

Testing Rank of Incomplete Unimodal Matrices

Rui Zhang , Junting Chen , Yao Xie , Member, IEEE, Alexander Shapiro , and Urbashi Mitra , Fellow, IEEE

Abstract—Several statistics-based detectors, based on unimodal matrix models, for determining the number of sources in a field are designed. A new variance-ratio statistic is proposed, and its asymptotic distribution is analyzed. The variance-ratio detector is shown to outperform the alternatives. It is shown that further improvements are achievable via optimally selected rotations. Numerical experiments demonstrate the performance gains of our detection methods over the baseline approach.

Index Terms—Detector, low-rank matrices, statistical hypothesis testing.

I. INTRODUCTION

SOURCE localization based on received signal strength (RSS) [1] is challenged in environments where universal signal models are absent, such as underwater scenarios. Recently, a semi-parametric method based on sparsely spatially sampled RSS values and unimodal matrix factorizations has shown strong promise [2] for localizing multiple sources. However, [2] requires *a priori* knowledge of the number of sources. In practice, such information is unknown and must be learned. Herein, we design methods for estimating the number of sources when the signal propagation model is unknown save for the general property that the signal strength decreases with range. While the rank of the observation matrix, in specific cases, can be indicative of the number of sources, this is not true in general. For example, a single source can yield an observation matrix with rank two or more. Alternatively, if sources are co-linear, the rank of the observation matrix can be less than the number of sources.

Prior art on source localization also required the number of sources (e.g., [3]). If specific parametric models for signal strength exist, they can be leveraged to estimate the number of sources as in [4], [5]. Our goal, herein, is to provide methods for very general signal models such as unimodality.

Manuscript received January 23, 2021; revised March 20, 2021; accepted March 21, 2021. Date of publication April 2, 2021; date of current version May 13, 2021. The work of Rui Zhang and Yao Xie was supported by the NSF CAREER Award under Grant CCF-1650913 and by the NSF under Grant DMS-1938106 and DMS-1830210. The work of Junting Chen was supported by the Guangdong Fund under Grant 2017ZT07X152 Grant 2019QN01X895 and by the Shenzhen Institute of AIRS under Grant AC01202005001. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Pu Wang. (*Corresponding author: Yao Xie.*)

Rui Zhang, Yao Xie, and Alexander Shapiro are with the H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: ruizhang_ray@gatech.edu; yao.xie@isye.gatech.edu; ashapiro@isye.gatech.edu).

Junting Chen is with the School of Science and Engineering, Chinese University of Hong Kong - Shenzhen, Shenzhen 407605, China (e-mail: ejtchen@connect.ust.hk).

Urbashi Mitra is with the Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90007 USA (e-mail: ubli@usc.edu).

Digital Object Identifier 10.1109/LSP.2021.3070524

This paper develops a statistical test on the rank of the sparse observation matrix to estimate the number of sources based on RSS measurements from different locations. Our proposed method does not require a parametric model for the signal, and we also assume the noise variance is unknown. We tackle the challenge of testing the rank of a low-rank matrix using incomplete observations while assuming the signal energy field is unimodal. The proposed method is based on extending the results in [6], [7] to unimodal matrices. Moreover, we develop a new *variance-ratio statistic* and the corresponding test procedure. We further derive the asymptotic distribution of our proposed statistic, enabling the design of a threshold to control the false alarm rate. The proposed detection statistic is asymptotically efficient compared to an alternative estimator with an independent numerator and denominator. Furthermore, to improve our detector performance, we compute an optimal rotation to avoid the scenario of worst-case co-linearity of sources [2]. Numerical experiments validate the theory and demonstrate the proposed procedure's good performance.

II. PROBLEM SETUP

Consider the problem of K -source localization. Each source emits a signal that is unimodal: maximum signal strength for a measurement made at the source location, with reduced signal strength for measurements made away from the source location, decaying with the distance from the source location. For a single source, k , there is an $N \times N$ signal strength matrix $H^{(k)} \in \mathbb{R}^{N \times N}$. Let the singular value decomposition (SVD) of $H^{(k)}$ be

$$H^{(k)} = \sum_{i=1}^N \lambda_{k,i} \mathbf{u}_{k,i} \mathbf{v}_{k,i}^\top, \quad k = 1, \dots, K,$$

where $\lambda_{k,1} \geq \lambda_{k,2}, \dots, \geq \lambda_{k,N}$. A key result from [2] is that if the source signal is unimodal, then the singular vectors $(\mathbf{u}_{k,1}$ and $\mathbf{v}_{k,1})$ associated with the dominant singular value are also unimodal. Thus, we approximate the signal matrix for source k as,

$$H^{(k)} \approx \lambda_{k,1} \mathbf{u}_{k,1} \mathbf{v}_{k,1}^\top, \quad k = 1, \dots, K.$$

With our approximation in hand, the composite energy field is the superposition of the matrices of the K sources, which can be written as:

$$\begin{aligned} H &:= \sum_{k=1}^K H^{(k)} = \sum_{k=1}^K \lambda_{k,1} \mathbf{u}_{k,1} \mathbf{v}_{k,1}^\top + \sum_{k=1}^K \sum_{i=2}^N \lambda_{k,i} \mathbf{u}_{k,i} \mathbf{v}_{k,i}^\top \\ &:= S + U, \end{aligned}$$

where the second double sum is due to contributions of the non-dominant singular values and $H \approx S$. The first term is referred to as the *structured* component, comprised of the sum of K rank-one matrices due to the dominant singular values, and U is the *unstructured* component.

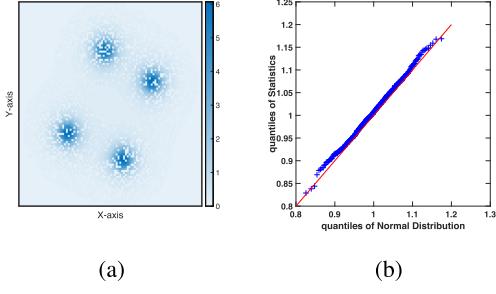


Fig. 1. (a) Partially observed energy field with four sources; (b) QQ-plot: $\hat{\sigma}_2^2(\mathbf{Z})/\hat{\sigma}_1^2(\mathbf{Z})$ against quantiles of $\mathcal{N}(1, (c+2)/(2cL))$; the parameter values are $c = 20$, $L = 150$.

We assume access to a limited number of measurements at a subset of all possible locations. The partially observed matrix M is then given by,

$$M_{i,j} = \begin{cases} S_{i,j} + \varepsilon_{i,j} & \forall (i,j) \in \Omega, \\ 0 & \forall (i,j) \notin \Omega, \end{cases} \quad (1)$$

where Ω is the set of observation locations, and ε is the measurement error on the (i,j) th observation which encapsulates both measurement noise as well as the contributions due to the unstructured components. Our goal is to identify the number of sources, K , given the partially observed matrix M . In Fig. 1(a), we see examples of unimodal energy fields with four sources.

III. METHODS

In this section, we propose two methods to identify the number of sources. As shown above, the rank of S is closely related to the number of sources. The rank of S is precisely equal to the number of sources if the unimodal property holds, and no rank-degeneracy is induced due to the summing of the contributions of the multiple sources, such as co-linearity. Therefore, determining the number of sources can be posed as finding a low-rank representation of M .

Our first method exploits the idea of rotation as proposed in [2] to mitigate the possibility of rank-degeneracy and thus improve performance. In the second strategy, we develop a heuristic method that considers multiple rotations and computes a statistic based on summing the dominant singular values obtained from each rotation.

A. Method I: Variance-Ratio Statistics With Optimal Rotation

Herein, we adapt our prior statistical analysis in [6] to the signal model herein Equation (1). Let $M, S \in \mathbb{R}^{n_1 \times n_2}$ and $r^* = \text{rank}(S)$. We assume $\varepsilon_{i,j} \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$, $\forall (i,j) \in \Omega$.

In practice, the variance of the noise is often unknown and needs to be estimated. In [7], a method to estimate the σ^2 and determine the rank heuristically is proposed. In contrast, here, we propose a statistic that does not depend on the scale of σ^2 and derive the asymptotic distribution. Therefore, we can provide a statistical hypothesis test for the rank.

To construct the statistic, we sub-sample from Ω to create a sequence of observations sets, such that $\Omega = \Omega_0 \supset \Omega_1 \supset \dots \supset \Omega_L$ and $|\Omega_l| - |\Omega_{l-1}| = c$, $\forall 1 \leq l \leq L$, where L is the number

of times we sub-sample the observations and c is the number of observations we “leave-out” between two consecutive sets.

Let SSE_l denote the the sum of squared errors of the matrix completion solution at rank- r :

$$\text{SSE}_l = \min_{Y: \text{rank}(Y)=r} \sum_{(i,j) \in \Omega_l} (M_{ij} - Y_{ij})^2. \quad (2)$$

According to Proposition IV.3 in [6], if $r = r^*$, as $\sigma^2 \rightarrow 0$,

$$\frac{Z_l}{\sigma^2} = \frac{\text{SSE}_{l-1} - \text{SSE}_l}{\sigma^2} \rightarrow \chi^2(c), \quad \forall 1 \leq l \leq L.$$

Let

$$\hat{\sigma}_1^2(\mathbf{Z}) = \frac{\sum_{l=1}^L Z_l}{cL}, \quad (3)$$

and

$$\hat{\sigma}_2^2(\mathbf{Z}) = \sqrt{\frac{\sum_{l=1}^L (Z_l - \bar{Z})^2}{2cL}}, \quad (4)$$

where $\mathbf{Z} = (Z_1, \dots, Z_L)$ and $\bar{Z} = (\sum_{l=1}^L Z_l)/L$. Here, we essentially use the method of moments to estimate the σ^2 : the first sample moment in Equation (3) for $\hat{\sigma}_1^2$ and the second sample moment in Equation (4) for $\hat{\sigma}_2^2$.

From the law of large numbers, $\hat{\sigma}_1^2 \rightarrow \sigma^2$ and $\hat{\sigma}_2^2 \rightarrow \sigma^2$, as $L \rightarrow \infty$. Moreover, we have the following theorem:

Theorem 1 (Asymptotic distribution of variance-ratio test statistic): Suppose Z_1, \dots, Z_L are *i.i.d.* random variables, and $Z_1/\sigma^2 \sim \chi^2(c)$. Then, $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ defined in (3) and (4), respectively, satisfy

$$\sqrt{\frac{cL}{2}} \left(\frac{\hat{\sigma}_2^2(\mathbf{Z})}{\hat{\sigma}_1^2(\mathbf{Z})} - 1 \right) \xrightarrow{d} \mathcal{N} \left(0, \frac{c+2}{4} \right),$$

as $L \rightarrow \infty$. \xrightarrow{d} denotes convergence in distribution.

The experiments in this paper and those in [7] show that the asymptotic result provides good performance under finite sample application. To illustrate the approximation of Theorem 1, we show an example in Fig. 1(b). In the QQ-plot, we can see the sample quantiles and target quantiles are very close. To statistically validate the approximation quality, we perform KS-test [8] and Anderson-Darling test [9]. In this example, KS-test and Anderson-Darling test accept that the sample and the target distribution are the same with significance levels 0.05 and 0.005, respectively.

Since the parameters of sub-sampling, c and L , which are defined above Equation (2), are known in practice, we can control the type-I error with the threshold determined by Theorem 1. From our numerical results, we observe that when the matrix approximation problem in Equation (2) is solved with the estimated rank smaller than the true rank, we tend to have $\hat{\sigma}_2^2 > \hat{\sigma}_1^2$. When the estimated rank r is larger than the true rank r^* , the proposed statistics $\hat{\sigma}_2^2/\hat{\sigma}_1^2$ tends to be close to 1 with larger variance than the statistics computed with the true rank. More discussion can be found in section IV-A. This observation suggests a rank-estimation strategy as shown in Algorithm 1: for each assumed rank r , we form the statistics and select the smallest r such that the statistic fails to reject the null hypothesis.

In the experiments, we used singular-value-thresholding algorithms [10]. However, the results are independent of the choice of the algorithm as long as it provides a good approximation for (2).

Algorithm 1: Rank Detection with Variance-Ratio.

Input: $M, \Omega, r_{\max}, c, L, b$.
 $\hat{r}, r = 0$;
while $r + 1 < r_{\max}$ **do**
 $r = r + 1$;
 compute $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ as in eq. (2), (3) and (4);
 if $\hat{\sigma}_2^2/\hat{\sigma}_1^2 < b$ **then**
 $\hat{r} = r$;
 exit while;
 end
end
Result: \hat{r}

1) *Relative Efficiency:* Notice that $\hat{\sigma}_1^2(\mathbf{Z})$ and $\hat{\sigma}_2^2(\mathbf{Z})$ are dependent. One can construct two independent estimates by performing more sub-sampling, i.e. $\Omega = \Omega_0 \supset \Omega_1 \supset \dots \supset \Omega_{2L}$. Let $\mathbf{Z}_{1:L}$ denote the first half of (Z_1, \dots, Z_{2L}) and $\mathbf{Z}_{L+1:2L}$ denote the second half. Then $\hat{\sigma}_1^2(\mathbf{Z}_{1:L})$ and $\hat{\sigma}_2^2(\mathbf{Z}_{L+1:2L})$ are independent. The following proposition shows that our proposed variance-ratio statistic $\hat{\sigma}_2^2(\mathbf{Z}_{1:L})/\hat{\sigma}_1^2(\mathbf{Z}_{1:L})$ is asymptotically more efficient than $\hat{\sigma}_2^2(\mathbf{Z}_{L+1:2L})/\hat{\sigma}_1^2(\mathbf{Z}_{1:L})$, i.e. the asymptotic variance of our proposed statistic is smaller.

Proposition 2 (Alternative variance-ratio statistic): Suppose $\mathbf{Z} = (Z_1, \dots, Z_L)$ and $\mathbf{X} = (X_1, \dots, X_L)$, where Z_i and X_i are *i.i.d.* random variables, and

$$\frac{Z_1}{\sigma^2} \stackrel{d}{=} \frac{X_1}{\sigma^2} \sim \chi^2(c),$$

$\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are constructed as in (3) and (4), respectively. Then as $L \rightarrow \infty$, we have

$$\sqrt{\frac{cL}{2}} \left(\frac{\hat{\sigma}_2^2(\mathbf{Z})}{\hat{\sigma}_1^2(\mathbf{X})} - 1 \right) \xrightarrow{d} \mathcal{N} \left(0, \frac{c+10}{4} \right).$$

Comparing Theorem 1 with Proposition 2, we can conclude that $\hat{\sigma}_2^2(\mathbf{Z})/\hat{\sigma}_1^2(\mathbf{Z})$ is asymptotically more efficient than $\hat{\sigma}_2^2(\mathbf{Z})/\hat{\sigma}_1^2(\mathbf{X})$. In practice, we only have a limited number of observations and can not sub-sample many times. With Theorem 1 and Proposition 2, we have shown that we can construct a better statistic, with fewer samples, and attain improved performance by exploiting the dependence of the statistics in the numerator and estimator of our test.

2) *Optimal Rotation:* To achieve a better result in practice, we incorporate a rotation step before applying Algorithm 1. In [2], it is observed that collinearity of sources can lead to rank degeneracy, and thus the number of sources may be larger than the rank of S . It was further shown in [2], that rotations can mitigate this issue and improve performance. Thus, herein we consider multiple rotations of various degrees, and E is a permutation matrix if and only if each row and column contains a single non-zero component of value one. Rotating a matrix $A \in \mathbb{R}^{N \times N}$ with degree θ can be achieved in the following way: $A(\theta) = \text{vec}^{-1}(E_\theta \text{vec}(A))$, where $\text{vec}(A)$ is the vectorization of A which stacks columns of A in a vector, vec^{-1} is the inverse function of vec , and $E_\theta \in \mathbb{R}^{N^2 \times N^2}$ is the permutation matrix corresponding the degree θ .

As stated previously, to avoid rank-degeneracy when applying the statistic in Theorem 1, we need to look for a rotation such that most of the sources are not co-linear. We employ the following

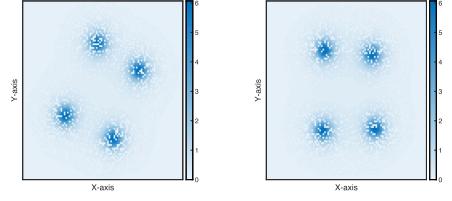


Fig. 2. Left: Rotation with degree θ_{opt} . Right: Rotation with degree θ_{max} .

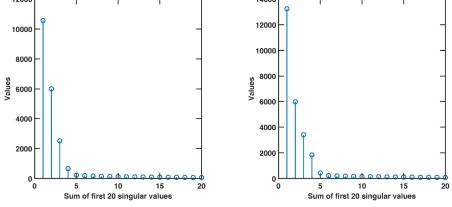


Fig. 3. Sum of first 20 singular values of 100 rotation degrees. Left: $K = 3$. Right: $K = 4$.

Algorithm 2: Rank Detection with Averaging Rotations.

Input: M, Ω, n, D, b , rotation degrees $(\theta_1, \dots, \theta_D)$.
while $i \leq D$ **do**
 rotation M with θ_i degree;
 complete M with nuclear-norm minimization;
 compute first n singular values: $\lambda_{i,1}, \lambda_{i,2}, \dots, \lambda_{i,n}$;
 $i = i + 1$
end
Result: $\hat{r} = \min \left\{ r : \frac{\sum_{j=1}^D \sum_{i=1}^r \lambda_{j,i}}{\sum_{j=1}^D \sum_{i=1}^n \lambda_{j,i}} > b \right\}$

strategy as suggested in [2] to determine the optimal rotation,

$$\theta_{\text{opt}} = \arg \min_{\theta} \frac{\lambda_1^2(H(\theta))}{\sum_{k=1}^N \lambda_k^2(H(\theta))}, \quad (5)$$

Numerical results suggest that the optimal rotation achieves the goal of avoiding rank degeneracy with high probability; furthermore, the use of θ_{opt} further improves the performance of our detector (see Theorem 1). As a validating example, we examine the effect of rotating the partially observed matrix by θ_{opt} in Equation (5), the optimal degree. Define $\theta_{\text{max}} = \arg \max_{\theta} \rho(\theta)$. Fig. 2 shows the matrix after rotation with degree θ_{opt} and θ_{max} , respectively. We can see rotation with θ_{opt} avoids the alignment of the sources; whereas rotating with θ_{max} actually aligns the sources leading to rank degeneracy.

B. Method II: Averaging Effects of Rotations

Here, we introduce a heuristic approach that can be applied in general. Instead of looking for an optimal rotation, we can rotate the matrix with multiple different degrees. For each rotation, matrix completion via nuclear norm minimization is conducted and the first few dominant singular values are computed. The resultant singular values are averaged over the multiple rotations. This averaging reduces the effect of degeneracy by reducing the effects of “bad degrees”. Fig. 3 shows the result for 100 rotation degrees. We clearly see that that there are K dominant singular values for $K = 3, 4$ respectively.

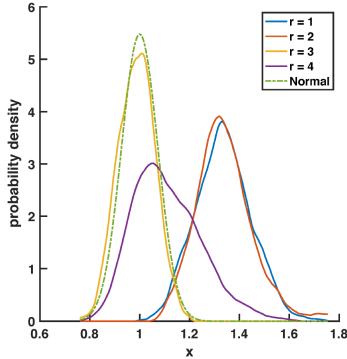


Fig. 4. A low-rank matrix with $r^* = 3$: Solid lines: empirical density estimation at each rank with 200 experiments. Dash line: probability density function of $\mathcal{N}(1, 0.005)$, which is the asymptotic approximation by Theorem 1. Clearly, it is a good fit when we compute our proposed statistics with the true rank.

IV. EXPERIMENTS

A. Theory Validation for Exact Low Rank Matrix

In this experiment, we show the distribution of our proposed statistics, when the data comes from the low rank matrix model in [7], i.e. $M_{i,j} = Y_{i,j} + \epsilon_{i,j}$, $\forall (i, j) \in \Omega$ and Y is a low rank matrix. We performed 200 experiments. In each experiment, a rank-3 matrix $Y \in \mathbb{R}^{100 \times 100}$ is generated and Ω is uniformly randomly chosen, s.t. $|\Omega| = 7500$. $\epsilon_{i,j} \sim \mathcal{N}(0, 1)$. Then, we compute the proposed statistics, under the assumption that the rank varies from to $1, \dots, 4$, with $c = 30$ and $L = 100$. In Fig. 4, we show the empirical density estimates of our proposed statistics for each rank. When $r = 3$, which is the actual rank, by Theorem 1, our proposed statistics can be approximated by a normal random variable with mean 1, and variance equal to 0.005. Fig. 4 shows a good fit for the empirical density corresponding to the true rank. For the rank less than the true rank, the density is well separated from the density with true rank. However, when the rank is larger than the true rank, it is not. Therefore, by choosing the smallest rank such that the p-value of proposed statistic is larger than some threshold, we can find the true rank with high probability.

B. Determining Number of Sources in Energy Field

We next examine the efficacy of our method for detecting the number of sources via the confusion matrix and the F_1 score. Given multiple classes, we use the average F_1 score in [11], which is defined as follows:

$$P_i = \frac{s_{ii}}{\sum_{x=1}^4 s_{ix}}, R_i = \frac{s_{ii}}{\sum_{x=1}^4 s_{xi}}, F_1 = \frac{1}{2} \sum_{i=2}^3 \frac{2P_i R_i}{P_i + R_i},$$

where i is the estimated number of sources, j is the true value and s_{ij} denotes the number of occurrences wherein the method estimates j sources given i . We consider both isotropic and skew energy sources.

- *Isotropic sources:* For isotropic sources, the sampled energy is purely range-dependent and thus independent of the rotation of the coordinator. In this experiment, we generate the energy field as the following. $af_{dB} =$

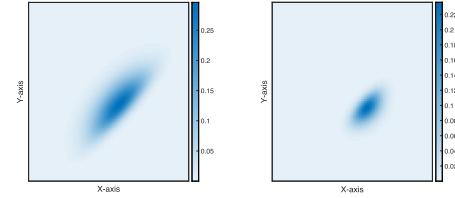


Fig. 5. Skewed energy source modeled by bivariate skew-normal distribution. Left: $\delta_1 = \delta_2 = 0.8$, $\omega = -0.5$. Right: $\delta_1 = \delta_2 = 0.35$, $\omega = -0.25$.

TABLE I
CONFUSION MATRICES OF DETECTION STATISTICS

		isotropic sources					skew sources						
		b	1	2	3	4	F_1	b	1	2	3	4	F_1
I	2	2.25	0	192	0	0	0.98	1.57	0	192	6	1	0.97
	3	1	0	193	0	0	(0.87)		3	0	196	1	
II	2	0.8	0	200	0	0	1.00	0.82	0	200	0	0	1.00
	3	0	0	200	0	0			0	0	200	0	
III	2	0.42	0	106	94	0	0.68	0.575	0	121	79	0	0.62
	3	0	30	170	0	0			0	71	129	0	

$0.11 \times f^2 / (1 + f^2) + 44 \times f^2 / (4100 + f^2) + 2.75 \times 10^{-4} \times f^2 + 0.003$ with $f = 5$. $a(\mathbf{x}) = d(\mathbf{x})^\alpha \times 10^{-af_{dB}/10 \times d(\mathbf{x})}$, $P(\mathbf{x}) = p/(a(\mathbf{x}) + 1)$, where $\alpha = 3$ and $p = 6$. The field is square with diameter 15 km.

- *Skew sources:* To generate the energy field with skew sources, we use the bivariate skew-normal distribution. According to Section 3 of [12], δ_1 , δ_2 and ω are the parameters to control the skewness. Fig. 5 displays examples of skew sources. In all experiments, δ_1 , δ_2 and ω are generated uniformly from $[-0.25, 0.25]$ for each energy source.

All the experiments are repeated 200 times each for the number of sources $K^* = 2, 3$. The standard variance of ambient noise is 0.01. The number of sensors is 4500, and we map them onto a matrix $M \in \mathbb{R}^{100 \times 100}$. Notice that with the mapping, $|\Omega| \leq 4500$. Our proposed methods are compared with a baseline method.

- *Method I: Variance-ratio statistic with optimal rotation.* In each experiment, we find the optimal rotation first, and then apply Algorithm 1 with $r_{\max} = 4$, $c = 2$ and $L = 750$.
- *Method II: Averaging effect of rotation.* In each experiment, we apply Algorithm 2 with $n = 20$, $D = 20$ and $\theta_i = (i - 1)\pi/(2D)$, $\forall i = 1, \dots, 20$.
- *Method III (baseline): Zero filling.* We apply singular value decomposition to the observation matrix M in eq.(1) and determine the number of source by thresholding the percentage of leading singular values, i.e., given threshold b , the estimated number of source of baseline method, $\hat{r}_{BL} = \min\{r : \sum_{i=1}^r \lambda_i / \sum_{i=1}^{100} \lambda_i > b\}$.

Results for isotropic and skew sources are provided in Table I. For the isotropic sources, the methods I, II, and baseline thresholds are 2.25, 0.8, and 0.42, respectively. For the skew sources, the thresholds of the Methods I, II, and baseline are 1.57, 0.82, and 0.575, respectively. For both types of energy sources, the proposed methods outperform the baseline. Furthermore, the use of the optimal rotation for isotropic sources also offers an improvement over no rotation. The F_1 score of it is shown in parenthesis, and the detailed results are shown in Table II in the appendix.

REFERENCES

- [1] C. Meesookho, U. Mitra, and S. Narayanan, "On energy-based acoustic source localization for sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 365–377, Jan. 2008.
- [2] J. Chen and U. Mitra, "Unimodality-constrained matrix factorization for non-parametric source localization," *IEEE Trans. Signal Process.*, vol. 67, no. 9, pp. 2371–2386, May 2019.
- [3] X. Sheng and Y.-H. Hu, "Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 44–53, Jan. 2005.
- [4] B. M. Ottersten, M. Viberg, P. Stoica, and A. Nehorai, "Exact and large sample ML techniques for parameter estimation and detection in array processing," in *Radar Array Processing*. Berlin, Germany: Springer, 1993, ch. 4.
- [5] M. Viberg and A. Lee Swindlehurst, "A Bayesian approach to auto-calibration for parametric array signal processing," *IEEE Trans. Signal Process.*, vol. 42, no. 12, pp. 3495–3507, Dec. 1994.
- [6] A. Shapiro, Y. Xie, and R. Zhang, "Matrix completion with deterministic pattern: A geometric perspective," *IEEE Trans. Signal Process.*, vol. 67, no. 4, pp. 1088–1103, Feb. 2018.
- [7] R. Zhang, A. Shapiro, and Y. Xie, "Statistical rank selection for incomplete low-rank matrices," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2019, pp. 2912–2916.
- [8] F. J. Massey, Jr., "The Kolmogorov-Smirnov test for goodness of fit," *J. Amer. Stat. Assoc.*, vol. 46, no. 253, pp. 68–78, 1951.
- [9] T. W. Anderson and D. A. Darling, "A test of goodness of fit," *J. Amer. Stat. Assoc.*, vol. 49, no. 268, pp. 765–769, 1954.
- [10] R. Mazumder, T. Hastie, and R. Tibshirani, "Spectral regularization algorithms for learning large incomplete matrices," *J. Mach. Learn. Res.*, vol. 11, no. 08, pp. 2287–2322, 2010.
- [11] J. Opitz and S. Burst, "Macro f1 and macro f1," 2019, *arXiv:1911.03347*.
- [12] A. Azzalini and A. Dalla Valle, "The multivariate Skew-normal distribution," *Biometrika*, vol. 83, no. 4, pp. 715–726, 1996.
- [13] G. Casella and R. L. Berger, *Statistical Inference*, 2nd ed. Boston, MA, USA: Thomson Learning, 2002.