Proceedings of the World Congress on Engineering and Computer Science 2017 Vol I
WCECS 2017, October 25-27, 2017, San Francisco, USA

Source Enumeration with Random Matrix Theory in the Low SNR Regime

Huayang Chen, Yechao Bai, Xinggan Zhang, Qiong Wang, and Lan Tang

***Abstract*—Source enumeration is a critical step in array processing in communication, radar, and so on. Many existing methods are proposed under the assumption that the number of array is fixed while the number of snapshots tends to infinity. Their performances degrade in the case of small sample sizes and low signal-to-noise ratios (SNRs). In this paper, we employ random matrix theory to obtain the asymptotic distributions of eigenvalues and the Frobenius norm of observed data. By using Bayes formula, we derive a method that combines the information of eigenvalues distributions and the Frobenius norm together to get the estimation of the number of sources which outperforms other methods at low SNRs. Simulation results illustrate the proposed method is capable of correctly detecting the number of sources in the low SNR regime.**

***Index Terms*—array processing, source enumeration, random matrix theory, information theoretic criteria.**

I. Introduction

IN array processing, the observation vector can be modeled as a superposition of a finite number of signals corrupted by additive noise. Estimation of the signal parameters sometimes requires the knowledge of the number of sources. As signal numbers may be unknown, source enumeration is a key issue and has received considerable attention in array processing during these years.

The conventional source enumeration methodologies vary from hypothesis testing to the information theoretic criterion (ITC). The hypothesis testing method is a subjective judgment as it requires for deciding on the threshold levels. Unlike the hypothesis testing, the ITCs [1], such as Akaikes information criterion (AIC), Bayesian information criterion (BIC), and minimum description length (MDL) criterion, which are proposed in [2], [3], are based on the application of the information theoretic criteria for model selection. These methods minimize the Kullback-Leibler distance between the hypothesized model and the observed data to measure how well the model fits the observed data. The number of sources is determined by computing any one of these criteria for all candidate models and choosing the model with the smallest description [4]. As the best trade-off is achieved when the score is minimized, the ITCs need no subjective judgment to estimate the source number. However, the statistical method based on the AIC and MDL proposed by Wax and Kailath have drawbacks. The AIC is not consistent and tends to asymptotically overestimate the number of signal sources, and its probability of error cannot reach zero even at a high signal-to-noise ratio (SNR) [5]. The MDL criterion is consistent but it performs poorly at low SNR and usually underestimates the number of signal sources [5].

Most of the aforementioned methods are proposed under the assumption that the number of array M is fixed while the number of snapshots N tends to infinity, which is referred to as the classical asymptotic regime. However, when the number of array elements and the number of snapshots are finite and comparable in magnitude, i.e., $M/N \rightarrow c \in (0, \infty)$, which is referred to as the general asymptotic regime, ITCs cannot properly work. To solve this problem, Nadakuditi and Edelman have devised the RMT-AIC criterion [6]. Lu and Zoubir proposed a two-step test [7] for source enumeration by employing random matrix theory which is capable of correctly determining the number of sources in the case of small sample sizes. We now prefer a source enumerator that always selects the true source number for the general asymptotic regime even at low SNRs. As the two-step test only uses the information of the extreme eigenvalues of signals, the performance of the algorithm is not good enough at low SNRs because of the loss information of other eigenvalues. In this paper, we are intended to improve the performance of the second-step test of two-step test by adding the Frobenius norm information of received data to the enumerator. Some results of the singular values and singular vectors in random matrix theory are used to derive the estimator. Simulations show that the proposed method outperforms many existing approaches under low SNRs.

The remainder of the paper is organized as follows. The observation data model is given in section II. The new method for source enumeration with general asymptotic regime using random matrix theory is described in detail in Section III. Simulation results that illustrate the superior performance of the new method are presented in Section IV. Finally, conclusions are drawn in Section V.

II. Array Signal Model

Consider an array of M sensors of uniform linear receiving $P(P < M)$ uncorrelated narrowband source signals from far field located in distinct directions, the array received vector can be written as

$$\mathbf{x} = \mathbf{A}\left(\theta\right)\mathbf{s} + \mathbf{n} \tag{1}$$

where $\mathbf{x} = \left[x_1, x_2, ..., x_M\right]^T$ is the observed signal vector, θ_j refers to as the direction-of-arrival (DOA) of the jth source, $a\left(\theta_j\right)$ represents the $M \times 1$ array steering vector, $\mathbf{A}\left(\theta\right) = \left[\begin{array}{ccc} a\left(\theta_1\right) & ... & a\left(\theta_P\right) \end{array}\right]$ is the $M \times P$ array manifold matrix of the uniform linear array (ULA), $\mathbf{s} = \left[s_1, s_2, ..., s_P\right]^T$ is the source signal vector, and the noise vector is $\mathbf{n} = \left[n_1, n_2, ..., n_M\right]^T$.

We handle the P signals $s_1, s_2, ..., s_P$ as being deterministic. The noise vector is assumed to be a complex, stationary

Manuscript received May 10th, 2017. This work is supported by the Basic Research Program of Jiangsu Province (BK20151391).

H. Chen, Y. Bai, X. Zhang, Q. Wang and L. Tang are with the School of Electronic Science and Engineering, Nanjing University, Nanjing, China. e-mail: mg1623072@smail.nju.edu.cn, ychbai@nju.edu.cn (corresponding author), zhxg@nju.edu.cn, wangqiong@nju.edu.cn, tanglan@nju.edu.cn.

ISBN: 978-988-14047-5-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

and ergodic Gaussian vector process, independent of the signals, with zero mean and covariance matrix $\sigma^2\mathbf{I}$, where σ^2 is an unknown scalar constant and $\mathbf{I}$ is the identity matrix. The matrix $\mathbf{A}$ is of rank P.

In practice, only a finite set of observations is available. The received N snapshots of independent and identically distributed (i.i.d.) complex data can be expressed as

$$\mathbf{X} = \mathbf{A}(\theta)\mathbf{S} + \mathbf{N} \tag{2}$$

where $\mathbf{X} = [\mathbf{x}(t_1), ..., \mathbf{x}(t_N)]^T$, $\mathbf{S} = [\mathbf{s}(t_1), ..., \mathbf{s}(t_N)]^T$, and $\mathbf{N} = [\mathbf{n}(t_1), ..., \mathbf{n}(t_N)]^T$.

The population covariance matrix of the received data is given by

$$\mathbf{R}_X = \mathrm{E}\left[\mathbf{X}\mathbf{X}^H\right] = \mathbf{A}\mathbf{R}_S\mathbf{A}^H + \sigma^2\mathbf{I}_M \tag{3}$$

where $\mathbf{R}_S = \mathrm{E}\left[\mathbf{S}\mathbf{S}^H\right]$ is the signal covariance, $E(\cdot)$ and $(\cdot)^H$ denote expectation and Hermitian transpose, $\mathbf{I}_M$ is a $M \times M$ identity matrix. Denoting the eigenvalues of $\mathbf{R}_X$ by

$$\lambda_1 \geq \lambda_2 \geq ...\lambda_P > \lambda_{P+1} = ... = \lambda_M = \sigma^2 \tag{4}$$

where the first P eigenvalues are contributed by the sources and noise, while the smallest $M - P$ eigenvalues of $\mathbf{R}_X$ are contributed by noise only.

As we do not have access to the population covariance matrix in practice, we can only obtain the sample covariance matrix, which is calculated by

$$\widehat{\mathbf{R}}_X = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}(t_i)\mathbf{x}(t_i)^H = \frac{1}{N}\mathbf{X}\mathbf{X}^H \tag{5}$$

The corresponding sample eigenvalues of $\widehat{\mathbf{R}}_X$ are given by

$$l_1 > l_2 > ... > l_P > l_{P+1} ... > l_M \tag{6}$$

which are all distinct with probability one, when $N \geq M$. If $N > M$, the smallest $M - P$ eigenvalues are 0.

To make an estimation of source number, most existing algorithms concentrate on the distribution of the sample eigenvalues which suffers significant performance degradations in the case of low SNRs and finite sample sizes. We are now focus on developing a new approach under low SNRs and short snapshots using random matrix theory.

III. Method

As the random matrix theory provides a more accurate approximation for the distribution of the sample eigenvalues in the case of low SNRs and finite sample sizes, it has been used in the source enumeration of small sample sizes recently. In this section, we present an approach of two steps using some results of random matrix theory and singular values and vectors of low rank perturbations of large rectangular random matrices which are given in Appendix A.

A. First Step

As we learn from Theorem A.1, $b = \sigma^2(1+\sqrt{c})^2$ is the convergence of the largest noise eigenvalue, and the detectable signal eigenvalues converge to a limit larger than b. Thus, the limit value $b = \sigma^2(1+\sqrt{c})^2$ is the asymptotic separator between noise eigenvalues and signal eigenvalues, which can be chosen as the threshold for the hypothesis test of the largest noise eigenvalue. The number of signals can be detected by:

$$P_0 = \min\left\{k : l_{k+1} < b = \sigma^2\left(1+\sqrt{c}\right)^2\right\}, k = 0, 1, ..., p-1 \tag{7}$$

which is referred to as the first step. According to the lemma given in [7], the test in (7) can achieve $P_0 = P$ with probability one as the sample size N goes to infinity. However, the test tends to underestimate P in the case of low SNRs and finite sample sizes, and will underestimate P by one when underestimate occurs, i.e., $P_0 = P - 1$.

As we tend to underestimate the number of source signals at low SNRs and finite sample sizes, we are supposed to reduce the underestimation probability or pull up the estimate P_0 from $P - 1$ to P to remedy the underestimation, as explained in the second step.

B. Second Step

In the first step, we derived an initial estimation P_0 for the number of signals. The true value of the source number $P = P_0$ or $P_0 + 1$, as we have explained in the first step. In the second step, we conduct a test between the following two hypotheses

$$\mathrm{H}_0\text{: } P_0 \text{ sources vs. } \mathrm{H}_1\text{: } P_0 + 1 \text{ sources}$$

which can be discriminated using the sample eigenvalues distributions.

As it is shown in [7], the joint probability density functions (pdfs) of sample eigenvalues under the hypotheses H_0 and H_1, which can be respectively written as $f(l_1, ..., l_M|H_0)$ and $f(l_1, ..., l_M|H_1)$, depend asymptotically on the eigenvalue l_{P_0+1} only, which is the technical basis of conducting Roys largest root test [10]. In other words, the eigenvalue l_{P_0+1}, which is referred to as the extreme eigenvalue, contains the richest statistical information and plays a dominant role in source number estimation. Thus, the information of the extreme eigenvalue is implemented to discriminate the two hypotheses.

However, the extreme eigenvalue l_{P_0+1} is greatly affected by noise in the case of low SNRs, which may decrease the accuracy of the estimation. To remedy this, we derive the Frobenius norm from the observed data and combine it with the information of the extreme eigenvalue using Bayes formula.

The second step contains three courses.

1) The information of the extreme eigenvalue: We calculate the prior probability of the hypotheses H_0 and H_1 using the information of the extreme eigenvalue l_{P_0+1}. We use $p(H_0)$ to denote the pdf of the eigenvalue l_{P_0+1} when it is the noise eigenvalue, and $p(H_1)$ denoting the probability when l_{P_0+1} is the signal eigenvalue.

It follows from the Marčenko-Pastur law that

$$p(H_0) = f_c(l_{P_0+1}) \tag{8}$$

where the pdf f_c is given in (21) in Theorem A.1.

The expression of $p(H_1)$ can be given as

$$\begin{aligned} p(H_1) &= \int_{\alpha_1}^{\alpha_2} f_\nu(l_{P_0+1}|\nu_{P_0+1}) f(\nu_{P_0+1})\, d\nu_{P_0+1} \\ &= \frac{1}{\alpha_2-\alpha_1}\int_{\alpha_1}^{\alpha_2} f_\nu(l_{P_0+1}|\nu_{P_0+1}) d\nu_{P_0+1} \end{aligned} \tag{9}$$

ISBN: 978-988-14047-5-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

with the bounds $\alpha_1 = \max\left(\sigma^2\sqrt{c}, l_{P_0+1} - b\right)$, $\alpha_2 = l_{P_0+1} - a$ and $f_\nu\left(l_{P_0+1}|\nu_{P_0+1}\right)$ given in (24) in Theorem A.2 and a,b given in (22).

2) The Frebenius norm of observed data: We will use the Frobenius norm of observed data, and combine it with the information of extreme eigenvalues we have obtained in (8) and (9) to raise the probability of correct estimation at low SNRs.

As it is given in (37) Appendix B, the Frobenius norm of observed data is denoted by Λ. The value of Λ contains richest statistical information of source numbers. Λ only contains the information of noise under the hypothesis H_0, but contains both of the noise and signal under the hypothesis H_1. Thus Λ follows different distribution under the hypothesis H_0 and H_1.

Λ can be written as

$$\Lambda = tr\left(N^H N\right) = \sum_{i=1}^{MN} {n_i}^2 \tag{10}$$

under the hypothesis H_0. As noise is assumed to be uncorrelated Gaussian noise with zero mean and variance $1/N$, we have the mean and variance of Λ

$$\begin{array}{c} E\left(\Lambda_{H_0}\right) = \mu_{H_0} = M \\ \mathrm{var}\left(\Lambda_{H_0}\right) = {\sigma_{H_0}}^2 = \frac{2M}{N} \end{array} \tag{11}$$

Λ under the hypothesis H_1 is given as

$$\Lambda = tr\left[\left(N + v_{P_0+1}\right)^H \left(N + v_{P_0+1}\right)\right] = \sum_{i=1}^{MN}\left(n_i + v_{P_0+1}\right)^2 \tag{12}$$

The mean and variance of Λ is given as follows:

$$\begin{array}{c} E\left(\Lambda_{H_1}\right) = \mu_{H_1} = M + NMv_{P_0+1}^2 \\ \mathrm{var}\left(\Lambda_{H_1}\right) = {\sigma_{H_1}}^2 = \frac{2M}{N} + 4Mv_{P_0+1}^2 \end{array} \tag{13}$$

where v_{P_0+1} is the P_0+1 singular value of all signals.

As Λ is the sum of a large number of the variables ${n_i}^2$, Λ follows the Gaussian distribution according to the central limit theorem. Let $p\left(\Lambda|H_0\right)$ and $H\left(\Lambda|H_1\right)$ be the pdf of Λ under hypothesis H_0 and H_1. We have:

$$p\left(\Lambda|H_0\right) = \frac{1}{\sqrt{2\pi{\sigma_{H_0}}^2}}\exp\left[-\frac{\left(\Lambda - \mu_{H_0}\right)^2}{2{\sigma_{H_0}}^2}\right] \tag{14}$$

where μ_{H_0} and σ_{H_0} is given in (11).

As the exact singular value v_{P_0+1} of signals in (13) cannot be calculated at low SNRs when ${\theta_{\mathbf{S}i}}^2 \le \sqrt{c}$, $p\left(\Lambda|H_1\right)$ is supposed to be written as an integration of a conditional probability $p\left(\Lambda|H_1, v_{P_0+1}\right)$ under the hypothesis H_1 and a prior density $p\left(v_{P_0+1}|H_1\right)$ which is chosen for v_{P_0+1}, which is given by

$$p\left(\Lambda|H_1\right) = \int p\left(\Lambda|v_{P_0+1}, H_1\right)p\left(v_{P_0+1}|H_1\right)dv_{P_0+1} \tag{15}$$

with

$$p\left(\Lambda|H_1, v_{P_0+1}\right) = \frac{1}{\sqrt{2\pi{\sigma_{H_1}}^2}}\exp\left[-\frac{\left(\Lambda - \mu_{H_1}\right)^2}{2{\sigma_{H_1}}^2}\right] \tag{16}$$

where μ_{H_1} and σ_{H_1} is given in (13).

For simplicity, we assume v_{P_0+1} to have uniform distribution in the interval $(\varepsilon_1, \varepsilon_2)$, that is

$$p\left(v_{P_0+1}|H_1\right) = \begin{cases} \frac{1}{\varepsilon_2 - \varepsilon_1} & \varepsilon_1 < v_{P_0+1} < \varepsilon_2 \\ 0 & otherwise \end{cases} \tag{17}$$

Then the pdf $p\left(\Lambda|H_1\right)$ can be expressed as

$$p\left(\Lambda|H_1\right) = \frac{1}{\varepsilon_2 - \varepsilon_1}\int_{\varepsilon_1}^{\varepsilon_2} p\left(\Lambda|H_1, v_{P_0+1}\right)dv_{P_0+1} \tag{18}$$

where $p\left(\Lambda|H_1, v_{P_0+1}\right)$ is given in (16). The bound ε_1 and ε_2 are the range of the smallest signal amplitude v_{P_0+1}, which we choose them to be $\varepsilon_1 = 0$ and $\varepsilon_2 = \sigma^2\sqrt{c}$.

3) Combination: Finally, to integrate the information of extreme eigenvalues and the Frobenius norm of observations, we use Bayes formula to get the combination tests $p\left(H_0|\Lambda\right)$ and $p\left(H_1|\Lambda\right)$ under the two hypotheses. They are given as

$$\begin{array}{c} p\left(H_0|\Lambda\right) = \frac{p(\Lambda|H_0)p(H_0)}{p(\Lambda)} \\ p\left(H_1|\Lambda\right) = \frac{p(\Lambda|H_1)p(H_1)}{p(\Lambda)} \end{array} \tag{19}$$

where $p\left(H_0\right)$ is given in (8), $p\left(H_1\right)$ is given in (9), $p\left(\Lambda|H_0\right)$ is given in (14) and $p\left(\Lambda|H_1\right)$ is given in (18). As $p\left(H_0|\Lambda\right)$ and $p\left(H_1|\Lambda\right)$ has the same denominator $p\left(\Lambda\right)$, it can be neglected.

Thus, the two hypotheses in (19) can be discriminated using the test

$$\frac{p\left(H_1|\Lambda\right)}{p\left(H_0|\Lambda\right)} \underset{H_0}{\overset{H_1}{\gtrless}} 1 \tag{20}$$

IV. SIMULATIONS

In this section, simulation results are presented to validate the proposed method and to demonstrate its performance.

In our simulations, we employ an array with $M = 50$ sensors and $P = 5$ source signals with directions-of-arrivals $20°, 22°, 25°, 27°$ and $30°$ .The case of uncorrelated complex deterministic signals contaminated by white Gaussian noise was considered. All simulation results were obtained based on 500 Monte Carlo runs. We denote the AIC method in [2] by “AIC”. Denote by “TwoStepTest-1” the first-step of the two-step test, by “TwoStepTest-2” the second-step of the two-step test. The first step of our proposed method is the same as “TwoStepTest-1”, and the second step of our method is denoted by “Proposed Method”. We will evaluate these methods in different experimental settings.

In the first set of simulations, we focus on the case that sample size N varied around array number M. We now have the SNRs for all sources are set as -10dB. The number of snapshots N is varied from 10 to 100.

As it is shown in Fig. 1, the method “AIC” only works when N is larger than M. The proposed method “Proposed Method” can get the estimation more accurately than the other methods in the case of short snapshots, even when N is much smaller than M. As the eigenvalues information loss appears when the sample size is relatively small, the probability of correct determination of two-step test “TwoStepTest-2” is not as high as the proposed method “Proposed Method”. It is proved in our simulation that the performance of algorithm gets significant improvement at short snapshots when we add the Frobenius norm information in our detection.

In the other set of simulations, we focus on the case that sample size N closed to the array number M when the SNRs for all sources are varied from -20dB to 0dB. The number of snapshots N is set as 40. The direction of arrivals and number of sensors are the same as that in the

ISBN: 978-988-14047-5-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

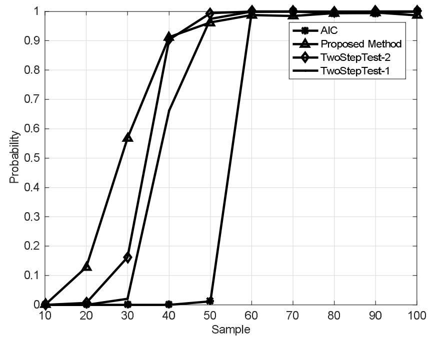

Fig. 1: Probability of correct estimation versus number of Samples

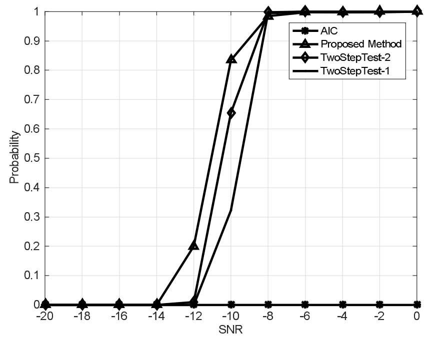

Fig. 2: Probability of correct estimation versus SNR

first simulation. It is seen in Fig. 2 that the proposed method "Proposed Method" shows obvious improvement of detecting the number of sources at a relatively short snapshots and low SNRs compared to the other methods. It is validated that the performance of algorithm gets significant improvement at low SNRs when we add the Frobenius norm of observations in our detection.

V. Conclusion

In this paper, we have proposed an approach for source enumeration in asymptotic regime under large-array condition at low SNRs by employing random matrix theory. We exploit the information from the distributions of the eigenvalues and the Frobenius norm of observations to get a better detection of number of signals. By using the threshold of the extreme eigenvalues of signals, we tend to underestimate the number of sources by one at low SNRs and short snapshots. Simulations have validated that the second-step in our proposed method does significant improvement in pulling up the estimation to the correct one. The proposed method outperforms some popular approaches at low SNRs and small sample size in the case that $M, N \to \infty$ and $M/N \to c > 0$. Nevertheless, the proposed method estimates the noise variance using noise data received from the array. It still remains a question to find an appropriate estimator to directly detect the noise variance using the observed data containing noise. And how to relax the assumption of i.i.d. Gaussian noise structure could be a part of future work.

Appendix A
Mathematical Preliminaries of Random Matrix Theory

Here we review some mathematical theories relevant to the problem at hand, in particular results from random matrix theory, and results about singular values and vectors of low rank perturbations of large rectangular random matrices.

Theorem A.1:[8] Let $\mathbf{N}$ be a $M \times N$ matrix with entries n_{ij} that are i.i.d. with mean zero and variance σ^2. Let $\widehat{\mathbf{R}}_N$ denote the sample covariance matrix of $\mathbf{N}$ with corresponding eigenvalues $l_i, i = 1, ..., M$. The density of the signal sample eigenvalue l_i converges w.p.1 to the Marčenko-Pastur density

$$f_c(l) = \max\left(0, 1 - \frac{1}{c}\right)\delta(l) + \frac{1}{2\pi c\sigma^2 l}\sqrt{(l-a)(b-l)}\mathbf{1}_{a \le l \le b} \tag{21}$$

where $c = M/N$ and

$$\begin{aligned} a &= \sigma^2(1 - \sqrt{c})^2 \\ b &= \sigma^2(1 + \sqrt{c})^2 \end{aligned} \tag{22}$$

The indicator function $\mathbf{1}_{a \le l \le b} = 1$ for $a \le l \le b$ and zero otherwise, $\delta(l)$ is the Dirac delta function.

The largest and smallest non-zero eigenvalues converge w.p.1 to the edges of the support of f_c:

$$\begin{aligned} l_1 &\to b \\ l_{\min(M,N)} &\to a \end{aligned} \tag{23}$$

Theorem A.2:[7] Let $\widehat{\mathbf{R}}_X$ denote a sample covariance matrix estimated from the $M \times N$ matrix of Gaussian observations with columns that are i.i.d. with mean zero and population covariance matrix $\mathbf{R}_X$. The eigenvalues of $\mathbf{R}_X$ and of $\widehat{\mathbf{R}}_X$ are respectively denoted by $\lambda_1 \ge \lambda_2 \ge ... \ge \lambda_P > \lambda_{P+1} = ... = \lambda_M = \sigma^2$ and $l_1, l_2, ..., l_M$. Denote the ith signal strength by $v_i = \lambda_i - \sigma^2$. In the joint limit $M, N \to \infty$ with $M/N \to c > 0$. For the ith signal strength $v_i = \lambda_i - \sigma^2 > \sigma^2\sqrt{c}$, the density of the corresponding signal sample eigenvalue l_i converges w.p.1 to the normal density

$$f_v(l) = \frac{1}{\delta\sqrt{2\pi}}\exp\left\{-\frac{(l-\tau)^2}{2\delta^2}\right\} \tag{24}$$

with

$$\begin{aligned} \tau &= \left(\nu + \sigma^2\right)\left(1 + \tfrac{c\sigma^2}{\nu}\right) \\ \delta &= \left(\nu + \sigma^2\right)\sqrt{\tfrac{2}{\beta N}\left(1 - \tfrac{c\sigma^4}{\nu^2}\right)} \end{aligned} \tag{25}$$

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

Theorem A.3:[9] Let $\mathbf{N}$ be a $M \times N$ complex matrix with independent, zero mean and normally distributed entries with variance $1/N$. $\mathbf{P}_M = \mathbf{A}(\theta)\mathbf{S}$ be a $M \times N$ deterministic matrix with P non-zero singular values $\theta_{\mathbf{S}_1} \ge ... \ge \theta_{\mathbf{S}_P}$ (P is independent of M, N). Let $\mathbf{X} = \mathbf{A}(\theta)\mathbf{S} + \mathbf{N}$ and $\mathbf{P}_M = \sum_{i=1}^{P}\theta_{\mathbf{S}_i}u_{\mathbf{S}_i}v_{\mathbf{S}_i}{}^H$. For any $\mathbf{P}_M$, and for any fixed $i \ge 1$, we have

$$\theta_{\mathbf{X}_i}(\mathbf{P}_M + \mathbf{N}) \xrightarrow{a.s.} \begin{cases} \sqrt{\frac{(1+\theta_{\mathbf{S}_i}{}^2)(c+\theta_{\mathbf{S}_i}{}^2)}{\theta_{\mathbf{S}_i}{}^2}} & i \le P, \theta_{\mathbf{S}_i} > c^{\frac{1}{4}} \\ 1 + \sqrt{c} & otherwise \end{cases} \tag{26}$$

where $c = M/N$, $\theta_{\mathbf{X}_i}$ is the ith singular value of $\mathbf{X}$.

The relationship of the singular vectors can be expressed as follow:

$$|\langle u_i, u_{\mathbf{S}_i}\rangle|^2 \xrightarrow{a.s.} \begin{cases} \gamma_1 & \theta_{\mathbf{S}_i} > c^{1/4} \\ 0 & otherwise \end{cases} \quad (27)$$

$$|\langle v_i, v_{\mathbf{S}_i}\rangle|^2 \xrightarrow{a.s.} \begin{cases} \gamma_2 & \theta_{\mathbf{S}_i} > c^{1/4} \\ 0 & otherwise \end{cases} \quad (28)$$

where $\gamma_1 = 1 - \frac{c(1+\theta_{\mathbf{S}_i}{}^2)}{\theta_{\mathbf{S}_i}{}^2(c+\theta_{\mathbf{S}_i}{}^2)}$ and $\gamma_2 = 1 - \frac{(c+\theta_{\mathbf{S}_i}{}^2)}{\theta_{\mathbf{S}_i}{}^2(1+\theta_{\mathbf{S}_i}{}^2)}$.

APPENDIX B
THE FROBENIUS NORM OF OBSERVED DATA

As the formulas in Theorem A.3 will be applied in the following calculation, the variance of $\mathbf{X}$ are supposed to be $1/N$. To normalize the variance, $\mathbf{X}$ is handled by

$$\mathbf{X}' = \mathbf{X}/\sqrt{N\sigma^2} = (\mathbf{AS} + \mathbf{N})/\sqrt{N\sigma^2} = \mathbf{AS}' + \mathbf{N}' \quad (29)$$

The dimension of $\mathbf{X}'$ is $M \times N$. And the rank of $\mathbf{X}'$ is $\min(M, N)$. Then by singular value decomposition (SVD), $\mathbf{X}'$ can be expressed by

$$\mathbf{X}' = \sum_{i=1}^{\min(M,N)} \theta_{\mathbf{X}i} u_{\mathbf{X}i} v_{\mathbf{X}i}^H \quad (30)$$

where $\theta_{\mathbf{X}i}$ is the singular value of $\mathbf{X}'$, $u_{\mathbf{X}i}$ and $v_{\mathbf{X}i}$ are left and right singular vector.

The dimension of $\mathbf{S}'$ is $P \times N$. As the P signals are uncorrelated and the rank of $\mathbf{A}$ is P, the rank of $\mathbf{AS}'$ is P, that is, there are $\min(M, N) - P$ singular values that are zero. Then using SVD and only considering non-zero singular values, $\mathbf{AS}'$ can be expressed by

$$\mathbf{AS}' = \sum_{i=1}^{P} \theta_{\mathbf{S}i} u_{\mathbf{S}i} v_{\mathbf{S}i}^H \quad (31)$$

where $\theta_{\mathbf{S}i}$ is the singular value of $\mathbf{AS}'$, $u_{\mathbf{S}i}$ and $v_{\mathbf{S}i}$ are left and right singular vector.

Using the result in the first step, the observation data can be expressed as

$$\mathbf{X}' = \mathbf{AS}' + \mathbf{N}' = \sum_{i=1}^{P_0} \theta_{\mathbf{S}i} u_{\mathbf{S}i} v_{\mathbf{S}i}^H + \theta_{\mathbf{S}_{P_0+1}} u_{\mathbf{S}_{P_0+1}} v_{\mathbf{S}_{P_0+1}}^H \mathbf{1}_{H_1} + \mathbf{N}' \quad (32)$$

where P_0 is the initial estimation derived in the first step indicating the number of eigenvalues that are larger than the threshold $b = \sigma^2(1+\sqrt{c})^2$. The indicator function $\mathbf{1}_{H_1} = 1$ for the hypothesis H_1 and zero otherwise.

To obtain the Frobenius norm of $\mathbf{S}_{P_0+1}\mathbf{1}_{H_1} + \mathbf{N}'$, we have

$$\begin{aligned} &\|\mathbf{S}_{P_0+1}\mathbf{1}_{H_1} + \mathbf{N}'\|_F^2 \\ &= tr\left[(\mathbf{S}_{P_0+1}\mathbf{1}_{H_1} + \mathbf{N}')^H (\mathbf{S}_{P_0+1}\mathbf{1}_{H_1} + \mathbf{N}')\right] \\ &= tr\left[(\mathbf{X}' - \mathbf{S}_{P_0})^H (\mathbf{X}' - \mathbf{S}_{P_0})\right] \\ &= tr\left[\mathbf{X}'^H\mathbf{X}' + \mathbf{S}_{P_0}^H\mathbf{S}_{P_0} - \mathbf{X}'^H\mathbf{S}_{P_0} - \mathbf{S}_{P_0}^H\mathbf{X}'\right] \end{aligned} \quad (33)$$

where

$$\mathbf{S}_{P_0} = \sum_{i=1}^{P_0} \theta_{\mathbf{S}i} u_{\mathbf{S}i} v_{\mathbf{S}i}^H, \ \mathbf{S}_{P_0+1} = \theta_{\mathbf{S}_{P_0+1}} u_{\mathbf{S}_{P_0+1}} v_{\mathbf{S}_{P_0+1}}^H \quad (34)$$

Putting (34) into (33) and using Λ to denote $\|\mathbf{S}_{P_0+1}\mathbf{1}_{H_1} + \mathbf{N}'\|_F^2$ under the estimation P_0, we have

$$\Lambda = \sum_{i=1}^{\min(M,N)} \theta_{\mathbf{X}i}{}^2 + \sum_{i=1}^{P_0} \left[\theta_{\mathbf{S}i}{}^2 - u_{\mathbf{S}i} u_{\mathbf{X}i}^H v_{\mathbf{S}i} v_{\mathbf{X}i}^H \left(\theta_{\mathbf{X}i}^H \theta_{\mathbf{S}i} + \theta_{\mathbf{S}i}^H \theta_{\mathbf{X}i}\right)\right] \quad (35)$$

According to (26) in Theorem A.3, when $\theta_{\mathbf{X}i} > 1 + \sqrt{c}$, $\theta_{\mathbf{S}i}{}^2$ can be solved approximately by equation

$$\theta_{\mathbf{S}i}{}^4 + \left(1 + c - \theta_{\mathbf{X}i}{}^2\right)\theta_{\mathbf{S}i}{}^2 + c = 0 \quad (36)$$

Using the result in (27), (28) and the solution of (36), Λ can be calculated by

$$\Lambda = \sum_{i=1}^{\min(M,N)} \theta_{\mathbf{X}i}{}^2 + \sum_{i=1}^{P_0} \left[\theta_{\mathbf{S}i}{}^2 - \sqrt{\gamma_1\gamma_2}\left(\theta_{\mathbf{X}i}^H \theta_{\mathbf{S}i} + \theta_{\mathbf{S}i}^H \theta_{\mathbf{X}i}\right)\right] \quad (37)$$

where γ_1 and γ_2 are given in (27) and (28).

REFERENCES

[1] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *IEEE Signal Processing Magazine*, vol. 21, no. 4, pp. 36–47, 2004.
[2] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 2, pp. 387–392, 1985.
[3] G. Schwarz *et al.*, "Estimating the dimension of a model," *The annals of statistics*, vol. 6, no. 2, pp. 461–464, 1978.
[4] S. Valaee and P. Kabal, "An information theoretic approach to source enumeration in array signal processing," *IEEE Transactions on Signal Processing*, vol. 52, no. 5, pp. 1171–1178, 2004.
[5] C.-M. Cho and P. M. Djuric, "Detection and estimation of doa's of signals via bayesian predictive densities," *IEEE Transactions on Signal Processing*, vol. 42, no. 11, pp. 3051–3060, 1994.
[6] R. R. Nadakuditi and A. Edelman, "Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 2625–2638, 2008.
[7] Z. Lu and A. M. Zoubir, "Source enumeration in array processing using a two-step test." *IEEE Trans. Signal Processing*, vol. 63, no. 10, pp. 2718–2727, 2015.
[8] R. Couillet and M. Debbah, "Signal processing in large systems: A new paradigm," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 24–39, 2013.
[9] F. Benaych-Georges and R. R. Nadakuditi, "The singular values and vectors of low rank perturbations of large rectangular random matrices," *Journal of Multivariate Analysis*, vol. 111, pp. 120–135, 2012.
[10] S. N. Roy, "On a heuristic method of test construction and its use in multivariate analysis," *The Annals of Mathematical Statistics*, pp. 220–238, 1953.

ISBN: 978-988-14047-5-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)