Source Enumeration in Sensor Array Processing: A Model Order Selection Problem

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

von
Dipl.-Ing. Zhihua Lu
geboren am 18.05.1983 in Zhejiang, China

Referent:	Prof. Dr.-Ing. Abdelhak M. Zoubir
Korreferent:	Prof. Dr. Visa Koivunen
Tag der Einreichung:	10.07.2012
Tag der mündlichen Prüfung:	25.10.2012

D 17
Darmstadt, 2012

Acknowledgments

I would like to thank all people who have helped me during my PhD study.

First of all, I would like to express my sincere gratitude to my supervisor Prof. Dr.-Ing. Abdelhak M. Zoubir for his invaluable guidance and continuous encouragement. I could not have imagined having a better supervisor and mentor for my PhD study, who shows such a high degree of enthusiasm and motivation.

I also acknowledge my gratitude to Prof. Dr.-Ing. Ralf Steinmetz, Prof. Dr. Visa Koivunen, Prof. Dr.-Ing. Jürgen Adamy and Prof. Dr.-Ing. Marius Pesavento who acted as the chair and examiners in the PhD committee.

I would like to thank my colleagues at the Signal Processing Group at TU Darmstadt. I treasure my memories of those joyous days. Thanks to Sara Al-Sayed, Mouhammad Alhumaidi, Nevine Demitri, Michael Fauss, Gökhan Gül, Jürgen Hahn, Philipp Heidenreich, Sahar Khawatmi, Stefan Leier, Michael Leigsnering, Ahmed Mostafa, Michael Muma, Waqas Sharif, Wassim Suleiman, Fiky Suratman, Gebremichael Teame, Christian Weiss, and Feng Yin, as well as Renate Koschella and Hauke Fath. I would also like to thank the former PhD students and postdocs Raquel Fandos, Christian Debes, Weaam Alkhaldi, Marco Moebus, Uli Hammes, and Yacine Chakhchoukh.

I want to thank Heike Hoffmann, Carina Schuster, Markus Lazanowski, Christine Baur, Christian Schmitt at the Graduate School of Computational Engineering, for their kind help.

I wish to thank my friends ZHONG hua, YIN Feng, ZHANG Weibin, ZHANG Weitao, LI Zhen, LIN Wei and FEI Tai for the happy time we spent together during the years in Darmstadt.

Last but not the least, I wish to thank my parents for their unconditional love and support throughout my life.

Darmstadt, 25.10.2012

Kurzfassung

In dieser Doktorarbeit wird eine der grundlegendsten Problemstellungen der Sensorgruppen-Signalverarbeitung untersucht, nämlich die Schätzung der Anzahl der Quellensignale. Als ein Problem der Bestimmung der Modellordnung kann die Anzahl der Quellensignale auf der Grundlage der Beobachtungsdaten am Ausgang der Sensorgruppe, genauer gesagt der Stichproben-Kovarianzmatrix oder deren entsprechenden Eigenwerte und Eigenvektoren, geschätzt werden. In den letzten drei Jahrzehnten wurde diesem Problem erhebliche Aufmerksamkeit gewidmet, was die Entwicklung zahlreicher Lösungsanstze zur Folge hatte.

Es wird aufgezeigt, dass die Verteilung der Stichproben-Eigenwerte statistische Information enthält, die für die Schätzung der Anzahl der Quellensignale von großer Bedeutung ist und von existierenden Lösungsansätzen nicht oder nur unzureichend berücksichtigt wird. Dies hat zur Folge, dass etablierte Ansätze in der Praxis häufig unzufriedenstellende Ergebnisse liefern, beispielsweise im Falle sehr kleiner Stichproben, eines sehr kleinen Signal-Rausch-Verhältnisses, sowie bei engem räumlichen Abstand und hoher Korrelation der Quellensignale. In dieser Doktorarbeit werden neue Lösungsansätze unter Zuhilfenahme der Verteilung der Stichproben-Eigenwerte eingeführt. Diese Verteilungen werden hierbei durch rechenintensive Resampling-Verfahren, beispielsweise dem Bootstrap-Verfahren, geschätzt oder aus der Theorie der multivariaten Statistik und Zufallsmatrizen hergeleitet.

In dieser Doktorarbeit werden vier neuartige Lösungsansätze im Rahmen von Hypothesentests und informationstheoretischen Kriterien eingeführt. Zunächst wird eine Verbesserung des Bootstrap-basierten Testverfahrens für Szenarien mit sehr kleiner Stichprobe und impulsivem Rauschen vorgestellt. Als zweites wird ein zweistufiges Testverfahren für extrem kleine Stichproben, einschließlich dem Fall wenn die Stichprobengröße kleiner ist als die Anzahl von Sensoren, entwickelt. Drittens wird, ausgehend von dem Bayesschen Informationskriteriums (BIC) durch die Einarbeitung eines zusätzlichen Parameters, ein anpassungsfähiges Detektionskriterium vorgeschlagen. Schließlich wird ein verallgemeinertes BIC vorgestellt bei dem, im Gegensatz zum herkömmlichen BIC, zusätzlich zu der Verteilungsfunktion der Beobachtungen auch die Verteilungsfunktion der Stichproben-Eigenwerte in die log-likelihood Funktion einfließt. Erwähnenswert ist die verallgemeinerte Natur der letzten beiden Ansätze gegenüber dem herkömmlichen BIC. Theoretische Analysen und numerische Simulationen zeigen, dass die vorgeschlagenen Lösungsansätze deutlich besser als die meisten der existierenden Lösungsansätze sind.

Abstract

In this PhD thesis, one of the most fundamental problems in sensor array processing is investigated, namely, determining the number of source signals impinging on a sensor array, which is referred to as source enumeration. As a problem of model order selection, source enumeration can be addressed using the information carried in the observed data at the array output, e.g., the sample covariance matrix of the observed data, or equivalently, its sample eigenvalues and eigenvectors. In the last three decades, this problem has received a large amount of attention and numerous approaches have been developed for it.

It is shown that the distribution of the sample eigenvalues contains statistical information which is critical for the problem of source enumeration. However, such information is not taken into account by most of the existing approaches. As a result, these approaches yield unsatisfactory performance in terms of correctly detecting the number of sources in some practical situations such as very small sample size, very low signal-to-noise ratio, close spacing and high correlation of source signals. Here, distinct distributions of the sample eigenvalues are used to construct new approaches for source enumeration. The distributions are either estimated by computer-intensive resampling algorithms, such as bootstrap techniques, or derived from classical multivariate statistical theory and modern random matrix theory.

As a consequence, four novel approaches are developed in a framework of hypothesis testing or information theoretic criteria. Firstly, the bootstrap-based test is improved in order to adapt itself to the case of impulsive noise or very small sample sizes. Secondly, based on random matrix theory, a two-step test procedure is developed for the case of extremely small sample sizes, including the case when the sample size is smaller than the array size. Thirdly, inspired by the performance analysis of the Bayesian information criterion (BIC), a flexible detection criterion is proposed by incorporating an extra parameter. Finally, a generalized BIC is proposed using the distributions of the sample eigenvalues and observations to construct the log-likelihood function, in contrast to the conventional BIC which contains only the distribution of the observations. Note that the last two approaches are more flexible and general than the conventional BIC. Theoretical analysis and numerical simulations show that the proposed approaches outperform significantly most of the existing approaches.

Contents

Chapter 1

Introduction

Nowadays, it is very common to use a sensor array to receive the signals emitted from a number of sources, in order to represent the spatial and temporal characteristics of the source signals. Practical applications of sensor arrays can be found in radar and sonar systems [1–4], radio astronomy [5], smart antenna in wireless communications [6], seismology and tomography in medical diagnosis [7], just to mention a few. A brief discussion of sensor array applications is given in [7, 8].

Due to wide applications of sensor arrays, sensor array processing, or simply array processing, has been an area of tremendous expansion and development over the last three decades [8–10]. Sensor array processing is concerning processing of the outputs of an array of sensors, i.e., observations. The related problems include beamforming, detecting the number of emitting sources, estimating the locations and the waveforms of emitting sources, as well as other signal parameters, and tracking multiple moving sources, etc.

Four assumptions are commonly made in sensor array processing. Firstly, the signal wave is propagated uniformly in all directions inside isotropic and non-dispersive medium. Secondly, the sources are located in far-field such that the radius of wave propagation is much greater than the array size and a plane wave is received at the array. Thirdly, there is no coupling between sensors and the calibration is perfect. Finally, both the noise and signal have zero mean.

In sensor array processing, the vector of observations can be modeled as a superposition of a finite number of source signals impinging on the array of sensors corrupted by additive noise. The number of sources impinging on the array is known as the model order. The problem of determining the unknown number of sources, called source enumeration in this thesis, can be generally referred to as a fundamental problem of model order selection.

The problem of source enumeration has received considerable attention during the last three decades since it is of high practical interest and utmost importance to follow-on array signal processing procedures. It is known that estimation of the signal spatial parameters requires the knowledge of the number of sources. For example, the number of sources is prerequisite to some popular algorithms of directions-of-arrival (DOAs)

estimation, e.g., MUSIC [11] and ESPRIT [12]. Thus, the problem of source enumeration becomes the first step in sensor array processing. Additionally, as a problem of matrix-based model order selection, it is similar to the problems of channel number estimation and blind source separation, etc. This thesis focuses on developing new algorithms for this problem.

1.1 State-of-the-Art

Most classical approaches for source enumeration are derived based on the sample eigenvalues calculated from the sample covariance matrix of the observations since the eigenvalues carry most of the useful information and the eigenvalues can be modeled in a much simpler way than the eigenvectors.

The problem of source enumeration is originally converted to testing multiple composite hypotheses, which requires subjective setting of the significance level for the test [13]. In [14] and [15], an approach was developed using a sequence of hypothesis tests to test equality of noise eigenvalues, and in [16] and [17], a test was constructed to detect the largest noise eigenvalue. The authors of [18–22] continued the research in this direction for small sample sizes. In [18] and [19], the distributions of the test statistics related to the noise eigenvalues were refined to a more precise χ^2 distribution for small sample sizes. The test procedures yield a satisfactory performance for detecting the number of signals. In [20], a heuristic result that the profile of the ordered noise eigenvalues approximately fits an exponential law for white Gaussian noise and a short data size, was used to construct the tests. It is extended to the multi-dimensional case using Tensors in [23]. The results are impressive in some cases but the algorithm is not analytically justified. In [21][22], the bootstrap [24–26] was applied to estimate the distributions of the test statistics of the sample eigenvalues. The performance of the test is improved at the expense of computational effort.

A further category of classical approaches is based on information theoretic criteria (ITC) [27], including Akaike's information criterion (AIC) [28] and Rissanen's minimum description length (MDL) criterion [29] (or equivalently, Schwarz's Bayesian information criterion (BIC) [30]). The information theoretic criteria choose a model that fits the data mostly from a family of possible models, by minimizing some distance, e.g., the Kullback-Leibler divergence, between all competing models and the true one. The formulation consists of two terms, namely, the log-likelihood function and a penalty function. After the MDL [29] (or BIC [30]) is first proposed for source enumeration in [31], it has become the standard approach, due to its computational simplicity and

large sample consistency. A refined version of the MDL has been developed in [32–34], which leads to a formulation which differs from the original formulation of the MDL [29] and the BIC [30]. For this reason, the approach in [31] is referred to as the BIC in this thesis. The statistical performance of the BIC has been extensively analyzed in [35–44].

The conventional BIC in [31] has been improved from diverse perspectives, e.g. refinement of the penalty function [45–47] or the log-likelihood function [48], extension from the non-parametric model to the parametric model [49–52], utilization of the distribution of the sample eigenvalues [53–55], and relaxation of the assumption of temporally and spatially independent and identically distributed (i.i.d.) noise [56–60].

In [45], it is shown that the proof for consistency of the BIC in [31] is not always valid. The authors corrected the consistency proof and suggested a more general penalty function which ensures a strongly consistent criterion for the cases where the noise variance is either known or unknown and the underlying distribution is not necessarily Gaussian. The resulting criterion is referred to as the efficient detection criterion (EDC). Its convergence rate was investigated in [61]. The BIC is a special case of the EDC. In [46], it is argued that only the eigenvectors need to be considered when counting the degrees of freedom for the penalty function, whereas the signal eigenvalues and the noise variance do no matter. The reduced penalty function results in a new criterion which slightly outperforms the BIC.

In [48], a more general and more flexible criterion with strong consistency, compared to the BIC, is proposed. The log-likelihood function is replaced by a flexible function with a positive integer r, and the penalty function is adjusted according to the integer r. The authors stated that a larger value of r leads to a faster convergence. Interestingly, the BIC corresponds to the special case when $r = 1$. For the approaches in [45] and [48], it is hard to find an optimal penalty function with theoretical justification.

In [49–51], a parametric model involving the DOAs of the source signals is used instead of the commonly used non-parametric model. Therein, the number of source signals and their DOAs are estimated simultaneously. The maximum likelihood (ML) estimate of the DOAs is used to represent the projection of the sample covariance matrix onto the signal and noise subspaces. The BIC is constructed by using the eigenvalues calculated from the signal and noise subspaces. The number of sources can be more accurately estimated by benefiting from the knowledge of the estimated DOAs. These parametric approaches are applicable to non-coherent and coherent signals, whereas the aforementioned non-parametric approaches break down for coherent signals. The drawback of parametric approaches is, however, the high computational cost.

It is well known that the log-likelihood function of the BIC is constructed using the joint density of the observations. In [53], the joint density of the sample eigenvalues is used to construct the log-likelihood function. Correspondingly, the penalty function is changed since the eigenvectors are excluded from the parameter space. This approach outperforms the conventional BIC. In [54] and [55], the density of the sample eigenvalues is combined with order statistics [62], which leads to performance improvement at the cost of high computational complexity.

Most of the works mentioned above are based on the assumption that both the signals and noise are i.i.d.. However, many contributions have emerged in relaxing this assumption and making the algorithms more robust in severe practical situations. The complicated noise structure was investigated in [56–60]. The case of correlated signals was studied based on spatial smoothing techniques in [63–65], and multidimensional search in [49][51].

Instead of using the eigenvalues to construct new algorithms, the eigenvectors are used in [66–68]. In [69], the eigenvectors are combined with the bootstrap technique to deal with the non-uniform noise.

It is worth mentioning that in [70–73], information theoretic criterion is formulated based on the output of the multistage Wiener filter. By doing this, the eigendecomposition of the sample covariance matrix is avoided, leading to a more computationally efficient algorithm. Also, in [74], the eigendecomposition is replaced by QR decomposition in order to reduce the computational cost.

1.2 Research context and contributions

At first, two classical approaches are investigated, which are based on hypothesis testing and information theoretic criteria, respectively, that is, the bootstrap-based test (BBT) and Bayesian information criterion (BIC). Their performance is analyzed in detail. After that, four new approaches with good performance are developed for source enumeration, by taking into account the relationship and distributions of the sample eigenvalues, which are provided by bootstrap techniques, classical multivariate statistical theory and modern random matrix theory. On the contrary, most of the existing approaches usually only check the magnitudes of the sample eigenvalues or do not fully make use of the distribution of the sample eigenvalues. The contributions of this thesis are listed as follows:

- **Hypothesis testing:**

    1. A robust bootstrap-based test is developed based on robust statistics in order to cope with impulsive noise environments. Moreover, a modified bootstrap-based test is developed by using a new test statistic for small sample sizes.
    2. A two-step test is developed for the case of extremely small sample sizes by using the distribution of the sample eigenvalues derived from random matrix theory.

- **Information theoretic criteria:**

    1. Two new procedures are proposed to analyze the performance of the conventional BIC based on random matrix theory.
    2. A flexible detection criterion (FDC) is developed by introducing an extra parameter, inspired by the procedure of performance analysis of the BIC.
    3. A generalized Bayesian information criterion (GBIC) is developed by incorporating the distributions of the observations and sample eigenvalues.

1.3 Publications

The following publications have been produced during the period of PhD candidacy.

Internationally Refereed Journal Articles

- Z. Lu and A.M. Zoubir, Generalized Bayesian Information Criterion for Source Enumeration in Array Processing, *IEEE Trans. Signal Processing*, accepted.

- Z. Lu and A.M. Zoubir, Flexible Detection Criterion for Source Enumeration in Array Processing, *IEEE Trans. Signal Processing*, under review.

- Z. Lu and A.M. Zoubir, Source Enumeration in Array processing Using a Two-step Test Procedure, *IEEE Trans. Signal Processing*, under review.

Internationally Refereed Conference Papers

- Z. Lu and A.M. Zoubir, Source Enumeration using the PDF of Sample Eigenvalues via Information Theoretic Criteria, in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'12)*, pp. 3361-3364, Kyoto, Japan, Mar. 2012.
- Z. Lu, A.M. Zoubir, F. Roemer, and M. Haardt, Source Enumeration using the Bootstrap for Very Few Samples, in *Proc. European Signal Processing Conference (EUSIPCO'11)*, pp. 976-979, Barcelona, Spain, Aug. 2011.
- Z. Lu, Y. Chakhchoukh, and A.M. Zoubir, Source Number Estimation in Impulsive Noise Environments using Bootstrap Techniques and Robust Statistics, in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'11)*, pp. 2712-2715, Prague, Czech Republic, May 2011.
- Z. Lu, P. Heidenreich and A.M. Zoubir, Objective Quality Assessment of Speech Enhancement Algorithms using Bootstrap-based Multiple Hypotheses Tests, in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'10)*, pp. 4102-4105, Dallas TX, USA, Mar. 2010.

1.4 Thesis overview

The thesis outline is as follows:

To formulate the problem, Chapter 2 introduces briefly the array signal model and two classical approaches for source enumeration, i.e., the bootstrap-based test (BBT) and Bayesian information criterion (BIC).

Chapter 3 extends the bootstrap-based test (BBT) to two harsh situations when impulsive noise is present or the sample size is extremely small, by introducing robust statistics or a new test statistic. As a result, two new approaches are proposed, which are referred to as the robust bootstrap-based test (RBBT) and modified bootstrap-based test (MBBT), respectively.

Chapter 4 uses some results of random matrix theory to construct a two-step test procedure that yields remarkable performance under small sample size constraints, including the extreme case when the sample size is smaller than the array size. The first-step test is a thresholding approach whereas the second-step test is a simplified likelihood ratio

test. Although both steps have simple implementation, their combination outperforms some popular approaches.

Chapter 5 analyzes the performance of the BIC thoroughly by proposing two new procedures. Therein, random matrix theory is proposed to replace classical multivariate statistical theory for modeling the distribution of the sample eigenvalues, in order to predict the BIC performance more precisely.

Chapter 6 presents the flexible detection criterion (FDC). Following the procedures of performance analysis in Chapter 5, the FDC introduces a crucial tuning parameter based on the concept of thresholding. By carefully choosing this parameter, the FDC is capable of substantially reducing the probability of underestimation of the number of sources, compared to the BIC.

Chapter 7 presents the generalized Bayesian information criterion (GBIC). The GBIC incorporates the distributions of the observations and sample eigenvalues, compared to the BIC which considers the distribution of the observations only. The GBIC outperforms significantly the BIC.

Finally, conclusions are drawn in Chapter 8.

Chapter 2

Problem Formulation

In this chapter, the fundamentals for the problem of source enumeration are given. Section 2.1 introduces the signal model for the sensor array. Section 2.2 reviews two classical approaches, that is, the bootstrap-based test (BBT) and Bayesian information criterion (BIC). These two approaches are the starting point of the work in this thesis.

2.1 Array signal model

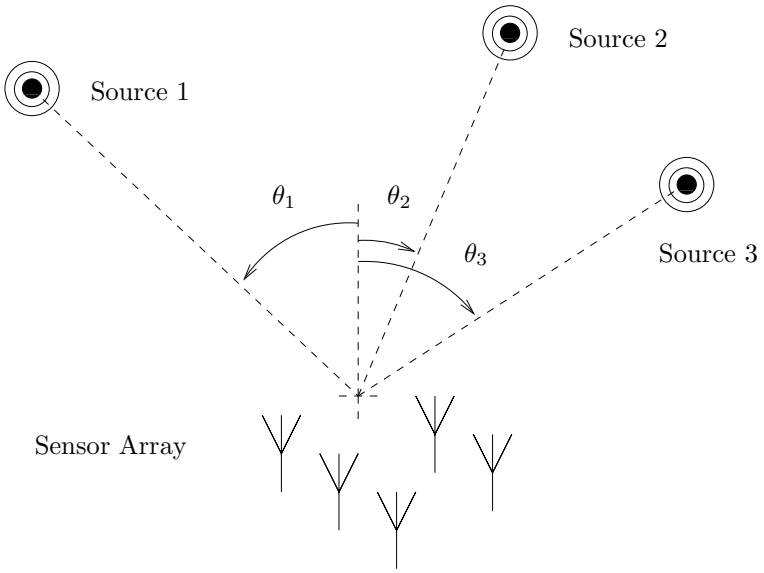

Figure 2.1. A simple sensor array configuration.

Consider an array of p sensors of arbitrary geometry (e.g., Fig. 2.1). Impinging on the array are q narrow-band signals from far-field sources located in distinct directions. In the presence of additive noise, the received vector can be expressed as

$$\begin{aligned} \mathbf{x} &= \sum_{j=1}^{q} \mathbf{a}(\theta_j) s_j + \mathbf{n} \\ &= \mathbf{A}(\boldsymbol{\theta})\mathbf{s} + \mathbf{n} \end{aligned} \tag{2.1}$$

with

$$\begin{aligned} \mathbf{x} &= [x_1, x_2, \cdots, x_p]^T \\ \boldsymbol{\theta} &= [\theta_1, \theta_2, \cdots, \theta_q]^T \\ \mathbf{A}(\boldsymbol{\theta}) &= [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \cdots, \mathbf{a}(\theta_q)] \\ \mathbf{s} &= [s_1, s_2, \cdots, s_q]^T \end{aligned} \tag{2.2}$$

where $[\cdot]^T$ denotes transpose, θ_j represents the direction-of-arrival (DOA) of the jth source signal, $\mathbf{a}(\theta_j)$ is a $p \times 1$ complex vector, referred to as the array steering vector, $\mathbf{A}(\boldsymbol{\theta})$ is the steering matrix of size $p \times q$, s_j is the complex envelope of the jth source signal, and $\mathbf{n}$ is a $p \times 1$ vector formed from the complex envelopes of the noise.

We assume that $q\,(q < p)$ signals $s_1, s_2, \cdots, s_q$ are complex, circular, ergodic Gaussian random processes, with zero mean and positive semi-definite covariance matrix. The noise vector $\mathbf{n}$ is assumed to be complex, circular, ergodic Gaussian vector process, independent of the signals, with zero mean and covariance matrix $\sigma^2\mathbf{I}$, where σ is an unknown scalar constant and $\mathbf{I}$ is the identity matrix. The matrix $\mathbf{A}(\boldsymbol{\theta})$ is of full rank.

To address the problem of determining the number of source signals q, we explore the structure of the covariance matrix of the observation vectors. The population covariance matrix of the received data is given by

$$\boldsymbol{R} = \mathrm{E}[\mathbf{x}\mathbf{x}^H] = \boldsymbol{A}\boldsymbol{R}_s\boldsymbol{A}^H + \sigma^2\boldsymbol{I} \tag{2.3}$$

where $\boldsymbol{R}_s = \mathrm{E}[\mathbf{s}\mathbf{s}^H]$ is the signal covariance, and $E[\cdot]$ and $(\cdot)^H$ denote expectation and Hermitian transpose, respectively. The population eigenvalues of $\boldsymbol{R}$ are given by

$$\lambda_1 \geq \cdots \geq \lambda_q > \lambda_{q+1} = \cdots = \lambda_p = \sigma^2 \tag{2.4}$$

where the first q eigenvalues are contributed by the sources and noise, which are called the signal eigenvalues. The remaining $p - q$ eigenvalues are contributed by noise only, and are called the noise eigenvalues. In general, the number of source signals q can be determined from the multiplicity of the smallest eigenvalues.

In practice, only a finite set of observations is available. Suppose that the received vector $\mathbf{x}$ is sampled at N time instants $t_1, t_2, \cdots, t_N$ independently. The resulting N snapshots form the $p \times N$ observation matrix $\mathbf{X} = [\mathbf{x}(t_1), \mathbf{x}(t_2), \cdots, \mathbf{x}(t_N)]$:

$$\mathbf{X} = \mathbf{A}(\boldsymbol{\theta})\mathbf{S} + \mathbf{N} \tag{2.5}$$

with $\mathbf{S} = [\mathbf{s}(t_1), \mathbf{s}(t_2), \cdots, \mathbf{s}(t_N)]$ and $\mathbf{N} = [\mathbf{n}(t_1), \mathbf{n}(t_2), \cdots, \mathbf{n}(t_N)]$. We do not have access to the population covariance matrix but its finite sample estimate, namely the

sample covariance matrix

$$\hat{\boldsymbol{R}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}(t_i)\mathbf{x}(t_i)^H = \frac{1}{N}\mathbf{X}\mathbf{X}^H. \tag{2.6}$$

The corresponding sample eigenvalues[1] of $\hat{\boldsymbol{R}}$ are

$$l_1 > \cdots > l_q > l_{q+1} > \cdots > l_p \tag{2.7}$$

which are all distinct with probability one. To enumerate the sources, most existing approaches exploit the structure, the relationships or the distributions of the sample eigenvalues in (2.7).

2.2 Classical approaches

As mentioned in Section 1.1, for the problem of source enumeration, there are mainly two categories of approaches which are based on hypothesis testing and information theoretic criteria, respectively. In the following, two examples are given, which are the bootstrap-based test (BBT) and Bayesian information criterion (BIC).

2.2.1 Bootstrap-based test (BBT)

The bootstrap-based test (BBT) is first proposed for source enumeration in [21]. In order to determine the multiplicity of the smallest eigenvalues, or the number of the smallest eigenvalues that are equal, the following set of hypotheses is constructed:

$$\begin{array}{lcll} \mathsf{H}_0 & : & \lambda_1 & = \cdots = \lambda_p \\ \vdots & & \vdots & \\ \mathsf{H}_k & : & \lambda_{k+1} & = \cdots = \lambda_p \\ \vdots & & \vdots & \\ \mathsf{H}_{p-2} & : & \lambda_{p-1} & = \lambda_p \end{array} \tag{2.8}$$

with corresponding alternatives $\mathsf{K}_0, \mathsf{K}_k, \ldots, \mathsf{K}_{p-2}$. The hypothesis H_k indicates that there exist $p-k$ noise eigenvalues and k signal eigenvalues. The tests in (2.8) are conducted sequentially with the procedure in Table 2.1[21].

[1] Without special statement, the eigenvalues mentioned in this thesis denote the sample eigenvalues, calculated from a sample covariance matrix.

Table 2.1. The sequential test procedure.

Step 1: Set $k = 0$.
Step 2: Test H_k.
Step 3: If H_k is accepted then set $\hat{q} = k$ and stop.
Step 4: If H_k is rejected and $k < p - 1$ then set $k = k + 1$ and return to Step 2. Otherwise set $\hat{q} = p - 1$ and stop.

Each hypothesis H_k in (2.8) contains a set of sub-hypotheses, which is given by

$$\mathsf{H}_k = \bigcap_{i,j} \mathsf{H}_{ij}, \quad i = k+1, \ldots, p-1, \; j = i+1, \ldots, p \tag{2.9}$$

where the sub-hypothesis

$$\mathsf{H}_{ij} : \lambda_i = \lambda_j \tag{2.10}$$

tests the difference of two eigenvalues λ_i and λ_j. The test result of H_k can be derived based on multiple hypotheses test (MHT) procedures [75] which test the group of sub-hypotheses H_{ij}'s simultaneously. In this thesis, Holm's sequentially rejective Bonferroni procedure (SRB) [76] is used. The null distribution of the test statistic H_{ij} is estimated by employing the bootstrap resampling algorithm [26][77]. This is due to the fact that, in the finite sample size case, the distribution of the sample eigenvalues and the test statistics is unknown analytically. More details about the bootstrap-based MHT can be found in [21][78].

2.2.2 Bayesian information criterion (BIC)

The information theoretic criterion is a general approach for choosing a model that fits the data mostly from a family of possible models, by minimizing the Kullback-Leibler divergence between all competing models and the true one. The criterion metric has a function of the log-likelihood of the maximum-likelihood estimator of the parameters of the competing model and a term which depends on the number of parameters of the model that penalizes the over-fitting of the model order [27]. The common procedure is

- Find a parameterized family of probability density functions (pdfs): $f(\mathcal{X}|\boldsymbol{\Theta}^{(k)})$, $\boldsymbol{\Theta}^{(k)} \in \boldsymbol{\Theta}$, given observations $\mathcal{X} = \{\mathbf{x}(t_1), \mathbf{x}(t_2), \cdots, \mathbf{x}(t_N)\}$ and corresponding unknown parameter vector $\boldsymbol{\Theta}^{(k)}$ with the possible size k in the parameter space $\boldsymbol{\Theta}$.

- Derive the maximum likelihood (ML) estimate

$$\hat{\mathbf{\Theta}}^{(k)} = \arg\min_{\mathbf{\Theta}^{(k)} \in \mathbf{\Theta}} \log f(\mathcal{X}|\mathbf{\Theta}^{(k)}).$$

- Select k such that

$$\hat{q} = \arg\min_{k} \left\{ -2\log f(\mathcal{X}|\hat{\mathbf{\Theta}}^{(k)}) + p(k) \right\} \tag{2.11}$$

  where $p(k)$ is a general penalty function associated with the kth family of distributions.

Thus, $\hat{\mathbf{\Theta}}^{(\hat{q})}$ is selected to contain the most probable parameters for representing the observations $\mathcal{X}$.

In [30], Schwarz approached the model order selection problem from the Bayesian point of view. He assumed that each model can be assigned to a prior probability, and proposed to select the model that yields the maximum posterior probability. The rule is called the Bayesian information criterion (BIC), which can be formulated as follows:

$$\hat{q} = \arg\min_{k} \left\{ \mathrm{BIC}(k) = -2\log f(\mathcal{X}|\hat{\mathbf{\Theta}}^{(k)}) + n_k \log N \right\} \tag{2.12}$$

given observations $\mathcal{X}$ and corresponding unknown parameter vector $\mathbf{\Theta}^{(k)}$ with the possible model order candidate k. n_k denotes the number of degrees of freedom of the space spanned by $\mathbf{\Theta}^{(k)}$. It is apparent that the BIC is a special case of (2.11) with a certain penalty function. A brief review for the derivation of the BIC can be found in [27]. In [29], the minimum description length (MDL) criterion is proposed by Rissanen. He assumed that each model can be used to encode the observed data and proposed to select the model that yields the minimum code length. The MDL in [29] results in the same formulation as in (2.12), i.e., the BIC. In [32][33], new description length formulae are developed for linear regression models so that the MDL differs from the BIC.

In [31], the criterion in (2.12) is used to estimate the number of sources impinging on the array. The observations $\mathcal{X} = \{\mathbf{x}(t_1), \mathbf{x}(t_2), \cdots, \mathbf{x}(t_N)\}$ are regarded as statistically i.i.d. circular symmetric complex Gaussian random vectors with zero mean. Define by $k \in \{0, 1, \cdots, p-1\}$ the set of all possible number of source signals, then the joint pdf of the observations is given by

$$\begin{aligned} f(\mathcal{X}|\mathbf{\Theta}^{(k)}) &= f(\mathbf{x}(t_1), \cdots, \mathbf{x}(t_N)|\mathbf{\Theta}^{(k)}) \\ &= \prod_{i=1}^{N} \frac{1}{\pi^p \det[\mathbf{R}^{(k)}]} \exp\left\{ -\mathbf{x}(t_i)^H \left[\mathbf{R}^{(k)}\right]^{-1} \mathbf{x}(t_i) \right\} \end{aligned} \tag{2.13}$$

with the population covariance matrix

$$\mathbf{R}^{(k)} = \sum_{i=1}^{k}(\lambda_i - \sigma^2)\mathbf{v}_i\mathbf{v}_i^H + \sigma^2\mathbf{I} \tag{2.14}$$

and the parameter vector of the model

$$\mathbf{\Theta}^{(k)} = (\lambda_1, \cdots, \lambda_k, \sigma^2, \mathbf{v}_1^T, \cdots, \mathbf{v}_k^T)^T \tag{2.15}$$

where $\lambda_1, \cdots, \lambda_k$ and $\mathbf{v}_1, \cdots, \mathbf{v}_k$ are the eigenvalues and eigenvectors of $\mathbf{R}^{(k)}$, respectively. The ML estimator [79][80]

$$\hat{\mathbf{\Theta}}^{(k)} = (l_1, \cdots, l_k, \frac{1}{p-k}\sum_{i=k+1}^{p} l_i, \mathbf{c}_1^T, \cdots, \mathbf{c}_k^T)^T \tag{2.16}$$

can be derived by maximizing the following log-likelihood function

$$\log f(\mathcal{X}|\mathbf{\Theta}^{(k)}) = -N\log\left\{\det[\mathbf{R}^{(k)}]\right\} - N\mathrm{tr}\left\{[\mathbf{R}^{(k)}]^{-1}\,\hat{\mathbf{R}}\right\} \tag{2.17}$$

where the sample covariance matrix in (2.6) can be re-written as

$$\hat{\mathbf{R}} = \sum_{i=1}^{p} l_i\,\mathbf{c}_i\mathbf{c}_i^H \tag{2.18}$$

with the ith sample eigenvalue l_i and eigenvector $\mathbf{c}_i$. From this, we obtain the BIC metric

$$\begin{aligned}\mathrm{BIC}(k) &= -2\log f(\mathcal{X}|\hat{\mathbf{\Theta}}^{(k)}) + n_k\log N \\ &= -2N(p-k)\log\left[\frac{\prod_{i=k+1}^{p} l_i^{1/(p-k)}}{\frac{1}{p-k}\sum_{i=k+1}^{p} l_i}\right] \\ &\quad + [k(2p-k)+1]\log N\end{aligned} \tag{2.19}$$

where $n_k = k(2p-k)+1$ is the number of degrees of freedom [31]. Then, we can determine the number of sources by

$$\hat{q} = \arg\min_k \mathrm{BIC}(k), \quad k = 0, \ldots, p-1. \tag{2.20}$$

Due to its computational simplicity and large sample consistency, the BIC has become the standard approach. However, the BIC performs unsatisfactorily when p/N is not so small, even breaks down when $p/N > 1$. In order to make it work for $p/N > 1$, an *ad hoc* modification

$$p = \min(p, N) \tag{2.21}$$

is needed. By doing this, the $p - N$ zero eigenvalues are totally removed and only N non-zero eigenvalues are used in the metric of (2.19).

Chapter 3

Variations of bootstrap-based test (BBT)

In this chapter, the BBT in [21] is refined in order to tackle some harsh situations when the impulsive noise is present or the sample size is extremely small. In Section 3.1, to combat impulsive noise more effectively, the concept of robust statistics [81–84] is used, more precisely, two robust estimators with high breakdown points are applied in combination with the bootstrap. In Section 3.2, to remove the inappropriate assumption of equality of the noise eigenvalues when the sample sizes is extremely small, a new test statistic is derived by using the exponential profile property of the noise eigenvalues.

3.1 Robust bootstrap-based test (RBBT)

Impulsive noise has been considered as a more accurate description for ambient noise in many communication channels such as urban and indoor radio channels, due to the impulsive nature of man-made electromagnetic interference and a large amount of natural noise [85]. The BBT in [21] relaxes the assumption of Gaussian data. However, its performance is very poor in low SNR regimes with the presence of impulsive noise causing outliers in real data. This is because, like most of the existing approaches, the eigenvalues are obtained from the sample covariance (e.g., (2.6)) of observations. The sensitivity of the sample covariance to outliers leads to the inaccuracy of eigenvalues estimation. More importantly, the bootstrap has a very low breakdown point since the bootstrap distribution may be severely affected by bootstrap samples with a higher proportion of outliers than the one in the original data set. To improve the robustness of the covariance estimator and prevent expanding outlier contamination in bootstrap samples, two robust estimators with a high breakdown point are adopted and combined with bootstrap techniques. Thus, two new approaches are proposed to estimate the number of sources in the context of small sample size, low SNR and heavy impulsive noise. In the first proposed approach, the minimum covariance determinant (MCD) estimator [84] is used to detect and discard outliers caused by impulsive noise. Then the outlier-free data are processed by the bootstrap-based test in [21]. In the second proposed approach, the traditional bootstrap is replaced by the fast and robust bootstrap (FRB)[86], where the MM-estimator is integrated. These two new approaches are referred as to robust bootstrap-based tests (RBBTs).

Note that the impulsive ambient noise $\boldsymbol{n}$ in (2.1) can be modeled by the ε-contaminated Gaussian mixture model with probability density function (pdf)

$$f = (1-\varepsilon)\mathcal{N}(0,\nu^2) + \varepsilon\mathcal{N}(0,\kappa\nu^2). \tag{3.1}$$

where $\nu > 0$, $0.01 \leq \varepsilon \leq 0.1$ and $10 \leq \kappa \leq 100$. Here, ε denotes the probability that impulses occur. The deviation from the Gaussian assumption of noise has a large distorting influence on the covariance estimator in (2.6), causing biased estimation of the number of sources, which necessitates use of robust statistics that are resistant against impulsive noise effects. In array processing, the received observation $\mathbf{x}_i$ is complex valued. Here, robust statistics are applied to the real vector $\mathbf{z}_i = [\mathsf{Re}(\mathbf{x}_i^T), \mathsf{Im}(\mathbf{x}_i^T)]^T$, which is a concatenation of the real and imaginary parts of the complex vector $\mathbf{x}_i$[1]. Then, the covariance estimator of $\mathbf{x}_i$ can be recovered from that of $\mathbf{z}_i$.

3.1.1 Formulation of the RBBT

For the first proposed approach, we apply a procedure of outlier detection as a preprocessing step. A standard approach to investigate whether a multivariate dataset contains outliers or atypical points is to calculate the Mahalanobis distances of the observations

$$\mathrm{MD}(\mathbf{x}_i) = \sqrt{(\mathbf{x}_i - \mathbf{t}_N)'\mathbf{S}_N^{-1}(\mathbf{x}_i - \mathbf{t}_N)} \quad i = 1,\ldots,n. \tag{3.2}$$

where $\mathbf{t}_N$ and $\mathbf{S}_N$ denote the sample mean and sample covariance matrix of the dataset. However, this approach does not suffice for detecting multiple outliers because of the masking effect, by which outliers do not necessarily have large Mahalanobis distances. It is due to the fact that the sample mean and sample covariance matrix are too sensitive to outliers. Hence it is better to use distances based on robust estimators.

Here, the robust Mahalanobis distance is calculated by involving the minimum covariance determinant (MCD) estimator [84]. The MCD estimator is based on h observations whose classical covariance matrix has the lowest determinant. A standard choice for h is $[(n+p+1)/2]$ which yields the maximal breakdown point[2]. Herein, these h observations are found by a fast algorithm called FAST-MCD [87] (or see Appendix 3.4.1). The MCD estimate of location is then the average of these h points, and the MCD estimate of scatter is their covariance matrix. The main steps of the first proposed approach are given in Table 3.1.

[1]To ease notation, $\mathbf{x}_i$ is used for $\mathbf{x}(t_i)$.

[2]The breakdown point of an estimator is the proportion of outliers an estimator can handle before giving an arbitrarily large result.

Table 3.1. The approach based on the MCD estimator and the bootstrap.

Step 1: Construct the real observations $\mathcal{Z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$ from the complex observations $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$.
Step 2: Obtain the sample mean $\mathbf{T}_N$ and the covariance matrix $\mathbf{C}_N$ based on h real observations derived by the FAST-MCD method.
Step 3: Calculate the robust Mahalanobis distances for $\mathbf{z}_i$: $\mathrm{RD}(\mathbf{z}_i) = \sqrt{(\mathbf{z}_i - \mathbf{T}_N)'\mathbf{C}_N^{-1}(\mathbf{z}_i - \mathbf{T}_N)},\ i = 1, \ldots, N$.
Step 4: Compare the robust Mahalanobis distances with a cutoff value $\sqrt{\chi^2_{2p,0.975}}$, and discard all the $\mathbf{z}_i$'s (and corresponding $\mathbf{x}_i$'s) with higher distance values than the cutoff value which are considered as outliers.
Step 5: Get the eigenvalues of the sample covariance (in (2.6)) of outlier-free complex-valued observations.
Step 6: Resample outlier-free complex observations, get the bootstrap samples of eigenvalues and construct the null distribution of the test statistics.
Step 7: Run the sequential test procedure in Table 2.1.

For the second proposed method, we use an MM-estimator, which is both highly robust against outliers and highly efficient for normal data [83]. It can be expressed as the solution of a fixed point equation [88]:

$$\hat{\boldsymbol{\Theta}}_N = \boldsymbol{f}_N(\hat{\boldsymbol{\Theta}}_N) \tag{3.3}$$

where $\hat{\boldsymbol{\Theta}}_N$ is a vector containing the MM-estimator of interest, e.g., location and shape. If the MM-estimator is used to calculate the covariance matrix of observations instead of the sample covariance estimator in (2.6), the result will be more robust against outliers. However, simply re-calculating MM-estimators for all bootstrap samples, i.e.,

$$\hat{\boldsymbol{\Theta}}^*_N = \boldsymbol{f}^*_N(\hat{\boldsymbol{\Theta}}^*_N) \tag{3.4}$$

will cause two main problems. Firstly, high computational cost would be needed for solving (3.4) for the required number of bootstrap samples. Secondly, in some bootstrap samples, the proportion of outliers which is much higher than the original data set, exceeds the breakdown value of the MM-estimator. And thus, it leads to unreliable MM-estimators for such bootstrap samples. To integrate the MM-estimator with the bootstrap compactly, the references [86][88] proposed an algorithm called the fast and robust bootstrap (FRB) (See Appendix 3.4.2). The MM-estimator for the bootstrap sample is given by

$$\hat{\boldsymbol{\Theta}}^*_N := \hat{\boldsymbol{\Theta}}_N + [\boldsymbol{I} - \nabla \boldsymbol{f}_N(\hat{\boldsymbol{\Theta}}_N)]^{-1}(\boldsymbol{f}^*_N(\hat{\boldsymbol{\Theta}}_N) - \hat{\boldsymbol{\Theta}}_N) \tag{3.5}$$

where the MM-estimator $\hat{\boldsymbol{\Theta}}^*_N$ for bootstrap samples is approximated by the MM-estimator $\hat{\boldsymbol{\Theta}}_N$ of the original observations which is calculated in (3.3). Since the

bootstrap sample of observations is only used to compute the function $\boldsymbol{f}_N^*$. Based on a linear correction in (3.5), computation of (3.4) is avoided. The main steps of the second proposed method are given in Table 3.2.

Table 3.2. The method based on the fast and robust bootstrap.

Step 1: Construct the real observations $\mathcal{Z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$ from the complex observations $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$.
Step 2: Calculate the MM-estimator of $\mathcal{Z}$ using (3.3), from which recover the MM-estimator of $\mathcal{X}$.
Step 3: Get the eigenvalues from the MM-estimator of $\mathcal{X}$.
Step 4: Resample $\mathcal{Z}$, calculate the MM-estimators $\hat{\boldsymbol{\Theta}}_N^{\mathbf{z}*}$ for the bootstrap samples $\mathcal{Z}^*$ using (3.5), from which recover the MM-estimators $\hat{\boldsymbol{\Theta}}_N^{\mathbf{x}*}$ for the bootstrap samples $\mathcal{X}^*$.
Step 5: Get the bootstrap samples of eigenvalues from $\hat{\boldsymbol{\Theta}}_N^{\mathbf{x}*}$ and construct the null distribution of the test statistics.
Step 6: Run the sequential test procedure in Table 2.1.

3.1.2 Simulation results

In our numerical simulations, a uniform linear array (ULA) with omni-directional sensors and inter-sensors spacing of half the wavelength was employed.[3] The steering vector can be characterized by

$$\mathbf{a}(\theta_j) = [1 \quad e^{-j\pi \sin(\theta_j)} \quad \cdots \quad e^{-j\pi (p-1) \sin(\theta_j)}]^T \tag{3.6}$$

where θ_j represents the DOA of the jth signal. The SNR is defined as

$$\text{SNR} = 10\log_{10}\frac{\sigma_s^2}{\sigma^2} \tag{3.7}$$

where σ_s^2 and σ^2 are the variance of the signals and the noise, respectively.

Simulation results were obtained based on 80 snapshots of a Gaussian source signal contaminated by impulsive noise, and 500 Monte Carlo runs. The numbers of samples, sensors and sources are denoted by N, p and q, respectively. The probability of correctly detecting q is denoted by P_c. "DOA" is short for the direction of arrival of a source. "SNR" is short for the signal-to-noise ratio. The number of bootstrap samples was chosen as $B = 200$, and a global level of significance $\alpha = 2\%$ was set for the multiple

[3]The same ULA will be employed in the simulations throughout the thesis.

hypotheses test. The SNR range in the experiment was focused on $[-16, 0]$ dB and $\kappa = 100$ was set for high impulsiveness. Since the percentage of impulsive noise is assumed to be less than 10% in our scenario, we choose $h = 0.75N$ for the MCD estimator to gain some efficiency. The breakdown value of 0.5 and the efficiency of 95% were set for the MM-estimator. The traditional bootstrap-based test [21] is denoted by "BBT", for which Holm's sequentially rejective Bonferroni procedure (SRB) is used. The two robust methods based on the MCD estimator and the FRB procedure are denoted by "MCD" and "FRB", respectively.

Suppose that we have an array with 4 sensors and 2 sources, which are located at $-10°$ and $10°$ with respect to broadside. The results are quantified by the empirical probability P_c versus SNR, see Figs. 3.1 and 3.2. In Fig. 3.1, $\varepsilon = 0$ means the noise is Gaussian. "MCD" performs similarly to "BBT" in the low SNR regime, while "MCD" suffers performance loss as the SNR increases. "FRB" shows large performance loss compared to "BBT". It is implied that it is improper to use robust statistics without impulsive noise since some observations are treated as outliers by the MCD estimator and MM-estimator. In Fig. 3.2, $\varepsilon = 1\%$ means a small fraction of impulsiveness. "BBT" suffers large performance loss compared to the case of Gaussian noise while the two proposed methods preserve their performance quite well. Generally speaking, with impulsive noise getting heavier, the performance of "BBT" degrades vary fast, and "MCD" preserves its performance better than "FRB" in the low SNR regime.

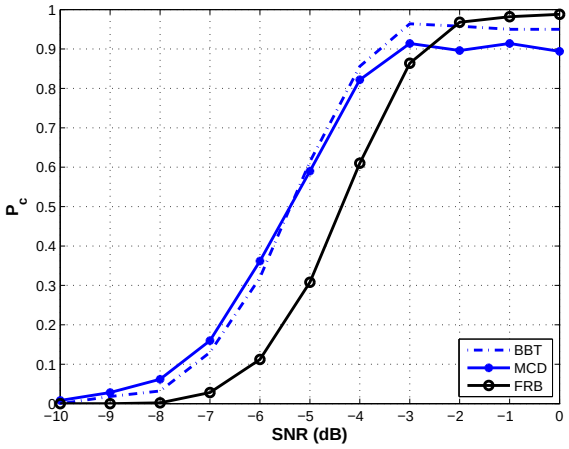

Figure 3.1. Probability of correct detection P_c in Gaussian noise with $p = 4$, $q = 2$.

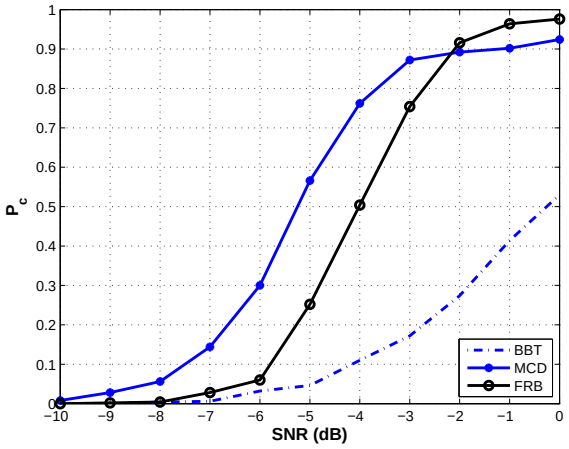

Figure 3.2. Probability of correct detection P_c for a small fraction of impulsiveness with $p = 4$, $q = 2$.

3.2 Modified bootstrap-based test (MBBT)

As we know, the bootstrap-based test in [21] tests the equality of the sample eigenvalues. The hypotheses in (2.8), which are constructed by intersections between sub-hypotheses in (2.9), are tested with a multiple hypotheses test (MHT) [75]. It means that 2^p hypotheses are tested simultaneously no matter how many sources are present. The computational cost increases exponentially with the array size p. Also, the bootstrap resampling algorithm [26] which is a computer-intensive method is employed to simulate the null distribution of the test statistic for each hypothesis. Bootstrap-based MHT formulations can be computationally expensive, especially for the high-dimensional array system. Therefore, a computationally simple test procedure is to be sought for.

In addition, the bootstrap-based test performs well when the sample size N is not extremely small, e.g., $N = 100$. When N is comparable to p, e.g., $N = p = 10$, the test breaks down since equality of the noise eigenvalues does not hold for very few samples due to the spreading of the noise eigenvalues. To remedy this problem, [21] introduced the concept of bias of the sample eigenvalues, following the result of [15]. It is necessary to reduce the bias which becomes quite significant in the very small sample size case. Through bias correction for the sample eigenvalues, the assumption that the noise-only sample eigenvalues have equal means, to some extent, is recalled. The bias-corrected bootstrap-based test continued to work for very few samples but with unsatisfactory performance.

3.2.1 Formulation of the MBBT

To improve the bootstrap-based test in [21], instead of using bias correction before the tests, we use a more accurate test statistic which reflects fluctuations of the noise eigenvalues, following the heuristic result in [20], that is, the profile of the ordered noise eigenvalues is seen to approximately fit an exponential law for white Gaussian noise and short data. Then, a relatively computationally simple bootstrap-based test procedure is constructed in order to infer the number of sources.

The authors in [20] show the approximate exponential profile of the ordered noise eigenvalues (see (2.7))

$$l_\alpha = l_\beta r_{M,N}^{\alpha-\beta}, \quad \alpha, \beta = q+1, \ldots, p \tag{3.8}$$

where $r_{M,N} = e^{-2a}(a > 0)$ denotes the exponential function of the number of the noise eigenvalues $M = p - q$ and the number of samples N. Thus the sequence of the ordered noise eigenvalues seems to be a geometric series. a can be derived based on the assumption,

$$\sum_{i=1}^{M} l_i = M\sigma^2 \tag{3.9}$$

where σ^2 denotes the noise variance, and an order-4 Taylor expansion of the hyperbolic tangent function. A corrected version [89] which removes the assumption $p \leq N$ is given as

$$a = \sqrt{\frac{1}{2}\left(\frac{15}{\mu^2+2} - \sqrt{\frac{225}{(\mu^2+2)^2} - \frac{180\mu}{\nu(\mu^2-1)(\mu^2+2)}}\right)} \tag{3.10}$$

where $\mu = \min\{M, N\}$ and $\nu = \max\{M, N\}$. It can be seen from the computation of $r_{M,N}$ that the relationship in (3.8) is valid for all sample sizes, with the extreme case of the noise eigenvalues becoming equal as N tends to infinity. Due to the relationships in (3.8) and (3.9), from preceding smaller observed noise eigenvalues, we can predict the next noise eigenvalue:

$$\tilde{l}_{p-i} = (i+1)\frac{1 - r_{i+1,N}}{1 - (r_{i+1,N})^{i+1}}\hat{\sigma}^2, \quad i = 1, \ldots, p-1 \tag{3.11}$$

with

$$\hat{\sigma}^2 = \frac{1}{i}\sum_{j=0}^{i-1} l_{p-j} \tag{3.12}$$

where $\hat{\sigma}^2$ is an estimator of the noise variance, according to (3.9). Then, we test the hypothesis

$$\begin{aligned} &\mathsf{H}_i : \lambda_{p-i} = \tilde{l}_{p-i} \quad \text{against} \\ &\mathsf{K}_i : \lambda_{p-i} \neq \tilde{l}_{p-i} \end{aligned} \tag{3.13}$$

where λ_{p-i} is the population eigenvalue with its estimate l_{p-i} which is obtained from the sample covariance matrix. The test is conducted by the bootstrap, see Table 3.3 [26], where $\theta = \lambda_{p-i}$, $\hat{\theta} = l_{p-i}$ and $\theta_0 = \tilde{l}_{p-i}$. If H_i is accepted, the observed noise eigenvalue still follows the theoretical exponential profile, that is, λ_{p-i} belongs to one of the noise eigenvalues. Otherwise, λ_{p-i} is one of the source eigenvalues. Following this statement, we construct a sequential test procedure in Table 2.1, in order to detect the number of the noise or source eigenvalues. This modified test is referred to as the modified bootstrap-based test (MBBT).

Table 3.3. The bootstrap-based test for the hypothesis $\mathsf{H} : \theta = \theta_0$ against $\mathsf{K} : \theta \neq \theta_0$.

Step 0. *Experiment.* Conduct the experiment and collect the data into the sample $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$. Calculate the test statistic

$$T_N = |\hat{\theta} - \theta_0|/\hat{\sigma},$$

where $\hat{\theta}$ is an estimator of θ and $\hat{\sigma}^2$ is an estimator of the variance σ^2 of $\hat{\theta}$.
Step 1. *Resampling.* Draw a random sample of size n, with replacement from $\mathcal{X}$

$$\mathcal{X}^* = \{\mathbf{x}_1^*, \mathbf{x}_2^*, \ldots, \mathbf{x}_N^*\}.$$

Step 2. *Calculation of the bootstrap statistic.* From $\mathcal{X}^*$, calculate

$$T_N^* = |\hat{\theta}^* - \hat{\theta}|/\hat{\sigma}^*,$$

where $\hat{\theta}^*$ and $\hat{\sigma}^*$ are computed in the same manner as $\hat{\theta}$ and $\hat{\sigma}$, but with the bootstrap sample $\mathcal{X}^*$ replacing $\mathcal{X}$.
Step 3. *Repetition.* Repeat Steps 1 and 2 many times to obtain a total of B bootstrap estimates $T_{n,1}^*, T_{n,2}^*, \ldots, T_{n,B}^*$.
Step 4. *Ranking.* Rank the collection $T_{N,1}^*, T_{N,2}^*, \ldots, T_{N,B}^*$ into increasing order to obtain

$$T_{N,(1)}^* \leq T_{N,(2)}^* \leq \ldots \leq T_{N,(B)}^*.$$

Step 5. *Test.* A bootstrap test has then the following form: reject H if $T_N > T_{(q)}^*$, where the choice of q determines the level of significance of the test and is given by $\alpha = (B + 1 - q)(B + 1)^{-1}$, where α is the nominal level of significance.

It is worth mentioning that the exponential fitting test (EFT) proposed in [20] used a more complicated test statistic, whose distribution is unknown. For this reason, the threshold for the hypothesis test was calculated by Monte Carlo simulations with a prior knowledge of the exact noise distribution. It is unrealistic in practice since it is not always possible to repeat the experiment for data collection or there is not enough a prior knowledge to run Monte Carlo simulations. In this case, the bootstrap is a proper alternative, due to its simple and attractive properties. This is validated by the

simulations.

3.2.2 Simulation results

For simplicity, the case of uncorrelated Gaussian source signals contaminated by Gaussian white noise was considered. Simulation results were obtained based on 500 Monte Carlo runs. The number of bootstrap samples was chosen as $B = 200$, and a level of significance $\alpha = 2\%$ was set for all involved hypothesis tests. The traditional bootstrap-based methods [21] without and with bias correction are denoted by "BBTe" and "BBTb"[4], respectively. Denote the methods proposed in [20] and [90] by "EFT" and "NAD", respectively. The method proposed in this section is denoted by "MBBT".

Suppose that we have an array with 8 sensors and 3 sources, which are located at $-10°, 5°, 15°$ with respect to broadside. The SNR range in this simulation was focused on $[0, 16]$ dB. Only 10 snapshots were used. The results are quantified by the empirical probability P_c of correctly detecting the source number versus SNR, see Fig. 3.3. For very few samples, the method "BBTe" breaks down completely. With bias correction of the eigenvalues, the method "BBTb" starts to work, although it performs unsatisfactorily. The method "EFT", with knowing the exact noise distribution a priori, has slightly lower detection rate than the well known method "NAD". The proposed method "MBBT" performs better than the other methods. It has highest convergence rate with respect to SNR. It is worth noting that the method "MBBT" suffers a large performance degradation at low SNRs (e.g., SNR < 0 dB).

Based on the above simulations for the very few sample case, we can see that the proposed method "MBBT" provides the best results in terms of source enumeration at a relatively high SNR when the number of samples are close to the number of sensors. It increases the performance gain substantially compared to the method "BBTb", while reducing the computational cost. The involved new test statistic is more effective in dealing with the spreading phenomenon of the noise eigenvalues than bias correction for the sample eigenvalues. It is worth mentioning that the bootstrap is a much more practical choice than Monte Carlo simulations for inferring the statistics numerically in our case. In addition, compared to the method "NAD" which has the lowest computational complexity, the minimal distributional assumptions are made for the method "BBTb".

[4]Herein, the jackknife is used for bias correction, for which Holm's sequentially rejective Bonferroni procedure (SRB) is used. More details about the jackknife are introduced in [24].

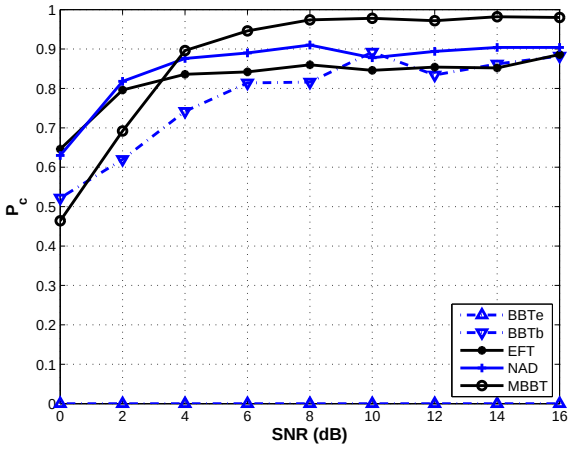

Figure 3.3. Probability of correct detection P_c vs. SNR with $N = 10$, $p = 8$, $q = 3$.

3.3 Conclusions

To cope with the case when impulsive noise is present, two robust estimators, which are highly resistant to outliers, i.e., the MCD estimator and the MM-estimator, have been incorporated into the BBT. The two proposed methods are very effective to suppress impulsive noise effects at low SNRs, since they inherit both the advantage of the bootstrap, i.e., distributional assumption relaxation and that of robust statistics, i.e., resistance towards outliers. The computational cost for the proposed methods is only sightly increased due to the fact that robust estimators are computed once for the original data rather than for all the bootstrap samples.

Moreover, a more computationally and functionally effective BBT, compared to the original one, has been proposed for the case of very few samples, i.e., the sample size is nearly equal to the array size. In light of the spreading phenomenon of the noise eigenvalues for very few samples, the property of the noise eigenvalues' exponential profile was used to construct the test statistic. Simulations show that the modified BBT outperforms the original BBT when the SNR is relatively high.

3.4 Appendix

3.4.1 FAST-MCD

Herein, a fast algorithm for the MCD estimator called FAST-MCD [87] is applied for finding the h observations whose covariance matrix has the lowest determinant. The

main scheme is given as follows.

Theorem 3.4.1 *Consider a dataset* $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ *of p-variate observations. Let* $H_1 \subset \{1, \ldots, N\}$ *with h numbers, and put* $\boldsymbol{T}_1 := \frac{1}{h}\sum_{i \in H_1} \mathbf{x}_i$ *and* $\boldsymbol{C}_1 := \frac{1}{h}\sum_{i \in H_1}(\mathbf{x}_i - \boldsymbol{T}_1)(\mathbf{x}_i - \boldsymbol{T}_1)'$. *If* $\det(\boldsymbol{C}_1) \neq 0$, *define the relative distances*

$$d_1(i) := \sqrt{(\mathbf{x}_i - \boldsymbol{T}_1)'\boldsymbol{C}_1^{-1}(\mathbf{x}_i - \boldsymbol{T}_1)}, \quad i = 1, \ldots, N, \tag{3.14}$$

which can be sorted in an ascending order. Now take the first h observation indices with the relatively small distances to construct H_2, *and compute* $\boldsymbol{T}_2$ *and* $\boldsymbol{C}_2$ *based on* H_2. *Then*

$$\det(\boldsymbol{C}_2) \leq \det(\boldsymbol{C}_1) \tag{3.15}$$

with equality if and only if $\boldsymbol{T}_2 = \boldsymbol{T}_1$ *and* $\boldsymbol{C}_2 = \boldsymbol{C}_1$.

On the basis of Theorem 1, a procedure called 'C-step' is constructed, and the corresponding scheme can be described as follows. Given the h-subset H_{old} or the pair $\boldsymbol{T}_{old}, \boldsymbol{C}_{old}$, perform the following:

1. Compute the distances d_{old} for $i = 1, \ldots, N$.
2. Sort these distances, which yields a permutation π for which $d_{old}(\pi(1)) \leq d_{old}(\pi(2)) \leq \cdots \leq d_{old}(\pi(N))$.
3. Put $H_{new} := \{\pi(1), \pi(2), \ldots, \pi(h)\}$.
4. Compute $\boldsymbol{T}_{new} := \text{ave}(H_{new})$ and $\boldsymbol{C}_{new} := \text{cov}(H_{new})$

The resulting algorithm of FAST-MCD:

1. Construct an initial h-subset H_1. Draw a random $(p+1)$-subset $\boldsymbol{J}$, and then compute $\boldsymbol{T}_0 := \text{ave}(\boldsymbol{J})$ and $\boldsymbol{C}_0 := \text{cov}(\boldsymbol{J})$. If $\det(\boldsymbol{C}_0) = 0$, then extend $\boldsymbol{J}$ by adding another random observation, and continue adding observations until $\det(\boldsymbol{C}_0) > 0$. Compute the distances $d_0(i) := \sqrt{(\mathbf{x}_i - \boldsymbol{T}_0)'\boldsymbol{C}_0^{-1}(\mathbf{x}_i - \boldsymbol{T}_0)}$ for $i = 1, \ldots, N$. Sort them into $d_0(\pi(1)) \leq \cdots \leq d_0(\pi(N))$ and put $H_1 := \{\pi(1), \pi(2), \ldots, \pi(h)\}$.
2. Carry out two C-steps.
3. Repeat Step 1 and 2, e.g., 500 times.
4. Find 10 h-subsets H_3 with lowest $\det(\boldsymbol{C}_3)$ and carry out C-steps until convergence.
5. Find the h-subset with lowest $\det(\boldsymbol{C})$ and corresponding $(\boldsymbol{T}, \boldsymbol{C})$.

3.4.2 Fast and robust bootstrap (FRB)

Multivariate MM-estimators of location, shape and covariance are now defined as follows.

Theorem 3.4.2 *Let $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ which are p-variate observations with $N \geq p+1$, and let $(\tilde{\boldsymbol{\mu}}_N, \tilde{\boldsymbol{\Sigma}}_N)$ be multivariate S-estimators of location and scatter, that is $(\tilde{\boldsymbol{\mu}}_N, \tilde{\boldsymbol{\Sigma}}_N)$ minimize $|\boldsymbol{C}|$ subject to*

$$\frac{1}{N}\sum_{i=1}^{N}\rho_0\left(\sqrt{(\mathbf{x}_i-\boldsymbol{T})'\boldsymbol{C}^{-1}(\mathbf{x}_i-\boldsymbol{T})}\right) = b \tag{3.16}$$

over all $(\boldsymbol{T}, \boldsymbol{C})$ corresponding to location and scatter estimates. b is set to ensure consistency at the multivariate normal distribution. Denote $\hat{\sigma}_N := |\tilde{\boldsymbol{\Sigma}}_N|^{1/(2p)}$. Then the multivariate MM-estimators for location and shape $(\hat{\boldsymbol{\mu}}_N, \hat{\boldsymbol{\Gamma}}_N)$ minimize

$$\frac{1}{N}\sum_{i=1}^{N}\rho_1\left(\sqrt{(\mathbf{x}_i-\boldsymbol{T})'\boldsymbol{G}^{-1}(\mathbf{x}_i-\boldsymbol{T})}/\hat{\sigma}_N\right) \tag{3.17}$$

over all $(\boldsymbol{T}, \boldsymbol{G})$ corresponding to location and shape estimators for which $|\boldsymbol{G}| = 1$. The MM-estimator for the covariance matrix is $\hat{\boldsymbol{\Sigma}}_N = \hat{\sigma}_N^2\hat{\boldsymbol{\Gamma}}_N$. The loss functions ρ_0, ρ_1 are defined by Tukey's biweight functions

$$\rho(x) = \begin{cases} \frac{x^2}{2} - \frac{x^4}{2c^2} + \frac{x^6}{6c^4}, & |x| \leq c \\ \frac{c^2}{6}, & |x| \geq c. \end{cases} \tag{3.18}$$

where $c > 0$ is a user-chosen tuning constant. By selecting different c for ρ_0,and ρ_1, expected breakdown point and efficiency could be obtained, respectively.

The multivariate MM-estimators as defined in Definition 1 can be expressed as a solution of fixed-point equations as follows:

$$\tilde{\boldsymbol{\mu}}_N = \left(\sum_{i=1}^{N}\tilde{w}_i\right)^{-1}(\tilde{w}_i\mathbf{x}_i) \tag{3.19}$$

$$\tilde{\boldsymbol{\Sigma}}_N = \frac{1}{Nb - \sum_{i=1}^{N}\tilde{v}_i}\left(\sum_{i=1}^{N}p\tilde{w}_i(\mathbf{x}_i-\tilde{\boldsymbol{\mu}}_N)(\mathbf{x}_i-\tilde{\boldsymbol{\mu}}_N)'\right) \tag{3.20}$$

$$\hat{\boldsymbol{\mu}}_N = \left(\sum_{i=1}^{N}w_i\right)^{-1}(w_i\mathbf{x}_i) \tag{3.21}$$

$$\hat{\boldsymbol{\Gamma}}_N = G\left(\sum_{i=1}^{N}w_i(\mathbf{x}_i-\hat{\boldsymbol{\mu}}_N)(\mathbf{x}_i-\hat{\boldsymbol{\mu}}_N)'\right) \tag{3.22}$$

where $G(A) = |A|^{-1/p}A$ for $p \times p$ matrix A, $\tilde{w}_i = \rho_0'(\tilde{d}_i)/\tilde{d}_i$, $w_i = \rho_1'(d_i/|\tilde{\Sigma}_N|^{1/(2p)})/d_i$, $\tilde{v}_i = \rho_0(\tilde{d}_i) - \rho_0'(\tilde{d}_i)\tilde{d}_i$ and $\tilde{d}_i = [(\mathbf{x}_i - \tilde{\boldsymbol{\mu}}_N)'\tilde{\boldsymbol{\Sigma}}_N^{-1}(\mathbf{x}_i - \tilde{\boldsymbol{\mu}}_N)]^{1/2}$, $d_i = [(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_N)'\hat{\boldsymbol{\Gamma}}_N^{-1}(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_N)]^{1/2}$. Given a bootstrap sample $\mathcal{X}^* = \{\mathbf{x}_1^*, \ldots, \mathbf{x}_N^*\}$, an approximation of recalculated estimators would be as follows:

$$\tilde{\boldsymbol{\mu}}_N^* = \left(\sum_{i=1}^N \tilde{w}_i^*\right)^{-1} (\tilde{w}_i^*\mathbf{x}_i^*) \tag{3.23}$$

$$\tilde{\boldsymbol{\Sigma}}_N^* = \frac{1}{Nb - \sum_{i=1}^N \tilde{v}_i^*}\left(\sum_{i=1}^N p\tilde{w}_i^*(\mathbf{x}_i^* - \tilde{\boldsymbol{\mu}}_N)(\mathbf{x}_i^* - \tilde{\boldsymbol{\mu}}_N)'\right) \tag{3.24}$$

$$\hat{\boldsymbol{\mu}}_N^* = \left(\sum_{i=1}^N w_i^*\right)^{-1} (w_i^*\mathbf{x}_i^*) \tag{3.25}$$

$$\hat{\boldsymbol{\Gamma}}_N^* = G\left(\sum_{i=1}^N w_i^*(\mathbf{x}_i^* - \hat{\boldsymbol{\mu}}_N)(\mathbf{x}_i^* - \hat{\boldsymbol{\mu}}_N)'\right) \tag{3.26}$$

Where $\tilde{w}_i^* = \rho_0'(\tilde{d}_i^{\,*})/\tilde{d}_i^{\,*}$,$w_i^* = \rho_1'(d_i^*/|\tilde{\Sigma}_N|^{1/(2p)})/d_i^*$, $\tilde{v}_i^* = \rho_0(\tilde{d}_i^{\,*}) - \rho_0'(\tilde{d}_i^{\,*})\tilde{d}_i^{\,*}$, and $\tilde{d}_i^{\,*} = [(\mathbf{x}_i^* - \tilde{\boldsymbol{\mu}}_N)'\tilde{\boldsymbol{\Sigma}}_N^{-1}(\mathbf{x}_i^* - \tilde{\boldsymbol{\mu}}_N)]^{1/2}$, $d_i^* = [(\mathbf{x}_i^* - \hat{\boldsymbol{\mu}}_N)'\hat{\boldsymbol{\Gamma}}_N^{-1}(\mathbf{x}_i^* - \hat{\boldsymbol{\mu}}_N)]^{1/2}$. Note that since we are keeping the estimators $\tilde{\boldsymbol{\mu}}_N, \tilde{\boldsymbol{\Sigma}}_N, \hat{\boldsymbol{\mu}}_N, \hat{\boldsymbol{\Gamma}}_N$ fixed on the right hand side of equations, these approximations will likely underestimate the variability of the MM-estimator. To remedy this a linear correction can be applied as follows. The equations (3.19)-(3.22) and (3.23)-(3.26) can be summarized respectively as

$$\hat{\boldsymbol{\Theta}}_N = \boldsymbol{f}_N(\hat{\boldsymbol{\Theta}}_N) \tag{3.27}$$

$$\hat{\boldsymbol{\Theta}}_N^{1*} = \boldsymbol{f}_N^*(\hat{\boldsymbol{\Theta}}_N) \tag{3.28}$$

where $\hat{\boldsymbol{\Theta}}_N := ((\tilde{\boldsymbol{\mu}}_N)', \mathrm{vec}(\tilde{\boldsymbol{\Sigma}}_N)', \mathrm{vec}(\hat{\boldsymbol{\mu}}_N)', \mathrm{vec}(\hat{\boldsymbol{\Gamma}}_N)')'$. The notation 'vec' means vectorization of the matrix. A Taylor expansion about limiting value of $\hat{\boldsymbol{\Theta}}_N$ can be derived:

$$\hat{\boldsymbol{\Theta}}_N = \boldsymbol{f}_N(\boldsymbol{\Theta}) + \nabla\boldsymbol{f}_N(\boldsymbol{\Theta})(\hat{\boldsymbol{\Theta}}_N - \boldsymbol{\Theta}) + \boldsymbol{R}_N \tag{3.29}$$

where $\boldsymbol{\Theta} := ((\tilde{\boldsymbol{\mu}})', \mathrm{vec}(\tilde{\boldsymbol{\Sigma}})', \mathrm{vec}(\hat{\boldsymbol{\mu}})', \mathrm{vec}(\hat{\boldsymbol{\Gamma}})')'$, $\boldsymbol{R}_N$ is the remainder term and $\nabla\boldsymbol{f}(.)$ is the matrix of partial derivatives. If $\boldsymbol{R}_N$ is sufficiently small, Equation (3.29) could be rewritten as

$$\sqrt{N}(\hat{\boldsymbol{\Theta}}_N - \boldsymbol{\Theta}) \sim [\boldsymbol{I} - \nabla\boldsymbol{f}_N(\boldsymbol{\Theta})]^{-1}\sqrt{N}(\boldsymbol{f}_N(\boldsymbol{\Theta}) - \boldsymbol{\Theta}) \tag{3.30}$$

where $\sim$ denotes both sides have the same limiting distribution. Under certain conditions $\sqrt{N}(\hat{\boldsymbol{\Theta}}_N^* - \hat{\boldsymbol{\Theta}}_N) \sim \sqrt{N}(\hat{\boldsymbol{\Theta}}_N - \boldsymbol{\Theta})$ and $\sqrt{N}(\boldsymbol{f}_N^*(\hat{\boldsymbol{\Theta}}_N) - \hat{\boldsymbol{\Theta}}_N) \sim \sqrt{N}(\boldsymbol{f}_N(\boldsymbol{\Theta}) - \boldsymbol{\Theta})$ hold. If $[\boldsymbol{I} - \nabla\boldsymbol{f}_N(\boldsymbol{\Theta})]^{-1}$ is further approximated by $[\boldsymbol{I} - \nabla\boldsymbol{f}_N(\hat{\boldsymbol{\Theta}}_N)]^{-1}$, we obtained

$$\sqrt{N}(\hat{\boldsymbol{\Theta}}_N^* - \hat{\boldsymbol{\Theta}}_N) \sim [\boldsymbol{I} - \nabla\boldsymbol{f}_N(\hat{\boldsymbol{\Theta}}_N)]^{-1}\sqrt{N}(\boldsymbol{f}_N^*(\hat{\boldsymbol{\Theta}}_N) - \hat{\boldsymbol{\Theta}}_N), \tag{3.31}$$

from which we can obtain

$$\hat{\boldsymbol{\Theta}}_N^{R*} := \hat{\boldsymbol{\Theta}}_N + [\boldsymbol{I} - \nabla\boldsymbol{f}_N(\hat{\boldsymbol{\Theta}}_N)]^{-1}(\hat{\boldsymbol{\Theta}}_N^{1*} - \hat{\boldsymbol{\Theta}}_N) \tag{3.32}$$

which is the approximation to $\hat{\boldsymbol{\Theta}}_N^*$.

Chapter 4
Two-step test procedure

In many state-of-the-art radar and sonar systems, it has become increasingly common that the number of sensors p is close to or even exceeds the available sample size N, e.g., $p/N > 1$, due to the increasing growth of system dimensions and the need for even faster computational speed. Also, in many situations, the sample size is forced to a small number due to the non-stationary nature of the observed process so that only quasi-stationary properties are assumed. Of practical interest is thus to detect the number of source signals using a limited number of samples N, which is comparable to or even smaller than the number of sensors p.

Most of the existing approaches rely on the asymptotic distribution of the sample eigenvalues known from classical multivariate statistical theory, e.g., Anderson's work [79], which is based on a fixed-system-size large-sample-size asymptotic regime, i.e., p fixed, $N \to \infty$. It is implied that the ratio $p/N \to 0$. With the support of large sample sizes, Anderson's analysis suggests that the sample eigenvalues are symmetrically centered around the population eigenvalues. When the sample size is sufficiently large, the sample eigenvalues, which have the same population eigenvalue, are approximately equal. However, when the sample size is extremely small, namely not-so-small ratio p/N, Anderson's large-sample asymptotics are no longer valid so that most approaches suffer significant performance degradation.

Large dimensional random matrix theory [91], which originates from the works of Wigner [92] and Marčenko and Pastur [93], has emerged recently as a basic tool to deal with the case of not-so-small p/N when classical multivariate statistical theory fails. Random matrix theory characterizes the distribution of the sample eigenvalues using the large-system-size large-sample-size asymptotic regime, i.e., $p, N \to \infty$ with $p/N \to c \in (0, \infty)$. It can describe the asymmetrical spreading phenomenon of the sample eigenvalues when the ratio p/N is not small. It is justified by the work in [43, 90, 94–99] that random matrix theory provides a more precise approximation for the distribution of the sample eigenvalues in finite sample size settings than classical multivariate statistical theory. In other words, random matrix theory is a more appropriate tool in terms of dealing with the small sample size case. The use of random matrix theory in wireless communications has attracted much attention recently. Interesting applications can be found in [100].

In this chapter, we use some results of random matrix theory to construct a two-step test procedure that yields remarkable performance under small sample size constraints.

For the first-step test, we find a threshold for discriminating the signal eigenvalues from the noise eigenvalues. This thresholding algorithm is quite simple in view of implementation, but at the same time has extremely good performance. In order to improve the first-step test result, we propose to conduct the second-step test, namely a likelihood ratio test, which is constructed based on the distribution of the sample eigenvalues. Simulations show that the proposed two-step test procedure substantially outperforms many existing methods in difficult situations, e.g., very small sample size, very low signal-to-noise ratio (SNR), close spacing and high correlation of source signals.

The remainder of the chapter is organized as follows. Some pertinent results of random matrix theory is introduced in Section 4.1, based on which a two-step test procedure is presented in Section 4.2. After that, simulation results are given in Section 4.3 before the conclusions are drawn in Section 4.4.

4.1 Random matrix theory

In most cases, the number of sources is much smaller than the number of sensors, i.e., $q \ll p$, which means that the population covariance matrix in (2.3) is a small rank perturbation of an identity matrix. Such a population model, which contains a purely noise model spiked with a few significant signal eigenvalues, has been called the spiked population model [101]. Under this model, we briefly introduce three theoretical results from large dimensional random matrix theory in this section, which will be used to derive new algorithms for source enumeration in the next section.

The first result is about the Marčenko-Pastur law, which can be used to model the distribution of the signal-free noise eigenvalues. This result is proved in [93, 102–105].

Theorem 4.1.1 *Let* $\mathbf{N}$ *be a* $p \times N$ *matrix whose entries* n_{ij} *are i.i.d. with mean zero and variance* σ^2*. Let* $\hat{\mathbf{R}}_n$ *be the sample covariance matrix of* $\mathbf{N}$ *with corresponding eigenvalues* $l_1 > l_2 > \cdots > l_p$*. The spectral distribution of* $\hat{\mathbf{R}}_n$*, that is, the empirical distribution function (edf) of the eigenvalues of* $\hat{\mathbf{R}}_n$*, is given by*

$$F_{\hat{\mathbf{R}}_n}(l) = \frac{1}{p} \sum_{i=1}^{p} \mathbf{1}_{l_i \leq l} \tag{4.1}$$

where the indicator function $\mathbf{1}_{l_i \leq l} = 1$ *for* $l_i \leq l$*. Assume the dimension to sample size ratio* $p/N \to c > 0$ *as* $N \to \infty$*. Under the assumption that the entries* n_{ij} *have a*

finite eighth moment, $F_{\hat{\mathbf{R}}_n}(l)$ converges w.p.1 to the Marčenko-Pastur distribution F_c for every l, with the density

$$\begin{aligned} F_c'(l) = f_c(l) = &\max\left(0, 1 - \frac{1}{c}\right)\delta(l) \\ &+ \frac{1}{2\pi c\sigma^2 l}\sqrt{(l-a)(b-l)}\mathbf{1}_{a\leq l\leq b} \end{aligned} \tag{4.2}$$

where

$$\begin{aligned} a =& \sigma^2(1-\sqrt{c})^2 \\ b =& \sigma^2(1+\sqrt{c})^2. \end{aligned} \tag{4.3}$$

Herein, w.p.1 denotes convergence with probability one (i.e., almost sure convergence). The indicator function $\mathbf{1}_{a\leq l\leq b} = 1$ for $a \leq l \leq b$ and zero otherwise, $\delta(l)$ is the Dirac delta function. For $0 < c \leq 1$, the density $f_c(l)$ has only the second term, whereas for $c > 1$, it has a mass point at the origin due to $p - N$ zero eigenvalues. The convergence rate of the edf of the eigenvalues of $\hat{\mathbf{R}}_n$ to the Marčenko-Pastur distribution F_c in probability is $O(N^{-2/5})$ and the convergence rate of the expected edf for $c < 1$ is $O(N^{-1/2})$ [106][104]. Moreover, suppose that the entries n_{ij} have a finite fourth moment, then the largest and smallest non-zero eigenvalues converge w.p.1 to the edges of the support of f_c [105, 107–111]:

$$\begin{aligned} l_1 &\to b \\ l_{\min(p,N)} &\to a. \end{aligned} \tag{4.4}$$

It is shown in [103] that the Marčenko-Pastur density still holds for the spiked model. However, (4.4) is not guaranteed and some eigenvalues are not necessarily in the support of the Marčenko-Pastur density.

For simplicity, we assume $c = p/N$. The Marčenko-Pastur density is plotted for different values of c in [90]. When c converges to 0, the noise eigenvalues are increasingly clustered around the true noise variance σ^2. When c is not small enough, the noise eigenvalues spread asymmetrically with respect to σ^2, which differs from Anderson's results [79]. The smaller the sample size N is, the more widely the noise eigenvalues spread. Although the Marčenko-Pastur density is asymptotic in terms of p and N, numerical experiments show that it remains a good approximation for small system and sample sizes, which makes the Marčenko-Pastur density promising in a wide range of application for different p and N.

The second result describes the so-called phase transition phenomenon for signal eigenvalues in the spiked model. If the signal strength is not larger than a certain threshold,

the corresponding signal eigenvalue converges to the upper limit b of the support of the Marčenko-Pastur density in Theorem 4.1.1, otherwise it is pulled up to a higher limit than b. Proofs can be found in [112] and [113].

Theorem 4.1.2 *Let $\hat{\mathbf{R}}$ denote a sample covariance matrix formed from the $p \times N$ matrix of observations, whose columns are i.i.d. with mean zero and covariance $\mathbf{R}$. Suppose that the fourth moment of the entries of $\hat{\mathbf{R}}$ exists. The corresponding population eigenvalues and sample eigenvalues are given as $\lambda_1 \geq \cdots \geq \lambda_q > \sigma^2 = \cdots = \sigma^2$ and $l_1 > l_2 > \cdots > l_p$, respectively. Denote the ith signal strength by $\nu_i = \lambda_i - \sigma^2$. Then as $p, N \to \infty$ with $p/N \to c \in (0, \infty)$, the ith signal sample eigenvalue l_i converges w.p.1 to*

$$l_i \to \begin{cases} (\nu_i + \sigma^2)(1 + c\sigma^2/\nu_i) & \text{if } \nu_i > \sigma^2\sqrt{c} \\ \sigma^2(1+\sqrt{c})^2 & \text{if } \nu_i \leq \sigma^2\sqrt{c} \end{cases} \tag{4.5}$$

for $i = 1, \ldots, q$.

The threshold $\sigma^2\sqrt{c}$ can be referred to as the non-parametric asymptotic limit of detection, which means that the asymptotically detectable signal has strength ν_i, which is larger than $\sigma^2\sqrt{c}$.

The third result describes the limiting fluctuations or distribution of the signal eigenvalue with strength $\nu_i > \sigma^2\sqrt{c}$. Such signal eigenvalues are distributed normally around the limiting value $(\nu_i + \sigma^2)(1 + c\sigma^2/\nu_i)$ given in Theorem 4.1.2. Proofs are given in [112], [114] and [115].

Theorem 4.1.3 *Let $\hat{\mathbf{R}}$ denote a sample covariance matrix estimated from the $p \times N$ matrix of Gaussian observations, whose columns are i.i.d. with mean zero and population covariance $\mathbf{R}$. The corresponding population eigenvalues and sample eigenvalues are given as $\lambda_1 > \cdots > \lambda_q > \sigma^2 = \cdots = \sigma^2$ and $l_1 > l_2 > \cdots > l_p$, respectively. In the joint limit $p, N \to \infty$ with $p/N \to c \in (0, \infty)$, for the ith signal strength $\nu_i = \lambda_i - \sigma^2 > \sigma^2\sqrt{c}$, at a convergence rate of $O(N^{-1/2})$, the density of the corresponding signal sample eigenvalue l_i converges w.p.1 to the normal density*

$$l_i \xrightarrow{\mathcal{D}} \mathcal{N}\left(\tau_i, \delta_i^2\right) \tag{4.6}$$

with

$$\begin{aligned} \tau_i &= (\nu_i + \sigma^2)(1 + c\sigma^2/\nu_i) \\ \delta_i &= (\nu_i + \sigma^2)\sqrt{\frac{2}{\beta N}(1 - c\sigma^4/\nu_i^2)} \end{aligned} \tag{4.7}$$

where $\beta = 2$ for complex-valued observations.

To illustrate the above, we plotted the Marčenko-Pastur density and the normal density in Fig. 4.1. The red curve represents the Marčenko-Pastur density, including the mass point at the origin. The blue curve corresponds to the normal density. The black dashed-dotted line is for the mean of the normal density τ_i, which is also the limiting value when $\nu_i > \sigma^2\sqrt{c}$ in Theorem 4.1.2.

We give some remarks for the preceding three theorems. Firstly, Theorems 4.1.2 and 4.1.3 have different convergence rates. To the best of our knowledge, the convergence rate for Theorem 2 is still unclear, but intuitively, it is supposed to be slower than the one for Theorem 3. Secondly, the entries of the involved matrix do not have to follow the Gaussian distribution in Theorems 4.1.1 and 4.1.2. The Gaussian assumption is needed only in Theorem 4.1.3. Thirdly, the signal eigenvalues must be of multiplicity one in Theorem 4.1.3 whereas not in Theorem 4.1.2. So if we want to use Theorem 4.1.3, the signal population eigenvalues in (2.4) should have multiplicity one, which, in fact, is also the implicit assumption for the BIC. Finally, the noise variance σ^2 is involved in all the three theorems above. In our work, it is assumed that $\sigma^2 = 1$, in order to represent the pdf of the eigenvalues precisely without impact of the noise variance estimate.

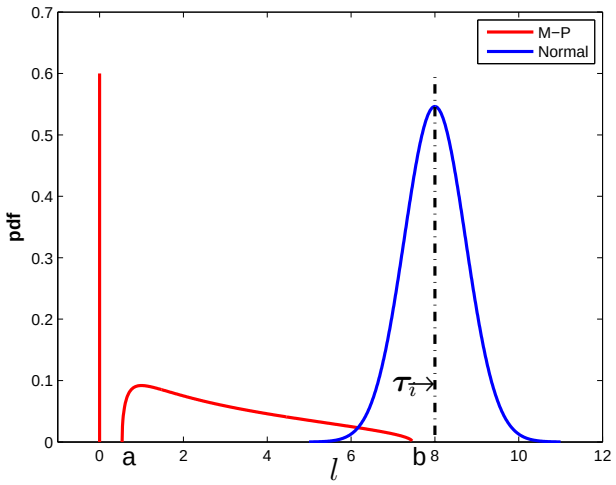

Figure 4.1. The Marčenko-Pastur density and the normal density ($N = 20, c = 3, \sigma^2 = 1, \nu_i = 3$).

4.2 Formulation of two-step test procedure

Recently, random matrix theory has been used to detect the number of source signals, or discriminate the signal eigenvalues from the noise eigenvalues. The performance

superiority over classical approaches in the case of finite sample sizes, can be found among others, in [43, 90, 94–99]. Two examples are introduced briefly in the sequel. They will be included for performance comparison in our simulations in Section 4.3.

The spreading phenomenon of noise eigenvalues can be modeled by the Marčenko-Pastur density with a simple closed-form expression. In [90], it is shown that based on the Marčenko-Pastur density, a noise-variance-independent statistic of the noise eigenvalues

$$q_k = \frac{\frac{1}{p-k}\sum_{i=k+1}^{p} l_i^2}{(\frac{1}{p-k}\sum_{i=k+1}^{p} l_i)^2} \tag{4.8}$$

has an asymptotic normal density $f(q_k)$, which is used to construct the log-likelihood function of the AIC metric, and leads to an AIC-like criterion

$$\hat{q} = \arg\min_k \left\{-2\log f(q_k) + 2(k+1)\right\}, \quad k = 0, \ldots, \min(p, N) - 1, \tag{4.9}$$

which does not rely on the noise variance σ^2. However, simulations show that the above criterion yields unsatisfactory results.

The fluctuation of the largest noise eigenvalue can be modeled by the Tracy-Widom distribution under the assumption of Gaussian data [116–118]. In [94], the number of source signals is found via detecting the largest noise eigenvalue. The sequential tests are conduced as follows:

$$\hat{q} = \min\left\{k: \quad l_{k+1} < \sigma^2 C(\alpha)\right\}, \quad k = 0, \ldots, \min(p, N) - 1, \tag{4.10}$$

where σ^2 denotes the noise variance, and $C(\alpha)$ depends on the false alarm rate α, as well as the centering and scaling parameters related to the Tracy-Widom distribution. Simply put, the Tracy-Widom distribution is used to construct the thresholds for the sequential tests. Nevertheless, there is no explicit closed-form expression for the Tracy-Widom distribution so that analytical inference becomes impossible in some cases.

Here, we present a two-step test procedure based on random matrix theory for source enumeration. The first-step test is constructed based on Theorems 4.1.1 and 4.1.2, and the second-step test is based on Theorems 4.1.1 and 4.1.3.

4.2.1 The first-step test

As we know, Theorem 4.1.1 also establishes the almost sure convergence of the largest noise eigenvalue to $b = \sigma^2(1 + \sqrt{c})^2$, see (4.4). Theorem 4.1.2 states that the asymptotically detectable signal eigenvalues converge to a limit which is much larger than b,

and the non-detectable signal eigenvalues converge to the limit b, which can be treated approximately as noise eigenvalues. Recall Fig. 4.1, where the non-detectable signal sample eigenvalues stay at the point b, whereas the detectable ones jump to the black dashed-dotted line. Therefore, the limit value $b = \sigma^2(1+\sqrt{c})^2$ is the asymptotic separator between noise eigenvalues and signal eigenvalues. It can be chosen as the threshold for the hypothesis test which tests the largest noise eigenvalue. The number of sources can be detected as follows:

$$q_0 = \min\left\{k: \quad l_{k+1} < b = \sigma^2(1+\sqrt{c})^2\right\}, \quad k = 0, \ldots, p-1. \tag{4.11}$$

This test is very simple in terms of implementation, but simulations show that it gives superior performance. It is worth noting that the test in (4.11) does not require subjective setting of the threshold level, i.e., the false alarm rate, which is different from classical hypothesis testing. Since there is no Gaussian data assumption in Theorems 4.1.1 and 4.1.2, this test can be used in the non-Gaussian case.

The test given in (4.11) and the test proposed in [94] are similar in spirit to Roy's largest root test [16], which finds the largest signal-free eigenvalue by checking the significance of the eigenvalues one by one. Nevertheless, the threshold of the test in (4.11) is derived based on the the Marčenko-Pastur density and the phase transition phenomenon, whereas the threshold used in [94] is based on the Tracy-Widom distribution.

It is noteworthy that the possible candidate k of the number of sources essentially has an upper bound which is smaller than $p-1$. According to Theorem 4.1.2, if the signal is detectable, the condition $\nu_i > \sigma^2\sqrt{c}$ should be fulfilled. Based on Weyl's inequality (see Appendix 4.5.1), we obtain[1]

$$\nu_i + \rho_{\min(p,N)} \leq l_i \leq \nu_i + \rho_1 \tag{4.12}$$

where ρ is the noise level, or the pure noise eigenvalue, with $\rho_i = l_i$ for $i = q+1, \ldots, \min(p, N)$. It follows from (4.4) and (4.12) that the corresponding l_i must be larger than $\sigma^2\sqrt{c} + a$. Then the possible candidate k of the number of sources is supposed to be not larger than d, as given by

$$k \leq d = \max\left\{i: \quad l_i > \sigma^2\sqrt{c} + a\right\}, \quad i = 1, \ldots, \min(p, N) \tag{4.13}$$

where d is useful in some cases, e.g., for estimating an initial noise variance.

Due to the fact that (4.4) and (4.5) hold asymptotically, the test in (4.11) is consistent in terms of detecting the number of sources q. The convergence rate is lower than $O(N^{-1/2})$. However, it is shown in the following that for a wide range of SNRs and sample sizes, the test

[1]Here, $\nu_i \approx \hat{\nu}_i$ is assumed, where $\hat{\nu}_i$ is the ML estimate in the finite sample size case.

1. tends to underestimate q in the case of low SNRs and finite sample sizes,

2. underestimates q mostly by one when underestimation occurs.

The validation of (4.4) and (4.5) needs an extremely large sample size. In the case of low SNRs and finite sample sizes, (4.4) and (4.5) are not precise enough, especially the latter, that is, the phase transition phenomenon is not significant. The smallest signal eigenvalues are very close to the largest noise eigenvalues, or in the worst case, they are mixed with each other. In such a situation, those smallest signal eigenvalues can be detected as noise eigenvalues. So the number of sources q is underestimated. The underestimation implies that the threshold b in (4.11) is higher than the true one, which is proven below.

Define l_{q+1} as the largest noise eigenvalue of the signal bearing covariance matrix, $\hat{l}_{q+1}$ as the $q+1$ largest eigenvalue of the signal-free covariance matrix, and $\tilde{l}_{q+1}$ as the largest eigenvalue of a $(p-q)\times(p-q)$ sub-matrix of the signal-free covariance matrix. Based on the Marčenko-Pastur law, $\tilde{l}_{q+1}$ converges asymptotically to

$$\tilde{b} = \sigma^2 \left(1 + \sqrt{\frac{p-q}{N}}\right)^2 \tag{4.14}$$

which is smaller than the limit $b = \sigma^2(1+\sqrt{\frac{p}{N}})^2$. Here, we assume $l_{q+1} \simeq \hat{l}_{q+1}$ since $q \ll p$. According to Cauchy's interlace theorem (see Appendix 4.5.2), we have $\hat{l}_{q+1} \leq \tilde{l}_{q+1}$. It follows that $l_{q+1} \leq \tilde{l}_{q+1}$, which indicates that l_{q+1} is not larger than $\tilde{b}$ approximately, or more precisely, the bound of l_{q+1} is closer to $\tilde{b}$ than b. Therefore, in (4.11), the threshold b is somewhat higher than the true value of l_{q+1}, which leads to underestimation of q, namely, referring to l_q as the noise eigenvalue in some cases when the value of l_q is close to l_{q+1} and lower than the threshold b. If there exists a true threshold T for estimating q such that the probability of correct estimation is 1 and define $T_1 = b$ as the threshold of the test in (4.11), we find

$$T_1 \geq T \tag{4.15}$$

where the equality holds asymptotically.

Before we show that the test in (4.11) underestimates q mostly by one when the underestimation occurs, we recall the performance of the BIC, which is extensively studied in [35]–[42]. In [38], it is shown that for a wide range of SNRs and sample sizes, assuming the same power for the source signals, the probability of incorrect estimation of the

BIC is mainly dominated by the probability of underestimation by one. In [39], the threshold for detecting q is approximately expressed as

$$l_q > T_{\text{BIC}} \simeq \hat{\sigma}^2 \left\{ 1 + \sqrt{\frac{2(p-q+1)[2(p-q)+1]}{p-q}} \cdot \sqrt{\frac{\log N}{N}} \right\} \tag{4.16}$$

where $\hat{\sigma}^2 = \frac{1}{p-k} \sum_{i=k+1}^{p} l_i$ is the ML estimate of the noise variance σ^2. When the sample size N is large enough, it is readily found that

$$T_{\text{BIC}} > T_1. \tag{4.17}$$

Since the BIC underestimates q by one in most cases and the test in (4.11) has a much lower threshold than the BIC, the test in (4.11) underestimates q by one as well. However, the test has a much lower underestimation probability than the BIC.

Thus far, we validated that the test in (4.11) tends to underestimate q by one in the case of low SNRs and finite sample sizes, i.e., $q_0 = q-1$. To remedy this, we reduce the underestimation probability, or pull up the estimate q_0 from $q-1$ to q, as explained in the second-step test.

4.2.2 The second-step test

In first-step test, we derived an initial value q_0 for the number of sources, and $q_0 = q-1$ or q. In the second-step test, we conduct a test between the following two hypotheses

$$\mathsf{H}_0 : \ q_0 \text{ sources} \quad \text{vs.} \quad \mathsf{H}_1 : \ q_0 + 1 \text{ sources} \tag{4.18}$$

which can be implemented based on the likelihood ratio test

$$\text{LRT} = \frac{f(l_1, \cdots, l_p|\mathsf{H}_1)}{f(l_1, \cdots, l_p|\mathsf{H}_0)} \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtrless}} \eta \tag{4.19}$$

where $f(l_1, \cdots, l_p|\mathsf{H}_0)$ and $f(l_1, \cdots, l_p|\mathsf{H}_1)$ denote the joint pdfs of the sample eigenvalues under the hypotheses H_0 and H_1, respectively, and η is the test threshold. Intuitively, the test checks whether the underestimation occurred in the first-step test.

The joint pdf of the eigenvalues $f(l_1, \cdots, l_p)$ can be derived based on the complex Wishart distribution of the sample covariance matrix of the observations in (2.6). From the viewpoint of classical multivariate statistical theory, all the eigenvalues are mutually dependent, which results in a rather complicated formulation of the joint pdf of the eigenvalues [53]. The dependence decays as the sample size increases. Under the

assumption that all the signal population eigenvalues have multiplicity one, that is, they are distinct, a simplified version of the joint pdf is given for large sample sizes in [64, Eq. (16)]. It then follows that the signal eigenvalues are approximately mutually independent, and normally distributed but with a distinct mean and variance. Moreover, the signal eigenvalues are independent of the noise eigenvalues. This is exactly the assumption which is made frequently in analyzing the performance of the BIC [35]–[42].

Nevertheless, the pdfs used in [53] and [64], provided by classical multivariate statistical theory, which are based on large-sample asymptotic, are inappropriate when the sample size is small and even break down when the sample size is smaller than the array size. Here, we employ the distributions of the eigenvalues from random matrix theory, which are more robust against small sample sizes. Approximately, it is implied from Theorem 4.1.3 that the signal eigenvalues are independently distributed with different normal densities, and independent of the noise eigenvalues, which coincides with the result in [64] under the same assumption on the signal population eigenvalues. Thus, separately from the joint pdf of noise eigenvalues, the joint pdf of signal eigenvalues can be derived based on the normal densities in Theorem 4.1.3, namely as the product of the marginal pdfs of signal eigenvalues. However, it is incorrect to construct the joint pdf of noise eigenvalues as the product of the Marčenko-Pastur densities, since the noise eigenvalues are fully correlated. Consequently, it is unclear how to construct the joint pdf of the eigenvalues using random matrix theory.

In [94], it is shown that the likelihood ratio test in (4.19) depends asymptotically on the eigenvalue l_{q_0+1} only, which is the technical basis of conducting Roy's largest root test [16]. In other words, the eigenvalue l_{q_0+1} contains the richest statistical information and plays a dominant role in the test in (4.19). It is implied that the marginal pdf of the eigenvalue l_{q_0+1} can be an alternative for testing the two hypotheses in (4.18) when the joint pdf of all the eigenvalues is uncertain. Following this statement, the two hypotheses in (4.18) can be discriminated using the test

$$\frac{f(l_{q_0+1}|\mathsf{H}_1)}{f(l_{q_0+1}|\mathsf{H}_0)} \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtrless}} \eta \tag{4.20}$$

where $f(l_{q_0+1}|\mathsf{H}_0)$ and $f(l_{q_0+1}|\mathsf{H}_1)$ denote the pdfs of the eigenvalue l_{q_0+1} as the noise eigenvalue and the signal eigenvalue, respectively. Note that the likelihood test in (4.20) is a sub-optimal test for finite sample sizes since it neglects the dependence between the eigenvalues. In what follows, we derive $f(l_{q_0+1}|\mathsf{H}_0)$ and $f(l_{q_0+1}|\mathsf{H}_1)$ based on random matrix theory.

It follows from the Marčenko-Pastur law that

$$f(l_{q_0+1}|\mathsf{H}_0) = f_c(l_{q_0+1}) \tag{4.21}$$

where the pdf f_c is given in (4.2) in Theorem 4.1.1.

Theorem 4.1.3 establishes a normal density for $f(l_{q_0+1}|\nu_{q_0+1})$ which is the pdf of the signal eigenvalue l_{q_0+1} conditioned on the signal strength ν_{q_0+1}. However, it is hard to estimate ν_{q_0+1} accurately, since we have only one sample of l_{q_0+1}. In a Bayesian framework, a prior density can be chosen for $f(\nu_{q_0+1})$, e.g., Jeffrey's prior [119], which is a non-informative (objective) prior distribution on the parameter space. It is proportional to the square root of the determinant of Fisher's information. In our case, the distribution in Theorem 4.1.3 is complicated with respect to ν_{q_0+1}, so that Jeffrey's prior for ν_{q_0+1} has a cumbersome closed-form expression. If numerical methods are used to calculate Jeffrey's prior, the computational complexity may be too high. For simplicity, we assume ν_{q_0+1} to have a uniform distribution in the interval (α_1, α_2), as in [97], that is,

$$f(\nu_{q_0+1}) = \begin{cases} \frac{1}{\alpha_2-\alpha_1} & \text{if } \alpha_1 < \nu_{q_0+1} < \alpha_2 \\ 0 & \text{otherwise.} \end{cases} \tag{4.22}$$

Then, the pdf $f(l_{q_0+1}|\mathsf{H}_1)$ can be expressed as

$$\begin{aligned} f(l_{q_0+1}|\mathsf{H}_1) &= \int_{\alpha_1}^{\alpha_2} f(l_{q_0+1}|\nu_{q_0+1}) f(\nu_{q_0+1}) d\nu_{q_0+1} \\ &= \frac{1}{\alpha_2 - \alpha_1} \int_{\alpha_1}^{\alpha_2} f(l_{q_0+1}|\nu_{q_0+1}) d\nu_{q_0+1} \end{aligned} \tag{4.23}$$

where the bounds α_1 and α_2 are still missing. A possible choice for α_1 and α_2 is given by

$$\begin{aligned} \alpha_1 &= \sigma^2\sqrt{c} \\ \alpha_2 &= l_{q_0+1} \end{aligned} \tag{4.24}$$

based on Theorems 4.1.2 and 4.1.3. The interval in (4.24) is quite loose. For accurate modeling of the density $f(\nu_{q_0+1})$, we need to find a tighter interval for ν_{q_0+1}. Using Weyl's inequality in (4.12), we obtain

$$\nu_{q_0+1} + \rho_{\min(p,N)} \leq l_{q_0+1} \leq \nu_{q_0+1} + \rho_1. \tag{4.25}$$

According to (4.4) in Theorem 4.1.1, the largest and smallest non-zero noise eigenvalues converge to the upper and lower limits of the support of the Marčenko-Pastur density, respectively, that is,

$$\begin{aligned} \rho_1 &\to b = \sigma^2(1+\sqrt{c})^2 \\ \rho_{\min(p,N)} &\to a = \sigma^2(1-\sqrt{c})^2. \end{aligned} \tag{4.26}$$

With some straightforward manipulations of (4.25) as well as (4.26), and incorporating (4.24), an asymptotic interval is obtained as follows:

$$\begin{aligned} \alpha_1 &= \max(\sigma^2\sqrt{c}, l_{q_0+1} - b) \\ \alpha_2 &= l_{q_0+1} - a \end{aligned} \tag{4.27}$$

which concludes the derivation of the pdf $f(l_{q_0+1}|\mathsf{H}_1)$ in (4.23). Since the interval (α_1, α_2) relies on the data, the procedure of integrating the unknown parameters out is not truly in a Bayesian sense. It can be referred to as an empirical Bayesian method. Note that the integration with respect to ν_{q_0+1} is taken numerically, which is the main computational burden for the test in (4.20).

After obtaining the pdfs under the null hypothesis H_0 and alternative hypothesis H_1, the threshold η needs to be found in order to conduct the test in (4.20). In our case, it is not easy to derive an explicit expression for the threshold η based on the likelihood ratio distribution, due to the fact that the pdfs $f(l_{q_0+1}|\mathsf{H}_0)$ and $f(l_{q_0+1}|\mathsf{H}_1)$ are rather unusual, more precisely, the pdf $f(l_{q_0+1}|\mathsf{H}_0)$ is a bounded function and $f(l_{q_0+1}|\mathsf{H}_1)$ contains a numerical integration. In [120], it is shown that the generalized likelihood ratio test is equivalent to the information theoretic criteria when a specific test threshold is adopted. Let the test in (4.20) be equivalent to the BIC rule, as given in [120]. We then use a specific threshold, which is independent of the false alarm rate, namely,

$$\eta = \sqrt{N} \tag{4.28}$$

where N is the sample size.

Since all the ingredients of the likelihood test in (4.20) are well established, we summarize the proposed two-step test procedure for source enumeration in Table 4.1.

Table 4.1. The two-step test procedure based on random matrix theory.

Step 1: Conduct the first-step test in (4.11) to find an initial estimate q_0 for the number of sources.
Step 2: Conduct the second-step test in (4.20) to discriminate the two hypotheses in (4.18), that is, whether the number of sources is q_0 or $q_0 + 1$.

1. Construct the pdf $f(l_{q_0+1}|\mathsf{H}_0)$ using (4.21).
2. Construct the pdf $f(l_{q_0+1}|\mathsf{H}_1)$ using (4.23).
3. Derive the threshold η in (4.28).

When the sample size is not extremely large, two pdfs in (4.2) and (4.6) for the second-step test are more reliable than two limits in (4.4) and (4.5) for the first-step test, since the the convergence rate of the former two is faster than the one of the latter two. As a result, the second-step test is able to reduce the underestimation probability of the first-step test, however, at the cost of slightly increasing the overestimation probability, as shown in simulations. Let T_2 denote the threshold of the two-step test. Following (4.17), we find

$$T_{\text{BIC}} > T_1 \geq T_2 \tag{4.29}$$

where the equality holds asymptotically. T_2 is sometimes smaller than the true threshold T in (4.15) when the two-step test overestimates the number of sources.

It is worth mentioning that the second-step test in (4.20) is suboptimal as the joint pdfs of the eigenvalues are unavailable and only the marginal pdfs are used. Its implementation is not tedious and it is free of a false alarm rate, as the first-step test is. Also, the proposed two-step test procedure is deduced based some asymptotic results, but it yields remarkable performance in the case of finite sample sizes. It is consistent due to the fact that $T_1 = T_2$ in (4.29) is asymptotically true.

4.3 Simulation results

A uniform linear array with omni-directional sensors and inter-sensors spacing of half the wavelength is employed for simulations. The case of complex circular Gaussian signals contaminated by white Gaussian noise is considered. All simulation results were obtained based on 1,000 Monte Carlo runs. The numbers of samples, sensors and sources are denoted by N, p and q, respectively. The probability of correctly determining q is denoted by P_c. "DOA" is short for the direction of arrival of a source. "SNR" is short for the signal-to-noise ratio.

We denote the BIC method with the re-formulation in (2.21) by "BIC". Denote by "NAD" the method using (4.9) proposed in [90], by "KN" the method using (4.10) proposed in [94], by "HT_1" the first-step test in (4.11), and "HT_2" the two-step test in Table 4.1. The false alarm rate is set as $\alpha = 1\%$, which is much higher than 0.1% in [94][2]. In what follows, we evaluate these methods in different experimental settings.

Setting 1: Sample size (see Figs. 4.2 and 4.3). We focus on the case that N is varied around p. For the high dimension case (see Fig. 4.2), we have $p = 100$ sensors and

[2]Therein, $\alpha \ll 1$ is assumed due to the technical need. $\alpha = 0.1\%$ yields poor performance in our simulations. So it is a compromise to set $\alpha = 1\%$.

$q = 5$ source signals, with the DOAs of 20°, 23°, 25°, 27° and 30°, with respect to the broadside. The SNRs for all sources are set as -8 dB. The number of snapshots N is varied from 10 to 100.

Although random matrix theory is derived for the high-dimension case, it is interesting to see its performance in the low-dimension case. For the low-dimension case (see Fig. 4.3), we have $p = 15$ sensors and $q = 2$ source signals. The sources are located at 20° and 23°. The SNRs for both sources are set as 1 dB. The number of snapshots N is varied from 6 to 20.

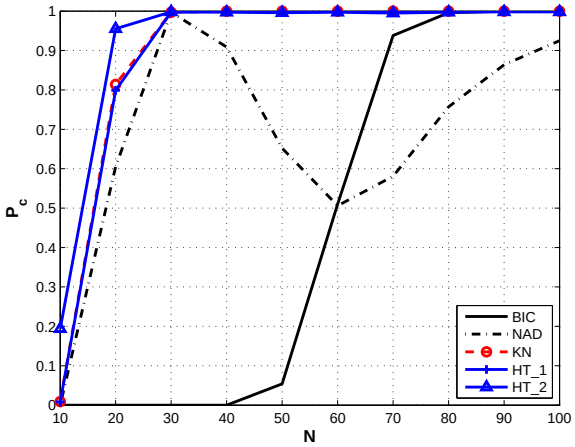

Figure 4.2. Probability of correct determination P_c versus number of samples N in the high dimensional array case ($p = 100, q = 5$).

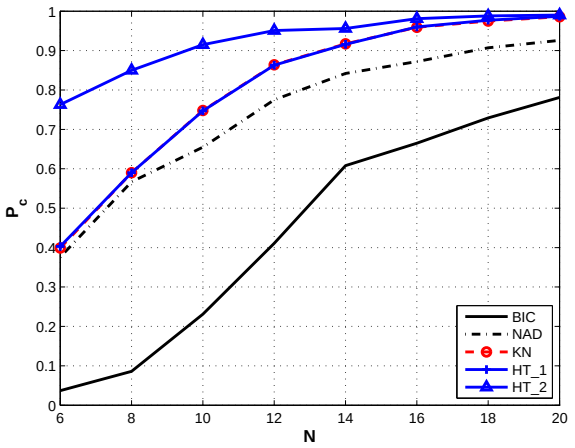

Figure 4.3. Probability of correct determination P_c versus number of samples N in the low dimensional array case ($p = 15, q = 2$).

It is seen that the proposed method "HT_2" can detect the number of source signals

more accurately than the other methods for both the high-dimension and the low-dimension cases. The BIC gives acceptable results only when N is very close to p, whereas the method "HT_2" gives excellent results even when N is much smaller than p. The methods "HT_1" and "KN" perform similarly.

Setting 2: SNR (see Figs. 4.4 and 4.5). We focus on the case that N is close to p. For the high-dimension case (see Fig. 4.4), we have $p = 50$ sensors and $q = 5$ source signals, with the DOAs of 20°, 23°, 25°, 27° and 30°. They have the same SNR, which is varied between $[-14, 0]$ dB. $N = 40$ snapshots are used.

In order to clarify the incorrect estimation probabilities of the methods "HT_1" and "HT_2", we tabulate the results in Fig. 4.4 into Table 4.2. Simply put, the method "HT_2" pulls up the estimated number of signals with respect to the method "HT_1". However, the method "HT_2" yields a slight probability to overestimate q, when the SNR gets higher.

For the low-dimension case (see Fig. 4.5), we have $p = 15$ sensors and $q = 2$ source signals. The sources are located at 20° and 22°. They have the same SNR, which is varied between $[-6, 8]$ dB. Only $N = 20$ snapshots are available.

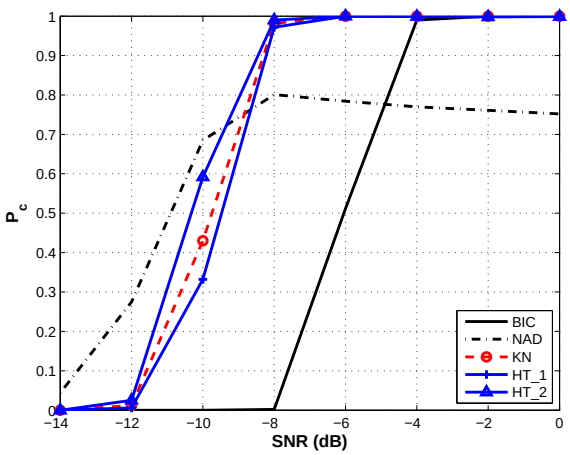

Figure 4.4. Probability of correct determination P_c versus SNR in the high dimensional array case ($p = 50, q = 5$).

Like Setting 1, remarkable performance improvement of the proposed method "HT_2" can be observed at a relatively small sample size and low SNR, compared to the BIC, for both the high-dimension and the low-dimension cases. For instance, in Figs. 4.4 and 4.5, the method "HT_2" has an improvement of nearly 4 dB over the BIC. It is worth noting that the performance improvement of the method "HT_2" is more

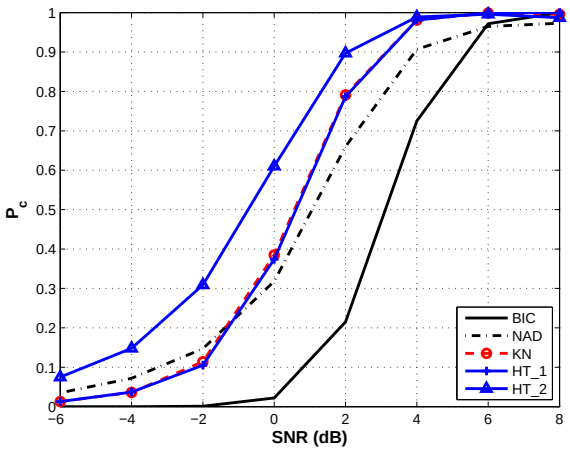

Figure 4.5. Probability of correct determination P_c versus SNR in the low dimensional array case ($p = 15, q = 2$).

significant in the low-dimension case than in the high-dimension case, as compared to the methods "KN" and "HT_1".

In our simulations, the performance of the method "NAD" is unstable in the case of $N < p$ (see Figs. 4.2, 4.3 and 4.4). This is due to the fact that all the eigenvalues, including the $p-N$ zero eigenvalues, are involved for constructing the method "NAD". If only the N non-zero eigenvalues are used, as most existing methods did, its performance curve becomes smooth and behaves like that in the case of $N \geq p$ (see Figs. 4.5, 4.6 and 4.7). Also, this AIC-like criterion is not consistent with respect to the SNR and N since it tends to overestimate the number of sources.

Setting 3: Angular resolution (see Fig. 4.6). We have $p = 10$ sensors and $q = 2$ source signals. One signal has the DOA of 20°, The other has a varying DOA in the range of $[22°, 36°]$. The SNRs for both sources are set as -3 dB. Only $N = 20$ snapshots are used.

Setting 4: Correlated signals (see Fig. 4.7). We have $p = 20$ sensors and $q = 2$ source signals, with the DOAs of 20° and 25°. The SNRs for both sources are set as -3 dB. Only $N = 30$ snapshots are used. The correlation coefficient between the two signals is varied within the interval $[0.6, 0.95]$.

In Fig. 4.6 and Fig. 4.7, it is apparent that the proposed method "HT_2" is superior in terms of two-signal-separation over the BIC, when their DOAs are very close to each other or they are highly correlated. The method "HT_2" is still the most competent one in such difficult situations.

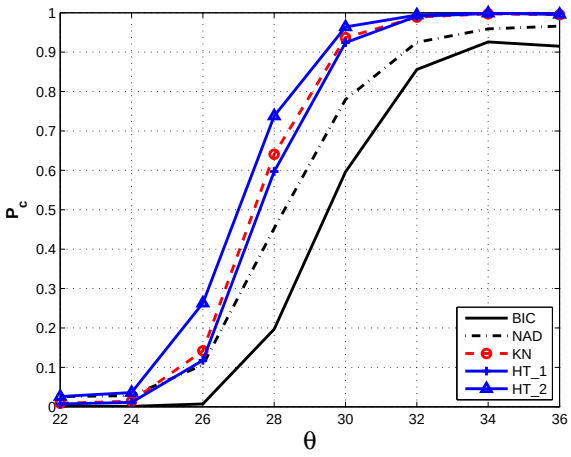

Figure 4.6. Probability of correct determination P_c versus DOA of the second signal θ $(p = 10, q = 2)$.

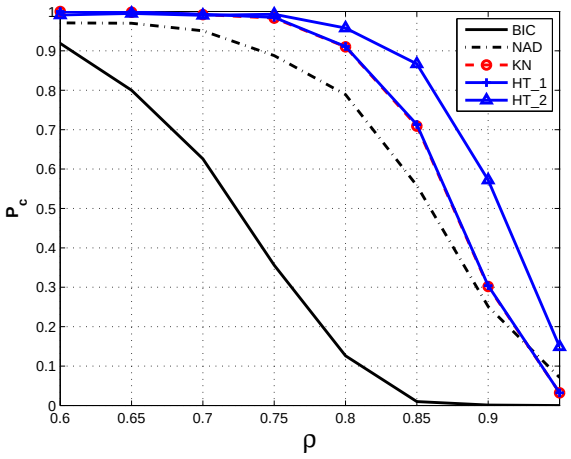

Figure 4.7. Probability of correct determination P_c versus correlation coefficient of two signals ρ. $(p = 20, q = 2)$.

Table 4.2. Probability of determining q versus SNR for the high-dimension case in Setting 2.

SNR(dB)		-14	-12	-10	-8	-6	-4	-2	0
BIC	$p(q<5)$	1	1	1	0.998	0.488	0.010	0	0
	$\mathbf{p(q=5)}$	**0**	**0**	**0**	**0.002**	**0.512**	**0.990**	**1**	**1**
	$p(q>5)$	0	0	0	0	0	0	0	0
KN	$p(q<5)$	1	0.988	0.570	0.018	0	0	0	0
	$\mathbf{p(q=5)}$	**0**	**0.012**	**0.430**	**0.982**	**1**	**1**	**1**	**1**
	$p(q>5)$	0	0	0	0	0	0	0	0
HT_1	$p(q<5)$	1	0.994	0.668	0.028	0	0	0	0
	$\mathbf{p(q=5)}$	**0**	**0.006**	**0.332**	**0.972**	**1**	**1**	**1**	**1**
	$p(q>5)$	0	0	0	0	0	0	0	0
HT_2	$p(q<5)$	1	0.975	0.408	0.010	0	0	0	0
	$\mathbf{p(q=5)}$	**0**	**0.025**	**0.592**	**0.990**	**1**	**0.999**	**0.999**	**1**
	$p(q>5)$	0	0	0	0	0	0.001	0.001	0

In the preceding simulations, the methods "KN", "HT_1" and "HT_2" used the prior knowledge of the noise variance σ^2, whereas the BIC and the method "NAD" could not. When σ^2 is unknown, we propose to use the estimator $\hat{\sigma}^2$ presented in [94] since it has much lower bias than the one in (2.16). The performance can be seen in Fig. 4.8 and Fig. 4.9, which are derived based on the same experimental settings for Fig. 4.3 and Fig. 4.4, respectively. Comparing Figs. 4.8 and 4.9 with Figs. 4.3 and 4.4, the performance is reduced as the estimator $\hat{\sigma}^2$ is substituted for σ^2. Generally, our methods "HT_1" and "HT_2" retain good performance by using the estimator $\hat{\sigma}^2$ in [94].

As we know, the BIC is derived based on the joint multivariate complex normal distribution of the observed data. It only takes the distribution of observations into account, while it ignores the distributions of the parameters, e.g., eigenvalues or eigenvectors. It yields poor performance when the environments are severe, e.g., low SNR or small N. Contrary to the BIC, the methods "KN", "HT_1" and "HT_2" exploit the information from the distributions of the eigenvalues rather than the distribution of the observations. Their performance is largely improved according to the simulations, which confirms that the distributions of the eigenvalues are more informative than the distribution of the observations.

The method "HT_1" is competitive with the method "KN", which indicates that the thresholds of both methods for the tests are very close to each other, at least in our case when the array size p is not very large. Note that the implementation of the method "HT_1" is much simpler. These two methods are beaten by the method "HT_2", which indicates that the threshold is refined by considering the difference between

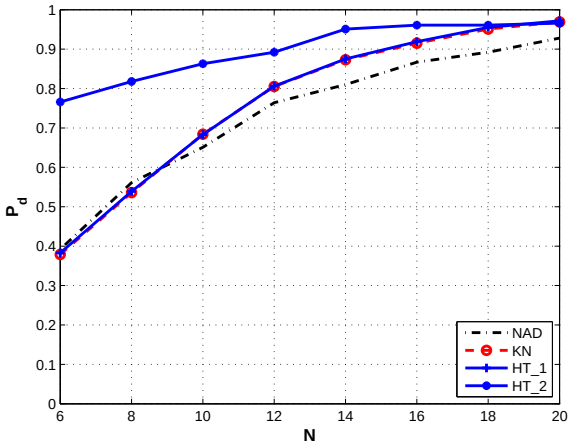

Figure 4.8. Probability of correct determination P_c versus number of samples N in the low dimensional array case with $\hat{\sigma}^2$ $(p = 15, q = 2)$.

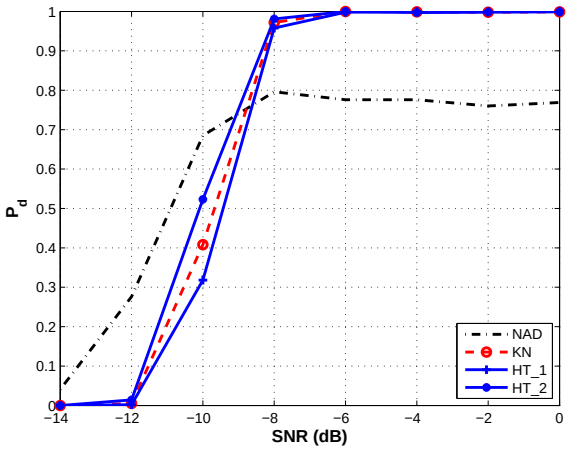

Figure 4.9. Probability of correct determination P_c versus SNR in the high dimensional array case with $\hat{\sigma}^2$ $(p = 50, q = 5)$.

the distributions of signal eigenvalues and noise eigenvalues. Simulations show that the method "HT_2" has a smaller underestimation probability than the method "HT_1" in the regime of relatively low SNR and small N, whereas the former overestimates slightly the number of sources in the regime of relatively high SNR and large N, which is in good agreement with the analysis in Section 4.2.

Although random matrix theory is based on the high-dimension and large-sample asymptotic regime, it yields good results in our simulations in the case of a low array dimension and a small sample size. Such results are very important in real applications, since the array dimension and the sample size are not very large in most cases. Moreover, of practical interest is to see that the proposed method is capable of resolving closely-spaced signals, and extremely robust against correlation between signals.

4.4 Conclusions

A two-step test procedure has been proposed for the problem of source enumeration by employing random matrix theory. As an asymptotic result, random matrix theory preserves high accuracy in the case of low array dimensions and small sample sizes, which renders our procedure applicable in a wide range in terms of array size as well as sample size, including the extreme case $p/N > 1$. It has been validated by simulations that the distributions of the eigenvalues from random matrix theory provide more accurate information than the distribution of the observations in difficult situations, e.g., very small sample size, very low SNR, close spacing and high correlation between signals. By exploiting the information from the distributions of the eigenvalues, we have constructed two consecutive tests in order to discriminate the signal eigenvalues from the noise eigenvalues. The first-step test is a thresholding approach whereas the second-step test is a simplified likelihood ratio test. We have validated theoretically that the first-step test tends to underestimate the number of sources by one for a low SNR and a small sample size, which necessitates the second-step test. Simulations have shown that the second-step test is competent to remedy the limitations of the first-step test by significantly reducing the underestimation probability. Although both tests have simple implementation, their combination outperforms some popular approaches.

It is worth noting that the first-step test alone yields notable performance so that it can be used as a pre-processing step for other approaches. We have derived the upper limit of the candidates for the number of sources and the tight interval for the signal population eigenvalue, which are valuable for other approaches as well. In addition, the attempt in this chapter shows that cautious utilization of random matrix theory

provides great potential to create novel and efficient algorithms for source enumeration, or even other problems in array processing.

4.5 Appendix

4.5.1 Weyl's inequality

In linear algebra, Weyl's inequality is a theorem about the changes to eigenvalues of a Hermitian matrix that is perturbed, or there is an uncertainty about the entries. Proofs are given in [121].

Theorem 4.5.1 *Let* $\mathbf{A}$, $\mathbf{B}$ *be the* $p \times p$ *Hermitian matrices, and let* $\mathbf{A}$ *be the exact matrix and* $\mathbf{B}$ *be a perturbation matrix that represents the uncertainty. Let the eigenvalues* $\lambda_i(\mathbf{A})$, $\lambda_i(\mathbf{B})$ *and* $\lambda_i(\mathbf{A}+\mathbf{B})$ *be arranged in decreasing order* $(\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p)$. *Then, for each* $i = 1, \ldots, p$, *we have*

$$\lambda_k(\mathbf{A}) + \lambda_p(\mathbf{B}) \leq \lambda_k(\mathbf{A}+\mathbf{B}) \leq \lambda_k(\mathbf{A}) + \lambda_1(\mathbf{B}). \tag{4.30}$$

We use the model in (2.5), where $\mathbf{A}$ is the covariance matrix of the product of the steering matrix and the signal matrix, which is perturbed by $\mathbf{B}$, the covariance matrix of the noise matrix.

4.5.2 Cauchy's interlace theorem

Cauchy's interlace theorem shows how the eigenvalues of a Hermitian matrix relate to those of a principal sub-matrix. Proofs can be found in [122].

Theorem 4.5.2 *Let* $\mathbf{A}$ *be a given Hermitian matrix of size* $p \times p$. *Partition* $\mathbf{A}$ *as*

$$\begin{bmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{C}^H & \mathbf{D} \end{bmatrix} \tag{4.31}$$

with $\mathbf{B}$ *of size* $(p-k) \times (p-k)$. *Let the eigenvalues* $\lambda_i(\mathbf{A})$*and* $\lambda_i(\mathbf{B})$ *be arranged in increasing order* $(\lambda_1 \leq \lambda_2 \leq \cdots)$. *Then, for each* $i = 1, \ldots, p-k$, *we have*

$$\lambda_i(\mathbf{A}) \leq \lambda_i(\mathbf{B}) \leq \lambda_{i+k}(\mathbf{A}). \tag{4.32}$$

Chapter 5

Performance analysis of Bayesian information criterion (BIC)

Under certain regularity conditions, the BIC is consistent, which ensures that it estimates the true number of sources with probability one as the number of snapshots increases to infinity. However, only a limited number of observations is available in practice. Thus, the estimation performance evaluated for a finite number of observations is of greater practical interest than the asymptotic behavior.

The non-parametric eigendecomposition based BIC [31] is derived using the eigenvalues which are calculated from the sample covariance matrix of the observed data, without knowledge of the array manifold and the desired signal waveforms. Its statistical performance for source enumeration has been extensively analyzed in [35–44]. In [35, 37, 38], the authors used the distributions of the eigenvalues, which were derived based on classical multivariate statistical theory [123]. When the sample size is not extremely large, these distributions are significantly biased, which leads to inaccurate analysis results. The bias of the distributions can be corrected [15]. Based on this correction, the methods in [42][43] are able to predict the theoretical performance of the BIC in close accordance to empirical results from simulations. In [41], the authors generalized the method in [38] based on the theory of misspecified models, which can deal with more complex formulations of the BIC, that involve e.g. the knowledge of the array geometry. In addition, it is shown in [42][41] that the performance of the BIC for source enumeration is approximately the same for models consisting of deterministic and stochastic signals.

In this chapter, we analyze the performance of the BIC from a different viewpoint: we propose to use distributions derived from random matrix theory [91]. They are found to be more accurate in modeling the eigenvalues than those from classical multivariate statistical theory. The remainder of the chapter is organized as follows. The problem of analyzing the performance of the BIC is formulated in Section 5.1. Section 5.2 summarizes some important statistical properties of the sample eigenvalues. Two new procedures are proposed for performance analysis of the conventional BIC in Sections 5.3. Numerical simulations that illustrate the performance of the proposed procedures are presented in Section 5.4. Finally, conclusions are drawn in Section 5.5.

5.1 Problem formulation

As shown in Section 2.2.2, the conventional BIC is given by

$$\begin{aligned}\hat{q} &= \arg\min_k \{-L(k) + y(k)\} \\ &= \arg\min_k \left\{-N(p-k)\log\frac{G_k}{A_k} + \frac{1}{2}k(2p-k)\log N\right\}, \quad k = 0, \ldots, p-1.\end{aligned} \tag{5.1}$$

where $L(k)$ is the log-likelihood term, $y(k)$ is the penalty term, $G_k = \prod_{i=k+1}^{p} l_i^{1/(p-k)}$ is the geometric mean of the $p-k$ smallest sample eigenvalues, and $A_k = \frac{1}{p-k}\sum_{i=k+1}^{p} l_i$ is the corresponding arithmetic mean. In what follows, we will introduce the fundamentals for evaluating the performance of the BIC, which are established in [38].

Denote by H_q the hypothesis that the true number of sources is q. Then, the probability of estimating the number of sources incorrectly is defined as

$$P_e = P(\hat{q} \neq q|\mathsf{H}_q) = P_m + P_f \tag{5.2}$$

with the probability of underestimation

$$P_m = P(\hat{q} < q|\mathsf{H}_q) \tag{5.3}$$

and the probability of overestimation

$$P_f = P(\hat{q} > q|\mathsf{H}_q). \tag{5.4}$$

For a wide range of signal-to-noise ratios (SNRs) and sample sizes, assuming that the source signals have the same power, Zhang *et al.* showed in [38] that for the BIC, the probability of incorrect estimation is mainly dominated by the probability of underestimation, that is,

$$P_e \simeq P_m \quad \text{and} \quad P_f \simeq 0, \tag{5.5}$$

and further, P_m can be approximated by

$$P_m \simeq P(\hat{q} = q-1|\mathsf{H}_q). \tag{5.6}$$

As in [38], if we define

$$\begin{aligned}\Delta L_q &= \frac{1}{N}\left[L(q) - L(q-1)\right] \\ \Delta y_q &= \frac{1}{N}\left[y(q) - y(q-1)\right],\end{aligned} \tag{5.7}$$

it follows that

$$P_e \simeq P_m \simeq P(\hat{q} = q-1|\mathsf{H}_q) = P(\Delta L_q < \Delta y_q) \tag{5.8}$$

which is the underlying principle for the performance analysis of the BIC. From (5.1), we can derive

$$\begin{aligned} \Delta L_q &= \log\left\{\frac{A_q}{l_q}\left[1+\frac{1}{p-q+1}\left(\frac{l_q}{A_q}-1\right)\right]^{p-q+1}\right\} \\ \Delta y_q &= \frac{(2p-2q+1)\log N}{2N} \end{aligned} \tag{5.9}$$

where l_q is the smallest signal sample eigenvalue, and A_q is the arithmetic mean of the noise sample eigenvalues, namely, $A_q = \frac{1}{p-q}\sum_{i=q+1}^{p} l_i$, which is also the maximum likelihood (ML) estimate of the noise variance σ^2.

The prerequisite for calculating P_e or P_m in (5.8) is to know the distribution of the random variable ΔL_q. Since ΔL_q is dependent of two random variables l_q and A_q (see (5.9)), their statistical properties are of utmost importance. Thus, in the following section, the statistical properties of l_q and A_q are investigated in detail.

5.2 Statistical properties of sample eigenvalues

In this section, we introduce the basis of performance analysis of the BIC, that is, the statistical structures of the sample eigenvalues, which are related to the distributions of the smallest signal eigenvalue l_q and the arithmetic mean of the noise eigenvalues A_q. There mainly exist three different statistical results which are based on (i) classical multivariate statistical theory [123], (ii) modern random matrix theory [91], and (iii) Lawley's theory [15].

Classical multivariate statistical theory characterizes the sample eigenvalues based on large-sample asymptotics, i.e., p fixed, $N \to \infty$, whereas modern random matrix theory assumes the high-dimension and large-sample asymptotic regime, i.e., $p, N \to \infty$ with $p/N \to c \in (0, \infty)$. However, it is justified by simulations that random matrix theory works also very well in the low-dimension case. Moreover, Lawley revised the result of classical multivariate statistical theory by considering the interaction among the eigenvalues. The application of classical multivariate statistical theory for the problem of source enumeration can be found in [54,55], while the applications of random matrix theory and Lawley's theory can be found in [90, 94, 95, 99] and [21], respectively.

All the three theories mentioned above state that the two statistics l_q and A_q follow Gaussian distributions with distinct means and variances. In order to see the difference, we table all the involved values into Table 5.1. Herein, "MST" denotes classical

Table 5.1. Comparison of three results for the expectation and variance of l_q and A_q.

	MST	RMT	LAW
μ_{l_q}	λ_q	$\lambda_q \left[1 + \frac{(p-q)\sigma^2}{N(\lambda_q-\sigma^2)}\right]$	$\lambda_q \left[1 + \frac{(p-q)\sigma^2}{N(\lambda_q-\sigma^2)} - \frac{1}{N}\sum_{j=1}^{q-1}\frac{\lambda_j}{\lambda_j-\lambda_q}\right]$
$\sigma^2_{l_q}$	$\frac{\lambda_q^2}{N}$	$\frac{\lambda_q^2}{N}\left[1 - \frac{(p-q)\sigma^4}{N(\lambda_q-\sigma^2)^2}\right]$	/
μ_{A_q}	σ^2	σ^2	$\sigma^2\left[1 - \frac{1}{N}\sum_{j=1}^{q}\frac{\lambda_j}{\lambda_j-\sigma^2}\right]$
$\sigma^2_{A_q}$	$\frac{\sigma^4}{N(p-q)}$	$\frac{\sigma^4}{N(p-q)}$	/

multivariate statistical theory, "RMT" random matrix theory and "LAW" Lawley's theory.

Comparing the expressions of μ_{A_q} and $\sigma^2_{A_q}$ of "MST" and "RMT", it follows that the two preceding theories have the same distribution of A_q, although it is derived in different contexts. Note that μ_{l_q} and $\sigma^2_{l_q}$ of "RMT" have the additional second term compared to those in "MST". When $N \gg p-q$, the second terms of μ_{l_q} and $\sigma^2_{l_q}$ of "RMT" vanish and the distribution of "RMT" decays to that of "MST". Due to the second terms, "RMT" is still accurate even when the number of samples N is not very large compared to $p-q$, whereas "MST" provides accurate results only when $N \gg p-q$. Hence, "RMT" is less sensitive to the number of samples than "MST", which is also validated in numerous works (see, e.g., [90, 94, 95, 99]).

It is worth noting that an implicit assumption for the result of "RMT" is that the number of sources is much smaller than the number of sensors, i.e., $q \ll p$, which fits the so-called spiked model. In addition, the signal population eigenvalues should be larger than $\sigma^2(1+\sqrt{c})$. The proofs for the result of "RMT" can be found in [112]–[125].

For μ_{l_q} of "LAW", the second term in brackets is the interaction between the signal population eigenvalue λ_q and the noise population eigenvalues $\lambda_{q+1}, \cdots, \lambda_p$ which are all equal to σ^2, and the third term is the interaction between λ_q and the other signal population eigenvalues $\lambda_1 \cdots, \lambda_{q-1}$. However, μ_{A_q} of "LAW" does not contain the interaction between the noise eigenvalues, since all the population eigenvalues $\lambda_{q+1}, \cdots, \lambda_p$ are equal. It is known that the sample eigenvalues are highly affected by a bias, which is caused by the interaction among the eigenvalues. Thus, the additional terms in "LAW", compared to "MST", are non-negligible for practical values of p and N.

Comparing the μ_{l_q}'s in Table 5.1, we can find that "LAW" takes into account the disturbance from the noise and the other signals, whereas "MST" does neither. "RMT" considers only the disturbance from the noise. Comparing the μ_{A_q}'s, only "LAW" considers the disturbance. Therefore, we propose to use random matrix theory for the performance analysis of the BIC if we have only access to the smallest signal population eigenvalue λ_q. If all the signal population eigenvalues $\lambda_1, \cdots, \lambda_q$ are known, we incorporate Lawley's theory into random matrix theory, more precisely, use the expectations from Lawley's theory and the variances from random matrix theory. This is referred to as "unbiased " random matrix theory.

5.3 Procedures for performance analysis

Following the introduction of the BIC in Section 5.1 and the previous work in [42] [38][43], we review a previous procedure and present two new procedures for analyzing the performance of the BIC, namely, computing the probability of underestimation P_m. It is assumed that the sample size and the SNR are both large enough and the source signals are located widely, to avoid a subspace swap, e.g., a crossover between the smallest signal sample eigenvalue and the largest noise sample eigenvalue [54][55].

5.3.1 ΔL_q-distribution-based method

This method is proposed in [38][43]. Following (5.8) directly, we convert the problem to the statistics related to the eigenvalues ΔL_q. Thus, the distribution of ΔL_q is used to compute the probability P_m. In what follows, the main work is to derive the distribution for ΔL_q.

It follows from (5.9) that ΔL_q is a function of l_q and A_q. Let us define $\Delta L_q = g(l_q, A_q)$, and its gradient vector

$$\nabla g(l_q, A_q) = \left[\frac{\partial g(l_q, A_q)}{\partial l_q}, \frac{\partial g(l_q, A_q)}{\partial A_q} \right]^T. \tag{5.10}$$

As introduced in Section 5.2, l_q and A_q are both independently Gaussian distributed. By employing the delta method [126], we can obtain that ΔL_q is asymptotically Gaussian distributed, that is,

$$\Delta L_q \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mu_{\Delta L_q}, \sigma^2_{\Delta L_q}\right) \tag{5.11}$$

with

$$\begin{aligned}\mu_{\Delta L_q} &= g(\mu_{l_q}, \mu_{A_q}) \\ \sigma^2_{\Delta L_q} &= \nabla g(\mu_{l_q}, \mu_{A_q})^T \begin{bmatrix} \sigma^2_{l_q} & 0 \\ 0 & \sigma^2_{A_q} \end{bmatrix} \nabla g(\mu_{l_q}, \mu_{A_q}).\end{aligned} \tag{5.12}$$

After some straightforward manipulations upon (5.9) and (5.10), we have

$$\nabla g(\mu_{l_q}, \mu_{A_q}) = \left[\frac{1}{\mu_{l_q}}, \frac{1}{\mu_{A_q}}\right]^T \cdot \frac{(p-q)(\mu_{l_q} - \mu_{A_q})}{\mu_{l_q} + (p-q)\mu_{A_q}}, \tag{5.13}$$

and further, (5.12) results in

$$\begin{aligned}\mu_{\Delta L_q} &= \log\left\{\frac{\mu_{A_q}}{\mu_{l_q}}\left[1 + \frac{1}{p-q+1}(\frac{\mu_{l_q}}{\mu_{A_q}} - 1)\right]^{p-q+1}\right\} \\ \sigma^2_{\Delta L_q} &= \left(\frac{\sigma^2_{l_q}}{\mu^2_{l_q}} + \frac{\sigma^2_{A_q}}{\mu^2_{A_q}}\right)\left[\frac{(p-q)(\mu_{l_q} - \mu_{A_q})}{\mu_{l_q} + (p-q)\mu_{A_q}}\right]^2.\end{aligned} \tag{5.14}$$

Based on the Gaussian probability density function (pdf) $f(\Delta L_q)$ with the mean and variance in (5.14), the probability of underestimation P_m can be easily calculated by

$$P_m \simeq P(\Delta L_q < \Delta y_q) = 1 - Q\left(\frac{\Delta y_q - \mu_{\Delta L_q}}{\sigma_{\Delta L_q}}\right) \tag{5.15}$$

where $Q(x) = \int_x^\infty \exp(-t^2/2)/\sqrt{2\pi}dt$, $\mu_{\Delta L_q}$ and $\sigma_{\Delta L_q}$ are given in (5.14). Simply put, calculation of P_m in (5.15) is nothing but calculating the cumulative distribution function (cdf) of ΔL_q, i.e., $F(\Delta L_q \leq \Delta y_q)$.

5.3.2 ρ-distribution-based method

It is seen from (5.9) that ΔL_q is a complicated function of l_q and A_q. In order to reduce the modeling error using the delta method, we prefer to use some simpler statistic of l_q and A_q. It can be readily shown that ΔL_q is a monotonically increasing function of the ratio $\frac{l_q}{A_q}$, since the first derivative of ΔL_q with respect to $\frac{l_q}{A_q}$ is larger than zero for $p > q$ and $\frac{l_q}{A_q} > 1$. So we use the ratio

$$\rho = \frac{l_q}{A_q} \tag{5.16}$$

as the statistic for the performance analysis. By using the delta method again, we find that ρ has an asymptotic Gaussian distribution,

$$\rho \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mu_\rho, \sigma^2_\rho\right) \tag{5.17}$$

with

$$\begin{aligned}\mu_\rho &= \frac{\mu_{l_q}}{\mu_{A_q}} \\ \sigma_\rho^2 &= \frac{\sigma_{l_q}^2}{\mu_{A_q}^2} + \frac{\mu_{l_q}^2 \sigma_{A_q}^2}{\mu_{A_q}^4}.\end{aligned} \tag{5.18}$$

Then, if we can find a value ρ_0 such that $\Delta L_q(\rho_0) = \Delta y_q$, the probability P_m can be computed by

$$P_m \simeq P(\rho < \rho_0) = 1 - Q\left(\frac{\rho_0 - \mu_\rho}{\sigma_\rho}\right) \tag{5.19}$$

where μ_ρ and σ_ρ are given in (5.18). Since a closed-form expression for ρ_0 does not exist for a finite sample size N, we use the Newton-Raphson method to find ρ_o numerically from the equation $\Delta L_q(\rho_0) = \Delta y_q$, which has the following detailed form

$$\log\left[\frac{1}{\rho_0}\left(1 + \frac{\rho_0 - 1}{p - q + 1}\right)^{p-q+1}\right] = \frac{(2p - 2q + 1)\log N}{2N}. \tag{5.20}$$

The initial value for ρ_0 is given by the approximated value in [40]

$$\rho_{0*} = 1 + \varepsilon\left\{\kappa\left[1 + \sqrt{\frac{2(\kappa - 1)}{\varepsilon\kappa}}\right] - 1\right\}$$

with $\varepsilon = p - q + 1$ and $\kappa = N^{\frac{2\varepsilon - 1}{2N(\varepsilon - 1)}}$. This method is very similar to the method proposed in [42]. However, the latter is not a unified method for all q.

It is worth noting that there are some more accurate methods, apart from the delta method, which can be used to derive the ratio distribution for two Gaussian random variables. For instance, in [127], the Mellin transform is used to derive the ratio distribution, which is a polynomial with infinite terms. In [128], the authors derived the distribution for the ratio of two correlated Gaussian random variables. In our case, the denominator variable A_q has a very small variance, which leads to good performance of the delta method.

Note that ρ is a much simpler statistic related to the eigenvalues than ΔL_q and its distribution of ρ is more precise than that of ΔL_q. Nevertheless, the ρ-distribution-based method requires a numerical method to find ρ_0 in (5.20). Most importantly, it can be observed from the ρ-distribution-based method that the BIC is simply a thresholding approach, that is, comparing the ratio ρ, which tells the difference between the smallest signal eigenvalue and the noise eigenvalues, with a threshold ρ_0. The threshold ρ_0 is determined by the penalty term of the BIC. If $\rho > \rho_0$, the BIC yields an accurate estimation result, namely, $\hat{q} = q$, otherwise the BIC underestimates q by one, namely, $\hat{q} = q - 1$. This observation, which is confirmed by simulations, motivates the new criterion in Chapter 6. It is worth mentioning that this property of the BIC may not be valid in other problems, such as model order selection for time series data.

5.3.3 Bootstrap-based method

As we know, Gaussian distributions of l_q and A_q render the preceding two methods to be implemented using the delta method. If l_q and A_q have non-Gaussian distributions, it is not easy to derive a simple form of distribution analytically for either ΔL_q or ρ. An alternative is to use resampling techniques, such as the bootstrap [25][77]. The application of the bootstrap for source enumeration can be found in [21]. The bootstrap is used to estimate the distribution of the test statistics of the eigenvalues, and the number of signals are detected by means of hypothesis testing. A good performance is obtained with the bootstrap for small sample sizes.

Here, the parametric bootstrap technique is employed to estimate the probability P_m, due to the fact that the distributions of l_q and A_q are assumed to be known a priori. This method is analytically simple at the expense of high computational complexity. The parametric bootstrap procedure for computing the probability P_m is given in Table 5.2.

Table 5.2. The parametric bootstrap procedure for computing the probability P_m.

Step 1. *Resampling.* Draw a random sample l_q^* and A_q^*, from the distributions F_{l_q} and F_{A_q}, respectively. For example, the distributions from random matrix theory can be used.
Step 2. *Calculation of the bootstrap statistic.* Calculate the ratio $\rho^* = \frac{l_q^*}{A_{q+1}^*}$, and the bootstrap statistic ΔL_q^* using (5.9), but with the bootstrap sample ρ^* replacing the sample ρ.
Step 3. *Repetition.* Repeat Steps 1 and 2 many times to obtain a total of B bootstrap estimates $\Delta L_{q1}^*, \Delta L_{q2}^*, \ldots, \Delta L_{qB}^*$.
Step 4. *Calculation of the probability P_m.* Calculate $P_m = \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}_{\{\Delta L_{qb}^* < \Delta y_q\}}$, where the indicator function $\mathbf{1}_{[\Delta L_{qb}^* < \Delta y_q]}$ is one if $\Delta L_{qb}^* < \Delta y_q$ is fulfilled, otherwise zero.

As stated in Section 5.2, we use the distributions based on random matrix theory instead of the distributions based on classical multivariate statistical theory to compute the probability P_m for all three methods, since random matrix theory characterizes the eigenvalues more precisely than classical multivariate statistical theory. Furthermore, the distributions based on "unbiased " random matrix theory are used if we have full access to all the signal population eigenvalues $\lambda_1, \cdots, \lambda_q$. The proposed methods here can also be used for performance analysis of other information theoretic criteria.

5.4 Simulation results

A uniform linear array with omni-directional sensors and inter-sensors spacing of half the wavelength is employed for simulations. The case of uncorrelated complex circular Gaussian source signals contaminated by white Gaussian noise is considered. For notations, the numbers of samples, sensors and sources are denoted by N, p and q, respectively. The DOAs are denoted by the vector $\boldsymbol{\theta}$. The probabilities of underestimation, overestimation and correct estimation of q are denoted by P_m, P_f and P_c, respectively.

The method presented in [43] by Nadler is used as the benchmark method, which corrects the bias of the result of "MST" based on Lawley's theory and a more sophisticated method than the delta method. Since the 3 methods in Section 5.3 perform similarly, only the ΔL_q-distribution-based method is plotted in the following figures. Denote by "RMT" the ΔL_q-distribution-based method using the result of random matrix theory, by "unbiased RMT" the ΔL_q-distribution-based method using the results of random matrix theory and Lawley's theory. Nadler's method and the method "RMT" employ only the smallest signal population eigenvalue λ_q whereas the method "unbiased RMT" uses all the signal population eigenvalues $\lambda_1, \cdots, \lambda_q$.

The theoretical probabilities of underestimation predicted by the proposed methods or the reference method are compared with the empirical probabilities of underestimation obtained by simulations. All the BIC simulation results are obtained based on 10,000 Monte Carlo trials. The true values for the signal population eigenvalues $\lambda_1, \cdots, \lambda_q$ are calculated from the covariance matrix based on 10^5 snapshots. The noise variance σ^2 is set always as one. Six different experimental settings are listed in the following:

- Varying the sample size N:
  - Setting 1 (see Fig. 5.1): $N \in [1000, 2000]$, $p = 25$, $q = 4$, DOAs: $\boldsymbol{\theta} = \{10^\circ, 13^\circ, 17^\circ, 20^\circ\}$, SNR $= -10$ dB.
  - Setting 2 (see Fig. 5.2): $N \in [100, 500]$, $p = 15$, $q = 3$, $\boldsymbol{\theta} = \{-5^\circ, 5^\circ, 10^\circ\}$, SNR $= -10$ dB.
  - Setting 3 (see Fig. 5.3): $N \in [100, 500]$, $p = 10$, $q = 2$, $\boldsymbol{\theta} = \{-5^\circ, 10^\circ\}$, SNR $= -11$ dB.
- Varying the SNR:
  - Setting 4 (see Fig. 5.4): SNR $\in [-10, -7]$ dB, $N = 1000$, $p = 15$, $q = 2$, $\boldsymbol{\theta} = \{0^\circ, 3^\circ\}$.

– Setting 5 (see Fig. 5.5): SNR $\in [-10, -5]$ dB, $N = 100$, $p = 15$, $q = 3$, $\boldsymbol{\theta} = \{-5°, 5°, 10°\}$.

– Setting 6 (see Fig. 5.6): SNR $\in [-10, -5]$ dB, $N = 50$, $p = 20$, $q = 3$, $\boldsymbol{\theta} = \{5°, 10°, 15°\}$.

In each setting all sources are assumed to have the same SNR.

Firstly, in Tables 5.3 and 5.4, the performance of the 3 methods presented in Section 5.3, i.e., the ΔL_q-distribution-based, ρ-distribution-based and Bootstrap-based methods, are compared quantitatively based on Settings 1 and 5, respectively. Herein, the result of random matrix theory is used for these 3 methods. For the bootstrap-based method, 300 Monte Carlo trials are used and the resampling size $B = 1000$. Surprisingly, the difference between the probabilities P_m's is less than 2%.

Table 5.3. P_m of the ΔL_q-distribution-based, ρ-distribution-based and Bootstrap-based methods in Setting 1.

N	1000	1200	1400	1600	1800	2000
Simulation (%)	97.93	88.76	64.88	33.26	11.59	2.82
ΔL_q	98.99	90.17	66.19	35.04	13.43	3.79
ρ	98.63	89.52	66.05	34.92	12.86	3.26
Bootstrap	98.54	89.39	66.02	35.09	13.09	3.44

Table 5.4. P_m of the ΔL_q-distribution-based, ρ-distribution-based and Bootstrap-based methods in Setting 5.

SNR (dB)	-10	-9	-8	-7	-6	-5
Simulation (%)	99.98	98.05	80.36	33.01	3.81	0.12
ΔL_q	99.99	98.59	79.13	32.19	5.15	0.34
ρ	99.97	98.16	78.73	32.04	4.71	0.25
Bootstrap	99.96	97.92	78.58	31.89	4.48	0.21

For the relatively large sample size, the results can be seen in Figs. 5.1 and 5.4. The proposed method "unbiased RMT" outperforms the other two methods and is in perfect agreement with the BIC simulation results. Nadler's method performs equivalently or even slightly better compared to the proposed method "RMT". In this case, Nadler's method provides a better solution of bias correction for the distribution of l_q and A_q than the method "RMT", since the former uses a more accurate estimate for the mean of the distribution of l_q and A_q than the delta method which is used by the latter. Additionally, the advantage of the method "RMT" is lost that random matrix theory gives most accurate variances of the distributions of l_q and A_q, due to $N \gg p - q$.

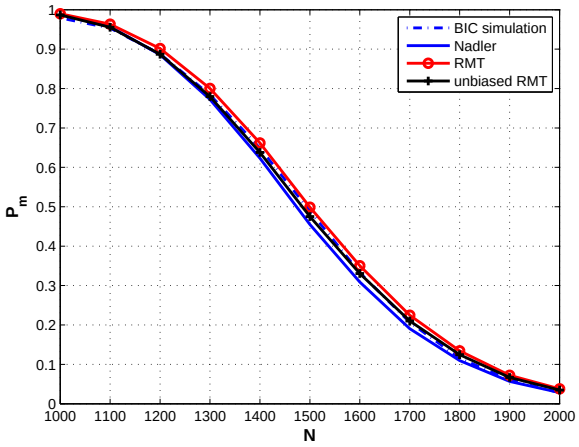

Figure 5.1. Probability of underestimation P_m versus N with $p = 25$, $q = 4$, $\boldsymbol{\theta} = \{10°, 13°, 17°, 20°\}$, SNR $= -9$ dB.

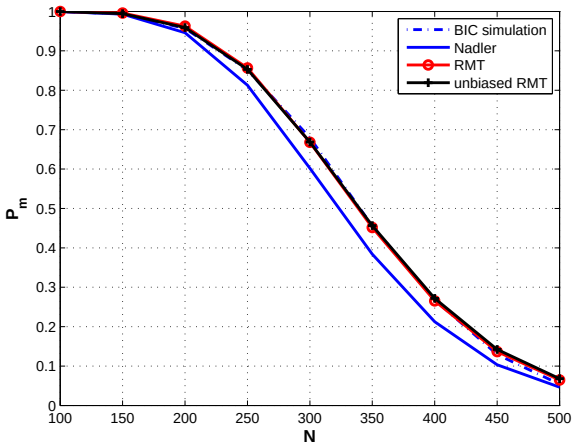

Figure 5.2. Probability of underestimation P_m versus N with $p = 15$, $q = 3$, $\boldsymbol{\theta} = \{-5°, 5°, 10°\}$, SNR $= -10$ dB.

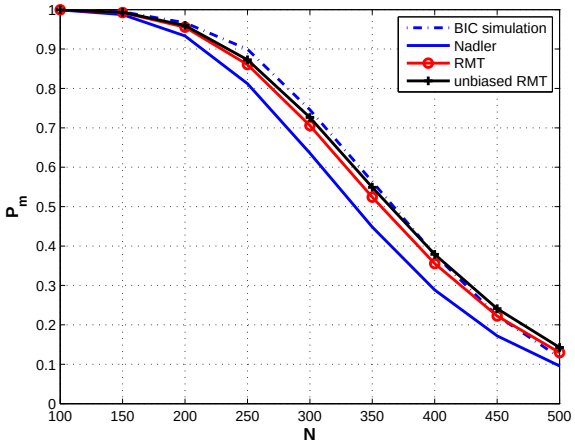

Figure 5.3. Probability of underestimation P_m versus N with $p = 10$, $q = 2$, $\boldsymbol{\theta} = \{-5°, 10°\}$, SNR $= -11$ dB.

For the moderate sample size, the results can be seen in Figs. 5.2, 5.3 and 5.5. Undoubtedly, the method "unbiased RMT" performs best and matches the BIC simulation results extraordinarily well. Nadler's method is defeated by the method "RMT". In some cases, e.g., in Figs. 5.2 and 5.5, the method "RMT" predicts the BIC simulation results very precisely, as the method "unbiased RMT".

For the relatively small sample size, the results can be seen in Figs. 5.6. Nadler's method deviates by far from the BIC simulation results, the method "RMT" deviates somewhat, and even the method "unbiased RMT" can not match the BIC simulation results exactly. In this case, all the asymptotic distributional results are not accurate enough, although random matrix theory is still better than classical multivariate statistical theory.

In general, the method "unbiased RMT" is superior to the other two methods in all cases we studied due to the fact that it employs all the signal population eigenvalues $\lambda_1, \cdots, \lambda_q$ and the bias of all the involved distributions is corrected. The performance of the method "RMT" and Nadler's method are highly dependent of the characteristics of the bias in the experimental settings, since they only correct partially the bias which is caused by the disturbance from the noise eigenvalues. In the case when the bias is not large or the disturbance from the signal eigenvalues is weak, the method "RMT" and Nadler's method yield satisfactory results. The former can handle the bias better than the latter in most cases. The method "RMT" yields a good match with the BIC simulation results in some conditions but slight deviation in other conditions where the distributions from random matrix theory have a biased sample mean.

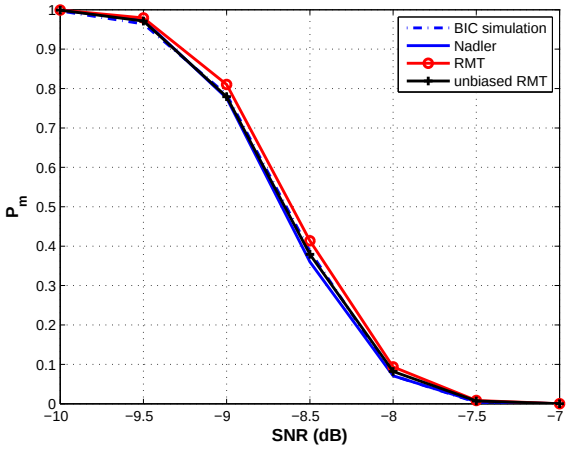

Figure 5.4. Probability of underestimation P_m versus SNR with $N = 1000$, $p = 15$, $q = 2$, $\boldsymbol{\theta} = \{0°, 3°\}$.

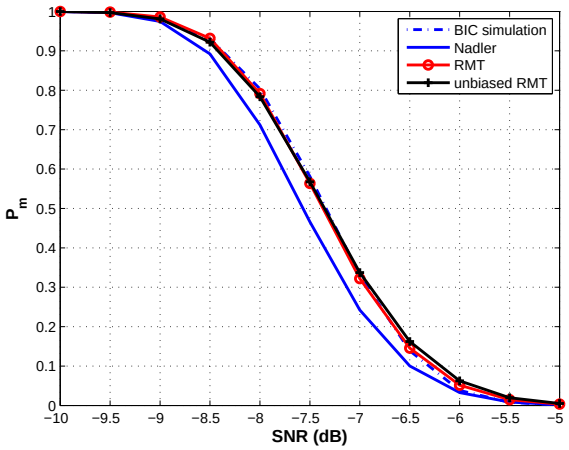

Figure 5.5. Probability of underestimation P_m versus SNR with $N = 100$, $p = 15$, $q = 3$, $\boldsymbol{\theta} = \{-5°, 5°, 10°\}$.

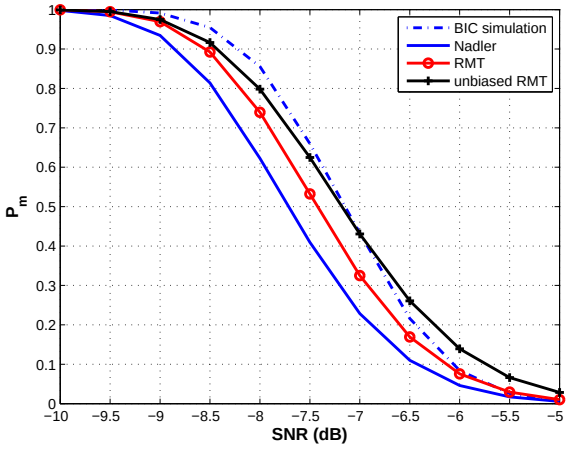

Figure 5.6. Probability of underestimation P_m versus SNR with $N = 50$, $p = 20$, $q = 3$, $\boldsymbol{\theta} = \{5°, 10°, 15°\}$.

It is worth mentioning that the BIC is widely used in practical situations when only a limited number of samples is available, although the BIC is derived based on asymptotic inference. This is motivated by the observation that the BIC yields accurate results for model selection in some cases of limited number of samples. Thus, our proposed methods are of great practical value since they are able to precisely predict the performance of the BIC in such cases.

5.5 Conclusions

In the context of source enumeration, we have analyzed theoretically the performance of the conventional BIC. Following the results of random matrix theory and Lawley's theory, 3 different methods have been presented to predict the probability of underestimation of the number of sources. These 3 proposed methods yield similar theoretical results which are in good agreement with the empirical results based on simulations. Simulations also validate that random matrix theory corrected by Lawley's theory characterizes the distributions of the eigenvalues perfectly, even for a moderate sample size.

Chapter 6

Flexible detection criterion (FDC)

The proposed procedures to evaluate theoretically the performance of the BIC in Chapter 5 shed light on the behavior of the BIC. In order to remedy the limitation of the BIC, that is, the BIC tends to underestimate the number of sources by one in most cases, an extra parameter is introduced into the BIC formulation, which results in a new criterion for source enumeration, referred as to the flexible detection criterion (FDC). By carefully choosing this parameter, the FDC is capable of substantially reducing the probability of underestimation. Note that the FDC is consistent and has low computational complexity, as the BIC.

The remainder of the chapter is organized as follows. Section 6.1 presents the formulation of the flexible detection criterion (FDC). An example is given in Section 6.2 and the choice of the parameter r is given in 6.3. Simulation results are given in Section 6.4 and conclusions are drawn in Section 6.5.

6.1 Formulation of FDC

The procedures for the performance analysis of the BIC in last section, especially the ρ-distribution-based method, provide insight into the mechanism of the BIC and show the means of improving the performance of the BIC. Starting from this, we develop a new criterion for source enumeration in this section.

As introduced in Chapter 5, the BIC tends to underestimate the number of sources by one, namely, $\hat{q} = q-1$. In order to improve the performance of the BIC, or equivalently increase its correct estimation probability of $P_c = P(\hat{q} = q)$, the underestimation probability $P_m = P(\hat{q} = q-1)$ has to be reduced. It is conspicuous from (5.19) that P_m becomes smaller if we reduce the threshold ρ_0 or increase the statistic ρ.

Most of previous works aim to reduce P_m by reducing ρ_0, more precisely, reducing the penalty term of the criterion in (5.1). It follows from (5.20) that ρ_0 is determined by the term on the right-hand side of the equation which is exactly Δy_q in (5.9), relying on the penalty term of the BIC, as shown in (5.7). It is found that reducing the penalty term leads to reducing ρ_0 and further P_m. Note that the log-likelihood term in (5.1) stays

intact in this case. Diverse penalty terms are implemented in [45] and [47]. However, finding the optimal penalty term is still an open question.

In this thesis, we reduce P_m by fixing the value of ρ_0 and replacing the statistic ρ by some other statistic, which has a higher value than ρ. This indicates that the penalty term of the criterion in (5.1) does not change and the log-likelihood term is modified. We replace l_i with l_i^r, for $r > 1$ and $i = 1, \ldots, p$. Consequently, the statistic related to the eigenvalues ρ is replaced by

$$\rho_r = \frac{l_q^r}{\frac{1}{p-q}\sum_{i=q+1}^{p} l_i^r} \tag{6.1}$$

which is larger than ρ for $r > 1$ since ρ_r is a monotonically increasing function of r. Then, the new criterion is given by

$$\hat{q} = \arg\min_k \left\{ -N(p-k)\log\frac{H_k}{B_k} + \frac{1}{2}k(2p-k)\log N \right\}, \quad k = 0, \ldots, p-1, \tag{6.2}$$

with

$$\begin{aligned} H_k &= \prod_{i=k+1}^{p} l_i^{r/(p-k)} \\ B_k &= \frac{1}{p-k}\sum_{i=k+1}^{p} l_i^r. \end{aligned} \tag{6.3}$$

The expression in (6.2) is referred to as the flexible detection criterion (FDC), since it contains a flexible parameter r. Clearly, the FDC reduces to the BIC when $r = 1$ and their computational costs are nearly the same. The FDC finds the optimal value of r subject to minimizing the probability of incorrect estimation P_e, namely,

$$r_c = \arg\min_r \{P_e(r) = P_m(r) + P_f(r)\} \tag{6.4}$$

where the underestimation probability P_m and the overestimation probability P_f are defined in (5.3) and (5.4), respectively. More precisely, P_m is given as

$$P_m = P(\rho_r < \rho_0). \tag{6.5}$$

To calculate P_f, the statistic related to the eigenvalues η_r is defined as

$$\eta_r = \frac{l_{q+1}^r}{\frac{1}{p-q-1}\sum_{i=q+2}^{p} l_i^r}, \tag{6.6}$$

and P_f is given as

$$P_f = P(\eta_r > \eta_0) \tag{6.7}$$

where $\eta_0 < \rho_0$. As introduced in Chapter 5, the distributions of ρ_r and η_r are prerequisite for calculating P_m and P_f. However, These distributions can not be derived readily for an arbitrary r. Only the distribution of ρ_r is given in Appendix 6.6.1 when r is an integer, such as $r = 1$ and 2. Also, these distributions contain two unknown parameters, namely, the noise variance σ^2 and the smallest signal population eigenvalue λ_q. Thus, it is unclear how to find the optimal value r_c using (6.4).

It is known that both ρ_r and η_r are monotonically increasing functions of r. Based on the definition of the probabilities P_m and P_f, the following lemma follows:

Lemma 6.1.1 *As the parameter r of the FDC increases, its underestimation probability P_m decreases and overestimation probability P_f increases.*

Note that the sum of the probabilities P_m and P_f is minimized at r_c so that both P_m and P_f are negligible at r_c in most cases when the sample size is large enough and the signal and noise eigenvalues are well separated. Thus, it follows from Lemma 6.1.1 that if $r \gg r_c$, the FDC yields a notable P_f and a negligible P_m, whereas if $r \ll r_c$ the FDC yields a notable P_m and a negligible P_f. The FDC is capable of controlling the probabilities P_m and P_f by choosing different values of r, which is impossible with the BIC. Furthermore, we can roughly find the distance between r and r_c based on the values of P_m and P_f if some prior information is available.

Since the BIC is a special case of the FDC when $r = 1$, Lemma 6.1.2 follows immediately from Lemma 6.1.1:

Lemma 6.1.2 *The FDC yields a lower underestimation probability P_m and a higher overestimation probability P_f than the BIC.*

In Appendix 6.6.2, this lemma is proven in a different context.

6.2 An example of FDC

To illustrate Lemmas 6.1.1 and 6.1.2, an example is given here for the experimental setting: $N = 100$ samples, $p = 15$ sensors, and $q = 3$ sources in the directions of

$-5°, 5°, 10°$ with SNR $= -9$ dB. In Table 6.1, simulation results are shown for the probabilities of underestimation P_m, overestimation P_f and correct estimation P_c with different values of r. As r increases from 1 to 2, P_m decreases and P_f increases, which coincides with Lemma 6.1.1. When r is close to 1, the probability of incorrect estimation P_e is primarily contributed by P_m, and when r is close to 2, P_f dominates P_e. The highest probability of correct estimation, i.e., $P_c = 0.9026$, is found at $r = 1.5$, which is significantly improved compared to $P_c = 0.0158$ at $r = 1$. In this case when $r = 1$ has a very low P_c, all the choices of r in our example have an improvement in terms of P_c, even $r = 1.1$.

In addition, Fig. 6.1 shows the pdfs $f(\rho_r)$ and $f(\eta_r)$ for $r = [1, 1.5, 1.7, 2]$ with respect to ρ_0.[1] $f(\rho_r)$ and $f(\eta_r)$ are plotted in blue and black, respectively. The red dashed-dotted line denotes the location of ρ_0. As r increases, both $f(\rho_r)$ and $f(\eta_r)$ are shifted to the right and become flat. Recall that the threshold ρ_0 is fixed and $P_m = P(\rho_r < \rho_0)$ and $P_f = (\eta_r > \rho_0)$. To minimize P_m and P_f, the blue curve should be located on the right side of the red line as far away as possible and the black curve goes in the inverse direction. This occurs at $r = 1.5$. The variation of the pdfs $f(\rho_r)$ and $f(\eta_r)$ in Fig. 6.1 intuitively reflects the underlying idea of designing the FDC.

Table 6.1. The probabilities P_m, P_f and P_c for different values of r.

r	1.0	1.1	1.2	1.3	1.4	**1.5**	1.6	1.7	1.8	1.9	2.0
P_m (%)	98.42	86.27	61.59	33.24	16.19	**6.81**	2.66	0.91	0.31	0.13	0
P_f	0	0	0	0.01	0.49	**2.93**	10.36	25.12	45.80	67.05	83.77
P_c	1.58	13.73	38.41	66.75	83.32	**90.26**	86.98	73.97	53.89	32.82	16.23

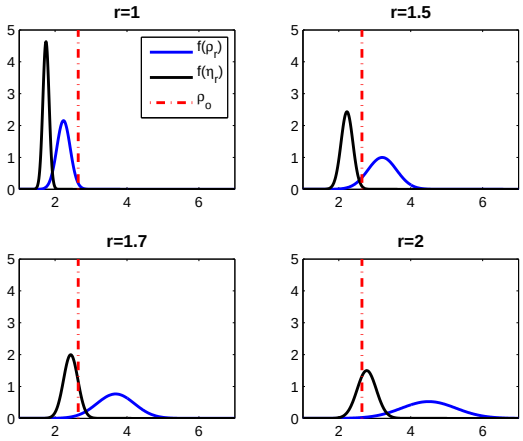

Figure 6.1. The pdfs $f(\rho_r)$ and $f(\eta_r)$ for different values of r.

[1]Here, η_0 is replaced by ρ_0 for simplicity, since η_0 is only slightly smaller than ρ_0.

6.3 Choice of r

In this section, we provide a selection procedure for r. First, a set of r is given, which ensures consistency of the FDC. Second, a typical value of r is given, which ensures superior performance of the FDC.

It is clear that the performance of the FDC relies on the value of r. The values of r should be close to r_c, in order to let the FDC perform well in difficult conditions, such as low SNR, small sample size and close-spacing between the sources. As shown in Fig. 6.1, r_c is found when ρ_0 is located between the two peaks of the pdfs $f(\rho_r)$ and $f(\eta_r)$. Thus, these values of r close to r_c are chosen such that $\mu_{\eta_r} < \eta_0 < \rho_0 < \mu_{\rho_r}$ is fulfilled. Herein, μ_{ρ_r} and μ_{η_r} are the mean values of ρ_r and η_r in (6.1) and (6.6), respectively. Then, r should reside in the interval (r_l, r_u). r_l and r_u are the lower and upper bounds, respectively. The lower bound of r is chosen such that

$$\rho_0 = \mu_{\rho_r}. \tag{6.8}$$

Analogous to μ_{ρ_1} given in (6.16), μ_{ρ_r} is unknown in practice, since λ_q is unknown. Hence, the lower bound of r cannot be obtained usually based on (6.8). For simplicity, we set the lower bound $r_l = 1$. The upper bound of r is chosen such that

$$\mu_{\eta_r} = \eta_0. \tag{6.9}$$

We elaborate on the relationship between r_u and r later.

The following lemma gives the asymptotic properties of the FDC when $1 < r < r_u$.

Lemma 6.3.1 *The FDC is a consistent estimator, that is, estimating the true number of sources with probability one as the sample size increases to infinity.*

Proofs are given in Appendix 6.6.3.

According to Corollary 11.1.2 in [129, page 558], $l_i^r, i = 1, \ldots, p$, are the eigenvalues of the matrix $\hat{\boldsymbol{R}}^r$, which leads to the FDC sharing similar asymptotic properties of the efficient detection criterion (EDC) in [45][61]. Besides, both criteria are the modification of the BIC, more precisely, the EDC has much freedom in constructing the penalty term where the FDC has much freedom in constructing the "log-likelihood term". Note that the FDC is developed primarily based on the concept of thresholding rather than information theoretic criteria. Thus, the first term of the FDC in (6.2) no longer relates to the log-likelihood function.

Although all the values of r chosen from the interval $(1, r_u)$ render the FDC outperforming the BIC in most cases, we attempt to find a value of r from this interval, which maximizes the performance of the FDC, that is, its underestimation probability P_m is minimized while its overestimation probability P_f is preserved to be nearly zero. According to Lemma 6.1.1, a low P_m is resulted from a high value of r, which implies that the value of r should be closer to r_u than 1. It is worth mentioning that, although $r < r_u$ guarantees consistency of the FDC for large sample sizes, a value r that is very close to r_u, i.e., $r \approx r_u$, leads to that P_f of the FDC does not decay to zero as the SNR increases. This undesired phenomenon is explained below.

Suppose that we change the SNR by changing the signal power and keeping the noise power intact, which is also the case of interest in practice. Assume that we have a value of r which is very close to r_u. Then, there is a notable P_f dominating P_e, according to Lemma 6.1.1, e.g., see the case when $r = 1.7$ in Fig. 6.1. As the SNR increases, ρ_r increases and ρ_0 stays the same such that $P_m = P(\rho_r < \rho_0)$ decreases. However, both η_r and η_0 stay the same such that $P_f = P(\eta_r > \eta_0)$ does not change. It indicates that P_f does not vanish as the SNR increases.

In order to minimize P_f, we need to keep the value of r a certain distance from r_u, that is,

$$r = r_u - \delta \tag{6.10}$$

where $\delta > 0$. Here comes the question: how to choose the value of δ? Theoretically, δ depends on the distribution of the random variable η_r in (6.6) or simply its variance $\sigma^2_{\eta_r}$. For example, similar to the 3-sigma rule, we can choose δ and r based on the following reformulation of (6.9), namely,

$$\mu_{\eta_r} + 3\sigma_{\eta_r} = \eta_0. \tag{6.11}$$

Then, it results in a fairly small P_f. The variance $\sigma^2_{\eta_r}$ can be calculated using the Tracy-Widom distribution [94]. We can solve (6.11) based on some optimization methods to derive r. However, the calculation of $\sigma^2_{\eta_r}$ is tedious due to the presence of r. To avoid this, we choose r based on the value r_s, which is given in (6.27) in Appendix 6.6.4.

Appendix 6.6.4 shows that the value r_s is smaller than r_u when the sample size is finite. However, r_s approaches to r_u as the sample size tends to infinity. The distance between r_s and r_u becomes larger as q increases. For example, when $q = 0$, r_s is approximately equal to r_u; when $q = 3$, r_s is considerably smaller than r_u. Based on simulations, we find that the value of r_s calculated using $q = 3$ in (6.27) is sufficiently smaller than those values of r_u for different q less than or equal to 4.[2] Then, the value of $r_s(q = 3)$

[2]Note that we assumed that $p \geq 10$, in order to meet the assumption $q \ll p$ made for (6.24).

can be used to construct the FDC to estimate the number of sources $q \leq 4$. If $q > 4$, we can calculate r_s with the supposed maximal number $q_{\max}$. The performance of the FDC with $r_s(q = 3)$ is given in detail in Section 6.4.

An alternative approach to optimize r would be to use a resampling technique, such as the bootstrap. This would increase the computational expense and thus, we recommend the approach above.

6.4 Simulation results

A uniform linear array with omni-directional sensors and inter-sensors spacing of half the wavelength is employed for simulations. We consider only the case of uncorrelated complex circular Gaussian source signals contaminated by complex circular white Gaussian noise. The numbers of samples, sensors and sources are denoted by N, p and q, respectively. The DOAs are denoted by the vector $\boldsymbol{\theta}$. "SNR" is short for the signal-to-noise ratio. The probability of correctly estimating q is denoted by P_c.

For evaluating the proposed FDC, we include the approach proposed in [94] for performance comparison, which is denoted by "KN" in the simulations. As an approach of hypothesis testing, the false alarm rate of "KN" is set to $\alpha = 0.1\%$, as suggested in [94]. The simulation results are based on 10,000 Monte Carlo trials. Four experimental settings are listed in the following:

- Varying the sample size N:
  - Setting 1 (see Fig. 6.2): $N \in [400, 2000]$, $p = 10$, $q = 2$, $\boldsymbol{\theta} = \{0°, 3°\}$, SNR $= -7$ dB, $r = 1.7092$.
  - Setting 2 (see Fig. 6.3): $N \in [400, 2000]$, $p = 15$, $q = 3$, $\boldsymbol{\theta} = \{-3°, 0°, 3°\}$, SNR $= -3$ dB, $r = 1.7836$.
- Varying the SNR:
  - Setting 3 (see Fig. 6.4): SNR $\in [-4, -4]$ dB, $N = 2000$, $p = 10$, $q = 3$, $\boldsymbol{\theta} = \{-4°, 0°, 4°\}$, $r = 1.7092$.
  - Setting 4 (see Fig. 6.5): SNR $\in [-8, 0]$ dB, $N = 3000$, $p = 15$, $q = 4$, $\boldsymbol{\theta} = \{-4°, 0°, 4°, 8°\}$, $r = 1.8288$.

Following the statement in Section 6.3, we calculate r_s using (6.27) with the available maximal sample size N and the array size $q = 3$. For Setting 1 and Setting 3, r is chosen to be r_s, which is calculated using (6.27) with substituting $N = 2000$, $p = 10$, $q = 3$. For Setting 2, r is equal to r_s, calculated with $N = 2000$, $p = 15$, $q = 3$, whereas for Setting 4, r is equal to r_s, calculated with $N = 3000$, $p = 15$, $q = 3$.

Figs. 6.2 and 6.3 show that both the FDC and "KN" are superior to the BIC, and have fast convergence with respect to increasing sample sizes. However, the FDC outperforms "KN" in the regime of relatively small sample sizes. Figs. 6.4 and 6.5 show that the FDC performs similar to "KN" and they yield significant performance improvement over the BIC. The BIC is dominated by tremendous probability of underestimation P_m, in contrast to the FDC and "KN". The performance of the FDC validates the way of choosing r proposed in Section 6.3.

The preceding figures also show that the FDC suffers slightly overestimating q, analogous to "KN", especially for high SNRs. The probability of overestimation P_f is less than 1% in most cases. If we want to eliminate P_f for high SNRs, the value of r can be selected to be e.g., $0.3 \sim 0.5$ smaller than r_s, however, at the cost of slight performance degradation for low SNRs. Note that all the values of r chosen between 1 and r_s ensure the FDC outperforming the BIC in most cases.

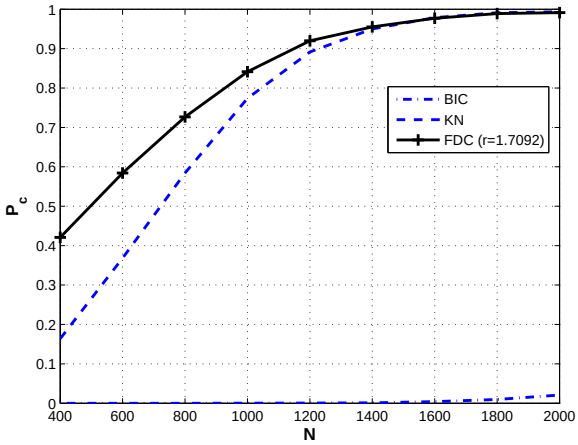

Figure 6.2. Probability of correct estimation P_c versus N with $p = 10$, $q = 2$, $\boldsymbol{\theta} = \{0°, 3°\}$, SNR $= -7$ dB.

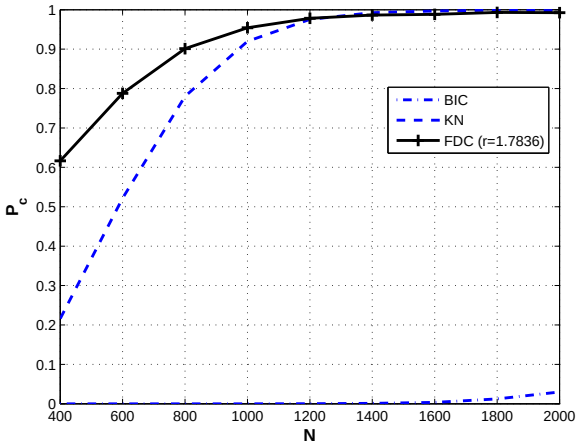

Figure 6.3. Probability of correct estimation P_c versus N with $p = 15$, $q = 3$, $\boldsymbol{\theta} = \{-3°, 0°, 3°\}$, SNR $= -3$ dB.

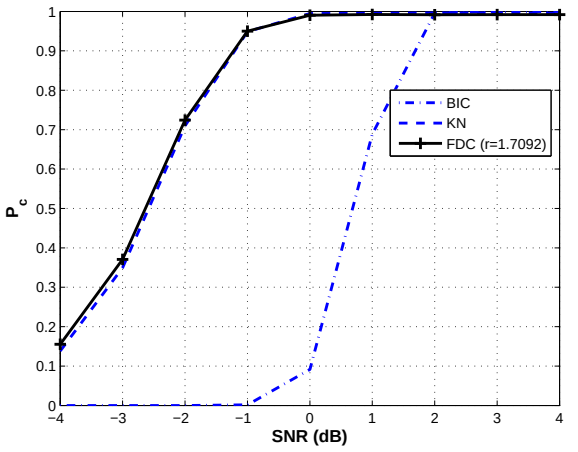

Figure 6.4. Probability of correct estimation P_c versus SNR with $N = 2000$, $p = 10$, $q = 3$, $\boldsymbol{\theta} = \{-4°, 0°, 4°\}$.

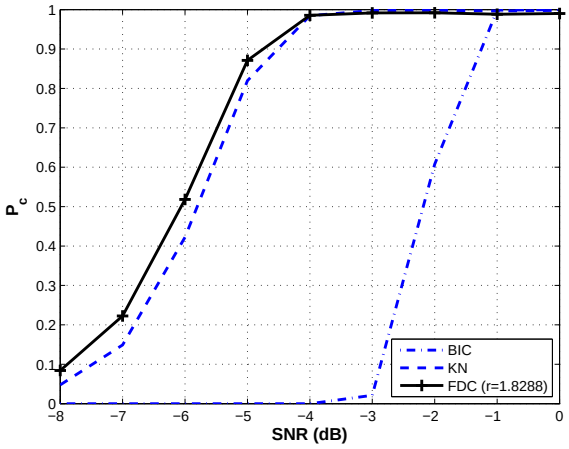

Figure 6.5. Probability of correct estimation P_c versus SNR with $N = 3000$, $p = 15$, $q = 4$, $\boldsymbol{\theta} = \{-4°, 0°, 4°, , 8°\}$.

6.5 Conclusions

We have developed the flexible detection criterion (FDC) which contains an extra parameter, motivated by the observation that the BIC can be intuitively regarded as a thresholding approach in the context of source enumeration. Some properties of the FDC and rules of choosing the parameter are suggested with theoretical justification. By tuning this parameter, its probabilities of underestimation and overestimation can be controlled to a very low level. The FDC significantly outperforms the BIC while their computational costs are comparable.

6.6 Appendix

6.6.1 Distributions of ρ_1 and ρ_2

Define $\rho_1 = \frac{l_q}{A_q}$ and $\rho_2 = \frac{\tau_q}{B_q}$ with $\tau_q = l_q^2$ and $B_q = \frac{1}{p-q}\sum_{i=q+1}^{p} l_q^2$. Assume $\sigma^2 = 1$ and $N \gg (p-q)$, that is, $0 < \gamma = \frac{p-q}{N} \ll 1$. Thus, we can ignore all the terms concerning the interaction between the eigenvalues in the distributions of l_q and A_q. The asymptotic Gaussian distributions have the following means and variances:

$$\mu_{l_q} = \lambda_q, \qquad \sigma_{l_q}^2 = \frac{\lambda_q^2}{N} \tag{6.12}$$

$$\mu_{A_q} = 1, \qquad \sigma_{A_q}^2 = \frac{1}{N(p-q)}. \tag{6.13}$$

Further, τ_q and B_q are found to be asymptotically Gaussian distributed with the following parameters:

$$\mu_{\tau_q} = \lambda_q^2, \qquad \sigma_{\tau_q}^2 = \frac{4\lambda_q^4}{N} \tag{6.14}$$

$$\mu_{B_q} = 1 + \gamma, \qquad \sigma_{B_q}^2 = \frac{2(2\gamma^2 + 5\gamma + 2)}{N(p-q)}. \tag{6.15}$$

Herein, μ_{τ_q} and $\sigma_{\tau_q}^2$ in (6.14) are derived based on the delta method, and μ_{B_q} and $\sigma_{B_q}^2$ in (6.15) are derived based on the Marčenko-Pastur density in [124] and [90], which cannot be derived using the delta method. By simply using the delta method for the ratio distribution, more precisely, using (5.18), ρ_1 and ρ_2 are also derived to be Gaussian distributed with the following parameters:

$$\begin{aligned} \mu_{\rho_1} &= \lambda_q \\ \sigma_{\rho_1}^2 &= \frac{\lambda_q^2}{N}(1 + \frac{1}{p-q}) \end{aligned} \tag{6.16}$$

and

$$\begin{aligned} \mu_{\rho_2} &= \frac{\lambda_q^2}{1+\gamma} \\ \sigma_{\rho_2}^2 &= \frac{\lambda_q^4}{N(1+\gamma)^2}\left[4 + \frac{2(2\gamma^2 + 5\gamma + 2)}{(1+\gamma)^2(p-q)}\right], \end{aligned} \tag{6.17}$$

respectively. It is noteworthy that we can use the preceding distributions to validate that the probability of underestimation P_m decreases when r is switched from 1 to 2, that is, $P(\rho_1 < \rho_0) > P(\rho_2 < \rho_0)$, which is out of the scope of this thesis.

6.6.2 Proof of Lemma 6.1.2

According to Jensen's inequality, for $r > 1$, we have

$$\frac{1}{p-q}\sum_{i=q+1}^{p} l_i^r \geq \left(\frac{1}{p-q}\sum_{i=q+1}^{p} l_i\right)^r \tag{6.18}$$

where the equality holds when $l_{q+1} = \cdots = l_p$. When the sample size N is relatively large, the noise sample eigenvalues $l_{q+1}, \cdots, l_p$ are approximately equal. Then, we have

$$\begin{aligned} \rho_r &= \frac{l_q^r}{\frac{1}{p-q}\sum_{i=q+1}^{p} l_i^r} \\ &\simeq \frac{l_q^r}{\left(\frac{1}{p-q}\sum_{i=q+1}^{p} l_i\right)^r} \\ &= \rho_1^r. \end{aligned} \tag{6.19}$$

Then, we reformulate the underestimation probability P_m of the FDC as

$$\begin{aligned} P_m &= P(\rho_r < \rho_0) \simeq P\left(\rho_1^r < \rho_0\right) \\ &= P\left(\rho_1 < \rho_0^{\frac{1}{r}}\right) \end{aligned} \tag{6.20}$$

which is lower than $P_m = (\rho_1 < \rho_0)$ of the BIC, since $\rho_0^{\frac{1}{r}} < \rho_0$ when $r > 1$. Note that $\rho_0 > 1$.

Similarly, the overestimation probability P_f can be expressed as

$$\begin{aligned} P_f &= P(\eta_r > \eta_0) \simeq P\left(\eta_1^r > \eta_0\right) \\ &= P\left(\eta_1 > \eta_0^{\frac{1}{r}}\right) \end{aligned} \tag{6.21}$$

which is higher than $P_f = P\left(\eta_1 > \eta_0\right)$ of the BIC, since $\eta_0^{\frac{1}{r}} < \eta_0$ when $r > 1$. It concludes the proof of Lemma 6.1.2.

6.6.3 Proof of Lemma 6.3.1

In order to prove large-sample consistency of the FDC with $1 < r < r_u$, we show below that the probability of incorrect estimation $P_e = P_m + P_f \to 0$ as the sample size $N \to \infty$.

As stated in Lemma 6.1.2, the FDC has a lower underestimation probability P_m than the BIC. Due to consistency of the BIC, P_m of the FDC approaches to zero as N goes to infinity.

As shown in Section 6.1, $r < r_u$ ensures that $\mu_{\eta_r} < \eta_0$. Similar to $\sigma^2_{\rho_r}$ in (6.16) and (6.17), the variance $\sigma^2_{\eta_r}$ of η_r is of order $O(N^{-1})$. It can be seen from (6.23) and (6.25) that $\eta_0 \approx 1 + O(\sqrt{\frac{\log N}{N}})$ and $\mu_{\eta_r} \approx 1 + O(N^{-r})$. As $N \to \infty$, $\sigma^2_{\eta_r}$ approaches to zero, whereas η_0 and μ_{η_r} approach to one. However, the convergence of $\sigma^2_{\eta_r}$ and μ_{η_r} is much faster than that of η_0. Then all the values of η_r tend to be smaller than η_0. It indicates that $P_f = P(\eta_r > \eta_0)$ tends to zero.

Then, it follows that $P_e = P_m + P_f \to 0$ as the sample size $N \to \infty$. The consistency of the FDC with $1 < r < r_u$ is proven. Also, it is noteworthy that the FDC can be easily proven to be consistent, following the same steps in [45].

6.6.4 The value of r_s

The true value of η_0 can be found by solving the following equation

$$\log\left[\frac{1}{\eta_0}\left(1+\frac{\eta_0-1}{p-q}\right)^{p-q}\right]=\frac{(2p-2q-1)\log N}{2N} \tag{6.22}$$

which has a similar formulation to (5.20). According to [39, Theorem 2], η_0 is approximately given by

$$\eta_0 \simeq 1+\sqrt{\frac{2(p-q)[2(p-q)-1]}{p-q-1}}\cdot\sqrt{\frac{\log N}{N}} \tag{6.23}$$

which, however, is much smaller than the true value of η_0, as shown in Fig. 2 in [40]. In the sequel, the true value of η_0 is used, instead of the approximation in (6.23).

According to the Marčenko-Pastur law in [93][102][105], the upper limit of the noise eigenvalues is $\sigma^2(1+\sqrt{c})^2$ as $p, N \to \infty$ with $p/N \to c \in (0,\infty)$. Then, in our case, the largest noise eigenvalue l_{q+1} converges asymptotically to $\sigma^2(1+\sqrt{c})^2$, more precisely,

$$l_{q+1} \leq \sigma^2(1+\sqrt{c})^2, \tag{6.24}$$

assuming $q \ll p$.

When the sample size N is large, μ_{η_r} can be approximated by

$$\begin{aligned}\mu_{\eta_r} &= \frac{\mu(l_{q+1}^r)}{\mu\left(\frac{1}{p-q-1}\sum_{i=q+2}^{p} l_i^r\right)} \\ &\leq \frac{\mu(l_{q+1}^r)}{\mu\left[\left(\frac{1}{p-q-1}\sum_{i=q+2}^{p} l_i\right)^r\right]} \\ &\leq \frac{\sigma^{2r}(1+\sqrt{c})^{2r}}{\sigma^{2r}} = (1+\sqrt{c})^{2r}\end{aligned} \tag{6.25}$$

where the second inequality follows from (6.18). The equality in (6.25) holds only for infinite N. Recall (6.9) and define r_s such that

$$(1+\sqrt{c})^{2r_s} = \eta_0 \tag{6.26}$$

where η_0 is calculated in (6.22). Then, we derive

$$r_s = \frac{\log\,\eta_0}{2\log\left(1+\sqrt{\frac{p}{N}}\right)}. \tag{6.27}$$

Obviously, it follows from (6.9), (6.25) and (6.26) that $r_s \leq r_u$, where the equality holds only for infinite N.

It is seen that r_s depends on the sample size N, the array size p and the number of sources q, more precisely, r_s increases as N or p increases, whereas r_s decreases as q increases. Most importantly, as q increases, l_{q+1} becomes increasingly smaller than the threshold $\sigma^2(1+\sqrt{c})^2$. Correspondingly, μ_{η_r} becomes increasingly smaller than $(1+\sqrt{c})^{2r}$. As a result, r_s becomes increasingly smaller than r_u as q increases. In order to let r_s be sufficiently smaller than r_u, we can calculate r_s using a large value of q.

Chapter 7

Generalized Bayesian information criterion (GBIC)

In this chapter, we present a very general rule for constructing the Bayesian information criterion, which we refer to as the generalized Bayesian information criterion (GBIC). The proposed GBIC overcomes the limitations of the conventional BIC by using more information from the available data. In contrast to the conventional BIC, GBIC not only uses the density of the observed data, but also the density of the sample eigenvalues or corresponding statistics. It consists of two expressions for different cases. As a representative example of the rule, we take into account the density of all the sample eigenvalues instead of various statistics based on them. Also, we compare the theoretical and numerical performance of the conventional BIC and the proposed GBIC, which provides a solid justification for the GBIC rule.

The remainder of the chapter is organized as follows. In Section 7.1, the generalized Bayesian information criterion (GBIC) rule is proposed, which contains two expressions. An example of GBIC is presented based on the density of the sample eigenvalues in Section 7.2. For a demonstration of the performance improvement of GBIC, simulation results and discussions are presented in Section 7.3, before conclusions are drawn in Section 7.4.

7.1 Formulation of GBIC

Following an original formulation of BIC in [27], we introduce two expressions of GBIC for different cases.

7.1.1 An original formulation of BIC

As introduced in Section 2.2.2, the conventional BIC involves only the density of the observed data. The density of the parameters, e.g., in (2.15) or (2.16), is neglected, which, however, is the essential ingredient of the Bayesian framework. In essence, a more precise formulation of BIC can be given as follows (see Eq. (84) in [27]):

$$\mathrm{BIC}(k) = -2\left[\log f(\mathcal{X}|\hat{\boldsymbol{\Theta}}^{(k)}) + \log f(\hat{\boldsymbol{\Theta}}^{(k)})\right] + n_k \log N \tag{7.1}$$

which contains the density function $f(\hat{\boldsymbol{\Theta}}^{(k)})$. The formulation in (7.1) can be simplified to the one in (2.12) based on two assumptions on $f(\boldsymbol{\Theta}^{(k)})$, that is,

- $f(\boldsymbol{\Theta}^{(k)})$ is flat around the ML estimator $\hat{\boldsymbol{\Theta}}^{(k)}$,
- $f(\boldsymbol{\Theta}^{(k)})$ is independent of the sample size N.

If these two assumptions, especially the second one, are not fulfilled, the density $f(\hat{\boldsymbol{\Theta}}^{(k)})$ can not be removed from (7.1). Hence, it is implied that $f(\hat{\boldsymbol{\Theta}}^{(k)})$ carries useful information for BIC. In most practical cases, it is not possible to judge if the preceding two assumptions hold since both $f(\boldsymbol{\Theta}^{(k)})$ and $f(\hat{\boldsymbol{\Theta}}^{(k)})$ are unknown. This is why the authors in [27] argued that the criterion in (7.1) has little practical value.

7.1.2 Part I of GBIC

In array processing, $f(\hat{\boldsymbol{\Theta}}^{(k)})$ is the joint density of the signal sample eigenvalues and eigenvectors as well as the ML noise variance estimator. Some simple expressions exist for the density of the sample eigenvalues and the ML noise variance estimator. However, the density of the sample eigenvectors is too cumbersome for general use, which sometimes renders the criterion in (7.1) unfeasible. If the density of the sample eigenvectors can be excluded, the construction of the criterion in (7.1) will be much easier. It is known from practice that the eigenvectors contain much less information for the problem of source enumeration than the eigenvalues such that exclusion of the eigenvectors from the formulation in (7.1) leads to slight performance degradation. This is confirmed theoretically in the sequel.

Let us recall the derivation of BIC in (2.19). Following the joint pdf of the observations in (2.13), the log-likelihood function is given by

$$\log f(\mathcal{X}|\boldsymbol{\Theta}^{(k)}) = -Np\log\pi - N\log\left\{\det[\mathbf{R}^{(k)}]\right\} - N\mathrm{tr}\left\{[\mathbf{R}^{(k)}]^{-1}\,\hat{\mathbf{R}}\right\}. \qquad (7.2)$$

By replacing the parameter vector $\boldsymbol{\Theta}^{(k)}$ with the ML estimator $\hat{\boldsymbol{\Theta}}^{(k)}$ in (2.16), we

obtain[1]

$$\begin{aligned}\log f(\mathcal{X}|\hat{\boldsymbol{\Theta}}^{(k)}) &= -Np\log\pi - N\log\left[\prod_{i=1}^{k}\lambda_i\cdot(\sigma^2)^{p-k}\right] - N\left[\sum_{i=1}^{k}\frac{l_i}{\lambda_i}+\sum_{i=k+1}^{p}\frac{l_i}{\sigma^2}\right] \\ &= -N\log\left[\prod_{i=1}^{k}l_i\cdot\left(\frac{1}{p-k}\sum_{i=k+1}^{p}l_i\right)^{p-k}\right] - Np\log\pi - Np \\ &= N(p-k)\log\left(\frac{\prod_{i=k+1}^{p}l_i^{1/(p-k)}}{\frac{1}{p-k}\sum_{i=k+1}^{p}l_i}\right) - Np(\log\pi+1) - N\log\left[\det(\hat{\mathbf{R}})\right]\end{aligned} \tag{7.3}$$

where the last equality holds due to the fact that $\prod_{i=1}^{k} l_i = \det(\hat{\mathbf{R}})/\prod_{i=k+1}^{p} l_i$. Its second and third terms are constants with respect to k so that they have no influence in the choice of q. It readily follows from (7.3) that the dependence of the likelihood function $f(\mathcal{X}|\hat{\boldsymbol{\Theta}}^{(k)})$ on the sample eigenvectors is inexistent, which confirms the unimportance of the eigenvectors for the problem of source enumeration.

Furthermore, $f(\mathcal{X}|\hat{\boldsymbol{\Theta}}^{(k)})$ can be reformulated based on different sets of parameters, e.g., the sample eigenvalues $l_1,\cdots,l_p$ and the ML noise variance estimator $\hat{\sigma}^2$ in (2.16). The second equality in (7.3) shows that $f(\mathcal{X}|\hat{\boldsymbol{\Theta}}^{(k)})$ depends on all the sample eigenvalues $l_1,\cdots,l_p$, whereas the third equality shows that $f(\mathcal{X}|\hat{\boldsymbol{\Theta}}^{(k)})$ depends on the noise sample eigenvalues $l_{k+1},\cdots,l_p$ only, ignoring the constant terms. If $\hat{\sigma}^2$ is substituted for $\frac{1}{p-k}\sum_{i=k+1}^{p} l_i$ in the second and third equalities in (7.3), the unknowns become $l_1,\cdots,l_k,\hat{\sigma}^2$ and $l_{k+1},\cdots,l_p,\hat{\sigma}^2$, respectively. Hence, following (7.3), the parameter vector $\hat{\boldsymbol{\Theta}}^{(k)}$ shrinks to $\mathcal{Z}^{(k)}$ which can take the following combinations:

- $l_1,\cdots,l_p$
- $l_{k+1},\cdots,l_p$
- $l_1,\cdots,l_k,\hat{\sigma}^2$
- $l_{k+1},\cdots,l_p,\hat{\sigma}^2$.

It is known that the number of elements of $\mathcal{Z}^{(k)}$ changes as the candidate of the number of sources k changes from 0 to $p-1$. To ease notation, $\mathcal{Z}$ is used instead of $\mathcal{Z}^{(k)}$ in the sequel. Then, following (7.3), we have

$$f(\mathcal{X}|\hat{\boldsymbol{\Theta}}^{(k)}) = f(\mathcal{X}|\mathcal{Z}). \tag{7.4}$$

[1]When $k=0$, all the population eigenvalues $\lambda_i=\sigma^2, i=1,\ldots,p$.

We conclude that after the ML estimator $\hat{\boldsymbol{\Theta}}^{(k)}$ is substituted into the density of the observations $f(\mathcal{X})$, $f(\mathcal{X})$ depends on $\mathcal{Z}$ only, rather than $\hat{\boldsymbol{\Theta}}^{(k)}$. Thus, the criterion in (7.1) can be reformulated as

$$\begin{aligned} \mathrm{GBIC}_1(k) &= -2\left[\log f(\mathcal{X}|\mathcal{Z}) + \log f(\mathcal{Z})\right] + n_k \log N \\ &= -2\left[\log f(\mathcal{X}|\hat{\boldsymbol{\Theta}}^{(k)}) + \log f(\mathcal{Z}|\hat{\boldsymbol{\Theta}}_z^{(k)})\right] + n_k \log N \end{aligned} \tag{7.5}$$

which is the first expression of GBIC. The second equality in (7.5) is obtained based on the relationship (7.4). Herein, $\hat{\boldsymbol{\Theta}}_z^{(k)}$ is the ML estimator based on the likelihood function $f(\mathcal{Z}|\boldsymbol{\Theta}_z^{(k)})$. Since $\mathcal{Z}$ is totally determined by the observations $\mathcal{X}$, $\boldsymbol{\Theta}_z^{(k)}$ is only a subset of the parameter vector $\boldsymbol{\Theta}^{(k)}$ in (2.15). Thus, the parameter space of the model does not change and n_k stays the same as the one in (2.19). There are different formulations for GBIC_1 depending on the choices of $\mathcal{Z}$, which is caused by the ambiguity of parametrization of the density of the observations. If $f(\mathcal{Z}|\hat{\boldsymbol{\Theta}}_z^{(k)})$ were known, GBIC_1 in (7.5) could be readily implemented.

GBIC_1 in (7.5) excludes the density of the sample eigenvectors and the log-likelihood function of GBIC_1 is much simpler than that of the criterion in (7.1). Derivation of (7.5) from (7.1) can also be explained intuitively using Fisher's argument [130], that is, inference should involve only the part of the likelihood function that is known, omitting the unknown part. Assume that the likelihood function w.r.t all parameters, i.e., $f(\hat{\boldsymbol{\Theta}}^{(k)})$ in the second term of (7.1), is not completely known. More precisely, the likelihood function w.r.t. the sample eigenvectors is unknown and only the likelihood function $f(\mathcal{Z}|\hat{\boldsymbol{\Theta}}_z^{(k)})$ is known. It follows from Fisher's argument that $f(\hat{\boldsymbol{\Theta}}^{(k)})$ in the second term of (7.1) should be replaced by $f(\mathcal{Z}|\hat{\boldsymbol{\Theta}}_z^{(k)})$, which results in (7.5). In [131], Fisher's argument is also used in the problem of model order selection, in order to remove the unknown part of the likelihood function.

In a more general sense, if some additional information $\mathcal{Z}$, in addition to the observations $\mathcal{X}$, is useful in terms of selecting the model order, it is necessary to involve it to construct the information theoretic criterion. For example, $\mathcal{Z}$ can be some statistics as a function of the noise eigenvalues, such as

$$\log \frac{\left(\frac{1}{p-k}\sum_{i=k+1}^{p} l_i\right)^{p-k}}{\prod_{i=k+1}^{p} l_i} \quad \text{and} \quad \frac{\frac{1}{p-k}\sum_{i=k+1}^{p} l_i^2}{\left(\frac{1}{p-k}\sum_{i=k+1}^{p} l_i\right)^2} \tag{7.6}$$

which carry statistical information of the noise eigenvalues and are used in some references as the test statistics for determining the multiplicity of the noise eigenvalues. Herein, the first expression in (7.6) follows a scaled χ^2 distribution [18], whereas the second one follows a Gaussian distribution [90]. Both are independent of the noise

variance σ^2. As mentioned above, $\mathcal{Z}$ can be the signal and noise sample eigenvalues, or the statistics based on them, such as the ML noise variance estimator $\hat{\sigma}^2$ and the two statistics in (7.6), which results in different formulations of GBIC_1 in (7.5).

It follows from the preceding discussion that the conventional BIC is generalized in a Bayesian framework by taking into account the density of the sample eigenvalues or corresponding statistics, in addition to the density of the observations. In what follows, these two densities are compared intuitively. Let us recall the mechanism of the conventional BIC for discriminating the signal eigenvalues from the noise eigenvalues, that is, the density of the observations in (7.3). It depends explicitly on the sample eigenvalues and provides the information of their magnitude distinctions rather than of the statistical characteristics. For example, the first term of the last equality in (7.3) which is the ratio of the geometric mean and the arithmetic mean with respect to the smallest eigenvalues, is nothing but a statistic for checking the multiplicity of the eigenvalues, see the first one in (7.6). It is known that the eigenvalue magnitudes fluctuate dramatically in critical situations, e.g., small sample sizes. This kind of statistic, more precisely, the density of the observations, is very sensitive to the fluctuations of the eigenvalue magnitudes and not effective in discriminating the sample eigenvalues. Contrarily, the density of the sample eigenvalues or corresponding statistics provides the statistical characteristics of the sample eigenvalues, which discriminate the sample eigenvalues by checking the probabilities rather than the magnitudes and is more robust against fluctuations of the eigenvalue magnitudes. Therefore, the density of the sample eigenvalues or corresponding statistics used by the proposed GBIC_1 remedies the limitations of the conventional BIC.

7.1.3 Part II of GBIC

The first expression of GBIC in (7.5), namely GBIC_1, is derived by excluding the density of the sample eigenvectors from the formulation of BIC in (7.1). Note that the eigenvectors still play a role in GBIC_1, as in BIC, due to the dimensionality of the parameter vector $\boldsymbol{\Theta}^{(k)}$, i.e., n_k, contained in the penalty function. n_k is mainly contributed by the number of free scale parameters associated with the eigenvectors [31]. If we want to further reduce the influence of the eigenvectors, we can drop the density of the observations $f(\mathcal{X}|\hat{\boldsymbol{\Theta}}^{(k)})$ from GBIC_1. As a result, we have

$$\text{GBIC}_2(k) = -2\log f(\mathcal{Z}|\hat{\boldsymbol{\Theta}}_z^{(k)}) + \tilde{n}_k \log N \tag{7.7}$$

which is the second expression of GBIC. Herein, $\tilde{n}_k$ is substituted for n_k and depends on $\mathcal{Z}$. Note that $\tilde{n}_k$ is much smaller than n_k since the eigenvectors are excluded from the

parameter vector of the model. The typical choice for $\tilde{n}_k$ is $k+1$, which is the number of parameters, including the signal population eigenvalues and the noise variance.

In [53], it is stated that more parameters of the model are included, higher uncertainty is introduced in estimation of these parameters. Estimation of the eigenvectors introduces a large amount of uncertainty for a relatively minor gain in terms of information content. This is reflected by three facts. First, the parameters contributed by the eigenvectors are dominant in the parameter space of the model. Second, the sample eigenvectors do not appear in (7.3). Third, the free parameters associated with the sample eigenvectors have intractable distributions. Then, the penalty function (a measure of the uncertainty) may not be offset by the log-likelihood function (a measure of the information gain) [53], which results in unsatisfactory performance of BIC. GBIC_1 suffers from the same problem, although its log-likelihood function is refined in contrast to that of BIC. This motivates one to reduce the uncertainty introduced by the eigenvectors, or only incorporate the information independent of the eigenvectors. Thus, we devise GBIC_2 in (7.7) based on $\mathcal{Z}$ only, ignoring the observations $\mathcal{X}$. By doing this, the eigenvectors explicitly play no role in GBIC_2. It is worth noting that the special property of the parameter space in our problem renders GBIC_2 applicable.

From the viewpoint of the density of the observations $f(\mathcal{X}|\hat{\mathbf{\Theta}}^{(k)})$, it is necessary to include the density of the sample eigenvalues or corresponding statistics $f(\mathcal{Z}|\hat{\mathbf{\Theta}}_z^{(k)})$ for constructing the information theoretic criterion, since the combination of the two densities provides more information than the first one alone in most cases. However, it is not necessarily true from the viewpoint of the density $f(\mathcal{Z}|\hat{\mathbf{\Theta}}_z^{(k)})$, since incorporating $f(\mathcal{X}|\hat{\mathbf{\Theta}}^{(k)})$ amounts to introduction of high uncertainty or a large penalty function, especially when the sample size is large.

Thus far, two expressions for the GBIC rule are derived, that is, GBIC_1 in (7.5) and GBIC_2 in (7.7), by reducing the impact of the eigenvectors in the criterion in (7.1). The selection between these two expressions is detailed based on an example of GBIC in the next section.

7.2 An example of GBIC

Among the aforementioned choices for $\mathcal{Z}$, the most informative one is $l_1, \cdots, l_p$, i.e., all the sample eigenvalues. Other choices originate from there, e.g., the ML noise variance estimator $\hat{\sigma}^2$ and the two statistics in (7.6) are derived based on the noise sample eigenvalues. In principle, the joint density of all the sample eigenvalues contains the

richest statistical information compared to those of the statistics derived based on them. In the sequel, we will implement GBIC_1 in (7.5) with the following example

$$\begin{aligned}\mathrm{GBIC}_1(k) =& -2\left[\log f(\mathcal{X}|\hat{\mathbf{\Theta}}^{(k)}) + \log f(l_1,\cdots,l_p)\right] + n_k \log N \\ =& -2N(p-k)\log\left[\frac{\prod_{i=k+1}^{p} l_i^{1/(p-k)}}{\frac{1}{p-k}\sum_{i=k+1}^{p} l_i}\right] \\ & -2\log f(l_1,\cdots,l_p) + [k(2p-k)+1]\log N\end{aligned} \tag{7.8}$$

where the joint pdf of the sample eigenvalues $f(l_1,\cdots,l_p)$[2] is given in the next subsections. Correspondingly, GBIC_2 in (7.7) becomes

$$\mathrm{GBIC}_2(k) = -2\log f(l_1,\cdots,l_p) + (k+1)\log N \tag{7.9}$$

where $k+1$ is substituted for $\tilde{n}_k$ in (7.7).

In fact, the criterion in (7.9) was first proposed in [53]. The used pdf of the sample eigenvalues $f(l_1,\cdots,l_p)$ was derived based on classical multivariate statistical theory, which has a rather complicated form. In this thesis, a simplified version of the pdf in [53] is used, which is given in the next subsection.

7.2.1 Density of sample eigenvalues

For constructing GBIC, the density of the sample eigenvalues is of utmost importance. Here, it is assumed that the smallest signal sample eigenvalue is much larger than the largest noise sample eigenvalue so that a subspace swap is avoided, e.g., a crossover between the smallest signal sample eigenvalue and the largest noise sample eigenvalue. Following the statement that the sample covariance matrix scaled by the number of samples follows the complex Wishart distribution [132], the density of the sample eigenvalues is given based on the hypogeometric function for the general case in [123] and [133]. In [39], the density is simplified for the case of distinct signal population eigenvalues and large sample sizes, which is given as follows:

Theorem 7.2.1 *Denote by* $\hat{\mathbf{R}}$ *a sample covariance matrix formed from the* $p \times N$ *matrix of Gaussian observations, e.g., the observed data with* k *source signals in (2.1), whose columns are i.i.d. with mean* $\mathbf{0}$ *and covariance* $\mathbf{R}$*. The corresponding population eigenvalues and sample eigenvalues are given as* $\lambda_1 > \cdots > \lambda_q > \sigma^2 = \cdots = \sigma^2$ *and*

[2]For simplicity of notation, the unknown parameters of the pdf are neglected.

$l_1 > l_2 > \cdots > l_p$, respectively. Then, as $N \to \infty$, the signal sample eigenvalues are asymptotically independent and Gaussian distributed, that is,

$$f(l_i|\lambda_i) = \frac{\sqrt{N}}{\sqrt{2\pi}\lambda_i} \exp\left[-\frac{N(l_i-\lambda_i)^2}{2\lambda_i^2}\right], \quad i = 1, \ldots, k, \tag{7.10}$$

and the noise sample eigenvalues have the following joint density

$$\begin{aligned} f(l_{k+1}, \cdots, l_p|\sigma^2) = & \prod_{i=k+1,\ i<j}^{p} \left[\sqrt{N}\left(\frac{l_i - l_j}{\sigma^2}\right)\right]^2 \\ & \cdot \frac{1}{\Gamma(1)\cdots\Gamma(p-k)} \\ & \cdot \prod_{i=k+1}^{p} \frac{\sqrt{N}}{\sqrt{2\pi}\sigma^2} \exp\left[-\frac{N(l_i-\sigma^2)^2}{2(\sigma^2)^2}\right]. \end{aligned} \tag{7.11}$$

It follows readily from Theorem 7.2.1 that the signal sample eigenvalues are mutually independent and follow Gaussian densities but with different means and variances, and the noise sample eigenvalues are dependent and the joint density contains a term concerning the interaction among the noise sample eigenvalues. Also, it is implied that the signal sample eigenvalues are independent of the noise sample eigenvalues. Thus, the joint density of all the sample eigenvalues is a product of the two densities in (7.10) and (7.11). It is noteworthy that the signal population eigenvalues $\lambda_1, \cdots, \lambda_k$ are distinct, which is also the assumption made for the conventional BIC.

In addition to classical multivariate statistical theory, e.g., Theorem 7.2.1, random matrix theory [93,102,112,115] can also be used to characterize the sample eigenvalues. It is known that random matrix theory provides more accurate statistical information than classical multivariate statistical theory in the case of finite sample sizes. However, it is still not clear how to derive the joint density of the sample eigenvalues from random matrix theory, given the marginal densities. In principle, random matrix theory yields similar statistical characteristics as classical multivariate statistical theory when the sample size is relatively large and the array size is relatively small.

7.2.2 Implementation issues

From Theorem 7.2.1, it follows that the densities of the sample eigenvalues are conditioned on the signal population eigenvalues $\lambda_1, \cdots, \lambda_k$ and the noise variance σ^2, which are usually unknown in practice. If these densities are plugged in (7.8) or (7.9)

to construct the GBIC, it is necessary to let the densities be free of these unknown parameters. The most common way is to estimate them based on the sample eigenvalues $l_1, \cdots, l_p$ using ML estimation. However, we have only one copy of $l_1, \cdots, l_p$ so that it is very hard to obtain an accurate estimation result based on the densities in Theorem 7.2.1. Moreover, it is almost impossible to derive a closed-form expression for the ML estimators so that a time-consuming multi-dimensional search is needed. Consequently, the ML estimators based on the densities in Theorem 7.2.1 are not feasible in our case.

Alternatively, we can resort to the Bayesian approach. A prior density e.g., the Jeffreys prior or the uniform prior, is assumed for each of the unknown parameters, namely, $\lambda_1, \cdots, \lambda_k, \sigma^2$, as in [134]. Then the unknown parameters can be integrated out by taking the integration of the joint pdf of the sample eigenvalues with respect to the unknown parameters. In addition, tight intervals where the unknown parameters are located need to be found in order to let the integration with respect to the unknown parameters be precise enough. The drawback is high computational complexity due to the fact that numerical integration is required.

As we know, based on the multivariate Gaussian density of the observations, ML estimators with a simple formulation have been already obtained for all the parameters of the model, see (2.16). Then, if the estimators

$$\begin{aligned} \hat{\lambda}_i &= l_i, i = 1, \ldots, k \\ \hat{\sigma}^2 &= \frac{1}{p-k} \sum_{i=k+1}^{p} l_i \end{aligned} \tag{7.12}$$

are substituted for the unknown parameters $\lambda_1, \cdots, \lambda_k$ and σ^2 in the densities of the sample eigenvalues in Theorem 7.2.1, we can derive the joint pdf

$$\begin{aligned} f(l_1, \cdots, l_p) &= f(l_1, \cdots, l_k | \hat{\lambda}_1, \cdots, \hat{\lambda}_p) \cdot f(l_{k+1}, \cdots, l_p | \hat{\sigma}^2) \\ &= \left(\frac{N}{2\pi}\right)^{p/2} \prod_{i=1}^{k} \frac{1}{l_i} \cdot \prod_{i=k+1,\ i<j}^{p} \left[\sqrt{N}\left(\frac{l_i - l_j}{\hat{\sigma}^2}\right)\right]^2 \\ &\quad \cdot \frac{1}{\Gamma(1)\cdots\Gamma(p-k)} \cdot \prod_{i=k+1}^{p} \frac{1}{\hat{\sigma}^2} \exp\left[-\frac{N(l_i - \hat{\sigma}^2)^2}{2(\hat{\sigma}^2)^2}\right] \end{aligned} \tag{7.13}$$

which is independent of unknown parameters.

In Appendices 7.5.1 and 7.5.2, the performance of BIC, GBIC_1 and GBIC_2 is compared analytically for large sample sizes. More precisely, their underestimation probabilities are compared since the underestimation probability is dominating in most cases of

interest. It is shown that BIC has the highest underestimation probability, whereas GBIC_2 has the lowest. Since the estimators in (2.16) converge to the true values as the sample size increases, GBIC_2 in (7.9) using the pdf in (7.13) can achieve the asymptotic estimation bound mentioned in Appendix 7.5.3. In Appendix 7.5.3, we compare quantitatively the asymptotic estimation bounds of BIC and GBIC_2. As the sample size increases, the discrepancy of the two estimation bounds increases. When the sample size is relatively large, GBIC_2 has a relatively lower estimation bound than BIC, and the bound of GBIC_1 is a compromise between the preceding two.

It is known that GBIC_1 and GBIC_2 are derived asymptotically, as BIC. In principle, they can be applied for large sample sizes only. However, it is of great practical interest to see their performance for small sample sizes. When the sample size is small, the estimators in (2.16) are not accurate enough and usually deviate from the true values by far. The density in (7.13) yields inaccurate probabilities for the sample eigenvalues such that GBIC_2 in (7.9) does not have satisfactory performance in terms of estimating the number of sources. Simulations demonstrate that the density $f(\mathcal{X}|\hat{\boldsymbol{\Theta}}^{(k)})$ is a good supplement of the density $f(l_1, \cdots, l_p)$ in (7.13), and their combination yields better results. As a result, GBIC_1 in (7.8) outperforms GBIC_2 in (7.9) for small sample sizes.

To summarize this section, as two parts of GBIC, GBIC_1 and GBIC_2 are constructed using (7.13) (or equivalently (7.10), (7.11) and (7.12)). GBIC_1 outperforms GBIC_2 in the case of small sample sizes, whereas GBIC_2 yields better performance than GBIC_1 in the case of large sample sizes. Note that these two cases of sample sizes are defined by accuracy of the density $f(l_1, \cdots, l_p)$ in (7.13). In the regime of large sample sizes, the density $f(l_1, \cdots, l_p)$ in (7.13) is accurate enough, whereas in the regime of small sample sizes, it is not. This issue will be elaborated in our simulations in the next section.

7.3 Simulation results

A uniform linear array with omni-directional sensors and inter-sensors spacing of half the wavelength is employed for simulations. We consider the scenario that the complex circular Gaussian source signals impinging on the array is contaminated by white Gaussian noise. Simulation results are obtained based on 10,000 Monte Carlo trials. The numbers of samples, sensors and sources are denoted by N, p and q, respectively. The probability of correctly estimating q is denoted by P_c. "DOA" is short for the direction of arrival of a source. "SNR" is short for the signal-to-noise ratio. The conventional BIC is denoted by "BIC". The proposed generalized BIC using (7.8) and

(7.9) are denoted by "GBIC1" and "GBIC2", respectively. When the prior knowledge of the noise variance σ^2 is used, "GBIC2" is modified to "GBIC2_σ^2". Moreover, the test proposed in [94] is included for performance comparison, denoted by "KN". The false alarm rate of "KN" is set to 0.1%, as suggested in [94]. In the following, these four approaches are compared in different experimental settings:

- Small sample sizes:
  - Setting 1: Number of samples (see Fig. 7.1). $N \in [20, 60]$, $p = 15$, $q = 2$, DOAs = $0°$, $3°$, SNR = -2 dB.
  - Setting 2: SNR (see Fig. 7.2). $N = 50$, $p = 12$, $q = 3$, DOAs = $-4°$, $0°$, $4°$, SNR $\in [-2, 12]$ dB.
  - Setting 3: Angular resolution (see Fig. 7.3). $N = 60$, $p = 15$, $q = 2$, DOA1 = $0°$, DOA2 $\in [1°, 8°]$, SNR = -6 dB.
- Large sample sizes:
  - Setting 4: Number of samples (see Fig. 7.4). $N \in [1000, 5000]$, $p = 15$, $q = 4$, DOAs = $-6°$, $-2°$, $2°$, $6°$, SNR = -3 dB.
  - Setting 5: SNR (see Fig. 7.5). $N = 1000$, $p = 10$, $q = 3$, DOAs = $-3°$, $0°$, $3°$, SNR $\in [2, 10]$ dB.

In each setting, all the sources are assumed to have the same SNR.

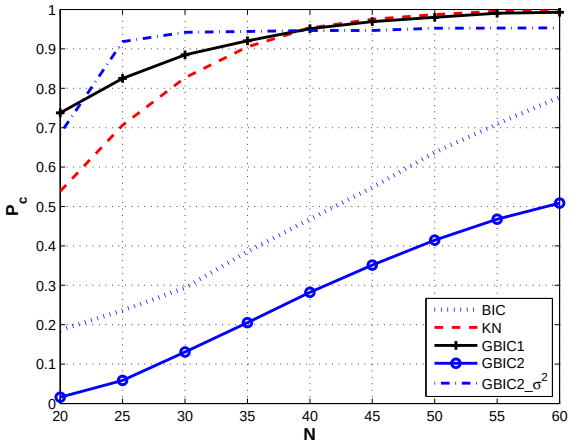

Figure 7.1. Probability of correct estimation P_c vs. sample size N with $p = 15$, $q = 2$.

Figs. 7.1–7.3 represent the case of small sample sizes. "GBIC2" yields poor performance, which is caused by inaccuracy of the estimators of unknown parameters in (2.16)

for small N, especially of the estimator of the noise variance $\hat{\sigma}^2$. Since the estimator $\hat{\sigma}^2$ is negatively biased, "GBIC2" overestimates q with a extremely high probability, as shown in Figs. 7.1–7.3. If the noise variance σ^2 can be estimated correctly, the performance of "GBIC2" can be improved. This is validated by good performance of "GBIC2_σ^2" using the prior knowledge of the noise variance σ^2. But "GBIC2_σ^2" still tends to overestimate q and has a very low convergence rate with respect to N and SNRs due to the fact that the signal population eigenvalues are simply estimated as the signal sample eigenvalues. Note that it is challenging to find accurate estimators for the signal population eigenvalues for small N.

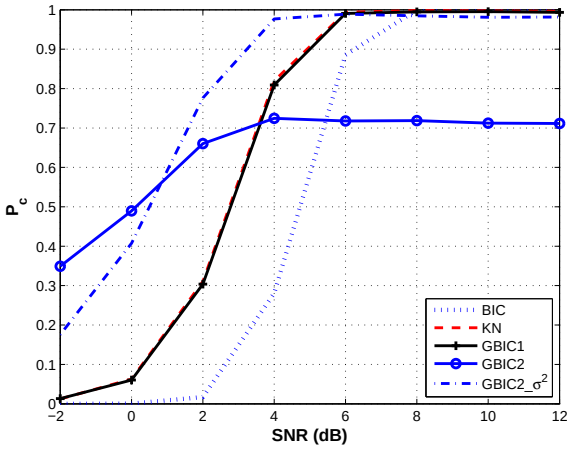

Figure 7.2. Probability of correct estimation P_c vs. SNR with $N = 50$, $p = 12$, $q = 3$.

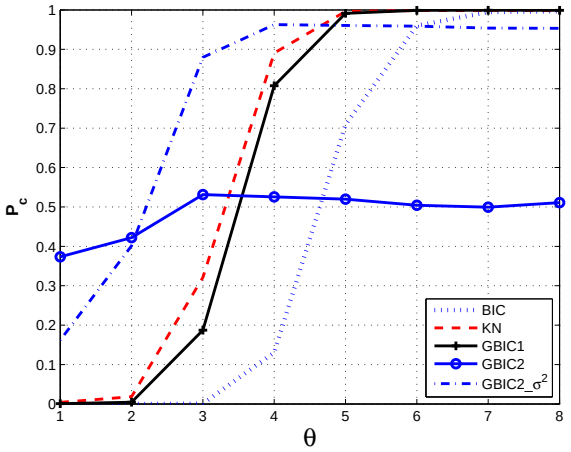

Figure 7.3. Probability of correct estimation P_c vs. DOA of the second source θ with $N = 60$, $p = 15$, $q = 2$.

The impact of the inaccurate estimators in (2.16) on the two preceding approaches can be alleviated by integrating the densities of the sample eigenvalues and the observa-

tions, as done in "GBIC1". In other words, the density of the observations somewhat compensates for the inaccurate estimators in (2.16) used in the density of the sample eigenvalues in (7.13). The two densities complement each other, which enhances accuracy of the log-likelihood function of the criterion. This can be readily observed from the performance of "GBIC1". "GBIC1" significantly outperforms "BIC" and "GBIC2" and even converges much faster with N and SNRs than "GBIC2_σ^2". It is known that "BIC" based on only the density of the observations is apt to underestimate q for small N or low SNRs whereas "GBIC2" yields overestimation of q. By adopting both densities, "GBIC1" overcomes the weakness of "BIC" and "GBIC2" and yields a much higher probability of correctly estimating q.

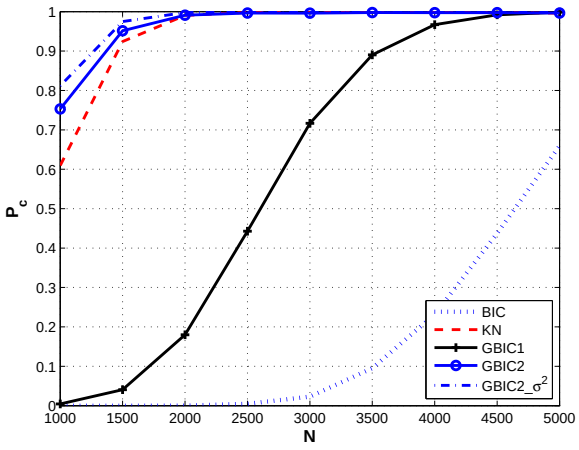

Figure 7.4. Probability of correct estimation P_c vs. sample size N with $p = 15$, $q = 4$.

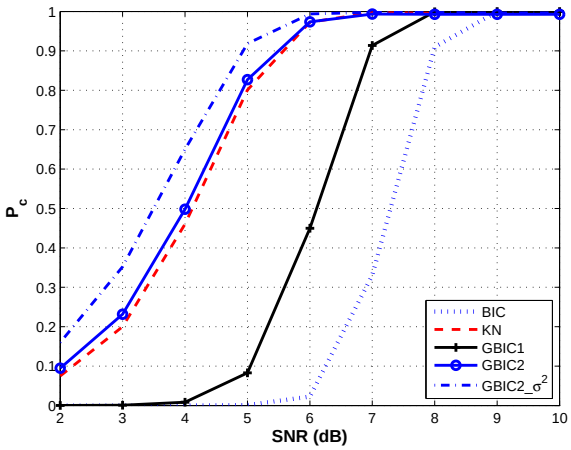

Figure 7.5. Probability of correct estimation P_c vs. SNR with $N = 1000$, $p = 10$, $q = 3$.

Figs. 7.4–7.5 represent the case of large sample sizes. As the sample size N increases, the involved density in (7.13) becomes more accurate, since the density of the sample eigenvalues is an asymptotic density and the estimators in (2.16) asymptotically converge to the true values. So "GBIC2" performs extraordinarily well and increasingly close to "GBIC2_σ^2". In this case, the density of the observation is not capable of enhancing accuracy of the density of the eigenvalues. As shown in Appendix 7.5.3, the performance gap between "BIC" and "GBIC2" increases as N increases. The combination of the two densities leads to the performance degradation of "GBIC1", compared to "GBIC2", although there is significant performance improvement over "BIC".

Generally, "GBIC1" is slightly inferior to "KN" for small sample sizes, whereas "GBIC2" is slightly superior to "KN" for large small sizes. "KN" keeps good performance for the case when the sample size is very small or close to the array size (i.e., $N \approx p$), whereas our proposed approaches do not. Also, "GBIC1" and "GBIC2" suffer performance degradation, as BIC, if the noise is non-Gaussian or non-uniform [66][72]. The important point to bear in mind is that the performance of "GBIC1" and "GBIC2" highly depends on accuracy of the density of the sample eigenvalues in Theorem 7.2.1. Any derivation of the assumptions made in Theorem 7.2.1, leading to inaccuracy of the density of the sample eigenvalues, may reduce the performance of "GBIC1" and "GBIC2".

In Section 7.2, "small sample sizes" and "large sample sizes" are the two situations which imply if the density in (7.13) is accurate or not. Here, these two situations are discriminated based on the performance of "GBIC2" since its performance is closely related to accuracy of the density in (7.13). In order to clarify the quantitative difference between "small sample sizes" and "large sample sizes", an example is given when the simulation is conducted in Setting 1 except that the sample size ranges from 100 to 700. We summarize the results of the aforementioned approaches in Table 7.1. Herein, both "BIC" and "GBIC1" yield 100% correction estimation. Both "KN" and "GBIC2_σ^2" have slight overestimation probability. "GBIC2" has less than 3% overestimation probability only when $N > 500$. In this example, $N > 500$ can be roughly considered as "large sample sizes" since "GBIC2" yields reliable performance. On the contrary, $N < 200$ can be regarded as "small sample sizes" since "GBIC2" has an overestimation probability larger than 10%. The regime $N \in [200, 500]$ corresponds to "moderate sample sizes", where the selection between "GBIC1" and "GBIC2" should be more cautious. Note that the quantitative definition of "small sample sizes" and "large sample sizes" is not the same if different densities of the sample eigenvalues are used. Some densities may be less sensitive to the sample size than the density in (7.13).

Table 7.1. Probability of correct estimation P_c vs. sample size N in Setting 1.

N	100	200	300	400	500	600	700
BIC; GBIC1 (%)	100	100	100	100	100	100	100
KN	99.84	99.70	99.74	99.71	99.78	99.76	99.77
GBIC2	72.48	89.04	94.09	95.46	97.01	97.31	97.69
GBIC2_σ^2	96.32	97.76	98.35	98.60	99.01	99.11	99.04

7.4 Conclusions

The problem of source enumeration in array processing has been addressed by proposing a generalized Bayesian information criterion (GBIC) rule, which suggests to incorporate the density of the sample eigenvalues or corresponding statistics in constructing the log-likelihood function of the criterion. Due to the relationship between the densities of the sample eigenvalues and the observations, we have presented an example of GBIC, which consists of two expressions, namely, GBIC_1 and GBIC_2. The former includes both densities and the latter includes only the density of the sample eigenvalues. It has been shown that GBIC_1 has a lower underestimation probability than BIC. GBIC_2 outperforms BIC and GBIC_1 only when the sample size is relatively large. The proposed criterion validates that the density of the sample eigenvalues is a supplement to information theoretic criteria.

The proposed GBIC can be applied to the parametric model as well. The parametric BIC performs significantly better than the non-parametric one. If we incorporate the density of the sample eigenvalues into the parametric BIC, the performance can be further improved. So the parametric GBIC can be a topic of future work.

7.5 Appendix

7.5.1 Performance comparison between BIC and GBIC_1

In this appendix, we compare the performance of BIC in (2.19) and GBIC_1 in (7.8) for large sample sizes.

Generally, GBIC_1 and GBIC_2 are consistent due to the fact that they belong to the BIC framework. Then, for these two criteria, the probability of incorrectly estimating the number of sources $P_e = P(\hat{q} \neq q)$, assuming that q sources exist, converges to

zero as the sample size increases to infinity. In other words, both the underestimation probability $P(\hat{q} < q)$ and the overestimation probability $P(\hat{q} > q)$ converge to zero. Moreover, it is empirically observed in the literature (e.g., see [135]) that for most of the existing information theoretic criteria in the regime of large sample sizes, including AIC, the underestimation probability gradually dominates P_e when the sample size and the SNR decrease[3]. In such cases, the overestimation probability is negligible as compared to the underestimation probability. This is also true for the proposed GBIC_1 and GBIC_2, as shown in our simulations. Simply put, in most cases of interest, the performance of distinct criteria can be measured by the underestimation probability. The lower underestimation probability the criteria have, the higher is the performance they achieve.

The performance of the conventional BIC has been analyzed in [35–44], by calculating the probability P_e. It is established in [38] that for a wide range of SNRs and sample sizes, the probability P_e is mainly dominated by the probability to estimate the number of sources as $q-1$, i.e., $P_e \simeq P(\hat{q} = q-1) = P(\text{BIC}(q) - \text{BIC}(q-1) > 0)$. In other words, BIC tends to underestimate q by one, namely, $\hat{q} = q-1$.

By incorporating the density of the sample eigenvalues or corresponding statistics, GBIC_1 has much lower underestimation probability than BIC. Simply put, in some cases when BIC chooses $\hat{q} = q-1$ due to $\text{BIC}(q) > \text{BIC}(q-1)$, that is,

$$\log f(\mathcal{X}|\hat{\boldsymbol{\Theta}}^{(q)}) - \log f(\mathcal{X}|\hat{\boldsymbol{\Theta}}^{(q-1)}) < \Delta = (n_q - n_{q-1})\log N/2, \tag{7.14}$$

GBIC_1 chooses $\hat{q} = q$ due to $\text{GBIC}_1(q) < \text{GBIC}_1(q-1)$, that is,

$$\Big[\log f(\mathcal{X}|\hat{\boldsymbol{\Theta}}^{(q)}) - \log f(\mathcal{X}|\hat{\boldsymbol{\Theta}}^{(q-1)})\Big] + \Big[\log f(\mathcal{Z}|\hat{\boldsymbol{\Theta}}_z^{(q)}) - \log f(\mathcal{Z}|\hat{\boldsymbol{\Theta}}_z^{(q-1)})\Big] > \Delta = (n_q - n_{q-1})\log N/2 \tag{7.15}$$

where $\mathcal{Z} = \{l_1, \cdots, l_p\}$. In (7.14), the gain of the probability of the observations under the true model over the probability under the underestimated model, is smaller than the value Δ which is related to the difference of the two penalty functions. It is the reason why BIC underestimates q. In (7.15), there is an additional positive term on the left-hand side, i.e., $\log f(\mathcal{Z}|\hat{\boldsymbol{\Theta}}_z^{(q)}) - \log f(\mathcal{Z}|\hat{\boldsymbol{\Theta}}_z^{(q-1)})$, which is the gain of the probability of the sample eigenvalues or the corresponding statistics under the true model over the probability under the underestimated model. With the presence of this positive term, the inequality in (7.15) holds with an increasing probability, which indicates that the underestimation probability of GBIC_1 is reduced.

[3]Note that this observation may not be true when these criteria are used in other model selection problems, such as model order selection for a polynomial signal.

7.5.2 Performance comparison between GBIC_1 and GBIC_2

In this appendix, we compare the performance of GBIC_1 in (7.8) and GBIC_2 in (7.9) for large sample sizes.

It is obvious that GBIC_1 in (7.8) can be decomposed as

$$\begin{aligned}\mathrm{GBIC}_1(k) = &\left\{-2\log f(\mathcal{X}|\hat{\mathbf{\Theta}}^{(k)}) + (n_k - k - 1)\log N\right\} \\ &+ \{-2\log f(l_1, \cdots, l_p) + (k+1)\log N\}\end{aligned} \tag{7.16}$$

where the function in the first braces is the modified BIC proposed in [46] and the function in the second braces is exactly GBIC_2 in (7.9). It is implied that GBIC_1 is a compromise of the above two criteria. It is known that the modified BIC in [46] tends to underestimate the number of sources but with a lower probability than the conventional BIC since its penalty function is reduced from n_k to $n_k - k - 1$. In the case of large sample sizes, it is shown that the density $f(l_1, \cdots, l_p)$ is accurate enough and GBIC_2 yields a much lower underestimation probability than BIC. Thus, as a compromise of GBIC_2 and the modified BIC in [46], GBIC_1 yields a probability of underestimation which is higher than that of GBIC_2 and lower than that of the modified BIC in [46], which also confirms that GBIC_1 outperforms the conventional BIC and is defeated by GBIC_2.

The performance difference between GBIC_1 and GBIC_2 can also be explained intuitively using the concept of information theoretic criteria. In the case of large sample sizes, GBIC_2 is capable of estimating the number of sources correctly, which indicates that its penalty function $(k+1)\log N$ is proper for the log-likelihood function $\log f(l_1, \cdots, l_p)$, whereas GBIC_1 suffers performance degradation due to the fact that its penalty function $n_k \log N$ is too high for $\log f(l_1, \cdots, l_p)$.

7.5.3 Asymptotic estimation bounds of BIC and GBIC_2

In this appendix, the asymptotic estimation bounds of BIC in (2.19) and GBIC_2 in (7.9) are approximately given by the resolvable contrast between the smallest signal sample eigenvalue and the noise variance estimator.

According to the performance of the conventional BIC, its most probable estimation results are either $q-1$ or q. It is shown in [39] and [40] that to ensure $\hat{q} = q$, namely,

$\mathrm{BIC}(q) - \mathrm{BIC}(q-1) < 0$, the ratio between the smallest signal sample eigenvalue l_q and the ML estimator of the noise variance $\hat{\sigma}^2 = \frac{1}{p-q}\sum_{i=q+1}^{p} l_i$, that is,

$$\rho = \frac{l_q}{\hat{\sigma}^2}, \tag{7.17}$$

should be large enough, e.g.,

$$\rho > T_1 \tag{7.18}$$

where T_1 denotes the estimation bound of the conventional BIC. Since there does not exist a closed-form expression for T_1, two different approximations are given in [39] and [40], respectively. For simplicity of the following inference, we use the one in [39] with the expression

$$T_1 \simeq 1 + \sqrt{\frac{2(p-q+1)[2(p-q)+1]}{p-q}} \cdot \sqrt{\frac{\log N}{N}}, \tag{7.19}$$

from which it can be seen that as N increases, the threshold T_1 of order $O(\sqrt{\frac{\log N}{N}})$ decreases. It is worth noting that the true value for the threshold T_1 is even larger than the approximation in (7.19).

For analyzing the asymptotic performance of GBIC_2, we resort to the distribution from random matrix theory, since random matrix theory has emerged recently as the most popular tool to model the density of the sample eigenvalues. Before an asymptotic estimation bound for random matrix theory is given, we recall the so-called phase transition phenomenon for the signal eigenvalues in Theorem 4.1.2 in Chapter 4. Asymptotically, if the signal population eigenvalue λ_i is smaller than or equal to the value $\sigma^2(1+\sqrt{c})$, its sample eigenvalue l_i converges to the limit $\sigma^2(1+\sqrt{c})^2$ which is exactly the upper bound of the noise sample eigenvalues [93][102]. In this case, the signal sample eigenvalues can not be discriminated from the noise sample eigenvalues. If λ_i is larger than the value $\sigma^2(1+\sqrt{c})$, l_i is pulled up to the limit $\lambda_i(1+\frac{c\sigma^2}{\lambda_i-\sigma^2})$ which is larger than $\sigma^2(1+\sqrt{c})^2$. It is implied that the signal sample eigenvalue l_i can be detected only if the corresponding population eigenvalue λ_i is larger than $\sigma^2(1+\sqrt{c})$. Hence, the value $\sigma^2(1+\sqrt{c})$ can be considered as the non-parametric asymptotic detection limit from the viewpoint of random matrix theory. Following this statement, we obtain

$$\begin{aligned} \rho = \frac{l_q}{\hat{\sigma}^2} &\simeq \frac{l_q}{\sigma^2} \\ &> T_2 = \frac{\lambda_q}{\sigma^2} = 1+\sqrt{c} = 1+\sqrt{\frac{p}{N}} \end{aligned} \tag{7.20}$$

where $\hat{\sigma}^2$ and l_q converge to σ^2 and λ_q, respectively, as N increases. GBIC2 in (7.9) may achieve the estimation bound T_2 asymptotically (or more precisely, of order $O(\sqrt{\frac{1}{N}})$) and GBIC_1 in (7.8) has an estimation bound which is located between T_2 and T_1.

Since both estimation bounds are given, it can be readily derived that T_1 is much larger than T_2 from the following inequality

$$\frac{T_1 - 1}{T_2 - 1} = \sqrt{\frac{4(p-q) + 6 + \frac{2}{p-q}}{p} \log N} > 2\sqrt{\log N}, \tag{7.21}$$

assuming $p \gg q$. When N increases, the gap between these two bounds becomes larger and the performance gain of GBIC_2 over BIC gets increasingly remarkable. Note that the lower estimation bound results in the lower underestimation probability, which coincides with the result in Appendix 7.5.2.

Chapter 8

Conclusions

This thesis has thoroughly investigated the problem of source enumeration in sensor array processing and proposed four novel approaches by incorporating the distinct distributions of the sample eigenvalues. This model order selection problem is addressed in a framework of hypothesis testing or information theoretic criteria.

The bootstrap-based test (BBT) is improved in order to adapt itself to the case of impulsive noise and very small sample sizes. In the presence of impulsive noise, the concept of robust statistics is integrated into the bootstrap techniques. Two robust estimators, i.e., the minimum covariance determinant (MCD) estimator and the MM-estimator, are used to detect and remove outliers or to suppress the distorting impact of outliers. The resulting tests which are robust against the outliers caused by impulsive noise, are referred to as the robust bootstrap-based tests (RBBTs). Moreover, when the sample size is very small, a new statistic is found based on the exponential profile of the sample eigenvalues, which is more accurate than the statistic checking the equality of the eigenvalues in [21]. Then the modified bootstrap-based test (MBBT) yields better performance. Note that the bootstrap-based test is analytically simple but computationally complex, and it can be widely used in non-Gaussian data case since the minimal distributional assumption is made.

The two-step test procedure has been developed for the case of extremely small sample sizes, including the case when the sample size is smaller than the the array size. An initial value for the number of sources is obtained by a simple thresholding approach in the first-step test, which is refined by a likelihood ratio test in the second-step test. The propose two-step test do not need to set the false alarm rate. The test is fairly simple to implement, whereas it yields superior performance. It is worth noting that the distribution of the sample eigenvalues involved in the test is derived from random matrix theory since it can capture this asymptotic distribution accurately and provide a good approximation even for a finite sample size.

The performance of the conventional Bayesian information criterion (BIC) has been analytically investigated. The results from random matrix theory and Lawley's theory are used to derive the distribution of the statistics related to the sample eigenvalues and then proposed two new procedures to evaluate theoretically the performance of the BIC, more precisely, to calculate the probability of underestimation of the number of

sources. The theoretical results given by the proposed procedures are in good agreement with the empirical results based on simulations, even when the sample size is modest.

The proposed performance analysis procedures provide insight into the behavior of the BIC, that is, the BIC tends to underestimate the number of sources by one in most cases. In order to remedy the limitation of the BIC, an extra parameter is introduced into the BIC formulation based on the concept of thresholding. It leads to the flexible detection criterion (FDC). By carefully choosing this parameter, the new criterion is capable of substantially reducing the probability of underestimation such that it yields superior performance over the BIC, which is validated by theoretical analysis and simulations.

The generalized Bayesian information criterion (GBIC) has been proposed by including the distribution of the sample eigenvalues as one part of its log-likelihood function. Such a distribution contains extra information and complements the distribution of the observations in constructing the GBIC. As a result, two different expressions for the GBIC are suggested, that is, only the distribution of the sample eigenvalues is involved for large sample sizes and both distributions are involved for small sample sizes. The performance superiority of the GBIC over the BIC can be observed numerically and theoretically. Note that the joint distribution of the sample eigenvalues is derived from classical multivariate statistical theory.

Note that the BIC is a special case of the FDC or the GBIC. The latter two are consistent and of comparable computational cost to the former.

List of Figures

List of Tables

List of Acronyms

AIC	Akaike's information criterion
BBT	bootstrap-based test
BIC	Bayesian information criterion
cdf	cumulative distribution function
edf	empirical distribution function
EFT	exponential fitting test
EDC	efficient detection criterion
FDC	flexible detection criterion
FRB	fast and robust bootstrap
DOA	direction-of-arrival
GBIC	generalized Bayesian information criterion
GLRT	generalized likelihood ratio test
i.i.d.	independent and identically distributed
ITC	information theoretic criterion
LLRT	log-likelihood ratio test
LRT	likelihood ratio test
MBBT	Modified bootstrap-based test
MCD	minimum covariance determinant
MDL	minimum description length
MHT	multiple hypotheses test
ML	maximum likelihood
MOS	model order selection
MST	multivariate statistical theory
pdf	probability density function

RBBT	robust bootstrap-based test
RMT	random matrix theory
SINR	signal-to-interference-plus-noise ratio
SNR	signal-to-noise ratio
SRB	Holm's sequentially rejective Bonferroni procedure

List of Symbols

$\mathbf{a}(\theta_j)$	array steering vector
$\mathbf{A}$	array steering matrix
$\mathbf{c}_i$	sample eigenvector
$f(\mathcal{X}\|\boldsymbol{\Theta}^{(k)})$	pdf of observations
$f(\mathcal{X}\|\hat{\boldsymbol{\Theta}}^{(k)})$	pdf of observations with ML estimator of the parameter vector
H	null hypothesis
$\mathbf{I}$	identity matrix
k	candidate for the number of sources
K	alternative hypothesis
l_i	the ith sample eigenvalues
$\boldsymbol{n}$	column vector of the noise
N	Number of samples (or snapshots, observations)
$\boldsymbol{N}$	matrix of noise
n_k	degree of freedom of the space spanned by $\boldsymbol{\Theta}^{(k)}$
$\mathbf{n}(t)$	noise column vector at time instant t
p	Number of sensors
q	Number of sources
$\boldsymbol{R}$	covariance matrix of observations
$\hat{\boldsymbol{R}}$	sample covariance matrix of observations
$\boldsymbol{R}_s$	covariance matrix of source signals
$\boldsymbol{s}$	column vector of source signals
$\boldsymbol{S}$	matrix of source signals
$\mathbf{v}_i$	population eigenvector
$\boldsymbol{x}$	column vector of the observed data
$\mathbf{X}$	matrix of observations
$\mathcal{X}$	set of observations
θ_j	DOA of the jth source
λ_i	the ith population eigenvalue
σ^2	noise variance
$\boldsymbol{\theta}$	column vector of the DOAs of all the sources
$\boldsymbol{\Theta}^{(k)}$	parameter vector of the pdf of observations
$\hat{\boldsymbol{\Theta}}^{(k)}$	ML estimator of the parameter vector of the pdf of observations

Bibliography

[1] D.K. Barton, *Modern Radar: Analysis, Evaluation, and System Design*, New York: Wiley, 1965.

[2] D.K. Barton and H.R. Ward, *Handbook of Radar Measurement*, New Jersey: Prentice Hall, 1969.

[3] M.I. Skolnik, *Radar Handbook*, New York: McGraw-Hill, 1970.

[4] M.I. Skolnik, *Introduction to Radar Systems*, New York: McGraw-Hill, 1980.

[5] J.D. Kraus, *Radio Astronomy*, New York: McGraw-Hill, 1966.

[6] J.H. Winters, "Smart antennas for wireless systems," *IEEE Personal Commun.*, vol. 1, pp. 22–27, Feb. 1998.

[7] S. Haykin, *Array Signal Processing*, Englewood Cliffs, NJ: Prentice Hall, 1985.

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

[9] J.F. Böhme, "Array processing," in *Advances in Spectrum Estimation and Array Processing*, S. Haykin, Ed. 1991, pp. 1–63, Englewood Cliffs, NJ: Prentice Hall.

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

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

[12] R. Roy, A. Paulraj, and T. Kailath, "ESPRIT – A subspace rotation approach to estimation of parameters of cisoids in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 5, pp. 1340–1342, Oct. 1986.

[13] S.M. Kay, *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory*, Prentice Hall, 1998.

[14] M.S. Bartlett, "A note on the multiplying factors for various χ^2 approximations," *J. R. Stat. Soc., Ser. B*, vol. 16, no. 2, pp. 296–298, 1954.

[15] D.N. Lawley, "Tests of significance for the latent roots of covariance and correlation matrices," *Biometrika*, vol. 43, pp. 128–136, 1956.

[16] S.N. Roy, "On a heuristic method of test construction and its use in multivariate analysis," *Ann. Math. Stat.*, vol. 24, no. 2, pp. 220–238, 1953.

[17] W. Chen, K.M. Wong, and J.P. Reilly, "Detection of the number of signals: A predicted eigen-threshold approach," *IEEE Trans. Signal Process.*, vol. 39, no. 5, pp. 1088–1098, 1991.

[18] D.B. Williams and D.H. Johnson, "Using the sphericity test for source detection with narrow-band passive arrays," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 11, pp. 2008–2014, Nov. 1990.

[19] A.A. Shah and D.W. Tufts, "Determination of the dimension of a signal subspace from short data records," *IEEE Trans. Signal Process.*, vol. 42, no. 9, pp. 2531–2535, Sep. 1994.

[20] A. Quinlan, J. Barbot, P. Larzabal, and M.Haardt, "Model order selection for short data: An exponential fitting test (EFT)," *EURASIP J. Adv. Signal Process.*, vol. 2007, 2007.

[21] R.F. Brcich, A.M. Zoubir, and P. Pelin, "Detection of sources using bootstrap techniques," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 206–215, Aug. 2002.

[22] P.-J. Chung, J.F. Bohme, C.F. Mecklenbrauker, and A.O. Hero, "Detection of the number of signals using the Benjamini-Hochberg procedure," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2497–2508, Jun. 2007.

[23] J.P.C.L. da Costa, *Parameter Estimation Techniques for Multi-Dimensional Array Signal Processing*, Ph.D. thesis, TU Ilmenau, 2010.

[24] B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*, Chapman & Hall, 1993.

[25] A.M. Zoubir, "Bootstrap methods for model selection," *AEÜ Int. J. Electron. Commun.*, vol. 53, no. 6, pp. 386–392, Dec. 1999.

[26] A.M. Zoubir and D.R. Iskander, *Bootstrap Techniques for Signal Processing*, Cambridge, U.K.: Cambridge Univ. Press, 2004.

[27] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, Jul. 2004.

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

[29] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.

[30] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, 1978.

[31] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 387–392, Apr. 1985.

[32] J. Rissanen, "MDL denoising," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2537–2543, Nov. 2000.

[33] J. Rissanen, *Information and Complexity in Statistical Modeling*, ser. Information Science and Statistics, 1st ed. New York: Springer, 2007.

[34] D.F. Schmidt and E. Makalic, "The consistency of MDL for linear regression models with increasing signal-to-noise ratio," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1508–1510, Mar. 2012.

[35] H. Wang and M. Kaveh, "On the performance of signal-subspace processing–Part I: Narrow-band systems," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 5, pp. 1201–1209, Oct. 1986.

[36] L. Zhao, P. Krishnaiah, and Z. Bai, "Remarks on certain criteria for detection of number of signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 2, pp. 129–132, Feb. 1987.

[37] 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 Process.*, vol. 35, no. 9, pp. 1350–1352, Sep. 1987.

[38] Q.-T. Zhang, K.M. Wong, P.C. Yip, and J.P. Reilly, "Statistical analysis of the performance of information theoretic criteria in the detection of the number of signals in array processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 10, pp. 1557–1567, Oct. 1989.

[39] W. Xu and M. Kaveh, "Analysis of the performance and sensitivity of eigendecomposition-based detectors," *IEEE Trans. Signal Process.*, vol. 43, no. 6, pp. 1413–1426, Jun. 1995.

[40] A.P. Liavas and P.A. Regalia, "On the behavior of information theoretic criteria for model order selection," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1689–1695, Aug. 2001.

[41] E. Fishler, M. Grosmann, and H. Messer, "Detection of signals by information theoretic criteria: general asymptotic performance analysis," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1027–1036, May 2002.

[42] F. Haddadi, M. Malek-Mohammadi, M.M. Nayebi, and M.R. Aref, "Statistical performance analysis of MDL source enumeration in array processing," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 452–457, Jan. 2010.

[43] B. Nadler, "Nonparametric detection of signals by information theoretic criteria: Performance analysis and an improved estimator," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2746–2756, May 2010.

[44] J.P. Delmas and Y. Meurisse, "Performance analysis of MDL criterion for the detection of noncircular or/and nonGaussian components," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'11)*, May 2011, pp. 4188–4191.

[45] L.C. Zhao, P.R. Krishnaiah, and Z.D. Bai, "On detection of the number of signals in presence of white noise," *J. Multivar. Anal.*, vol. 20, no. 1, pp. 1–25, 1986.

[46] D.B. Williams, "Counting the degrees of freedom when using AIC and MDL to detect signals," *IEEE Trans. Signal Process.*, vol. 42, no. 11, pp. 3282–3284, Nov. 1994.

[47] P. Chen, T.-J. Wu, and J. Yang, "A comparative study of model selection criteria for the number of signals," *IET Radar Sonar Navig.*, vol. 2, no. 3, pp. 180–188, Jun. 2008.

[48] Y. Yin and P.R. Krishnaiah, "On some nonparametric methods for detection of the number of signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 11, pp. 1533–1538, Nov. 1987.

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

[50] M. Wax, "Detection and localization of multiple sources via the stochastic signals model," *IEEE Trans. Signal Process.*, vol. 39, no. 11, pp. 2450–2456, Nov. 1991.

[51] S. Valaee and P. Kabal, "An information theoretic approach to source enumeration in array signal processing," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1171–1178, May 2004.

[52] Q. Wu and D.R. Fuhrmann, "A parametric method for determining the number of signals in narrow-band direction finding," *IEEE Trans. Signal Process.*, vol. 39, no. 8, pp. 1848–1857, Aug. 1991.

[53] K.M. Wong, Q.-T. Zhang, J.P. Reilly, and P.C. Yip, "On information theoretic criteria for determining the number of signals in high resolution array processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 11, pp. 1959–1971, Nov. 1990.

[54] E. Fishler and H. Messer, "Order statistics approach for determining the number of sources using an array of sensors," *IEEE Signal Process. Lett.*, vol. 6, no. 7, pp. 179–182, Jul. 1999.

[55] E. Fishler and H. Messer, "On the use of order statistics for improved detection of signals by the MDL criterion," *IEEE Trans. Signal Process.*, vol. 48, no. 8, pp. 2242–2247, Aug. 2000.

[56] 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. Multivar. Anal.*, vol. 20, no. 1, pp. 26–49, 1986.

[57] Q.T. Zhang and K.M. Wong, "Information theoretic criteria for the determination of the number of signals in spatially correlated noise," *IEEE Trans. Signal Process.*, vol. 41, no. 4, pp. 1652–1663, Apr. 1993.

[58] M. Wax, J. Sheinvald, and A.J. Weiss, "Detection and localization in colored noise via generalized least squares," *IEEE Trans. Signal Process.*, vol. 44, no. 7, pp. 1734–1743, Jul. 1996.

[59] P. Stoica and M. Cedervall, "Detection tests for array processing in unknown correlated noise fields," *IEEE Trans. Signal Process.*, vol. 45, no. 9, pp. 2351–2362, Sep. 1997.

[60] E. Fishler and H.V. Poor, "Estimation of the number of sources in unbalanced arrays via information theoretic criteria," *IEEE Trans. Signal Process.*, vol. 53, no. 9, pp. 3543–3553, Sep. 2005.

[61] Z.-D. Bai, P.R. Krishnaiah, and L.-C. Zhao, "On rates of convergence of efficient detection criteria in signal processing with white noise," *IEEE Trans. Inf. Theory*, vol. 35, no. 2, pp. 380–388, Mar. 1989.

[62] H.A. David and H.N. Nagaraja, *Order Statistics*, 3rd ed. New Jersey: John Wiley & Sons, 2003.

[63] J.H. Cozzens and M.J. Sousa, "Source enumeration in a correlated signal environment," *IEEE Trans. Signal Process.*, vol. 42, no. 2, pp. 304–317, Feb. 1994.

[64] G. Xu, R.H. Roy III, and T. Kailath, "Detection of number of sources via exploitation of centro-symmetry property," *IEEE Trans. Signal Process.*, vol. 42, no. 1, pp. 102 –112, Jan. 1994.

[65] C. Ma and C. Teng, "Detection of coherent signals using weighted subspace smoothing," *IEEE Trans. Antennas Propag.*, vol. 44, no. 2, pp. 179–187, Feb. 1996.

[66] H. Wu, J. Yang, and F. Chen, "Source number estimators using transformed Gerschgorin radii," *IEEE Trans. Signal Process.*, vol. 43, no. 6, pp. 1325–1333, Jun. 1995.

[67] R. Badeau, B. David, and G. Richard, "A new perturbation analysis for signal enumeration in rotational invariance techniques," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 450–458, Feb. 2006.

[68] J.-M. Papy, L. De Lathauwer, and S. Van Huffel, "A shift invariance-based order-selection technique for exponential data modelling," *IEEE Signal Process. Lett.*, vol. 14, no. 7, pp. 473–476, Jul. 2007.

[69] S. Aouada, *Source Detection and Parameter Estimation in Array Processing in the Presence of Nonuniform Noise*, Ph.D. thesis, TU Darmstadt, 2006.

[70] L. Huang and S. Wu, "Low-complexity MDL method for accurate source enumeration," *IEEE Signal Process. Lett.*, vol. 14, no. 9, pp. 581–584, Spt. 2007.

[71] L. Huang, S. Wu, and X. Li, "Reduced-rank MDL method for source enumeration in high-resolution array processing," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5658–5667, Dec. 2007.

[72] L. Huang, T. Long, and S. Wu, "Source enumeration for high-resolution array processing using improved Gerschgorin radii without eigendecomposition," *IEEE Trans. Signal Process.*, vol. 56, no. 12, pp. 5916–5925, Dec. 2008.

[73] L. Huang, T. Long, E. Mao, and H.C. So, "Mmse-based mdl method for robust estimation of number of sources without eigendecomposition," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 4135–4142, Oct. 2009.

[74] J. Xin, N. Zheng, and A. Sano, "Simple and efficient nonparametric method for estimating the number of signals without eigendecomposition," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1405–1420, Apr. 2007.

[75] Y. Hochberg and A. Tamhane, *Multiple Comparison Procedures*, New York: Wiley, 1987.

[76] S. Holm, "A simple sequentially rejective multiple test procedure," *Scund. J. Stat.*, vol. 6, pp. 65–70, 1979.

[77] A.M. Zoubir and D.R. Iskander, "Bootstrap methods and applications," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 10–19, Jul. 2007.

[78] A.M. Zoubir and J.F. Bohme, "Bootstrap multiple tests applied to sensor location," *IEEE Trans. Signal Process.*, vol. 43, no. 6, pp. 1386–1396, Jun. 1995.

[79] T.W. Anderson, "Asymptotic theory for principal component analysis," *Ann. Math. Stat.*, vol. 34, no. 1, pp. 122–148, 1963.

[80] P. Stoica and M. Jansson, "On maximum likelihood estimation in factor analysis– An algebraic derivation," *Signal Processing*, vol. 89, no. 6, pp. 1260–1262, 2009.

[81] P.J. Huber and E.M. Ronchetti, *Robust Statistics*, New Jersey: John Wiley & Sons, 2nd ed edition, 2009.

[82] F.R. Hampel, E.M. Ronchetti, and P.J. Rousseeuw, *Robust Statistics: The Approach Based on Influence Functions*, John Wiley & Sons, 1986.

[83] R.A. Maronna, R.D. Martin, and V.J. Yohai, *Robust Statistics: Theory and Methods*, John Wiley & Sons, 2006.

[84] P.J. Rousseeuw and A.M. Leroy, *Robust Regression and Outlier Detection*, New York: Wiley, 1987.

[85] D. Middleton, "Statistical-physical models of electromagnetic interference," *IEEE Trans. Electromagn. Compat.*, vol. 19, no. 3, pp. 106–127, Aug. 1977.

[86] M. Salibián-Barrera, S. Van Aelst, and G. Willems, "Fast and robust bootstrap," *Statistical Methods and Applications*, vol. 17, pp. 41–71, 2008.

[87] P.J. Rousseeuw and K. van Driessen, "A fast algorithm for the minimum covariance determinant estimator," *Technometrics*, vol. 41, no. 3, pp. 212–223, 1999.

[88] M. Salibián-Barrera, S. Van Aelst, and G. Willems, "PCA based on multivariate MM-estimators with fast and robust bootstrap," *J. Am. Statist. Assoc.*, vol. 101, pp. 1198–1211, 2005.

[89] J.P.C.L. da Costa, M. Haardt, F. Römer, and G. Del Galdo, "Enhanced model order estimation using higher-order arrays," in *Proc. 41-st Asilomar Conf. on Signals, Systems, and Computers*, Nov. 2007, pp. 412–416.

[90] R.R. Nadakuditi and A. Edelman, "Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2625–2638, Jul. 2008.

[91] Z.D. Bai and J.W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, Science Press, 2006.

[92] E. Wigner, "Characteristic vectors of bordered matrices with infinite dimensions," *Ann. Math.*, vol. 62, no. 3, pp. 548–564, Nov. 1955.

[93] V.A. Marčenko and L.A. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Math USSR-Sbornik*, vol. 1, no. 4, pp. 457–483, Apr. 1967.

[94] S. Kritchman and B. Nadler, "Non-parametric detection of the number of signals: Hypothesis testing and random matrix theory," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3930–3941, Oct. 2009.

[95] R.R. Nadakuditi and J.W. Silverstein, "Fundamental limit of sample generalized eigenvalue based detection of signals in noise using relatively few signal-bearing and noise-only samples," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 3, pp. 468–480, Jun. 2010.

[96] P.O. Perry and P.J. Wolfe, "Minimax rank estimation for subspace tracking," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 3, pp. 504–513, Jun. 2010.

[97] R. Everson and S. Roberts, "Inferring the eigenvalues of covariance matrices from limited, noisy data," *IEEE Trans. Signal Process.*, vol. 48, no. 7, pp. 2083–2091, Jul. 2000.

[98] M.O. Ulfarsson and V. Solo, "Dimension estimation in noisy PCA with SURE and random matrix theory," *IEEE Trans. Signal Process.*, vol. 56, no. 12, pp. 5804–5816, Dec. 2008.

[99] R. Couillet, J.W. Silverstein, Z. Bai, and M. Debbah, "Eigen-inference for energy estimation of multiple sources," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2420–2439, Apr. 2011.

[100] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*, 1st ed. New York, USA: Cambridge Univ. Press, 2011.

[101] I.M. Johnstone, "On the distribution of the largest eigenvalue in principal components analysis," *Ann. Statist.*, vol. 29, no. 2, pp. 295–327, 2001.

[102] J.W. Silverstein, "Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices," *J. Multivar. Anal.*, vol. 55, no. 2, pp. 331–339, 1995.

[103] J.W. Silverstein and Z.D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *J. Multivar. Anal.*, vol. 54, no. 2, pp. 175–192, 1995.

[104] Z. Bai, J. Hu, and W. Zhou, "Convergence rates to the Marčenko-Pastur type distribution," *Stoch. Proc. Appl.*, vol. 122, no. 1, pp. 68–92, 2012.

[105] Z.D. Bai and J.W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices," *Ann. Probab.*, vol. 26, no. 1, pp. 316–345, 1998.

[106] Z.D. Bai, B. Miao, and J. Yao, "Convergence rates of spectral distributions of large sample covariance matrices," *SIAM J. Matrix Anal. Appl.*, vol. 25, pp. 105–127, Jan. 2003.

[107] S. Geman, "A limit theorem for the norm of random matrices," *Ann. Probab.*, vol. 8, no. 2, pp. 252–261, 1980.

[108] J.W. Silverstein, "The smallest eigenvalue of a large dimensional wishart matrix," *Ann. Probab.*, vol. 13, no. 4, pp. 1364–1368, 1985.

[109] Y.Q. Yin, "Limiting spectral distribution for a class of random matrices," *J. Multivar. Anal.*, vol. 20, no. 1, pp. 50–68, 1986.

[110] Y.Q. Yin, Z.D. Bai, and P.R. Krishnaiah, "On the limit of the largest eigenvalue of the large dimensional sample covariance matrix," *Prob. Theory Related Fields*, vol. 78, pp. 509–521, 1988.

[111] Z.D. Bai and Y.Q. Yin, "Limit of the smallest eigenvalue of a large dimensional sample covariance matrix," *Ann. Probab.*, vol. 21, no. 3, pp. 1275–1294, 1993.

[112] J. Baik, G. Ben Arous, and S. Péché, "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices," *Ann. Probab.*, vol. 33, no. 5, pp. 1643–1697, 2005.

[113] J. Baik and J.W. Silverstein, "Eigenvalues of large sample covariance matrices of spiked population models," *J. Multivar. Anal.*, vol. 97, no. 6, pp. 1382–1408, 2006.

[114] D. Paul, "Asymptotics of sample eigenstructure for a large dimensional spiked covariance model," *Statistica Sinica*, vol. 17, no. 4, pp. 1617–1642, 2007.

[115] Z. Bai and J. Yao, "Central limit theorems for eigenvalues in a spiked population model," *Ann. Inst. Poincaré*, vol. 44, pp. 447–474, 2008.

[116] C. Tracy and H. Widom, "Level-spacing distributions and the airy kernal," *Commun. Math. Phys.*, vol. 159, pp. 151–174, 1994.

[117] C. Tracy and H. Widom, "On orthogonal and symplectic matrix ensembles," *Commun. Math. Phys.*, vol. 177, pp. 727–754, 1996.

[118] N. El Karoui, "A rate of convergence result for the largest eigenvalue of complex white wishart matrices," *Ann. Probab.*, vol. 34, no. 6, pp. 2077–2117, 2006.

[119] P.M. Lee, *Bayesian Statistics: An Introduction*, Oxford, U.K.: Oxford Univ. Press, 1989.

[120] P. Stoica, Y. Selen, and J. Li, "On information criteria and the generalized likelihood ratio test of model order selection," *IEEE Signal Process. Lett.*, vol. 11, no. 10, pp. 794–797, Oct. 2004.

[121] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge, U.K.: Cambridge Univ. Press, 1985.

[122] B.N. Parlett, *The Symmetric Eigenvalue Problem*, Englewood Cliffs, NJ: Prentice-Hall, 1980.

[123] R.J. Muirhead, *Aspects of Multivariate Statistical Theory*, New York: Wiley, 1982.

[124] Z.D. Bai and J.W. Silverstein, "CLT for linear spectral statistics of large-dimensional sample covariance matrices," *Ann. Prob.*, vol. 32, no. 1, pp. 553–605, 2004.

[125] I. Dumitriu and A. Edelman, "Global spectrum fluctuations for the β-Hermite and β-Laguerre ensembles via matrix models," *J. Math. Phys.*, vol. 47, no. 6, 2006.

[126] S. Zacks, *The Theory of Statistical Inference*, New York: Wiley, 1971.

[127] M.D. Springer, *The Algebra of Random Variables*, John Wiley & Sons, 1979.

[128] D.V. Hinkley, "On the ratio of two correlated normal random variables," *Biometrika*, vol. 56, no. 3, pp. 635–639, 1969.

[129] G.H. Golub and C.F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: The Johns Hopkins Univ. Press, 1996.

[130] Sir M. Kendall and A. Stuart, *The Advanced Theory of Statistics, Vol. 2: Inference and Relationship*, New York: Macmillan, 4th edition, 1977.

[131] S. Kay, "Exponentially embedded families - new approaches to model order estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 1, pp. 333–345, Jan. 2005.

[132] N.R. Goodman, "Statistical analysis based on a certain multivariate complex gaussian distribution (an introduction)," *Ann. Math. Stat.*, vol. 34, no. 1, pp. 152–177, 1963.

[133] A.T. James, "Distributions of matrix variates and latent roots derived from normal samples," *Ann. Math. Stat.*, vol. 35, no. 2, pp. 475–501, 1964.

[134] Z. Lu and A.M. Zoubir, "Source enumeration using the pdf of sample eigenvalues via information theoretic criteria," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'12)*, Mar. 2012, pp. 3361–3364.

[135] S. Tu and L. Xu, "A theoretical investigation of several model selection criteria for dimensionality reduction," *Pattern Recog. Lett.*, vol. 33, no. 9, pp. 1117–1126, 2012.

Curriculum vitae

Name:	Zhihua Lu
Date of birth:	18.05.1983
Place of birth:	Zhejiang, China
Family status:	single
Nationality:	Chinese

Education

11/2009 – 10/2012	Technische Universität Darmstadt, Germany Information- and Communication Engineering (PhD)
10/2006 – 06/2009	Technische Universität Darmstadt, Germany Information- and Communication Engineering (Diplom)
09/2000 – 06/2004	Zhejiang University, China Information- and Communication Engineering (Bachelor)

Work experience

08/2004 – 05/2005	UTStarcom Telecom CO., Ltd., Hangzhou, China

Erklärung laut §9 der Promotionsordnung

Ich versichere hiermit, dass ich die vorliegende Dissertation allein und nur unter Verwendung der angegebenen Literatur verfasst habe. Die Arbeit hat bisher noch nicht zu Prüfungszwecken gedient.

Darmstadt, 25. Oktober 2012

_____

Zhihua Lu