Abstract—This paper presents a method based on Minimum Description Length (MDL) criterion to enumerate the incident waves impinging on a large array using a relatively small number of samples. The proposed scheme exploits the statistical properties of eigenvalues of the Sample Covariance Matrix (SCM) of Gaussian processes. We use a number of moments of noise eigenvalues of the SCM in order to separate noise and signal subspaces more accurately. In particular, we assume a Marcenko-Pastur probability density function (pdf) for the eigenvalues of SCM associated with the noise subspace. We also use an enhanced noise variance estimator to reduce the bias leakage between the subspaces. Numerical simulations demonstrate that the proposed method estimates the true number of signals for large arrays and a relatively small number of snapshots. In particular, the proposed method requires less number of samples to achieve the same correct enumeration probability compared to the state-of-the-art methods. We evaluated the assumed pdf in order to justify the limitation and the behavior of the proposed method for small number of snapshots and array sizes.

Index Terms—Array signal processing, Minimum Description Length (MDL), Random Matrix Theory.

I. INTRODUCTION

Array signal processing has many applications ranging from radar and sonar to mobile communications. The observation vector in many signal processing applications can be modeled as a superposition of a finite number of signals embedded in additive white noise. In many applications the objective is to determine the number of superimposed sources, embedded in additive white noise, observed by an array, which is often called source enumeration [1], [2]. It is also the starting point of many signal parameter estimation algorithms. Two commonly used information theoretic criteria exploited for enumeration of sources are the Akaike Information Criterion (AIC) and the Minimum Descriptive Length (MDL) criterion [3]–[5]. These criteria are related to the Kullback-Leibler distance between the observed data and the model [5], [6]. The MDL chooses the model which has the minimum code length, i.e., the best model is assumed to give most information with least code length. In contrast to the MDL, the AIC is not consistent and tends to overestimate the number of sources even at high Signal-to-Noise-Ratios, (SNR)s. The MDL source enumerator [5], has received a considerable attention over the last two decades. Using the eigenvalue decomposition of the Sample Covariance Matrix (SCM), the eigenvalues are decomposed into signal and noise components. The MDL criterion can be viewed as the Log-Likelihood Function (LLF) evaluated at the Maximum Likelihood (ML) estimate of the unknown parameters of the model plus a penalty term which depends on the number of parameters of the model. This penalty term prevents the overestimation of the order. In [7], [8], the LLF is derived in terms of the eigenvalues of the SCM assuming that all signal sources are jointly Gaussian. Several versions of MDL are proposed to improve its performance, e.g., see [9]–[11]. A more accurate estimator for parameters is proposed in [9] using different LLF and penalty function employing the marginal distribution of the noise eigenvalues of the SCM. In [12] a statistical performance analysis of the MDL and AIC estimators has been presented and an improved AIC-type estimator with a better detection performance than MDL and a negligible probability of over-estimation has been proposed. The performance of the early methods are poor for limited number of data samples. Therefore, some properties of the ordered eigenvalues are exploited to improve the enumeration in [10]. A recursive MDL algorithm is proposed in [13] for online applications which outperforms the traditional MDL but has higher computational complexity.

Most of these enumerators have been derived assuming that the number of data samples is significantly larger than the number of antennas. This assumption is not always valid; especially when large antenna arrays are employed or when number of samples is small for some practical reasons, e.g., see [10], [14], [15]. In such cases, the performance of these source enumerators is not acceptable. Recently some schemes based on new results in random matrix theory has been derived to improve the performance of source enumerator algorithms in sample starved conditions [14], [16], [17]. New enumerators are proposed in [16], [17] using the second moment of noise eigenvalues of the SCM. In [14], [17], the number of signals are estimated by a sequence of hypothesis tests in terms of a user chosen confidence level (whereas the scheme in [16] chooses the model with the minimum AIC). Thus, the enumerator proposed in [14], [17] is a parameter free scheme, however, it is not consistent even at very high SNRs.

In this paper, we develop an algorithm based on the moments of the eigenvalues of the SCM. To the best knowledge of authors, this is the first algorithm which employs the higher order moments of eigenvalues. Signals and the noise are assumed to be stationary, ergodic and Gaussian processes. In this
case, the SCM of $n$ observations has Wishart distribution [18], thus, we can approximate the distribution of the eigenvalues of signal-free SCM by the Marcenko-Pastur distribution [19]. We exploit random matrix theory and propose a new LLF, employing moments of noise eigenvalues which improves the enumeration performance when the number of samples is not large compared with the dimension of the array.

II. Problem Formulation

Assume an array of $m$ sensors with $q$ signals impinging on it. The $i$th snapshot of the array observation is given by

$$x_i = A s_i + n_i, \quad i = 1, \cdots, n,$$

where $x_i \in \mathbb{C}^{m \times 1}$ is the observation vector, $s_i \in \mathbb{C}^{q \times 1}$ is the zero-mean signal vector with a Gaussian distribution with $E[s_i s_i^H] = S \delta_k$, where $S$ is a $q \times q$ nonsingular covariance matrix, $n_i \in \mathbb{C}^{m \times 1}$ is a zero-mean additive white Gaussian noise process with $E[n_i n_i^H] = \sigma^2 \delta_k \mathbf{I}$, where $\mathbf{I}$ is an $m \times m$ identity matrix and $A \in \mathbb{C}^{m \times q}$ is an unknown deterministic array manifold matrix. We assume that the signal and noise vectors are independent of each other, the covariance matrix of $x_i$, $R = A S A^H + \sigma^2 \mathbf{I}$. Assuming that the matrix $A$ is full rank and the covariance matrix $S$ is nonsingular, it follows that the rank of $A S A^H$ is $q$ and the $m-q$ smallest eigenvalues of $R$ are equal to $\sigma^2$. In such case, we denote the eigenvalues of $R$ by

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_q > \lambda_{q+1} = \cdots = \lambda_m = \sigma^2. \quad (2)$$

Obviously, the multiplicity of the smallest eigenvalue of $R$ gives the number of signals, $q$. The covariance matrix $R$ is estimated using $X = [x_1, \cdots, x_n]$ as the SCM which is also the ML estimate of $R$ as follows

$$\hat{R} = \frac{1}{n} \sum_{i=1}^{n} x_i x_i^H = \frac{1}{n} XX^H. \quad (3)$$

Since $X$ has a zero-mean Gaussian distribution, the matrix $\hat{R}$ has a Wishart distribution [18]. For $n \gg m + 1$, Anderson [20] has shown that the eigenvalues of $\hat{R}$, $\ell_1 \geq \ell_2 \geq \cdots \geq \ell_m$ are symmetrically centered around the eigenvalues of $R$ and has also characterized their asymptotic probability distribution. Assuming $q$ signal sources are present, Wax and Kailath [7] proposed an MDL criterion for source enumeration as follows

$$\text{MDL}(q) = - \ln(E[XX^H]) + \frac{1}{2} k \ln(n), \quad (4)$$

where $\hat{\Phi}_q$ is the Maximum Likelihood (ML) estimate of the parameter vector and $k$ is the number of free elements of the parameter vector. In this criterion, the term $\ln(E[XX^H])$ is the LLF of the observation vector maximized over the parameters and the second term is added as a penalty function to prevent over estimation. Using the Gaussian distribution of the received signal and the ML estimation of eigenvalues and eigenvectors of the SCM, the MDL proposed by Wax and Kailath estimates the number of sources as

$$\hat{q} = \text{argmin}_q \text{WK-MDL}(q),$$

$$= \text{argmin}_q n(m-q) \ln \left( \frac{m-q}{m} \sum_{i=q+1}^{m} \ell_i \right) + \frac{q(2m-q)}{2} \ln(n), \quad (5)$$

where WK-MDL($q$) is the description length of the observed data assuming $q$ signal sources are present. Simplicity and consistency of this estimator are among the main advantages of this source enumerator. However, this enumerator performs poorly when the number of observed snapshots is relatively small compared to the number of antenna array elements. In such a condition, the performance of the algorithms which use asymptotic estimation of the eigenvalues degrade drastically since the eigenvalues of the SCM are not closely centered around their true values.

Figure 1 compares the 20 eigenvalues of $R$ and those of the SCM for two number of snapshots 200 and 20 when $q = 4$ signal eigenvalues present. We observe that the distinction between signal and noise eigenvalues becomes less clear as the number of snapshots is reduced, i.e., in sample starved conditions. To overcome this limitation, we use some recent results from random matrix theory which characterize the distribution of the eigenvalues of the SCM more accurately and propose an enhanced source enumeration algorithm. Nadakuditi and Edelman [16] have recently used the joint distribution of the first and second moments of the noise eigenvalues and developed a test statistic whose distribution is invariant with the unknown noise variance $\sigma^2$, and proposed the following AIC enumerator

$$\hat{k} = \text{argmin}_k \left( \frac{1}{2} \left( \frac{m}{m-n} \right)^2 \ell^2_k + 2k \right), \quad (6)$$

$$t_k = \left( \frac{m-k}{\sum_{i=k+1}^{m} \ell^2_i} \left( \frac{m}{m-k} \right)^2 \right) - \left( 1 + \frac{m}{n} \right) \times m. \quad (7)$$

The computational complexity of this enumerator is comparable to the enumerator proposed by Wax and Kailath [7] while performs better for small number of snapshots. This enumerator becomes also consistent, employing MDL criteria instead of AIC just by adding a different penalty term as follows

$$\hat{k} = \text{argmin}_k \text{NE-MDL}(k), \quad (8)$$

$$= \text{argmin}_k \left( \frac{t^2_k}{2(\delta^2)} + (k+1) \ln(n) \right).$$

In this paper, we use the statistics of the higher moments of noise eigenvalues and a simple noise estimator in order to obtain further improvement in source enumeration. We exploit the MDL criterion in order to achieve consistent enumerator, whose probability of incorrect decision tends to zero in large sample size conditions.

III. SOME EIGENVALUE THEOREMS

In this section, we summarize some of the characteristics of eigenvalues in random matrices and then make use of these theorems to develop our new estimator. Note that the analytical results from random matrix theory, describe the characteristics of eigenvalues for large $m$ and $n$, i.e., as $(m,n \rightarrow \infty)$. This is in contrast to classical methods where the size parameter $m$ is fixed and only $n$ is large. The empirical distribution function (edf) of the eigenvalues of a non-negative Hermitian matrix $A$ with $m$ real eigenvalues is defined as:
\[ F^A(x) = \frac{1}{m} \{ \text{Number of eigenvalues of } A \leq x \}. \]

For a broad class of random matrices, this function converges to a non-random distribution as \( m, n \to \infty \) [21]. In our problem, the matrix \( R \) has Wishart pdf and the empirical distribution function of the signal-free SCM converges as demonstrated by the following proposition.

Proposition 1 (see [19], [22]): Let \( \hat{R} = \frac{1}{m} XX^H \) denote a signal-free sample covariance formed from \( m \times n \) matrix \( X \) with i.i.d. zero-mean Gaussian samples where all the eigenvalues of \( R \) are equal to \( \sigma^2 \). In such a case, the e.d.f., \( F^R(x) \), converges almost surely to Marcenko-Pastur distribution \( F^W(x) \) as \( m, n \to \infty, m/n = c > 0 \), i.e.,

\[
\begin{align*}
\frac{dF^W(x)}{dx} &= f^W(x), \\
&= \frac{\sqrt{(x-a_+)(x-a_-)}}{2\pi\sigma^2} \prod_{a_+,a_-} + \max \left( \frac{c}{1}, \frac{1}{c} \right) \delta(x),
\end{align*}
\]

where \( a_+ = \sigma^2 \left( 1 + \sqrt{c} \right)^2 \), \( a_- = \sigma^2 \left( 1 - \sqrt{c} \right)^2 \), \( \delta(x) \) is the Dirac delta function, and the random variable \( \sum_{a_+,a_-} \) converges weakly to a Gaussian random vector.

\[ \Pi_n, a, b \]

From (13), (11) and (12), it is proved that \( \ell_j \) tends to \( \lambda_j \) as \( m, n \to \infty \) and \( \lim_{m,n \to \infty} \frac{m-n}{m} = c \) [26]. In the next section, we exploit these distributions in order to derive enhanced source enumerators.

IV. PROPOSED SCHEME FOR SOURCE ENUMERATION

Assuming \( q \) signal sources, the parameter vector denoted by \( \phi_q = [\lambda_1, \lambda_2, \cdots, \lambda_q, \sigma^2] \) has a length of \( q+1 \) for different possible values of \( q = 0, 1, \cdots, m-1 \). The problem is to determine the number of sources \( q \). As mentioned before, the MDL principle chooses the model with minimum description length based on \( \Phi_q \), the estimate of the parameter vector, among candidate models. As mentioned before moments of Marcenko-Pastur distribution uniquely determine the distribution, so utilizing (10) a moment based test statistic can be developed employing \( \hat{z}_r(q) \) as follows

\[
\text{MDL}_q(q) = -\ln \left( f(\hat{z}_r(q)|\phi_q) \right) + \frac{q+1}{2} \ln(n). \quad (12)
\]

This criterion is obtained from (4) by using the distribution of \( \hat{z}_r(q) \). Employing (12) requires an accurate estimate of the unknown noise variance. In the presence of \( q \) signals, the ML estimation of noise variance is given by

\[
\hat{\sigma}^2_{\text{ML}} = \frac{1}{m-q} \sum_{i=q+1}^{m} \ell_i. \quad (13)
\]

From (13), (11) and \( E \left( \text{tr}(\hat{R}) \right) = \sum_{i=1}^{m} \lambda_i \) [27], we obtain

\[
E \left( \hat{\sigma}_{\text{ML}}^2 \right) = \sigma^2 - \frac{1}{n} \sum_{q=1}^{m} \lambda_i \sigma^2 + O(n^{-2}). \quad (14)
\]
This shows that the estimator in (13) is biased and underestimates the noise variance. The ML estimate of the parameters are conventionally used in MDL criteria. However, the underestimation of (13) in return results in a drastic overestimation of the number of signals, since the larger eigenvalues are compared with an underestimated value. To resolve this problem some enhanced noise variance estimators are proposed, e.g., see [14] and [9]. Wong, Zhang, Reilly and Yip [9] derived an estimator using the asymptotic joint pdf of eigenvalues of SCM, which requires to solve the following equations

$$\lambda_i = \ell_i - \frac{m-q}{n} \frac{\hat{\alpha}_i^2}{\lambda_i - \sigma^2} - \frac{1}{n} \sum_{j=1, j \neq i}^{q} \hat{\lambda}_j \lambda_i, \quad i = 1, \ldots, q$$ (15)

$$\hat{\sigma}^2 = \frac{1}{m-q} \sum_{i=1}^{m} \ell_i + \frac{1}{n} \sum_{j=1}^{q} \lambda_j \sigma^2$$ (16)

with the initial value of $$\hat{\lambda}_i = \ell_i, i = 1, \ldots, q$$ and $$\hat{\sigma}^2 = \frac{1}{m-q} \sum_{i=1}^{m} \ell_i$$. The Newton method is used to solve these nonlinear equations iteratively, to simultaneously estimate the signal eigenvalues and noise power. Thus, the Computational Complexity (CC) of this algorithm is much more than that of the ML estimator.

Here motivated by (14), we propose and use the following simple estimator for the noise variance

$$\hat{\sigma}^2 = \hat{\sigma}^2_{\text{ML}} + \frac{1}{n} \sum_{j=1}^{q} \ell_j \hat{\sigma}^2_{\text{ML}}$$ (17)

This estimator is easily implemented and can reduce the leakage from the signal subspace into noise subspace. The additional CC of the improved noise estimator in (17) is of order of $$\min(m,n)$$ and is much less than that of the method in (16) because it doesn’t estimate signal and noise eigenvalues simultaneously. Note that this additional load is insignificant compared to the computational load required for the calculation of eigenvalues $$\{\ell_j\}$$ from the input data.

Figure 2 shows simulation results for a linear array with $$m = 100$$ sensors, $$n = 100$$ snapshots, the noise variance of $$\sigma^2 = 1$$ versus number of signal sources which arrive at 0°, ±6°, ±12°, …, and the SNR of all sources are set to -17dB, which causes the signal eigenvalues to be greater than $$\sigma^2(1 + \sqrt{3})$$ and are detectable. In this case we observe that for a large number of sources our method results in smallest bias among other methods excluding for $$q = 5, 6$$ where the method in (16) is preferred. We must note that the main contribution of this paper is employing higher moments of eigenvalues which has not been addressed before. In this case, such a noise eigenvalue estimator provides additional improvement.

Assuming that the random vector $$\mathbf{z}_r(q)$$ has a normal distribution as in Theorem 1 and using the MDL criterion (12), we propose the following enumerator

$$\text{MDL}_r(q) = g(\mathbf{z}_r(q)) + \frac{1}{2} \ln |2\pi Q_r| + \frac{q}{2} + \frac{1}{2} \ln(n).$$ (18)

where $$g(\mathbf{z}_r(q)) = \frac{1}{2} (\mathbf{z}_r(q) - \mu_r) Q_r^{-1} (\mathbf{z}_r(q) - \mu_r)^T$$. Dropping off the terms that are not a function of $$q$$, sources are enumerated by minimizing $$\text{MDL}_r(q)$$ as follows

$$\hat{q} = \arg\min_{0 \leq q < \min(n,m)} g(\mathbf{z}_r(q)) + r(r+1) \ln(\sigma^2) + \frac{q}{2} \ln(n).$$ (19)

V. SIMULATION RESULTS

In our simulations, we consider a linear array of $$m$$ sensors exposed to $$q$$ planar narrowband uncorrelated signal sources arriving from different angles. The spacing between consecutive sensors is half of the wavelength [28], [29]. We compare the WK-MDL proposed in [7], improved AIC proposed in [12], the NE-AIC method proposed in [16] and its modified version using MDL criteria described in (8), the KN method proposed in [14] and our new criteria for various values of $$r$$ to detect the number of signals. All of the above schemes are parameter free estimators except for the KN method which requires a user chosen confidence level and is set to $$\alpha = 0.01$$ in our simulations. The improved AIC method proposed in [12] employs a different penalty function compared with the traditional AIC method. For various values of $$n$$, $$m$$ and SNR, we count the number of times that each of these estimators correctly enumerates the number of signals out of $$10^5$$ independent trials used for each scenario.

The probability of correct enumeration is plotted in Figures 3a and 3b as a function of SNR and number of snapshots, respectively. Figure 3a plots the results obtained for $$m = 100$$, $$n = 150$$, $$q = 3$$ signal sources and different values of SNR in the numerical simulations. In this scenario we observe that the NE-MDL (8) has an improvement of about 3.4 dB in SNR compared with the WK-MDL. The proposed method for $$r = 3, 4, 6$$ and 8 outperforms the WK-MDL about 4.2, 4.6, 4.85 and 4.95 dB in SNR, respectively, where the probability of correct decision is greater than 0.7 in low SNR regime. It can be observed that the KN method has comparable performance with the proposed method for $$r = 6$$ and $$r = 8$$. As shown in this figure, the NE-AIC method proposed in [16] is not consistent while achieves a higher correct detection probability in low SNRs; i.e., it suffers from overestimation when the number of snapshots is comparable to the number of array elements. This is because the penalty term of AIC criteria is smaller than that of the MDL which results in overestimation. This figure shows that the improved AIC method has better probability of correct detection comparing with the traditional MDL while it’s performance is inferior to that of the random matrix based schemes.

Figure 3b plots the probability of correct decision when the number of sources is $$q = 3$$, for $$m = 80$$, fixed SNR and different values of $$n$$ between 80 and 750. The results illustrate the enhanced performance of new source enumerators for relatively few number of samples. This figure shows that the random matrix based enumerators requires much less samples comparing with the traditional ones for the same probability of correct enumeration. In particular, the proposed MDL requires significantly less number of samples comparing with NE-MDL, i.e., the proposed method for $$r = 3, 4, 6$$ and 8, respectively, requires 45%, 54%, 59% and 61% less number of samples to achieve the same correct enumeration probability compared to the NE-MDL.

In another scenario with more sources $$q = 16$$, Figure 4 plots the probability of correct decision for $$m = 90$$ and $$n = 130$$. In this case, the proposed scheme still outperforms the traditional enumerators and its performance is comparable with that of
the KN method.

The proposed method is derived using the asymptotic regime in Theorem 1, i.e., where \( m \) and \( n \) are large numbers. However, Figure 5 demonstrates the performance obtained versus different SNRs, for \( m = 20 \) and \( n = 40 \) and 6 sources. Comparing with previous figures for a moderate number of antennas and snapshots, we observe that the proposed method remains superior to the traditional schemes. Interestingly, the proposed method significantly outperforms the KN method for small array dimensions.

VI. DISCUSSION

Our simulations show that the proposed algorithm results in significant improvement as the number of employed moments \( r \) is increased initially. However for larger values of \( r \) and limited number of samples, this improvement becomes saturated and false detections become more frequent. This is because in Section III we assumed that \( \hat{z}_r(q) \) converge to a multivariate Gaussian distribution. However, this assumption is only accurate asymptotically where \( m \) and \( n \) are large. We observe that higher order moments \( r \) require larger values of \( m \) and \( n \). This simply means that the distribution of \( \hat{z}_r(q) \) deviates more from the Gaussian assumption for limited \( m, n \) when \( r \) is too large. To observe such a deviation, Figure 6(a) shows the empirical estimated distribution of \( u_k = \frac{z_k - \mu_k}{\sqrt{\text{Cov}(z_k,z_k)}} \) for \( k = 1 \) and 4, 8, 16 where \( z_k = \sum_{i=1}^{m} \theta_i^k \), \( m = 100, n = 125 \) in noise only case. The estimated pdf of the random variable \( u_k \) represents the normalized pdf of \( z_k \) using the assumed mean and covariance in (10). The empirical probability distribution function is obtained using \( \hat{f}(u_k) = \frac{1}{2\pi \sigma^2} \sum_{i=1}^{N} \exp\left(-\frac{(u_k-\mu_k)^2}{2\sigma^2}\right) \) [30] where \( \sigma^2 = 0.05 \) and \( u_{k,i} \) are \( N = 10^4 \) randomly generated and normalized samples. The assumption in (10) implies that \( u_k \) must have a zero-mean Gaussian pdf with variance one. From these curves, we observe that the assumptions are reasonable. However, the right tail of the estimated pdf are heavier for higher order moments (as \( k \) increases), i.e., the larger \( k \) the more deviation from the assumption. Note that such a heavier right tails increases the chance that a noise eigenvalue to deviates more from the Gaussian assumption for limited \( m, n \). As a consequence, the noise eigenvalues are more likely mistaken with the signal source.

Many authors (for example see [16], [25], [26]) have assumed Marcenko-Pastur distribution for noise (small) eigenvalues in spiked models. This assumption means that the power of the source signals have no impact on the distribution of the smaller eigenvalues (which are assumed to correspond to the noise subspace). In fact most existing literature implicitly ignore the power leakage between signal and noise subspaces. As a known example in spike models, the mean of the noise eigenvalues in (14) demonstrates this power leakage which appears as an estimation bias and is a function of the signal eigenvalues. Figure 6(b) illustrates the empirical estimated distribution of \( u_k = \frac{z_k - \mu_k}{\sqrt{\text{Cov}(z_k,z_k)}} \) for \( k = 1 \) and 4, 8, 16 when \( m = 100, n = 125, q = 1 \) and SNR of -5dB. It can be seen from these curves that the mean of moments of eigenvalues in signal plus noise condition has a negative bias compared with the theoretic asymptotic distribution and this deviation will affect the performance of source enumeration schemes using these models.

The criterion in (19) is a quadratic function of \( \hat{z}_r(q) \). Figure 7 shows the condition number of \( Q_r \) versus \( c \) defined by \( \kappa(Q_r) = \log_{10}\frac{\lambda_{\text{max}}(Q_r)}{\lambda_{\text{min}}(Q_r)} \) where \( \lambda_{\text{max}}(Q_r) \) and \( \lambda_{\text{min}}(Q_r) \) are the largest and smallest eigenvalues of \( Q_r \), respectively for \( \sigma^2 = 1 \) in (10). We observe that the curves of the condition number significantly increases (moves up) as the matrix size increases. The increased condition number amplifies the inaccuracies of \( \hat{z}_r(q) \) and affects the MDL criteria in (19) which appears as increased false alarms. For \( r \geq 10 \) in Figure 7, the matrix \( Q_r \) is ill-conditioned for some values of \( 0 < c \leq 1 \) and cannot be inverted and used in (19). Thus, we recommend to use \( r \leq 8 \).

Figures 8a and 8b show the effect of sample size \( n \) and the array size \( m \) on the performance of the source enumerator for \( q = 1 \) and different values of \( r = 3, 4, 6 \) and 8. In these figures, we have \( c = \frac{n}{m} = 0.8, \lambda_1 = 3 \) and \( \sigma^2 = 1 \). Figure 8a shows that the probability of incorrect decision decreases as the number of snapshots (proportionally the number of antennas) increases for different values of \( r \). Figure 8b shows that our proposed method is more efficient if \( m \) and \( n \) are larger. In particular, the larger values of \( r \) can be used to obtain more significant improvement only for larger values of \( m, n \). This is because the pdf of the higher moments of eigenvalues deviate more from the assumed theoretic asymptotic distribution for smaller \( m, n \). Figure 8b illustrates that, the overestimation probability of the proposed enumerators decrease as \( m, n \) increase. The overestimation probability of the proposed enumerators increases as \( r \) increases and is less than that of NE-AIC and NE-MLD for \( r = 3, 4 \).

The proposed enumerator has the same order of CC as the WK-MDL and the NE-MDL and KN schemes. This is because all these methods commonly require the eigenvalues of the SCM where the CC for the computation of these eigenvalues is the dominant component of the CC. Thus for \( m < n \), let us compare the additional CC of these methods by excluding the common computations required for the calculation of the eigenvalues. Interestingly, the CC of (19) can be reduced by absorbing the variance terms \( \sigma^2 \) in \( Q^{-1} \), and \( \mu_k \) into \( \hat{z}_r(q) \), i.e., by first calculating \( \hat{z}_r(q) = \left[ \frac{1}{m\sigma^2} \sum_{i=1}^{m} \ell_i, \ldots, \frac{1}{m\sigma^2} \sum_{i=1}^{m} \ell_i \right] \) using cumulative sum algorithm which has a CC of order of \( O(mr) \) and then calculating \( g(\hat{z}_r(q)) = \frac{1}{2} (z')^T (Q^{-1} z') - \frac{1}{2} \) with a CC order of \( O(r^2) \) for each \( q \) where \( \mu_k' \) and \( Q^{-1} \) are obtained from (10) by substituting \( \sigma^2 = 1 \). Note that \( \mu_k' \) and \( Q^{-1} \) are constant and can be calculated in advance. In addition, the condition number of \( Q^{-1} \) is not data dependent; thus the impact of the finite precision numerical errors is considerably reduced by using \( Q^{-1} \). Therefore, the total additional CC for the proposed method is of order of \( O(mr + \sum_{q=1}^{r} r^2) = O(mr^2) \). The additional CC of the traditional enumerator and the NE scheme are of order of \( O(m) \) which is less
than the proposed enumerator because they do not require to calculate the moments of the eigenvalues. However, the KN scheme sequentially estimate the noise eigenvalues and requires $O((m + q^2)\nu)$ additional calculations, where $\nu$ is the average number of iterations of the noise eigenvalue estimator to converge, i.e., the additional CC of the KN scheme depends on a number of parameters such as the number of sources $q$ and the SNRs. To have a rough comparison, let’s consider the case where $\hat{q} = q$. In this case, for small number of sources, the KN scheme has a comparable CC as other schemes. However, its CC rapidly increases more than that of the proposed enumerator as the number of active signals increases. Nevertheless, these methods have the same order of CC, since their additional CC costs are insignificant compared with that of their common part which is the computation of the eigenvalue from the input data.

VII. CONCLUSION

In this paper, we introduced an information theoretic algorithm for enumeration of Gaussian sources in white Gaussian noise using a large array and a relatively small number of snapshots. To derive this enumerator, we used Minimum Description Length criterion and the statistical properties of the eigenvalues of the sample covariance matrix which are known asymptotically from the random matrix theory. In particular, we assumed that the vector of higher order moments of the eigenvalues of the SCM has a multivariate Gaussian distribution, on a number of parameters such as the number of sources $q$ and the SNRs. To have a rough comparison, let’s consider the case where $\hat{q} = q$. In this case, for small number of sources, the KN scheme has a comparable CC as other schemes. However, its CC rapidly increases more than that of the proposed enumerator as the number of active signals increases. Nevertheless, these methods have the same order of CC, since their additional CC costs are insignificant compared with that of their common part which is the computation of the eigenvalue from the input data.

REFERENCES


Fig. 1. Eigenvalues of $R$ and the Sample Covariance Matrix $\hat{R}$ for different number of samples, $n = 200$ and $n = 20$ ($m = 20$, $q = 4$ and $\sigma^2 = 2$).

Fig. 2. Normalized Bias of noise estimation of different schemes for different number of sources and for $n = m = 100$, SNR=-17dB.

Fig. 3. Probability of correct decision in present of 3 signal sources (a) as a function of received SNR and for $m = 100$, $n = 150$ (b) as a function of number of snapshots and for SNR=-14 dB.

Fig. 4. Probability of correct decision in present of 16 signal sources as a function of received SNR, for $m = 90$, $n = 140$. 
Fig. 5. Probability of correct decision in present of 16 signal sources as a function of received SNR, for $m = 20, n = 40$.

Fig. 6. Estimated pdf $\hat{f}(u_k) = \frac{1}{2\pi\sigma} \sum_{i=1}^{N} \exp\left(-\frac{(u_k - u_{k,i})^2}{2\sigma^2}\right)$ of $u_k = \frac{z_k^* - \mu_k}{\sqrt{\text{Cov}(z_k,z_k)}}$ for $k = 1$ and 4, 8, 16 where $z_k = \sum_{i=1}^{m} \ell_i$ and $\ell_i$ are eigenvalues of SCM for $m = 100, n = 120$, a) a noise only SCM, b) a signal plus noise SCM $q = 1$, SNR=-5dB.

Fig. 7. Condition number of $Q_r$ defined by $\kappa(Q_r) = \log_{10} \frac{\lambda_{\text{max}}(Q_r)}{\lambda_{\text{min}}(Q_r)}$ for $\sigma^2 = 1$ and $r = 2, 4, 6, 8$.

Fig. 8. Probability of (a) correct decision (b) overestimation, as a function of number of sensors $m$/snapshots $n$ (where $c = \frac{m}{n} = 0.8$), for one signal source $q = 1$, $\lambda_1 = 3$ and $\sigma^2 = 1$. 