

Grid-less Variational Direction of Arrival Estimation in Heteroscedastic Noise Environment

Qi Zhang, Jiang Zhu, Yuantao Gu and Zhiwei Xu

Abstract

Horizontal line arrays are often employed in underwater environments to estimate the direction of arrival (DOA) of a weak signal. Conventional beamforming (CB) is robust, but has wide beamwidths and high-level sidelobes. High-resolution methods such as minimum-variance distortionless response (MVDR) and subspace based MUSIC algorithm, produce low sidelobe levels and narrow beamwidths, but are sensitive to signal mismatch and require many snapshots and the knowledge of number of sources. In addition, noise at arrays may be heteroscedastic due to nonstationary environments and degrade the conventional methods significantly. This paper studies DOA in heteroscedastic noise (HN) environment, where the variance of noise is varied across the snapshots and the antennas. By treating the DOAs as random variables and the nuisance parameters of the noise variance different across the snapshots and the antennas, multi-snapshot variational line spectral estimation (MVALSE) dealing with heteroscedastic noise (MVHN) is proposed, which automatically estimates the noise variance, nuisance parameters of the prior distribution, number of sources, and providing the uncertain degrees of DOA estimates. Besides, it is shown that MVALSE can be applied to deal with the incomplete measurement case and measurement with outliers. When the noise variance is only varied across the snapshots or the antennas, the variants of MVHN, i.e., MVHN-S and MVHN-A can be naturally developed. Finally, substantial numerical experiments are conducted to illustrate the performance of the proposed algorithms, including a real data set in DOA application.

keywords: Variational Bayesian inference, von Mises distribution, grid-less, DOA estimation, heteroscedastic noise

Qi Zhang, Jiang Zhu and Zhiwei Xu are with the Ocean College, Zhejiang University, Zhoushan, CHINA, email: {zhangqi13, jiangzhu16, xuzw}@zju.edu.cn. Yuantao Gu is with Department of Electronic Engineering, Tsinghua University, Beijing, China, email: gyt@tsinghua.edu.cn.

I. INTRODUCTION

Horizontal line arrays (HLA) are often used for target bearing estimation based on the direction of arrival (DOA) of the received signal. For the far-field scenario where the source and receiver is large compared with the array aperture, the signals are assumed to arrive as plane waves. A popular method of array signal processing is conventional beamforming (CBF), in which the received signals are delayed and summed based on the sensor configuration relative to the signal look direction. When the steering angle matches with the DOA of a given target, the target signals are coherently summed. In contrast, noise between the receivers is uncorrelated and is summed incoherently. As a consequence, the signal to noise ratio (SNR) at the output of the beamformer is enhanced, and the SNR gain compared to the input SNR is termed as the array gain (AG).

The advantage of CBF, as is well known, is the robustness against signal mismatch between the assumed and the actual signal wave front. The disadvantages of CBF are the wide beamwidths which makes it difficult to detect closely targets within the same beam, and the high sidelobe levels, which makes it hard to detect the weak source in the presence of a loud interferer. To overcome the deficit of CBF, high-resolution methods such as minimum-variance distortionless response (MVDR) and subspace based MUSIC algorithm are proposed [1]. These algorithms yield low sidelobe levels and narrow beamwidths, while they are sensitive to the signal mismatches at the same time. Besides, all the methods require the knowledge of the number of sources.

Often, the main assumption used in most DOA problem is that noise is wide sense stationary, i.e., the noise variance is constant across both antennas and snapshots. However, for long observation times, the noise variance may change either in antennas or snapshots, or both. Thus, developing algorithms taking noise variation into account will be beneficial for DOA. Mathematically, the modelings of the noise variation can be classified into four cases [2]:

Case I: Noise is wide-sense stationary in both antennas and snapshots.

Case II: Noise is wide-sense stationary only in antennas, and is varied across the snapshots.

Case III: Noise is wide-sense stationary only in snapshots, and is varied across the antennas.

Case IV: Noise is heteroscedastic across both antennas and snapshots.

Case I corresponding to wide-sense stationary noise in both antennas and snapshots is the most widely used assumption, and many well-known methods are developed. For the other three cases, practical scenarios also exist. For example, for Noise Case II, underwater acoustic channel is time-varying, and it varies in seasons, areas and situation of sea surface. In addition, spatial movement of the source and/or changes in the propagation conditions such as sound speed profile, which lead to channel variation. As

a result, one can model the additive noise with time-varying variance [3, 4]. For the Noise Case III, it is also known as nonuniform noise which models the sensors with hardware nonidealities in receiving channels [5], as well as for arrays with position dependent noise (for example, hydrophones near to the surface have a larger noise variance because of passing ocean wave on the surface). As for Noise Case IV, it may correspond to the situation where a moving target/interference with high bearing rate dominates a few nearby sensors (spatial varying noise) [6, 7].

Most works focus on studying the DOA under noise Case I assumption. Since the sources are sparse in the spatial domain, many compressed sensing based DOA estimation methods are proposed and can be classified into three cases: on-grid, off-grid and grid-less. On-grid refers to discretize the DOAs belonging to $[-90, 90]^\circ$ into a number of grids, off-grid is based on the on-grid approach with an additional grid refinement. Grid-less treating the frequency as the continuous parameter completely overcomes the model mismatch [14], compared to on-grid and off-grid methods, and thus has attracted much more attention in recent years. It is also worth noting that both CBF and MVDR are on-grid methods, while MUSIC is grid-less.

It has already been shown that traditional DOA estimation methods under Case I noise assumption are sensitive to the noise models [8], and their performance degrades significantly when the assumption is not met [9]. For example, the number of effective samples for constructing the covariance matrix can be relatively small and leads to the so-called snapshot deficient condition [6]. As a result, several algorithms have been proposed to tackle the DOA problem under the three noise modeling cases [2, 10–13]. For Case II, an adaptive sparse Bayesian learning (SBL) algorithm is proposed, and improved source localization performance is demonstrated with experimental data [10]. The deterministic maximum likelihood (ML) estimator [11] and stochastic ML estimator [12] are proposed for the nonuniform white noise with an arbitrary diagonal covariance matrix as Case III, respectively. In [2], Case IV is studied, and on-grid SBL is proposed to estimate the heteroscedastic noise process and performance is demonstrated numerically. In contrast, this paper rigorously develops the grid-less variational line spectral estimation (Valse) based approach in heteroscedastic noise environments.

A. Related Work

In [15], Valse is proposed under homogenous noise, which automatically estimates the number of sources, the nuisances parameters of the prior distribution and noise variance. In addition, in contrast to the previous works which outputs the point estimates of DOAs only, Valse treats the DOAs as random parameters, and outputs the posterior probability density function (PDF) of the DOAs. In [16], multi-snapshot Valse (MVALSE) is developed for Case I, and sequential MVALSE (Seq-MVALSE)

is also proposed to perform sequential estimation. While in this work, noise assumptions corresponding to Case II-IV are studied, and three algorithms termed as MVALSE under heteroscedastic noise of snapshots (MVHN-S), MVALSE under heteroscedastic noise of antennas (MVHN-A), MVALSE under heteroscedastic noise of both snapshots and antennas (MVHN) are proposed. Compare to VALSE and MVALSE under homogenous noise, the variant of MVALSE, i.e., the proposed three algorithms are rederived. In particular, intermediate quantities defined later in (30) is coupled with the noise variance, whereas intermediate quantities of MVALSE defined in [16, eq. (21)] is independent of the noise variance.

B. Main Contributions

This paper studies the DOAs in heteroscedastic noise environment, including noise Case II-IV. Although the noise variance is a nuisance parameter that we are not interested in, estimating the noise variance is beneficial to DOA estimation. In particular, three algorithms termed as MVHN-S, MVHN-A and MVHN corresponding to noise Case II-IV are proposed. It is shown that the three algorithms can be derived in a unified way, and each algorithm estimates the noise variance in its own way. In addition, MVALSE, MVHN-S and MVHN-A can be obtained through MVHN by averaging the noise variances estimates over antennas and snapshots, antennas, snapshots, respectively. To provide a benchmark performance of the three algorithms, the lower bound of unbiased estimator, i.e., the Cramér Rao bound (CRB) is derived. Given that noise variance follows either Case II, Case III or Case IV, the corresponding proposed three algorithms perform better than MVALSE which does not take the variation of noise variance into account. Besides, MVHN and its variants are also extended to deal with the incomplete measurements scenario and measurements with outliers. Substantial numerical experiments are conducted to demonstrate the superior performance of the proposed algorithms, and it is shown that for Case I-III, the corresponding algorithms approach the CRB. While for Case IV, there exists a gap between MVHN and the corresponding CRB. Finally, performance of MVHN and its variants are demonstrated by applying on a real experimental data set focusing on DOA application.

Notation: Let \mathcal{M} and \mathcal{N} be the subsets of $\{1, \dots, M\}$ and $\{1, \dots, N\}$, and $|\mathcal{M}|$ denotes its cardinality. For a matrix $\mathbf{A} \in \mathbb{C}^{M \times N}$, let $\mathbf{A}_{\mathcal{M}, \mathcal{N}}$ and $[\mathbf{A}]_{\mathcal{M}, \mathcal{N}}$ denote the submatrix by choosing the rows and columns indexed by \mathcal{M} and \mathcal{N} , respectively. For $M = N$ and $\mathcal{M} = \mathcal{N}$, $\mathbf{A}_{\mathcal{M}}$ and $[\mathbf{A}]_{\mathcal{M}}$ denote the submatrix by choosing both rows and columns indexed by \mathcal{M} . Let \mathbf{I}_L denote the identity matrix of dimension L . $(\cdot)^*$, $(\cdot)^T$ and $(\cdot)^H$ are the conjugate, transpose and Hermitian transpose operator, respectively. Let $\Re\{\cdot\}$ return the real part. $\mathcal{CN}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote the complex normal distribution of \mathbf{x} with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. $\|\cdot\|_0$ denotes the number of nonzero elements.

II. PROBLEM SETUP

Consider a HLA with M antennas uniformly spaced with a half wavelength separation $d = \lambda/2$ and L snapshots are available. For the l th snapshot, the noisy measurement $\mathbf{y}_l \in \mathbb{C}^M$ can be described as

$$\mathbf{y}_l = \tilde{\mathbf{z}}_l + \tilde{\mathbf{w}}_l, \quad l = 1, \dots, L, \quad (1)$$

where $\{\tilde{\mathbf{z}}_l\}_{l=1}^L$ denotes the noiseless signal defined by

$$\tilde{\mathbf{z}}_l \triangleq \sum_{k=1}^K \mathbf{a}(\tilde{\theta}_k) \tilde{x}_{k,l}, \quad (2)$$

$\tilde{\theta}_k \in [-90, 90]^\circ$ and $\tilde{x}_{k,l}$ denote the k th frequency and the complex weight coefficient, respectively, $\mathbf{a}(\theta)$ is the array steering vector defined as

$$\begin{aligned} \mathbf{a}(\theta) &= [1, e^{j2\pi d \sin \theta / \lambda}, \dots, e^{j2\pi(M-1)d \sin \theta / \lambda}]^T \\ &= [1, e^{j\pi \sin \theta}, \dots, e^{j\pi(M-1) \sin \theta}]^T, \end{aligned} \quad (3)$$

$\tilde{\mathbf{w}}_l \in \mathbb{C}^{M \times 1}$ is the additive white Gaussian noise independent of the snapshot l . Let $\tilde{w}_{m,l}$ denote the m th element of $\tilde{\mathbf{w}}_l$ satisfying $\tilde{w}_{m,l} \sim \mathcal{CN}(\tilde{w}_{m,l}; 0, \nu_{m,l})$. For the variance assumptions of four cases described in Section I, the variances can be mathematically formulated as follows [2]:

- Case I: Noise variance $\nu \triangleq \nu_{m,l}, \forall m, \forall l$ are the same for both different antennas and snapshots.
- Case II: Noise variance $\nu_l \triangleq \nu_{m,l}, \forall m$ are the same for different antennas, i.e., noise variance depends only on snapshots.
- Case III: Noise variance $\nu_m \triangleq \nu_{m,l}, \forall l$ are the same for different snapshots, i.e., noise variance depends only on antennas.
- Case IV: Noise variance $\nu_{m,l}$ is heteroscedastic across both antennas and snapshots, i.e., noise variance depends on both snapshots and antennas.

The goal of this paper is to estimate the number of sources K , DOAs $\tilde{\boldsymbol{\theta}}$, the complex weight coefficients $\{\tilde{\mathbf{x}}_l\}_{l=1}^L$, and the noiseless signal

$$\tilde{\mathbf{Z}} \triangleq [\tilde{\mathbf{z}}_1, \dots, \tilde{\mathbf{z}}_L] \in \mathbb{C}^{M \times L} \quad (4)$$

for Case II-IV. In the ensuing Section, MVHN for Case IV is derived and the relationship between MVHN and its variants are then revealed.

Before deriving the MVHN, the ratio of the number of unknowns to the number of measurements defined as γ are calculated and summarized in Table I. In general, the larger the γ is, the more likely the algorithm tend to overfit. Thus it is expected that once the spectral is estimated well, the uncertain degrees of the frequencies will be the smallest. Besides, when the noise assumptions are I, II or III, the

TABLE I
THE NUMBER OF UNKNOWN PARAMETERS FOR CASE I-IV

Scenario	Case I	Case II	Case III	Case IV
γ	$K/(2ML) + K/M + 1/(2\mathbf{M}\mathbf{L})$	$K/(2ML) + K/M + 1/(2\mathbf{M})$	$K/(2ML) + K/M + 1/(2\mathbf{L})$	$K/(2ML) + K/M + 1/2$

estimates of MVHN is more likely to be biased as the noise variances are estimated for each snapshot and antenna.

III. MVHN ALGORITHM

In this section, MVHN algorithm for noise Case IV is developed. First, the probabilistic formulation similar to [15] is introduced. Then, MVHN is developed.

Before deriving the MVHN algorithm, we reparameterize the model by defining

$$\omega = \pi \sin \theta, \quad (5)$$

where $\omega \in [-\pi, \pi]$ is termed as the frequency. Note that ω and θ is a one-to-one correspondence, and θ can be calculated through ω as $\theta = \sin^{-1}(\omega/\pi)$, where $\sin^{-1}(\cdot)$ denotes the inverse of $\sin(\cdot)$. In the following, ω is inferred instead.

A. Probabilistic Formulation

Since the number of sources K is unknown, an over complete model is imposed, i.e., the number of sources is assumed to be N and $K \leq N \leq M$. As a result, a pseudo observation model for the l th snapshot is obtained as

$$\mathbf{y}_l = \sum_{k=1}^N \mathbf{a}(\omega_k) x_{k,l} + \mathbf{w}_l = \mathbf{A}(\boldsymbol{\omega}) \mathbf{x}_l + \mathbf{w}_l, \quad l = 1, \dots, L, \quad (6)$$

where \mathbf{x}_l denote the complex amplitude of the l th snapshot and $\mathbf{x}_l = [x_{1,l}, \dots, x_{N,l}]^T \in \mathbb{C}^N$, $\mathbf{A}(\boldsymbol{\omega}) \in \mathbb{C}^{M \times N}$ is

$$\mathbf{A}(\boldsymbol{\omega}) = [\mathbf{a}(\omega_1), \dots, \mathbf{a}(\omega_N)] \in \mathbb{C}^{M \times N}. \quad (7)$$

For the notation simplicity, the array model can be described as

$$\mathbf{Y} = \mathbf{A}(\boldsymbol{\omega}) \mathbf{X} + \mathbf{W}, \quad (8)$$

where $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_L]$ is the measurements, $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_L]$ is the complex weight coefficient matrix and $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_L]$ is the noise.

For the k th source, let the prior distribution be $p(\omega_k)$. In general, uninformative prior distribution is used and $p(\omega_k) = 1/(2\pi)$. For the proposed over complete model, binary hidden variables $\{s_k\}_{k=1}^N \in \{0, 1\}^N$ are introduced to promote the sparsity. Specifically, let $s_k = 1$ denote the k th frequency being active, i.e., the complex weight coefficient $\mathbf{X}_{k,:}$ satisfies $\|\mathbf{X}_{k,:}\|_2 \neq 0$, otherwise deactive and $\|\mathbf{X}_{k,:}\|_2 = 0$. Given s_k , the complex weight coefficient $\mathbf{X}_{k,:}$ is supposed to follow Bernoulli Gaussian distribution

$$p(\mathbf{X}_{k,:}|s_k;\tau) = (1 - s_k)\delta(\mathbf{X}_{k,:}) + s_k\mathcal{CN}(\mathbf{X}_{k,:}; \mathbf{0}, \tau\mathbf{I}_L). \quad (9)$$

For the prior distribution of s_k , Bernoulli distribution is used. Let ρ denote the probability that the k th component is active, i.e.,

$$p(s_k; \rho) = \rho^{s_k}(1 - \rho)^{1-s_k}. \quad (10)$$

From measurement model (6), the likelihood function $p(\mathbf{Y}|\boldsymbol{\omega}, \mathbf{X}; \boldsymbol{\nu})$ is

$$p(\mathbf{Y}|\boldsymbol{\omega}, \mathbf{X}; \boldsymbol{\nu}) = \prod_{l=1}^L \mathcal{CN}(\mathbf{y}_l; \mathbf{A}(\boldsymbol{\omega})\mathbf{x}_l, \boldsymbol{\Sigma}_l). \quad (11)$$

where $\boldsymbol{\nu} = [\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_L] \in \mathbb{C}^{M \times L}$, $\boldsymbol{\Sigma}_l = \text{diag}(\boldsymbol{\nu}_l)$ and $\boldsymbol{\nu}_l = [\nu_{l,1}, \dots, \nu_{l,M}]^T$. As a result, the type II ML estimation of the nuisance parameters are

$$(\hat{\rho}_{\text{ML}}, \hat{\tau}_{\text{ML}}, \hat{\boldsymbol{\nu}}_{\text{ML}}) = \underset{\rho, \tau, \boldsymbol{\nu}}{\text{argmax}} p(\mathbf{Y}; \rho, \tau, \boldsymbol{\nu}), \quad (12)$$

where $p(\mathbf{Y}; \rho, \tau, \boldsymbol{\nu})$ is the marginalized likelihood function

$$\begin{aligned} p(\mathbf{Y}; \rho, \tau, \boldsymbol{\nu}) \\ = \int p(\mathbf{Y}|\boldsymbol{\omega}, \mathbf{X}; \boldsymbol{\nu}) \prod_{k=1}^N (p(\omega_k)p(\mathbf{X}_{k,:}|s_k;\tau)p(s_k;\rho)) \text{d}\mathbf{s} \text{d}\mathbf{X} \text{d}\boldsymbol{\omega}. \end{aligned} \quad (13)$$

Given that $(\hat{\rho}_{\text{ML}}, \hat{\tau}_{\text{ML}}, \hat{\boldsymbol{\nu}}_{\text{ML}})$ are estimated, the maximum a posterior (MAP) estimate is

$$(\hat{\boldsymbol{\omega}}_{\text{MAP}}, \hat{\mathbf{X}}_{\text{MAP}}, \hat{\mathbf{s}}_{\text{MAP}}) = \underset{\boldsymbol{\omega}, \mathbf{X}, \mathbf{s}}{\text{argmax}} p(\boldsymbol{\omega}, \mathbf{X}, \mathbf{s}|\mathbf{Y}; \hat{\rho}_{\text{ML}}, \hat{\tau}_{\text{ML}}, \hat{\boldsymbol{\nu}}_{\text{ML}}), \quad (14)$$

where the posterior PDF $p(\boldsymbol{\omega}, \mathbf{X}, \mathbf{s}|\mathbf{Y}; \hat{\rho}_{\text{ML}}, \hat{\tau}_{\text{ML}}, \hat{\boldsymbol{\nu}}_{\text{ML}})$ is

$$p(\boldsymbol{\omega}, \mathbf{X}, \mathbf{s}|\mathbf{Y}; \hat{\rho}_{\text{ML}}, \hat{\tau}_{\text{ML}}, \hat{\boldsymbol{\nu}}_{\text{ML}}) \propto p(\mathbf{Y}|\boldsymbol{\omega}, \mathbf{X}; \hat{\boldsymbol{\nu}}_{\text{ML}}) \prod_{k=1}^N (p(\omega_k)p(\mathbf{X}_{k,:}|s_k;\hat{\tau}_{\text{ML}})p(s_k;\hat{\rho}_{\text{ML}})) \quad (15)$$

Obviously, solving either (12) or (15) is intractable. As a result, a variational Bayesian approach is adopted.

Let $\boldsymbol{\Theta} = (\omega_1, \dots, \omega_N, (\mathbf{X}, \mathbf{s}))$ be the set of all latent variables. MVHN iteratively optimizes \mathcal{L} over each factor $q(\boldsymbol{\Theta}_i|\mathbf{Y})$, $i = 1, \dots, N + 1$ separately with others being fixed. For the MVHN algorithm, it

approximates the posterior PDF $p(\Theta|\mathbf{Y})$ as $q(\Theta|\mathbf{Y})$ by minimizing $\text{KL}(q(\Theta|\mathbf{Y})||p(\Theta|\mathbf{Y}))$ which equals to maximize [18, pp. 732-733]

$$\mathcal{L}(q(\Theta|\mathbf{Y})) = \mathbb{E}_{q(\Theta|\mathbf{Y})} \left[\ln \frac{p(\mathbf{Y}, \Theta)}{q(\Theta|\mathbf{Y})} \right]. \quad (16)$$

Here $q(\Theta|\mathbf{Y})$ is supposed to have the following structure

$$q(\Theta|\mathbf{Y}) = \prod_{k=1}^N q(\omega_k|\mathbf{Y})q(\mathbf{X}|\mathbf{Y}, \mathbf{s})q(\mathbf{s}), \quad (17)$$

where $q(\mathbf{s})$ is a delta function given by $q(\mathbf{s}) = \delta(\mathbf{s} - \mathbf{s}_0)$, and the joint PDF of \mathbf{Y} and Θ is

$$p(\mathbf{Y}, \Theta) = p(\mathbf{Y}|\omega, \mathbf{X}; \nu) \prod_{k=1}^N (p(\omega_k)p(\mathbf{X}_{k,:}|s_k; \tau)p(s_k; \rho)). \quad (18)$$

Maximizing \mathcal{L} with respect to all the factors is intractable. Thus $q(\Theta_i|\mathbf{Y})$ is calculated separately and the posterior approximation $q(\Theta|\mathbf{Y})$ is calculated as [18, pp. 735, eq. (21.25)]

$$\ln q(\Theta_i|\mathbf{Y}) = \mathbb{E}_{q(\Theta \setminus \Theta_i|\mathbf{Y})} [\ln p(\mathbf{Y}, \Theta)] + \text{const}, \quad (19)$$

where the expectation is taken with respect to all the variables Θ except Θ_i .

Before deriving the MVHN algorithm, some definitions are introduced. $\hat{\omega}_i$ is defined as the mean direction of $e^{j\omega_i}$ [17] and $\hat{\mathbf{a}}_i$ is the estimation of $\mathbf{a}(\tilde{\omega}_i)$ which will be used to give the estimation of weights and reconstructed signals, i.e.,

$$\hat{\omega}_i = \arg \left(\mathbb{E}_{q(\omega_i|\mathbf{Y})} [e^{j\omega_i}] \right), \quad (20a)$$

$$\hat{\mathbf{a}}_i = \mathbb{E}_{q(\omega_i|\mathbf{Y})} [\mathbf{a}(\omega_i)], \quad i \in \{1, \dots, N\}. \quad (20b)$$

We denote $\hat{\mathbf{A}} = [\hat{\mathbf{a}}_1, \dots, \hat{\mathbf{a}}_N]$. The posterior PDF of \mathbf{X} is

$$q(\mathbf{X}|\mathbf{Y}) = \int q(\mathbf{X}, \mathbf{s}|\mathbf{Y})\delta(\mathbf{s} - \mathbf{s}_0)d\mathbf{s} = q(\mathbf{X}|\mathbf{Y}; \mathbf{s}_0). \quad (21)$$

Analogously, the mean and covariance of the weights for the l th snapshot are estimated as

$$\hat{\mathbf{x}}_l = \mathbb{E}_{q(\mathbf{X}|\mathbf{Y})} [\mathbf{x}_l], \quad (22a)$$

$$\hat{\mathbf{C}}_l = \mathbb{E}_{q(\mathbf{X}|\mathbf{Y})} [\mathbf{x}_l \mathbf{x}_l^H] - \hat{\mathbf{x}}_l \hat{\mathbf{x}}_l^H, \quad l = 1, \dots, L, \quad (22b)$$

and $\hat{\mathbf{X}} = [\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_L]$. Let \mathcal{S} be the set of the active element indices of \mathbf{s} , i.e.,

$$\mathcal{S} = \{i | 1 \leq i \leq N, s_i = 1\}, \quad (23)$$

and $\hat{\mathbf{s}}$ be the estimate of \mathbf{s} , then the estimated model order is $\hat{K} = |\hat{\mathcal{S}}|$. According to (6), the noise-free signal is reconstructed as

$$\hat{\mathbf{Z}} = \hat{\mathbf{A}}_{:, \hat{\mathcal{S}}} \hat{\mathbf{X}}_{\hat{\mathcal{S}}, :}. \quad (24)$$

B. Inferring the Posterior PDF of Frequencies

In this section, we maximize \mathcal{L} with respect to the factor $q(\omega_i|\mathbf{Y})$ for $i = 1, \dots, N$. For $i \notin \mathcal{S}$, $q(\omega_i|\mathbf{Y})$ is kept unchanged during iteration. According to (19), for $i \in \mathcal{S}$, $\ln q(\omega_i|\mathbf{Y})$ can be calculated as

$$\begin{aligned}
\ln q(\omega_i|\mathbf{Y}) &= \mathbb{E}_{q(\boldsymbol{\Theta} \setminus \omega_i|\mathbf{Y})} [\ln p(\mathbf{Y}, \boldsymbol{\Theta})] + \text{const} \\
&= \mathbb{E}_{q(\boldsymbol{\Theta} \setminus \omega_i|\mathbf{Y})} [\ln(p(\boldsymbol{\omega})p(\mathbf{s})p(\mathbf{X}|\mathbf{s})p(\mathbf{Y}|\boldsymbol{\omega}, \mathbf{X}))] + \text{const} \\
&= \mathbb{E}_{q(\boldsymbol{\Theta} \setminus \omega_i|\mathbf{Y})} \left[\sum_{j=1}^N \ln p(\omega_j) + \sum_{j=1}^N \ln p(s_j) + \ln p(\mathbf{X}|\mathbf{s}) \right] + \mathbb{E}_{q(\boldsymbol{\Theta} \setminus \omega_i|\mathbf{Y})} \left[\sum_{l=1}^L \ln p(\mathbf{y}_l|\boldsymbol{\omega}, \mathbf{x}_l; \boldsymbol{\Sigma}_l) \right] + \text{const} \\
&= \mathbb{E}_{q(\boldsymbol{\Theta} \setminus \omega_i|\mathbf{Y})} \left[\sum_{l=1}^L \left(\mathbf{y}_l - \sum_{k \in \mathcal{S}} \mathbf{a}(\omega_k) x_{k,l} \right)^H \boldsymbol{\Sigma}_l^{-1} \left(\mathbf{y}_l - \sum_{k \in \mathcal{S}} \mathbf{a}(\omega_k) x_{k,l} \right) \right] + \ln p(\omega_i) + \text{const} \\
&= \ln p(\omega_i) + \Re \{ \boldsymbol{\eta}_i^H \mathbf{a}(\omega_i) \} + \text{const}, \tag{25}
\end{aligned}$$

where the complex vector $\boldsymbol{\eta}_i$ is

$$\boldsymbol{\eta}_i = \sum_{l=1}^L \boldsymbol{\eta}_{i,l}, \tag{26}$$

and $\boldsymbol{\eta}_{i,l}$ is

$$\boldsymbol{\eta}_{i,l} = 2\boldsymbol{\Sigma}_l^{-1} \left[\left(\mathbf{y}_l - \sum_{j \neq i} \hat{\mathbf{a}}_j \hat{x}_{j,l} \right) \hat{x}_{i,l}^* - \sum_{j \neq i} [\hat{\mathbf{C}}_l]_{j,i} \hat{\mathbf{a}}_j \right], \tag{27}$$

which can be viewed as the weighted sum of $2((\mathbf{y}_l - \sum_{j \neq i} \hat{\mathbf{a}}_j \hat{x}_{j,l}) \hat{x}_{i,l}^* - \sum_{j \neq i} [\hat{\mathbf{C}}_l]_{j,i} \hat{\mathbf{a}}_j)$ with respect to the inverse noise variance $\text{diag}(\boldsymbol{\Sigma}_l^{-1})$. Given the posterior distribution (25), (20) can not be evaluated in closed form. Thus, by referring to [15, Heuristic 2], $q(\omega_i|\mathbf{Y})$ is approximated as a von Mises distribution

$$q(\omega_i|\mathbf{Y}) \approx \mathcal{VM}(\omega; \hat{\mu}_i, \hat{\kappa}_i), \tag{28}$$

which yields analytical results (20), where

$$\mathcal{VM}(\omega; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\omega - \mu)}, \tag{29}$$

where μ and κ are the mean direction and concentration parameters, $I_p(\cdot)$ is the modified Bessel function of the first kind and the order p [17, p. 348]. For a given von Mises distribution, $\arg(\mathbb{E}_{\mathcal{VM}(\omega; \mu, \kappa)}[e^{j\omega}]) =$

$\arg\left(e^{j\mu\frac{I_1(\kappa)}{I_0(\kappa)}}\right) = \mu = E_{\mathcal{VM}(\omega;\mu,\kappa)}[\omega]$. In addition, $E[e^{jm\omega}] = e^{jm\mu}(I_m(\kappa)/I_0(\kappa))$ [17, pp. 26]. Therefore, $\hat{\omega}_i$ and $\hat{\mathbf{a}}_i$ (20) can be calculated analytically¹.

C. Inferring the Posterior PDF of Weights and Support

Then \mathcal{L} is maximized w.r.t. $q(\mathbf{X}, \mathbf{s}|\mathbf{Y})$. For $l = 1, \dots, L$, define the matrices \mathbf{J}_l and \mathbf{H} as

$$[\mathbf{J}_l]_{i,j} = \begin{cases} \text{tr}(\boldsymbol{\Sigma}_l^{-1}), & i = j \\ \hat{\mathbf{a}}_i^H \boldsymbol{\Sigma}_l^{-1} \hat{\mathbf{a}}_j, & i \neq j \end{cases}, \quad i, j \in \{1, \dots, N\}, \quad (30a)$$

$$\mathbf{H}_{:,l} = \hat{\mathbf{A}}^H \boldsymbol{\Sigma}_l^{-1} \mathbf{y}_l, \quad (30b)$$

where $[\mathbf{J}_l]_{i,j}$ denotes the (i, j) th element of \mathbf{J}_l .

According to (19), $q(\mathbf{X}, \mathbf{s}|\mathbf{Y})$ can be calculated as

$$\begin{aligned} \ln q(\mathbf{X}, \mathbf{s}|\mathbf{Y}) &= E_{q(\boldsymbol{\Theta}|\mathbf{X}, \mathbf{s}|\mathbf{Y})} [\ln p(\mathbf{Y}, \boldsymbol{\Theta})] + \text{const} \\ &= E_{q(\omega|\mathbf{Y})} \left[\sum_{i=1}^N \ln p(s_i) + \ln p(\mathbf{X}|\mathbf{s}) + \ln p(\mathbf{Y}|\omega, \mathbf{X}) \right] + \text{const} \\ &= \|\mathbf{s}\|_0 \ln \frac{\rho}{1-\rho} + \|\mathbf{s}\|_0 L \ln \frac{1}{\pi\tau} - \frac{1}{\tau} \sum_{k \in \mathcal{S}} \mathbf{X}_{k,:} \mathbf{X}_{k,:}^H + \text{const} \\ &\quad - \sum_{l=1}^L E_{q(\omega|\mathbf{Y})} \left[\left(\mathbf{y}_l - \sum_{k \in \mathcal{S}} \mathbf{a}(\omega_k) x_{k,l} \right)^H \boldsymbol{\Sigma}_l^{-1} \left(\mathbf{y}_l - \sum_{k \in \mathcal{S}} \mathbf{a}(\omega_k) x_{k,l} \right) \right] \\ &\stackrel{a}{=} \|\mathbf{s}\|_0 \ln \frac{\rho}{1-\rho} + \|\mathbf{s}\|_0 L \ln \frac{1}{\pi\tau} - \frac{1}{\tau} \sum_{l=1}^L \mathbf{X}_{S,l}^H \mathbf{X}_{S,l} + \sum_{l=1}^L \left[2\Re\{\mathbf{X}_{S,l}^H \mathbf{H}_{S,l}\} - \mathbf{X}_{S,l}^H [\mathbf{J}_l]_{\mathcal{S}} \mathbf{X}_{S,l} \right] + \text{const} \\ &= \sum_{l=1}^L (\mathbf{X}_{S,l} - \hat{\mathbf{X}}_{S,l})^H [\hat{\mathbf{C}}_l]_{\mathcal{S}}^{-1} (\mathbf{X}_{S,l} - \hat{\mathbf{X}}_{S,l}) + \text{const}, \end{aligned} \quad (31)$$

where $\stackrel{a}{=}$ utilizes $\sum_{k \in \mathcal{S}} \mathbf{X}_{k,:} \mathbf{X}_{k,:}^H = \sum_{l=1}^L \mathbf{X}_{S,l}^H \mathbf{X}_{S,l}$ and

$$\hat{\mathbf{X}}_{S,l} = [\hat{\mathbf{C}}_l]_{\mathcal{S}} \mathbf{H}_{S,l}, \quad (32a)$$

$$[\hat{\mathbf{C}}_l]_{\mathcal{S}} = \left([\mathbf{J}_l]_{\mathcal{S}} + \frac{\mathbf{I}_{|\mathcal{S}|}}{\tau} \right)^{-1}, \quad l = 1, \dots, L. \quad (32b)$$

¹An alternative approach is to approximate $q(\omega_i|\mathbf{Y})$ as $q(\omega_i|\mathbf{Y}) \approx \delta(\omega_i - \hat{\mu}_i)$, which corresponds to the point estimates of the frequencies, and then $\hat{\mathbf{a}}_i$ is calculated to be $\mathbf{a}(\hat{\omega}_i)$. This estimation approach yields the VALSE-pt algorithm [15]. Actually, the larger κ of von Mises distribution, the closer $|E_{q(\omega|\mathbf{Y})}[e^{j\omega}]|$ is to one. Conversely, the smaller κ is, the closer $|E_{q(\omega|\mathbf{Y})}[e^{j\omega}]|$ is to zero. Hence, for VALSE algorithm, $\hat{\mathbf{a}}_i$ reflects the uncertainty of frequency estimation which will influence the model order estimation. For VALSE-pt, it gives full certainty for all frequencies which might overestimate the model order. As shown in [15], VALSE-pt performs worse than VALSE. Hence, the first approach is adopted in this paper.

According to (21), to calculate $q(\mathbf{X}|\mathbf{Y})$, \mathbf{s}_0 has to be given. Plugging the postulated PDF $q(\boldsymbol{\Theta}|\mathbf{Y})$ (17) in (16), one has

$$\begin{aligned} \ln Z(\mathbf{s}_0) &\triangleq \mathcal{L}(q(\boldsymbol{\Theta}|\mathbf{Y}); \mathbf{s}_0) = \mathbb{E}_{q(\boldsymbol{\Theta}|\mathbf{Y})} \left[\ln \frac{p(\mathbf{Y}, \boldsymbol{\Theta}; \mathbf{s}_0)}{q(\boldsymbol{\Theta}|\mathbf{Y}; \mathbf{s}_0)} \right] \\ &= \mathbb{E}_{q(\boldsymbol{\Theta}|\mathbf{Y})} [\ln p(\mathbf{s}) + \ln p(\mathbf{X}|\mathbf{s})] + \mathbb{E}_{q(\boldsymbol{\Theta}|\mathbf{Y})} [\ln p(\mathbf{Y}|\boldsymbol{\omega}, \mathbf{X}) - \ln q(\mathbf{X}|\mathbf{Y})] + \text{const} \\ &= \|\mathbf{s}_0\|_0 \ln \frac{\rho}{1-\rho} - \sum_{l=1}^L [\ln \det([\mathbf{J}_l]_{\mathcal{S}_0} + \frac{1}{\tau} \mathbf{I}_{|\mathcal{S}_0|}) + \|\mathbf{s}_0\|_0 \ln \frac{1}{\tau} + (\mathbf{H}_{\mathcal{S}_0, l}^H([\mathbf{J}_l]_{\mathcal{S}_0} + \frac{1}{\tau} \mathbf{I}_{|\mathcal{S}_0|})^{-1} \mathbf{H}_{\mathcal{S}_0, l})] + \text{const}. \end{aligned} \quad (33)$$

where \mathcal{S}_0 is the set of the active element indices of \mathbf{s}_0 . Thus \mathbf{s}_0 should be chosen to maximize $\ln Z(\mathbf{s}_0)$ (33), i.e.,

$$\hat{\mathbf{s}}_0 = \underset{\mathbf{s}_0}{\operatorname{argmax}} \ln Z(\mathbf{s}_0), \quad (34)$$

A naive approach to solve the above problem is to enumerate all the possible binary values of \mathbf{s}_0 , which costs $O(2^N)$ and is impractical for typical values of N . To reduce the computation complexity, a greedy iterative search strategy is proposed to find a local optimum. Given \mathbf{s}_0 , the strategy proceeds as follows: For each $k = 1, \dots, N$, calculate $\Delta_k = \ln Z(\mathbf{s}_0^k) - \ln Z(\mathbf{s}_0)$, where \mathbf{s}_0^k is the same as \mathbf{s}_0 except that the k th element of \mathbf{s}_0 is flipped. Let $k^* = \underset{k}{\operatorname{argmax}} \Delta_k$. If $\Delta_{k^*} > 0$, we update \mathbf{s}_0 with the k^* th element flipped, and \mathbf{s}_0 is updated, otherwise $\hat{\mathbf{s}}_0$ is obtained as \mathbf{s}_0 , and the algorithm is terminated. In fact, Δ_k can be easily calculated and the details are provided in Appendix IX-A. Numerically, the computational complexity of the greedy approach is about $O(\hat{K})$.

D. Estimating the Model Parameters

After updating the frequencies and weights, the model parameters $\boldsymbol{\beta} = \{\boldsymbol{\nu}, \rho, \tau\}$ is estimated via maximizing the lower bound $\mathcal{L}(q(\boldsymbol{\Theta}|\mathbf{Y}); \boldsymbol{\beta})$. Plugging the postulated PDF (17) in (16), $\mathcal{L}(q(\boldsymbol{\Theta}|\mathbf{Y}); \boldsymbol{\beta})$ is

$$\begin{aligned} \mathcal{L}(q(\boldsymbol{\Theta}|\mathbf{Y}); \boldsymbol{\beta}) &= \mathbb{E}_{q(\boldsymbol{\Theta}|\mathbf{Y})} \left[\ln \frac{p(\mathbf{Y}, \boldsymbol{\Theta}; \boldsymbol{\beta})}{q(\boldsymbol{\Theta}|\mathbf{Y})} \right] \\ &= \mathbb{E}_{q(\boldsymbol{\Theta}|\mathbf{Y})} [\ln p(\mathbf{s}) + \ln p(\mathbf{X}|\mathbf{s}) + \ln p(\mathbf{Y}|\boldsymbol{\omega}, \mathbf{X})] + \text{const} \\ &= \|\hat{\mathbf{s}}\|_0 \ln \rho - \|\hat{\mathbf{s}}\|_0 \ln(1-\rho) + \|\hat{\mathbf{s}}\|_0 L \ln \frac{1}{\pi\tau} - \mathbb{E}_{q(\mathbf{X}|\mathbf{Y})} \left[\frac{1}{\tau} \operatorname{tr}(\mathbf{X}_{\hat{\mathcal{S}},:} \mathbf{X}_{\hat{\mathcal{S}},:}^H) \right] - \sum_{l=1}^L \mathbb{E}_{q(\mathbf{X}|\mathbf{Y})} \left[\mathbf{X}_{\hat{\mathcal{S}}, l}^H [\mathbf{J}_l]_{\hat{\mathcal{S}}} \mathbf{X}_{\hat{\mathcal{S}}, l} \right] \\ &\quad + \sum_{l=1}^L \left(\ln \frac{1}{\det(\pi \boldsymbol{\Sigma}_l)} - \mathbf{y}_l^H \boldsymbol{\Sigma}_l^{-1} \mathbf{y}_l + 2\Re\{(\hat{\mathbf{X}}_{\hat{\mathcal{S}}, l}^H \mathbf{H}_{\hat{\mathcal{S}}, l})\} \right) + \text{const}. \end{aligned} \quad (35)$$

Substituting $E_{q(\mathbf{X}|\mathbf{Y})}[\text{tr}(\mathbf{X}_{\hat{\mathcal{S}},:} \mathbf{X}_{\hat{\mathcal{S}},:}^H)] = \text{tr}(\hat{\mathbf{X}}_{\hat{\mathcal{S}},:}^H \hat{\mathbf{X}}_{\hat{\mathcal{S}},:}) + \sum_{l=1}^L \text{tr}([\hat{\mathbf{C}}_l]_{\hat{\mathcal{S}}})$ and $E_{q(\mathbf{X}|\mathbf{Y})}[\mathbf{X}_{\hat{\mathcal{S}},l}^H [\mathbf{J}_l]_{\hat{\mathcal{S}}} \mathbf{X}_{\hat{\mathcal{S}},l}] = \text{tr}([\mathbf{J}_l]_{\hat{\mathcal{S}}}(\hat{\mathbf{X}}_{\hat{\mathcal{S}},l} \hat{\mathbf{X}}_{\hat{\mathcal{S}},l}^H + [\hat{\mathbf{C}}_l]_{\hat{\mathcal{S}}}))$ in (35), we have

$$\begin{aligned} & \mathcal{L}(q(\Phi|\mathbf{Y}); \beta) \\ &= \sum_{l=1}^L \left(\ln \frac{1}{\det(\pi \Sigma_l)} - \mathbf{y}_l^H \Sigma_l^{-1} \mathbf{y}_l + 2\Re\{(\hat{\mathbf{X}}_{\hat{\mathcal{S}},l}^H \mathbf{H}_{\hat{\mathcal{S}},l})\} \right) + N \ln(1 - \rho) \\ & - \sum_{l=1}^L \text{tr}([\mathbf{J}_l]_{\hat{\mathcal{S}}}(\hat{\mathbf{X}}_{\hat{\mathcal{S}},l} \hat{\mathbf{X}}_{\hat{\mathcal{S}},l}^H + [\hat{\mathbf{C}}_l]_{\hat{\mathcal{S}}})) \|\hat{\mathbf{s}}\|_0 \left(\ln \frac{\rho}{1 - \rho} - L \ln \tau \right) \\ & - \frac{1}{\tau} \left[\text{tr}(\hat{\mathbf{X}}_{\hat{\mathcal{S}},:} \hat{\mathbf{X}}_{\hat{\mathcal{S}},:}^H) + \sum_{l=1}^L \text{tr}([\hat{\mathbf{C}}_l]_{\hat{\mathcal{S}}}) \right] + \text{const}, \end{aligned} \quad (36)$$

Setting $\frac{\partial \mathcal{L}}{\partial \rho} = 0$ and $\frac{\partial \mathcal{L}}{\partial \tau} = 0$, we have

$$\hat{\rho} = \frac{\|\hat{\mathbf{s}}\|_0}{N}, \quad \hat{\tau} = \frac{\text{tr}(\mathbf{X}_{\hat{\mathcal{S}},:}^H \mathbf{X}_{\hat{\mathcal{S}},:}) + \sum_{l=1}^L \text{tr}([\hat{\mathbf{C}}_l]_{\hat{\mathcal{S}}})}{L \|\hat{\mathbf{s}}\|_0}. \quad (37)$$

Setting $\frac{\partial \mathcal{L}}{\partial \nu_{m,l}} = 0$, $m = 1, \dots, M$, $l = 1, \dots, L$, we obtain

$$\hat{\nu}_{m,l} = |y_{m,l} - \hat{\mathbf{A}}_{i,\hat{\mathcal{S}}} \hat{\mathbf{X}}_{\hat{\mathcal{S}},l}|^2 + \hat{\mathbf{A}}_{i,\hat{\mathcal{S}}} [\hat{\mathbf{C}}_l]_{\hat{\mathcal{S}}} \hat{\mathbf{A}}_{i,\hat{\mathcal{S}}}^H + \sum_{i \in \hat{\mathcal{S}}} |\hat{x}_{i,l}|^2 (1 - |\hat{\mathbf{A}}_{m,i}|^2). \quad (38)$$

It is worth noting that $\hat{\nu}_{m,l}$ consists of three terms, where the first term is the fitting (residue) error, the second term is the error coming from the complex weight \mathbf{W} , and the last term from the frequencies ω . Given that the fitting is perfect, the weight estimate or the frequency is estimated exactly ($\hat{\mathbf{C}}_l \rightarrow \mathbf{0}$ or $\hat{\kappa}_i \rightarrow \infty$ (28)), the corresponding three terms will diminish.

The initializations of MVHN are the same as [16] and MVHN is outlined in Algorithm 1.

Algorithm 1 Outline of MVHN algorithm

Input: Signal matrix \mathbf{Y}

Output: The model order estimate \hat{K} , frequencies estimate $\hat{\omega}_{\hat{\mathcal{S}}}$, complex weights estimate $\hat{\mathbf{X}}_{\hat{\mathcal{S}},:}$ and reconstructed signal $\hat{\mathbf{Z}}$

- 1: Initialize $\hat{\nu}, \hat{\rho}, \hat{\tau}$ and $q(\omega_i|\mathbf{Y}), i \in \{1, \dots, N\}$
 - 2: **repeat**
 - 3: Calculate $q(\mathbf{X}, \mathbf{s}|\mathbf{Y})$ and update $\hat{\mathbf{s}}, \hat{\mathbf{X}}_{\hat{\mathcal{S}},:}$ (Section III-C)
 - 4: Update the parameters $\hat{\rho}, \hat{\tau}$ (37) and $\hat{\nu}_{l,m}$, $l = 1, \dots, L$, $m = 1, \dots, M$ (38)
 - 5: Calculate $q(\omega|\mathbf{Y})$ and update $\hat{\omega}$ (Section III-B)
 - 6: **until** stopping criterion
 - 7: **return** $\hat{K}, \hat{\omega}_{\hat{\mathcal{S}}}, \hat{\mathbf{X}}_{\hat{\mathcal{S}},:}$ and $\hat{\mathbf{Z}}$
-

The computational complexity of Algorithm 1 mainly depends on the estimation of model order in Sec. III-C and the approximation of $q(\omega_i|\mathbf{Y}), i \in \hat{\mathcal{S}}$ as von Mises distribution. According to [15, 16], the complexity is $O(NL\hat{K} + MNL\hat{K})$ for each iteration.

IV. INCOMPLETE MEASUREMENTS AND OUTLIER CASES

The MVHN have been developed in the previous section. Here we show that MVHN can be easily extended to deal with incomplete measurements case or measurements with sparse outliers case.

A. Incomplete Measurement Case

The incomplete measurement model is described as

$$y_{m,l} = \begin{cases} \tilde{z}_{m,l} + \tilde{w}_{m,l}, & \Omega_{m,l} = 1 \\ 0, & \Omega_{m,l} = 0 \end{cases},$$

$$m = 1, \dots, M, l = 1, \dots, L, \quad (39)$$

where Ω is a subset and \mathbf{Y}_Ω denotes the incomplete measurements. Note that DOA estimation from sparse array can be abstracted as the above problem. For the incomplete measurement model (39), the likelihood function is

$$p(\mathbf{Y}|\mathbf{Z}) = \prod_{(m,l) \in \Omega} \mathcal{CN}(y_{m,l}; \tilde{z}_{m,l}, \nu_{m,l}). \quad (40)$$

Substituting (2) and (40) in (12), the marginalized likelihood function is obtained. Then, using the variational approach and (18), we perform the inference process. Fortunately, by carefully comparing the likelihood (40) and (11), it can be shown that (40) can be written in the same form as (11), i.e.,

$$p(\mathbf{Y}|\mathbf{Z}) = \prod_{(m,l) \in \Omega} \mathcal{CN}(y_{m,l}; \tilde{z}_{m,l}, \nu_{m,l}) \prod_{(m',l') \notin \Omega} 1_{m'l'} = \prod_{(m,l) \in \Omega} \mathcal{CN}(y_{m,l}; \tilde{z}_{m,l}, \nu_{m,l}). \quad (41)$$

As a result, only two minor differences are made to ensure that MVHN works in this setting. First, for the initialization of MVHN, we initialize $\nu_{m,l} = \infty, \forall (m,l) \notin \Omega$. Second, the noise variance estimation step is rederived. The noise variance estimate $\hat{\nu}_{m,l}$ is updated as (38) $\forall (m,l) \in \Omega$. For $(m,l) \notin \Omega$, we set $\hat{\nu}_{m,l} = \infty$.

To provide the benchmark performance of the four algorithms for noise Case I-IV, the CRB is derived and is postponed to Appendix IX-B.

B. Robust to Outliers

Here we also study the robust LSE where some measurements is corrupted with (arbitrarily) large errors described by

$$\mathbf{Y} = \tilde{\mathbf{Z}} + \tilde{\mathbf{O}} + \tilde{\mathbf{W}}, \quad (42)$$

where $\tilde{\mathbf{O}}$ is the sparse error matrix with arbitrarily large coefficients, $\tilde{\mathbf{W}}$ is the dense error matrix with small amplitudes.

Note that MVHN has the following properties:

- MVHN estimates the noise variance for each snapshot and measurement.
- Given that MVHN estimates the line spectral with high accuracy, the noise variance estimate (38) is approximately equal to the residue error. Therefore, for the measurements corrupted with outliers, the noise variance estimate of MVHN will be large. As a result, MVHN adaptively enhance the contribution of measurements with small errors and suppress the contribution of measurements with large outliers.

Therefore, it is believed that MVHN is robust to outliers.

Furthermore, the noise variance estimate results is beneficial to identify the outliers. For the measurements with outliers, the noise variance estimates are more likely to be large. Therefore, a threshold can be used to estimate the position of the outliers. Here we design a variant of MVHN algorithm termed as MVHN-ADI, which adaptively identify (ADI) the position of outliers and use the remaining (incomplete) measurements to perform LSE in each iteration. Later, a numerical simulation is conducted to demonstrate the effectiveness of both MVHN and MVHN-ADI.

V. THE VARIANT ALGORITHMS OF MVHN

Now we focus on noise Case II-III, which are special cases of Case IV. The variants of MVHN algorithm, termed as MVHN-S and MVHN-A are derived and relationship between MVALUE, MVHN-S, MVHN-A and MVHN are revealed.

The only one minor difference between MVHN and other three algorithms is the estimation of noise variances in (38). For MVALUE of Case I in [16], we have $\mathbf{J}_1 = \dots = \mathbf{J}_L = \mathbf{J}$ and $\hat{\mathbf{C}}_1 = \dots = \hat{\mathbf{C}}_L = \hat{\mathbf{C}}$. The estimate of noise variance is

$$\hat{\nu} = \|\mathbf{Y} - \hat{\mathbf{A}}_{:, \hat{\mathcal{S}}} \hat{\mathbf{X}}_{\hat{\mathcal{S}},:}\|_{\text{F}}^2 / (ML) + \text{tr}(\mathbf{J}_{\hat{\mathcal{S}}} \hat{\mathbf{C}}_{\hat{\mathcal{S}}}) / M + \sum_{i \in \hat{\mathcal{S}}} \sum_{l=1}^L |\hat{x}_{i,l}|^2 (1 - \|\hat{\mathbf{a}}_i\|_2^2 / M) / L. \quad (43)$$

For Case II, we have $\nu_{1,l} = \dots = \nu_{M,l} = \nu_l$. The estimates of noise variances are

$$\hat{\nu}_l = \frac{1}{M} \sum_{m=1}^M \left(|y_{m,l} - \hat{\mathbf{A}}_{i,\hat{\mathcal{S}}} \hat{\mathbf{X}}_{\hat{\mathcal{S}},l}|^2 + \hat{\mathbf{A}}_{i,\hat{\mathcal{S}}} [\hat{\mathbf{C}}_l]_{\hat{\mathcal{S}}} \hat{\mathbf{A}}_{i,\hat{\mathcal{S}}}^H \right) + \frac{1}{M} \sum_{m=1}^M \sum_{i \in \hat{\mathcal{S}}} |\hat{x}_{i,l}|^2 (1 - |\hat{\mathbf{A}}_{m,i}|^2). \quad (44)$$

For Case III, we have $\nu_{m,l} = \dots = \nu_{m,L} = \nu_m$. The estimates of noise variances are

$$\hat{\nu}_m = \frac{1}{L} \sum_{l=1}^L \left(|y_{m,l} - \hat{\mathbf{A}}_{i,\hat{\mathcal{S}}} \hat{\mathbf{X}}_{\hat{\mathcal{S}},l}|^2 + \hat{\mathbf{A}}_{i,\hat{\mathcal{S}}} [\hat{\mathbf{C}}_l]_{\hat{\mathcal{S}}} \hat{\mathbf{A}}_{i,\hat{\mathcal{S}}}^H \right) + \frac{1}{L} \sum_{l=1}^L \sum_{i \in \hat{\mathcal{S}}} |\hat{x}_{i,l}|^2 (1 - |\hat{\mathbf{A}}_{m,i}|^2). \quad (45)$$

According to (43), (44), (45) and (38), the variance estimates under Case I-III can be derived from Case IV, i.e.,

$$\hat{\nu} = \sum_{m=1}^M \sum_{l=1}^L \hat{\nu}_{m,l} / (ML), \quad (46a)$$

$$\hat{\nu}_l = \sum_{m=1}^M \hat{\nu}_{m,l} / M, \quad (46b)$$

$$\hat{\nu}_m = \sum_{l=1}^L \hat{\nu}_{m,l} / L, \quad (46c)$$

Thus, the estimate of noise variance for Case I is the average of Case IV on both antennas and snapshots from (46a). In addition, the noise variance estimates of Case II and Case III are the average of Case IV on antennas and snapshots from (46b) and (46c), respectively. To sum up, the three algorithms MVALUE, MVHN-S and MVHN-A can be viewed as the variants of MVHN.

For incomplete measurement case in Section IV-A, The noise variance estimates of MVHN-S, MVHN-A and MVALUE algorithms are ²

$$\hat{\nu} = \frac{\sum_{m,l} \Omega_{m,l} \hat{\nu}_{m,l}}{\sum_{m,l} \Omega_{m,l}}, \quad (47a)$$

$$\hat{\nu}_l = \frac{\sum_m \Omega_{m,l} \hat{\nu}_{m,l}}{\sum_m \Omega_{m,l}}, \quad (47b)$$

$$\hat{\nu}_m = \frac{\sum_l \Omega_{m,l} \hat{\nu}_{m,l}}{\sum_l \Omega_{m,l}}, \quad (47c)$$

respectively, which is the weighted averaged of $\hat{\nu}_{m,l}$.

²We define the appropriate arithmetic operations with $0 \cdot \infty = \infty \cdot 0 = 0$.

VI. NUMERICAL SIMULATION

In this section, substantial numerical experiments are conducted to evaluate the performance of the four algorithms, i.e., MVALUE, MVHN-S, MVHN-A and MVHN algorithms. The normalized mean squared error (NMSE) of $\hat{\mathbf{Z}}$ is $\text{NMSE}(\hat{\mathbf{Z}}) \triangleq 10\log(\|\hat{\mathbf{Z}} - \tilde{\mathbf{Z}}\|_{\text{F}}^2 / \|\tilde{\mathbf{Z}}\|_{\text{F}}^2)$ and MSE of $\hat{\omega}$ is $\text{MSE}(\hat{\omega}) \triangleq 10\log(\|\hat{\omega} - \tilde{\omega}\|_2^2)$, the correct model order estimated probability $P(\hat{K} = K)$ are adopted as the performance metrics. The frequency estimation error is averaged over the trials in which $\hat{K} = K$ for a given simulation point.

Simulation Setup: The magnitudes and phases of the complex weight coefficients are generated i.i.d. from normal distribution $\mathcal{N}(1, 0.2)$ and uniform distribution $\mathcal{U}(-\pi, \pi)$, respectively. The frequencies and complex amplitudes are fixed for each MC trial. We define nominal signal-to-noise ratio (SNR) as $\text{SNR} \triangleq 10\log(\|\tilde{\mathbf{Z}}\|_{\text{F}}^2 / (\tilde{\nu}_0 ML))$, where $\tilde{\nu}_0$ is the nominal noise variance and we denote $\tilde{\nu}_{0,\text{dB}} = 10\log \tilde{\nu}_0$. The noise variance of four cases are generated as follows:

Case I: Noise variance $\tilde{\nu}$ equals to nominal noise variance, i.e., $\tilde{\nu} = \tilde{\nu}_0$.

Case II: $10\log \tilde{\nu}_l$ is generated from uniform distribution $\mathcal{U}(\tilde{\nu}_{0,\text{dB}} - \Delta_\nu/2, \tilde{\nu}_{0,\text{dB}} + \Delta_\nu/2)$, $l = 1, \dots, L$.

Case III: $10\log \tilde{\nu}_m$ is generated from uniform distribution $\mathcal{U}(\tilde{\nu}_{0,\text{dB}} - \Delta_\nu/2, \tilde{\nu}_{0,\text{dB}} + \Delta_\nu/2)$, $m = 1, \dots, M$.

Case IV: $10\log \tilde{\nu}_{m,l}$ is generated from uniform distribution $\mathcal{U}(\tilde{\nu}_{0,\text{dB}} - \Delta_\nu/2, \tilde{\nu}_{0,\text{dB}} + \Delta_\nu/2)$, $m = 1, \dots, M$, $l = 1, \dots, L$,

where Δ_ν is the strength of noise fluctuation which characterizes the fluctuation of noise variances. The Algorithm 1 stops when $\|\hat{\mathbf{X}}^{(t-1)} - \hat{\mathbf{X}}^{(t)}\|_2 / \|\hat{\mathbf{X}}^{(t-1)}\|_2 < 10^{-6}$ or $t > 500$, where t is the number of iteration.

Six numerical simulations are conducted in this section. For the first simulation, the posterior PDFs output by all the algorithms under Case I-IV are illustrated. For the second and third simulations, the performances of all the algorithms versus the nominal SNR or the fluctuation strength of the noise variance are investigated. Then, estimation from incomplete measurements or measurements with outliers are evaluated for the fourth and fifth simulation. For the last simulation, DOA estimation using real data is performed.

A. The Posterior PDF Output by the Four Algorithms

Before moving on, we first look into the detailed reconstructed results and present the posterior PDFs of all the algorithms for the four Noise Cases. Parameters are set as follows: $M = N = L = 20$, $K = 2$ and the true frequencies are $\tilde{\omega} = [-0.1; 0.5]^T$. From Table I, it can be calculated that γ are

0.1038, 0.1275, 0.1275 and 0.6025 for Case I-IV, respectively. Thus MVHN is more likely to overfit the measurements than the other three algorithms, as shown in the ensuing simulation. The nominal SNR is $\text{SNR} = 0$ dB and the strength of noise fluctuation is $\Delta_\nu = 20$ dB. 50 MC trials are performed and results are averaged over only $\hat{K} = K$ trials. The model order recovery probability $P(\hat{K} = K)$ and the reconstruction results are presented in Table II and Fig. 1, respectively. It can be seen that

Case I All the algorithms successfully estimate the model order in all the trials. From Fig. 1(a), all the algorithms estimates the frequencies well. In addition, compared to the other three algorithms, the posterior PDF output by the MVHN is the most peaked, which means that the uncertain degrees of the frequency estimates is very small, and the frequency estimates are biased.

Case II MVHN-S achieves the highest model order estimation probability, followed by MVHN, MVHN-A and MVALUE. According to Fig. 1(b), MVHN-S and MVHN estimate the frequencies well, while the frequencies estimation error of both MVHN-A and MVALUE are large. In addition, the posterior PDF output by the MVHN is much more peaked than MVHN-S, and the estimation bias of MVHN is larger than that of MVHN-S.

Case III MVHN-A has the highest model order estimation probability, followed by MVHN, MVHN-S and MVALUE. From Fig. 1(c), all the algorithms estimates the frequencies well. Besides, the posterior PDF output by the MVHN is the most peaked, followed by MVHN-A, MVALUE, MVHN-S. Similarly, the estimation bias of MVHN is the largest.

Case IV MVHN achieves the highest model order estimation probability, followed by MVHN-S, MVHN-A and MVALUE. According to Fig. 1(d), all algorithms estimate the frequencies well. In addition, the posterior PDF output by the MVHN is the most peaked, followed by MVHN-A, MVHN-S, MVALUE. As for the estimation bias, MVHN is the smallest among the four algorithms.

To sum up, for Case I, all algorithms work well. While for Case II-IV, the corresponding designed algorithms work best. For Case I or III, the estimation bias of MVHN is the largest. For all the cases, the posterior PDF output by the MVHN is the most peaked, implying the least uncertain degrees of frequency estimates. The reason may be that MVHN tend to overfit the measurements because the number of parameters that needs to be estimated is much larger than that of MVALUE.

B. Performance Versus the Nominal SNR

In this section, the performances of the four algorithms (MVHN, MVHN-S, MVHN-A, MVALUE) under Case I-IV with varied nominal SNR are investigated. Parameters are set as follows: $M = N = 20$, $L = 10$, $K = 3$ and the true frequencies are $\tilde{\omega} = [-0.1; 0.5; 2.1]^T$. The strength of noise fluctuation is

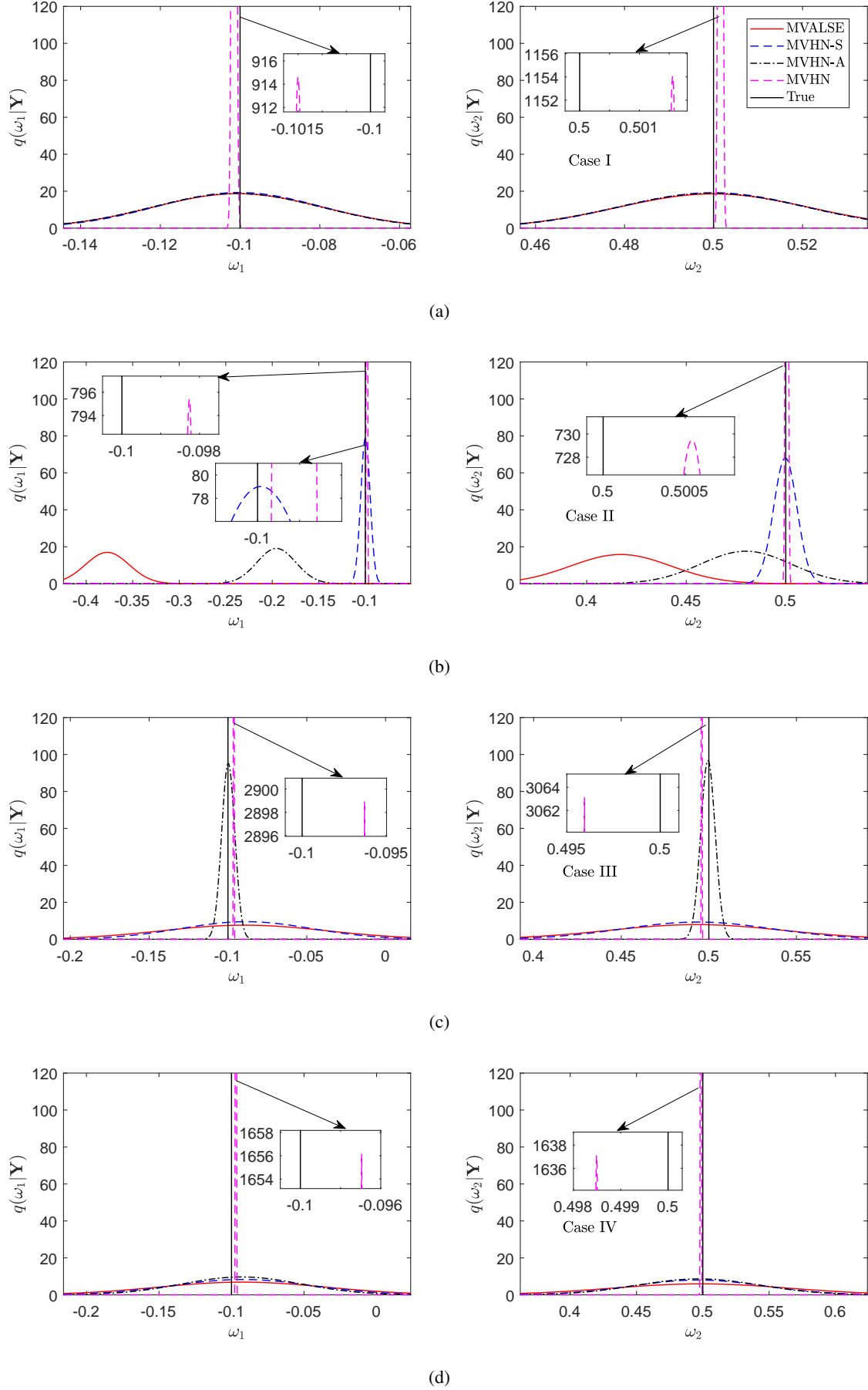


Fig. 1. The posterior PDF output of all four algorithms for Case I-IV.

TABLE II
THE RECOVERY PROBABILITY OF $\hat{K} = K$ OF THE ALGORITHMS FOR FOUR GENERATION CASES.

	MVALUE	MVHN-S	MVHN-A	MVHN
Case I	100%	100%	100%	100%
Case II	12%	72%	22%	58%
Case III	20%	22%	36%	34%
Case IV	14%	16%	12%	20%

$\Delta_\nu = 15$ dB for Case II-IV. While for Case I, the noise variance is fixed and thus $\Delta_\nu = 0$ dB. All the results are averaged over 300 MC trials.

The results are shown in Fig. 2. From Fig. 2(a), MVALUE, MVHN-S and MVHN-A achieve almost the same performance and asymptotically approach the CRB, while MVHN shows some performance degradation under Case I. The reason may be that the estimation bias of MVHN is very large, as discussed in subsection VI-A. For Case II, Fig. 2(b) shows that MVHN-S performs best and the frequency estimation error is close to CRB as SNR increases. As for model order estimation probability and frequency estimation error, MVHN outperforms MVALUE and MVHN-A while it performs worst in signal reconstruction. By carefully looking into the detailed reconstruction results, it is found that MVHN estimates $\hat{\mathbf{X}}$ worst. For Case III, Fig. 2(c) shows that MVHN-A performs best, followed by MVHN, MVHN-A and MVALUE. The frequency estimation error of MVHN-A is very close to the CRB. For Case IV in Fig. 2(d), MVHN performs best in terms of signal reconstruction error and frequency estimation error, while its recovery probability is smaller than MVHN-S and MVHN-A. In addition, there exists an obvious performance gap between the frequency estimation error and the CRB.

C. Performance Versus Δ_ν

The performances of the four algorithms under Case II-IV are investigated with varied strength of noise fluctuation Δ_ν . Parameters are set as follows: $M = N = 20$, $L = 5$, $K = 3$ and $\tilde{\omega} = [-0.1; 0.5; 2.1]^T$. The nominal SNR is $\text{SNR} = 5$ dB and 300 MC trials are performed. Results are shown in Fig. 3.

From Fig. 3(a) under Case II, as $\Delta_\nu \leq 4$ dB, both MVHN-S and MVALUE performs best, followed by MVHN-A and MVHN. As Δ_ν increases, the signal reconstruction error and model order probability of MVALUE and MVHN-A degrade more quickly than the other two algorithms, which demonstrate the robustness of MVHN. The frequency estimation error of MVHN-S is always close to the CRB. While for the other three algorithms, there exist obvious performance gaps compared to the CRB $\Delta_\nu \geq 4$

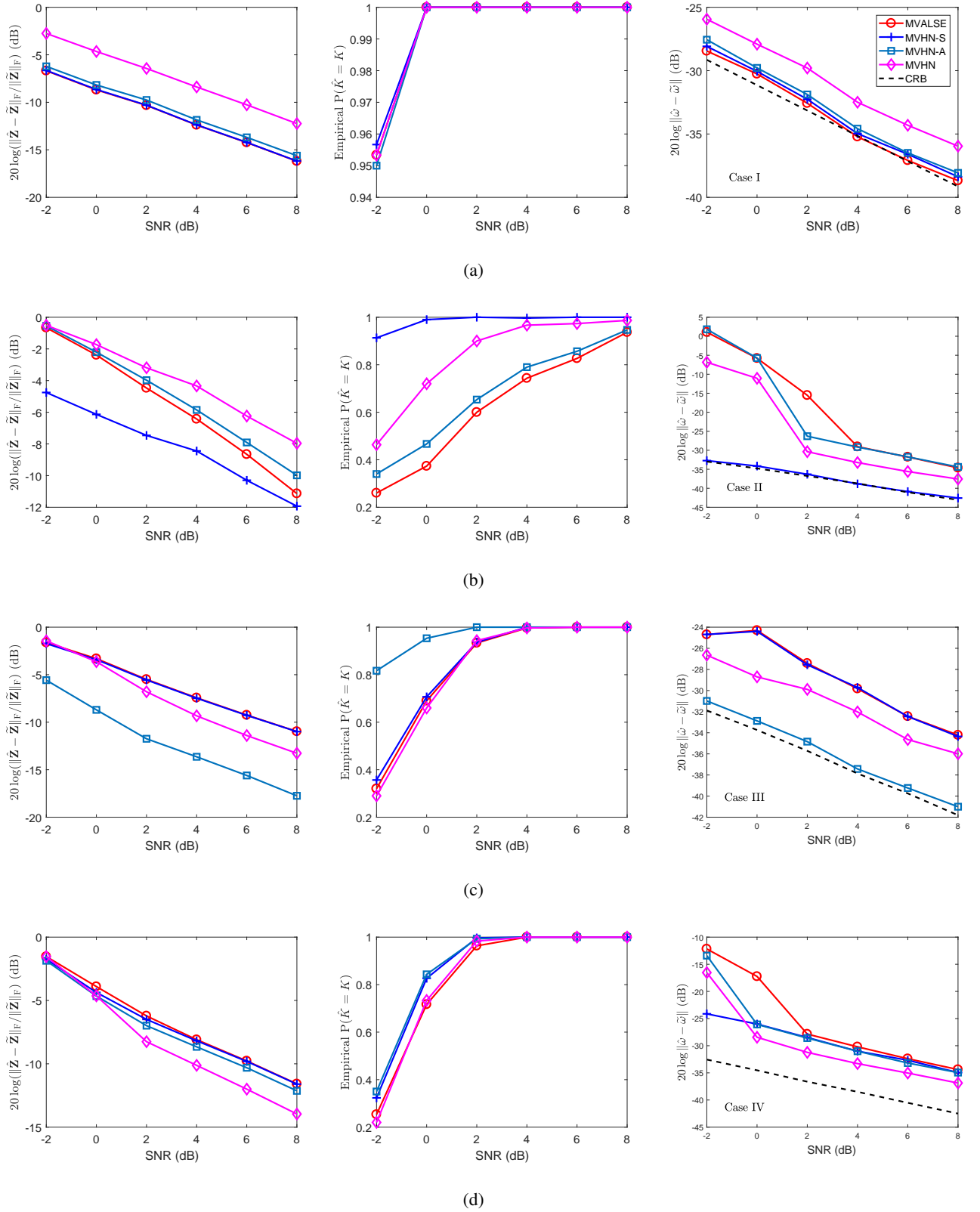


Fig. 2. Performance versus SNR for $M = 20$, $L = 10$. Fig. 2(a), 2(b), 2(c), 2(d) correspond to Noise Case I-IV, respectively.

dB. For Case III shown in Fig. 3(b), as $\Delta_\nu \leq 8$ dB, MVALUE and MVHN-S performs best. As Δ_ν increases, MVALUE and MVHN-S degrades quickly and perform worse than MVHN-A and MVHN. Similar to Case II, the robustness of MVHN is also demonstrated. Compared to CRB, gaps exist for all the algorithms. As for Case IV shown in Fig. 3(c), the phenomenon are basically the same as Fig. 3(b) for $\Delta_\nu \geq 16$ dB. As Δ_ν increases, the performance of MVHN is still stable, while the other three algorithms degrade quickly. To sum up, the algorithm matched to the noise Case always work well, and MVHN is robust for Case II-IV.

D. Performance from Incomplete Measurements

The performances of the four algorithms dealing with incomplete measurements are investigated. Parameters are set the same as Sec. VI-B except $\Delta_\nu = 10$ dB, and the number of incomplete measurements are $|\Omega| = 0.1ML = 20$, which is sampled uniformly at random. Fig. 4 shows the results of all the algorithms. It can be seen that all the four algorithms perform well. Besides, the performances of MVALUE, MVHN-S and MVHN-A are almost the same, and better than that of MVHN. The frequency estimation errors of MVALUE, MVHN-S and MVHN-A are close to the CRB.

E. Performance with Outliers

The robustness of MVHN and its ability to identify outliers are demonstrated. The noises are generated as Case I. Parameters are set the same as Sec. VI-B except $\Delta_\nu = 10$ dB. We set $\|\tilde{\mathbf{O}}\|_0 = 0.1ML = 20$. For the nonzero elements of $\tilde{\mathbf{O}}$, we drawn i.i.d. from $\mathcal{CN}(0, 50)$. According to the signal generation model, it can be calculated that the mean and variance of \tilde{Z}_{ij} is 0 and approximately 3 (equal to the number of spectral K), respectively, thus the value of outliers is much larger than the average value of \tilde{Z}_{ij} , and can be regarded as outliers. The performance of four algorithms are evaluated: MVALUE, MVHN, MVHN-ADI developed in subsection IV-B which adaptively identifies the outliers and use the remaining measurements to perform estimation, MVHN-AWARE which uses the measurements without outliers. The CRB corresponding to the measurements without outliers is also evaluated. Thus it is expected that MVHN-AWARE will perform best and approach the CRB asymptotically. For the implementation of MVHN-ADI, the measurements corresponding to top 15% estimated variances are recognized as outliers in each iteration, and MVALUE is applied using the remaining measurements. For the next iteration, MVHN is applied to estimate the variance corresponding to each measurement, and the outliers is reidentified.

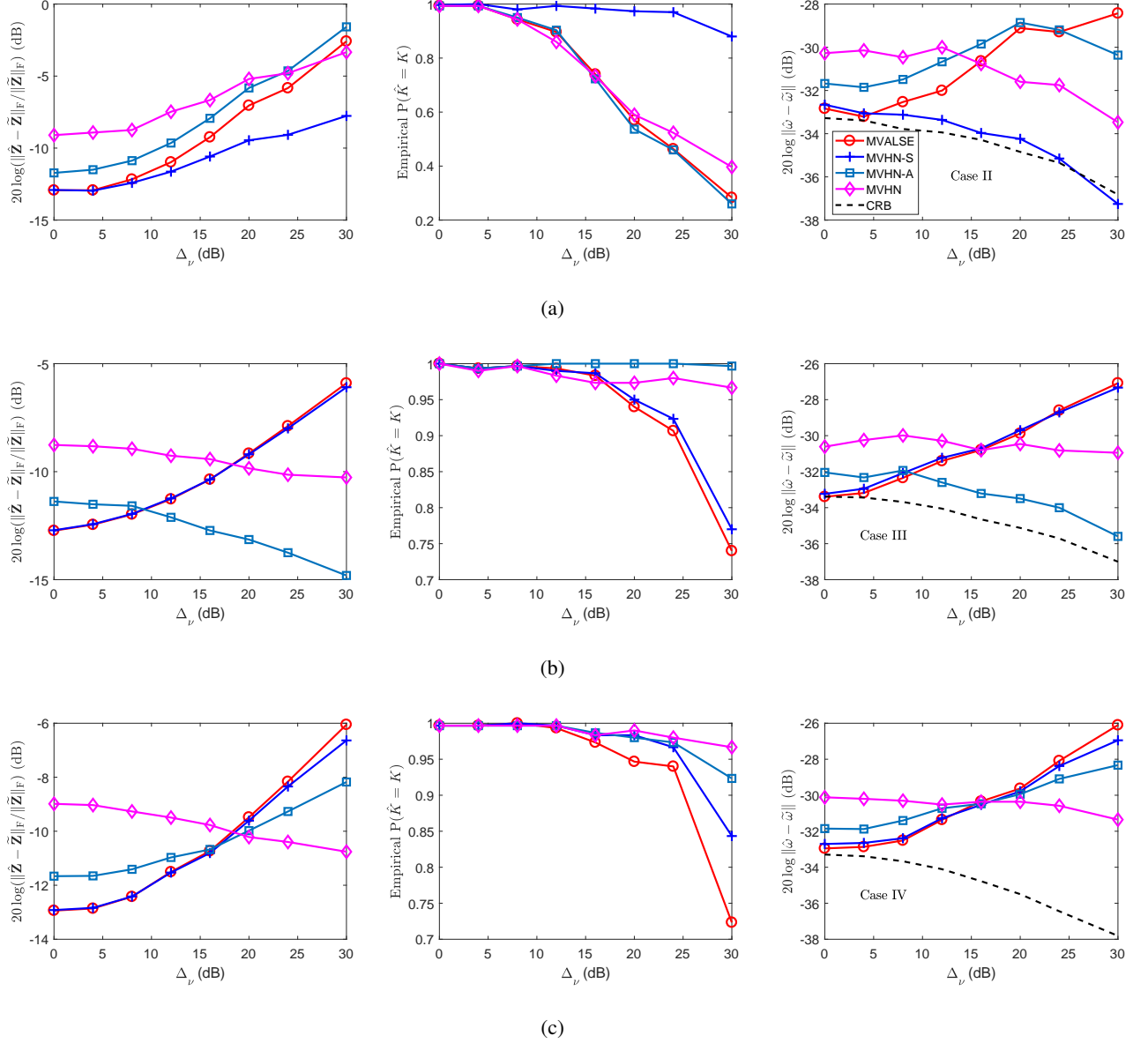


Fig. 3. Performance of algorithms by varying Δ_ν for Case II-IV corresponding to Fig. 3(a), 3(b), 3(c). The number of measurements is $M = 20$ and the number of snapshots is $L = 5$.

At first, we conduct a single experiment and give the noise variance estimate results of MVHN presented in Fig. 5. It can be seen that the position of outliers can be identified (detected) with very high accuracy, and only few outliers marked with red rectangle are missed.

The results are shown in Fig. 6. MVHN-AWARE and MVHN-ADI perform significantly better than MVHN and MVALUE. As $\text{SNR} \geq 4$ dB, MVHN performs better than MVALUE. For the signal reconstruction error, the performance of MVALUE does not improve as SNR increases. While for the

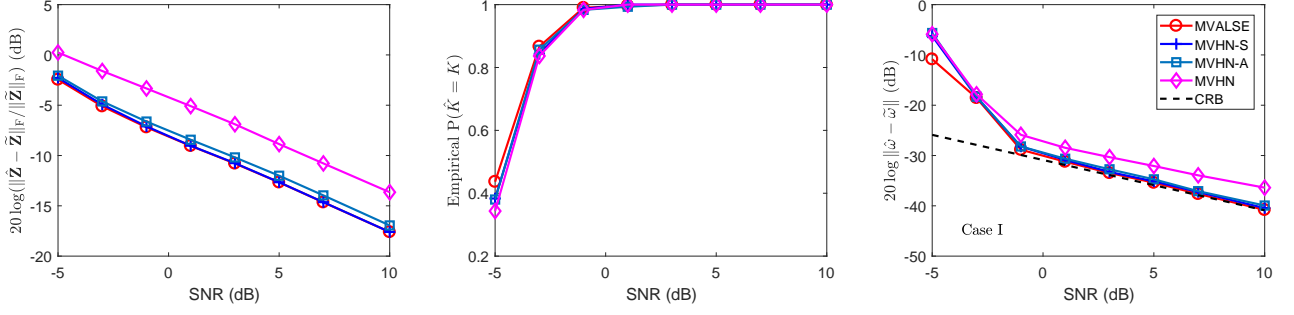


Fig. 4. Estimation performance of algorithms from incomplete measurements with varied nominal SNR. The noises are generated as Case I. The number of measurements is $M = 20$ and the number of snapshots is $L = 10$.

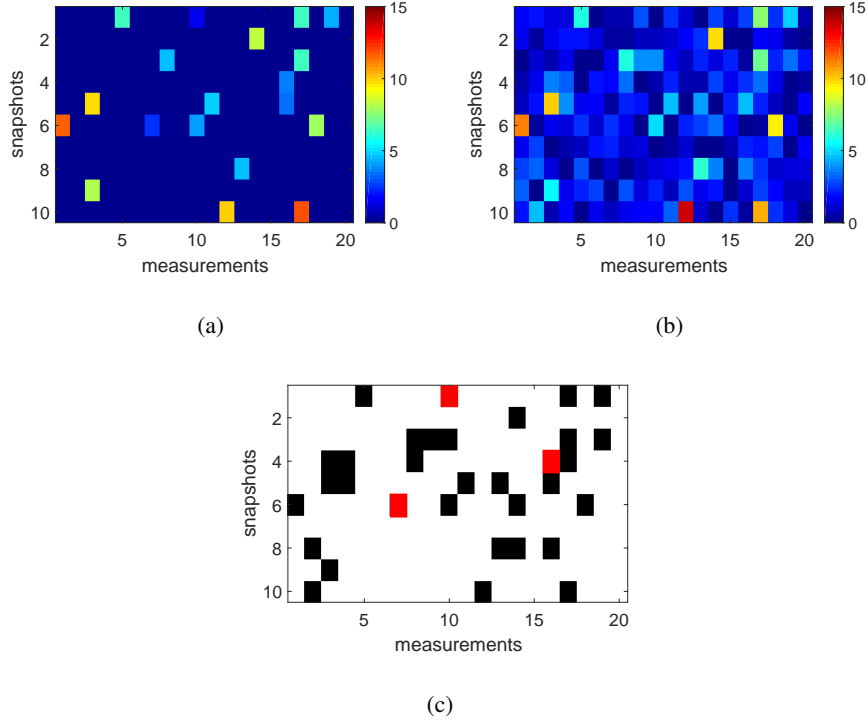


Fig. 5. The noise variance estimation results in a single realization. Fig. 5(a) shows the true value of outliers, and Fig. 5(b) gives the standard deviation $\sqrt{\hat{\nu}_{ml}}$ of the estimates output by MVHN. As for Fig. 5(c), we preserve the top 15% of $\sqrt{\hat{\nu}_{ml}}$ for better illustration.

other three algorithms, the signal reconstruction errors decrease linearly with respect to SNR. As SNR increases, the frequency estimation error of MVHN-AWARE approaches the CRB.

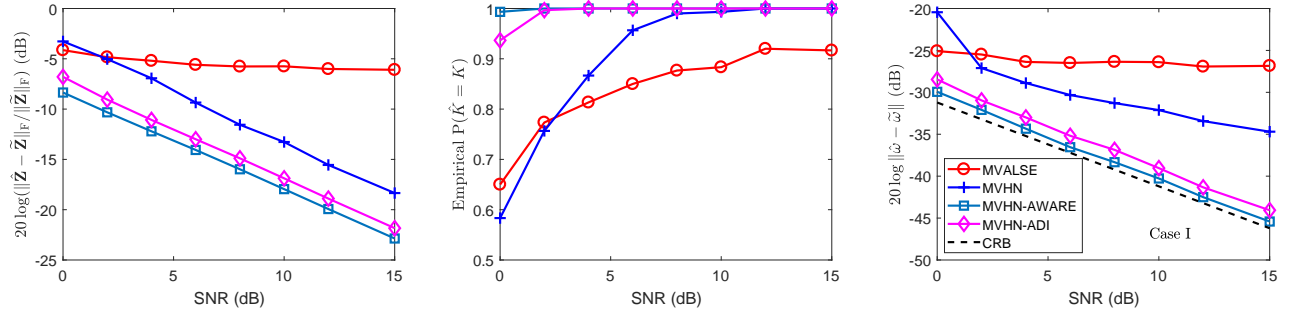


Fig. 6. Robustness of MVHN against outliers. The noise is generated as Case I. The number of measurements is $M = 20$ and the number of snapshots is $L = 10$.

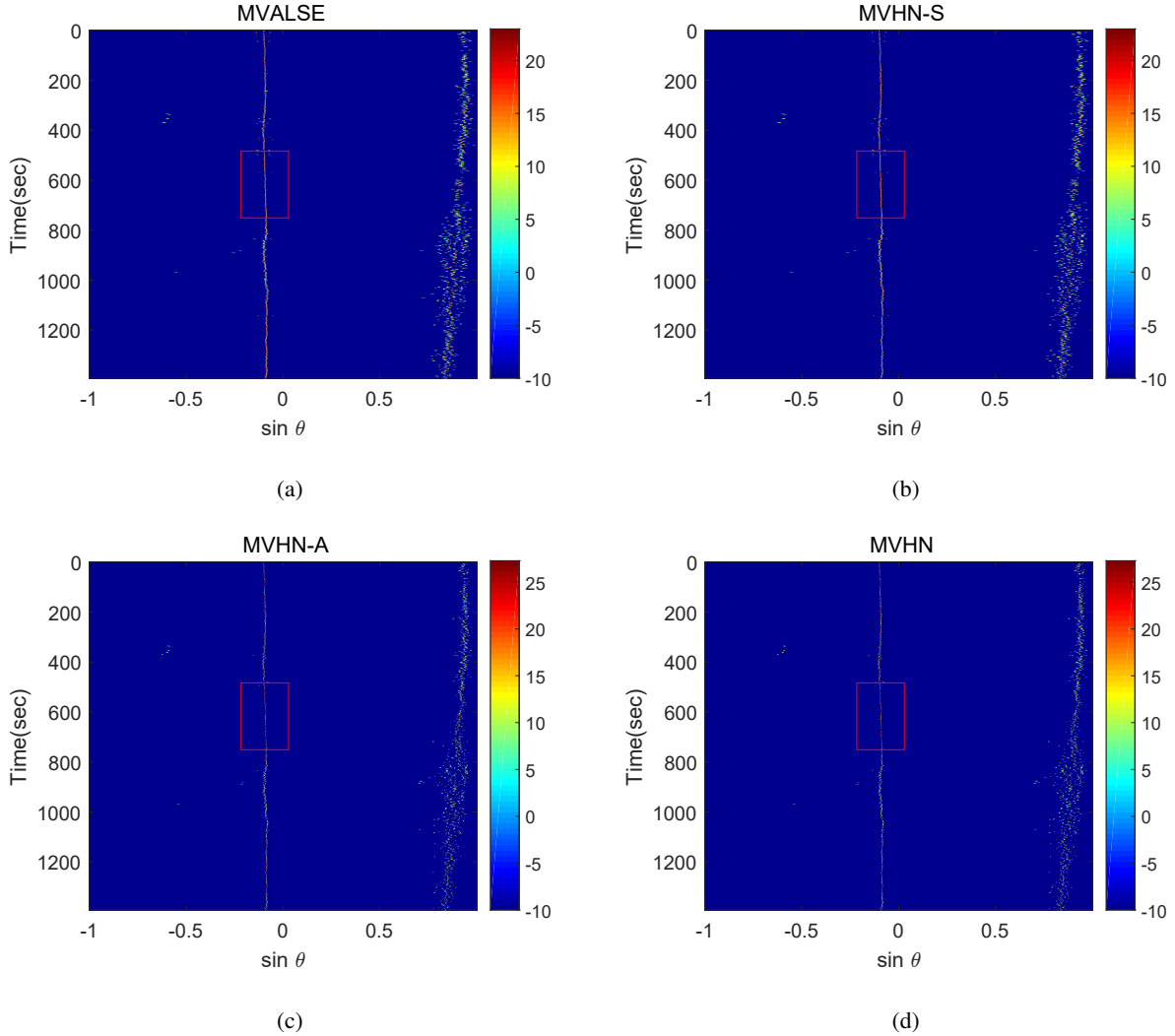


Fig. 7. The synthesized posterior PDF of $\sin \theta$ for real data.

VII. APPLICATION: DOA ESTIMATION

In this section, the performances of MVHN and its variants are evaluated with real experimental acoustic data collected at sea. The data were collected by the horizontal linear arrays (HLA) deployed on the seafloor [19]. The array had 32 elements uniformly spaced with an aperture of 15 m (design frequency 50 Hz). A low-frequency J-15-3 acoustic source towed by the R/V Endeavor is at around ~ 12 m depth and broadcasts continuous tones at 50 Hz. The ship (with the towed source) was moving outward from the HLA at a direction near broadside of the array. The 10 snapshots are arranged as a batch and the data generated from the first 16 elements of the array are chosen.

For each algorithm, the synthesized posterior PDF of frequencies $\sum_{i=1}^{|\hat{\mathcal{S}}|} q(\omega_i | \mathbf{Y}) / |\hat{\mathcal{S}}|$ are evaluated. For the DOA problem, $\omega_i = \pi \sin \theta_i$, where θ_i denotes the DOAs. Thus Fig. 7 displays the results with respect to $\sin \theta$. It can be seen that all the algorithms perform well. The towed source signal is at $\sin \theta \approx -0.1$ and there is a dynamic signal near the endfire direction which is assumed to be a moving ship. Compared to MVALSE and MVHN-S, the uncertain degrees output by MVHN-A and MVHN are more smaller, especially in marked red rectangles. In addition, the results produced by MVHN-A and MVHN are more smoother, which may imply that the noise is more likely to be varying across antennas.

VIII. CONCLUSION

In this paper, the line spectral estimation problem under heteroscedastic noise is studied. The MVHN and its variants are proposed and derived in a unified way. In addition, the proposed algorithms are also extended to deal with the incomplete measurements scenario and measurements with outliers. Finally, numerical experiments demonstrate the effectiveness of the proposed algorithms, including on a real data set.

IX. APPENDIX

A. Finding a Local Maximum of $\ln Z(\mathbf{s})$

Finding the globally optimal binary sequence \mathbf{s} of (33) is hard in general. As a result, a greedy iterative search strategy is adopted. We proceed as follows: In the p th iteration, we obtain the k th test sequence \mathbf{t}_k by flipping the k th element of $\mathbf{s}^{(p)}$. Then we calculate $\Delta_k^{(p)} = \ln Z(\mathbf{t}_k) - \ln Z(\mathbf{s}^{(p)})$ for each $k = 1, \dots, N$. If $\Delta_k^{(p)} < 0$ holds for all k we terminate the algorithm and set $\hat{\mathbf{s}} = \mathbf{s}^{(p)}$, else we choose the t_k corresponding to the maximum $\Delta_k^{(p)}$ as $\mathbf{s}^{(p+1)}$ in the next iteration.

When $k \notin \mathcal{S}$, that is, $s_k = 0$, we activate the k th component of \mathbf{s} by setting $s'_k = 1$. Now, $\mathcal{S}' = \mathcal{S} \cup \{k\}$.

$$\begin{aligned} \Delta_k &= \ln Z(\mathbf{s}') - \ln Z(\mathbf{s}) \\ &= \sum_{l=1}^L \left(\ln \det([\mathbf{J}_l]_{\mathcal{S}} + \frac{1}{\tau} \mathbf{I}_{|\mathcal{S}|}) - \ln \det([\mathbf{J}_l]_{\mathcal{S}'} + \frac{1}{\tau} \mathbf{I}_{|\mathcal{S}'|}) \right) \\ &\quad + \sum_{l=1}^L \left(\mathbf{H}_{\mathcal{S}',l}^H([\mathbf{J}_l]_{\mathcal{S}'} + \frac{1}{\tau} \mathbf{I}_{|\mathcal{S}'|})^{-1} \mathbf{H}_{\mathcal{S}',l} - \mathbf{H}_{\mathcal{S},l}^H([\mathbf{J}_l]_{\mathcal{S}} + \frac{1}{\tau} \mathbf{I}_{|\mathcal{S}|})^{-1} \mathbf{H}_{\mathcal{S},l} \right) \\ &\quad + L \ln \frac{1}{\tau} + \ln \frac{\rho}{1-\rho}. \end{aligned} \quad (48)$$

By using the block-matrix determinant formula, one has

$$\ln \det([\mathbf{J}_l]_{\mathcal{S}'} + \frac{1}{\tau} \mathbf{I}_{|\mathcal{S}'|}) = \ln \det([\mathbf{J}_l]_{\mathcal{S}} + \frac{1}{\tau} \mathbf{I}_{|\mathcal{S}|}) + \ln \left(\text{tr}(\mathbf{\Sigma}_l^{-1}) + \frac{1}{\tau} - [\mathbf{J}_l]_{\mathcal{S},k}^H([\mathbf{J}_l]_{\mathcal{S}} + \frac{1}{\tau} \mathbf{I}_{|\mathcal{S}|})^{-1} [\mathbf{J}_l]_{\mathcal{S},k} \right). \quad (49)$$

By the block-wise matrix inversion formula, one has

$$\mathbf{H}_{\mathcal{S}',l}^H([\mathbf{J}_l]_{\mathcal{S}'} + \frac{1}{\tau} \mathbf{I}_{|\mathcal{S}'|})^{-1} \mathbf{H}_{\mathcal{S}',l} = \mathbf{H}_{\mathcal{S},l}^H([\mathbf{J}_l]_{\mathcal{S}} + \frac{1}{\tau} \mathbf{I}_{|\mathcal{S}|})^{-1} \mathbf{H}_{\mathcal{S},l} + \frac{|u_{k,l}|^2}{v_{k,l}}, \quad (50)$$

where

$$\begin{aligned} v_{k,l} &= \left(\text{tr}(\mathbf{\Sigma}_l^{-1}) + \frac{1}{\tau} - [\mathbf{J}_l]_{\mathcal{S},k}^H([\mathbf{J}_l]_{\mathcal{S}} + \frac{1}{\tau} \mathbf{I}_{|\mathcal{S}|})^{-1} [\mathbf{J}_l]_{\mathcal{S},k} \right)^{-1}, \\ u_{k,l} &= v_{k,l} \left(h_{k,l} - [\mathbf{J}_l]_{\mathcal{S},k}^H([\mathbf{J}_l]_{\mathcal{S}} + \frac{1}{\tau} \mathbf{I}_{|\mathcal{S}|})^{-1} \mathbf{H}_{\mathcal{S},l} \right). \end{aligned} \quad (51)$$

Inserting (49) and (50) into (48), Δ_k can be simplified as

$$\Delta_k = \sum_{l=1}^L \left(\ln \frac{v_{k,l}}{\tau} + \frac{|u_{k,l}|^2}{v_{k,l}} \right) + \ln \frac{\rho}{1-\rho}. \quad (52)$$

Given that \mathbf{s} is changed into \mathbf{s}' , the mean $\hat{\mathbf{X}}'_{\mathcal{S}',l}$ and covariance $[\hat{\mathbf{C}}'_l]_{\mathcal{S}'}$ of the l th snapshot can be updated from (32), i.e.,

$$[\hat{\mathbf{C}}'_l]_{\mathcal{S}'} = ([\mathbf{J}_l]_{\mathcal{S}'} + \frac{1}{\tau} \mathbf{I}_{|\mathcal{S}'|})^{-1}, \quad (53a)$$

$$\hat{\mathbf{X}}'_{\mathcal{S}',l} = [\hat{\mathbf{C}}'_l]_{\mathcal{S}'} \mathbf{H}_{\mathcal{S}',l}. \quad (53b)$$

In fact, the matrix inversion can be avoided when updating $\hat{\mathbf{X}}'_{\mathcal{S}',l}$ and $[\hat{\mathbf{C}}'_l]_{\mathcal{S}'}$. It can be shown that

$$\begin{aligned} \begin{bmatrix} [\hat{\mathbf{C}}'_l]_{\mathcal{S}} & [\hat{\mathbf{C}}'_l]_{\mathcal{S},k} \\ [\hat{\mathbf{C}}'_l]_{k,\mathcal{S}} & [\hat{\mathbf{C}}'_l]_{k,k} \end{bmatrix} &= \begin{bmatrix} [\mathbf{J}_l]_{\mathcal{S}} + \frac{1}{\tau} \mathbf{I}_{|\mathcal{S}|} & [\mathbf{J}_l]_{\mathcal{S},k} \\ [\mathbf{J}_l]_{k,\mathcal{S}} & \text{tr}(\mathbf{\Sigma}_l^{-1}) + \frac{1}{\tau} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} [\hat{\mathbf{C}}_l]_{\mathcal{S}}^{-1} & [\mathbf{J}_l]_{\mathcal{S},k} \\ [\mathbf{J}_l]_{k,\mathcal{S}} & \text{tr}(\mathbf{\Sigma}_l^{-1}) + \frac{1}{\tau} \end{bmatrix}^{-1} = \begin{bmatrix} [\hat{\mathbf{C}}_l]_{\mathcal{S}} + v_{k,l} [\hat{\mathbf{C}}_l]_{\mathcal{S}} [\mathbf{J}_l]_{\mathcal{S},k} [\mathbf{J}_l]_{k,\mathcal{S}} [\hat{\mathbf{C}}_l]_{\mathcal{S}} & -v_{k,l} [\hat{\mathbf{C}}_l]_{\mathcal{S}} [\mathbf{J}_l]_{\mathcal{S},k} \\ -v_{k,l} [\mathbf{J}_l]_{k,\mathcal{S}} [\hat{\mathbf{C}}_l]_{\mathcal{S}} & v_{k,l} \end{bmatrix}. \end{aligned} \quad (54)$$

Furthermore, the weight $\hat{\mathbf{X}}'_{S',l}$ is updated as

$$\begin{bmatrix} \hat{\mathbf{X}}'_{S',l} \\ \hat{x}'_{k,l} \end{bmatrix} = \begin{bmatrix} [\hat{\mathbf{C}}'_l]_S & [\hat{\mathbf{C}}'_l]_{S,k} \\ [\hat{\mathbf{C}}'_l]_{k,S} & [\hat{\mathbf{C}}'_l]_{k,k} \end{bmatrix} \begin{bmatrix} \mathbf{H}_{S,l} \\ h_{k,l} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{X}}_{S,l} - [\hat{\mathbf{C}}_l]_S [\mathbf{J}_l]_{S,k} u_{k,l} \\ u_{k,l} \end{bmatrix}.$$

For the deactive case with $\mathbf{s}_k = 1$, $\mathbf{s}'_k = 0$ and $S' = S \setminus \{k\}$, $\Delta_k = \ln Z(\mathbf{s}') - \ln Z(\mathbf{s})$ is the negative of (52), i.e.,

$$\Delta_k = - \sum_{l=1}^L \left(\ln \frac{v_{k,l}}{\tau} + \frac{|u_{k,l}|^2}{v_{k,l}} \right) - \ln \frac{\rho}{1-\rho}. \quad (55)$$

Similar to (54), the posterior mean and covariance update equation from S' to S case of l th snapshot can be rewritten as

$$\begin{bmatrix} [\hat{\mathbf{C}}'_l]_{S'} + v_{k,l} [\hat{\mathbf{C}}'_l]_{S'} [\mathbf{J}_l]_{S',k} [\mathbf{J}_l]_{k,S'} [\hat{\mathbf{C}}'_l]_{S'} & -v_{k,l} [\hat{\mathbf{C}}'_l]_{S'} [\mathbf{J}_l]_{S',k} \\ -v_{k,l} [\mathbf{J}_l]_{k,S'} [\hat{\mathbf{C}}'_l]_{S'} & v_{k,l} \end{bmatrix} = \begin{bmatrix} [\hat{\mathbf{C}}_l]_{S'} & [\hat{\mathbf{C}}_l]_{S',k} \\ [\hat{\mathbf{C}}_l]_{k,S'} & [\hat{\mathbf{C}}_l]_{k,k} \end{bmatrix} \quad (56)$$

$$\begin{bmatrix} \hat{\mathbf{X}}'_{S',l} - [\hat{\mathbf{C}}'_l]_{S'} [\mathbf{J}_l]_{S',k} u_{k,l} \\ u_{k,l} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{X}}_{S',l} \\ \hat{x}_{k,l} \end{bmatrix}, \quad (57)$$

According to (56) and (57), one has

$$[\hat{\mathbf{C}}'_l]_{S'} + v_{k,l} [\hat{\mathbf{C}}'_l]_{S'} [\mathbf{J}_l]_{S',k} [\mathbf{J}_l]_{k,S'} [\hat{\mathbf{C}}'_l]_{S'} = [\hat{\mathbf{C}}_l]_{S'}, \quad (58a)$$

$$-v_{k,l} [\hat{\mathbf{C}}'_l]_{S'} [\mathbf{J}_l]_{S',k} = [\hat{\mathbf{C}}_l]_{S',k} \quad (58b)$$

$$v_{k,l} = [\hat{\mathbf{C}}_l]_{k,k}, \quad (58c)$$

$$\hat{\mathbf{X}}'_{S',l} - [\hat{\mathbf{C}}'_l]_{S'} [\mathbf{J}_l]_{S',k} u_{k,l} = \hat{\mathbf{X}}_{S',l}, \quad (58d)$$

$$u_{k,l} = \hat{x}_{k,l}. \quad (58e)$$

Thus, $\hat{\mathbf{C}}'_{S',l}$ can be updated by substituting (58b) and (58c) in (58a), i.e.,

$$[\hat{\mathbf{C}}'_l]_{S'} = [\hat{\mathbf{C}}_l]_{S'} - \frac{[\hat{\mathbf{C}}_l]_{S',k} [\hat{\mathbf{C}}_l]_{k,S'}}{[\hat{\mathbf{C}}_l]_{k,k}}. \quad (59)$$

Similarly, $\hat{\mathbf{X}}'_{S',l}$ can be updated by substituting (58b) and (58e) in (58d), i.e.,

$$\hat{\mathbf{X}}'_{S',l} = [\hat{\mathbf{C}}'_l]_{S'} [\mathbf{J}_l]_{S',k} u_{k,l} + \hat{\mathbf{X}}_{S',l} = \hat{\mathbf{X}}_{S',l} - \frac{[\hat{\mathbf{C}}_l]_{S',k}}{[\hat{\mathbf{C}}_l]_{k,k}} \hat{x}_{k,l}. \quad (60)$$

According to $v_{k,l} = [\hat{\mathbf{C}}_l]_{k,k}$ (58c) and $u_{k,l} = \hat{x}_{k,l}$ (58e), Δ_k (55) can be simplified as

$$\Delta_k = - \sum_{l=1}^L \left(\ln \frac{[\hat{\mathbf{C}}_l]_{k,k}}{\tau} + \frac{|\hat{x}_{k,l}|^2}{[\hat{\mathbf{C}}_l]_{k,k}} \right) - \ln \frac{\rho}{1-\rho}. \quad (61)$$

B. The Derivation of CRB

Here we calculate the CRB for the general Case IV, then we specialize it to the other three Cases ³. Let $g_{k,l}$ and $\phi_{k,l}$ be the amplitude and phase of $x_{k,l}$, $\forall k = 1, \dots, K, l = 1, \dots, L$, i.e., $g_{k,l} = |x_{k,l}|$ and $\phi_{k,l} = \angle x_{k,l}$. Then we obtain matrices \mathbf{G} and Φ . Let $\boldsymbol{\kappa}$ be $\boldsymbol{\kappa} = [\boldsymbol{\omega}^T, \mathbf{g}^T, \boldsymbol{\phi}^T]^T$, where $\mathbf{g} = \text{vec}(\mathbf{G})$ and $\boldsymbol{\phi} = \text{vec}(\Phi)$. Then the FIM is calculated according to [20, 21]

$$\mathbf{I}(\boldsymbol{\kappa}) = \sum_{m=1}^M \sum_{l=1}^L \frac{2}{\nu_{m,l}} \left(\frac{\partial \Re\{Z_{m,l}\}}{\partial \boldsymbol{\kappa}} \left(\frac{\partial \Re\{Z_{m,l}\}}{\partial \boldsymbol{\kappa}} \right)^T + \frac{\partial \Im\{Z_{m,l}\}}{\partial \boldsymbol{\kappa}} \left(\frac{\partial \Im\{Z_{m,l}\}}{\partial \boldsymbol{\kappa}} \right)^T \right). \quad (62)$$

By defining $\mathbf{g}_l = [g_{1,l}, \dots, g_{K,l}]^T$ and $\boldsymbol{\phi}_l = [\phi_{1,l}, \dots, \phi_{K,l}]^T$, we have

$$\frac{\partial \Re\{Z_{m,l}\}}{\partial \boldsymbol{\kappa}} = \begin{bmatrix} \frac{\partial \Re\{Z_{m,l}\}}{\partial \boldsymbol{\omega}} \\ \mathbf{0}_{(l-1)K} \\ \frac{\partial \Re\{Z_{m,l}\}}{\partial \mathbf{g}_l} \\ \mathbf{0}_{(L-l)K} \\ \mathbf{0}_{(l-1)K} \\ \frac{\partial \Re\{Z_{m,l}\}}{\partial \boldsymbol{\phi}_l} \\ \mathbf{0}_{(L-l)K} \end{bmatrix}, \quad \frac{\partial \Im\{Z_{m,l}\}}{\partial \boldsymbol{\kappa}} = \begin{bmatrix} \frac{\partial \Im\{Z_{m,l}\}}{\partial \boldsymbol{\omega}} \\ \mathbf{0}_{(l-1)K} \\ \frac{\partial \Im\{Z_{m,l}\}}{\partial \mathbf{g}_l} \\ \mathbf{0}_{(L-l)K} \\ \mathbf{0}_{(l-1)K} \\ \frac{\partial \Im\{Z_{m,l}\}}{\partial \boldsymbol{\phi}_l} \\ \mathbf{0}_{(L-l)K} \end{bmatrix}, \quad (63)$$

where

$$\begin{aligned} \frac{\partial \Re\{Z_{m,l}\}}{\partial \omega_k} &= -(m-1)g_{k,l} \sin[(m-1)\omega_k + \phi_{k,l}], \\ \frac{\partial \Re\{Z_{m,l}\}}{\partial g_{k,l}} &= \cos[(m-1)\omega_k + \phi_{k,l}], \\ \frac{\partial \Re\{Z_{m,l}\}}{\partial \phi_{k,l}} &= -g_{k,l} \sin[(m-1)\omega_k + \phi_{k,l}], \\ \frac{\partial \Im\{Z_{m,l}\}}{\partial \omega_k} &= (m-1)g_{k,l} \cos[(m-1)\omega_k + \phi_{k,l}], \\ \frac{\partial \Im\{Z_{m,l}\}}{\partial g_{k,l}} &= \sin[(m-1)\omega_k + \phi_{k,l}], \\ \frac{\partial \Im\{Z_{m,l}\}}{\partial \phi_{k,l}} &= g_{k,l} \cos[(m-1)\omega_k + \phi_{k,l}]. \end{aligned}$$

Substituting (63) in (62), the FIM $\mathbf{I}(\boldsymbol{\kappa})$ is obtained. The CRB is $\text{CRB}(\boldsymbol{\kappa}) = \mathbf{I}^{-1}(\boldsymbol{\kappa})$ and CRB of frequencies are $[\text{CRB}(\boldsymbol{\kappa})]_{1:K, 1:K}$, which will be used as the performance metrics.

Substituting $\nu_{m,l} = \nu$, $\forall m, l$, $\nu_{m,l} = \nu_m$, $\forall l$, $\nu_{m,l} = \nu_l$, $\forall m$ in (62), we obtain the FIM for Case I, II, III, respectively. Taking the inverse of the FIM yields the CRB.

³The CRB under incomplete scenario is straightforward and is omitted here for simplicity. The numerical experiments we have conducted indeed evaluate the CRB from incomplete measurement, see subsection VI-D and subsection VI-E.

X. ACKNOWLEDGEMENT

The authors would like to thank T. C. Yang for valuable discussions and suggestions on this work.

REFERENCES

- [1] H. L. Van Trees, "Optimum array processing," New York, NY, USA: Wiley, 2002.
- [2] P. Gerstoft, S. Nannuru, C. F. Mecklenbräuker and G. Leus, "DOA estimation in heteroscedastic noise," *Signal Process.*, vol. 161, pp. 63-73, 2019.
- [3] H. Amiri, H. Amindavar, R. Kirlin, "Array signal processing using garch noise modeling," in *IEEE Int. Conf. on Acoust., Speech, and Signal Process.*, vol. 2, 2004.
- [4] H. Amiri, H. Amindavar, M. Kamarei, "Underwater noise modeling and direction-finding based on heteroscedastic time series," *EURASIP J. Adv. Signal Process.*, 2007.
- [5] U. Nickel, "On the influence of channel errors on array signal processing methods," *Int. J. Electron. Commun.*, vol. 47, no. 4, pp. 209-219, 1993.
- [6] B. Baggeroer and H. Cox, "Passive sonar limits upon nulling multiple moving ships with large aperture arrays," in *Proc. IEEE 33rd Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Oct. 1999, pp. 103-108.
- [7] N. Riahi, P. Gerstoft, "The seismic traffic footprint: tracking trains, aircraft, and cars seismically," *Geophys. Res. Lett.*, vol. 42, no. 8, pp. 2674-2681, 2015.
- [8] H. Sinath and V. U. Reddy, "Analysis of MUSIC algorithm with sensor gain and phase perturbations," *Signal Process.*, vol. 23, no. 3, pp. 245-256, 1991.
- [9] L. Fu and R. Vaccaro, "Performance degradation of DOA estimators due to unknown noise fields," *IEEE Trans. Signal Process.*, vol. 40, no. 3, pp. 686-690, Mar. 1992.
- [10] K. L. Gemba, S. Nannuru and P. Gerstoft, "Robust ocean acoustic localization with sparse Bayesian learning," *IEEE J. Sel. Topics. Signal Process.*, vol. 13, no. 1, Mar. 2019.
- [11] M. Pesavento and A. M. Gershman, "Maximum-likelihood direction-of-arrival estimation in the presence of unknown nonuniform noise," *IEEE Trans. Signal Process.*, vol. 49, no. 7, pp. 1310-1324, Jul. 2001.
- [12] C. E. Chan, F. Lorenzelli, R. E. Hudson and K. Yao, "Stochastic maximum-likelihood DOA estimation in the presence of unknown uniform noise," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3038-3044, Jul. 2008.
- [13] Y. Wu, C. Hou, G. liao and Q. Guo, "Direction-of-arrival estimation in the presence of unknown nonuniform noise fields," *IEEE J. Oceanic Eng.*, vol. 31, no. 2, pp. 504-510, 2006.

- [14] Y. Chi, L. L. Scharf, A. Pezeshki and R. Calderbank, “Sensitivity of basis mismatch to compressed sensing,” *IEEE Trans. on Signal Process.*, vol. 59, pp. 2182 - 2195, 2011.
- [15] M. A. Badiu, T. L. Hansen and B. H. Fleury, “Variational Bayesian inference of line spectral,” *IEEE Trans. Signal Process.*, vol. 65, no. 9, pp. 2247-2261, 2017.
- [16] J. Zhu, Q. Zhang, P. Gerstoft, M. A. Badiu and Z. Xu, “Grid-less variational Bayesian line spectral estimation with multiple measurement vectors,” *Signal Process.*, vol. 61, pp. 155-164, 2019.
- [17] K. V. Mardia and P. E. Jupp, *Directional Statistics*. New York, NY, USA: Wiley, 2000.
- [18] K. P. Murphy, *Machine Learning A Probabilistic Perspective*. MIT Press, 2012.
- [19] T. C. Yang, “Deconvolved conventional beamforming for a horizontal line array,” *IEEE J. Oceanic Eng.*, vol. 43, no. 1, pp. 160-172, 2018.
- [20] H. Fu and Y. Chi, “Quantized spectral compressed sensing: Cramér-Rao bounds and recovery algorithms,” *IEEE Trans. on Signal Process.*, vol. 66, no. 12, pp. 3268-3279, 2018.
- [21] J. Zhu, L. Han, R. S. Blum and Z. Xu, “Multi-snapshot Newtonalized orthogonal matching pursuit for line spectrum estimation with multiple measurement vectors,” *Signal Process.*, vol. 165, pp. 175-185, 2019.