Estimating the Number of Sources in White Gaussian Noise: Simple Eigenvalues Based Approaches

IET Signal Processing

Research Article

Estimating the number of sources in white Gaussian noise: simple eigenvalues based approaches

ISSN 1751-9675
Received on 8th March 2016
Revised 8th August 2016
Accepted on 7th February 2017
doi: 10.1049/iet-spr.2016.0128
www.ietdl.org

Ahmed Badawy[1] ✉, Tara Salman[2], Tarek Elfouly[3], Tamer Khattab[4], Amr Mohamed[3], Mohsen Guizani[3]

Q1 *[1]Department of Electronics and Telecommunications, Politecnico di Torino, Torino 10129, Italy*
Q2 *[2]Washington University in St. Louis, Saint Louis, Missouri 63130-4899, USA*
[3]Department of Computer Science and Engineering, Qatar University, Doha, Ad Dawhah 02713, Qatar
[4]Department of Electrical Engineering, Qatar University, Doha 2713, Qatar
✉ *E-mail: ahmed.badawy@polito.it*

Abstract: Estimating the number of sources is a key task in many array signal processing applications. Conventional algorithms such as Akaike's information criterion (AIC) and minimum description length (MDL) suffer from underestimation and overestimation errors. In this study, the authors propose four algorithms to estimate the number of sources in white Gaussian noise. The authors' proposed algorithms are categorised into two main categories; namely, sample correlation matrix (CorrM) based and correlation coefficient matrix (CoefM) based. Their proposed algorithms are applied on the CorrM and CoefM eigenvalues. They propose to use two decision statistics, which are the moving increment and the moving standard deviation of the estimated eigenvalues as metrics to estimate the number of sources. For their two CorrM based algorithms, the decision statistics are compared to thresholds to decide on the number of sources. They show that the conventional process to estimate the threshold is mathematically tedious with high computational complexity. Alternatively, they define two threshold formulas through linear regression fitting. For their two CoefM based algorithms, they re-define the problem as a simple maximum value search problem. Results show that the proposed algorithms perform on par or better than AIC and MDL as well as recently modified algorithms at medium and high signal-to-noise ratio (SNR) levels and better at low SNR levels and low number of samples, while using a lower complexity criterion function.

Q3

1 Introduction

Accurate and efficient estimation of number of sources is a critical task in many signal processing applications. Moreover, several applications in telecommunications and security assume a number of sources to be known a priori. As a result, many algorithms such as direction of arrival (DoA) [1, 2], blind source and channel order separation [3, 4] rely on the robustness of estimating the number of sources in order to function properly. As an example, DoA estimation algorithms such as MUltiple SIgnal Classification assume the number of sources to be known a priori [5]. Thus, having number of sources accurately estimated is critical for correct DoA estimation. Furthermore, DoA estimation is involved in many further applications such as localisation and tracking of objects and beamforming in wireless networks [6].

Many algorithms have been proposed to estimate the number of sources. According to [7], these algorithms can be classified into: information theoretic based [8], threshold based [9], eigenvector based [10, 11] and data based estimations [12]. The first two are the most popular, which we summarise below:

- *Information theoretic based estimation:* algorithms such as Akaike's information criterion (AIC), Bayesian information criterion (BIC) and minimum description length (MDL) use information theoretic criteria to estimate the number of sources [8, 13]. They search for minimum value of their log-likelihood function and an added penalty term. AIC and BIC tend to overestimate the number of sources and their error rate does not reach zero even at high signal-to-noise ratio (SNR) levels [14]. MDL underestimates the number of sources and has a poor performance at low SNR levels [14].In [15], the authors propose a design strategy for the penalty terms in both AIC and BIC with a finite number of samples. This is achieved by analysing the probability of model selection of algorithms, which gave an insight on the performance of underestimation and overestimation in the algorithms. They design the penalty based on a tradeoff between the overestimation and underestimation probabilities. Such approach would minimise the error resulting from underestimation and keep the overestimation probability under a certain level. Results show the effectiveness of the new optimised penalty terms when compared to traditional AIC and BIC algorithms, however, the new penalty depends on the maximum probability of overestimation along with the overestimated number of sources, which were assumed to be given. In practical scenarios, these parameters should be measured empirically a priori and shall differ from one scenario to another.

In addition, Nadakuditi and Edelman [16] propose a modification of AIC based on random matrix theory (RMT), which is denoted by RMT_AIC, which changes the log-likelihood minimisation criterion to include an approximation of noise log-likelihood in finite number of samples. Similarly, Seghouane [17] presents a corrected AIC, called AIC_{C3}, that adds a simplified approximation of bootstrap penalty term. Bootstrap penalty term, as presented in [18], tends to perform well with finite number of samples; however, it is complex due to the usage of cross-validation approximation process. Hence, AIC_{C3} simplifies the cross-correlation calculation by approximating it using linear regression estimator. RMT_AIC solves the problem of overestimation in high SNR while AIC_{C3} has a better performance at low SNR values.

Moreover, a modification to the classical MDL is presented in [19]. The authors modify the MDL criterion for number of sources estimation utilising the Gaussian assumption of the noise subspace and the identity structure of the covariance matrix. Using linear shrinking and the Gaussian assumption of the observations, the algorithm finds the covariance matrix of the noise subspace and its corresponding eigenvalues which can be used in the MDL criterion instead of the full sample covariance matrix eigenvalues in traditional MDL. Results show that the new algorithm outperforms existing techniques in many cases and were in line with the theoretical analysis presented in this paper. However, the algorithm requires a second multidimensional operation to calculate the linear shrinking values, which adds to traditional algorithms complexity.

- *Threshold based estimation:* the number of sources in this category is estimated by setting a threshold on some statistics of the estimated eigenvalues. Exceeding that threshold indicates the transition to signal eigenvalues and the number of sources can be detected at that point. The threshold can be set based on the upper bound of noise eigenvalues as in [20] or based on the difference between two consecutive eigenvalues as in [21]. The main drawback in most of the techniques in this category is that the formula needed to estimate the threshold has an adjustment coefficient, which needs to be set beforehand. This adjustment coefficient is estimated through extensive computer simulation for each pair of antenna elements and number of collected samples. In other words, if the number of antenna elements, the number of collected samples or both change, the coefficient has to be reconfigured accordingly, which adds a considerable burden to the system.

  In [22], the authors derived the profile of ordered noise eigenvalues, which they found to approximately fit an exponential law. Thus, they derived a recursive algorithm to estimate the number of sources by detecting a mismatch between the observed eigenvalues and the theoretical exponentially behaving eigenvalues. When this mismatch exceeds a certain threshold, it indicates that this is a signal eigenvalue and the number of sources can be detected at this point. A major drawback in their proposed technique is that their threshold is estimated through extensive Monte Carlo simulation. To estimate the threshold, they generate a large number of noise only matrices, on which they apply an eigenvalue decomposition (EVD) operation in order to take the mean of the noise only eigenvalues. Therefore, for a different noise variance, number of samples or number of antenna array elements, this exhaustive process needs to be repeated a priori.

  In a recent publication [23], the authors solve the problem of number of sources estimation in a two-step threshold algorithm. In the first step, they estimate a threshold that distinguishes between the signal and noise subspace based on the noise variance, as in [20], however, that tends to underestimate the number of sources in most cases. Hence, they developed the second step, which is based on the likelihood ratio test and a second threshold. The second threshold is based on maximum-a-posteriori probability (MAP) rule, which maximises the probability of correct estimation. The joint probability density functions of the sample eigenvalues used in the likelihood ratio test are for two hypotheses; one assumes the correct number of sources was estimated in the first step, while the second assumes that it was underestimated. Results showed that the two-step algorithm outperforms traditional ones. The improvement in performance comes at the cost of additional computational complexity from the second step. In addition, in their second step, the probabilities of the hypotheses used in their MAP rule are assumed to be known a priori, which is not the case in many applications.

To address the overestimation and underestimation errors of information theoretic algorithm and the reconfigurability problem in the threshold based algorithms, we propose some simple yet efficient solutions to estimate the number of sources. Our proposed algorithms can be categorised into two main categories based on the matrix used in estimating the number of sources. Namely, we propose sample correlation matrix (CorrM) based algorithms and correlation coefficient matrix (CoefM) based algorithms. We define two decision statistics, namely moving increment (MI) and moving standard deviation (MS) of the estimated eigenvalues, which are used as the metrics to estimate the number of sources. In other words, in each category, we propose two algorithms. We first estimate the selected matrix, apply an EVD operation and then estimate the decision statistics from its eigenvalues. In the first category, the decision statistics are compared to a preset threshold, while in the second a simple maximisation technique is proposed. Our main contributions in this work as compared to others include:

- Proposing four novel algorithms that use two different matrices to estimate the number of sources.
- Exploiting two different decision statistics to distinguish between noise and signal eigenvalues, i.e. MI and MS.
- For the two CorrM based algorithms, we define two non-reconfigurable formulas to estimate the threshold for each decision statistic. First, we find the distribution of the probability of false alarm of the MI case. We show it is a mathematically tedious process to estimate the threshold through this conventional process. We then derive the thresholds using regression analysis.
- For CoefM based algorithms, we redefine the problem as a simple maximisation problem.
- We compare the performance of our proposed algorithms to the traditional AIC and MDL algorithms and some of their recent modifications. We show that our proposed algorithms have comparable performance at medium and high levels SNR and better performance at low SNR values as well as low number of samples.
- We compare the required number of floating point operation (flops) of the criterion function of our proposed algorithms to AIC and MDL and show that our proposed algorithms have less computational complexity.

To the best of the authors' knowledge, estimating the number of sources through a maximisation approach applied on the estimated eigenvalues of the CoefM, or by a single threshold formula without an adjusting coefficient applied on the eigenvalues of the CorrM, have not been presented in the literature before. While in most existing literature, the received signal is modelled as stochastic with zero mean, non-zero mean signals are a combination of zero mean signals and unknown deterministic constants [24]. There exist several applications of non-zero mean signals [Refer to [24] for further details on such applications. Moreover, in case of non-Q4 RF signals such as acoustic (such as in [25]), the wireless stochastic signal is generally non-zero mean.]. Since the methodology in our paper is meant to be more generalised towards any array signal processing for DoA and number of sources estimation in any wireless communications system, we keep the more generalised form of non-zero mean signals. In the following, we denote samples by small symbols, vectors by small bold symbols and matrices by capital bold symbols. The remaining of this paper is organised as follows, Section 2 presents the system model while Section 3 discusses the proposed algorithms. Section 4 reviews some existing techniques. Section 5 presents simulation results and comparisons and finally, conclusion is presented in Section 6.

2 System model

In our system model, we assume that the receiver is equipped with M-sensor UCA antenna. Considering K signals are impinging on the receiver's array, the received signal at an instant of time t can be expressed as

$$\boldsymbol{y}(t) = \sum_{k=1}^{K} \boldsymbol{a}(\phi_k)\boldsymbol{s}_k(t) + \boldsymbol{w}(t), \tag{1}$$

where $\boldsymbol{a}(\phi_k)$ is the steering vector for the signal arriving at azimuth angle ϕ_k, $\boldsymbol{s}_k(t)$ is the impinging signal from the kth source at time t, and $\boldsymbol{w}(t)$ is the additive white Gaussian noise (AWGN). In the matrix notation, (1) can be represented as

$$\boldsymbol{Y} = \boldsymbol{AS} + \boldsymbol{W}, \tag{2}$$

where $\boldsymbol{Y} \in \mathbb{C}^{M \times N}$, $\boldsymbol{A} \in \mathbb{C}^{M \times K}$, $\boldsymbol{S} \in \mathbb{C}^{K \times N}$, $\boldsymbol{W} \in \mathbb{C}^{M \times N}$, with N being the total number of collected samples and $\mathbb{C}$ is the set of complex numbers. The matrix of steering vectors is

$$\boldsymbol{A} = [\boldsymbol{a}(\phi_1), \boldsymbol{a}(\phi_2), \ldots, \boldsymbol{a}(\phi_k)]\,. \tag{3}$$

The steering vector $\boldsymbol{a}(\boldsymbol{f}_k)$ for a uniform circular array (UCA) can be represented as

$$\boldsymbol{a}(\phi_k) = \mathrm{e}^{(2\pi/\eta r \sin(\theta)(\cos(\phi - \gamma)))}, \tag{4}$$

with waveform η, radius r and γ is $360/N * (0{:}N-1)$, θ is the elevation angle which we assume to be orthogonal to the array and hence $\sin(\theta)$ is 1 for the remaining of this paper. The CorrM of the received data can be expressed as

$$\boldsymbol{R}_{YY} = \mathbb{E}\big[\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{H}}\big] = \boldsymbol{A}\boldsymbol{R}_{SS}\boldsymbol{A}^{\mathrm{H}} + \boldsymbol{R}_{WW} \tag{5}$$

where $\mathbb{E}[\cdot]$ denotes the expectation operation, H denotes the Hermitian operation, $\boldsymbol{R}_{SS}$ is the correlation matrix of the impinging signal, $\boldsymbol{R}_{WW} = \sigma^2\boldsymbol{I}$ is the auto covariance matrix of the receiver's AWGN with σ^2 is the noise variance and $\boldsymbol{I}$ is $M \times M$ identity matrix. The received signal is assumed to be non-zero mean as in [24, 26]. In practice, the sample CorrM is estimated instead of the CorrM. We express $\boldsymbol{R}'_{YY}$ as the CorrM of N observations as

$$\boldsymbol{R}'_{YY} = \frac{1}{N}\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{H}} \tag{6}$$

where $\boldsymbol{R}'_{YY}$ converges to $\boldsymbol{R}_{YY}$ for large number of samples.

3 Proposed algorithms

As mentioned earlier, we propose four algorithms to estimate the number of sources in which we categorise them into CorrM algorithms and CoefM algorithms [Note that for stochastic signals with zero mean, the two measures will collapse into a single method.]. We propose to use MI and MS as the metrics to estimate the number of sources. We first formulate the problem and define CorrM and CoefM matrices. Then, we introduce our two decision statistics. We present the intuition behind our proposed algorithms. We then present our four proposed algorithms categorised into two subsections, one for each matrix category.

3.1 Problem formulation

3.1.1 CorrM matrix: We first apply and EVD on $\boldsymbol{R}'_{YY}$, which leads to

$$\boldsymbol{R}'_{YY} = \boldsymbol{U}_Y\boldsymbol{\Lambda}_{\boldsymbol{Y}}\boldsymbol{U}_Y^{\mathrm{H}} = \boldsymbol{U}_S\boldsymbol{\Lambda}_S\boldsymbol{U}_S^{\mathrm{H}} + \boldsymbol{U}_W\boldsymbol{\Lambda}_W\boldsymbol{U}_W^{\mathrm{H}}, \tag{7}$$

where $\boldsymbol{U}_S$ and $\boldsymbol{U}_W$ are signal and noise subspaces unitary matrices, respectively, and $\boldsymbol{\Lambda}_S$ and $\boldsymbol{\Lambda}_W$ are diagonal matrices of the eigenvalues of the signal and noise, respectively. $\boldsymbol{\Lambda}_Y$ can be expressed as

$$\begin{aligned}\boldsymbol{\Lambda}_Y &= \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_M)\\ &= \mathrm{diag}(0, \ldots, 0, \lambda_1, \lambda_2, \ldots, \lambda_K) + \sigma^2\boldsymbol{I},\end{aligned} \tag{8}$$

where $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_K$. The total eigenvalues $(\lambda_1, \lambda_2, \ldots, \lambda_M)$ with their corresponding eigenvectors $(\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_M)$ define the noise and signal subspaces as $\boldsymbol{U}_W = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{M-K}]$ and $\boldsymbol{U}_S = [\boldsymbol{e}_{M-K+1}, \ldots, \boldsymbol{e}_M]$, respectively. The problem then becomes estimating the value of K, i.e. the number of impinging signals, given the estimated $(\lambda_1, \lambda_2, \ldots, \lambda_M)$.

3.1.2 CoefM matrix: Two of our proposed techniques exploit CoefM rather than CorrM to estimate the number of impinging sources. To define CoefM, we first define the covariance matrix as

$$\begin{aligned}\boldsymbol{V}_{YY} &= \mathbb{E}\big[(\boldsymbol{Y} - \boldsymbol{\mu}_Y)(\boldsymbol{Y} - \boldsymbol{\mu}_Y)^{\mathrm{H}}\big]\\ &= \boldsymbol{A}\boldsymbol{R}_{SS}\boldsymbol{A}^{\mathrm{H}} + \boldsymbol{R}_{WW} - \boldsymbol{\mu}_Y\boldsymbol{\mu}_Y^{\mathrm{H}},\end{aligned} \tag{9}$$

where $\boldsymbol{\mu}_Y = \mathbb{E}[\boldsymbol{Y}]$. The elements in the diagonal of $\boldsymbol{V}_{YY}$ are the variances of $\boldsymbol{Y}$. With $\boldsymbol{\mu}'_Y$ being the sample mean, the sample version of $\boldsymbol{V}_{YY}$ is then given by

$$\boldsymbol{V}'_{YY} = \frac{1}{N}(\boldsymbol{Y} - \boldsymbol{\mu}'_Y)(\boldsymbol{Y} - \boldsymbol{\mu}'_Y)^{\mathrm{H}}\,. \tag{10}$$

CoefM is then given by

$$\boldsymbol{C}_{YY} = \left(\mathrm{diag}(\boldsymbol{V}'_{YY})\right)^{-(1/2)}\boldsymbol{V}'_{YY}\left(\mathrm{diag}(\boldsymbol{V}'_{YY})\right)^{-(1/2)}. \tag{11}$$

We then apply the EVD on $\boldsymbol{C}_{YY}$ which leads to

$$\boldsymbol{C}_{YY} = \boldsymbol{U}_C\boldsymbol{\Lambda}_C\boldsymbol{U}_C^{\mathrm{H}}, \tag{12}$$

where

$$\begin{aligned}\boldsymbol{\Lambda}_C &= \mathrm{diag}\left(\lambda_1^C, \lambda_2^C, \ldots, \lambda_M^C\right)\\ &= \mathrm{diag}\left(0, \ldots, 0, \lambda_1^C, \lambda_2^C, \ldots, \lambda_K^C\right)\\ &\quad + \left(\mathrm{diag}(\boldsymbol{V}'_{YY})\right)^{-(1/2)}\left(\sigma^2\boldsymbol{I} - \boldsymbol{\mu}_Y\boldsymbol{\mu}_Y^{\mathrm{H}}\right)\left(\mathrm{diag}(\boldsymbol{V}'_{YY})\right)^{-(1/2)},\end{aligned} \tag{13}$$

where $\lambda_1^C \le \lambda_2^C \le \cdots \le \lambda_K^C$. The total eigenvalues $(\lambda_1^C, \lambda_2^C, \ldots, \lambda_M^C)$ with their corresponding eigenvectors $(\boldsymbol{e}_1^C, \boldsymbol{e}_2^C, \ldots, \boldsymbol{e}_M^C)$ define the noise and signal subspaces. The problem is then estimating the value of K given the estimated $(\lambda_1^C, \lambda_2^C, \ldots, \lambda_M^C)$.

3.2 Proposed decision statistics

In our algorithms, we use two decision statistics, namely MI and MS. We first arrange the eigenvalues in an ascending order, rather than a descending order as in the case of AIC and MDL. Hence, eigenvalues are arranged from the beginning as $(\lambda_1, \lambda_2, \ldots, \lambda_M)$, where $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_M$ and $(\lambda_1, \lambda_2, \ldots, \lambda_{M-K})$ are in the noise subspace while $(\lambda_{M-K+1}, \ldots, \lambda_M)$ are in the signal subspace.

3.2.1 MI decision statistic: The first decision statistic is the MI, $\boldsymbol{\delta}$, which is simply the difference between two consecutive eigenvalues. It can be expressed as

$$\delta_i = \lambda_i - \lambda_{i-1} \quad \text{for} \quad i = 2, 3, \ldots, M\,. \tag{14}$$

3.2.2 MS decision statistic: Our second proposed decision statistic used as a metric to decide on the number of sources is the moving standard deviation of the estimated eigenvalues, $\boldsymbol{\alpha}$. The sample standard deviation is calculated by

$$s_M = \sqrt{\frac{1}{M-1}\sum_{i=1}^{M}(x_i - u)^2}, \tag{15}$$

where x_i is the value of the ith sample and u is the mean. The sample standard deviation of two consecutive eigenvalues can be given by

$$\mathrm{STD}(i) = \sqrt{(\lambda_i - u)^2 + (\lambda_{i-1} - u)^2}, \tag{16}$$

where u is the mean of the two eigenvalues calculated as

$$u = \frac{\lambda_i + \lambda_{i-1}}{2}\,. \tag{17}$$

We define our second decision statistic α as the difference between two consecutive STDs

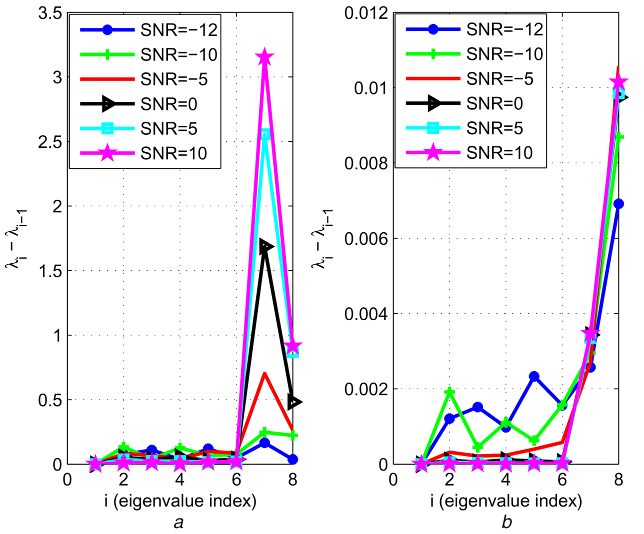

Fig. 1 *MI of the estimated eigenvalues for $M = 8$, $K = 2$ and $N = 1024$ at different SNR levels for*
(a) CoefM, ***(b)*** CorrM

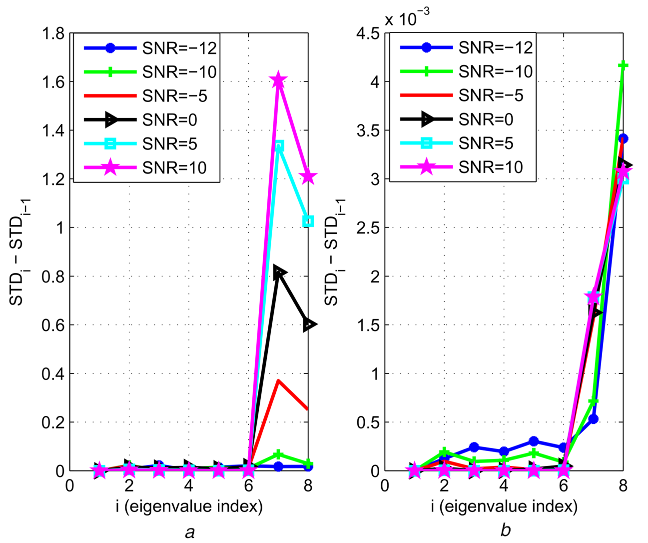

Fig. 2 *MS of the estimated eigenvalues for $M = 8$, $K = 2$ and $N = 1024$ at different SNR levels for*
(a) CoefM, ***(b)*** CorrM

$$\alpha_i = \mathrm{STD}(i) - \mathrm{STD}(i-1), \quad \text{for } i = 3, 4, \ldots, M. \tag{18}$$

Proposition 1: α_i can be rewritten as

$$\alpha_i = \frac{1}{\sqrt{2}}(\lambda_i - 2\lambda_{i-1} + \lambda_{i-2}). \tag{19}$$

Proof: The proof for Proposition 1 is obtained by substituting (17) into (16) and with some simple algebraic manipulation, we have

$$\mathrm{STD}(i) = \frac{\lambda_i - \lambda_{i-1}}{\sqrt{2}}. \tag{20}$$

Applying the same for $\mathrm{STD}(i-1)$ and substituting into (18), α_i can be written as in (19). □

3.3 Proposed algorithm principle

The objective of our algorithm is to estimate the point at which the eigenvalues move from the noise subspace to the signal subspace; hence the intuition behind exploiting MI and MS of the eigenvalues. MS is a measure of how disperse the eigenvalues are.

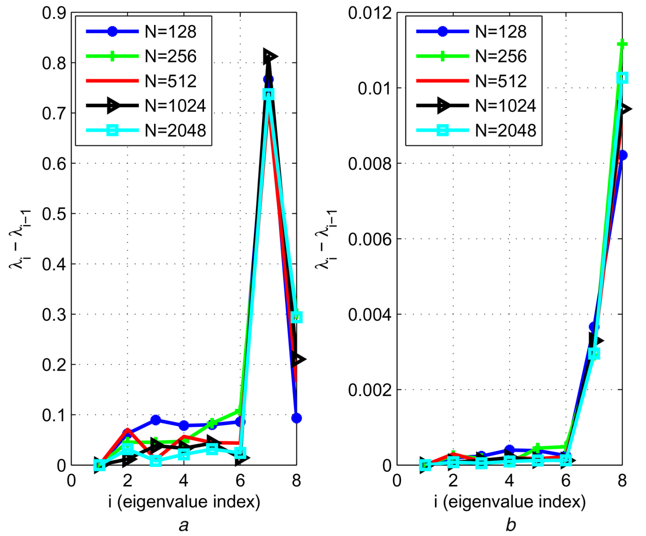

Fig. 3 *MI of the estimated eigenvalues for $M = 8$, $K = 2$ and $SNR = -7$ dB at different number of samples for*
(a) CoefM, ***(b)*** CorrM

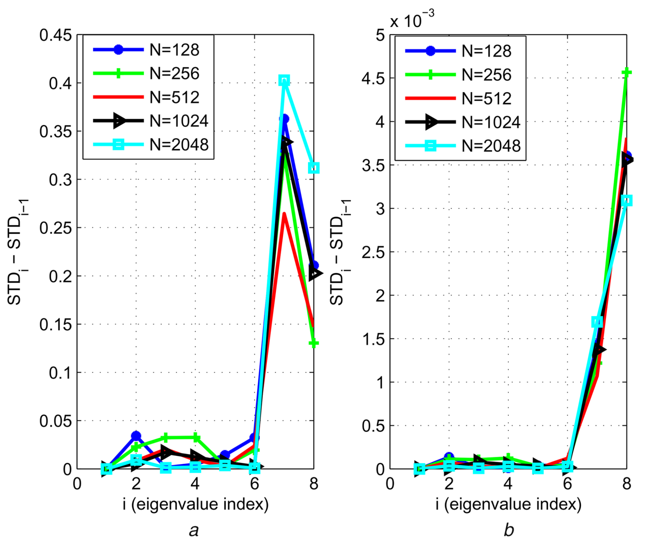

Fig. 4 *MS of the estimated eigenvalues for $M = 8$, $K = 2$ and $SNR = -7$ dB at different number of samples for*
(a) CoefM, ***(b)*** CorrM

Therefore, it is expected to increase at the transition point between the noise and signal subspaces. It can be inferred from (8) and (13) that the values of sources' signal eigenvalues are expected to be considerably higher than noise eigenvalues at moderate and high SNR levels. At the same time, the noise eigenvalues are expected to be comparable to one another. We plot MI and MS of the estimated eigenvalues of CoefM versus CorrM for different SNR levels in Figs. 1 and 2 and different number of collected samples in Figs. 3 and 4. The simulation parameters in Figs. 1 and 2 are circular antenna array with $M = 8$, $K = 2$ and $N = 1024$ samples at different SNR levels. Same simulation parameters are used in Figs. 3 and 4 except that the SNR was kept fixed at -7 dB and N changed from 128 to 2048 samples.

The main difference between using CoefM in (13) and CorrM in (8) is that the signal eigenvalues of CoefM are closer to one another than CorrM signal eigenvalues. Therefore, the change in both MI and MS of CoefM eigenvalues when first moving from the noise subspace to the signal subspace, i.e. between eigenvalue indices 6 and 7, is always the highest. MI and MS then start to decrease. In other words, the highest increment in both MI and MS always happens when moving from the noise subspace to the signal subspace. This implies that when using MI or MS of CoefM, the problem is transformed into a simple maximisation problem, where the index at which the highest increment occurs is searched for.

On the contrary, MI and MS of CorrM eigenvalues consistently increase when moving from the noise subspace to the signal subspace. Hence, a threshold needs to be set to compare MI or MS to. This threshold can be estimated either with a configurable parameter as in traditional threshold based techniques [9]. The configurable parameter requires extensive simulation a priori to be set. We propose a simple single formula to estimate the threshold for CorrM eigenvalues below.

3.4 CoefM based algorithms

For MI of the CoefM eigenvalues, the index at which the shift from the noise subspace to the signal's subspace occurs can be estimated as

$$j = \arg\max_i \delta_i . \tag{21}$$

In this case, the estimated number of sources can be given by $\hat{K} = M - j + 1$. Similarly, for MS of CorrM eigenvalues, expressed in (19), the index at which the highest increment occurs can be estimated as

$$l = \arg\max_i \alpha_i \tag{22}$$

Consequently, the number of sources can be given by $\hat{K} = M - l + 1$.

3.5 CorrM based algorithms

As stated earlier, MI and MS of CorrM eigenvalues should be compared to preset thresholds in order to estimate the number of sources. For MI case, we denote the threshold by ξ_n. At index, i, at which $\delta_i \geq \xi_n$, the number of sources is estimated as $\hat{K} = M - i + 1$. To estimate the threshold ξ_n, the probability distribution of δ_i has to be derived first. The probability of false alarm, P_f, is defined as

$$P_\mathrm{f} = \Pr\left(\delta_i \geq \xi_n | i \leq M - K\right). \tag{23}$$

Thus, for a given P_f, ξ_n can be estimated accordingly. In case of noise only, i.e. $K = 0$, the received samples follow $\mathcal{N}(0, \sigma^2)$, and therefore $\boldsymbol{R}'_{YY}$ follows a Wishart distribution with N degrees of freedom and variance Σ, i.e. $\boldsymbol{R}'_{YY}$ follows $\mathcal{W}(N, \Sigma)$ [27]. The empirical distribution function of the noise eigenvalues can be expressed by [16, 27]

$$F^R(\lambda) = \frac{\text{number of eigenvalues of } \boldsymbol{R}'_{YY} \leq \lambda}{M} \tag{24}$$

according to [16], F^R converges to f^W with high probability when the number of samples $N \to \infty$. f^W follows a Marcenko–Pastur density function [28], which can be expressed as

$$\begin{aligned} f_W(\lambda) = dF^W(\lambda) = {} & \max\left(0, (1 - G)\right)\delta_d(\lambda) \\ & + \frac{\sqrt{(\lambda - a_-)}\sqrt{(a_+ - \lambda)}}{2\pi\sigma\lambda(1/G)}\Pi_{[a_-, a_+]}(\lambda), \end{aligned} \tag{25}$$

where G is the samples to elements ratio (N/M), $a_\pm = \sigma(1 \pm 1/\sqrt{G})^2$, $\delta_d(\lambda)$ is the Dirac delta function, and the function $\Pi_{[a,b]}(\lambda)$ has the value 1 for $a \leq \lambda \leq b$ and 0 otherwise. We derive the probability distribution of the noise eigenvalues of δ_i in Proposition 2.

Proposition 2: The probability distribution of δ_i can be given by (see (26))

Proof: The proof for Proposition 2 is provided in the Appendix. □

Consequently, (23) can be rewritten as

$$P_\mathrm{f} = \Pr\left(\delta_i \geq \xi_n | i \leq M - K\right) = 1 - F_{\delta_i}(\xi_n). \tag{27}$$

Conventionally, in order to estimate the threshold ξ_n, (27) has to be solved for a desired P_f. This requires calculating the integral in (26), which is highly prohibitive due to the following:

i. For a given P_f, it requires the total number of noise eigenvalues ($M - K$) to be known a priori in order to solve for ξ_n. This is not feasible since the total number of sources (K) is our unknown to be estimated.
ii. Another approach is to try to minimise P_f by differentiating it twice, once with respect to K and another with respect to ξ_n. Then equating the outputs to zero to solve for the two unknowns. This, as can be seen from (26), is a mathematically tedious process and will have a high computational complexity.

Alternatively, we estimate the threshold in (23) through multiple linear regression, which is a popular machine learning approach [29, 30]. Among the class of multiple linear regression techniques, we use the least square fitting. It should be noted that other machine learning techniques, such as logistic regression or neural networks [31], can results in less error in the threshold function, however, we choose linear regression due to its simplicity and applicability in our context. Analogous to the approaches in [20–22], we find the threshold through extensive simulation results. However, we empirically [Extensive simulation data is collected for all possible scenarios and regions of interest.] find a formula for the threshold rather than having an adjustment coefficient that changes with the change of any parameter the threshold depends on such as the number of samples or SNR.

A multiple linear regression model for ξ_n is defined as

$$\xi_n = \eta_0 + \sum_j \eta_j Z_j, \tag{28}$$

where the dependent variable is the threshold that we need to model having multiple other independent variables Z_j, such as SNR value, number of samples, number of antenna elements and the received power and η_j ($j = 0, 1, 2, \ldots$) are the parameters to be estimated through multiple linear regression. This model approximates the threshold in one equation using least square fitting by minimising the sum of the squares of the residual between the observed value and the fitted value [30–32].

Using the previous concept in Section 3, depicted in Figs. 1*B* and 3*B*, the threshold can be defined with multiple regions using

$$\xi_n = P_\mathrm{s} \times \begin{cases} \sigma^2, & \text{for} \quad \mathrm{SNR} > 2\,\mathrm{dB} \\ \sigma^2/8, & \text{for} \quad N > 10000 \\ \sigma^2/2, & \text{for} \quad \mathrm{SNR} \geq 2\,\mathrm{dB} \quad N < 100 \\ \sigma^2/6, & \text{for} \quad \mathrm{SNR} \leq 6\,\mathrm{dB} \quad N > 1000 \\ \sigma^2/4, & \text{for} \quad \text{elsewhere} \end{cases} \tag{29}$$

where P_s is the power of the received signal. Taking some samples from those regions and applying the concept of multiple linear least squares regression [33] to estimate the threshold function, the threshold can be defined by

$$\xi_n = P_\mathrm{s}(-7.75 \times 10^{-3} \cdot \mathrm{SNR} - 3.77 \times 10^{-5} \cdot N + 1.05) \tag{30}$$

This threshold takes into consideration specific regions of interest that include: SNR values $\mathrm{SNR} \in [-20\,\mathrm{dB}{:}40\,\mathrm{dB}]$, received power $P_\mathrm{s} \in [0\,\mathrm{dBm}{:} -100\,\mathrm{dBm}]$, number of samples $N \in [2^6{:}2^{14}]$, and number of antenna array element of <30. Outside these regions the threshold might fail to estimate the number of sources correctly and hence requires further analysis. As a matter of fact, it is difficult for these parameters to be outside the specified regions in practical scenarios. For MS of CorrM eigenvalues, we denote the threshold by γ_n. At index, i, at which $\alpha_i \geq \gamma_n$, the number of sources is estimated as $\hat{K} = M - i + 1$. Another technique is to use MS instead of MI and set a threshold to distinguish between noise and signal eigenvalues. Hence, the decision is taken on the number of sources: The probability of false alarm for MS case can be given by

$$P_{\mathrm{f}} = \Pr\left(\alpha_i \geq \gamma_n | i \leq M - K\right). \tag{31}$$

As in the MI case, in order to estimate the threshold γ_n, the probability distribution function $F_{\alpha_i}(\gamma_n)$ has to be expressed first, then solve (31) for a given P_{f}. As a matter of fact, $F_{\alpha_i}(\gamma_n)$ is expected to be more complicated to solve than $F_{\delta_i}(\xi_n)$. Again, we find the threshold in (31) through a linear least squares regression fitting approach [33] that takes into consideration the number of samples, the number of elements and the received power. Using the previous concept in Section 3, depicted in Figs. 2*B* and 4*B*, the threshold can be defined with multiple regions using

$$\gamma_n = \begin{cases} 0.45P_{\mathrm{s}}, & \text{for} \quad 500 < N < 1000 \quad M \leq 8 \\ P_{\mathrm{s}}, & \text{for} \quad 100 < N < 500 \quad M > 8 \\ 0.75P_{\mathrm{s}}, & \text{elsewhere} \end{cases} \tag{32}$$

Taking some samples from those regions and using a linear least squares regression to estimate the threshold function, the threshold can be defined by

$$\gamma_n = P_{\mathrm{s}}(2.5 \times 10^{-5} \cdot N - 0.12 \times 10^{-4} \cdot M + 0.66). \tag{33}$$

This threshold takes into consideration specific regions of interest including received power $P_{\mathrm{s}} \in [0\,\mathrm{dBm}: -100\,\mathrm{dBm}]$, number of samples $N \in [2^6 : 2^{14}]$, and number of antenna array elements of <30. Outside these regions the threshold might fail to estimate the number of sources correctly and hence needs further analysis.

4 Existing techniques

AIC and MDL are the most widely used number of sources estimation techniques. They are model order selection information theoretic approaches that use the eigenvalues of the sample covariance matrix to determine how many smallest eigenvalues are approximately equal. These eigenvalues lie in the noise subspace, while others lie in the signal subspace. Both algorithms involve minimising a criterion of log-likelihood over the number of signals that are detectable. We compared our algorithms to traditional AIC and MDL as well as other new modified versions of AIC and MDL, which were presented in the recent literature. Namely, we compared to corrected AIC, $\mathrm{AIC}_{\mathrm{C3}}$ presented in [17], RMT_AIC presented in [16], LS_MDL presented in [19] and traditional AIC and MDL presented in [13]. Table 1 summarises the criteria for each algorithm, where the covariance matrix eigenvalues are ordered in a descending order, i.e. $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M$.

5 Simulation results

We evaluate the performance of our algorithms versus existing algorithms of Table 1 in different scenarios. The simulated scenarios include different SNR levels, different number of samples, different number of sources and different number of array elements. Performance metric used for comparison is the percentage error rate, which can be expressed as

$$\text{error rate} = \left(1 - \frac{\text{number of successes}}{\text{number of runs}}\right) \times 100. \tag{34}$$

Except for different array element simulation results, the array that is used is a UCA with $M = 8$ elements. It is worth noting that our proposed algorithms are independent of the antenna array formation. Rather, they can operate with any formation, linear, circular, planar or any antenna array formation where the steering vector of the antenna array can be estimated either analytically or experimentally. We estimate the error rate through 10,000 iterations.

For simplicity and arrangement of figures, Table 2 represents the notations of the legends that are used in all the following figures. It shows the notations for our proposed algorithms only, while other algorithms notations are the same as given in Table 1.

$$\begin{aligned} F_{\delta_i}(\xi_n) = {} & \left(\frac{G}{2\pi\sigma}\right)^i \int_{-\infty}^{\xi_n} \frac{(M-K)!}{(i-2)!(M-K-i)!} \int_{a_-}^{a_+} \frac{\sqrt{(\lambda - a_-)}\sqrt{(a_+ - \lambda)}}{\lambda} \\ & \times \left\{ \frac{1}{4\sqrt{a_-}\sqrt{a_+}} \left[2\arcsin\left(\frac{-2\lambda + a_- + a_+}{a_- - a_+} \right) a_-^{3/2}\sqrt{a_+} \right.\right. \\ & + 2\arcsin\left(\frac{-2\lambda + a_- + a_+}{a_- - a_+} \right) a_+^{3/2}\sqrt{a_-} \\ & + 4a_+a_-\arctan\left(\frac{1}{2} \frac{2a_-a_+ - a_+\lambda - a_-\lambda}{\sqrt{a_-}\sqrt{a_+}\sqrt{(a_+ - \lambda)}\sqrt{(\lambda - a_-)}} \right) \\ & + \pi a_-^{3/2}\sqrt{a_+} + \pi\sqrt{a_-}a_+^{3/2} - 2a_-a_+\pi \\ & \left.\left. + 4\sqrt{a_-}\sqrt{a_+}\sqrt{\lambda - a_-}\sqrt{a_+ - \lambda} \right] \right\}^{i-2} \left(\frac{\sqrt{(\lambda + \delta) - a_-}\sqrt{a_+ - (\lambda + \delta)}}{\lambda + \delta} \right) \\ & \times \left(1 - \frac{G}{2\pi\sigma} \left\{ \frac{1}{4\sqrt{a_-}\sqrt{a_+}} \left[2\arcsin\left(\frac{-2(\lambda + \delta) + a_- + a_+}{a_- - a_+} \right) a_-^{3/2}\sqrt{a_+} \right.\right.\right. \\ & + 2\arcsin\left(\frac{-2(\lambda + \delta) + a_- + a_+}{a_- - a_+} \right) a_+^{3/2}\sqrt{a_-} + 4a_+a_- \\ & \arctan\left(\frac{1}{2} \frac{2a_-a_+ - a_+(\lambda + \delta) - a_-(\lambda + \delta)}{\sqrt{a_-}\sqrt{a_+}\sqrt{(a_+ - (\lambda + \delta))}\sqrt{((\lambda + \delta) - a_-)}} \right) \\ & + \pi a_-^{3/2}\sqrt{a_+} + \pi\sqrt{a_-}a_+^{3/2} - 2a_-a_+\pi \\ & \left.\left.\left. + 4\sqrt{a_-}\sqrt{a_+}\sqrt{(\lambda + \delta) - a_-}\sqrt{a_+ - (\lambda + \delta)} \right] \right\} \right)^{M-K-i} \mathrm{d}\lambda\, \mathrm{d}\delta \end{aligned} \tag{26}$$

Table 1 Criterion functions of RMT_AIC, AIC_{C3}, LS_MDL, AIC and MDL

Abbreviation	Algorithm
AIC [13]	$\hat{K} = \arg\min_k \left(2N(M-k)\log\left(\frac{(1/M-k)\sum_{i=k+1}^{M}\lambda_i}{\Pi_{i=k+1}^{M}\lambda_i^{(1/M-k)}}\right) + 2k(2M-k)\right)$
MDL [13]	$\hat{K} = \arg\min_k \left(N(M-k)\log\left(\frac{(1/M-k)\sum_{i=k+1}^{M}\lambda_i}{\Pi_{i=k+1}^{M}\lambda_i^{(1/M-k)}}\right) + \frac{1}{2}k(2M-k)\log(N)\right)$
LS_MDL [19]	$\hat{K} = \arg\min_k \left(N(M-k)\log\left(\frac{(1/M-k)\sum_{i=k+1}^{M}\varsigma_i^{(k)}}{\left(\Pi_{i=k+1}^{M}\varsigma_i^{(k)}\right)^{(1/M-k)}}\right) + \frac{1}{2}k(k-1)\log(N)\right)$ $\varsigma_i^{(k)} = \beta^{(k)}\tau^{(k)} + (1-\beta^{(k)})\lambda_i \quad \text{and} \quad i = k+1:M$ $\tau^{(k)} = \frac{1}{M-k}\sum_{i=k+1}^{M}\lambda_i \quad \text{and} \quad \beta^{(k)} = \max(1, \nu^{(k)})$ $\nu^{(k)} = \frac{\sum_{i=k+1}^{M}\lambda_i^2 + \left(\sum_{i=k+1}^{M}\lambda_i\right)^2}{N+1\left(\sum_{i=k+1}^{M}\lambda_i^2 - \left(\sum_{i=k+1}^{M}\lambda_i\right)^2/m-k\right)}$
AIC_{C3} [17]	$\hat{K} = \arg\min_k \left(2N(M-k)\log\left(\frac{(1/M-k)\sum_{i=k+1}^{M}\lambda_i}{\Pi_{i=k+1}^{M}\lambda_i^{(1/M-k)}}\right) + \frac{2(v_k+1)(\hat{m}+v_k+2)}{\hat{m}-v_k-2}\right)$ $v_k = k(2M-k) \quad \text{and} \quad \hat{m} = 2MN$
RMT_AIC [16]	$\hat{K} = \arg\min_k \left(\frac{N}{2}\left(\frac{\sum_{i=k+1}^{M}\lambda_i^2}{\left((1/M-k)\sum_{i=k+1}^{M}\lambda_i\right)^2}\right)^2 + 2k\right)$

Table 2 Figures legend abbreviations

Abbreviation	Algorithm
MS_{COEF}	proposed MS for CoefM algorithm
MI_{COEF}	proposed MI for CoefM algorithm
MS_{CORR}	proposed MS for CorrM algorithm
MI_{CORR}	proposed MI for CorrM algorithm

5.1 Algorithms' performance at various SNR levels

We first evaluate the performance of our proposed algorithms at various SNR levels. In our simulation, SNR values ranged from −20 to 15 dB, while the number of samples was fixed to $N = 1024$ samples and the actual number of sources was $K = 2$.

Fig. 5*a* shows the error rate for our proposed algorithm as well as the algorithms presented in Table 1 versus SNR. As shown in Fig. 5*a*, the proposed algorithms outperform MDL algorithms at low SNR levels and outperform AIC and AIC_{C3} at high SNR levels. For SNR ≤ 10 dB, the proposed algorithms have a comparable performance to AIC and better performance than MDL, with MI_{COEF} achieving the least error rate at SNR ≤ 15 dB. For SNR ≥ 10 dB, the performance of MDL, LS_MDL, RMT_AIC and the proposed algorithms is indistinguishable with a minimum error rate, i.e. almost 0%, while AIC and AIC_{C3} kept their error rate at about 10%. The reason why AIC and AIC_{C3} are not achieving lower error rate is the overestimation of the number of sources, which tends to occur at relatively high SNR levels. This overestimation is due to the added penalty term as presented in [34].

In Fig. 5*b*, we plot the error rate for all algorithms using the same simulation setting as shown in Fig. 5*a*, except at a low number of samples of $N = 100$. CoefM based algorithms outperform others at low SNR levels, i.e. ≤10 dB. This goes back to the better contrast between the signal and the noise eigenvalues that is introduced by exploiting CoefM and hence better number of sources' estimation.

5.2 Algorithms' performance at various number of samples

We evaluate the efficiency of our proposed algorithms in terms of error rate versus number of samples. It is desirable in practical scenarios to process the lower number of samples while achieving an adequate performance. Fig. 6 depicts the error rate for our proposed algorithms as well as the algorithms presented in Table 1 versus the number of samples at SNR = −5 dB and $K = 2$. The *x*-axis is the number of samples represented by $\log_2 N$.

Our proposed algorithms outperform MDL algorithms and have a comparable performance to AIC algorithms at low number of samples. MDL and LS_MDL algorithms underestimate the number of sources when using a low number of samples. As in changing SNR case, MDL, LS_MDL, RMT_AIC and the proposed algorithms have the same performance for $N > 256$ samples, i.e. $N > 2^8$, where the error rates are almost 0%. AIC and AIC_{C3}, on the other hand, overestimate the number of sources and hence had their almost 10% error rate.

5.3 Algorithms' performance with different number of sources

Different algorithms have different sensitivities in terms of the number of sources they can estimate. Hence, in Fig. 7*a*, we plot the error rate of all algorithms against different number of sources at SNR = −5 dB and $N = 1024$. At higher number of sources, i.e. $K > 5$, AIC algorithms outperform others by achieving a slightly lower error rate. However, at $K < 5$, AIC algorithms have a consistent 10% error rate. Moreover, all algorithms fail to achieve an acceptable error rate when $K > 5$. On the other hand, MI_{COEF} has the worst performance among all algorithms, achieving a high error rate at $K > 3$. MDL, LS_MDL, MS_{COEF}, MS_{CORR} and MI_{CORR} have a comparable performance.

5.4 Algorithms' performance at different number of array elements

We then examine the effect of increasing the number of elements that construct the array at SNR = −5 dB, $N = 100$ samples and $K = 2$ in Fig. 7*b*. MS_{COEF}, MI_{CORR} and LS_MDL show the best performance among others for $M > 8$. As M increases the error rate decreases until it approximately approaches 0% for MDL, LS_MDL, RMT_AIC and all proposed algorithms except for MI_{COEF}, which occurs at $M > 12$ elements.

5.5 Complexity comparison

Our proposed algorithms as well as the algorithms presented in Table 1 require an estimation of the CorrM or the CoefM followed by and EVD operation. These two steps have a complexity in the order of $\mathcal{O}\left(M^2N + M^3\right)$ [19]. However, each algorithm has a different criterion function. We compared the required number of

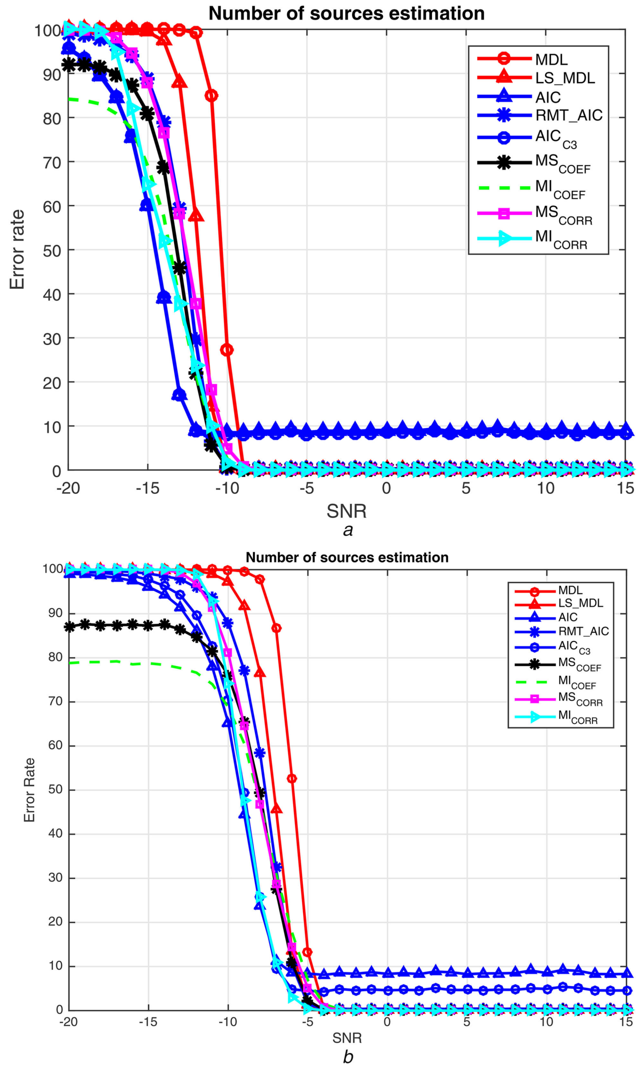

Fig. 5 *Error rate for our proposed algorithm as well as algorithms presented in Table 1 versus SNR at*
(a) $N = 1024$ samples and ***(b)*** $N = 100$ samples

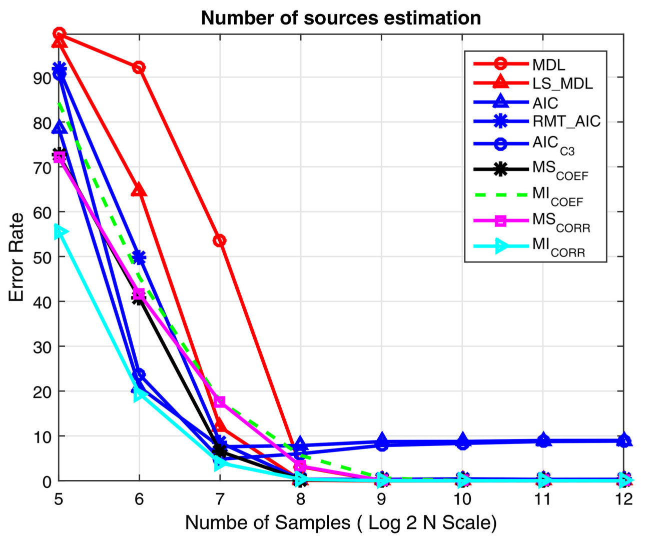

Fig. 6 *Error rate for our proposed algorithm as well as algorithms presented in Table 1 versus number of samples at SNR = −5 dB*

floating point operation (flops) of the criterion function of our proposed algorithms to AIC and MDL in Table 3. It is assumed that each multiplication, addition or comparison requires one flop. Our algorithms require less number of flops to implement the criterion function than AIC and MDL. The complexity of our

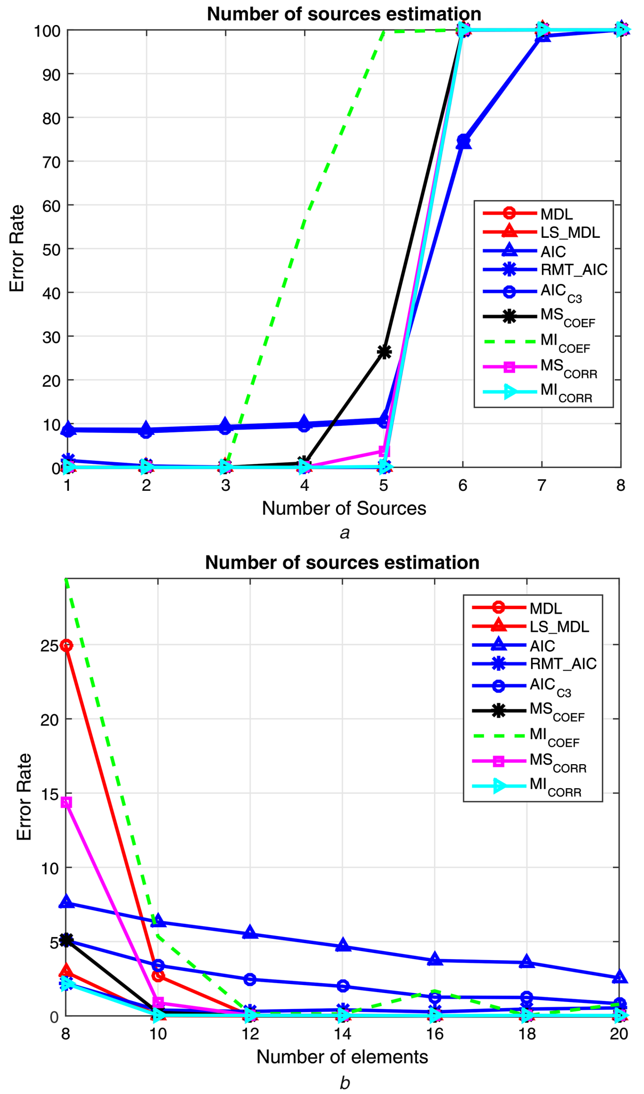

Fig. 7 *Error rate for our proposed algorithm as well as algorithms presented in Table 1*
(a) Versus different number of sources at SNR = −5, $N = 1024$ and $M = 8$ and ***(b)*** Versus number of antenna array elements at $N = 100$ samples

Table 3 Required number of flops for our proposed algorithms, AIC and MDL to implement the criterion function

Algorithm	Number of flops	Complexity
MS_{COEF}	$4M - 9$	$\mathcal{O}(M)$
MI_{COEF}	$2M - 3$	$\mathcal{O}(M)$
MS_{CORR}	$4M + 2$	$\mathcal{O}(M)$
MI_{CORR}	$2M + 4$	$\mathcal{O}(M)$
AIC [35]	$\frac{M^2 + M}{2} + 6M$	$\mathcal{O}(M^2)$
MDL [36]	$M^2 + 16M$	$\mathcal{O}(M^2)$

proposed algorithms is in the order of $\mathcal{O}(M)$, while the complexity of AIC and MDL is $\mathcal{O}(M^2)$.

We plot the simulation run-time for the criterion functions in Fig. 8 versus the number of antenna elements. Our proposed algorithms have drastically improved the simulation run-time, while AIC and MDL algorithms are having a comparable run-time. Our CoefM algorithms achieve simulation run-time that is <5% of that achieved using LS_MDL at low number of antenna elements and <2% at higher number of antenna elements, while our CorrM algorithms are achieving simulation run-time that is <25% of that achieved using LS_MDL at low number of antenna elements and <10% at higher number of antenna elements.

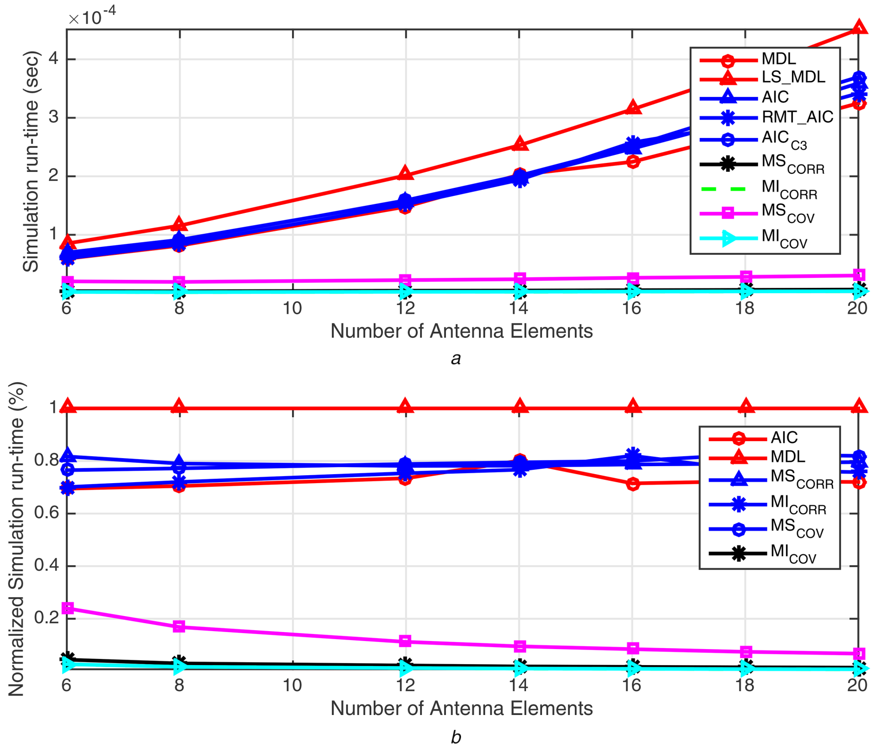

Fig. 8 *Simulation run-time versus number of antenna elements*
(a) Actual run-time in seconds, ***(b)*** Run-time normalised to LS_MDL run-time

6 Conclusion

In this paper, we presented four new techniques for number of sources estimation based on the eigenvalues decomposition. The proposed algorithms operate on the estimated eigenvalues of CorrM and CoefM. Simple decision statistics were presented to estimate the number of sources. CorrM based algorithms use a threshold, which is derived using least squares linear regression fitting, while CoefM algorithms rely on a simple search for maximum value of the decision statistic, which makes them more robust and suitable for hardware implementation and practical scenarios. Results showed that our algorithms have better performance in comparison to MDL at low SNR and number of samples conditions and better than AIC at high SNR conditions. Moreover, the complexity of the criterion functions of our proposed algorithms has lower computational complexity than those for the information theoretic ones.

7 References

[1] Tuncer, T., Friedlander, B.: '*Classical and modern direction-of-arrival estimation*' (Elsevier Science, 2009). Available at: https://books.google.com.qa/books?id=1aQbxKJI2CsC

[2] Najjar-Atallah, L., Marcos, S.: 'Subspace-based approach for doa estimation using pilot symbol channel identification', *IEE Proc., Vis. Image Signal Process.*, 2005, **152**, (1), pp. 20–28

[3] Bell, A.J., Sejnowski, T.J.: 'An information-maximization approach to blind separation and blind deconvolution', *Neural Comput.*, 1995, **7**, (6), pp. 1129–1159

[4] Liavas, A., Regalia, P., Delmas, J.-P.: 'Blind channel approximation: effective channel order determination', *IEEE Trans. Signal Process.*, 1999, **47**, (12), pp. 3336–3344

[5] Schmidt, R.: '*A signal subspace approach to multiple emitter location and spectral estimation*' (Stanford University, 1981). Available at: http://books.google.com.qa/books?id=mLKUnQEACAAJ

[6] Baig, N.: '*Estimation of direction of arrival via beamforming*' (Lap Lambert Academic Publishing GmbH KG, 2012). Available at: https://books.google.com.qa/books?id=k2n1ugAACAAJ

[7] Jiang, J.-S., Ingram, M.-A.: 'Robust detection of number of sources using the transformed rotational matrix'. IEEE Wireless Communications and Networking Conf., 2004, WCNC 2004, March 2004, vol. **1**, pp. 501–506

[8] Akaike, H.: 'A new look at the statistical model identification', *IEEE Trans. Autom. Control*, 1974, **19**, (6), pp. 716–723

[9] Jing, W., Jianguo, H., Chengbing, H., *et al.*: 'Detection of the number of signals by a threshold based on peak-to-average ratio'. IEEE Int. Conf. on Signal Processing, Communications and Computing (ICSPCC), 2011, September 2011, pp. 1–4

[10] Di, A., Tian, L.: 'Matrix decomposition and multiple source location'. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, ICASSP '84, March 1984, vol. **9**, pp. 722–725

[11] Gu, J.F., Wei, P., Tai, H.M.: 'Detection of the number of sources at low signal-to-noise ratio', *IET Signal Process.*, 2007, **1**, (1), pp. 2–8

[12] Di, A.: 'Multiple source location – a matrix decomposition approach', *IEEE Trans. Acoust. Speech Signal Process.*, 1985, **33**, (5), pp. 1086–1091

[13] Wax, M., Kailath, T.: 'Detection of signals by information theoretic criteria', *IEEE Trans. Acousti. Speech Signal Process.*, 1985, **33**, (2), pp. 387–392

[14] Cho, C.-M., Djuric, P.: 'Detection and estimation of doa's of signals via Bayesian predictive densities', *IEEE Trans. Signal Process.*, 1994, **42**, (11), pp. 3051–3060

[15] Mariani, A., Giorgetti, A., Chiani, M.: 'Model order selection based on information theoretic criteria: design of the penalty', *IEEE Trans. Signal Process.*, 2015, **63**, (11), pp. 2779–2789

[16] Nadakuditi, R., Edelman, A.: 'Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples', *IEEE Trans. Signal Process.*, 2008, **56**, (7), pp. 2625–2638

[17] Seghouane, A.-K.: 'Asymptotic bootstrap corrections of aic for linear regression models', *Signal Process.*, 2010, **90**, (1), pp. 217–224

[18] Seghouane, A.-K., Bekara, M.: 'A small sample model selection criterion based on kullback's symmetric divergence', *IEEE Trans. Signal Process.*, 2004, **52**, (12), pp. 3314–3323

[19] Huang, L., So, H.: 'Source enumeration via mdl criterion based on linear shrinkage estimation of noise subspace covariance matrix', *IEEE Trans. Signal Process.*, 2013, **61**, (19), pp. 4806–4821

[20] Chen, W., Wong, K.M., Reilly, J.: 'Detection of the number of signals: a predicted eigenthreshold approach', *IEEE Trans. Signal Process.*, 1991, **39**, (5), pp. 1088–1098

[21] Hu, O., Zheng, F., Faulkner, M.: 'Detecting the number of signals using antenna array: a single threshold solution'. Proc. of the Fifth Int. Symp. on Signal Processing and its Applications, 1999, ISSPA '99, 1999, vol. **2**, pp. 905–908

[22] Quinlan, A., Barbot, J.-P., Larzabal, P., *et al.*: 'Model order selection for short data: an exponential fitting test (eft)', *EURASIP J. Appl. Signal Process.*, 2007, **2007**, (1), pp. 201–201. Available at: http://dx.doi.org/10.1155/2007/71953

[23] Lu, Z., Zoubir, A.: 'Source enumeration in array processing using a two-step test', *IEEE Trans. Signal Process.*, 2015, **63**, (10), pp. 2718–2727

[24] Li, T., Nehorai, A.: 'Maximum likelihood direction finding in spatially colored noise fields using sparse sensor arrays', *IEEE Trans. Signal Process.*, 2011, **59**, (3), pp. 1048–1062

Q5 [25] Li, T., Nehorai, A.: 'Maximum likelihood direction-of-arrival estimation of underwater acoustic signals containing sinusoidal and random components', *IEEE Trans. Signal Process.*, 2011, **59**, (11), pp. 5302–5314

[26] Chang, C.-I., Du, Q.: 'Estimation of number of spectrally distinct signal sources in hyperspectral imagery', *IEEE Trans. Geosci. Remote Sens.*, 2004, **42**, (3), pp. 608–619

[27] Xiao, M., Wei, P., Tai, H.-M.: 'Estimation of the number of sources based on hypothesis testing', *J. Commun. Netw.*, 2012, **14**, (5), pp. 481–486

[28] Marenko, V.A., Pastur, L.A.: 'Distribution of eigenvalues for some sets of random matrices', *Mathematics of the USSR-Sbornik*, 1967, **1**, (4), p. 457. Available at: http://stacks.iop.org/0025-5734/1/i=4/a=A01

[29] Hansen, P., Pereyra, V., Scherer, G.: '*Least squares data fitting with applications*', ser. Least Squares Data Fitting with Applications (Johns Hopkins University Press, 2012). Available at: https://books.google.com.qa/books?id=ynMi1ND6YjYC

[30] Kelleher, J., Namee, B., D'Arcy, A.: '*Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies*' (MIT Press, 2015). Available at: https://books.google.com/books?id=3EtQCgAAQBAJ

[31] Hastie, T., Tibshirani, R., Friedman, J.: '*The elements of statistical learning*', ser. Springer Series in Statistics (Springer New York Inc., New York, NY, USA, 2001)

[32] Zhu, X., Goldberg, A.: '*Introduction to semi-supervised learning*', ser. Synthesis Lectures on Artificial Intelligence and Machine Learning (Morgan & Claypool, 2009). Available at: https://books.google.com/books?id=cnhaJrQ0ScAC

[33] Freedman, D.: '*Statistical models: theory and practice*' (Cambridge University Press, 2009). Available at: https://books.google.com.qa/books?id=fWn9BV5Wpf8C

[34] Fishler, E., Grosmann, M., Messer, H.: 'Detection of signals by information theoretic criteria: general asymptotic performance analysis', *IEEE Trans. Signal Process.*, 2002, **50**, (5), pp. 1027–1036

[35] Wu, H.-T., Yang, J.-F., Chen, F.-K.: 'Source number estimators using transformed Gerschgorin radii', *IEEE Trans. Signal Process.*, 1995, **43**, (6), pp. 1325–1333

[36] Xin, J., Zheng, N., Sano, A.: 'Simple and efficient nonparametric method for estimating the number of signals without eigendecomposition', *IEEE Trans. Signal Process.*, 2007, **55**, (4), pp. 1405–1420

[37] David, H., Nagaraja, H.: '*Order statistics*', ser. Wiley Series in Probability and Statistics (Wiley, 2004). Available at: https://books.google.com.qa/books?id=bdhzFXg6xFkC

8 Appendix

8.1 Proof for Proposition 2

The joint probability density function of the difference between two ordered independent random variables, X_r and X_S with $1 \le r < s < n$, $W_{rs} = X_s - X_r$ is [37]

$$f_{W_{rs}}(w_{rs}) = D_{rs}\int_{-\infty}^{\infty} F(x)^{r-1} f(x)(F(x+w_{rs}) \\ -F(x))^{s-r-1} f(x+w_{rs})(1-F(x+w_{rs}))^{n-s}\,\mathrm{d}x\,. \tag{35}$$

In our case, we need to find the joint the probability density function of $\delta_i = \lambda_i - \lambda_{i-1}$, hence, we let $r = i-1$, $s = i$ and $\delta_i = W_{rs}$ and let $D = D_{rs}$. In this way, (35) for $\delta_i = \delta$ and $\lambda_i = \lambda$ can be redefined as

$$f_{\delta_i}(\delta) = D\int_{-\infty}^{\infty} F(\lambda)^{i-2} f(\lambda)(F(\lambda+\delta) \\ -F(\lambda))^{i-i+1-1} f(\lambda+\delta)(1-F(\lambda+\delta))^{M-K-i}\,\mathrm{d}\lambda\,. \tag{36}$$

Since $i - i + 1 - 1 = 0$, the term $(F(\lambda+\delta) - F(x))^{i-i+1-1}$ will be 1 and (36) can be rewritten as

$$f_{\delta_i}(\delta) = D\int_{-\infty}^{\infty} F(\lambda)^{i-2} f(\lambda) f(\lambda+\delta)(1-F(\lambda+\delta))^{M-K-i}\,\mathrm{d}\lambda,$$

where D is a constant defined as

$$\begin{aligned} D &= \frac{(M-K)!}{(r-1)!(s-r-1)!(M-K-s)!} \\ &= \frac{(M-K)!}{(i-2)!(i-i+1-1)!(M-K-i)!} \\ &= \frac{(M-K)!}{(i-2)!(M-K-i)!} \end{aligned} \tag{37}$$

$f(\lambda)$ is defined in (25). On the assumption that G is >1, i.e. $N > M$, the first term in (25) will be cancelled and (25) will be represented by its second term. $F(\lambda)$ can be defined as

$$\begin{aligned} F(\lambda) &= \int_{-\infty}^{\lambda} f(\lambda)\,\mathrm{d}\lambda = \int_{-\infty}^{\lambda} \frac{\sqrt{(\lambda-a_-)}\sqrt{(a_+-\lambda)}}{2\pi\sigma\lambda(1/G)}\Pi_{[a_-,a_+]}(\lambda)\,\mathrm{d}\lambda \\ &= \int_{-\infty}^{\lambda} \frac{G}{2\pi\sigma}\frac{\sqrt{(\lambda-a_-)}\sqrt{(a_+-\lambda)}}{\lambda}\Pi_{[a_-,a_+]}(\lambda)\,\mathrm{d}\lambda \\ &= \frac{G}{2\pi\sigma}\int_{a_-}^{\lambda} \frac{\sqrt{(\lambda-a_-)}\sqrt{(a_+-\lambda)}}{\lambda}\,\mathrm{d}\lambda \end{aligned} \tag{38}$$

Assuming that $0 \le a_- \le \lambda \le a_+$:

$$\begin{aligned} F(\lambda) = \frac{G}{2\pi\sigma}\frac{1}{4}\frac{1}{\sqrt{a_-}\sqrt{a_+}}&\left(2\arcsin\left(\frac{-2\lambda+a_-+a_+}{a_- - a_+}\right)a_-^{3/2}\sqrt{a_+}\right. \\ &+2\arcsin\left(\frac{-2\lambda+a_-+a_+}{a_- - a_+}\right)a_+^{3/2}\sqrt{a_-} \\ &+4a_+a_-\arctan\left(\frac{1}{2}\frac{2a_-a_+ - a_+\lambda - a_-\lambda}{\sqrt{a_-}\sqrt{a_+}\sqrt{(a_+-\lambda)}\sqrt{(\lambda-a_-)}}\right) \\ &+\pi a_-^{3/2}\sqrt{a_+} + \pi\sqrt{a_-}a_+^{3/2} - 2a_-a_+\pi + 4\sqrt{a_-}\sqrt{a_+}\sqrt{\lambda-a_-}\sqrt{a_+} - \end{aligned} \tag{39}$$

By that, (36) can be rewritten as

$$\begin{aligned} {}_{\delta_i}(\delta) &= \frac{(M-K)!}{(i-2)!(M-K-i)!}\int_{-\infty}^{\infty}\left(\frac{G}{2\pi\sigma}\int_{a_-}^{\lambda}\frac{\sqrt{(\lambda-a_-)}\sqrt{(a_+-\lambda)}}{\lambda}\mathrm{d}\right. \\ &\quad\times\left(\frac{G}{2\pi\sigma}\frac{\sqrt{(\lambda-a_-)}\sqrt{(a_+-\lambda)}}{\lambda}\Pi_{[a_-,a_+]}(\lambda)\right) \\ &\quad\times\left(\frac{G}{2\pi\sigma}\frac{\sqrt{(\lambda+\delta)-a_-}\sqrt{a_+-(\lambda+\delta)}}{\lambda+\delta}\Pi_{[a_-,a_+]}(\lambda+\delta)\right) \\ &\quad\times\left(1-\frac{G}{2\pi\sigma}\int_{a_-}^{\lambda+\delta}\frac{\sqrt{(\lambda+\delta-a_-)}\sqrt{(a_+-\lambda-\delta)}}{\lambda+\delta}d\lambda\right)^{M-K-i} \\ &= \frac{(M-K)!}{(i-2)!(M-K-i)!}\left(\frac{G}{2\pi\sigma}\right)^{1+i-2+1} \\ &\quad\times\int_{a_-}^{a_+}\frac{\sqrt{(\lambda-a_-)}\sqrt{(a_+-\lambda)}}{\lambda} \\ &\quad\times\left(\int_{a_-}^{\lambda}\frac{\sqrt{(\lambda-a_-)}\sqrt{(a_+-\lambda)}}{\lambda}\mathrm{d}\lambda\right)^{i-2} \\ &\quad\times\left(\frac{\sqrt{(\lambda+\delta)-a_-}\sqrt{a_+-(\lambda+\delta)}}{\lambda+\delta}\right) \\ &\quad\times\left(1-\frac{G}{2\pi\sigma}\int_{a_-}^{\lambda+\delta}\frac{\sqrt{(\lambda+\delta-a_-)}\sqrt{(a_+-\lambda-\delta)}}{\lambda+\delta}\mathrm{d}\lambda\right)^{M-K-i} \\ &= \frac{(M-K)!}{(i-2)!(M-K-i)!}\left(\frac{G}{2\pi\sigma}\right)^{i}\int_{a_-}^{a_+}\frac{\sqrt{(\lambda-a_-)}\sqrt{(a_+-\lambda)}}{\lambda} \\ &\quad\times\left(\int_{a_-}^{\lambda}\frac{\sqrt{(\lambda-a_-)}\sqrt{(a_+-\lambda)}}{\lambda}\mathrm{d}\lambda\right)^{i-2} \\ &\quad\times\left(\frac{\sqrt{(\lambda+\delta)-a_-}\sqrt{a_+-(\lambda+\delta)}}{\lambda+\delta}\right) \\ &\quad\times\left(1-\frac{G}{2\pi\sigma}\int_{a_-}^{\lambda+\delta}\frac{\sqrt{(\lambda+\delta-a_-)}\sqrt{(a_+-\lambda-\delta)}}{\lambda+\delta}\mathrm{d}\lambda\right)^{M-K-i} \end{aligned} \tag{40}$$

Substituting (39) into (40) will lead to

$$
\begin{aligned}
f_{\delta_i}(\delta) = {} & \frac{(M-K)!}{(i-2)!(M-K-i)!}\left(\frac{G}{2\pi\sigma}\right)^i \int_{a_-}^{a_+} \frac{\sqrt{(\lambda - a_-)}\sqrt{(a_+ - \lambda)}}{\lambda} \\
& \times \left\{ \frac{1}{4\sqrt{a_-}\sqrt{a_+}} \left[2\arcsin\left(\frac{-2\lambda + a_- + a_+}{a_- - a_+}\right) a_-^{3/2}\sqrt{a_+} \right.\right. \\
& + 2\arcsin\left(\frac{-2\lambda + a_- + a_+}{a_- - a_+}\right) a_+^{3/2}\sqrt{a_-} \\
& + 4a_+a_-\arctan\left(\frac{1}{2}\frac{2a_-a_+ - a_+\lambda - a_-\lambda}{\sqrt{a_-}\sqrt{a_+}\sqrt{(a_+ - \lambda)}\sqrt{(\lambda - a_-)}}\right) \\
& + \pi a_-^{3/2}\sqrt{a_+} + \pi\sqrt{a_-}a_+^{3/2} - 2a_-a_+\pi \\
& \left.\left. + 4\sqrt{a_-}\sqrt{a_+}\sqrt{\lambda - a_-}\sqrt{a_+ - \lambda} \right] \right\}^{i-2} \\
& \times \left(\frac{\sqrt{(\lambda + \delta) - a_-}\sqrt{a_+ - (\lambda + \delta)}}{\lambda + \delta} \right) \\
& \times \left(1 - \frac{G}{2\pi\sigma} \left\{ \frac{1}{4\sqrt{a_-}\sqrt{a_+}} \left[2\arcsin\left(\frac{-2(\lambda + \delta) + a_- + a_+}{a_- - a_+}\right) a_-^{3/2} \right.\right.\right. \\
& + 2\arcsin\left(\frac{-2(\lambda + \delta) + a_- + a_+}{a_- - a_+}\right) a_+^{3/2}\sqrt{a_-} \\
& + 4a_+a_-\arctan\left(\frac{1}{2}\frac{2a_-a_+ - a_+(\lambda + \delta) - a_-(\lambda + \delta)}{\sqrt{a_-}\sqrt{a_+}\sqrt{(a_+ - (\lambda + \delta))}\sqrt{((\lambda + \delta) - a_-)}}\right) \\
& + \pi a_-^{3/2}\sqrt{a_+} + \pi\sqrt{a_-}a_+^{3/2} - 2a_-a_+\pi \\
& \left.\left.\left. + 4\sqrt{a_-}\sqrt{a_+}\sqrt{(\lambda + \delta) - a_-}\sqrt{a_+ - (\lambda + \delta)} \right] \right\} \right)^{M-K-i} \mathrm{d}\lambda
\end{aligned} \tag{41}
$$

which leads directly to (26) for the probability distribution function.