

Fourth-Order Cumulant based Sources Number Estimation from Mixtures of Unknown Number of Sources

Fanglin Gu

College of Electronic Science and Engineering
National University of Defense Technology
Changsha, China 410073
Email: gu.fanglin@gmail.com

Wei Li

Unit 75753 of PLA
Guangzhou, China
Email: jackleewei@sina.com

Wenwu Wang

Department of Electronic Engineering
University of Surrey
Guildford, U.K. GU2 7XK
Email: w.wang@surrey.ac.uk

Abstract—We consider the source number estimation problem in the presence of unknown spatially nonuniform noise and underdetermined mixtures. We develop a new attractive source number estimator by replacing the covariance with the Fourth-Order (FO) cumulant of the observations. Furthermore, a Modified Minimum Description Length (MMDL) principle is proposed to reduce the divergence of the singular values corresponding to noise thereby improving the performance of the estimator. Simulation experiments validate the superiority of the proposed FO cumulant MMDL method to the conventional Minimum Description Length (MDL) method.

I. INTRODUCTION

Knowing the number of sources is a precondition for Blind Source Separation (BSS) and Blind Identification (BI). However, this condition is usually unsatisfied in practice. Hence estimating accurately the number of sources is crucial.

The most common techniques for estimating the source number are to apply information theoretic criteria, such as Akaike Information Criterion (AIC) [1] and Minimum Description Length (MDL) [2-3] principles or their modifications [4-9], on the received data. The MDL estimator has become the standard tool when it was first suggested in 1985 [2] for estimating the number of narrowband sources impinging on an array of sensors. However, there are still several unsolved problems. Firstly, most existing works are based on the assumption that the population sensor covariance matrix has P distinct eigenvalues, where P is the number of the sources, while the remaining eigenvalues are identical eigenvalues. Secondly, it is often assumed that the number of sensors $N-1$ is larger than the number of the source signals P . However, these assumptions may not be satisfied in practice, because (i) the noise components may be spatially correlated and vary considerably, and (ii) the underdetermined mixtures are common in practical situations, such as in radio communication over dense urban areas, the reception of more sources than sensors is possible and its probability increases with the reception bandwidth.

In this paper, we study the problem of estimating the number of source signals using Fourth-Order (FO) cumulant [10] of the observations. The reasons for using the FO cumulant

are twofold. Firstly, the FO cumulant of Gaussian noise is zero, regardless the Gaussian noise is white or colored. This property is useful to solve the source number estimation problem in unbalanced noise environments. Secondly, the FO cumulant can be applied to extend array virtually, which is equivalent to obtaining more observations. In such a way, the number of source signals can also be estimated, even if the mixture is underdetermined. Thus, a MDL enumerator that can use the FO cumulant is desirable.

To achieve this goal, the conventional MDL can be directly extended to FO cumulant by replacing the eigenvalue decomposition with Singular Value Decomposition (SVD) to remove the effect of the sign of FO cumulant. For practical reasons, the FO cumulant matrix and its singular values must be estimated from a finite set of snapshots, which will result in divergence of the singular values corresponding to noise. Thus, if the MDL criterion is applied under such an environment, severe degradation in detection performance may occur. To address this problem, we propose a Modified MDL (MMDL) criterion to estimate the number of sources, in which the divergence is reduced by replacing the singular values with its complement.

II. PROBLEM FORMULATION

Considering the following linear mixture model

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{w}(t) \quad (1)$$

where t is discrete time index. The stochastic vector $\mathbf{x}(t) \in \mathbb{C}^N$ represents observation signals, the components of the stochastic vector $\mathbf{s}(t) \in \mathbb{C}^P$ correspond to the unobserved source signals, and $\mathbf{w}(t) \in \mathbb{C}^N$ denotes Gaussian additive noise. The matrix $\mathbf{A} \in \mathbb{C}^{N \times P}$ characterizes the way the sources are mixed and it is unknown. In addition, the noise components are independent of the signals, with zero mean and correlation matrix given by $\text{diag}([\sigma_1^2, \dots, \sigma_N^2])$, which denotes a diagonal matrix. This correlation matrix represents the scenario in which each sensor potentially captures different level of noise.

According to the properties of the FO cumulant, we further make the following assumptions:

- The mixture can be underdetermined, but the number of sources P should be less than the squares of the number of sensors N^2 , that is, $P < N^2$.
- The sources have FO autocumulants with the same sign.

III. FOURTH-ORDER CUMULANT MMDL METHOD

A. Fourth-Order Cumulant

In this paper, the FO statistics of the data are characterized by the $N^2 \times N^2$ quadricovariance matrices $\mathbf{Q}_x(\tau_1, \tau_2, \tau_3)$, whose elements $Q_x(\tau_1, \tau_2, \tau_3)[i, j, k, l]$ ($1 \leq i, j, k, l \leq N$) are defined by

$$\begin{aligned} Q_x(\tau_1, \tau_2, \tau_3)[i, j, k, l] \\ \triangleq \text{cum}(x_i(t), x_j(t - \tau_1)^*, x_k(t - \tau_2)^*, x_l(t - \tau_3)) \\ = \text{E}\{x_i(t)x_j(t - \tau_1)^*x_k(t - \tau_2)^*x_l(t - \tau_3)\} - \\ \text{E}\{x_i(t)x_j(t - \tau_1)^*\}\text{E}\{x_k(t - \tau_2)^*x_l(t - \tau_3)\} - \\ \text{E}\{x_i(t)x_k(t - \tau_2)^*\}\text{E}\{x_j(t - \tau_1)^*x_l(t - \tau_3)\} - \\ \text{E}\{x_i(t)x_l(t - \tau_3)\}\text{E}\{x_j(t - \tau_1)^*x_k(t - \tau_2)^*\} \end{aligned} \quad (2)$$

where $\text{E}\{\cdot\}$ represents the expectation operator, $(\cdot)^*$ denotes the complex conjugate operator, and $x_i(t)$ is the i th component of $\mathbf{x}(t)$. Substituting (1) into (2) and assuming that $Q_x(\tau_1, \tau_2, \tau_3)[i, j, k, l]$ is the element $[N(i - 1) + j, N(k - 1) + l]$ of the matrix $\mathbf{Q}_x(\tau_1, \tau_2, \tau_3)$, we obtain the expression of the latter, which is given, under an additive Gaussian noise assumption, by

$$\mathbf{Q}_x(\tau_1, \tau_2, \tau_3) = [\mathbf{A} \otimes \mathbf{A}^*] \mathbf{Q}_s(\tau_1, \tau_2, \tau_3) [\mathbf{A} \otimes \mathbf{A}^*]^H \quad (3)$$

Under the assumption of statistically independent sources, the matrix $\mathbf{Q}_s(\tau_1, \tau_2, \tau_3)$ contains at least $P^4 - P$ zeros, and expression (3) degenerates into a simpler one given by

$$\begin{aligned} \mathbf{Q}_x(\tau_1, \tau_2, \tau_3) &= \sum_{p=1}^P c_p(\tau_1, \tau_2, \tau_3) [\mathbf{a}_p \otimes \mathbf{a}_p^*] [\mathbf{a}_p \otimes \mathbf{a}_p^*]^H \\ &= \mathbf{A}_Q \mathbf{C}_s(\tau_1, \tau_2, \tau_3) \mathbf{A}_Q^H \end{aligned} \quad (4)$$

where \mathbf{a}_p is the p th column of the mixing matrix \mathbf{A} , \mathbf{A}_Q is the $N^2 \times P$ matrix defined by $\mathbf{A}_Q \triangleq [\mathbf{a}_1 \otimes \mathbf{a}_1^*, \mathbf{a}_2 \otimes \mathbf{a}_2^*, \dots, \mathbf{a}_P \otimes \mathbf{a}_P^*]$ and $\mathbf{C}_s(\tau_1, \tau_2, \tau_3)$ is the $P \times P$ diagonal matrix defined by $\mathbf{C}_s(\tau_1, \tau_2, \tau_3) = \text{diag}[c_1(\tau_1, \tau_2, \tau_3), c_2(\tau_1, \tau_2, \tau_3), \dots, c_P(\tau_1, \tau_2, \tau_3)]$, with

$$c_p(\tau_1, \tau_2, \tau_3) \triangleq \text{cum}(s_p(t), s_p(t - \tau_1)^*, s_p(t - \tau_2)^*, s_p(t - \tau_3)) \quad (5)$$

To simplify the notations, we first define $\mathbf{Q}_x = \mathbf{Q}_x(0, 0, 0)$, $\mathbf{C}_s = \mathbf{C}_s(0, 0, 0)$, and $c_p = c_p(0, 0, 0)$, then we obtain (6) from (4)

$$\mathbf{Q}_x = \mathbf{A}_Q \mathbf{C}_s \mathbf{A}_Q^H \quad (6)$$

B. Modified MDL Criterion

Let $\lambda_i, i = 1, \dots, N^2$ be the singular values of \mathbf{Q}_x , and λ_i is shown in a descending order, then the singular values of \mathbf{Q}_x are given by

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_P \geq \lambda_{P+1} = \dots = \lambda_{N^2} = 0 \quad (7)$$

SNR(dB)	-5	0	5	10	15
λ_1	0.4548	0.4433	0.4310	0.4343	0.4334
λ_2	0.2825	0.2785	0.2780	0.2765	0.2759
λ_3	0.1055	0.0564	0.0248	0.0252	0.0301
λ_4	0.0967	0.0358	0.0155	0.0086	0.0076
λ_5	0.0684	0.0282	0.0087	0.0019	$8.1584e^{-4}$
λ_6	0.0531	0.0220	0.0042	0.0010	$7.4490e^{-4}$
λ_7	0.0441	0.0193	0.0019	$4.4630e^{-4}$	$9.1073e^{-5}$
λ_8	0.0423	0.0108	0.0017	$1.4857e^{-4}$	$8.6111e^{-5}$
λ_9	0.0198	0.0069	$4.5486e^{-4}$	$9.2304e^{-5}$	$2.4764e^{-6}$

TABLE I
THE SINGULAR VALUES OF FO CUMULANT MATRIX VERSUS SNR WHEN $P = 2$ NARROWBAND SOURCES ARE RECEIVED BY A UCA OF $Q = 3$ IDENTICAL SENSORS.

If the singular values of the received signal FO cumulant matrix \mathbf{Q}_x are available, the determination of P is straightforward. We only need to count the number of the singular values which are equal to zero. For practical reasons, the FO cumulant matrix and its singular values must be estimated from a finite set of snapshots. Therefore, the singular values become

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_P \geq \lambda_{P+1} \geq \dots \geq \lambda_{N^2} \quad (8)$$

For example, two narrowband source signals are received by a Uniform Circular Array (UCA) with three identical sensors (more details can be found in Simulation section), the singular values of the sample FO cumulant matrix versus SNR is shown in Table I. We can observe that the singular values $\lambda_{P+1}, \dots, \lambda_{N^2}$ corresponding to noise are close to zero but diverge severely. However, the existing criteria, such as AIC, MDL, etc, require that the eigenvalues of covariance matrix corresponding to noise be approximately equivalent or at least at the same magnitude. Hence, it will result in overestimate of the source number if the existing criteria are directly applied to the FO cumulant matrix.

In order to address the above mentioned problem, a MMDL criterion is proposed to estimate the source number using the FO cumulant matrix. In the MMDL criterion, a new statistic, which is complementary to the singular values, is used to decrease the divergence degree of the singular values corresponding to the noise. Then, the number of the source signals can be estimated by applying the MDL criterion with the new statistic.

Define

$$\mu_i = \lambda_{max} - \lambda_i \quad (9)$$

where λ_{max} is the maximum of the singular values. Because $\lambda_{max} \gg \lambda_{min}$, and the values of λ_i corresponding to noise are close to zero, then μ_i are mainly affected by the singular values corresponding to source signals, the values of $\mu_i, i = P + 1, \dots, N^2$ are fallen into the neighborhood of λ_{max} . In this way, the divergence degree of corresponding to noise decreases dramatically. Note that the MDL criterion is related to the divergence degree of the singular values, but not

related to the magnitude of the singular values. Therefore, the MDL criterion is also feasible even if the singular values are sorted in an ascending way, i.e., the singular values with small values correspond to the source signals while the singular values with large values corresponding to the noise.

Therefore, we propose a MMDL criterion based on the newly defined statistic μ_i as follows

$$\begin{aligned} \psi(k) = & T(M-k) \log \left(\frac{\frac{1}{M-k} \sum_{i=k+1}^M \mu_i}{\left(\prod_{i=k+1}^M \mu_i \right)^{\frac{1}{M-k}}} \right) \\ & + \frac{1}{2} k(2M-k) \log(T) \end{aligned} \quad (10)$$

where $M = N^2$, T is the number of snapshots. When k is an integer in the closed interval $[0, M-1]$, the number of sources is determined by the value of k for which $\psi(k)$ is minimum. Hence, the MMDL using the FO cumulant matrix is able to estimate the number of sources even in the underdetermined mixtures as long as the number of sources is less than the squares of the number of sensors.

The FO cumulant MMDL method can be summarized as follows:

- Compute the FO cumulant $Q_x(0, 0, 0)[i, j, k, l]$ of the observations and store the results in the quadricovariance matrix \mathbf{Q}_x .
- Compute the SVD of the FO cumulant matrix \mathbf{Q}_x to obtain the singular values λ_i , and calculate the new statistic μ_i .
- Estimate the number of sources using the MMDL criterion depicted in (10).

C. Computational Complexity

In this subsection, we aim to give some insight into the relative numerical complexity of the conventional MDL method and the proposed FO cumulant MMDL method for given values of T , N and P . The numerical complexity of the methods is presented in terms of the number of multiplication operations required to estimate the number of sources from T snapshots of data. The complexity of the proposed methods can be divided into two parts: one is used to calculate the covariance matrix or FO cumulant matrix, the other is used to implement the MDL criterion. Firstly, calculating the FO cumulant matrix \mathbf{Q}_x requires $7TN^4$ times multiplication operations when the length of snapshots is T , because each element in \mathbf{Q}_x requires $7T$ multiplication operations according to (2), whereas calculating covariance matrix only requires $2TN^2$ multiplication operations. Secondly, the SVD of a $K \times K$ matrix requires $4K^3$ multiplication operations using the QR iteration, hence, the SVD of the FO cumulant matrix \mathbf{Q}_x requires $4N^6$ operations, while the one for covariance matrix only requires $4N^3$ operations. Finally, implementing the MDL principle requires $N(N+1)/2 + 6N$ multiplications, whereas implementing the MMDL principle requires $N^2(N^2+1)/2 + 6N^2$ multiplications.

We can see that the FO cumulant MMDL method improves the performance at the expense of an increased computation complexity.

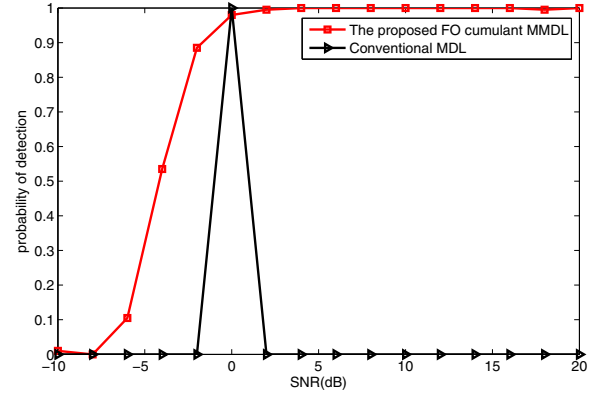


Fig. 1. The probability of detection in two-user scenario with unbalanced SNRs.

IV. SIMULATION AND ANALYSIS

In this section, simulation experiments are used to illustrate the performance of the proposed FO cumulant MMDL method. The experiments refer to the scenario that P narrowband source signals are received by a UCA of N identical sensors of radius R_a . Considering free space propagation, the entries of the mixing matrix \mathbf{A} are given by

$$a_{np} = \exp(2\pi j(\alpha_n \cos(\theta_p) \cos(\phi_p) + \beta_n \cos(\theta_p) \sin(\phi_p))) \quad (11)$$

where $\alpha_n = (R_a/\lambda) \cos(2\pi(n-1)/N)$, $\beta_n = (R_a/\lambda) \sin(2\pi(n-1)/N)$, and $j = \sqrt{-1}$. Assume $R_a/\lambda = 0.55$. The sources are unit-variance 4-QAM with uniform distribution, shaped by a raise cosine pulse shape filter with a roll-off $\rho = 0.3$. All sources have the same symbol duration $T = 4T_e$, where T_e is the sample period. The probability of detection is defined as the ratio of the number of trials at which the true number of sources has been observed to the total number of trials.

In the first experiment, we consider a UCA with five sensors ($N = 5$) and two sources ($P = 2$) with Directions-Of-Arrival (DOA) of the different sources given by $\theta_1 = 3\pi/10$, $\theta_2 = 3\pi/10$, and $\phi_1 = 7\pi/10$, $\phi_2 = 9\pi/10$. Additive zero-mean complex Gaussian noise is added to the observations. Firstly, we investigate the influence of the unbalanced SNRs of the observations on the performance of the proposed algorithm. In the simulation, the SNRs of the observations are 0 dB except one which ranges from -10 dB to 20 dB, and the number of symbols is 4000. The conventional MDL method is used as the compared method. 200 Monte Carlo experiments are run.

Fig. 1 shows that the performance of the conventional MDL method is degraded severely when the SNRs of the observations deviate from asymptotic assumption. The main reason is that the sources number will be overestimated when the SNRs of the observations deviate from asymptotic assumption. However, the performance of the proposed FO cumulant MMDL method is robust even in the unbalanced noise environment due to the fact that the fourth-order cumulant of Gaussian

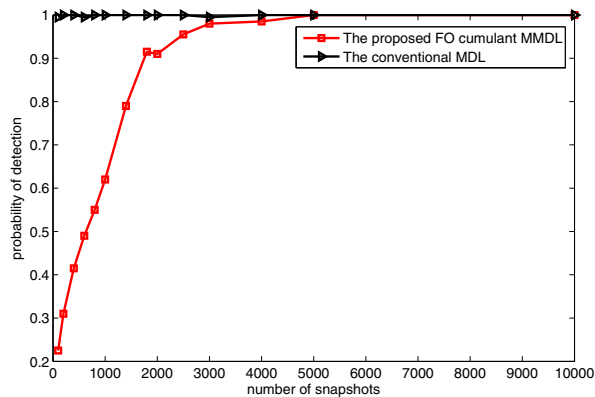


Fig. 2. The probability of detection as a function of the number of snapshots in two-user scenario.

noise is zero. On the other hand, the influence of the sample size of data on the performance of the proposed method is investigated. In the simulation, the number of symbols ranges from 100 to 10000, and the SNRs of the observations are 0 dB. The probabilities of detection versus the number of snapshots are shown in Fig. 2. As the FO cumulant requires longer data set to obtain an accurate estimation, the conventional MDL method outperforms the new proposed FO cumulant MMDL method when the number of samples is small.

In the second experiment, we investigate the influence of the number of sources on the performance of the proposed method. We consider a UCA with 4 sensors and 3, 4, 5 sources respectively. The DOAs of the first three sources are given by $\theta_1 = 3\pi/10$, $\theta_2 = 3\pi/10$, $\theta_3 = 2\pi/5$ and $\phi_1 = 7\pi/10$, $\phi_2 = 9\pi/10$, $\phi_3 = 3\pi/5$. The fourth and fifth sources are added at $\theta_4 = 0$, $\theta_5 = \pi/10$ and $\phi_4 = 4\pi/5$, $\phi_5 = 3\pi/5$ respectively. In the simulation, the SNRs of the observations are range from 0 dB when the number of symbols is 5000. The probabilities of detection of the FO cumulant MMDL method for the three cases are shown in Fig. 3.

It is clear that the proposed FO cumulant MMDL method can successfully estimate the number of sources even if the number of sources is close to or larger than the number of sensors as long as the SNRs of the observations and the length of the snapshot is long enough to acquire accurately estimate of FO cumulant. However, the conventional MDL method fails in this situation due to the fact that its conditions are not satisfied any longer.

V. CONCLUSION

We have presented a FO cumulant MMDL method to estimate the number of sources in BSS. Simulation results showed that the FO cumulant MMDL method, in comparison with the conventional MDL method, has the following advantages: i) the FO cumulant MMDL method has robust performance in an unbalanced noise environment, and ii) is able to estimate the number of sources when its number is close to or larger than the number of sensors at the expense of an increased

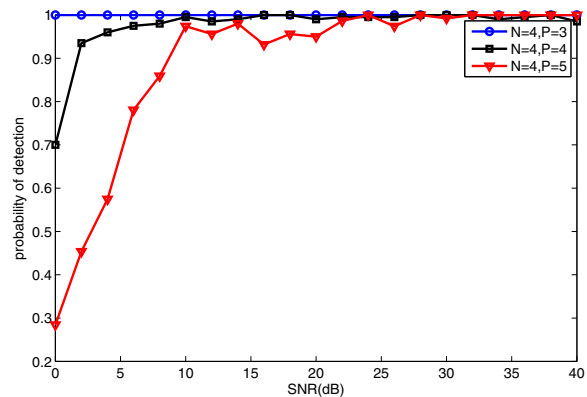


Fig. 3. The probability of detection of FO cumulant MMDL method for a UCA with 4 sensors and 3, 4, 5 sources.

computational complexity and longer data set. The future work is to extend the source number estimation method to higher-order cumulant for solving underdetermined mixtures with less restriction on the number of sensors.

REFERENCES

- [1] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716-723, 1974.
- [2] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions on Acoustics Speech Signal Processing*, vol. 33, no. 2, pp. 387-392, 1985.
- [3] M. Wax and I. Ziskind, "Detection of the number of coherent signals by the MDL principle," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 8, pp. 1190-1196, 1989.
- [4] E. Fishler and H. Messer, "Order statistics approach for determining the number of sources using an array of sensors," *IEEE Signal Processing Letters*, vol. 6, no. 7, pp. 179-182, 1999.
- [5] Y. Wu, K. Tam, and F. Li, "Determination of the number of sources with multiple arrays in correlated noise fields," *IEEE Transactions on Signal Processing*, vol. 50, no. 6, pp. 1257-1260, 2002.
- [6] Q. T. Zhang and K. M. Wong, "Information theoretic criteria for the determination of the number of signals in spatially correlated noise," *IEEE Transactions on Signal Processing*, vol. 41, no. 4, pp. 1652-1663, 1993.
- [7] E. Fishler and H. V. Poor, "Estimation of the number of sources in unbalanced arrays via information theoretic criteria," *IEEE Transactions on Signal Processing*, vol. 53, no. 9, pp. 3543-3553, 2005.
- [8] X. Bai and B. He, "Estimation of number of independent brain electric sources from the scalp EEGs," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 10, pp. 1883-1892, 2006.
- [9] J. Gu, P. Wei, and H. Tai, "Detection of the number of sources at low signal-to-noise ratio," *IET Signal Processing*, vol. 1, no. 1, pp. 2-8, 2007.
- [10] J. M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications," *Proceedings of the IEEE*, vol. 79, no. 3, pp. 278-305, 1991.