Source Number Estimation via Entropy Estimation of Eigenvalues (EEE) in Gaussian and Non-Gaussian Noise

Hamid Asadi, and Babak Seyfe, Senior Member, IEEE

Abstract—In this paper, a novel method based on the entropy estimation of the observation space eigenvalues is proposed to estimate the number of the sources in Gaussian and Non-Gaussian noise. In this method, the eigenvalues of correlation matrix of the observation space will be divided by two sets: eigenvalues of signal subspace and eigenvalues of noise subspace. We will use estimated entropy of eigenvalues to determine the number of sources. In this method we do not need any a priori information about signals and noise. The advantages of the proposed algorithm based on the performance is compared with the existing methods in the presence of Gaussian and Non-Gaussian noise. We have shown that our proposed method outperforms those methods in the literature, for different values of observation time and Signal to Noise Ratio, i.e. SNR. It is shown that the algorithm is consistent and also its probability of false alarm and probability of missed detection tend to zero for long observation time.

Index Terms—Source enumeration, eigenvalues, entropy estimation, kernel estimation, Non-Gaussian.

I. INTRODUCTION

Signal enumeration has an important role in several fields such as brain imaging [1], neural networks [2], audio signals separation [3] and finance [4]. Also, the most important problem in sensor array signal
processing is to estimate the coordinates of the sources that are emitting signals. Since the azimuth and elevation angles of each point in three dimensional spaces are estimated by the measurement of direction of arrival (DOA) by an array of sensors, the number of sources must be known [5].

Some of the proposed solutions for array signal processing case are based on information theoretic criteria. Akaike-Information Criterion (AIC) [6], Minimum Description Length (MDL) [7], Bayesian Information Criterion (BIC) [8] and Predictive Description Length (PDL) [9] can be considered as four such ways in this approach. AIC and MDL minimize the Kullback-Leibler distance between the observations and the data model that is estimated by maximum likelihood estimator. MDL estimator fails to estimate number of sources in low SNR or small number of observation samples [10] but it is consistent and for large observation times, the probability of error that is equal to the probability of under estimation tends to zero, approximately [11]. While AIC estimator has a better performance in low SNR regime and small number of observation samples [10], this estimator is not consistent in the same conditions and its error probability or probability of overestimation does not tend to zero [11]. The proposed PDL method uses the projection of sample covariance matrix on to the signal and noise subspaces to estimate the number of sources based on the maximum likelihood (ML) estimate of the past data and at each time, new data will be updated by a recursive algorithm [9].

In [12], the probability of missed detection in MDL method is derived using the Tracy-Widom distribution for the largest eigenvalues and Gaussian distribution by updating a threshold using a first order Taylor series. Also it is shown that the statistical performance of the MDL is approximately the same under deterministic signal models [12]. In [13], new frameworks for analytically evaluating the statistical performance of eigen decomposition based detectors are considered. Also, in there, the exact and approximate asymptotic bounds of the overestimation probabilities of AIC and MDL are discussed. The number of sources is approximated by Fishler, Grosmann and Messer [14] via considering the case of both Gaussian sources and Gaussian noise. They use the asymptotic distribution of the eigenvalues of the empirical correlation matrix and it is shown
that the performance of MDL estimator is not very sensitive to the actual distribution of source signals. In [15] the problem of optimal rank estimation is considered by developing a decision-theoretic rank estimation such as a min-max algorithm. In [16], the number of sources is estimated using hypothesis testing as a Neyman-Pearson method and at each step testing the significance of eigenvalues as arising from a signal will be evaluated. Also, the probability of overestimation and its bounds are computed. In [17], a nonparametric method that depends on choosing a function of estimated eigenvalues from sample covariance matrix that is proposed to enumerate the number of sources. In [18], the number of sources is estimated based on the correlation matrix decomposition method and it uses the null space of correlation matrix to estimate the directions of arrival of the sources. In [19], the problem of signal enumeration is considered in an information theoretic criteria framework. In there, probability density function of eigenvalues is estimated based on the observation samples and used to signal enumeration in addition to the probability density function of observation. In [20], signal enumeration is investigated via bootstrap method to estimate the source number in an environment with a Gaussian noise mixture. Also, in [21] signal enumeration problem using a criterion based on generalized Bayesian information Criterion, i.e. GBIC, is investigated.

In the most published works [6-19, 21], the authors assumed that the noise is Gaussian and white both spatially and temporally. But in many cases due to impulsive nature of noise, such as sonar [22-23] and impulsive man-made noise in urban and indoor wireless systems [24-25], the noise model is often non-Gaussian or impulsive. Based on the most knowledge of the authors of this paper, the only work that is assumed non-Gaussian noise in the literature is [20].

In this paper, we present and analyze a new estimation algorithm in Gaussian and non-Gaussian noise based on the information theoretic criteria using the concept of amount of information earned from eigenvalues of the empirical correlation matrix of the observations. In our proposed method, called *Entropy Estimation of Eigenvalues (EEE)*, we do not need to know the probability distribution or any a priori information about the observations. The difference between two sets of eigenvalues (noise subspace
eigenvalues and signal subspace eigenvalues) can be considered by using the entropy as a measure of uncertainty. We only assume that the sources are non-coherent, which is an advantage in comparison with other methods using additional assumptions [15], [16]. The entropy measures the uncertainty of random variables or amount of information of them without using any additive assumption [26]. We have shown that the proposed method is comparable and better than the best methods reported in literature in the presence of Gaussian noise and it outperforms them in the presence of Non-Gaussian noise drastically. Also, it is shown that the algorithm has an outstanding accuracy in limited number of observations compared with other methods.

This paper is organized as followings: in section II, we present models and preliminaries. In section III, we present our motivation to signal enumeration and an estimator to entropy estimation. Also, the entropy estimation of eigenvalues will be demonstrated as a rank estimator in this section. Sections IV and V, consist of the performance analysis of the proposed method, and simulation results, respectively. Finally, section VI concludes the paper.

**Notation:** In this paper, we use matrix and vectors by uppercase and lowercase, bold, respectively. The form (\( \hat{\cdot} \)) means the estimated unknown parameters. Also \( \mathcal{N}(0, \Lambda) \) denotes the Gaussian distribution with zero vector mean and positive definite covariance matrix \( \Lambda \). \( \hat{H}(\lambda) \) is the estimated Shannon entropy of random variable \( \lambda \) based on the observed samples \( \hat{\lambda}, ..., \hat{\lambda} \).

## II. SIGNAL SUBSPACE MODEL

### A. Problem Formulation

Assume that \( K \) sources send their signals, independently, to \( P (P > K) \) sensors at the receiver. The observed signal at the receiver in time \( t_i \) is denoted by

\[
x(t_i) = A s(t_i) + n(t_i), \quad 1 \leq i \leq N,
\]
where $\mathbf{x}(t_i)$ and $\mathbf{A}$ are the observed vector with the dimension of $P$ in time $t_i$ and steering matrix with the size of $P \times K$, respectively. Also the components of the vector $\mathbf{s}(t_i) = [s_1(t_i), ..., s_K(t_i)]^T$ are white, real, stationary processes with zero mean and positive definite covariance matrix $\mathbf{R_s} = \text{diag}(l_{S,1}, ..., l_{S,K})$ where $l_{S,i}, 1 \leq i \leq K$ is the received power from the $i^{th}$ source. The noise samples are assumed to be additive white zero mean (Gaussian or non-Gaussian type) with unknown noise power ($\sigma^2$) which are independent from $\mathbf{s}(t_i)$.

Noting (1), the population covariance matrix of the received signal from eigen-decomposition can be formulated as follows [7]

\[
\mathbf{C} = E[\mathbf{x}(t_i)\mathbf{x}^H(t_i)] \\
= \Lambda^H \mathbf{R_s} \Lambda + \sigma^2 \mathbf{I}_P \\
= \mathbf{U} \Lambda \mathbf{U}^H,
\]

where in (2), $\mathbf{I}_P$ denotes the $P$-dimensional identity matrix, superscript $H$ shows the Hermitian of matrix, $\Lambda = \text{diag}(\lambda_1, \lambda_2, ..., \lambda_P)$ is the eigenvalue matrix of the observation space and $\mathbf{U}$ is the eigenvector matrix of the observations with the size of $K \times P$. Equation (2) shows that $\lambda_1 > \lambda_2 > ... > \lambda_P$ denotes the eigenvalues of the observation space and they are ordered in descending order of magnitudes [10]. It is shown that the first $K$ eigenvalues, i.e. $\lambda_1, \lambda_2, ..., \lambda_K$, belongs to the signal subspace and the other smallest ($P-K$) eigenvalues assumed as the eigenvalues of the noise subspace which are equal to $\sigma^2$[21].

In a limited observation time, we can approximate the sample covariance matrix of $\mathbf{x}(.)$ at $N$ observation times, by

\[
\hat{\mathbf{C}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}(t_i)\mathbf{x}^H(t_i) \\
= \hat{\mathbf{U}} \hat{\Lambda} \hat{\mathbf{U}}^H, \quad \hat{\Lambda} = \text{diag}(\hat{\lambda}_1, ..., \hat{\lambda}_P).
\]

By the weak law of large numbers (WLLN), $\hat{\mathbf{C}}$ and $\hat{\Lambda}$ in (3) tend to the covariance matrix $\mathbf{C}$, and eigenvalue matrix, $\Lambda$ asymptotically as $N \to \infty$, respectively [27].

Now we define a new variable $\hat{\lambda}$ with the observed samples of $\left(\hat{\lambda}_1, ..., \hat{\lambda}_P\right)$. We would like to bring the
attentions on that the $\lambda$ is different from $N$-observation time to another $N$-observation time. Then we can assume that $\lambda$ is a random variable. We consider this randomness via entropy estimation as criterion to estimate the number of independent sources.

III. DETECTING THE NUMBER OF SIGNALS BASED ON ENTROPY ESTIMATION

A. Motivation: uncertainty or amount of information earned from estimated eigenvalues

We draw the attentions on that $\hat{\Lambda}$ is a random matrix in the limited time of observation ($N<\infty$). This allows us to use entropy of eigenvalues for our algorithm to enumerate the number of the source signals. To measure the uncertainty or amount of information that can be earned from eigenvalues, we use a criterion based on information theoretic measures such as entropy [27]. Based on entropy properties, we can see that the amount of information of eigenvalues vector equals to only first $(K+1)$ eigenvalues, because

$$
(\lambda_1,\ldots,\lambda_p) = \left[ \lambda_1, \ldots, \lambda_K, \frac{\sigma^2_{P-K}}{\sigma^2} \right],
$$

and it is seen that the last $(P-K)$ eigenvalues equal to $\sigma^2$ and based on entropy properties, the amount of information earned from one $\sigma^2$, equals to the amount of information observed from multiple $\sigma^2$, so we can see that

$$
\hat{H}(\lambda)_{k+1}^p = \hat{H}(\lambda)_{k+2}^p \\
\vdots \\
= \hat{H}(\lambda)^p,
$$

or

$$
\hat{H}(\lambda)_{1}^{k+1} = \hat{H}(\lambda)_{1}^{k+2} \\
\vdots \\
= \hat{H}(\lambda)_{1}^p.
$$

Equations (5) and (6) will be our motivation to enumerate the number of signals impinging to a sensor
array. So we need to estimate the entropy of eigenvalues.

**B. Entropy Estimation using Kernel Function**

Because of our above motivations, we need to estimate the entropy of the eigenvalues based on the observations. In [30], Paninski showed that the estimation of probability spaces, which are needed for an entropy estimator, is an important challenge and could be a much difficult problem, and the solution to solve this problem is: *nonparametric estimators*. Also, three algorithms were investigated by Paninski to estimate the entropy: maximum likelihood (MLE) [31-32], Miller-Madow bias correction [33], and jackknifed-MLE [34]. All of the above investigated algorithms are nonparametric estimators that it means they need to estimate the probability space of measured under consideration data. We devise an estimator that estimate the entropy directly from data without any need to know about its probability distribution function. This method is based on the using of the kernels which are developed by Parzen and Rosenblatt [35-36]. Kernels are used for example in regression for estimation of random variables when conditioned, or in probability density estimation to estimate the PDF of random variables [37].

Based on the difficulties of the density estimation to estimate the entropy in high dimensions and with few samples, the kernel functions [37], will be useful functions to estimate entropy directly from samples without any need to PDF estimation. Kernel-based entropy estimator from eigenvalues, will be assumed with the following properties [34]

\[
\mathcal{K}(x) \geq 0, \quad (7)
\]

\[
\int_{\mathbb{R}} \mathcal{K}(x) \, dx = 1, \quad (8)
\]

\[
\lim_{x \to \infty} \left| x \mathcal{K}(x) \right| = 0, \quad (9)
\]

where the kernel \(\mathcal{K}(x)\) is a symmetric function. A scaled kernel function, \(\mathcal{K}_h(x)\) with smoothing parameter \(h > 0\) that is called *bandwidth*, can be defined as
\[ \mathcal{K}_h(x) = \frac{1}{h} \mathcal{K}\left(\frac{x}{h}\right). \] (10)

Smoothing parameter, window width or bandwidth is a free parameter that affects the estimation and estimation of the bandwidth for a kernel is considered in some articles [41-43].

Some of the popular kernels are listed in [37, 40]. It can be readily seen that for most of listed kernels in [40] and all of the kernels listed in [37], \( \mathcal{K}(x) = 0 \) for \( |x| > 1 \).

To estimate the \( \hat{H}(\lambda)_{ji} \), the following equation will be proposed as a (kernel) entropy estimator [37]

\[ \hat{H}(\lambda)_{ji} = \left( \frac{-1}{j-i+1} \right) \sum_{k=i}^{j} \log \left( \frac{1}{j-i+1} \sum_{l=i}^{j} \mathcal{K}_h(\hat{\lambda}_k - \hat{\lambda}_l) \right). \] (11)

In [37], it is shown that the equation (11) can be used as the Shannon entropy of the \( \lambda \) based on \( (\hat{\lambda}_1, \ldots, \hat{\lambda}_j) \) samples of observations.

C. Entropy Based estimation of the Number of Sources

We define

\[ F(i) = \hat{H}(\lambda)_{pi} \] (12)

We consider \( F(i) \) in the joint limit \( P, N \to \infty \) with \( P/N \) fixed and show that it is a concave \((\sqcap)\) function of \( i \) for \( i \leq K+1 \) (Here, concavity means that the second order differentiation of \( F(i) \) as \( P, N \to \infty \) with \( P/N = c > 0 \), that is discrete in time, is negative for all values of \( i \leq K+1 \).

**Lemma 1:** For large values of number of sensors and observation time, i.e., \( P, N \to \infty \), \( F(i) \) is a concave function \((\sqcap)\) of \( i \) for \( i \leq K+1 \).

**Proof:** Define the first order difference of \( F(i) \) as follows

\[ \Delta F(i) = F(i+1) - F(i) \]
\[ = \hat{H}(\lambda)_{p(i+1)} - \hat{H}(\lambda)_{pi}. \] (13)
Based on the definition of kernel functions,
\[ K_h(\hat{\lambda}_j - \hat{\lambda}_k) = \left(\frac{1}{h}\right) K\left(\frac{\hat{\lambda}_j - \hat{\lambda}_k}{h}\right), \]  \tag{14}

\[ F(i) \] can be written as the following

\[ F(i) = \hat{H}(\hat{\lambda})_i^P \]
\[ = \left(\frac{-1}{P-i+1}\right) \sum_{j=i}^{P} \log \left( \frac{1}{P-i+1} \sum_{k=i}^{P} K_h(\hat{\lambda}_j - \hat{\lambda}_k) \right) \]
\[ = \left(\frac{-1}{P-i+1}\right) \sum_{j=i}^{P} \log \left( \frac{1}{h(P-i+1)} \left( K(0) + \sum_{k=1}^{P} K\left(\frac{\hat{\lambda}_j - \hat{\lambda}_k}{h}\right) \right) \right). \]  \tag{15}

After some simplifications, the above equation can be written as

\[ F(i) = \log(P-i+1) + \log(h) + \left(\frac{-1}{P-i+1}\right) \sum_{j=i}^{P} \log \left( K(0) + \sum_{k=1}^{P} K\left(\frac{\hat{\lambda}_j - \hat{\lambda}_k}{h}\right) \right). \]  \tag{16}

So \( \Delta F(i) \) is as following

\[ \Delta F(i) = \log\left(\frac{P-i}{P-i+1}\right) + \left(\frac{-1}{P-i}\right) \sum_{j=i}^{P} \log \left( K(0) + \sum_{k=1}^{P} K\left(\frac{\hat{\lambda}_j - \hat{\lambda}_k}{h}\right) \right) \]
\[ - \left(\frac{-1}{P-i+1}\right) \sum_{j=i+1}^{P} \log \left( K(0) + \sum_{k=1}^{P} K\left(\frac{\hat{\lambda}_j - \hat{\lambda}_k}{h}\right) \right). \]  \tag{17}

In [44], for optimum values of the smoothing parameter, i.e. the bandwidth \( h \), it is mention that “the ideal window width will converge to zero as the sample size increases” that it means \( h \to 0 \) when \( P, N \to \infty \).

Also The first \((K+1)\) samples of \( \lambda \) are different from each other, and then for large values of observations, i.e. \( N \), and based on the equation (9), we have
\begin{align}
\lim_{\lambda \to 0} \mathcal{K} \left( \frac{\hat{\lambda}_j - \hat{\lambda}_k}{h} \right) &= 0 \quad \text{for } j \neq k \text{ and } j, k \leq (K+1). 
\end{align}

So we have

\begin{align}
\lim_{N \to \infty} \Delta F(i) &= \lim_{P \to \infty} \log \left( \frac{P-i}{P-(i-1)} \right), \quad i \leq K+1.
\end{align}

It is seen that the first order differentiation of $F(i)$, e.g. $\Delta F(i)$, for $i \leq K+1$ is a negative decreasing function of $i$, asymptotically, and then the second differentiation of $F(i)$ is negative for all values of $i \leq K+1$, then it is a concave function ($\cap$) in this area.

\begin{itemize}
\item \textbf{Theorem 1:} The number of independent sources (noise excluded) in a sensor array can be estimated as follows for large number of observations, i.e. $N \to \infty$
\end{itemize}

\begin{align}
\forall 1 \leq i \leq P-1: \hat{K} &= \arg \min_{\hat{\lambda}} \left( \hat{H}(\hat{\lambda})^i_{i+1} - \hat{H}(\hat{\lambda})^i_{i} \right) \text{,} 
\end{align}

or

\begin{align}
\forall 1 \leq i \leq P-1: \hat{K} &= \arg \max_{\hat{\lambda}} \left( \hat{H}(\hat{\lambda})^{i+1}_{i+1} - \hat{H}(\hat{\lambda})^{i+1}_{i} \right) \text{,}
\end{align}

where $\hat{H}(\cdot)$ denotes the estimation of Shannon Entropy.

\begin{itemize}
\item \textbf{Note:} In the theorem 1, (20) and (21) are originated from equations (5) and (6), respectively. In the following, we prove (20) and the prove procedure of (21) is the same as (20), and then we neglect it.
\end{itemize}

\begin{itemize}
\item \textbf{Proof of Theorem 1:} As it is proved for lemma 1, $F(i)$ is a concave function ($\cap$) of $i$ for $i \leq K+1$. So for large values of observation time, $N$, we have
\end{itemize}

\begin{align}
i \leq K+1: & \quad \Delta F(i) < 0, \quad \text{and} \quad F(i+1) < F(i) .
\end{align}

For $K+1 \leq i \leq P$, $\hat{\lambda}_i$'s equal to $\hat{\sigma}^2$ and then we can write
\[
\lim_{N \to \infty} \Delta F(i) = \lim_{N \to \infty} \left[ \frac{-1}{P-i} \sum_{j=i+1}^{P} \log \left( \frac{1}{P-i} \sum_{k=i+1}^{P} \mathcal{A}_h(\lambda_j - \lambda_k) \right) - \left( \frac{-1}{P-i+1} \right) \sum_{j=i}^{P} \log \left( \frac{1}{P-i+1} \sum_{k=i}^{P} \mathcal{A}_h(\lambda_j - \lambda_k) \right) \right]
\]
\[
= \left( \frac{-1}{P-i} \right) \sum_{j=i+1}^{P} \log \left( \mathcal{A}_h(0) \right) - \left( \frac{-1}{P-i+1} \right) \sum_{j=i}^{P} \log \left( \mathcal{A}_h(0) \right)
\]
\[
= \log \left( \mathcal{A}_h(0) \right) - \log \left( \mathcal{A}_h(0) \right) = 0,
\]
\[
K + 1 \leq i \leq P
\]

The equations (22) and (23), mean that \( F(i) \) for \( i \leq K+1 \) is a decreasing function of \( i \), and for \( i \geq K+1 \) is a non-decreasing function of \( i \), respectively, and then we can conclude that \( F(i) \) reaches to its minimum at \( i=K+1 \).

For \( N \to \infty \), \( F(i) \) is a concave function for \( i \leq K+1 \) and has a constant values for \( i \geq K+1 \), then we can say that \( \Delta F(i) \) has only one minima at \( i=K \) and then the following criterion can be used to determine the number of sources.

\[
\forall 1 \leq i \leq P - 1: \hat{K} = \arg \min_i \left( \hat{H}(\lambda)_{i+1}^P - \hat{H}(\lambda)_{i}^P \right).
\]

Based on equation (11), (24) can be rewritten as

\[
\forall 1 \leq i \leq P - 1:
\hat{K} = \arg \min_i \left( \frac{-1}{P-i} \sum_{j=i+1}^{P} \log \left( \frac{1}{P-i} \sum_{k=i+1}^{P} \mathcal{A}_h(\lambda_j - \lambda_k) \right) - \left( \frac{-1}{P-i+1} \right) \sum_{j=i}^{P} \log \left( \frac{1}{P-i+1} \sum_{k=i}^{P} \mathcal{A}_h(\lambda_j - \lambda_k) \right) \right).
\]

We call this criterion as *Entropy Estimation of Eigenvalues (EEE)*. The performance analysis of EEE (probability of missed detection and probability of false alarm) will be considered in the next section.

### IV. PERFORMANCE ANALYSIS OF EEE

In this section, we show the performance of proposed algorithm, EEE, analytically when \( N \) tends to infinity for equation (20). Because the performance of EEE algorithm based on (20) and (21) are the same, we analyze
the performance of the proposed algorithm based on (20). This section consists of two parts: probability of false alarm and probability of missed detection.

A. Probability of false alarm (over estimation)

In this subsection, the false alarm (over estimation) probability of the proposed algorithm will be discussed. For the start, we rewrite the criterion for signal enumeration that is

\[
\hat{K} = \min_{1 \leq k \leq P-1} \left( \hat{H}(\lambda)_{i=1}^{k} - \hat{H}(\lambda)_{i=1}^{p} \right)
\]

= \min_{1 \leq k \leq P-1} (\Delta F(i)).

(26)

For true number of sources \(K\), the overestimation occurs when

\[
\Delta F(K+n) < \Delta F(K-m), \quad \forall \ 0 \leq m \leq K - 1, \forall \ 1 \leq n \leq P - K,
\]

and then the probability of false alarm will be

\[
P_{fa} = P(\Delta F(K+n) < \Delta F(K-m)), \quad \forall \ 0 \leq m \leq K - 1, \forall \ 1 \leq n \leq P - K.
\]

(28)

Based on equation (23), \(\lim_{N \to \infty} \Delta F(K+n) = 0\), and we can write

\[
\lim_{N \to \infty} P_{fa} = \lim_{N \to \infty} P(0 < \Delta F(K-m)), \quad \forall \ 0 \leq m \leq K - 1.
\]

(29)

Based on equation (19), we have

\[
\lim_{N \to \infty} P_{fa} = \lim_{N \to \infty} P \left( 0 < \log \left( \frac{P-(K-m)}{P-(K-m)+1} \right) \right), \quad \forall \ 0 \leq m \leq K - 1.
\]

(30)

But it is seen that \(\left( \frac{P-(K-m)}{P-(K-m)+1} \right)\) is always smaller than 1, and then

\[
\lim_{N \to \infty} P_{fa} = 0.
\]

(31)
Note 1: For criterion given in (21), false alarm occurs when $\Delta F(K + n) > \Delta F(K - m)$, for $0 \leq m \leq K - 1$, and $1 \leq n \leq P - K$. For computation of the false alarm probability of this criterion, we should obey a same procedure as it is mentioned for (20).

B. Probability of Missed Detection

Missed detection occurs when $\hat{K} < K$. It means that this event happens when $\Delta F(K - k) < \Delta F(K + l)$, for $1 \leq k \leq K - 1$, and $0 \leq l \leq P - K$. Based on [11], we will assume $l = 0$. If we define $P_{\text{missed}}$ as missed detection probability, then

\[
P_{\text{missed}} = P(\Delta F(K - k) < \Delta F(K))
= P(\Delta F(K - k) - \Delta F(K) < 0), \forall 1 \leq k \leq K - 1.
\]

(32)

We need to expand $\Delta F(K - k) - \Delta F(K)$. Based on equation (19), we can write

\[
\lim_{N \to \infty} (\Delta F(K - k) - \Delta F(K)) = \log \frac{P - (K - k)}{P - (K - k) + 1} - \log \frac{P - K}{P - K + 1}
= \log \left(\frac{P - (K - k)}{P - (K - k) + 1} \times \frac{P - K + 1}{P - K}\right), \forall 1 \leq k \leq K - 1.
\]

(33)

It is known that $((P - K + 1)(P - (K - k)))/(P - K)(P - (K - k) + 1)$ is always greater than 1 for $1 \leq k \leq K - 1$ and then $\Delta F(K - k) - \Delta F(K) > 0$ and based on our assumptions, when $N$ tends to infinity

\[
\lim_{N \to \infty} P_{\text{missed}} = 0.
\]

(34)

Note 2: For criterion given in (21), missed detection occurs when $\Delta F(K - k) > \Delta F(K + l)$, for $1 \leq k \leq K - 1$, and $0 \leq l \leq P - K$. For the computation of the missed detection probability, we should obey a same procedure as it is mentioned for (20).
V. SIMULATION RESULTS

In this section we study the performance of the proposed algorithms via computer simulations. Consider a linear array of sensors with complex-valued observations and uncorrelated sources. We compare the performance of the proposed method, i.e. EEE, with MDL, AIC and the methods developed in [16] and [21], denoted by “RMT” and “GBIC1”, respectively. The false alarm of RMT, as it is recommended by the authors, is set to 0.1%. We do this by running 10000 independent trials for each point.

Also for the simulations in the presence of non-Gaussian noise we use the Gaussian mixture noise with the probability density function (pdf) given by [38]

\[ p(x) = (1 - \varepsilon) \mathcal{N}(0, \sigma^2) + \varepsilon \mathcal{N}(0, \eta \sigma^2) \]  

(35)

where \( \varepsilon \) is the probability of impulses in noise and \( \mathcal{N}(0, \sigma^2) \) and \( \mathcal{N}(0, \eta \sigma^2) \) denote zero mean Gaussian pdf’s with variances \( \sigma^2 \) and \( \eta \sigma^2 \) respectively. The above model, \( p(x) \), is an approximation of the Middleton (Class A) model for impulsive noise [38].

The kernel function that is used for simulations, is Gaussian kernel. The Gaussian kernel is often used because of its convenient mathematical properties [45]. For Gaussian kernel we have

\[ \mathcal{K}_G(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad \mathcal{K}_{G,h}(x) = \frac{1}{\sqrt{2\pi h}} e^{-\frac{1}{2} \left( \frac{x}{h} \right)^2} \]  

(36)

An optimum value of smoothing parameter or bandwidth, i.e. \( h \) for Gaussian Kernel has been studied in [44], as following

\[ h = 1.06 \frac{\hat{\sigma}_\lambda}{P^2} \]  

(37)

where \( \hat{\sigma}_\lambda \) is the estimated variance of the \( \lambda \) variable. We will use the same value in non-Gaussian noise as well.
The first simulation is shown in Fig. 1, and we examine the probability of detection with respect to the number of snapshots in Gaussian noise by assuming $SNR=8dB$, $P=10$, and $K=5$. The probability of detection is considered as the number of runs with true estimated number of sources normalized by the number of runs. It shows that EEE outperforms other methods when the number of observations increases, especially for $N \geq 10$.

The second simulation that is depicted in Fig. 2, is the probability of false alarm versus number of snapshots in Gaussian noise by assuming $SNR=8dB$, $P=10$, and $K=5$. In Fig. 3, the probability of missed detection has been shown. This figure shows that, when the number of snapshots increases the probability of missed detection tends to zero. Notice that, the major error for AIC, MDL and GBIC1 is from the false alarm and their probability of missed detection are negligible.

In the Fig. 4, detection probability is shown as a function of SNR in the presence of Gaussian noise. As we see in this figure, EEE has an obvious advantage in performance compared other methods. Our observations shows that the EEE algorithm has a good performance in a few observation samples. This makes that EEE has low complexity of computations in defined performance with comparison of other algorithms.

Fig. 5, shows the probability of detection when the number of sources changes. It is seen that all other algorithms lose their performance when the number of sources increases. The only algorithm that has a negligible degradation in performance is EEE.

In the simulations by assuming the non-Gaussian noise, in Fig. 6 we compare the probability of detection for all algorithms (AIC, MDL, RMT, GBIC1 and EEE) as a function of $SNR$ for $N=100$ and non-Gaussian noise parameters $\varepsilon=0.01$, $\eta=100$ and $\eta=1000$, respectively. This figure shows that EEE method is the only algorithm that has acceptable performance in impulsive or non-Gaussian noise.

In Fig. 7 and Fig. 8, the performance of the estimators in various degrees of noise impulsiveness has been shown. They show that EEE has the best performance among the studied methods. We bring the attentions on that, the performance of other algorithms degrade by increasing the impulsiveness of noise. Note that by
increasing the probability of impulses, i.e. $\epsilon \to 1$, the Gaussian mixture noise goes to the Gaussian model and then, Fig. 8 shows that the performance of other methods than EEE will improve by increasing $\epsilon$ to more than 0.02. This happens when the performance of EEE is not affected by increasing the impulsiveness degree of noise. In these figures, the detection probability of GBIC$_1$ and MDL is the same as each other.

Fig. 9, shows the probability of detection for different snapshots in non-Gaussian noise. From this figure it can be easily seen that there is a noticeable difference between EEE and other algorithms. In this figure, the GBIC$_1$ and MDL have a same performance.

VI. CONCLUSIONS

In this paper, we present new algorithm based on information theoretic method to determine the number of sources in Gaussian and non-Gaussian noise. We call this algorithm entropy estimation of eigenvalues (EEE). Our approach to entropy estimation is based on kernels and it is shown that we do not need any a priori information of noise. Also we show that the performances of the proposed method outperforms other methods in the literature in Gaussian and non-Gaussian noise.

ACKNOWLEDGMENT

The authors would like to thank to B. Nadler and S. Kritchman for their RMT [16] and MDL simulation codes.

REFERENCES

[1] M. McKeown, S. Makeig, S. Brown, T. P. Jung, S. Kindermann, A. J. Bell, V. Iragui and T. Sejnowski, “Blind separation of functional magnetic resonance imaging (fMRI) data,” Human Brain Mapping, vol. 6, pp. 368-372, 1998.

[2] E. Oja, “Principal components, minor components, and linear neural networks,” Neural Networks, vol. 5, pp. 927-935, 1992.

[3] T.W. Lee, A. J. Bell and R. Lambert, “Blind source separation of delayed and convolved sources,” Advances in Neural Information Processing Systems, vol. 9, pp. 758-764, 1997.
[4] K. Kiviluoto and E. Oja, “Independent component analysis for parallel financial time series,” In Proc. Int. Conf. on Neural Information Process. (ICONIP’98), vol. 2, Tokyo, Japan, 1998, pp. 895-898.

[5] P. S. Naidu, Sensor Array Signal Processing, CRC Press, 2001.

[6] H. Akaike, “A new look at the statistical model identification,” IEEE Trans. Automat. Contr., vol. AC-19, no. 6, pp. 716-723, Dec. 1974.

[7] 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.

[8] G. Schwarz, “Estimating the dimension of a model,” Ann. Statist., vol. 6, pp. 461-464, Mar. 1978.

[9] 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.

[10] 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, May 1991.

[11] Q. T. Zhang, K. M. Wong, P. C. Yip and I. 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.

[12] F. Hadaddi, 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.

[13] W. Xu and M. Kaveh, “Analysis of the performance and sensitivity of eigen decomposition-based detectors,” IEEE Trans. Signal Process., vol. 43, no. 6, pp. 1413-1426, Jun. 1995.

[14] 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.

[15] P. O. Perry and P. J. Wolfe, “Minmax rank estimation for subspace tracking,” IEEE Journal of Selected Topics in Signal Process., vol. 4, no. 3, pp. 504-513, Jun. 2010.

[16] S. Kritchman and B. Nadler, “Nonparametric 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.

[17] Y. Q. Yin and P. R. Krishnaiah, “On some nonparametric methods for detection of the number of signals,” IEEE Trans. Acoust., Speech, Signal Process., vol. ASSP-35, no. 11, pp. 1533-1538, Nov. 1987.

[18] A. Di, “Multiple source location-a matrix decomposition approach,” IEEE Trans. Acoust., Speech, Signal Process., vol. 33, no. 4, pp. 1086-1091, Oct. 1985.
[19] 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), Kyoto, Japan, Mar. 2012, pp. 3361-3364.

[20] Z. Lu, 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), Prague, May 2011, pp. 2712-2715.

[21] Z. Lu, and A. M Zoubir, “Generalized Bayesian information criterion for source enumeration in array processing,” IEEE Trans. Signal Process., vol. 61, no. 6, pp. 1470-1480, Mar. 2013.

[22] D. Brady and J. C. Preisig, “Underwater acoustic communications,” in Wireless Communications: Signal Processing Perspectives, H. V. Poor and G. W. Wornell, Eds. Upper Saddle River, NJ: Prentice-Hall, 1998.

[23] P. L. Brockett, M. Hinich and G. R. Wilson, “Nonlinear and non-Gaussian ocean noise,” Journal of Acoust. Soc. Amer., vol. OE-12, pp. 4-28, 1987.

[24] K. L. Blackard, T. S. Rappaport and C. W. Bostain, “Measurements and models of radio frequency impulsive noise for indoor wireless communications,” IEEE Journal on Selected Areas in Commun., vol. 11, pp. 991-1001, Sep. 1993.

[25] D. Middleton and A. D. Spaulding, “Elements of weak signal detection in non-Gaussian noise,” in Advances in Statistical Signal Processing, vol. 2: Signal Detection, H. V. Poor and J. B. Thomas, Eds. New York: JAI, 1993.

[26] G. D. Tourassi, “Application of mutual information criterion for feature selection in computer-aided diagnosis,” Medical Physics, vol. 28, no. 12, pp. 2394-2402, Dec. 2001.

[27] T. M. Cover and J. A. Thomas, Element of Information Theory, John Wiley & Sons, Inc. (Second Edition). 2006.

[28] T. W. Anderson, “Asymptotic theory for principal component analysis,” Ann. J. Math. Stat., vol. 34, pp. 122-148, 1963.

[29] L. Huang, T. Long, E. Mao, and H. C. So, “MMSE-based MDL method for accurate source number estimation,” IEEE Signal Process. Letters, vol. 16, no. 9, pp. 798-801, Sep. 2009.

[30] L. Paninski, “Estimation of entropy and mutual information,” Neural Computation, vol. 15, no. 6, pp. 1191-1253, Jun. 2003.

[31] A. Antos, and I. Kontoyiannis, “Convergence properties of functional estimates for discrete distributions,” Random Structures and Algorithms, vol. 19, issue 3-4, pp. 163–193, Oct. – Dec. 2001.

[32] S. P. Strong, R. Koberle, R. R. van Steveninck, and W. Bialek, “Entropy and information in neural spike trains,” Physical Review Letters, vol. 80, no. 1, pp. 197–200, Jan. 1998.

[33] G. Miller, “Note on the bias of information estimates,” In Information Theory in Psychology: Problems and Methods II-B, H. Quastler, editor, pages 95–100. Free Press, Glencoe, IL, 1955.

[34] B. Efron, and C. Stein, “The Jackknife estimate of variance,” The Annals of Statistics, vol. 9, no. 3, pp. 586-596, May 1981.
[35] E. Parzen, “Statistical inference on time series by Hilbert space methods,” Tech. Report 23, Stat. Dept., Stanford Univ., 1959.

[36] M. Rosenblatt, “Remarks on Some Nonparametric Estimates of a Density Function,” *Annals of Mathematical Statistics*, vol. 27, issue 3, pp. 832-837, Sep. 1956.

[37] J. C. Principe, *Information Theoretic Learning*, Information Science and Statistics, Springer, LLC 2010.

[38] D. Middleton, “Man-made noise in urban environments and transportation systems,” *IEEE Trans. Commun.*, vol. COM-21, pp.1232-1241, Dec. 1973.

[39] E. Fishler and H. Vincent 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.

[40] W. Zucchini, Applied smoothing techniques, Part 1: Kernel Density Estimation, Oct. 2003, Seen at: www.stateok.wismo.uni-goettingen.de, 20 March 2014.

[41] B. U. Park, and J. S. Marron, “Comparison of data-driven bandwidth selectors,” *Journal of the American Statistical Association*, vol. 85, no. 409, pp. 66-72, Mar. 1990.

[42] B. U. Park, and B. A. Turlach, “Practical performance of several data driven bandwidth selectors,” *Computational Statistics* vol.7, no.3, pp. 251-270, Feb. 1992.

[43] M. C. Jones, J. S. Marron, and S. J. Sheather, “A brief Survey of bandwidth selection for Density Estimation,” *Journal of the American Statistical Association*, vol. 91, issue 43, pp. 401-407, Mar. 1996.

[44] B. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London, 1998.

[45] P. F. Evangelista, M. J. Embrechts, and B. K. Szymanski, “Some properties of the Gaussian kernel for one class learning,” in *Proc. Intern. Conference on Artificial Neural Networks, ICANN07 Porto*, Portugal, Sep. 2007, pp. 269-278.
Fig. 1. Probability of detection as a function of number of snapshots in Gaussian noise: $K = 5$, $P = 10$ and $SNR = 8$ dB.

Fig. 2. Probability of false alarm as a function of number of snapshots in Gaussian noise: $K = 5$, $P = 10$ and $SNR = 8$ dB.
Fig. 3. Probability of missed detection as a function of number of snapshots in Gaussian noise: $K = 5$, $P = 10$ and $SNR = 8 \, dB$.

Fig. 4. Probability of detection as a function of SNR in Gaussian noise: $K = 5$, $P = 10$, and $N = 12$. 
Fig. 5: Probability of detection in Gaussian noise as a function of number of sources: $P=12$, $N=14$, $SNR=8dB$.

Fig. 6. Probability of detection in Gaussian mixture noise as a function of SNR: $K=5$, $P=10$, $N=100$, $\varepsilon=0.01$, $\eta=100$, and $\eta=1000$. 
Fig. 7. Probability of detection in Gaussian mixture noise as a function of $\eta$: $K=5$, $P=10$, $N=100$, $\varepsilon=0.01$, $SNR=20dB$.

Fig. 8. Probability of detection in Gaussian mixture noise as a function of $\varepsilon$: $K=5$, $P=10$, $N=100$, $\eta=100$, $SNR=20dB$. 
Fig. 9. Probability of detection in Gaussian mixture noise as a function of number of snapshots: $K=5$, $P=10$, $\eta=100$, $\varepsilon=0.01$, $SNR=20dB$. 