODE [93].

The MLDOA algorithm is also subject to some limitations, namely with regard to the convexity of the cost function in (7.8). In [58], the issue of existence of the global minimum of the negative LL function is addressed. It is shown that for the typical case of spatially nonuniform noise, the number of nuisance parameters in the deterministic LL function increases through the different noise powers, and thus induces a form of indeterminacy. The latter is responsible for a sub-optimal performance of the MLDOA estimator. Recall that in the uniform noise case, the DML is already known for being inefficient as the number of unknown signal parameters increases without bound [51]. The indeterminacy in the model, rather than a failure of the ML approach itself, translates to a nonexistence of a global maximum of the LL objective function, and highlights saddle points instead. This limitation is also partly responsible for the occasional degradation in performance when increasing the number of iterations beyond two.

7.2.2 Weighted Least Squares Approach

The MLDOA cost function can analogously be obtained by following a Weighted Least Squares (WLS) approach leading to the following cost function [63]:

$$\mathcal{V}(\boldsymbol{\eta}_p) = \sum_{t=1}^{L} \left\| \tilde{\mathbf{x}}(t) - \tilde{\mathbf{A}}(\boldsymbol{\theta})\mathbf{s}(t) \right\|^2 \tag{7.10}$$

where the scaled data and steering matrix $\tilde{\mathbf{x}}(t)$ and $\tilde{\mathbf{A}}(\boldsymbol{\theta})$ are defined by (7.2) and (7.3), respectively. Owing to this scaling, the solution to the above cost function becomes similar to the standard uniform noise case.

For a fixed $\tilde{\mathbf{A}}(\boldsymbol{\theta})$, it is straightforward to see that the estimate of the signal samples is obtained as

$$\hat{\mathbf{s}}(t) = \tilde{\mathbf{A}}^{\#}(\boldsymbol{\theta})\tilde{\mathbf{x}}(t). \tag{7.11}$$

Substituting (7.11) into (7.10) leads to the following modified cost function:

$$\mathcal{V}(\boldsymbol{\theta}) = \sum_{t=1}^{L} \left\|\mathbf{\Pi}_{\tilde{\mathbf{A}}}^{\perp}(\boldsymbol{\theta})\tilde{\mathbf{x}}(t)\right\|^2 \tag{7.12}$$

which, for a fixed $\mathbf{Q}$, is equivalent to the ML cost function described previously.

The minimization of the cost function $\mathcal{V}(\boldsymbol{\theta})$ is also iterative, i.e., the noise covariance matrix is initialized first, yielding initial DOA estimates, which in turn are used to improve noise parameter estimates, and so on.

7.2.3 Power Domain Solution

The Power Domain Solution (PDS) was suggested in [63]. It follows from the observation that the columns of $\mathbf{R} - \mathbf{Q}$ are solely dependent on the signal subspace. These columns, denoted $\boldsymbol{\kappa}_{s_m}$, $m = 1, \ldots, M$, have the following structure:

$$\boldsymbol{\kappa}_{s_m} := \mathbf{r}_m - \sigma_m^2 \mathbf{i}_m \tag{7.13}$$

where, as indicated previously, $\mathbf{r}_m$ is the m-th column of $\mathbf{R}$, σ_m^2 is th noise power at the m-th sensor and vector $\mathbf{i}_m$ is defined as

$$\mathbf{i}_{m_i} := \begin{cases} 1 & ; \quad i = m \\ 0 & ; \quad i \neq m. \end{cases} \tag{7.14}$$

It follows from the above that the solution corresponding to the minimum sum of errors for a given vector of unknowns $\boldsymbol{\eta}_p$, is given by

$$\mathcal{V}(\boldsymbol{\eta}_p) = \sum_{m=1}^{M} \boldsymbol{\kappa}_{s_m}^H \mathbf{\Pi}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})\boldsymbol{\kappa}_{s_m} \tag{7.15}$$

where one makes use of the property

$$\mathbf{\Pi}_{\mathbf{A}}^{\perp H}(\boldsymbol{\theta})\mathbf{\Pi}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta}) = \mathbf{\Pi}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})\mathbf{\Pi}_{\mathbf{A}}^{\perp H}(\boldsymbol{\theta}) = \mathbf{\Pi}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta}). \tag{7.16}$$

Using partial derivation, the minimization of (7.15) with respect to the individual noise powers results in

$$\hat{\sigma}_m^2 = \frac{\mathbf{i}_m^H \mathbf{\Pi}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})\mathbf{r}_m + \mathbf{r}_m^H \mathbf{\Pi}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})\mathbf{i}_m}{2\mathbf{i}_m^H \mathbf{\Pi}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})\mathbf{i}_m} \tag{7.17}$$

for $m = 1, \ldots, M$.

By substituting (7.17) into (7.15), the final PDS cost function reduces to

$$\mathcal{V}(\boldsymbol{\theta}) = \sum_{m=1}^{M} \left(\mathbf{r}_m^H \mathbf{\Pi}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})\mathbf{r}_m - \frac{\left[\Re\left(\mathbf{i}_m^H \mathbf{\Pi}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})\mathbf{r}_m\right)\right]^2}{\mathbf{i}_m^H \mathbf{\Pi}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})\mathbf{i}_m} \right). \tag{7.18}$$

Note that the above cost function does not depend on the unknown noise powers. The PDS method involves a considerably smaller computational cost as compared to MLDOA. In particular, both methods require $O(p^3)$ computations to estimate the signal parameters, however MLDOA further needs many more calculations to estimate the noise powers over each iteration. In addition, the PDS solution can be improved by increasing the number of iterations, whereas MLDOA suffers from the non-convexity of its cost-function. MLDOA usually resorts to comprehensive optimization routines based on Genetic Algorithms (GA) to ensure convergence, whereas PDS can rely on considerably simpler Newton-based optimization techniques, similarly to the conventional DML [63]. It is worth mentioning that the PDS is also related to the Covariance Matching Estimation Technique (COMET) [73].

7.3 DOA Estimation using Approximate Subspace Separation

Standard beamforming methods suffer from a poor spatial resolution, which worsens with a lower SNR or a smaller sample size. If the noise powers are very different, this limited resolution is amplified as noise nonuniformity drives the model further away from the asymptotic ideal case. As a result, the obtained DOA estimates include a significant bias and are inefficient.

Subspace separation methods, similarly to the source detectors described previously, require a sufficiently accurate separation criterion. This accuracy can be severely limited by a higher WNPR, following the analysis of Chapter 3. Unless the

SNR is very high and the sources are well separated in space, methods like standard MUSIC will result in erroneous DOA estimates as illustrated in Figure 3.1.

When a criterion testing the rank deficiency of the covariance matrix of the data is available, it is possible to separate the signal and noise subspaces, and hence, use high resolution estimators to retrieve the DOAs, $\boldsymbol{\theta}$. Recall that when the noise is spatially nonuniform, this separation is only approximate. Thus, when the error on the estimated signal subspace is below the threshold described in Section 3.3, one can think of applying a MUSIC-like estimation technique. As opposed to ML methods, subspace techniques alleviate the difficulty related to the convexity of the cost function and can reduce the problem to a liner search over a single dimension.

Following the notation of Section 4.2, let $\mathbf{V}^{(M)}$ be a matrix whose columns are the $M-1-p$ eigenvectors of $\mathbf{R}^{(M)}$, corresponding to the smallest $M-1-p$ Gerschgorin radii $\rho^{(M)}$, i.e.,

$$\mathbf{V}^{(M)} = \left[\mathbf{v}_{p+1}^{(M)}, \ldots, \mathbf{v}_{M-1}^{(M)}\right] \tag{7.19}$$

where $\mathbf{e}_m^{(M)}$, defined in (4.4), is not necessarily equal to $\mathbf{v}_m^{(M)}$, for $m = 1, \ldots, M-1$. Under (3.13), $\mathbf{V}^{(M)}$ approximately spans the noise subspace, or equivalently, the null space of $\mathbf{A}^{(M)}(\boldsymbol{\theta})$. Hence, a Nonuniform MUSIC (NU-MUSIC) objective function can be defined as follows:

$$\mathcal{S}^{(M)}(\theta) := \frac{1}{\mathbf{a}^{(M)H}(\theta)\mathbf{V}^{(M)}\mathbf{V}^{(M)H}\mathbf{a}^{(M)}(\theta)} \tag{7.20}$$

where again, superscript (M) represents the index of the removed array element. The objective function $\mathcal{S}^{(M)}(\theta)$ provides a solution over the range of θ and the estimates are obtained through a 1D search over the range of DOAs as follows:

$$\hat{\theta} := \arg\max_{\theta} \mathcal{S}^{(M)}(\theta). \tag{7.21}$$

Obviously, other high resolution estimators can also be used in a similar way to MUSIC, as the required separation between the signal and noise subspaces is provided by (4.10).

The computational complexity of the NU-MUSIC algorithm as summarized by (7.20), is mainly dominated by the eigen-decomposition of the $(M-1) \times (M-1)$-dimensional covariance matrix $\hat{\mathbf{R}}^{(M)}$ which involves $O\left((M-1)^3\right)$ computations.

Since the noise powers are not equal from one sensor to another, accuracy of the NU-MUSIC estimator depends on the index of the particular array element to be removed. It is clear that M distinct NU-MUSIC objective functions can be obtained from the same array and an alternative estimator can be formulated by averaging the result over the M DOA estimates as follows:

$$\bar{\mathcal{S}}(\theta) = \frac{1}{M} \sum_{m=1}^{M} \mathcal{S}^{(m)}(\theta) \tag{7.22}$$

with

$$\hat{\theta} = \arg\max_{\theta} \bar{\mathcal{S}}(\theta). \tag{7.23}$$

The averaged NU-MUSIC in (7.22), involves M times the computational cost of a single estimator.

It is important to stress that because of the limits of approximate subspace separation, application of NU-MUSIC makes sense only under the scenario described by (3.13) and not (3.12), i.e., when the SNR and the sample size are sufficiently high, and for a limited noise nonuniformity range (WNPR).

Another drawback of the NU-MUSIC is the reduced angular resolution of the estimated sources, which results from a discarded array element. In particular, the estimation quality degrades significantly in instances where the number of sources p is very close to the number of sensors M.

7.4 Nonuniform Approximate Maximum Likelihood

In what follows, we consider a stochastic ML estimator for the specific case of a ULA configuration of M sensors. The proposed method attempts to estimate the noise parameters and scale the data in a single step. It is based on a linear expansion of the noise covariance matrix and signal parameter mapping.

7.4.1 Noise Modeling

In the following, the spatial covariance matrix of the noise, $\mathbf{Q}$, is assumed to have the following general structure:

$$\mathbf{Q} := \sum_{m=1}^{M} \sigma_m^2 \mathbf{\Psi}_m. \tag{7.24}$$

When the noise in the different sensors is assumed to be uncorrelated with different power levels, matrices $\mathbf{\Psi}_m$ are defined by

$$\mathbf{\Psi}_m := \mathbf{i}_m \mathbf{i}_m^T \tag{7.25}$$

where vectors $\mathbf{i}_m$ are defined in (7.14).

The structure of $\mathbf{\Psi}_m$ is defined *a priori*. Note that other cases of known noise covariance structures can be considered and modeled by (7.24) [42]. One way of choosing the latter is to regard the noise as a continuous external source whose power is assumed to be dependent on the DOAs [41, 42]. The noise covariance matrix is therefore defined as

$$\mathbf{Q} := \int_0^{2\pi} \mathbf{a}(\theta)\sigma_n^2(\theta)\mathbf{a}^H(\theta)d\theta \tag{7.26}$$

where $\sigma_n^2(\theta)$ is the periodic DOA-dependent power of the noise. This periodicity (with period 2π) makes it possible to expand the power function using Fourier series, resulting in

$$\sigma_n^2(\theta) = \sum_{k=0}^{\infty} g_k e^{jk\theta}. \tag{7.27}$$

The coefficients of the harmonic terms g_k will be small with increasing k, depending on the smoothness of $\sigma_n^2(\theta)$. This suggests a truncation of the Fourier series expansion to an order $k = M - 1$, leading to the following expression

$$\mathbf{\Psi}_k = \int_0^{2\pi} \mathbf{a}(\theta)e^{jk\theta}\mathbf{a}^H(\theta)d\theta. \tag{7.28}$$

Other choices for the structure of base matrices $\mathbf{\Psi}_k$ and a more thorough description are referred to in [41, 42] and the references therein.

7.4.2 Parameter Space Mapping

Consider a ULA of M sensors receiving p NB signals from sources with unknown DOAs $\boldsymbol{\theta}$.

The signal waveforms are assumed to be random zero-mean Gaussian processes [76, 91], satisfying the assumptions of Section 2.2. Hence, the joint pdf of the data is given by (4.21).

From (2.16), a consistent estimate of $\mathbf{R_s}$ can be obtained by fixing the irrelevant parameters and nulling the derivative of the LL function with respect to the unknown matrix of interest, leading to the following ML estimate [42]:

$$\hat{\mathbf{R}}_{\mathbf{s}} = \mathbf{A}^{\#}(\boldsymbol{\theta}) \left(\hat{\mathbf{R}} - \mathbf{Q}\right) \mathbf{A}^{\#H}(\boldsymbol{\theta}). \tag{7.29}$$

Replacing (7.29) into (2.16) leads to

$$\mathbf{R}(\boldsymbol{\eta}_p) \cong \mathbf{\Pi}_{\mathbf{A}}(\boldsymbol{\theta}) \left(\hat{\mathbf{R}} - \mathbf{Q}\right) \mathbf{\Pi}_{\mathbf{A}}(\boldsymbol{\theta}) + \mathbf{Q} \tag{7.30}$$

where $\boldsymbol{\eta}_p$ is the vector of unknown parameters as defined in Section 3.2.

First, we seek an estimate of the unknown noise parameter $\mathbf{q}$, as defined by (2.15). Thus, let us define the set $\boldsymbol{D}$ of parameters $\mathbf{d} = [1, d_1, \ldots, d_p]^T$, such that in the complex plane, we have [93]

$$z^p + d_1 z^{(p-1)} + \ldots + d_p = \prod_{q=1}^{p} (z - e^{j\pi \cos(\theta_q)}) \tag{7.31}$$

with

$$\boldsymbol{D} := \left\{ d_q \Big/ \sum_{q=1}^{p} d_q z^{p-q} \neq 0 \ \text{ for } \ |z| \neq 1 \right\}. \tag{7.32}$$

Similarly to IQML and MODE, observe that the set of unknown parameters $\boldsymbol{\theta}$ can be reparameterized in terms of $\mathbf{d}$ since the mapping from $\{\theta_q\} \in \mathbb{R}$ to $\{d_q\} \in \boldsymbol{D}$ is an application, under the condition that we eliminate the non-uniqueness related to the case $d_0 \neq 1$.

Since the polynomial (7.31) has all of its zeros on the unit circle, its coefficients satisfy the conjugate symmetry constraint [53, 93] as follows:

$$d_q = d^*_{p-q}, \qquad q = 0, \ldots, p. \tag{7.33}$$

with $(\cdot)^*$ standing for complex conjugate.

From the analysis of [93], it is clear that the above constraint is necessary but not sufficient for $\{d_q\} \in \boldsymbol{D}$. Nevertheless, sticking to the constraint (7.33) implies a very minor loss of performance in practice [57]. The commonly used normalization is $d_0 = 1$ [57].

Define matrix $\mathbf{D}$ as

$$\mathbf{D} := \begin{bmatrix} d_p & \dots & d_1 & 1 & & \dots & 0 \\ 0 & d_p & \dots & d_1 & 1 & \dots & 0 \\ \vdots & & \ddots & & \ddots & & \vdots \\ 0 & \dots & & d_p & \dots & d_1 & 1 \end{bmatrix} \tag{7.34}$$

and observe that, due to the ULA configuration, we have

$$\mathbf{D}\mathbf{A}(\boldsymbol{\theta}) = \mathbf{0}. \tag{7.35}$$

7.4.3 Parameter Estimation

Using (7.35), we define the following (where we drop the dependence on $\boldsymbol{\eta}_p$ for now):

$$\begin{aligned} \mathbf{R_d} &:= \mathbf{D}\mathbf{R}(\boldsymbol{\eta}_p)\mathbf{D}^H \\ &= \mathbf{D}\mathbf{Q}\mathbf{D}^H. \end{aligned} \tag{7.36}$$

Furthermore, taking into account properties of the trace$(\cdot)$ and vec$(\cdot)$ operators [19], we define the following WLS cost function:

$$\begin{aligned} \mathcal{V}(\mathbf{q}, \mathbf{d}) &= \text{trace}\left\{ \left(\mathbf{R_d} - \hat{\mathbf{R}}_\mathbf{d}\right) \mathbf{W} \left(\mathbf{R_d} - \hat{\mathbf{R}}_\mathbf{d}\right) \mathbf{W} \right\} \\ &= \text{vec}^H \left(\mathbf{R_d} - \hat{\mathbf{R}}_\mathbf{d}\right) \text{vec} \left[\mathbf{W} \left(\mathbf{R_d} - \hat{\mathbf{R}}_\mathbf{d}\right) \mathbf{W} \right] \\ &= \left(\mathbf{r_d} - \hat{\mathbf{r}}_\mathbf{d}\right)^H \left(\mathbf{W}^T \otimes \mathbf{W}\right) \left(\mathbf{r_d} - \hat{\mathbf{r}}_\mathbf{d}\right) \end{aligned} \tag{7.37}$$

where $\mathbf{W}$ is a weighting matrix, $\otimes$ denotes Kronecker product and $\hat{\mathbf{R}}_\mathbf{d} = \mathbf{D}\hat{\mathbf{R}}\mathbf{D}^H$, with

$$\hat{\mathbf{r}}_\mathbf{d} := \text{vec}\left(\hat{\mathbf{R}}_\mathbf{d}\right) \tag{7.38}$$

and

$$\begin{aligned}\mathbf{r_d} &:= \operatorname{vec}(\mathbf{R_d}) \\ &:= (\mathbf{D}^* \otimes \mathbf{D}) \operatorname{vec}(\mathbf{Q}) \\ &:= (\mathbf{D}^* \otimes \mathbf{D}) \mathbf{F}\mathbf{q} \\ &:= \mathbf{G}(\mathbf{d})\mathbf{q} \end{aligned} \tag{7.39}$$

and matrix $\mathbf{F}$ defined as

$$\mathbf{F} := \left[\operatorname{vec}\{\mathbf{\Psi}_1\} \quad \ldots \quad \operatorname{vec}\{\mathbf{\Psi}_M\} \right] \tag{7.40}$$

where matrices $\mathbf{\Psi}_m$, $m = 1, \ldots, M$, are defined in (7.25).

Using properties of the trace$(\cdot)$ and vec$(\cdot)$ operators [19], we can further simplify cost function (7.37) to

$$\mathcal{V}(\mathbf{q}, \mathbf{d}) = (\mathbf{r_d} - \hat{\mathbf{r}}_\mathbf{d})^H \tilde{\mathbf{W}} (\mathbf{r_d} - \hat{\mathbf{r}}_\mathbf{d}) \tag{7.41}$$

where $\tilde{\mathbf{W}} = \mathbf{W}^T \otimes \mathbf{W}$.

Furthermore, from (7.37), (7.38), and (7.39), cost function (7.41) becomes

$$\mathcal{V}(\mathbf{q}, \mathbf{d}) = (\mathbf{G}(\mathbf{d})\mathbf{q} - \hat{\mathbf{r}}_\mathbf{d})^H \tilde{\mathbf{W}} (\mathbf{G}(\mathbf{d})\mathbf{q} - \hat{\mathbf{r}}_\mathbf{d}) . \tag{7.42}$$

Keeping $\mathbf{d}$ fixed and nulling the quadratic expression (7.42) with respect to $\mathbf{q}$, we obtain the following WLS solution:

$$\begin{aligned}\hat{\mathbf{q}}(\mathbf{d}) &= \left[\mathbf{G}^H(\mathbf{d})\tilde{\mathbf{W}}\mathbf{G}(\mathbf{d})\right]^{-1} \mathbf{G}^H(\mathbf{d})\tilde{\mathbf{W}}\hat{\mathbf{r}}_\mathbf{d} \\ &:= \mathbf{\Delta}(\mathbf{d})\hat{\mathbf{r}}_\mathbf{d}. \end{aligned} \tag{7.43}$$

For this WLS problem, it is shown in [81, 42] that the choice for the weighting matrix as $\mathbf{W} = \mathbf{R}_\mathbf{d}^{-1}$ ensures asymptotic convergence to the CRB on the estimation error. It is also shown that these asymptotic properties can be maintained by replacing $\mathbf{R}_\mathbf{d}^{-1}$ by its consistent estimator $\hat{\mathbf{R}}_\mathbf{d}^{-1}$.

From (7.35), since rank $(\mathbf{A}(\boldsymbol{\theta})) = p$ and rank$(\mathbf{D}) = M - p$, it results that

$$\begin{aligned}\mathbf{\Pi}_\mathbf{A}(\boldsymbol{\theta}) &= \mathbf{I} - \mathbf{D}^H \left(\mathbf{D}\mathbf{D}^H\right)^{-1} \mathbf{D} \\ &= \mathbf{\Pi}_{\mathbf{D}^H}^{\perp}. \end{aligned} \tag{7.44}$$

Thus, replacing (7.43) and (7.44) into (7.30) results in

$$\hat{\mathbf{R}}(\mathbf{d}) \cong \mathbf{\Pi}_{\mathbf{D}^H}^{\perp} \left(\hat{\mathbf{R}} - \operatorname{diag}\{\hat{\mathbf{q}}(\mathbf{d})\}\right) \mathbf{\Pi}_{\mathbf{D}^H}^{\perp} + \operatorname{diag}\{\hat{\mathbf{q}}(\mathbf{d})\}. \tag{7.45}$$

Note that $\tilde{\mathbf{R}}(\mathbf{d})$ in (7.45) is completely parameterized in terms of $\mathbf{d}$. The obtained expression (7.45) can then be replaced into the initial stochastic LL cost function to give the following new reparameterized LL function:

$$\mathcal{L}(\mathbf{d}) \cong \ln\left(\det\left(\hat{\mathbf{R}}(\mathbf{d})\right)\right) + \operatorname{trace}\left\{\hat{\mathbf{R}}^{-1}(\mathbf{d})\hat{\mathbf{R}}\right\} \tag{7.46}$$

Finally, as the above approximate LL function is only parameterized in terms of the unknown coefficients $\mathbf{d}$, the algorithm reduces to the single step of minimizing (7.46). The parameters of interest $\boldsymbol{\theta}$ can then be easily obtained from the previously described application form $\mathbb{R}$ into $\boldsymbol{D}$, by reconstructing the polynomial of (7.31) using the estimated values $\hat{\mathbf{d}}$ and solving for its roots, where

$$\begin{aligned} \hat{\mathbf{d}} &= [1, \hat{d}_1, \ldots, \hat{d}_p]^T \\ &= \arg \min_{\{d_q\} \in \boldsymbol{D}} \mathcal{L}(\mathbf{d}). \end{aligned} \tag{7.47}$$

Similarly to MODE and IQML, the approximate constraint of (7.33) is usually eliminated by reparameterizing the cost function in terms of a real-valued vector $\boldsymbol{\beta} \in \mathbb{R}^{(p+1)\times 1}$, satisfying

$$\mathbf{d} := \mathbf{K}\boldsymbol{\beta} \tag{7.48}$$

with $\mathbf{K}$ being a square $(p+1) \times (p+1)$-dimensional matrix of appropriately set elements 0, 1 and $\pm j$, and $\beta_0 = 1$ [57].

The main difference to MODE and IQML however, is that the proposed ML approach is stochastic rather than deterministic. Also, because of the noise nonuniformity, cost function (7.47) is considerably more sensitive to initial values, in a way similar to MLDOA. Because of the noise covariance matrix reparameterization described in Section 7.4.1, the proposed ML estimator is similar to the AML approach [42]. The method offers at least a simpler solution as compared to MLDOA, as it does not estimate the DOAs and the nuisance parameters iteratively. Similarly to the PDS, it also retains similar features to the COMET class of estimators [73].

7.5 Identifiability and the Cramér-Rao Bound

Identifiability of the unknown parameters translates to the bound on the number of equations versus the number of unknowns in the model [46]. It is obvious that identifiability is not guaranteed when the number of unknowns exceeds the number of equations. Recall that for DOA identifiability, the number of sources p is assumed to be known *a priori*. Taking into account its rank p, and dimension $M \times M$, it follows that matrix $\mathbf{A}(\boldsymbol{\theta})\mathbf{R}_{\mathbf{s}}\mathbf{A}^H(\boldsymbol{\theta})$ is parameterized by $p(2M-p)$ real entries, as it is nonnegative definite. Keeping in mind the diagonal structure of $\mathbf{Q}$, it is clear that the number of independent equations in $\mathbf{R}$ is also $p(2M-p)$ [18, 46]. On the other hand, there are p^2 unknowns in $\mathbf{R}_{\mathbf{s}}$, p unknowns in $\boldsymbol{\theta}$, and M unknowns in $\mathbf{Q}$. Therefore, given the system of (2.6), it is necessary for the identifiability of the unknown parameters that the number of equations satisfies the inequality $p(2M-p) \geq p^2 + p + M$, which leads to the condition $p < M$.

Derivation of the NU-MUSIC algorithm of Section 7.3 is based on a reduced model after an arbitrary sensor is discarded from the original array. The first consequence of this reduction is a more restrictive identifiability condition. Specifically, in addition to $\mathbf{A}(\boldsymbol{\theta})$, the reduced matrix $\mathbf{A}^{(M)}(\boldsymbol{\theta})$ must be of full column rank, meaning that $p < M-1$, which follows from the number of independent equations in the reduced covariance matrix $\mathbf{R}^{(M)}$.

The second consequence appears in the asymptotic performance of the NU-MUSIC estimator. More specifically, the variance of the estimated DOAs $\hat{\boldsymbol{\theta}}$, will not reach the CRB corresponding to the original covariance matrix $\mathbf{R}$. In the best case, a specific NU-MUSIC leads to estimates that can asymptotically converge to the CRB corresponding to $\mathbf{R}^{(M)}$ (noted here $\mathrm{CRB}^{(M)}$). It is worth reminding that the identifiability conditions mentioned above ensure the nonsingularity of Fisher's Information Matrix (FIM) and thus the existence of the CRB [39].

Obviously, for a given number of nuisance parameters, a decreasing number of sensors inevitably results in an increasing value of $\mathrm{CRB}^{(M)}$ [39]. At the same time, as the noise covariance matrix $\mathbf{Q}$ differs from one model to another, variation of the value of $\mathrm{CRB}^{(M)}$ depends on the distribution of the noise powers and more particularly the value of σ_M^2, corresponding to the removed sensor. Indeed, by discarding one sensor, the nuisance parameter σ_M^2 is discarded from the data model,

leading to a decreasing $\text{CRB}^{(M)}$ [76, 91]. It is clear therefore that the relation between $\text{CRB}^{(M)}$ and the original CRB depends on the number of sensors, as well as the actual values of the noise variances. Such a relation is shown in Figure 7.2, where the value of CRB varies with the number of sensors M for two examples. In both examples, the DOAs are fixed at $\boldsymbol{\theta} = [0°, 45°]^T$ and the number of samples at $L = 100$. The noise powers are defined by individual SNRs (in dB) by vectors $\mathbf{q}_1$ and $\mathbf{q}_2$, for examples (a) and (b), respectively, as follows:

$$\begin{aligned}\mathbf{q}_1 &= [-1, -6, -2, -5, 0, -55, -55, -57, 15]^T, \\ \mathbf{q}_2 &= [-1, -6, -2, -5, 0, -55, 35, 35, 37]^T.\end{aligned}$$

Every smaller array is considered by simply removing the last sensor from the larger one (therefore discarding one nuisance parameter). Note how the CRB varies with an increasing M, describing the value of $\text{CRB}^{(M)}$ as opposed to CRB. Both examples show the same values of CRB up to $M = 6$. It appears clearly that addition of more sensors does not necessarily improve the estimation accuracy in a significant manner. Equivalently, if the removed sensor undergoes *intensive* noise, the cost for a reduced spatial diversity can be negligible (asymptotically).

The combination of the number of sensors and the values of the corresponding noise variances describes the *degree of nonuniformity* of the sensor noise.

Derivation of the approximate ML estimator of the previous section implies more restrictive identifiability requirements. In addition to the requirement that $\mathbf{R}$ is of rank p, the WLS system of (7.37) must be completely determined [81], and the $(M-p)^2 \times p$-dimensional matrix $\boldsymbol{\Delta}(\mathbf{d})$ of (7.43) must be tall, or equivalently $(M-p)^2 \geq p$ must be satisfied. Note that the latter condition on the number of sensors M is more restrictive than the usual $M > p$ condition.

For reference, the uniform MUSIC estimator is a large sample realization of the uniform ML estimator [91], with $M > p$. At the same time, the quantitative results of [91, 92] show that for close sources and a fixed SNR, efficiency of the MUSIC estimator can drop below 50% for typical values of $p = 2$ and $M = 5, 10$; which are above the threshold $(M-p)^2 \geq p$. This uniform MUSIC and ML performances further degrade as the sources become more correlated and the noise deviates from the uniform model. Therefore, in such scenarios the requirement of $(M-p)^2 \geq p$ can be acceptable given the improvement in performance that can be induced by

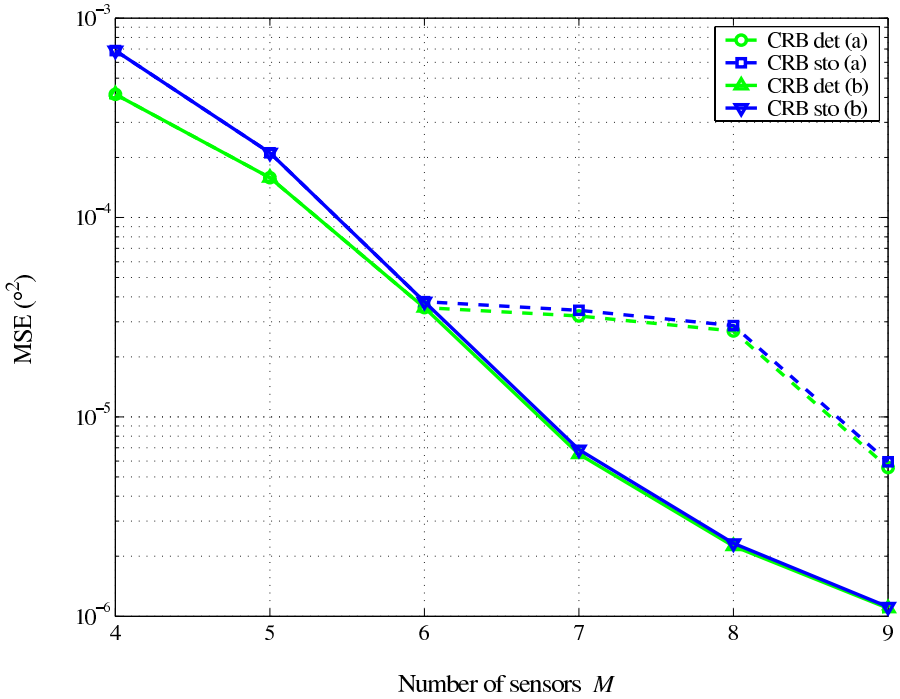

Figure 7.1: Effect of noise-related nuisance parameters on the CRB, through the variation of M.

the nonuniform ML approach.

In the general case of a covariance matrix $\mathbf{R}$, depending on whether the data signals $\mathbf{s}(t)$ are modeled as deterministic unknowns, or normally distributed random variables, the deterministic and stochastic CRBs for nonuniform noise were derived in [76], and are respectively given by:

$$\mathrm{CRB}_{\mathrm{det}}(\boldsymbol{\theta}) = \frac{1}{2L}\left\{\Re\left(\tilde{\mathbf{A}}'^{H}(\boldsymbol{\theta})\boldsymbol{\Pi}^{\perp}_{\tilde{\mathbf{A}}}(\boldsymbol{\theta})\tilde{\mathbf{A}}'(\boldsymbol{\theta})\right)\odot\hat{\mathbf{R}}_{\mathbf{s}}\right\}^{-1} \tag{7.49}$$

and

$$\mathrm{CRB}_{\mathrm{sto}}(\boldsymbol{\theta}) = \frac{1}{L}\left\{2\Re\left[\left(\mathbf{R}_{\mathbf{s}}\tilde{\mathbf{A}}^{H}(\boldsymbol{\theta})\tilde{\mathbf{R}}^{-1}\tilde{\mathbf{A}}(\boldsymbol{\theta})\mathbf{R}_{\mathbf{s}}\right)\right.\right.$$
$$\left.\left.\odot\left(\tilde{\mathbf{A}}'^{H}(\boldsymbol{\theta})\boldsymbol{\Pi}^{\perp}_{\tilde{\mathbf{A}}}(\boldsymbol{\theta})\tilde{\mathbf{R}}^{-1}\tilde{\mathbf{A}}'(\boldsymbol{\theta})\right)^{T}\right]-\mathfrak{M}\mathfrak{T}^{-1}\mathfrak{M}^{T}\right\}^{-1} \tag{7.50}$$

where

$$\tilde{\mathbf{A}}'(\boldsymbol{\theta}) = \mathbf{Q}^{-1/2}\mathbf{A}'(\boldsymbol{\theta}) \tag{7.51}$$

$$\mathbf{A}'(\boldsymbol{\theta}) = \left[\left.\frac{d\mathbf{a}(\theta)}{d\theta}\right|_{\theta_1} \quad \cdots \quad \left.\frac{d\mathbf{a}(\theta)}{d\theta}\right|_{\theta_p} \right] \tag{7.52}$$

$$\hat{\mathbf{R}}_{\mathbf{s}} = \frac{1}{L}\mathbf{S}\mathbf{S}^H \tag{7.53}$$

$$\mathfrak{M} = 2\Re\left\{ \left(\tilde{\mathbf{R}}^{-1}\tilde{\mathbf{A}}(\boldsymbol{\theta})\mathbf{R}_{\mathbf{s}}\right)^T \odot \left(\tilde{\mathbf{A}'}^H(\boldsymbol{\theta})\boldsymbol{\Pi}^{\perp}_{\tilde{\mathbf{A}}}(\boldsymbol{\theta})\right) \right\} \tag{7.54}$$

$$\mathfrak{T} = \left(\tilde{\mathbf{R}}^{-1}\right)^* \odot \tilde{\mathbf{R}}^{-1} - \left(\boldsymbol{\Pi}_{\tilde{\mathbf{A}}}(\boldsymbol{\theta})\tilde{\mathbf{R}}^{-1}\right)^* \odot \left(\boldsymbol{\Pi}_{\tilde{\mathbf{A}}}(\boldsymbol{\theta})\tilde{\mathbf{R}}^{-1}\right) \tag{7.55}$$

and $\odot$ stands for Schur-Hadamard matrix product, i.e., elementwise product.

The CRB expressions in (7.49) and (7.50) are also derived under a more general framework in [42, 103].

We extended this derivation for the 2D case when evaluating the 2D-AML for the specific case of localizing low-flying targets. As detailed in Appendices A and B.7, the linear expansion of the noise covariance matrix using known base matrices is exploited, in a similar way to (7.24).

7.6 Numerical Examples

In what follows we show the performance of the nonuniform stochastic ML estimator of Section 7.4, as well as the NU-MUSIC estimator of Section 7.3, and compare it to the stochastic and deterministic CRB. Although designed for the deterministic signal scenario, we also illustrate the behavior of the PDS of Section 7.2.1 and MLDOA of Section 7.2.3, along with the uniform DML and conventional MUSIC of Sections 2.5.1 and 2.5.3, respectively, to visualize the degradation of performance in nonuniform noise due to model mismatch. For conciseness purposes, we illustrate the results corresponding to the second source only, as estimates of both DOAs have similar properties.

Note that the performance of both nonuniform ML estimators is very similar. The number of sources is set to $p = 2$ and a ULA with $M = 6$ sensors is used. All the examples illustrate the estimators' performance in terms of the Root Mean Square Error (rMSE) of the estimated DOAs, over 200 Monte Carlo runs.

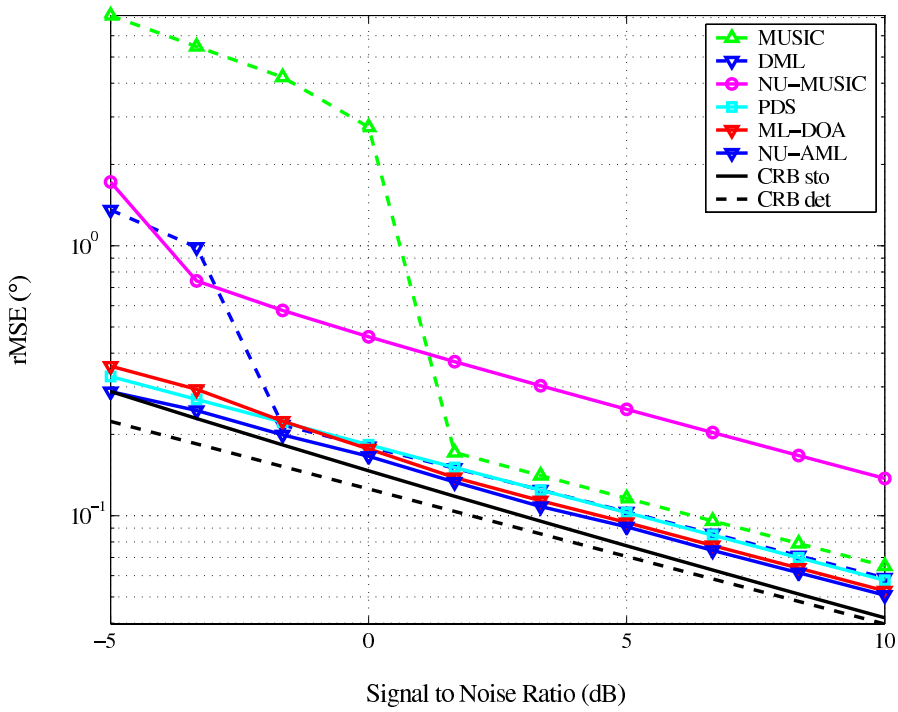

Figure 7.2: Performance of the estimators vs SNR in nonuniform noise.

- Figure 7.2 illustrates the performance with respect to the SNR. The fixed parameters are the number of snapshots $L = 100$ and the DOAs $\boldsymbol{\theta} = [0°, 17°]^T$, whereas the SNR varies from -5 dB to 10 dB. The WNPR is fixed at 10, and the noise powers are generated from a uniform distribution over the interval $[1, \text{WNPR}]$ and normalized by their mean values. As expected, both uniform estimators, i.e. MUSIC and DML, fail to resolve the DOAs at low SNR, whereas all other detectors are more suitable to varying noise-powers. Note however that due to the reduced array size, the NU-MUSIC finds its relative advantage over the uniform MUSIC only when SNR is very low, otherwise it induces a significant bias in the estimated DOA, which does not improve with increasing SNR, but is rather dependent on the spatial separation between the sources. All other methods are considerably more robust to noise-power variation. The performance shows a slight bias in the estimates that is due to the choice of the optimization routine and the preset convergence criterion. Here we used standard nonlinear programming for a relatively low number of iterations. For the ML estimators, both stochastic and deterministic, accuracy

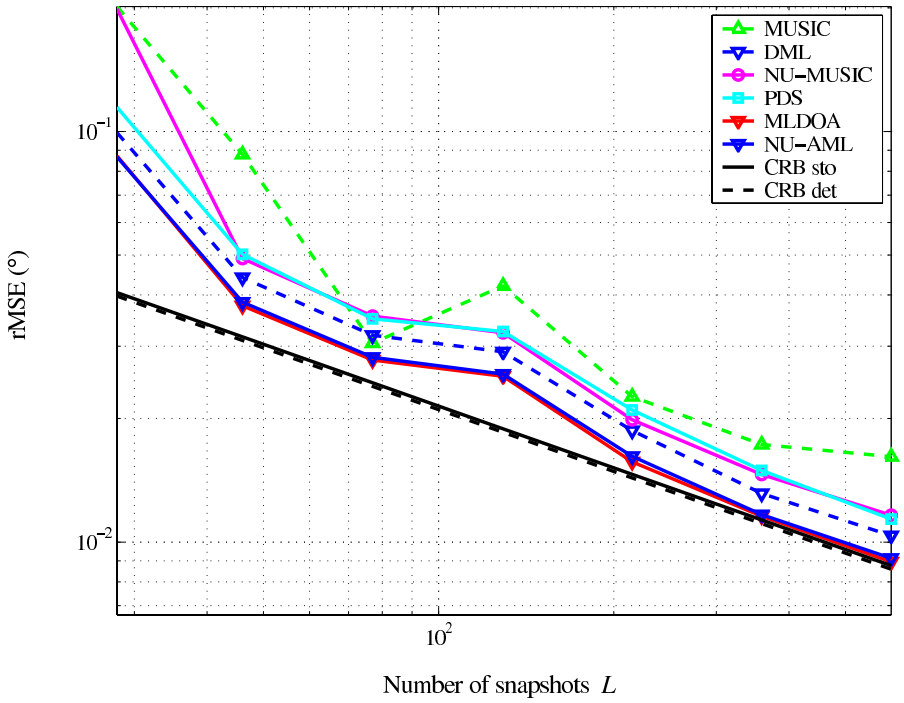

Figure 7.3: Performance of the estimators vs L in nonuniform noise.

of the estimation is dependent on the choice of the initial DOA values. These initial estimates are obtained using MUSIC under the assumption that $\mathbf{Q} = \mathbf{I}$. The PDS methods emerges as an accurate method, which is the fastest to converge to a minimum.

- Figure 7.3 illustrates the performance with respect to the number of snapshots L which varies from 30 to 600. The fixed parameters are SNR=10 dB and the DOAs $\boldsymbol{\theta} = [0°, 17°]^T$. The same noise powers as for the previous example are used, thus the same WNPR applies. Similarly, MUSIC fails to retrieve the DOAs. The ML estimators perform considerably better and show a similar relative performance. Note that as the number of samples increases, the number of unknown signal parameters increases accordingly. As a result, the deterministic estimators require a considerably higher amount of iterations, or equivalently a more restrictive convergence criterion, to yield efficient estimates.

- Figure 7.4 illustrates the performance with respect to the spatial resolution

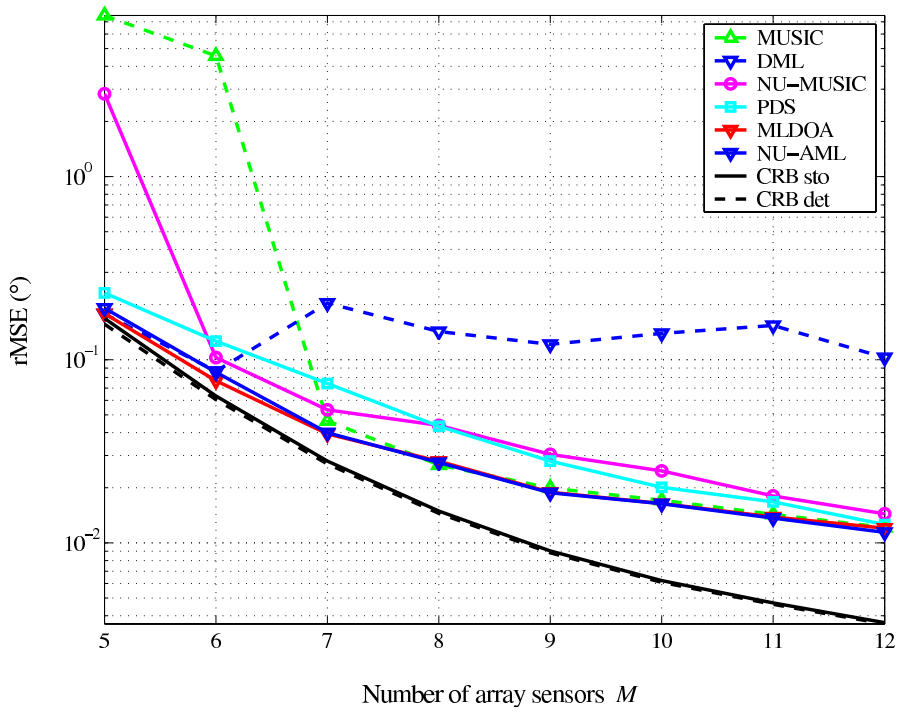

Figure 7.4: Performance of the estimators vs M in nonuniform noise.

of the array, where the number of sensors varies from 5 to 12. The fixed parameters are the DOAs $\boldsymbol{\theta} = [0°, 17°]^T$, SNR=10 dB, WNPR=10, and the number of snapshots $L = 100$. The noise powers are generated in the same way as in the previous examples. Note that as the number of sensors increases, the beamwidth of the array decreases, thus increasing the angular resolution of the sources. As expected, when the number of sensors is low, MUSIC and NU-MUSIC, like most spectrum-based methods, fail to resolve the DOAs. Moreover, as outlined in [76], for a small number of sensors, the number of unknown parameters associated with nonuniform noise is very close to the number of available equations in the model [46], which leads to very similar performance levels among the ML-based methods. Note also that while PDS is a faster method, it induces a larger bias in the estimated DOAs for a small number of iterations.

- Figure 7.5 illustrates the performance with respect to the WNPR. The fixed parameters are SNR=10 dB, the number of snapshots $L = 100$ and the DOAs $\boldsymbol{\theta} = [0°, 17°]^T$. The noise parameters are chosen such that the WNPR is varied

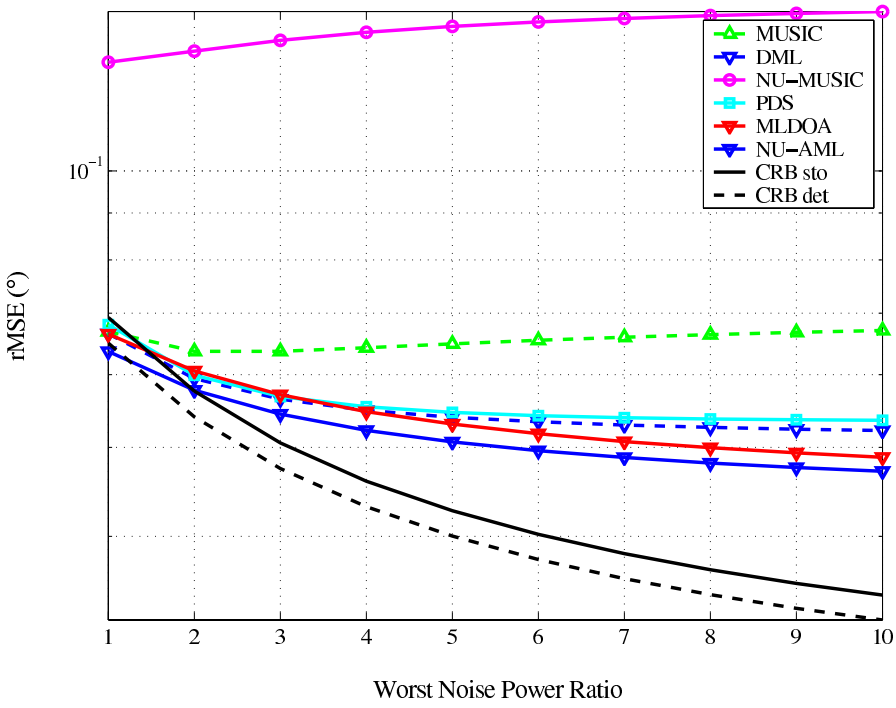

Figure 7.5: Performance of the estimators vs WNPR in nonuniform noise.

from 1 to 10. Note that the performance of all the estimators degrades as the noise becomes more nonuniform, however the relative advantage of the ML estimators becomes more obvious with a higher WNPR. Note also that the high estimation error induced by NU-MUSIC is of the same order as in the previous examples and does not improve with the variation of the WNPR. This is essentially due to the reduced array size, and the low quality of the estimated signal subspace, as outlined in Chapter 3. Note also that in the general case, the WNPR does not provide an information on the specific distribution of the noise powers.

7.7 Conclusion

We investigated the application of the Gerschgorin-based approximate subspace separation criterion to DOA estimation. The obtained NU-MUSIC retains the advantages of the classical MUSIC in terms of reduced computational cost and easy implementation. However, NU-MUSIC suffers from a limited angular resolution, which

results from a discarded array element. Moreover, the overall performance of NU-MUSIC is strongly sensitive to a high SNR and noise nonuniformity range (WNPR) and a small sample size.

Alternatively, we proposed a new nonuniform stochastic AML estimator of the DOAs. This algorithm relies on the knowledge of the structure of the covariance matrix of the noise and employs a linear expansion of the latter. It is possible to approximate the LL function by incorporating a consistent estimate of this covariance matrix. In the case of a ULA (or more generally a Vandermonde-type steering matrix), a reparameterization of the unknowns is applied to facilitate the isolation of nuisance parameters in the global cost function. This approximation avoids an iterative concentration of the LL function and the associated propagation of the estimation error. However, the method remains sub-optimal and is sensitive to initial conditions.

Chapter 8

Conclusions

The problem of estimating the dimension of the signal subspace solely from the data received by an unbalanced array of sensors has been addressed. When the unbalance in the array is caused by unequal individual noise perturbations over the sensors, the problem of estimating the number of sources is non-identifiable. However, the only theoretical example of non-identifiability is unlikely to occur in practice.

In our analysis, we opted for practically applicable approaches, thus favoring simple and sufficiently accurate detectors. The main idea behind this approach was to identify and bypass the elements responsible for performance degradation of classical detectors in spatially nonuniform noise. Hence, under the assumptions of ideal uniform noise, and both Gaussian signals and noise, the asymptotic behavior of the estimated signal subspace has been analyzed. It was deduced that the classical criterion of eigenvalue-equality is very sensitive to noise nonuniformity. In addition, a threshold on the values of the SNR, sample size and noise-power perturbation has been highlighted, beyond which subspace separation is theoretically impossible and any approximation becomes problematic.

Based on the above analysis, an alternative criterion of approximate subspace separation, based on Gerschgorin's theorem, has been proposed. The employed criterion is in fact a measure of the closeness of the estimated signal subspace to its theoretical reference. Up to a threshold on the variation of the estimation error, this criterion is more relaxed than the classical eigenvalue-equality criterion. The ESQ detector is a hypothesis test based on the above measure. It follows from a transformation of the covariance matrix of the data, resulting from array element

suppression to cope with potential noise-power discrepancies.

When no *a priori* knowledge is available on the distribution of the data, the ESQ test has been applied with a variety of test statistics whose null distributions were empirically estimated using the bootstrap. The algebraic structure of the employed test statistics reduces significantly the effect of a possible bias in the value of the estimated Gerschgorin radii, and thus saves the computational cost of the whole testing procedure. In order to validate the approach, an example with real data has been successfully used.

Similarly to the ESQ test, the proposed information theoretic criteria, i.e., NU-AIC and NU-MDL, follow from array transformation and Gerschgorin?s theorem. They are characterized by the same increased robustness to small noise power variation over the array sensors. Both theoretical and numerical performance analysis have been proposed to support the superiority of these criteria over classical eigenvalue-based detectors in the unconditional Gaussian case (asymptotically).

Nevertheless, due to the required array transformation, a good performance in spatially perturbed noise comes at a cost of reduced spatial diversity and stricter identifiability conditions on the number of resolvable sources with respect to array elements.

When the SNR and the sample size are sufficiently high, and for a limited noise nonuniformity range, the above approximate subspace separation method can be used to estimate the signal parameters, provided that the number of sources is known. However, while offering a smaller computational load, this approach is limited by a reduced angular resolution of the estimated sources, as a result of a discarded array element.

As an alternative, we have investigated a new stochastic nonuniform ML estimator of the DOAs. As compared to existing approaches, this algorithm avoids a stepwise concentration of the cost function with respect to the different unknown parameters. In particular, isolation of the nuisance parameters in the global cost function follows from the injection of a consistent estimate of the noise covariance matrix, as well as a reparameterization of the source parameters. The method remains sub-optimal and is sensitive to initialization.

Taking into account the nature of the addressed problem, its importance in practice and the above concluding remarks, several issues can be further investigated

beyond the scope of this dissertation. These aspects include the following:

- Similarly to the analysis of Chapter 6, it is interesting to analyze the effect of a possible bias in the estimated Gerschgorin radii in the Gaussian data case of Chapter 4. In particular, one is interested in deriving an expression for the bias and investigating simple and efficient methods to circumvent its effect in scenarios where the sample size is small or the SNR is low. Note that the settings of the examples in Chapter 4 correspond to a case where additional processing to account for the bias is not required.

- Observe that the WNPR is only a rough measure of noise-power nonuniformity. It is interesting to further investigate the effect of particular noise-power distributions on the performance of existing methods, as well as the proposed algorithms.

- Note that the comparison between the original ESQ hypothesis test of Chapter 4 and the bootstrap version of Chapter 6 is made indirectly by considering the performance of NU-MDL as a common reference. However, further analysis can be conducted to better identify the limits of the methods and their relative advantages in different scenarios, especially for deviations from asymptotic cases. In particular, one is interested in assessing the limits of validity of the normality properties of the test statistics of Chapter 4, and verifying the robustness of the bootstrap as an alternative.

- One direction for a possible performance improvement of both source detection and parameter estimation, is to investigate a different array transformation that allows full use of the available spatial diversity offered by the array. Alternatively, the simultaneous use of a number of sub-arrays, in a way similar to the generalized ESPRIT method can be analyzed. This approach, especially when supported by real-data experiments, can be of particular interest in smart antennae or spatially distributed antenna networks.

- Another direction is to conduct a thorough performance analysis, both for asymptotic and finite sample cases, of the existing and new DOA estimation methods, with an emphasis on potential equivalences between ML and other

sub-optimal estimators. Useful approximations are of particular interest, suggesting which ML estimator is better in practice.

- Finally, a full analysis of the problem is worth being conducted beyond the threshold on the estimated subspace quality, which is described in Chapter 3, i.e., under the more restrictive scenario of (3.12).

Appendix A

Joint Estimation of Angles and Frequencies in Partially Unknown Noise

A.1 Approximate Maximum Likelihood (AML)

Given the model (2.16), assume that the noise covariance matrix has a band structure, and consider the $(2M-1)$-dimensional vector of unknown noise parameters $\mathbf{q}$, given as $\mathbf{q} = [\mathrm{q}_0, \Re(\mathrm{q}_1), \Im(\mathrm{q}_1), \ldots, \Re(\mathrm{q}_{M-1}), \Im(\mathrm{q}_{M-1})]^T$, where $\mathrm{q}_0 = \sigma^2$ is the spatially uniform noise power, corresponding to the diagonal elements of the noise covariance matrix $\mathbf{Q}(\mathbf{q})$. More generally, $\mathbf{Q}(\mathbf{q})$ can be modeled as a linear combination of $2M-1$ known weighting matrices $\mathbf{\Psi}_i$ and a set of unknown real parameters q_i (the elements of $\mathbf{q}$) [14, 42], i.e.,

$$\mathbf{Q}(\mathbf{q}) = \sum_{i=1}^{2M-1} q_i \mathbf{\Psi}_i. \tag{A.1}$$

In [42], an application to underwater acoustics of the model (A.1) is described. Other examples include cases where the noise covariance matrix is approximated by a sum of weighted Fourier coefficients and the base matrices $\mathbf{\Psi}_i$ are functions of the known array geometry (see Section 7.4.1). This approximation is a result of a Fourier series expansion of the spatial noise power function [36]. More general models are also considered where the noise is spatially correlated, with a banded

and Toeplitz covariance matrix. Due to the Hermitian structure of $\mathbf{Q}(\mathbf{q})$, the base matrices $\boldsymbol{\Psi}_i$ have diagonals composed of zeros, ± 1, and $\pm j$ [37].

The AML algorithm [42] is based on the application of the properties of the vec$(\cdot)$ operator to (2.16) and using the expansion in (A.1),

$$\text{vec}(\mathbf{R}) = \text{vec}\left(\mathbf{A}(\boldsymbol{\theta})\mathbf{R}_{\mathbf{H}}\mathbf{A}^H(\boldsymbol{\theta})\right) + \text{vec}(\mathbf{Q}(\mathbf{q})). \tag{A.2}$$

Defining $\boldsymbol{\mathcal{A}}(\boldsymbol{\theta}) = [\mathbf{A}^*(\boldsymbol{\theta}) \otimes \mathbf{A}(\boldsymbol{\theta})]$ and $\mathbf{r}_{\mathbf{s}} = \text{vec}(\mathbf{R}_{\mathbf{s}})$, (A.2) can be written as

$$\text{vec}(\mathbf{R}) = \begin{bmatrix} \boldsymbol{\mathcal{A}}(\boldsymbol{\theta}) & \boldsymbol{\mathcal{P}} \end{bmatrix} \begin{bmatrix} \mathbf{r}_{\mathbf{s}} \\ \mathbf{q} \end{bmatrix} \tag{A.3}$$

where $\boldsymbol{\mathcal{P}} = [\text{vec}(\boldsymbol{\Psi}_1) \ \ldots \ \text{vec}(\boldsymbol{\Psi}_{2M-1})]$.

Define the following:

$$\boldsymbol{\mathcal{M}}(\boldsymbol{\theta}) := \begin{bmatrix} \boldsymbol{\mathcal{A}}(\boldsymbol{\theta}) & \boldsymbol{\mathcal{P}} \end{bmatrix} \tag{A.4}$$

$$\boldsymbol{p} := \begin{bmatrix} \mathbf{r}_{\mathbf{s}} \\ \mathbf{q} \end{bmatrix}. \tag{A.5}$$

From (A.3), a consistent estimate $\hat{\boldsymbol{p}}$ can be obtained as [42]

$$\hat{\boldsymbol{p}} = \left[\boldsymbol{\mathcal{M}}^H\left(\hat{\mathbf{R}}^{-T} \otimes \hat{\mathbf{R}}^{-1}\right)\boldsymbol{\mathcal{M}}\right]^{-1} \boldsymbol{\mathcal{M}}^H\left(\hat{\mathbf{R}}^{-T} \otimes \hat{\mathbf{R}}^{-1}\right)\hat{\mathbf{r}} \tag{A.6}$$

where $\hat{\mathbf{R}}$ is the sample covariance matrix of the data, defined in (2.24), and $\hat{\mathbf{r}} = \text{vec}(\hat{\mathbf{R}})$.

Substituting (A.6) into (A.2) leads to the compressed LL function

$$\mathcal{L}(\boldsymbol{\theta}) = \ln\left(\det\left(\breve{\mathbf{R}}(\boldsymbol{\theta})\right)\right) + \text{trace}\left\{\breve{\mathbf{R}}^{-1}(\boldsymbol{\theta})\hat{\mathbf{R}}\right\} \tag{A.7}$$

where $\breve{\mathbf{R}}(\boldsymbol{\theta}) = \mathbf{R}(\boldsymbol{\theta}, \hat{\boldsymbol{p}})$. Consequently, estimation of the parameters reduces to solving the following optimization problem

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left\{\mathcal{L}(\boldsymbol{\theta})\right\}. \tag{A.8}$$

A.2 Extension to 2D: Application to Low-Flying Target Localization

Target localization and tracking in radar systems involve the estimation of Doppler frequencies, DOAs in azimuth/elevation, and TDOAs. Tracking low-flying targets is

usually complicated because the received echo signal reaches the radar site via two or more paths. This is due to the close proximity of the target to a reflecting surface. A detailed model for the surface reflection in different situations has been established in [8], [110]. The reflected signal consists of both specular and diffuse components. When the ray grazing angle is small, the specular component dominates. In the case of a two path scenario, the radar spots two targets, i.e., the true target and its image created by the reflected echo. For an elevation measurement, the reflected image usually has an angular position that appears to be below the surface, and a range difference often less than the range resolution of the radar. In addition, the direct and reflected signals add constructively or destructively due to their phase difference. Another problem that arises in low-angle tracking is the small angular separation between the impinging echoes (in terms of the beamwidth) [8]. Making use of classical beamforming to spatially separate the two signals is therefore problematic.

The various techniques that have been proposed to solve the problem include parametric approaches where the problem at hand can be looked at in three different ways. First, the target and its image can be considered as two separate sources, second, the target and its image can be linked via a geometric relation and third, some *a priori* knowledge of the reflection coefficients can be used. These techniques are basically ML estimators [52]. The advantage of ML is that it is less sensitive to signal coherence as compared to subspace decomposition techniques, provided that the stochastic properties of the data are known.

Introducing *a priori* knowledge in ML considerably improves the performance [62]. This knowledge is derived from a highly deterministic multipath model, using geometric information and a set of physical parameters like the refractivity gradient, the reflection coefficient, the specular and diffuse scattering coefficients, and the divergence factor [15]. Based on this information, a refined model (RM) has been developed for low angle tracking [15, 96], leading to a refined ML (RML) technique, with a variant accounting for incoherent multipath signals [15]. Compared to the classical DOA model, with unknown pairs of angles, amplitudes and phases, the RM model has fewer unknown parameters (signal amplitude, signal phase, and height of the target) as the other parameters (range, height of the receiver, grazing angle, etc) are expressed in terms of the unknowns. This model assumes known environment parameters and spatially white noise and/or nonfluctuating complex amplitudes that

account for reflection coefficients and radar cross section (RCS) factors [62, 96]. In practice however, the environment parameters are difficult to obtain or unavailable.

In what follows, we summarize the problem of localization for low-elevation targets. We consider a specular reflection model, which is directional and follows the laws of classical optics. The RCS is assumed to be fluctuating from one pulse repetition interval (PRI) to another. We address joint ML estimation of DOAs and Doppler frequencies in the presence of a partially unknown noise field. The proposed approach uses *a priori* knowledge on the structure of the noise as it employs a linear expansion of its covariance matrix [14].

A.3 Data Model

~~Consider~~ a ULA of M identical antenna elements. We assume that the radar observes its received waveforms over K successive PRIs. The following scenario is considered: The radar receives a direct signal and a reflected echo from unknown angles.

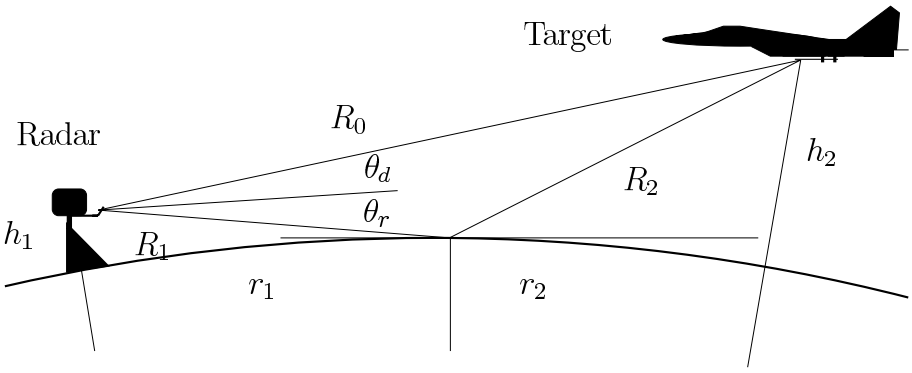

Figure A.1: Reflection model for a spherical smooth earth.

Thus, for $m = 1, \ldots, M$, the output of each antenna element can be given by

$$x_m(t) = A_{d_m}\xi_d s_d(t) + A_{r_m}\xi_r s_r(t) + n_m(t). \tag{A.9}$$

where A_{d_m} and A_{r_m} are the antenna responses to the direct and the reflected signals, $s_d(t)$ and $s_r(t)$, respectively. ξ_d and ξ_r are complex factors related to both signals and accounting for the RCS of the target [11], and $n_m(t)$ is the additive noise.

In our low-elevation scenario, for a small grazing angle, the direct and reflected echoes fall in the same range gate, and the two signals differ by their amplitudes and phase shifts. Hence, considering the set of all M antenna elements, using vector notation, the array output can be written as

$$\mathbf{x}(t) = \mathbf{A}(\boldsymbol{\theta})\mathrm{diag}(\boldsymbol{\xi})\mathbf{s}(t, \boldsymbol{f}) + \mathbf{n}(t) \tag{A.10}$$

where $\boldsymbol{\xi} = [\xi_d \; \xi_r]^T$, and $\mathbf{s}(t, \boldsymbol{f}) = \left[e^{j2\pi f_d t} \quad e^{j2\pi f_r t}\right]^T$. In the case of a ULA with element spacing of half-wavelength, the spatial steering matrix is defined as $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{A}_d \; \mathbf{A}_r]$ with

$$\mathbf{A}_d = \left[1 \quad e^{-j\pi\cos(\theta_d)} \ldots e^{-j(M-1)\pi\cos(\theta_d)}\right]^T \tag{A.11}$$

$$\mathbf{A}_r = \left[1 \quad e^{-j\pi\cos(\theta_r)} \ldots e^{-j(M-1)\pi\cos\theta_r}\right]^T \tag{A.12}$$

$$\boldsymbol{\theta} = [\theta_d \; \theta_r]^T \qquad \text{and} \qquad \boldsymbol{f} = [f_d \; f_r]^T \tag{A.13}$$

where θ_d, f_d and θ_r, f_r are the DOA and Doppler frequency of the direct and reflected signals, respectively.

The noise $\mathbf{n}(t)$ depends on physical variables such as thermal noise, clutter and possibly jammer interference.

A.3.1 Model Transformation

The samples under consideration are available in three processing dimensions, allowing the discrimination of multiple echoes in azimuth/elevation, range and Doppler frequency. In our case, we are interested in the joint DOA and Doppler frequency estimation, while the third dimension (range) is used to reduce the effect of the power nulling problem of the signals received from the low-elevation target [11]. We use several consecutive PRIs ($K \geq 2$) to calculate the estimates. We assume for simplicity that the DOAs and Doppler frequencies are invariant throughout the considered PRIs[1], while the complex factors $\boldsymbol{\xi}_k$, $k = 0, \ldots, K-1$, vary from one PRI to another.

Over each of the K PRIs, we assume that $\mathbf{x}(t)$ is sampled at a high rate, and that L samples are collected. The observation samples are denoted by $x_m(t_{l,k})$, with

[1]The PRI is assumed to be comprised within the coherent pulse interval (CPI).

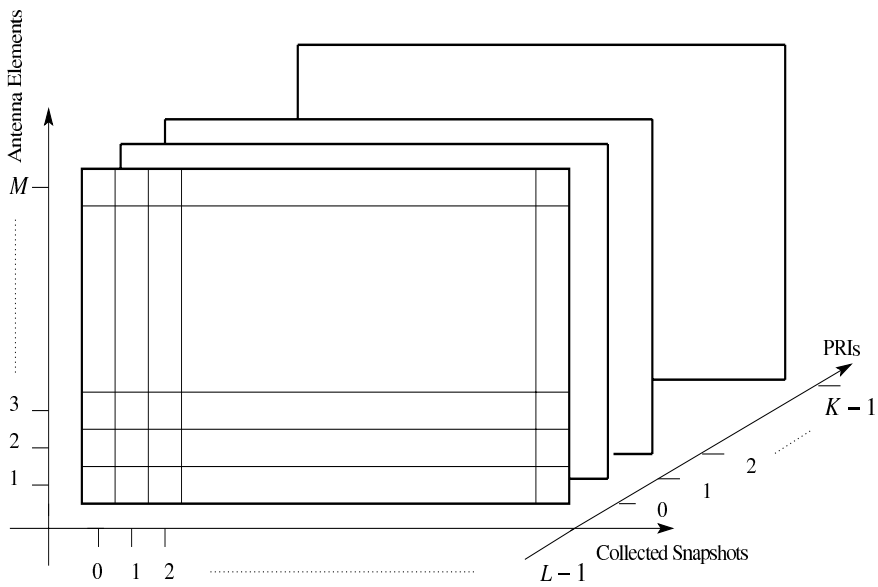

Figure A.2: Data processing over three dimensions.

$m = 1, \ldots, M$; $l = 0, \ldots, L-1$ and $k = 0, \ldots, K$. In other words, $x_m(t_{l,k})$ is the l-th sample of the m-th antenna output over the k-th PRI.

Suppressing the dependence on $\boldsymbol{\theta}$ and $\boldsymbol{f}$, the received data over the k-th PRI can be arranged in matrix form as follows

$$\mathbf{X}_k = \mathbf{A}\ \mathrm{diag}(\boldsymbol{\xi}_k)\ \mathrm{diag}(\boldsymbol{\phi}_k)\mathbf{F} + \mathbf{N}_k \tag{A.14}$$

where

$$\boldsymbol{\phi}_k := \begin{bmatrix} \phi_d^{(k-1)T_R} & \phi_r^{(k-1)T_R} \end{bmatrix}^T \tag{A.15}$$

and

$$\mathbf{F} := \begin{bmatrix} 1 & \phi_d^{\Delta t} & \ldots & \phi_d^{(l-1)\Delta t} & \ldots & \phi_d^{(L-1)\Delta t} \\ 1 & \phi_r^{\Delta t} & \ldots & \phi_r^{(l-1)\Delta t} & \ldots & \phi_r^{(L-1)\Delta t} \end{bmatrix} \tag{A.16}$$

with

$$\phi_d = e^{j2\pi f_d} \quad \text{and} \quad \phi_r = e^{j2\pi f_r} \tag{A.17}$$

and

$$\mathbf{N}_k = \left[\mathbf{n}(t_{0,k}), \ldots, \mathbf{n}(t_{(L-1),k})\right]. \tag{A.18}$$

After sampling, the columns of $\mathbf{X}_k$ are concatenated in a LM-dimensional vector $\mathbf{y}_k$ as follows:

$$\mathbf{y}_k = \mathrm{vec}(\mathbf{X}_k). \tag{A.19}$$

For arbitrary matrices $\mathbf{A}$, $\mathbf{C}$ and a diagonal matrix $\mathrm{diag}(\mathbf{b})$, it can be readily verified that $\mathrm{vec}(\mathbf{A}\mathrm{diag}(\mathbf{b})\mathbf{C}) = \left(\mathbf{C}^T \circ \mathbf{A}\right)\mathbf{b}$, where $\circ$ denotes Khatri-Rao product (a column-wise Kronecker product). Hence, it is straightforward to write (A.14) as

$$\begin{aligned} \mathbf{y}_k &= \mathrm{vec}\left\{\mathbf{A}\,\mathrm{diag}(\boldsymbol{\xi}_k)\,\mathrm{diag}(\boldsymbol{\phi}_k)\,\mathbf{F} + \mathbf{N}_k\right\} \\ &= \left(\mathbf{F}^T \circ \mathbf{A}\right)\left(\boldsymbol{\xi}_k \odot \boldsymbol{\phi}_k\right) + \mathrm{vec}(\mathbf{N}_k). \end{aligned} \tag{A.20}$$

Repeating the same operation over all the K PRIs and arranging the obtained data vectors in a matrix leads to

$$\begin{aligned} \mathbf{Y} &= \begin{bmatrix} \mathbf{y}_1 & \dots & \mathbf{y}_K \end{bmatrix} \\ &= \mathbf{U}(\boldsymbol{\theta}, \boldsymbol{f})\mathbf{H} + \boldsymbol{\Upsilon} \end{aligned} \tag{A.21}$$

where the space-frequency steering matrix is defined as $\mathbf{U}(\boldsymbol{\theta}, \boldsymbol{f}) = \mathbf{F}^T \circ \mathbf{A}$, and $\mathbf{H} = \boldsymbol{\Xi} \odot \boldsymbol{\Phi}$, with $\boldsymbol{\Xi} = [\boldsymbol{\xi}_1 \dots \boldsymbol{\xi}_K]$. The effect of the additive noise is given by $\boldsymbol{\Upsilon} = [\mathrm{vec}(\mathbf{N}_1) \dots \mathrm{vec}(\mathbf{N}_K)]$. Depending on whether the environment parameters are assumed steady or fluctuating, $\mathbf{H}$ is modeled as deterministic unknown or as random with probability densities described in [95]. For simplicity and without loss of generality, we focus only on the Swerling I case [95]. Thus, the pulse amplitude over the collected PRIs is considered as a single random variable with a Rayleigh distribution. The initial pulse phase is assumed to be uniformly-distributed over [0 2π].

The unknown parameters $\boldsymbol{\theta}$ and $\boldsymbol{f}$ depend on the array geometry and the target velocity, respectively. As they vary, matrix $\mathbf{U}$ in (A.21) describes a space-frequency manifold. Note that it is straightforward to extend the model to the case of more than 2 paths. Processing the data received over several PRIs exploits the diversity among the multidimensional measurements and reduces the effect of power nulling in the scenario of a low grazing angle, making the estimation of more paths than antennas possible. Indeed, if the number of paths is p, the dimension of the manifold matrix $\mathbf{U}$ is $LM \times p$, and the necessary condition on $\mathbf{U}$ for the identifiability of the unknown parameters is $LM > p$ and $\mathbf{U}$ is full rank, p. Thus, if $p > 2$, only

$M = 2$ antennas are sufficient to recover the unknown paths, provided the number of samples over each PRI, L, is high enough.

In what follows we assume that the data is modeled as a Gaussian stochastic process, with a covariance matrix $\mathbf{R}$, given by

$$\mathbf{R} = E\left\{\mathbf{y}_k\mathbf{y}_k^H\right\} = \mathbf{U}\mathbf{R}_{\mathbf{H}}\mathbf{U}^H + \mathbf{Q}(\mathbf{q}) \tag{A.22}$$

where $\mathbf{Q}(\mathbf{q})$ is the covariance matrix of the noise $\boldsymbol{\Upsilon}$, parameterized by the vector of unknown real elements $\mathbf{q}$. The following assumptions on the noise samples $\boldsymbol{\upsilon}_k$ (columns of $\boldsymbol{\Upsilon}$) are considered

$$E\left\{\boldsymbol{\upsilon}_k\boldsymbol{\upsilon}_l^H\right\} = \mathbf{Q}(\mathbf{q})\delta_{kl} \;\; ; \qquad E\left\{\boldsymbol{\upsilon}_k^T\boldsymbol{\upsilon}_l\right\} = 0. \tag{A.23}$$

Above, $\mathbf{R}_{\mathbf{H}}$ is the covariance matrix of the unknown scaled samples in $\mathbf{H}$, collected over the K PRIs. The columns $\mathbf{h}_k$ of $\mathbf{H}$ are assumed to be independent from the noise and satisfy the following:

$$E\left\{\mathbf{h}_k\mathbf{h}_l^H\right\} = \mathbf{R}_{\mathbf{H}}\delta_{kl} \;\; ; \qquad E\left\{\mathbf{h}_k^T\mathbf{h}_l\right\} = 0. \tag{A.24}$$

A.4 Joint Maximum Likelihood Estimation

Based on the above assumptions, the unconstrained negative LL function of the observed data, after normalization and omitting constant terms, is given by [42, 81]

$$\mathcal{L}(\boldsymbol{\eta}) = \ln\left(\det\left(\mathbf{R}(\boldsymbol{\eta})\right)\right) + \mathrm{trace}\{\mathbf{R}^{-1}(\boldsymbol{\eta})\hat{\mathbf{R}}\} \tag{A.25}$$

where, for the transformed model case, the sample covariance matrix $\hat{\mathbf{R}}$ is given by

$$\begin{aligned}\hat{\mathbf{R}} &:= \frac{1}{K}\sum_{k=1}^{K}\mathbf{y}_k\mathbf{y}_k^H \\ &= \frac{1}{K}\mathbf{Y}\mathbf{Y}^H\end{aligned} \tag{A.26}$$

and $\boldsymbol{\eta} = [\boldsymbol{\theta}^T, \boldsymbol{f}^T, \boldsymbol{\rho}^T, \mathbf{q}^T]^T$ is the vector of unknown real parameters, with $\boldsymbol{\rho}$ being the vector of real entries (real and imaginary parts) of the elements of $\mathbf{R}_{\mathbf{H}}$. At this stage, we seek a further simplified LL cost function by replacing $\mathbf{Q}(\mathbf{q})$ by a favorable estimate $\hat{\mathbf{Q}}(\mathbf{q})$.

Applying the linear expansion (A.1) to matrix $\mathbf{Q}(\mathbf{q})$ (where we have $2LM-1$ real entries instead of $2M-1$), and replacing matrix $\boldsymbol{\mathcal{A}}(\boldsymbol{\theta})$ in (A.3), by the 2D-parametrized matrix $\boldsymbol{\mathcal{U}}(\boldsymbol{\theta},\boldsymbol{f})$, defined as

$$\boldsymbol{\mathcal{U}}(\boldsymbol{\theta},\boldsymbol{f}) := \mathbf{U}^*(\boldsymbol{\theta},\boldsymbol{f}) \otimes \mathbf{U}(\boldsymbol{\theta},\boldsymbol{f}) \tag{A.27}$$

we easily arrive to the following modified negative LL function:

$$\mathcal{L}(\boldsymbol{\theta},\boldsymbol{f}) = \ln\left(\det\left(\breve{\mathbf{R}}(\boldsymbol{\theta},\boldsymbol{f})\right)\right) + \operatorname{trace}\left\{\breve{\mathbf{R}}^{-1}(\boldsymbol{\theta},\boldsymbol{f})\hat{\mathbf{R}}\right\} \tag{A.28}$$

and estimation of the unknown parameter-pairs of interest reduces to solving the following optimization problem:

$$\hat{\boldsymbol{\theta}},\hat{\boldsymbol{f}} = \arg\min_{\boldsymbol{\theta},\boldsymbol{f}} \left\{\mathcal{L}(\boldsymbol{\theta},\boldsymbol{f})\right\}. \tag{A.29}$$

A.5 Cramér-Rao Bound

Derivation of the stochastic CRB on the unknown pairs of interest, corresponding to the model (A.2) follows from the general analysis of [39] and is conveniently based on the linear expansion of the noise covariance matrix in (A.19).

Lemma A.5.1. *Under the assumptions of Section A.3, the stochastic CRB on the pairs* $(\boldsymbol{\theta},\boldsymbol{f})$ *given the model (A.14) is given by*

$$CRB_{sto}(\boldsymbol{\theta},\boldsymbol{f}) = \frac{1}{K}\left(\boldsymbol{\mathfrak{L}} - \boldsymbol{\mathfrak{M}}\boldsymbol{\mathfrak{T}}^{-1}\boldsymbol{\mathfrak{M}}\right)^{-1} \tag{A.30}$$

with

$$\boldsymbol{\mathfrak{L}} = 2\Re\left\{\left[\mathbf{1}_{22}\otimes\left(\mathbf{R}_{\mathbf{H}}\tilde{\mathbf{U}}^H\tilde{\mathbf{R}}^{-1}\tilde{\mathbf{U}}\mathbf{R}_{\mathbf{H}}\right)\right]^T \odot \left[\tilde{\mathbf{D}}^H\boldsymbol{\Pi}^{\perp}_{\tilde{\mathbf{U}}}\tilde{\mathbf{D}}\right]\right\} \tag{A.31}$$

$$\boldsymbol{\mathfrak{M}} = 2\Re\left\{(\mathbf{1}_2\otimes\boldsymbol{\mathcal{I}})^T\left(\tilde{\mathbf{D}}^H\boldsymbol{\Pi}^{\perp}_{\tilde{\mathbf{U}}}\right)\otimes\left(\mathbf{R}_{\mathbf{H}}{}^T\tilde{\mathbf{U}}^T\tilde{\mathbf{R}}^{-T}\right)\tilde{\boldsymbol{\mathcal{P}}}^*\right\} \tag{A.32}$$

$$\boldsymbol{\mathfrak{T}} = 2\Re\left\{\tilde{\boldsymbol{\mathcal{P}}}^H\left(\tilde{\mathbf{R}}^{-T}\otimes\boldsymbol{\Pi}^{\perp}_{\tilde{\mathbf{U}}}\right)\boldsymbol{\mathcal{P}}\right\} - \boldsymbol{\mathcal{P}}^H\left[\boldsymbol{\Pi}^{\perp T}_{\tilde{\mathbf{U}}}\otimes\boldsymbol{\Pi}^{\perp}_{\tilde{\mathbf{U}}}\right]\boldsymbol{\mathcal{P}} \tag{A.33}$$

where

$$\tilde{\mathbf{R}} = \mathbf{Q}^{-1/2}\mathbf{R}\mathbf{Q}^{-1/2} \tag{A.34}$$

$$\tilde{\mathbf{U}} = \mathbf{Q}^{-1/2}\mathbf{U} \tag{A.35}$$

$$\tilde{\mathbf{D}} = \mathbf{Q}^{-1/2}\mathbf{D} \tag{A.36}$$

$$\tilde{\boldsymbol{\mathcal{P}}} = \left[vec(\tilde{\boldsymbol{\Psi}}_1) \;\; \ldots \;\; vec(\tilde{\boldsymbol{\Psi}}_{2LM-1})\right] \tag{A.37}$$

$$\tilde{\boldsymbol{\Psi}}_i = \mathbf{Q}^{-1/2}\boldsymbol{\Psi}_i\mathbf{Q}^{-1/2}. \tag{A.38}$$

Proof. See Appendix B.9. □

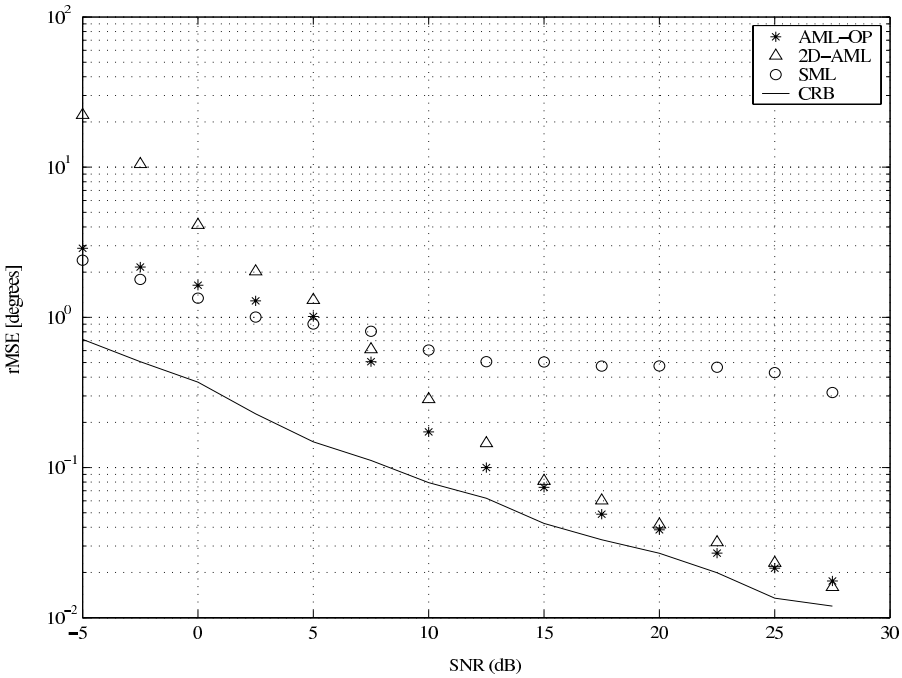

Figure A.3: rMSE of the DOA vs SNR for the different methods. $M = 12$, $K = 200$, $L = 3$, $\rho = 0.9$.

A.6 Numerical Examples

We use the following simulation scenario: Assume a 4/3-earth model [11]; a radar over a smooth earth is operating at a height of 3 m. The array is a ULA of M sensors. The target is at a distance of 12 km, flying at constant altitude of 200 m. The range is assumed roughly known. L is set to a value around the range bin of the target echo[2].

The actual noise covariance matrix is modeled as $\mathbf{Q}_{\mathrm{m,n}} = \sigma^2 \rho^{|\mathrm{m-n}|} e^{j\zeta(\mathrm{m-n})}$, where ρ is the correlation coefficient and ζ is a scaling factor.

In the following, the noise vectors $\boldsymbol{\upsilon}_k$ in (A.21) are modeled as a combination of two parts as follows:

$$\begin{aligned} \boldsymbol{\upsilon}_k &:= \mathbf{z}_k + \mathbf{w}_k \\ &= \mathbf{G}\boldsymbol{\omega}_k + \mathbf{w}_k. \end{aligned} \tag{A.39}$$

The first part $\mathbf{w}_k$, is an unstructured noise generated internally by the electronics

[2]The method proposed in this Appendix serves as a benchmark to the AML-OP estimator [29], which is derived under the scope of [28].

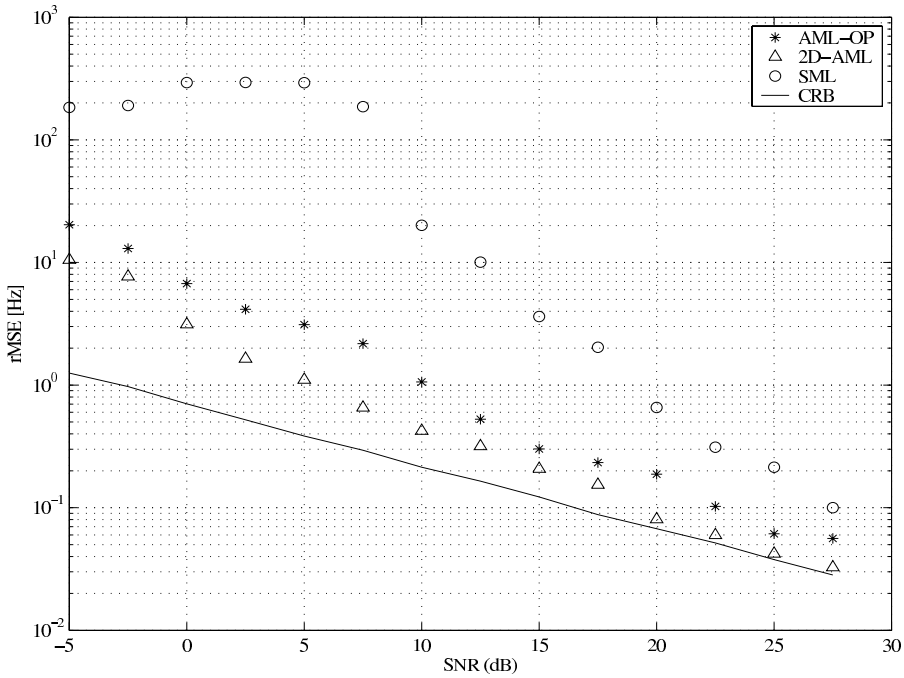

Figure A.4: rMSE of the Doppler frequency vs SNR for the different methods. $M = 12$, $K = 200$, $L = 3$, $\rho = 0.9$.

of the receiver. The second part is an external structured noise $\mathbf{z}_k$, resulting from filtering an unknown white process $\boldsymbol{\omega}_k$ through a known linear system $\mathbf{G} \in \mathbb{C}^{(LM \times P)}$, with $LM \geq P$ and P known. This linear process can be regarded as a set of P base matrices spanning the noise subspace (structured). Matrix $\mathbf{G}$ models the effect of any non-noiselike signal that interferes with the transmitted pulse. It can be estimated off-line. This can be done, for example, using known target-free signals or, a known land clutter (environment) model. For instance, it is shown in [84] that low-angle land clutter is non-noiselike. Note that knowledge of $\mathbf{G}$ up to a unitary matrix, say $\mathbf{T}$, is sufficient to retrieve the structured noise subspace, the latter being excited by a process $\tilde{\boldsymbol{\omega}}_k = \mathbf{T}^{-1}\boldsymbol{\omega}_k$ which is also white.

In the simulations, the unstructured noise component $\mathbf{w}_k$ is modeled as a white noise with variance $\sigma_{\mathbf{w}}^2$, where $\sigma_{\mathbf{w}} < \sigma$. The structure of the noise covariance matrix is of the form $\mathbf{Q} = \mathbf{G}\mathbf{G}^H + \sigma_{\mathbf{w}}^2\mathbf{I}$. Thus, the process $\mathbf{G}$ is chosen to be the square-root matrix of $\mathbf{Q} - \sigma_{\mathbf{w}}^2\mathbf{I}$. Note that this is not a unique choice. In practice $\mathbf{G}$ can be estimated up to a unitary matrix from known target-free signals or known land-clutter models. The value of the SNR is $10\log_{10}(\sigma_{\mathbf{s}}^2/\sigma^2)$, where $\sigma_{\mathbf{s}}^2$ is the signal

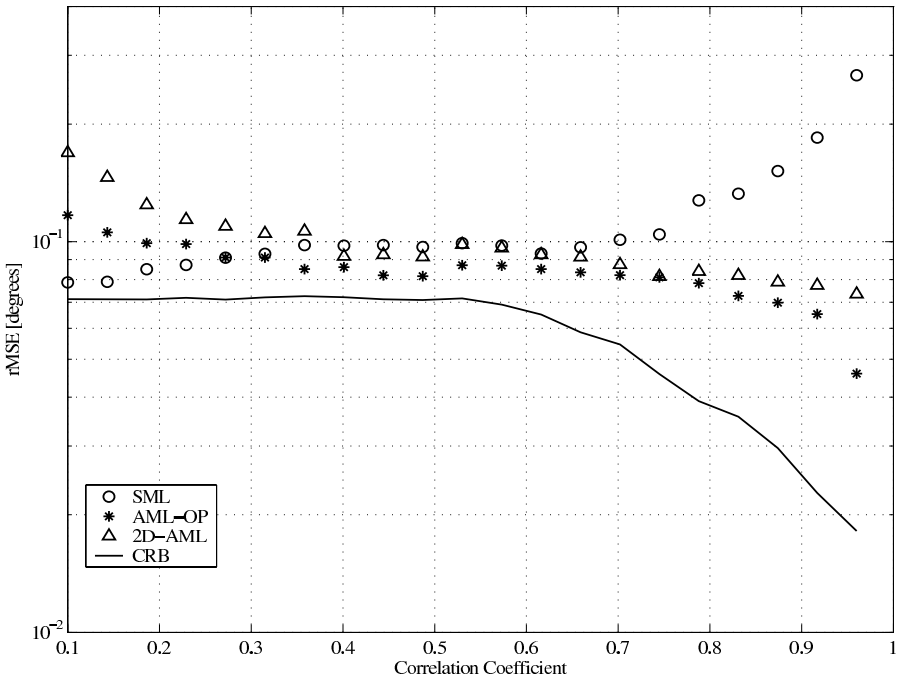

Figure A.5: rMSE of the DOA vs the correlation coefficient for the different methods. $M = 14$, $K = 200$, $L = 3$, SNR $= 10$ dB.

power.

Alternatively, the noise vector $\mathbf{q}$ clearly contains $2LM - 1$ real elements, i.e., σ^2 for the main diagonal of $\mathbf{Q}$, $\sigma^2\rho^i\cos(i\zeta)$ for the real part of the i-th super- and sub-diagonals of $\mathbf{Q}$, and $\pm\sigma^2\rho^i\sin(i\zeta)$ for the imaginary part of the i-th super- and sub-diagonals of $\mathbf{Q}$, respectively, with $i = 1, \ldots, LM - 1$. The corresponding base matrices $\mathbf{\Psi}_m$, $m = 1, \ldots, 2LM - 1$, can be easily deduced to have diagonals composed of zeros, 1, and $\pm j$, as described in [37].

We illustrate the performance of the proposed approach in terms of the rMSE versus the SNR, the spatial correlation coefficient ρ, the number of collected PRIs K, and the number of sensors M. We also illustrate the effect of the range on the DOA tracking. We compare the performance of the proposed technique to the classical SML (see Section 2.5.1), where the noise is assumed spatially and temporally white, as well as to the Oblique-Projector AML (AML-OP) [29] and the stochastic CRB, which is provided by (A.30). The parameter sets are indicated in the corresponding figure captions. The obtained results are averaged over 200 Monte Carlo runs. The results for the direct and reflected echoes are similar, therefore only those

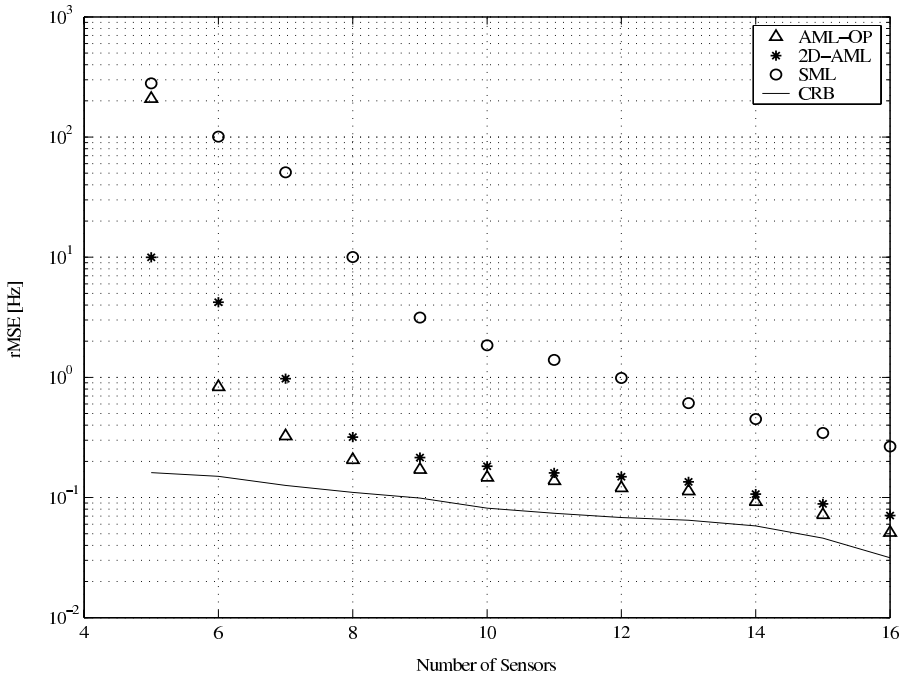

Figure A.6: rMSE of the Doppler frequency vs the number of sensors for the different methods. $M = 400$, $\rho = 0.95$, $L = 3$, SNR $= 10$ dB.

corresponding to the reflected echo are shown.

- Figures A.3 and A.4 illustrate the performance of AML-OP and 2D-AML versus SNR, in terms of rMSE. The AML-OP performs slightly better than the 2D-AML for DOA estimation and inversely, 2D-AML outperforms the AML-OP for Doppler frequency estimation. Overall, the AML-OP is observed to perform similarly to SML for SNR$<$ 10dB and similarly to 2D-AML for SNR$>$ 10dB. Also, AML-OP outperforms 2D-AML for low SNR. For the above settings the two estimators do not reach the CRB as the results show a bias in the estimates.

- Figure A.5 shows the variation of the rMSE of the estimated DOA as the correlation coefficient ρ varies. Note that SML diverges as ρ approaches one. As expected, AML-OP and 2D-AML perform better at higher values of ρ. Variation of the rMSE of the Doppler frequency is similar, therefore it is not shown for conciseness purposes. More generally, the same remarks as for the previous example can be made here, especially with respect to the relative

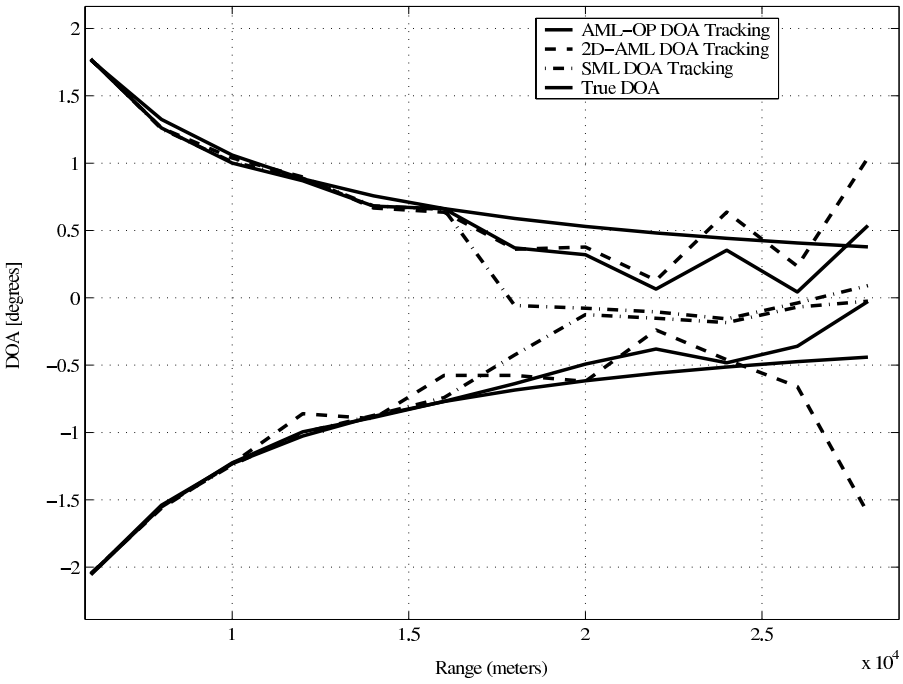

Figure A.7: Tracking of the direct and reflected DOAs for the different methods. $M = 12$, $K = 200$, $\rho = 0.95$, $L = 3$, SNR $= 10$ dB.

performance of AML-OP and 2D-AML.

- Figure A.6 illustrates the performance of the Doppler frequency estimation as we increase the number of antenna elements. Similarly, the same remarks as above apply for this example too.

- Figure A.7 shows the improvement of the performance as the angle difference between the direct and reflected paths increases (as the target approaches the radar site). The other parameters are kept fixed. Observe that 2D-AML and AML-OP still provide separable DOAs at any range unlike SML which fails to separate two DOAs that are merged around 0° at a range greater than 20 km.

Appendix B

Proofs of Lemmas and Results

B.1 Proof of Lemma 3.3.1

The proof of Lemma 3.3.1 partly follows from the development of [104]. Consider the following definition:

$$\cos(\phi_m(p)) = \frac{\left|\mathbf{e}_m^H \hat{\mathbf{e}}_m\right|}{\|\mathbf{e}_m\| \, \|\hat{\mathbf{e}}_m\|} = \left|\mathbf{e}_m^H \hat{\mathbf{e}}_m\right| \tag{B.1}$$

and let $\tilde{\mathbf{e}}_m$ denote the unnormalized sample eigenvector, such that $\hat{\mathbf{e}}_m = \tilde{\mathbf{e}}_m / \|\tilde{\mathbf{e}}_m\|$. Thus, we have

$$\cos(\phi_m(p)) = \left|\mathbf{e}_m^H \tilde{\mathbf{e}}_m\right| / \|\tilde{\mathbf{e}}_m\|. \tag{B.2}$$

To proceed with the proof, we need to consider separately the two cases, $m = 1, \ldots, p$, and $m = p+1, \ldots, M$.

1. <u>$m = 1, \ldots, p$:</u>

   For $m = 1, \ldots, p$, it is shown in [2] that when $L \to \infty$, the sample eigenvectors $\tilde{\mathbf{e}}_m$ are asymptotically independent and satisfy

   $$E\left\{\tilde{\mathbf{e}}_m\right\} = \mathbf{e}_m + O(L^{-1}) \tag{B.3}$$

   and the covariance matrix of $\tilde{\mathbf{e}}_m$ is

   $$\begin{aligned} \operatorname{cov}\left\{\tilde{\mathbf{e}}_m\right\} &= \frac{1}{L} \sum_{\substack{q=1 \\ q \neq m}}^{M} \frac{\lambda_m \lambda_q}{(\lambda_m - \lambda_q)^2} \mathbf{e}_q \mathbf{e}_q^H + O(L^{-2}) \\ &\quad \text{for } m = 1, \ldots, p \end{aligned} \tag{B.4}$$

Note that under H_p, for $m = 1, \ldots, p$, the rank of cov $\{\tilde{\mathbf{e}}_m\}$ is $M-1$ [104], due to $p-1$ distinct eigenvalues given by $(q \neq m)$

$$l_q = \frac{\lambda_m \lambda_q}{L(\lambda_m - \lambda_q)^2} \tag{B.5}$$

and one eigenvalue of multiplicity $M-p$, given by

$$l_p = \frac{\lambda_m \sigma^2}{L(\lambda_m - \sigma^2)^2}. \tag{B.6}$$

From (B.4), it is clear that $\lim_{L\to\infty} |\mathbf{e}_m \tilde{\mathbf{e}}_m| = 1$, and $\lim_{L\to\infty} \cos(\phi_m(p)) = \|\tilde{\mathbf{e}}_m\|^{-1}$. Moreover, from the theory of sampling distributions and quadratic forms [20, 79], we have

$$\tilde{\mathbf{e}}_m^H \tilde{\mathbf{e}}_m \xrightarrow{\mathcal{D}} X_{\mathbf{e}} \tag{B.7}$$

where

$$X_{\mathbf{e}} \sim 1 + \sum_{\substack{q=1 \\ q\neq m}}^{p} l_q \chi_2^2 + l_p \chi_{2(M-p)}^2. \tag{B.8}$$

It follows that when $L \to \infty$, we have

$$\cos(\phi_m(p)) \xrightarrow{\mathcal{D}} X_{\phi_m} \tag{B.9}$$

where

$$X_{\phi_m} \sim \left[1 + \sum_{\substack{q=1 \\ q\neq m}}^{p} l_q \chi_2^2 + l_p \chi_{2(M-p)}^2 \right]^{-1/2}. \tag{B.10}$$

Since for a large L, a large M $(M \gg p)$ and a relatively high SNR, the ideal value of $\cos(\phi_m(p))$ is 1, implying that $\sum_{q=1 q\neq m}^{p} l_q \chi_2^2 + l_p \chi_{2(M-p)}^2$ is very small, we can use the following approximation:

$$\cos(\phi_m(p)) \xrightarrow{\mathcal{D}} X_{\phi} \tag{B.11}$$

where

$$X_{\phi} \sim 1 - \frac{1}{2} \sum_{\substack{q=1 \\ q\neq m}}^{p} l_q \chi_2^2 - \frac{1}{2} l_p \chi_{2(M-p)}^2 \tag{B.12}$$

and finally, using applying the generalized central limit theorem to a large set of χ^2_ν-distributed variables, with mean ν and variance 2ν, directly obtain, when $L \to \infty$, the asymptotic distribution (3.2), for $m = 1, \ldots, p$.

2. $m = p+1, \ldots, M$:

   Second, for $m = p+1, \ldots, M$, let $\mathbf{E}_n = [\mathbf{e}_{p+1}, \ldots, \mathbf{e}_M]$, i.e., the matrix whose columns are the eigenvectors corresponding to the eigenvalue of multiplicity greater than one. Similarly, consider the unnormalized matrix $\tilde{\mathbf{E}}_n = [\tilde{\mathbf{e}}_{p+1}, \ldots, \tilde{\mathbf{e}}_M]$.

   Unlike the signal eigenvectors, the columns of $\tilde{\mathbf{E}}_n$ are not asymptotically independent. However, given our particular case where only one eigenvalue is of multiplicity greater than one and is ideally equal to σ^2, it is possible to deduce asymptotic marginal distributions with enough accuracy.

   It is shown in [2] that under the assumption of Gaussian data, when $L \to \infty$, we have asymptotically

$$\tilde{\mathbf{E}}_n = \mathbf{E}_n \mathbf{\Delta} + L^{-1/2} \sum_{q=1}^{p} \mathbf{e}_q \mathbf{\Gamma}_q \tag{B.13}$$

   where the limiting distribution of $\mathbf{\Delta}$ as $L \to \infty$, is the conditional Haar invariant distribution, or equivalently 2^M times the Haar measure over orthogonal matrices with positive diagonal elements. Moreover, the limiting distribution of $\mathbf{\Gamma}_q$ as $L \to \infty$, is normal with zero mean and variance $\lambda_q \sigma^2 / (\lambda_q - \sigma^2)^2$, with $q = 1, \ldots, p$.

   Using the above, as $L \to \infty$, it is easy to verify that $\mathbf{\Delta}$ converges asymptotically to $\mathbf{I}_M$ and consequently, for $m = p+1, \ldots, M$, the sample eigenvalues satisfy marginally the following:

$$E\{\tilde{\mathbf{e}}_m\} = \mathbf{e}_m + O(L^{-1}) \tag{B.14}$$

   and

$$\begin{aligned} \operatorname{cov}\{\tilde{\mathbf{e}}_m\} &= \frac{1}{L} \sum_{q=1}^{p} \frac{\lambda_q \sigma^2}{(\lambda_q - \sigma^2)^2} \mathbf{e}_q \mathbf{e}_q^H + O(L^{-2}) \\ &\quad \text{for } m = p+1, \ldots, M. \end{aligned} \tag{B.15}$$

Hence, similarly to the previous case with $m = 1, \ldots, p$, from (B.15) one can see that under H_p, for $m = p+1, \ldots, M$, the rank of $\mathrm{cov}\{\tilde{\mathbf{e}}_m\}$ is p, due to p distinct eigenvalues given by $l_q = \lambda_q \sigma^2 / (L(\lambda_q - \sigma^2)^2)$ with $q = 1 \ldots, p$, and when $L \to \infty$, it leads to [20, 79]

$$\cos(\phi_m(p)) \xrightarrow{\mathcal{D}} X_{\phi_n} \tag{B.16}$$

where

$$X_{\phi_n} \sim \left[1 + \sum_{q=1}^{p} l_q \chi_2^2\right]^{-1/2}. \tag{B.17}$$

Using the same approximations as for $m = 1, \ldots, p$, along with the generalized central limit theorem, when $L \to \infty$, the asymptotic distribution (3.6) follows for the case $m = p+1, \ldots, M$.

B.2 Proof of Corollary 3.3.2

Ideally, under H_p, if the noise subspace is completely disjoint from any signal attributes, the angle between $\mathbf{e}_m$, $m = p+1, \ldots, M$, and a subspace defined only by the signal parameters must be $\pi/2$. Thus, using the unnormalized sample eigenvectors $\tilde{\mathbf{e}}_m$, it can be easily deduced that $\gamma_m(p) = \pi/2 - \phi_m(p)$. This implies that

$$\cos^2(\gamma_m(p)) = 1 - \cos^2(\phi_m(p)). \tag{B.18}$$

Hence, form (B.17), when $L \to \infty$, we have

$$1 - \cos^2(\gamma_m(p)) \xrightarrow{\mathcal{D}} X_i \tag{B.19}$$

where

$$X_i \sim \left[1 + \sum_{q=1}^{p} l_q \chi_2^2\right]^{-1}. \tag{B.20}$$

For a large L, a large M ($M \gg p$) and a relatively high SNR, the value of $\cos(\gamma_m(p))$ is very small. Also, using the approximation leading from (B.10) to (B.12), i.e., $\sum_{q=1}^{p} l_q \chi_2^2$ being small, we can approximate (B.20) by

$$1 - 2\cos(\gamma_m(p)) \xrightarrow{\mathcal{D}} X_j \tag{B.21}$$

where

$$X_j \sim 1 - \sum_{q=1}^{p} l_q \chi_2^2 \tag{B.22}$$

or

$$\cos(\gamma_m(p)) \xrightarrow{\mathcal{D}} X_k \tag{B.23}$$

where

$$X_k \sim \frac{1}{2} \sum_{q=1}^{p} l_q \chi_2^2 \tag{B.24}$$

and finally using the central limit theorem, when $L \to \infty$, we obtain the asymptotic distribution (3.16), for $m = p+1, \ldots, M$.

B.3 Proof of Lemma 5.2.1

Similarly to [106], for some estimated source number $\hat{p}$, define the following difference

$$\Delta \;=\; (\Delta_{\mathcal{L}} + \Delta_{\mathcal{P}}) \tag{B.25}$$

with $\Delta_{\mathcal{L}} = \mathcal{L}^{(M)}(p) - \mathcal{L}^{(M)}(\hat{p})$ and $\Delta_{\mathcal{P}} = \mathcal{P}(p) - \mathcal{P}(\hat{p})$.

Assume first that $\hat{p} < p$ and consider $\lim_{L\to\infty} \Delta/L$. Since $r_{(M,M)}$, $\lambda_m^{(M)}$ and $\rho_m^{(M)}$, $m = 1, \ldots, \hat{p}, \ldots, p$, are all non-negative, $\mathcal{L}^{(M)}(p)$ in (5.5) is a monotonically decreasing function of p. As a result, $\lim_{L\to\infty} \Delta_{\mathcal{L}}/L < 0$ holds with probability one. Also, note that if $\Delta_{\mathcal{P}}$ is bounded with respect to L, then $\lim_{L\to\infty} \Delta_{\mathcal{P}}/L = 0$ with probability one. It follows that $\lim_{L\to\infty} \Delta < 0$ holds with probability one when $\hat{p} < p$. This is the case for both NU-MDL$^{(M)}(p)$ and NU-AIC$^{(M)}(p)$ as defined in Section 5.2.

Assume next that $p < \hat{p}$ and consider $\lim_{L\to\infty} -2\Delta$. Under $\mathsf{H}_{\hat{p}}$, we have the vector of unknown parameters $\boldsymbol{\eta}_{\hat{p}} = [\hat{p}, \mathbf{p}_{\hat{p}}^T, \sigma^2]^T$. Consider the theoretical case of a similar array of M elements with spatially nonuniform noise and $\hat{p}$ sources, such that every source of index $q > p$, is received by a single array-sensor only, and the power of every such source added to the power of the noise at the corresponding array-sensor

is equal to σ^2. In this case, under $\mathsf{H}_{\hat{p}}$, the covariance matrix of the data $\mathbf{R}(\bar{\boldsymbol{\eta}}_{\hat{p}})$ is parameterized by the vector of unknowns $\bar{\boldsymbol{\eta}}_{\hat{p}} = [\hat{p}, \bar{\mathbf{p}}_{\hat{p}}^T, \sigma^2, \boldsymbol{\epsilon}^T]^T$, where $\boldsymbol{\epsilon}$ is the vector of appropriately assigned noise-power perturbations over the sensors.

Observe that the model where $\mathbf{R}(\boldsymbol{\eta}_p)$, with $\boldsymbol{\eta}_p = [p, \mathbf{p}_p^T, \sigma^2]^T$, with spatially uniform noise, is one special case of the above nonuniform noise model, and $\mathbf{R}(\boldsymbol{\eta}_p) = \mathbf{R}(\bar{\boldsymbol{\eta}}_{\hat{p}})$ due to the model ambiguity associated with noise nonuniformity [34], i.e., the joint distribution of the data in both cases is the same. At the same time, it follows that vectors $\boldsymbol{\eta}_{\hat{p}}$ and $\bar{\boldsymbol{\eta}}_{\hat{p}}$ are inner points of the same subspace $\boldsymbol{H}_{\hat{p}}$ of candidate models [117]. Consequently, the term $-2\Delta_{\mathcal{L}}$ is equivalent to the LR between two hierarchically nested models. Indeed, it is twice the LL of the ML parameter estimator under the hypotheses H_p and $\mathsf{H}_{\hat{p}}$, respectively. This LR statistic approximately follows a χ^2_ν distribution, with degrees of freedom ν equal to the number of additional parameters in the more complex model [106], i.e., the difference of the dimensions of the parameter spaces under the two hypotheses. In our case, this difference is $\nu = \hat{p} + \hat{p}^2 - (p + p^2)$. Thus, as the sample size L increases, the probability that $-2\Delta_{\mathcal{L}}$ exceeds $-2\Delta_{\mathcal{P}}$ in (B.25) is given by the area in the tail from $-2\Delta_{\mathcal{P}}$ of the $\chi^2_{(\hat{p}+\hat{p}^2-p-p^2)}$ distribution. Hence, if $\lim_{L\to\infty} -2\Delta_{\mathcal{P}} = \infty$, then the tail area of the $\chi^2_{(\hat{p}+\hat{p}^2-p-p^2)}$ distribution tends to zero. As a result, $\lim_{L\to\infty} -\Delta > 0$ holds with probability one when $p < \hat{p}$. This is clearly verified for NU-MDL$^{(M)}(p)$, but not for NU-AIC$^{(M)}(p)$ since the corresponding probability that $-\Delta$ is positive is not one as L tends to infinity when $p < \hat{p}$.

Combination of the two cases ($p > \hat{p}$ and $p < \hat{p}$) implies that NU-MDL$^{(M)}(\hat{p})$ is minimized for $\hat{p} = p$, as L increases, making it, as expected, a consistent source number estimator, whereas NU-AIC$^{(M)}(p)$ is not consistent.

B.4 Proof of Result 5.2.2

Using the expressions (5.5), (5.20) and the definition of $W(p)$ in (5.23), it is straightforward to see that for $p = 1, \ldots, M-2$

$$P_M(p) = P\left\{ \ln\left[1 - \frac{\left(\hat{\rho}_p^{(M)}\right)^2}{\hat{W}(p)}\right] > \frac{-\Delta\mathcal{P}(p)}{L} \,\middle|\, \mathsf{H}_p \right\}. \tag{B.26}$$

where $\hat{W}(p)$ is defined by (5.23) with the sample eigenvalues and sample Gerschgorin radii instead of their ideal counterparts.

Note that under H_p, for $m = 1, \ldots, p-1$, the values $\hat{r}_{(M,M)}$, $\hat{\rho}_m^{(M)}$ and $\hat{\lambda}_m^{(M)}$ all contain signal components and are significantly greater than zero. Also, given the asymptotic properties of the independent signal sample eigenvalues [104], it follows that asymptotically, as $L \to \infty$, $\hat{W}(p)$ converges to $W(p)$. Thus, asymptotically, we can write

$$P_M(p) = P\left\{ \ln\left[1 - \frac{\left(\hat{\rho}_p^{(M)}\right)^2}{W(p)}\right] > \frac{-\Delta\mathcal{P}(p)}{L} \middle| \mathsf{H}_p \right\}. \tag{B.27}$$

Under H_p, when $L \to \infty$, the value of $\hat{\rho}_p^{(M)}$ is very small (ambiguity leading to under-modeling). It results that the value of $\cos(\gamma_p(p))$ is also very small and asymptotically, it reasonably satisfies (3.16). Moreover, since $\hat{\rho}_p^{(M)}$ is small, it is possible to linearize $\ln(1-u)$ around $u_0 = 0$, and (B.27) can be approximated by

$$P_M(p) \approx P\left\{ \frac{\left(\hat{\rho}_p^{(M)}\right)^2}{W(p)} < \frac{\Delta\mathcal{P}(p)}{L} \middle| \mathsf{H}_p \right\} \tag{B.28}$$

or

$$P_M(p) = P\left\{ \cos(\gamma_p(p)) < \|\mathbf{r}^{(M)}\|^{-1}\sqrt{\frac{W(p)\Delta\mathcal{P}(p)}{L}} \middle| \mathsf{H}_p \right\}. \tag{B.29}$$

Hence, given the asymptotic distribution of (3.16), when $L \to \infty$, the expression of $P_M(p)$ in (5.20) follows directly.

In the same way, from (5.5), (5.20) and the definition of $W(p)$ in (5.23), for $p = 0, \ldots, M-2$, we have

$$P_F(p) = P\left\{ \ln\left[1 - \frac{\left(\hat{\rho}_{p+1}^{(M)}\right)^2}{W(p+1)}\right] < \frac{-\Delta\mathcal{P}(p+1)}{L} \middle| \mathsf{H}_p \right\}. \tag{B.30}$$

Under H_p, when $L \to \infty$, the value of $\hat{\rho}_{p+1}^{(M)}$ is very small, and similarly to (B.28), the above equation can be approximated by

$$P_F(p) \approx P\left\{ \frac{\left(\hat{\rho}_{p+1}^{(M)}\right)^2}{W(p+1)} > \frac{\Delta\mathcal{P}(p+1)}{L} \middle| \mathsf{H}_p \right\}. \tag{B.31}$$

Under H_p, recall that when $L \to \infty$, we have $\cos(\gamma_{p+1}(p)) \sim \mathcal{N}\left(\mu_\gamma(p), \varsigma_\gamma^2(p)\right)$, where $\lambda_p^{(M)} > \lambda_{p+1}^{(M)} = \sigma^2$. Thus, for $p = 1, \ldots, M-2$, we obtain

$$P_F(p) = P\left\{ \cos(\gamma_{p+1}(p)) > \|\hat{\mathbf{r}}^{(M)}\|^{-1} \sqrt{\frac{W(p+1)\Delta\mathcal{P}(p+1)}{L}} \middle| \mathsf{H}_p \right\}. \tag{B.32}$$

Hence, given the asymptotic distribution (3.16), for $p = 1, \ldots, M-2$, the expression of $P_F(p)$ in (5.21) is deduced directly.

When $p = 0$, it is not possible to divide by $\|\mathbf{r}^{(M)}\|$ as it is equal to zero. However, once can see that $\lim_{L\to\infty} \hat{\mathbf{R}}^{(M)} = \sigma^2 \mathbf{I}_{M-1}$. Therefore, we also have $\lim_{L\to\infty} \hat{\mathbf{e}}_m^{(M)} = \mathbf{i}_m$, for $m = 1, \ldots, M-1$, where $\mathbf{i}_m$ is an $(M-1)$-dimensional vector of zeros except for its m-th element which is equal to one. Thus, asymptotically, we can reasonably approximate $\hat{c}_1^{(M)}$ by $\hat{r}_{(1,M)}$ as given by (4.3), bearing in mind that the elements $\hat{c}_m^{(M)}$ are ordered according to their magnitudes as in (4.10). Moreover, considering the normal product distribution (4.18), when $L \to \infty$, we can use the generalized central limit theorem to asymptotically obtain

$$\hat{c}_1^{(M)} \approx \hat{r}_{(1,M)} \xrightarrow{\mathcal{D}} X_c \tag{B.33}$$

where

$$X_c \sim \mathcal{N}(0, \sigma^4/L). \tag{B.34}$$

Equivalently, since $|\hat{c}_1^{(M)}| = \hat{\rho}_1^{(M)}$, for the case of real additive noise and for a large L, we have asymptotically

$$\frac{L}{\sigma^4}\left(\hat{\rho}_1^{(M)}\right)^2 \xrightarrow{\mathcal{D}} X_\rho \tag{B.35}$$

where

$$X_\rho \sim \chi_1^2. \tag{B.36}$$

Consequently, $\hat{\rho}_1^{(M)}\sqrt{L}/\sigma^2$ has a χ_1 distribution with mean $\sqrt{2/\pi}$ and variance $(\pi - 2)/\pi$. It is also a half-normal distribution with parameter $\vartheta = \sqrt{\pi/2}$. Therefore, from (B.31), when $L \to \infty$, we have

$$P_F(0) = P\left\{ \frac{\sqrt{L}}{\sigma^2}\hat{\rho}_1^{(M)} > \frac{\sqrt{W(1)\Delta\mathcal{P}(1)}}{\sigma^2} \middle| \mathsf{H}_0 \right\} \tag{B.37}$$

and given the half-normal distribution of $\hat{\rho}_1^{(M)}\sqrt{L}/\sigma^2$, expression (5.22) follows.

B.5 Proof of Result 5.2.3

Observe that relation (5.26) can be reformulated as

$$\frac{\|\mathbf{r}^{(M)}\|^{-1}\sqrt{\Delta\mathcal{P}(2)W(2)/L}-\mu_\gamma(1)}{\varsigma_\gamma(1)} \geq \sqrt{2\Delta\mathcal{P}(2)} \tag{B.38}$$

or

$$\sqrt{\Delta\mathcal{P}(2)W(2)} \geq \frac{\|\mathbf{r}^{(M)}\|\lambda_1^{(M)}\sigma^2}{\sqrt{L}(\lambda_1^{(M)}-\sigma^2)^2}\left[1+\sqrt{2\Delta\mathcal{P}(2)}\right]. \tag{B.39}$$

Under H_1, we have $\lambda_1^{(M)} > \lambda_2^{(M)} = \sigma^2$. Note that most penalty functions can be factorized as $\mathcal{P}(p) = \nu \cdot C(L)$, where ν is the number of free parameters in the data model, and $C(L)/L$ tends to zero as L increases. Thus, using the definitions of $\mu_\gamma(1)$, $\varsigma_\gamma^2(1)$, $\Delta\mathcal{P}(2)$, and $W(2)$, it can be easily seen that for a large L, relation (5.26) is approximately satisfied.

B.6 Proof of Result 5.2.4

Given the definitions of $W(1)$, $\mu_\gamma(1)$ and $\varsigma_\gamma(1)$, and using $\lambda_1^{(M)} = [1+(M-1)\text{SNR}]\sigma^2$, we have

$$W(1) = [1+(M-1)\text{SNR}]\sigma^2 r_{(M,M)} \tag{B.40}$$

$$\mu_\gamma(1) = \frac{1+M\text{SNR}}{L(M\text{SNR})^2} \tag{B.41}$$

$$\varsigma_\gamma(1) = \frac{1+M\text{SNR}}{L(M\text{SNR})^2}. \tag{B.42}$$

When $M\text{SNR} \ll 1$, we can rewrite $P_M(1)$ as

$$P_M(1) \approx \text{erf}\left(\frac{\|\mathbf{r}^{(M)}\|^{-1}\sqrt{\sigma^2 r_{(M,M)}\Delta\mathcal{P}(1)/L}}{1/L(M\text{SNR})^2}-1\right). \tag{B.43}$$

At the same time, using the expression of $P_{M_\mathbf{\Lambda}}(1)$ in (5.24), we obtain

$$P_{M_\mathbf{\Lambda}}(1) \approx \text{erf}\left(\sqrt{2\Delta P(1)}-\sqrt{L}M\text{SNR}\right). \tag{B.44}$$

Remark that under H_1, we are in a scenario where λ_1 and $\lambda_1^{(M)}$ are very close to σ^2 (under-modeling) and $\|\mathbf{r}^{(M)}\|$ is very small but different form zero. Hence, given that $\text{erf}(\cdot)$ is a non-decreasing function and $L(M\text{SNR})^2 \ll 1$, by comparing (B.43) and (B.44), relation (5.27) is verified straightforwardly.

B.7 Proof of Lemma 5.2.5

Rewrite the criterion as $\mathrm{GAIC}^{(M)}(p) = \tilde{\mathcal{L}}^{(M)}(p) + \tilde{\mathcal{P}}^{(M)}(p)$ with

$$\tilde{\mathcal{L}}^{(M)}(p) = -L(M-1-p)\ln\left\{\frac{\left(\prod_{m=p+1}^{M-1}\lambda_m^{(M)}\right)^{1/(M-1-p)}}{\frac{1}{M-1-p}\sum_{m=p+1}^{M-1}\lambda_m^{(M)}}\right\} \tag{B.45}$$

and

$$\tilde{\mathcal{P}}^{(M)}(p) \;=\; L\ln\left(r_{(M,M)} - \sum_{m=1}^{p}\frac{\left(\rho_m^{(M)}\right)^2}{\lambda_m^{(M)}}\right) + p^2 + p. \tag{B.46}$$

Observe that $\tilde{\mathcal{L}}^{(M)}(p)$ in (B.45) is the conventional eigenvalue-based LL [106, 113], i.e., the ratio of the geometric mean to the arithmetic mean of the noise-only eigenvalues. It is shown in [114] that $\tilde{\mathcal{L}}^{(M)}(p)$ satisfies the following:

$$\tilde{\mathcal{L}}^{(M)}(\hat{p}) = \begin{cases} O(L\ln(\ln(L))) & ; \quad \hat{p} < p, \\ O(\ln(\ln(L))) & ; \quad \hat{p} \geq p. \end{cases} \tag{B.47}$$

therefore, it can be seen that

$$\lim_{L\to\infty} \left(L\ln(\ln(L))\right) / \tilde{\mathcal{P}}(\hat{p}) \;=\; \infty \;;\; \hat{p} < p, \tag{B.48}$$

$$\lim_{L\to\infty} \tilde{\mathcal{P}}(\hat{p}) / \left(\ln(\ln(L))\right) \;=\; \infty \;;\; \hat{p} \geq p. \tag{B.49}$$

Combining (B.47), (B.48) and (B.49), it is clear that $\mathrm{GAIC}^{(M)}(\hat{p})$ is minimized for $\hat{p} = p$ and therefore is a consistent source number estimator.

B.8 Proof of Result 5.2.6

From [104] and the approximations of Section 5.2.1, under H_1, for $L \to \infty$, we asymptotically have

$$\left(\tilde{\mathcal{L}}^{(M)}(1) - \tilde{\mathcal{L}}^{(M)}(0)\right) \;\approx\; \frac{L}{2}\left(\frac{\hat{\lambda}_1}{\sigma^2} - 1\right)^2 \tag{B.50}$$

$$\left(\tilde{\mathcal{L}}^{(M)}(2) - \tilde{\mathcal{L}}^{(M)}(1)\right) \;\approx\; -\frac{L}{2}\left(\frac{\hat{\lambda}_2}{\sigma^2} - 1\right)^2 \tag{B.51}$$

where $\tilde{\mathcal{L}}^{(L)}(p)$ is defined by (B.45).

Hence, defining $P_{M_{\text{GLE}}}(p)$ and $P_{F_{\text{GLE}}}(p)$ similarly to (5.14) and (5.15), respectively, it can be readily verified that, for $L \to \infty$, we have

$$P_{M_{\text{GLE}}}(1) \approx P\left\{ \frac{\left(\hat{\rho}_1^{(M)}\right)^2}{W(1)} + \frac{1}{2}\left(\frac{\hat{\lambda}_1}{\sigma^2} - 1\right)^2 < \frac{\Delta P(1)}{L} \middle| \mathsf{H}_1 \right\} \tag{B.52}$$

$$P_{F_{\text{GLE}}}(1) \approx P\left\{ \frac{\left(\hat{\rho}_2^{(M)}\right)^2}{W(2)} - \frac{1}{2}\left(\frac{\hat{\lambda}_2}{\sigma^2} - 1\right)^2 > \frac{\Delta P(2)}{L} \middle| \mathsf{H}_1 \right\}. \tag{B.53}$$

Observe that in both the under- and overmodeling cases, $(\hat{\lambda}_1/\sigma^2 - 1) > 0$ and $(\hat{\lambda}_2/\sigma^2 - 1) > 0$. Thus, using the results of [104] leading to expressions (5.24) and (5.25), and the derivations of Appendix B.4 leading to expressions (5.20) and (5.21), and taking into account that $\text{erf}(\cdot)$ is a non-decreasing function, Result 5.2.6 follows directly.

B.9 Proof of Lemma A.5.1

In what follows, we present the details relative to the derivation of the stochastic CRB. Recall that the CRB is given by the inverse of the FIM of the unknown parameters of interest. Considering the vector of unknown real parameters $\boldsymbol{\eta}$ defined in Section A.4, the (i,j)-th element of the FIM, $\boldsymbol{\mathcal{F}}_{\boldsymbol{\eta}}$, is given by [81]

$$\boldsymbol{\mathcal{F}}_{i,j} := K\text{trace}\left\{ \mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \eta_i}\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \eta_j} \right\}. \tag{B.54}$$

Let $\boldsymbol{\nu} := \left[\boldsymbol{\theta}^T \ \boldsymbol{f}^T\right]^T$, and $\mathbf{D} := [\mathbf{D}_{\boldsymbol{\theta}} \ \ \mathbf{D}_{\boldsymbol{f}}]$, with

$$\mathbf{D}_{\boldsymbol{\theta}} := \mathbf{F}^T \circ \tfrac{\partial \mathbf{A}}{\partial \boldsymbol{\theta}} \quad \text{and} \quad \mathbf{D}_{\boldsymbol{f}} := \tfrac{\partial \mathbf{F}^T}{\partial \boldsymbol{f}} \circ \mathbf{A}. \tag{B.55}$$

Using the expression of $\mathbf{D}$ and the results of [76, 91, 103], note that

$$\begin{aligned}\left[\frac{\partial \mathbf{R}}{\partial \theta_i} \ \ \frac{\partial \mathbf{R}}{\partial \theta_j}\right] &= \mathbf{D}\left\{\mathbf{1}_2 \otimes \left(\mathbf{i}_i\mathbf{i}_i^T\mathbf{R}_{\mathbf{H}}\mathbf{U}^{\mathbf{H}}\right)\right\} + \\ &\quad \left\{\left(\mathbf{U}\mathbf{R}_{\mathbf{H}}\mathbf{i}_i\mathbf{i}_i^T\right) \otimes \mathbf{1}_2^T\right\}\mathbf{D}^H\end{aligned} \tag{B.56}$$

where $\mathbf{1}_2 = [1 \ \ 1]^T$, and $\mathbf{i}_i$ is defined in (7.14).

From (B.56) and using results of [39, 76, 103], we have the following:

$$\begin{aligned}\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta}} &= K\boldsymbol{\mathcal{I}}^T\left[\mathbf{D}_{\boldsymbol{\theta}}^T\otimes\left(\mathbf{R}\mathbf{U}^H\right)+\left(\mathbf{R}^T\mathbf{U}^T\right)\otimes\mathbf{D}_{\boldsymbol{\theta}}^H\right]\\ &\quad\left[\left(\mathbf{R}^T\right)^{-1}\otimes\left(\mathbf{R}\right)^{-1}\right]\\ &\quad\left[\mathbf{D}_{\boldsymbol{\theta}}^T\otimes\left(\mathbf{R}\mathbf{U}^H\right)+\left(\mathbf{R}^T\mathbf{U}^T\right)\otimes\mathbf{D}_{\boldsymbol{\theta}}^H\right]^H\boldsymbol{\mathcal{I}}\end{aligned}\tag{B.57}$$

$$\begin{aligned}\boldsymbol{\mathcal{F}}_{\boldsymbol{ff}} &= K\boldsymbol{\mathcal{I}}^T\left[\mathbf{D}_{\boldsymbol{f}}^T\otimes\left(\mathbf{R}\mathbf{U}^H\right)+\left(\mathbf{R}^T\mathbf{U}^T\right)\otimes\mathbf{D}_{\boldsymbol{f}}^H\right]\\ &\quad\left[\left(\mathbf{R}^T\right)^{-1}\otimes\left(\mathbf{R}\right)^{-1}\right]\\ &\quad\left[\mathbf{D}_{\boldsymbol{f}}^T\otimes\left(\mathbf{R}\mathbf{U}^H\right)+\left(\mathbf{R}^T\mathbf{U}^T\right)\otimes\mathbf{D}_{\boldsymbol{f}}^H\right]^H\boldsymbol{\mathcal{I}}\end{aligned}\tag{B.58}$$

where $\boldsymbol{\mathcal{I}} = \left[\mathrm{vec}(\mathbf{i}_1\mathbf{i}_1^T)\ \ \mathrm{vec}(\mathbf{i}_2\mathbf{i}_2^T)\right]$. From the structure of $\boldsymbol{\mathcal{I}}$, it can be easily shown that $\boldsymbol{\mathcal{I}}^T\left[\mathbf{AB}\right]\boldsymbol{\mathcal{I}} = \mathbf{A}\odot\mathbf{B}$. Using this property with (B.56) and the identity $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = \mathbf{AC}\otimes\mathbf{BD}$, it can be readily verified that $\boldsymbol{\mathcal{F}}_{\boldsymbol{\nu\nu}}$ can be written as [39, 76]

$$\begin{aligned}\boldsymbol{\mathcal{F}}_{\boldsymbol{\nu\nu}}=2K\Re\Big\{&\left[\mathbf{1}_{22}\otimes\left(\mathbf{R}_{\mathbf{H}}\mathbf{U}^H\mathbf{R}^{-1}\mathbf{U}\mathbf{R}_{\mathbf{H}}\right)\right]\odot\left(\mathbf{D}^H\mathbf{R}^{-1}\mathbf{D}\right)^T+\\ &\left[\mathbf{1}_2\otimes\left(\mathbf{R}_{\mathbf{H}}\mathbf{U}^H\mathbf{R}^{-1}\mathbf{D}\right)\right]\odot\left[\left(\mathbf{R}_{\mathbf{H}}\mathbf{U}^H\mathbf{R}^{-1}\mathbf{D}\right)^T\otimes\mathbf{1}_2^T\right]\Big\}\end{aligned}\tag{B.59}$$

with $\mathbf{1}_{22} = \mathbf{1}_2\mathbf{1}_2^T$.

Using results of [39, 76, 103], from the expansion of $\mathbf{Q}$ in (A.14), we get

$$\boldsymbol{\mathcal{F}}_{\mathbf{qq}} = K\boldsymbol{\mathcal{P}}^T\left[\left(\mathbf{R}^T\right)^{-1}\otimes\mathbf{R}^{-1}\right]\boldsymbol{\mathcal{P}}\tag{B.60}$$

where $\boldsymbol{\mathcal{P}}$ is defined in (A.4), applied to $\mathbf{R}_{\mathbf{H}}$ instead of $\mathbf{R}_{\mathbf{s}}$.

Note that $\mathbf{R}_{\mathbf{H}}$ has a Hermitian structure and is completely unknown. This suggests that $\mathbf{R}_{\mathbf{H}}$ can be written as a linear combination of p^2 base matrices, in the same way as $\mathbf{Q}$, where p is the number of sources, and in our case is equal to 2. Moreover, it can be shown that the parameterization of $\mathbf{R}_{\mathbf{H}}$ in terms of $\boldsymbol{\Psi}_i$'s does not have an impact on the final expression of the CRB on $\boldsymbol{\nu}$ [39]. Thus, similarly to $\boldsymbol{\mathcal{F}}_{q_iq_j}$, we obtain

$$\begin{aligned}\boldsymbol{\mathcal{F}}_{\boldsymbol{\rho\rho}} = K\boldsymbol{\mathcal{P}}_{\boldsymbol{\rho}}{}^T&\left[\mathbf{U}^T\otimes\mathbf{U}^H\right]\left[\left(\mathbf{R}^T\right)^{-1}\otimes\mathbf{R}^{-1}\right]\\ &\left[\mathbf{U}^*\otimes\mathbf{U}\right]\boldsymbol{\mathcal{P}}_{\boldsymbol{\rho}}\end{aligned}\tag{B.61}$$

with $\boldsymbol{\mathcal{P}}_{\boldsymbol{\rho}} = [\mathrm{vec}(\boldsymbol{\Psi}_1) \ldots \mathrm{vec}(\boldsymbol{\Psi}_{p^2})]$ and $p = 2$.

Using (B.59), (B.60) and (B.61), the cross-terms $\boldsymbol{\mathcal{F}}_{\boldsymbol{\nu}\mathbf{q}}$, $\boldsymbol{\mathcal{F}}_{\boldsymbol{\nu}\boldsymbol{\rho}}$ and $\boldsymbol{\mathcal{F}}_{\boldsymbol{\rho}\mathbf{q}}$ can be easily deduced.

From the obtained expressions of the auto- and cross-terms along with properties of block-matrix inversion, we get the following expression for the CRB:

$$\begin{aligned}
\mathrm{CRB}_{\boldsymbol{\nu}}^{-1} &= \boldsymbol{\mathcal{F}}_{\boldsymbol{\nu}\boldsymbol{\nu}} + \boldsymbol{\mathcal{F}}_{\boldsymbol{\nu}\boldsymbol{\rho}} \left(\boldsymbol{\mathcal{F}}_{\boldsymbol{\rho}\boldsymbol{\rho}} - \boldsymbol{\mathcal{F}}_{\boldsymbol{\rho}\mathbf{q}} \boldsymbol{\mathcal{F}}_{\mathbf{q}\mathbf{q}}^{-1} \boldsymbol{\mathcal{F}}_{\mathbf{q}\boldsymbol{\rho}}\right)^{-1} \\
&\quad \left(\boldsymbol{\mathcal{F}}_{\boldsymbol{\rho}\mathbf{q}} \boldsymbol{\mathcal{F}}_{\mathbf{q}\mathbf{q}}^{-1} \boldsymbol{\mathcal{F}}_{\mathbf{q}\boldsymbol{\nu}} - \boldsymbol{\mathcal{F}}_{\boldsymbol{\rho}\boldsymbol{\nu}}\right) + \\
&\quad \boldsymbol{\mathcal{F}}_{\boldsymbol{\nu}\mathbf{q}} \left(\boldsymbol{\mathcal{F}}_{\mathbf{q}\mathbf{q}} - \boldsymbol{\mathcal{F}}_{\mathbf{q}\boldsymbol{\rho}} \boldsymbol{\mathcal{F}}_{\boldsymbol{\rho}\boldsymbol{\rho}}{}^{-1} \boldsymbol{\mathcal{F}}_{\boldsymbol{\rho}\mathbf{q}}\right)^{-1} \\
&\quad \left(\boldsymbol{\mathcal{F}}_{\mathbf{q}\boldsymbol{\rho}} \boldsymbol{\mathcal{F}}_{\boldsymbol{\rho}\boldsymbol{\rho}}{}^{-1} \boldsymbol{\mathcal{F}}_{\boldsymbol{\rho}\boldsymbol{\nu}} - \boldsymbol{\mathcal{F}}_{\mathbf{q}\boldsymbol{\nu}}\right). \qquad \text{(B.62)}
\end{aligned}$$

Using the results of [39, 76, 90] and applying them to (B.62), the closed-form expression of the stochastic CRB of (A.30) can be deduced straightforwardly.

Bibliography

[1] H. Akaike, "A New Look at the Statistical Model Identification", *IEEE Trans. Automatic Control*, Vol. AC-19, No. 6, pp. 716-723, Dec. 1974.

[2] T. W. Anderson, "Asymptotic Theory for Principal Component Analysis", *Ann. Statist.*, Vol. 34, pp. 151-158, 1960.

[3] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Wiley, New York, 2003.

[4] R. Aouada, "Qualification et Estimation de Dérives de Données de Process (EDF)", Master's thesis, SUPELEC, Paris, France, Jun. 2005. (In French).

[5] A. J. Barabell, "Improving the Resolution Performance of Eigenstructure-based Direction Finding Algorithms", *Proc. IEEE ICASSP'83*, pp. 336-339, Boston, MA, 1983.

[6] M. S. Bartlett, "Smoothing Periodograms from Time Series with Continuous Spectra", *Nature*, Vol. 161, pp. 686-687, 1948.

[7] M. S. Bartlett, "A Note on the Multiplying Factors for Various χ^2 Approximations", *J. Royal Stat. Soc.*, B Vol. 16, pp. 296-298, 1954.

[8] D. K. Barton, "Low Angle Radar Tracking", *Proc. IEEE*, Vol. 62, No. 6, pp 687-704, Jun. 1974.

[9] R. T. Behrens and L. L. Scharf, "Signal Processing Applications of Oblique Projection Operators", *IEEE Trans. Signal Processing*, Vol. 42, No. 6, pp. 1413-1424, Jun. 1994.

[10] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso and E. Moulines, "A Blind Source Separation Technique Using Second-Order Statistics", *IEEE Trans. Signal Processing*, Vol. 45, No. 2, pp. 434-444, Feb. 1997.

[11] L. V. Blake, "Prediction of Radar Range", in M. I. Skolnik editor, *Radar Handbook*, McGraw-Hill, New York, 1990, 2nd ed.

[12] J. F. Böhme, "Array Processing", *Advances in Spectrum Analysis and Array Processing*, pp. 1-63, Prentice Hall, New York, 1991.

[13] J. F. Böhme, "Estimation of Source Parameters by Maximum Likelihood and Nonlinear Regression", *Proc. IEEE ICASSP'84*, pp. 7.3.1-7.3.4, San Diego, CA, 1984.

[14] J. F. Böhme and D. Kraus, "On Least Squares Methods for Direction of Arrival Estimation in the Presence of Unknown Noise fields", *Proc. IEEE ICASSP'88*, pp. 2785-2788, New York, Apr. 1988.

[15] E. Bossé, R. M. Turner and E. S. Riseborough, "Model-Based Multifrequency Array Signal Processing for Low-Angle Tracking", *IEEE Trans. Aerospace Electronic Syst.*, Vol. AES-31, pp. 194-210, Jan. 1995.

[16] G. E. P. Box, "A General Distribution Theory for a Class of Likelihood Criteria", *Biometrika*, Vol. 36, pp. 317-346, Dec. 1949.

[17] R. F. Brcich, A. M. Zoubir and P. Pelin, "Detection of Sources Using Bootstrap Techniques", *IEEE Trans. Signal Processing*, Vol. 50, No. 2, pp.206-215, Feb. 2002.

[18] Y. Bresler and A. Macovski, "On the Number of Signals Resolvable by a Uniform Linear Array", *IEEE Trans. Acoust. Speech, Signal Processing*, Vol. ASSP-34, No. 12, pp. 1361-1375, Dec. 1986.

[19] J. W. Brewer, "Kronecker Products and Matrix Calculus in System Theory", *IEEE Trans. Circuits and Systems*, Vol. CAS-25, No. 9, pp. 772-781, Sep. 1978.

[20] D. R. Brillinger, *Time Series: Data Analysis and Theory*, Expanded ed., Holden-Day, San Francisco, CA, 1981.

[21] C. G. Broyden, "The Convergence of a Class of Double-Rank Minimization Algorithms", *J. Isnt. Math. Applic.*, Vol. 6, pp. 76-90, 1970.

[22] J. P. Burg, "The Relationship between Maximum Entropy Spectra and Maximum Likelihood Spectra" *Geophysics*, Vol. 37, pp. 375-376, 1972.

[23] J. A. Cadzow, "A High Resolution Direction-of-Arrival Algorithm for Narrow-Band Coherent and Incoherent Sources", *IEEE Trans. Acoust. Speech, Signal Processing*, Vol. ASSP-36, No. 7, pp. 965-979, Jul. 1988.

[24] J. A. Cadzow, "Multiple Source Location - The Signal Subspace Approach", *IEEE Trans. Acoust. Speech, Signal Processing*, Vol. ASSP-38, No. 7, pp. 1110-1125, Jul. 1990.

[25] B. H. Cantrell, W. B. Gordon and G. V. Trunk, "Maximum Likelihood Elevation Angles Estimation of Radar Targets using Subapertures", *IEEE Trans. Aerospace Electronic Syst.,* Vol. AES-17, pp. 213-221, 1981.

[26] J. F. Cardoso and A. Souloumiac, "Blind Beamforming for Non-Gaussian Signals", *IEE-F Proc.*, Vol. 140, No. 6, pp. 362-370, Dec. 1993.

[27] W. Chen, K. M. Wong and J. P. Reilly, "Detection of the Number of Signals: A Predicted Eigen-Threshold Approach", *IEEE Trans. Signal Processing*, Vol. 39, No. 5, pp. 1088-1098, May 1991.

[28] M. Djeddou, "Estimation des Paramètres de Trajets Multiples dans les Systèmes Radar", PhD thesis, Ecole Nationale Polytechnique (ENP), Algiers, Algeria, Jul. 2005. (In French).

[29] M. Djeddou, A. Belouchrani and S. Aouada, "Maximum Likelihood Angle-Frequency Estimation in Partially Known Correlated Noise for Low-Elevation Targets", *IEEE Trans. Signal Processing*, Vol. 53, No. 8, pp. 3057-3064, Aug. 2005.

[30] G. d'Urso, P. Prieur and C. Vincent, "Blind Identification Methods Applied to Electricité de France's Civil Works and Power Plants Monitoring", *Proc. IEEE HOS'97*, pp. 82-86, Banff, Canada, Jun. 1997.

[31] B. Efron and R. Tibshirani, *An Introduction to Bootstrap*, Chapman & Hall, New York, 1993.

[32] M. Feder and E. Weinstein, "Parameter Estimation of Superimposed Signals using the EM Algorithm", *IEEE Trans. Acoust. Speech, Signal Processing*, Vol. ASSP-36, No. 2, pp. 477-489, Feb. 1988.

[33] E. Fishler and H. Messer, "On the use of Order Statistics for Improved Detection of Signals by the MDL Criterion", *IEEE Trans. Signal Processing*, Vol. 48, No. 8, pp. 2242-2247, Aug. 2000.

[34] E. Fishler and H. V. Poor, "Estimation of the Number of Sources in Unbalanced Arrays via Information Theoretic Criteria", *IEEE Trans. Signal Processing*, Vol. 53, No. 9, pp. 3543-3553, Sep. 2005.

[35] R. Fletcher and M. J. D. Powell, "A Rapidly Convergent Descent Method for Minimization", *Computer Journal*, Vol. 6, pp. 163-168, 1963.

[36] B. Friedlander and A. J. Weiss, "Direction Finding using Noise Covariance Modeling", *IEEE Trans. Signal Processing*, Vol. 43, No. 7, pp. 1557-1567, Jul. 1995.

[37] J. J. Fuchs, "Estimation of the Number of Signals in the Presence of Unknown Correlated Noise", *IEEE Trans. Signal Processing*, Vol. 40, No. 5, pp. 1053-1061, May 1992.

[38] A. B. Gershman, A. L. Matveyev and J. F. Böhme, "Maximum Likelihood Estimation of Signal Power in Sensor Array in the Presence of Unknown Noise Field", *Proc. IEE Radar, Sonar, Navig.*, Vol. 142, No. 5, pp.218-224, Oct. 1995.

[39] A. B. Gershman, P. Stoica, M. Pesavento and E. G. Larsson, "Stochastic Cramér-Rao Bound for Direction Estimation in Unknown Noise Fields", *Proc. IEE Radar, Sonar, Navig.*, Vol. 149, No. 1, pp. 2-8, Feb. 2002.

[40] D. Goldfarb, "A Family of Variable Metric Updates Derived by Variational Means", *Mathematics of Computing*, Vol. 24, pp. 23-26, 1970.

[41] B. Göransson, "Robust Direction Estimation in the Presence of Spatially Correlated Noise", *Procs. IEEE SSAP'94*, pp. 157-160, Québec, Canada, Jun. 1994.

[42] B. Göransson and B. Ottersten, "Direction Estimation in Partially Unknown Noise Fields", *IEEE Trans. Signal Processing*, Vol. 47, No. 9, pp. 2375-2385, Sep. 1998.

[43] P. Hall, "On the Number of Bootstrap Simulations Required to Construct a Confidence Interval", *The Annals of Statistics*, Vol. 14, No. 4, pp. 1453-1462, Apr. 1986.

[44] S. Haykin, *Array Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, 1984.

[45] Y. Hochberg and A. Tamhane, *Multiple Comparison Procedures*. Wiley, New York, 1987.

[46] B. Hochwald and A. Nehorai, "Identifiability in Array Processing Models with Vector-Sensor Applications", *IEEE Trans. Signal Processing*, Vol. 44, No. 1, pp. 83-95, Jan. 1996.

[47] H. M. Huizenga, J. C. de Munck, L. J. Waldorp and R. P. P. P. Grasman, "Spatiotemporal EEG/MEG Source Analysis based on a Parametric Noise Covariance Model", *IEEE Trans. Biomedical Eng.* Vol. 49, No. 6, pp. 533-539, Jun. 2002.

[48] A. Hyvärinen and E. Oja, "A Fast Fixed-Point Algorithm for Independent Component Analysis", *Neural Computing*, Vol. 9, No. 7, pp. 1483-1492, 1997.

[49] H. R. Karimi and A. Manikas, "Manifold of a Planar Array and its Effects on the Accuracy of Direction-finding Systems", *Proc. IEE Radar, Sonar, Navig.*, Vol. 143, No. 6, pp. 349-357, Dec. 1996.

[50] M. Kaveh, H. Wang and H. Hung, "On the Theoretical Performance of a Class of Estimators of the Number of Narrow-Band Sources", *IEEE Trans. Acoust. Speech, Signal Processing*, Vol. ASSP-35, No. 9, pp. 1350-1352, Oct. 1987.

[51] H. Krim and M. Viberg, "Two Decades of Array signal Processing Research", *IEEE Signal Processing Magazine*, Vol. 13, No. 4, pp. 67-94, Jul. 1996.

[52] A. A. Ksienski and R. B. McGhee, "A Decision Theoretic Approach to the Angular Resolution and Parameter Estimation Problem for Multiple Targets", *IEEE Trans. Aerospace Electronic Syst.,* Vol. AES-4, pp. 443-455, 1968.

[53] R. Kumaresan, L. L. Scharf and A. K. Shaw, "An Algorithm for Pole-Zero Modeling and Spectral Analysis", *IEEE Trans. Acoust. Speech, Signal Processing*, Vol. ASSP-34, No. 6, pp. 637-649, Jul. 1986.

[54] D. Lawley, "Test of Significance for the Latent Roots of Covariance and Correlation Matrices", *Biometrika*, Vol. 43, pp. 128-136, 1956.

[55] J. LeCadre, "Parametric Methods for Spatial Signal Processing in Presence of Unknown Colored Noise Fields", *IEEE Trans. Acoust. Speech, Signal Processing*, Vol. ASSP-37, No. 7, pp. 965-983, Jul. 1989.

[56] A. Lemma, A. J. Van der Veen and E. F. Deprettere, "Joint Angle-Frequency Estimation using Multi-Resolution ESPRIT", *Proc. ICASSP'98*, Seattle, WA, pp.1957-1960, May 1998.

[57] J. Li, P. Stoica and Z. S. Liu, "Comparative Study of IQML and MODE Direction-of-Arrival Estimators", *IEEE Trans. Signal Processing*, Vol. 46, No. 1, pp. 149-160, Jan. 1998.

[58] J. Li, P. Stoica and Z. S. Liu, "On Nonexistence of the Maximum Likelihood Estimate in Blind Multichannel Identification", *IEEE Signal Processing Magazine*, Vol. 22, No. 4, pp. 99-101, Jul. 2005.

[59] F. Li, K. W. Tam and Y. Wu, "Determining the Number of Sources in Signal Processing", *Proc. IEEE ICSP'00*, pp. 413-420, Beijing, China, 2000.

[60] A. P. Liavas and P. A. Regalia, "On the Behavior of Information Theoretic Criteria for Model Order Selection", *IEEE Trans. Signal Processing*, Vol. 49, No. 8, pp. 1689-1695, Aug. 2001.

[61] W. S. Ligget Jr., "Passive Sonar: Fitting Models to Multiple Time Series", *Signal Processing: Proc. NATO ASI Signal Processing*, J. W. R. Griffiths and P. L. Stocklin editors, pp. 327-345, Academic, New York, 1973.

[62] T. Lo and J. Litva, "Use of a Highly Deterministic Multipath Signal Model in Low-Angle Tracking", *Proc. IEE-F*, Vol. 138, No. 2, Apr. 1991.

[63] D. Madurasinghe, "A New DOA Estimator in Nonuniform Noise", *IEEE Signal Processing Letters*, Vol. 12, No. 4, pp. 337-339, Apr. 2005.

[64] A. Manikas, H. R. Karimi and I. Dacos, "Study of the Detection and Resolution Capabilities of One-Dimensional Array of Sensors by using Differential Geometry", *Proc. IEE Radar, Sonar, Navig.*, Vol. 141, No. 2, pp. 83-92, Feb. 1994.

[65] G. Miller and D. R. Fuhrmann, "Maximum Likelihood Narrow-Band Direction Finding and the EM Algorithm", *IEEE Trans. Acoust. Speech, Signal Processing*, Vol. ASSP-38, No. 9, pp. 1560-1577, Sep. 1990.

[66] R. J. Murihead, *Aspects of Multivariate Statistical Theory*, Wiley, New York, 1982.

[67] K. Nakatsuka, "Two-Beam Technique for Tracking a Target at Low Elevation Angles", *Proc. IEE-F*, Vol. 137, No. 6, pp. 397-406, Dec. 1990.

[68] U. Nickel, "Algebraic Formulation of Kumaresan-Tufts Superresolution Method, Showing Relation to ME and MUSIC Methods", *Proc. IEE-F*, Vol. 135, pp. 7-10, 1988.

[69] U. Nickel, "On the Influence of Channel Errors on Array Signal Processing Methods", *J. Electron. Commun.*, Vol. 47, No. 4, pp. 209-219, 1993.

[70] H. T. Ong and A. M. Zoubir, "Bootstrap-based Detection of Signals with Unknown Parameters in Unspecified Correlated Interference", *IEEE Trans. Signal Processing*, Vol. 51, No. 1, pp. 135-141, Jan. 2003.

[71] B. Ottersten, M. Viberg, P. Stoica and A. Nehorai, "Exact and Large Sample ML Techniques for Parameter Estimation and Detection in Array Processing", *Radar Array Processing*, Haykin, Litva and Shepher, editors, pp. 99-151, Springer-Verlag, Berlin, Germany, 1993.

[72] B. Ottersten, M. Viberg and T. Kailath, "Analysis of Subspace Fitting and ML Techniques for Parameter Estimation from Sensor Array Data", *IEEE Trans. Signal Processing*, Vol. 40, No.3, pp. 590-600, Mar. 1992.

[73] B. Ottersten, P. Stoica and R. Roy, "Covariance Matching Estimation Techniques for Array Signal Processing Applications", *Digital Signal Processing*, Vol. 8, pp. 185-210, 1998.

[74] A. Paulraj and T. Kailath, "Eigenstructure Methods for Direction of Arrival Estimation in the Presence of Unknown Noise Fields", *IEEE Trans. Acoust. Speech, Signal Processing*, Vol. ASSP-34, No. 1, pp. 13-20, Jan. 1986.

[75] C. Pell, "Phased-Array Radars", *IEE Review*, Vol. 34 , No. 9, pp. 363-367, Oct. 1988.

[76] M. Pesavento and A. B. Gershman, "Maximum-Likelihood Direction-of-Arrival Estimation in the presence of Unknown Nonuniform Noise", *IEEE Trans. Signal Processing*, Vol. 49, No.7, pp. 1310-1324, Jul. 2001.

[77] S. U. Pillai and B. H. Kwon, "Forward/Backward Spatial Smoothing Techniques for Coherent Signal Identification", *IEEE Trans. Acoust. Speech, Signal Processing,* Vol. ASSP-37, No. 1, pp. 8-15, Jan. 1989.

[78] D. Politis, J. Romano and M. Wolf, *Subsampling*, Springer-Verlag, New York, 1999.

[79] C. R. Rao, *Linear Statistical Inference and its Applications*, 2nd ed., Wiley, New York, 1973.

[80] J. Rissanen, "Modeling by Shortest Data Description", *Automatica*, Vol. 14, pp. 465-471, 1978.

[81] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, Addison-Wesley, Reading, MA, 1992.

[82] L. L. Scharf, "Blind Adaptation of Zero Forcing Projections and Oblique Pseudo-Inverses for Subspace Detection and Estimation When Interference Dominates Noise", *IEEE Trans. Signal Processing*, Vol. 50, No. 12, pp. 2938-2946, Dec. 2002.

[83] R. O. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation", *Proc. RADC Spectrum Estimation Workshop*, pp. 243-258, Oct. 1979.

[84] M. I. Skolnik, *Introduction to Radar Systems*, McGraw-Hill, New York, 1980.

[85] R. J. Serfling, *Approximation Theorems of Mathematical Statistics*, Wiley, New York, 2002.

[86] T. J. Shan, M. Wax and T. Kailath, "On Spatial Smoothing for Direction of Arrival Estimation of Coherent Signals", *IEEE Trans. Acoust. Speech, Signal Processing,* Vol. ASSP-33, No. 4, pp. 806-811, Apr. 1985.

[87] D. F. Shanno, "Conditioning of Quasi-Newton Methods for Function Minimization", *Mathematics of Computing*, Vol. 24, pp. 647-656, 1970.

[88] D. Starer and A. Nehorai, "Newton Algorithms for Conditional and Unconditional Maximum Likelihood Estimation of the Parameters of Exponential Signals in Noise", *IEEE Trans. Signal Processing*, Vol. 40, No. 6, pp. 1520-1534, Jun. 1992.

[89] P. Stoica and M. Cedervall, "Detection Tests for array Processing in Unknown Correlated Noise Fields", *IEEE Trans. Signal Processing*, Vol. 45, No. 9, pp. 2351-2362, Sep. 1997.

[90] P. Stoica, E. G. Larsson and A. B. Gershman, "The Stochastic CRB for Array Processing: A Textbook Derivation", *IEEE Signal Processing Letters*, Vol. 8, No. 5, pp.148-150, May 2001.

[91] P. Stoica and A. Nehorai, "MUSIC, Maximum Likelihood and the Cramér-Rao Bound", *IEEE Trans. Acoust. Speech, Signal Processing*, Vol. ASSP-37, No. 5, pp. 720-741, May 1989.

[92] P. Stoica and A. Nehorai, "Performance Study of Conditional and Unconditional Direction-of-Arrival Estimation", *IEEE Trans. Acoust. Speech, Signal Processing*, Vol. ASSP-38, No. 10, pp. 1783-1795, Oct. 1990.

[93] P. Stoica and K. C. Sharman, "Novel Eigenanalysis Method for Direction Estimation", *Proc. IEE-F*, Vol. 137, No. 1, pp. 19-26, Feb. 1990.

[94] P. Stoica and K. C. Sharman, "Maximum Likelihood Methods for Direction-Of-Arrival Estimation", *IEEE Trans. Acoust. Speech Signal Processing*, Vol. ASSP-38, No7, pp.1132-1143, Jul. 1990.

[95] P. Swerling, "Probability of Detection for Fluctuating Targets, Special Monograph", *IRE Trans. Information Theory*, Vol. IT-6, No. 2. pp. 269-308, Apr. 1960.

[96] R. M. Turner and E. Bossé, "Maximum Likelihood Tracking using a Highly Refined Multipath Model", *Proc. Asilomar Conf. on Signals, Systems and Computers,* Pacific Grove, CA, Nov. 1987.

[97] R. J. Urick, *Principles of Underwater Sound*, 2nd ed., McGraw Hill, New York, 1975.

[98] M. C. Vanderveen, A. J. Van der Veen and A. Paulraj, "Estimation of Multipath Parameters in Wireless Communications", *IEEE Trans. Signal Processing*, Vol. 46, No. 3, pp. 682-690, Mar. 1998.

[99] H. L. Van Trees, *Detection, Estimation and Modulation Theory*-Part I, John Wiley & Sons, New York, 1968.

[100] M. Viberg and B. Ottersten, "Sensor Array Processing based on Subspace Fitting", *IEEE Trans. Signal Processing*, Vol. 39, No. 5, pp. 1110-1121, May 1991.

[101] M. Viberg, B. Ottersten and T. Kailath, "Detection and Estimation in Sensor Arrays using Weighted Subspace Fitting", *IEEE Trans. Acoust. Speech, Signal Processing*, Vol. ASSP-39, No. 11, pp. 2436-2449, Nov. 1991.

[102] M. Viberg and A. L. Swindlehurst, "A Bayesian Approach to Auto-Calibration for Parametric Array Signal Processing", *IEEE Trans. Signal Processing*, Vol. 42, No. 12, pp. 3495-3507, Dec. 1994.

[103] S. A. Vorobyov, A. B. Gershman and K. M. Wong, "Maximum Likelihood Direction-of-Arrival Estimation in Unknown Noise Fields Using Sparse Sensor Arrays", *IEEE Trans. Signal Processing*, Vol. 53, No. 1, pp. 34-42, Jan. 1995.

[104] H. Wang and M. Kaveh, "On the Performance of Signal-Subspace Processing - Part I: Narrow-Band Systems", *IEEE Trans. Acoust. Speech, Signal Processing*, Vol. ASSP-34, No. 5, pp. 1201-1209, Oct. 1986.

[105] M. Wax, "Detection and Localization of Multiple Sources in Noise with Unknown Covariance", *IEEE Trans. Acoust. Speech, Signal Processing*, Vol. ASSP-40, No. 1, pp. 245-249, Jan. 1992.

[106] M. Wax and T. Kailath, "Detection of Signals by Information Theoretic Criteria", *IEEE Trans. Acoust. Speech, Signal Processing*, Vol. ASSP-33, No. 2, pp. 387-392, Feb. 1985.

[107] M. Wax and T. Kailath "Optimal Localization of Multiple Sources by Passive Arrays", *IEEE Trans. Acoust. Speech, Signal Processing*, Vol. ASSP-31, No. 5, pp. 1210-1218. May 1983.

[108] M. Wax and I. Ziskind, "Detection of the Number of Coherent Signals by the MDL Principle", *IEEE Trans. Acoust. Speech, Signal Processing*, Vol. ASSP-37, No. 8, pp, 1190-1196, Aug. 1989.

[109] P. H. Westfall and S. S. Young, *Resampling-based Multiple Testing*, John Wiley & Sons, New York, 1993.

[110] W. D. White, "Low Angle Radar Tracking in the Presence of Multi-Path", *IEEE Trans. Aerospace Electronic Syst.*, Vol. AES-10, pp. 335-352, 1974.

[111] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, pp-71-72, Oxford, Clarendon, 1965.

[112] D. B. Williams and D. H. Johnson, "Using the Sphericity Test for Source Detection with Narrow-Band Passive Arrays", *IEEE Trans. Acoust. Speech, Signal Processing*, Vol. 38, No. 11, pp. 2008-2014, Nov. 1990.

[113] H. T. Wu, J. F. Yang and F. K. Chen, "Source Number Estimation Using Transformed Gerschgorin Radii", *IEEE Trans. Signal Processing*, Vol. 43, No. 6, pp. 1325-1333, Jun. 1995.

[114] G. Xu, R. H. Roy and T. Kailath, "Detection of the Number of Sources via Exploitation of Centro-Symmetry Property", *IEEE Trans. Signal Processing*, Vol. 42, No. 1, pp. 102-112, Jan. 1994.

[115] L. C. Zhao, P. R. Krishnaiah and Z. D. Bai, "On Detection of the Number of Signals in Presence of White Noise", *J. Multivariate Analysis*, Vol. 20, No. 1, pp. 1-25, Feb. 1986.

[116] L. C. Zhao, P. R. Krishnaiah and Z. D. Bai, "On Detection of the Number of Signals when the Noise Covariance Matrix is Arbitrary", *J. Multivariate Analysis*, Vol. 20, No. 1, pp. 26-49, Feb. 1986.

[117] L. C. Zhao, P. R. Krishnaiah and Z. D. Bai, "Remarks on Certain Criteria for Detection of Number of Signals", *IEEE Trans. Acoust. Speech, Signal Processing*, Vol. ASSP-35, No. 2, pp. 129-132, Feb. 1987.

[118] I. Ziskind and M. Wax, "Maximum Likelihood Localization of Multiple Sources by Alternating Projections", *IEEE Trans. Acoust. Speech, Signal Processing*, Vol. ASSP-36, No. 10, pp. 1553-1560, Oct. 1988.

[119] I. Ziskind and M. Wax, "Maximum Likelihood Localization of Diversely Polarized Sources by Simulated Annealing", *IEEE Trans. Antennas Propagation*, Vol. 38, No. 7, pp. 1111-1114, Jul. 1990.

[120] A. M. Zoubir, "Bootstrap Methods for Model Selection", *J. Electron. Commun.*, Vol. 53, No. 6, pp. 386-392, 1999.

[121] A. M. Zoubir and D. R. Iskander, *Bootstrap Techniques for Signal Processing*, Cambridge University Press, Cambridge, U.K., 2004.

Saïd AOUADA

Born: 9 September 1975 in Termez (U.S.S.R.)
Citizenship: Algerian – Russian

Education:

2001 – 2006: **Doktor-Ingenieur** – Electrical Engineering / Signal Processing (Distinct.: Very Good)
Darmstadt University of Technology Darmstadt (Germany)
Curtin University of Technology – Transferred to Germany, 2003 Perth (Australia)

1998 – 2000: **Magister** – Electronics / Information Processing (Distinct.: Very Good)
Ecole Nationale Polytechnique (ENP) Algiers (Algeria)

1993 – 1998: **Ingénieur d'Etat** – Electrical Engineering / Automatic Control (Distinct.: Very Good)
National Institute of Electricity & Electronics (INELEC) Boumerdès (Algeria)

Professional Experience:

2003 – 2006 : **DARMSTADT UNIVERSITY OF TECHNOLOGY** Darmstadt (Germany)
Signal Processing Group (SPG)

Research Associate

1999 – 2001: **SCHNEIDER ELECTRIC** Algiers (Algeria)

2000 – 2001: Industrial Control Activity Manager
1999 – 2000: Industrial Control Product-Promotion Engineer

Sessional Academic Work:

2001 – 2003: **CURTIN UNIVERSITY OF TECHNOLOGY** Perth (Australia)
Communications & Signal Processing Group (CSP)

Teaching and Research Assistant (PhD candidate)

1998 – 1999: **BLIDA UNIVERSITY OF TECHNOLOGY** Blida (Algeria)

Teaching Assistant (Magister candidate)

Internships:

2000 – 2001: **SCHNEIDER ELECTRIC** Paris, Grenoble (France), Algiers (Algeria)
Schneider Corporate, Sales Techniques I & II, Marketing, various technical trainings

1998: **Entreprise Nationale de Géophysique (ENAGEO – SONATRACH)** Algiers (Algeria)
Seismic data processing (Identification of hydrocarbon reservoirs)

1997: **Entreprise des Ciments et Dérivés d'Ech-Cheliff (ECDE)** Chlef (Algeria)
Process control (Operation of the largest cement-plant in Africa)

1995: **Entreprise Nationale des Industries Electroniques (ENIE)** Blida (Algeria)
Company management (Audio & video equipment)

Languages (1 – Fluent, 2 - Average):

Arabic (1) Russian (1) French (1) English (1) Spanish (2) German (2)

Scholarships:

2001 – 2003: Curtin International Student Scholarship (CISS) Perth (Australia)
2001 – 2003: Australian Telecommunications Cooperative Research Centre (ATCRC) Top-Up Perth (Australia)