ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-6654-0540-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICASSP43922.2022.9746180

LEARNING-AIDED INITIALIZATION FOR VARIATIONAL BAYESIAN DOA ESTIMATION

Yongsung Park, Florian Meyer, and Peter Gerstoft

University of California San Diego, La Jolla, CA, USA

ABSTRACT

We present a sparsity-promoting method for the detection and estimation of the directions of arrival (DOAs) of source signals. The proposed method is based on the recently introduced variational Bayesian line spectral estimation (VALSE) approach, which is gridless. However, the performance of VALSE is sensitive to an initial guess of the measurement noise variance and potential DOAs. Thus, we propose a sparse Bayesian learning-aided initialization. Simulation results show that this learning-aided VALSE outperforms state-of-the-art DOA estimation methods as well as the conventional VALSE. We also evaluate the proposed method using acoustic data from an ocean acoustics experiment.

Index Terms— Direction-of-arrival (DOA) estimation, sparse signal recovery, variational Bayesian estimation.

1. INTRODUCTION

Direction-of-arrival (DOA) estimation is finding the direction of sources arriving at an array of sensors. Sparse signal processing, e.g., compressive sensing (CS), provides improved DOA estimation performance over traditional DOA methods [1–3]. DOAs exist on a continuous angular domain, and DOAs off the angular search grid points deteriorate the performance. We propose a Bayesian method, which makes it possible to perform sparsity-promoting gridless DOA estimation.

Promoting sparse solutions has shown its superiority over traditional DOA estimators: a high-resolution capability, dealing with single snapshot data [4, 5], and performing well with coherent arrivals [2, 6, 7]. Most DOA estimators solve non-linear estimation of DOAs on a continuous angular domain with a discretized angular search grid of potential DOAs ("steering vectors"). The mismatch among true DOAs and the grid points degrades the DOA estimates, which is called "basis mismatch" [8]. To overcome the basis mismatch, gridless CS are utilized for DOA estimation [9–13].

Sparse Bayesian learning (SBL) [14], a Bayesian estimation framework for sparse signal processing, has been introduced for DOA estimation [15–19]. Using statistical information for DOA estimation works well with sparsity-enforcing prior distributions. To address the off-grid DOAs, SBL-based methods have been proposed [20–24]. These are grid-based methods that use a grid but deal with DOAs off the grid.

A variational Bayesian method, variational line spectral estimation (VALSE) algorithm [25, 26], has been introduced as a sparse gridless method, and it promotes sparsity using a Bernoulli-Gaussian prior model. We interpret DOA estimation in the probabilistic formulation [27, 28] and apply VALSE to DOA estimation [29, 30]. The VALSE is an iterative algorithm, and accurate initialization improves the performance and leads convergence to a local optimum. The original VALSE builds a Toeplitz structured estimate of the sample covariance matrix and obtains an initial guess for potential DOAs and noise variance (the average of the lower quarter of the eigenvalues of the matrix). This initial guess results in poor convergence with incorrect DOAs / the noise variance even in high signal-to-noise ratio (SNR) regimes.

We suggest the grid-based SBL for the initial guess for potential DOAs and noise variance. The SBL-aided variational Bayesian method provides gridless sparse DOA estimation. In addition, many DOA estimators assume that the number of DOAs ("model order") is known or embed additional model order selection methods [31]. The presented method has sparsity-controlling Bernoulli variables that provide the number of DOAs. Numerical results demonstrate that our method improves performance compared to state-of-the-art techniques and the original VALSE. We apply the presented method to real data from an ocean acoustics experiment.

2. SIGNAL MODEL

We consider K narrowband sources with complex signal amplitude $w_k \in \mathbb{C}$ with DOAs $\theta_k \in \Theta \triangleq [-90°, 90°)$, $k = 1, \dots, K$. The sources arrive at a linear array with M sensors from the far-field. The observed data $\mathbf{y} \in \mathbb{C}^M$ is modeled as

$$\mathbf{y} = \sum_{k=1}^{K} w_k \mathbf{a}(\theta_k) + \mathbf{e}, \tag{1}$$

where $\mathbf{e} \in \mathbb{C}^M$ is the additive noise, and $\mathbf{a}(\theta_k) \in \mathbb{C}^M$ is the steering vector. The steering vector is given by (λ is the signal wavelength and d_m is the distance from sensor 1 to sensor m), $\mathbf{a}(\theta_k) = \left[1 \; e^{-j\frac{2\pi}{\lambda}d_2 \sin\theta_k} \; \dots \; e^{-j\frac{2\pi}{\lambda}d_M \sin\theta_k}\right]^{\mathrm{T}}$.

This research was supported by the Office of Naval Research under Grant N00014-21-1-2267.

Authorized licensed use limited to: Univ of Calif San Diego. Downloaded on February 21,2023 at 21:18:58 UTC from IEEE Xplore. Restrictions apply.

3. VARIATIONAL BAYESIAN DOA ESTIMATION

Let us consider the signal model (1) as a random process,

$$\mathsf{y} = \sum_{l=1}^{L} \mathsf{w}_l \mathbf{a}(\Theta_l) + \mathbf{e}. \tag{2}$$

Random variables are displayed in sans serif, upright fonts; their realizations in Latin modern, italic fonts, e.g., y is a random vector, and $\mathbf{y}$ is its realization. L is the number of potential DOAs and we consider $L = M$. We introduce independent Bernoulli variables $\mathsf{s} = (\mathsf{s}_1, \ldots, \mathsf{s}_L)$ for w (i.e., $s_l = 0$ implies $w_l = 0$). DOA estimation is represented by the minimum mean square error (MMSE) estimation of $\mathbf{\Phi} = \{\boldsymbol{\theta}, \mathbf{w}, \mathbf{s}\}$,

$$\left\{\hat{\boldsymbol{\theta}}, \hat{\mathbf{w}}, \hat{\mathbf{s}}\right\} = \hat{\mathbf{\Phi}} \triangleq \int \mathbf{\Phi} f(\mathbf{\Phi}|\mathbf{y})\, \mathrm{d}\mathbf{\Phi}, \tag{3}$$

where, using Bayes' rule and the chain rule for pdfs, $f(\boldsymbol{\theta}, \mathbf{w}, \mathbf{s}|\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\theta}, \mathbf{w}) \prod_{l=1}^{L} f(\theta_l) f(\mathbf{w}_l|s_l)\, p(s_l)$. Unfortunately, we cannot directly obtain (3) due to intractable integrals and summation over possible cases of binary variable $\mathbf{s}$.

VALSE [25, 26] is a variational Bayesian estimation method and finds a variational pdf $q(\boldsymbol{\theta}, \mathbf{w}, \mathbf{s}|\mathbf{y})$ that approximates $f(\boldsymbol{\theta}, \mathbf{w}, \mathbf{s}|\mathbf{y})$ well. By assuming that the noise $\mathbf{e}$ has statistically independent, circular-symmetric, and zero-mean Gaussian entries with variance ν. It gives the probability density function (pdf) of the likelihood from (1), i.e.,

$$f(\mathbf{y}|\boldsymbol{\theta}, \mathbf{w}, \mathbf{s}) = f_{\mathrm{CN}}\left(\mathbf{y}; \sum_{k=1}^{K} w_k \mathbf{a}(\theta_k), \nu \mathbf{I}_M\right). \tag{4}$$

Given (2), VALSE approximately represents marginal pdfs $f(\theta_l|\mathbf{y})$ and pmfs $p(s_l|\mathbf{y})$ by VM pdfs and Kronecker delta functions, respectively, i.e., $q(\theta_l|\mathbf{y}) = f_{\mathrm{VM}}(\theta_l|\mathbf{y})$ and $q(s_l|\mathbf{y}) = \delta(s_l - \hat{s}_l)$, $\hat{s}_l \in \{0, 1\}$. (s_l, w_l) has a Bernoulli-Gaussian distribution, i.e.,

$$f(w_l|s_l) = (1 - s_l)\delta(w_l) + s_l f_{\mathrm{CN}}(w_l; 0, \tau), \tag{5}$$

$$p(s_l) = \rho^{s_l}(1 - \rho)^{(1 - s_l)}. \tag{6}$$

3.1. Variational Bayesian Approach

The VALSE, a variational Bayesian approach, approximates $f(\boldsymbol{\theta}, \mathbf{w}, \mathbf{s}|\mathbf{y})$ with a variational pdf $q(\boldsymbol{\theta}, \mathbf{w}, \mathbf{s}|\mathbf{y})$ by minimizing the Kullback-Leibler (KL) divergence [32, pp. 732],

$$\begin{aligned}&\mathrm{KL}(q(\boldsymbol{\theta}, \mathbf{w}, \mathbf{s}|\mathbf{y})||f(\boldsymbol{\theta}, \mathbf{w}, \mathbf{s}|\mathbf{y}))\\ &= \int q(\boldsymbol{\theta}, \mathbf{w}, \mathbf{s}|\mathbf{y}) \ln \frac{q(\boldsymbol{\theta}, \mathbf{w}, \mathbf{s}|\mathbf{y})}{f(\boldsymbol{\theta}, \mathbf{w}, \mathbf{s}|\mathbf{y})}\, \mathrm{d}\boldsymbol{\theta}\mathrm{d}\mathbf{w}\mathrm{d}\mathbf{s}. \end{aligned}\tag{7}$$

For any assumed pdf $q(\boldsymbol{\theta}, \mathbf{w}, \mathbf{s}|\mathbf{y})$, the log marginal likelihood (model evidence) $\ln f(\mathbf{y})$ [32, pp. 732–733] is

$$\begin{aligned}\ln f(\mathbf{y}) = &\mathrm{KL}(q(\boldsymbol{\theta}, \mathbf{w}, \mathbf{s}|\mathbf{y})||f(\boldsymbol{\theta}, \mathbf{w}, \mathbf{s}|\mathbf{y}))\\ &+ \mathcal{L}\left(q(\boldsymbol{\theta}, \mathbf{w}, \mathbf{s}|\mathbf{y})\right), \end{aligned}\tag{8}$$

$$\mathcal{L}\left(q(\boldsymbol{\theta}, \mathbf{w}, \mathbf{s}|\mathbf{y})\right) = \mathbb{E}_{q(\boldsymbol{\theta}, \mathbf{w}, \mathbf{s}|\mathbf{y})}\left[\ln \frac{f(\mathbf{y}, \boldsymbol{\theta}, \mathbf{w}, \mathbf{s})}{q(\boldsymbol{\theta}, \mathbf{w}, \mathbf{s}|\mathbf{y})}\right]. \tag{9}$$

For a given data $\mathbf{y}$, $\ln f(\mathbf{y})$ is constant w.r.t. $q(\boldsymbol{\theta}, \mathbf{w}, \mathbf{s}|\mathbf{y})$ and $\mathrm{KL}(q(\boldsymbol{\theta}, \mathbf{w}, \mathbf{s}|\mathbf{y})||f(\boldsymbol{\theta}, \mathbf{w}, \mathbf{s}|\mathbf{y})) \geq 0$, minimizing the KL divergence is equivalent to maximizing $\mathcal{L}\left(q(\boldsymbol{\theta}, \mathbf{w}, \mathbf{s}|\mathbf{y})\right)$ (8).

To factorize $q(\boldsymbol{\theta}, \mathbf{w}, \mathbf{s}|\mathbf{y})$, we assume that the DOAs $\boldsymbol{\theta}$ are independent mutually and of the other variables; and the posterior of the support variable $q(\mathbf{s}|\mathbf{y})$ has all its mass at $\hat{\mathbf{s}}$, i.e., $q(\mathbf{s}|\mathbf{y}) = \delta(\mathbf{s} - \hat{\mathbf{s}})$, $\hat{s}_l \in \{0, 1\}$, where δ is the Kronecker delta function. Then the factorization $q(\boldsymbol{\theta}, \mathbf{w}, \mathbf{s}|\mathbf{y})$ is

$$q(\boldsymbol{\theta}, \mathbf{w}, \mathbf{s}|\mathbf{y}) = \prod_{l=1}^{L} q(\theta_l|\mathbf{y}) q(\mathbf{w}|\mathbf{y}, \mathbf{s}) \delta(\mathbf{s} - \hat{\mathbf{s}}). \tag{10}$$

3.2. Variational Bayesian Solution

We maximize the lower bound $\mathcal{L}$ (9) with $q(\boldsymbol{\theta}, \mathbf{w}, \mathbf{s}|\mathbf{y})$ (10). As maximizing $\mathcal{L}$ over all factors $\mathbf{\Phi} = \{\theta_1, \ldots, \theta_L, \{\mathbf{w}, \mathbf{s}\}\}$ together is intractable, we utilize alternating optimization. $\mathcal{L}$ is optimized over the parameters of each marginal distribution $q(\Phi_i|\mathbf{y})$, $i = 1, \ldots, L + 1$. Maximizing $\mathcal{L}$ over each of the factors $\mathbf{\Phi}$ is performed in turn while keeping the others fixed [32, pp. 735, Eq. (21.25)],

$$\ln q(\Phi_i|\mathbf{y}) = \mathbb{E}_{\sim q(\Phi_i|\mathbf{y})}[\ln f(\mathbf{\Phi}, \mathbf{y})] + \text{const}, \tag{11}$$

where $\mathbb{E}_{\sim q(\Phi_i|\mathbf{y})}[\ln f(\mathbf{\Phi}, \mathbf{y})]$ is the expectation over $\ln f(\mathbf{\Phi}, \mathbf{y})$ with respect to all the variables except for Φ_i. The alternating optimization is performed: inferring DOAs $\boldsymbol{\theta}$ [Sec. 3.2.A], inferring source amplitudes $\mathbf{w}$ and support $\mathbf{s}$ [Sec. 3.2.B], and estimating model parameters $\boldsymbol{\beta}(= \{\nu, \tau, \rho\})$ [Sec. 3.2.C].

A. Inferring DOAs $\boldsymbol{\theta}$. For each $i = 1, \ldots, L$, we estimate θ_i from (11) with $\mathbb{E}_{\sim q(\theta_i|\mathbf{y})}[\ln f(\mathbf{\Phi}, \mathbf{y})]$. For the resulting update, see [25, Eq. (16)] and [26, Appendix A.1]. $q(\theta_i|\mathbf{y})$ is well approximated by VM pdfs, and we use [25, Heuristic 1].

B. Inferring $\mathbf{w}$ and $\mathbf{s}$. We estimate $\mathbf{w}$ and $\mathbf{s}$ by using (11) with $\mathbb{E}_{\sim q(\mathbf{w}, \mathbf{s}|\mathbf{y})}[\ln f(\mathbf{\Phi}, \mathbf{y})]$. For the resulting update, see [25, Eq. (20)]. We obtain $\hat{\mathbf{s}}$ by plugging the postulated pdf (10) in (9), see [25, Algorithm 4] and [26, Appendix A.2].

C. Estimating model parameters $\{\nu$ (4), τ (5), ρ (6)$\}$. We obtain the model parameters $\boldsymbol{\beta}(= \{\nu, \tau, \rho\})$ by using (11) with $\mathbb{E}_{\sim q(\boldsymbol{\theta}, \mathbf{w}, \mathbf{s}|\mathbf{y})}[\ln f(\mathbf{\Phi}, \mathbf{y}; \boldsymbol{\beta})]$. For the resulting update, see [25, Eqs. (24), (25)] and [26, Appendix A.3].

4. PROPOSED LEARNING-AIDED INITIALIZATION

The VALSE requires an initial guess of the noise variance ν and $\{\mathbf{a}(\theta_i)\}_{i=1}^{L}$ for inferring DOAs [Section 3.2.A]. SBL provides a good initial guess based on a discrete angular search grid. A linear model for DOA uses the grid points of $N \gg K$ potential DOAs $\{\bar{\theta}_1 \ldots \bar{\theta}_N\}$ and the corresponding amplitude vector $\mathbf{x} \in \mathbb{C}^N$. With the dictionary $\mathbf{A} = [\mathbf{a}(\bar{\theta}_1) \ldots \mathbf{a}(\bar{\theta}_N)]$, the alternative linear model for (1) is given by

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}. \tag{12}$$

SBL is a Bayesian approach with a likelihood and a prior. The likelihood is the same as in VALSE (4) but with (12), i.e.,

$$f(\mathbf{y}|\mathbf{x}; \nu) = f_{\mathrm{CN}}\big(\mathbf{y}; \mathbf{A}\mathbf{x}, \nu \mathbf{I}_M\big). \tag{13}$$

Authorized licensed use limited to: Univ of Calif San Diego. Downloaded on February 21,2023 at 21:18:58 UTC from IEEE Xplore. Restrictions apply.

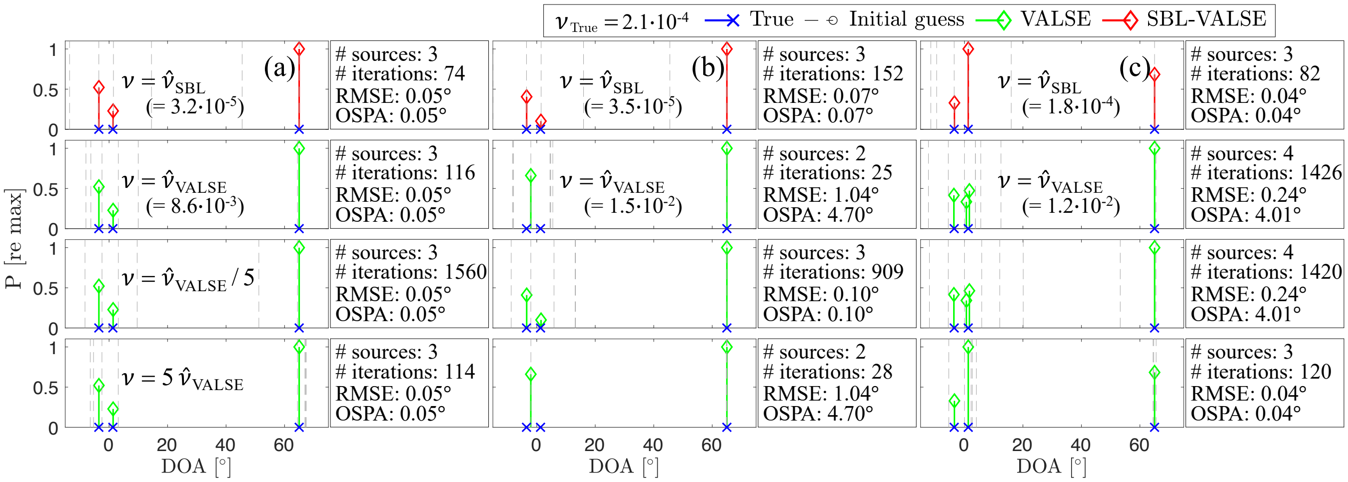

Fig. 1. Effect of varying the initial guess of $\boldsymbol{\theta}$ and ν. The scenario described in Sec. 5 with an SNR of 40 dB is considered. (a) shows a case where all considered values for $\boldsymbol{\theta}$ and ν were successful in detecting the three sources. (b) shows a case where the original VALSE fails to detect all three sources, but all three sources are detected after scaling ν by a factor of $1/5$. (c) shows a case where the original VALSE over-estimates the number of sources, but after scaling ν by a factor of 5, the number of sources is estimated correctly.

The prior distribution of $\mathbf{x}$ is assumed multi-variate Gaussian with zero-mean and variance with DOA-dependent hyperparameter $\boldsymbol{\gamma} \in \mathbb{R}^N$, i.e.,

$$f(\mathbf{x};\boldsymbol{\gamma}) = f_{\mathrm{CN}}\big(\mathbf{x};\mathbf{0},\boldsymbol{\Gamma}\big), \tag{14}$$

where $\boldsymbol{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma}) \in \mathbb{R}^{N\times N}$. From (13) and (14), the evidence $f(\mathbf{y})$ is Gaussian, i.e., $f(\mathbf{y}) = f_{\mathrm{CN}}\big(\mathbf{y};\mathbf{0},\boldsymbol{\Sigma}_{\mathbf{y}}\big)$, where $\boldsymbol{\Sigma}_{\mathbf{y}} = \nu\mathbf{I}_M + \mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^{\mathrm{H}}$, and SBL obtains $\{\boldsymbol{\gamma},\nu\}$ by maximizing the evidence,

$$\{\hat{\boldsymbol{\gamma}},\hat{\nu}\} = \underset{\boldsymbol{\gamma}\geq 0,\,\nu>0}{\arg\max}\; f\big(\mathbf{y};\boldsymbol{\gamma},\nu\big) = \underset{\boldsymbol{\gamma}\geq 0,\,\nu>0}{\arg\max}\; \frac{e^{-\mathbf{y}^{\mathrm{H}}\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{y}}}{\det\boldsymbol{\Sigma}_{\mathbf{y}}}. \tag{15}$$

The SBL iteratively updates $\{\hat{\boldsymbol{\gamma}},\hat{\nu}\}$ [3, 16, 19],

$$\gamma_n^{\mathrm{new}} = \gamma_n^{\mathrm{old}}\frac{\|\mathbf{y}^{\mathrm{H}}\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{a}_n\|_2^2}{\mathbf{a}_n^{\mathrm{H}}\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{a}_n}, \tag{16}$$

$$\nu^{\mathrm{new}} = \frac{\mathrm{tr}\left[\left(\mathbf{I}_M - \mathbf{A}_{\mathcal{M}}\mathbf{A}_{\mathcal{M}}^{+}\right)\mathbf{y}\mathbf{y}^{\mathrm{H}}\right]}{M-K}, \tag{17}$$

where $\mathcal{M}$ is a given set of K active DOAs; $\mathbf{A}_{\mathcal{M}}$ is the corresponding active steering matrix; and $\mathbf{A}_{\mathcal{M}}^{+}$ denotes the Moore-Penrose pseudo-inverse [3, 16, 19].

The DOA performance of the VALSE algorithm depends on the initial guess of $\{\rho,\tau,\nu,\{\mathbf{a}(\theta_i)\}_{i=1}^{L}\}$. A better initial guess of DOAs $\{\theta_i\}_{i=1}^{L}$ provides more accurate DOAs with a faster convergence, see Fig. 1 (the scenario in Sec. 5). For the initial guess of $\{\theta_i\}_{i=1}^{L}$, the original VALSE uses an improper flat prior for $\mathbf{w}$ due to a lack of knowledge of $\mathbf{w}$. It uses noncoherent estimation [25, Eq. (38)] for $\{q(\theta_i|\mathbf{y})\}_{i=1}^{L}$ and constructs L components in a greedy manner. Regarding the initial guess for ν, the original VALSE builds a Toeplitz estimate of the $\mathbb{E}[\mathbf{y}\mathbf{y}^{\mathrm{H}}]$ by computing the average of the diagonal elements of $\mathbf{y}\mathbf{y}^{\mathrm{H}}$. Then it initializes ν with the average of the lower quarter of the eigenvalues of the Toeplitz matrix.

The noise variance ν controls the sparsity level of the solution. The solution becomes sparse for large ν as it admits more noise and fits the data with sparser components. With a good initial guess of $\boldsymbol{\theta}$, the VALSE shows stable DOA estimation performance with ν, see Fig. 1(a). (Too small ν requires more iterations to converge.) Even with high SNR, the original VALSE scheme has a poor initial DOA guess. The poor initial DOA guess can be improved by adjusting ν. If too few DOAs are detected, a smaller initial noise variance ν leads to a less sparse solution, and missing DOA may be recovered, see Fig. 1(b). If the method over-estimates the number of sources, a larger initial ν promotes a sparser solution, and the correct number of sources may be recovered, see Fig. 1(c).

We use the values $\{\hat{\gamma}_{\mathrm{SBL}},\hat{\nu}_{\mathrm{SBL}}\}$ in (16) and (17) provided by SBL as and initialization for VALSE. Then, we construct $\{\mathbf{a}(\theta_i)\}_{i=1}^{L}$ with θ_i corresponding to the location of the L strongest peaks in $\hat{\gamma}_{\mathrm{SBL}}$. SBL is known to under-estimates the noise variance. Thus, we set the noise variance used for the initialization of VALSE to $\nu = 10\hat{\nu}_{\mathrm{SBL}}$, which showed favorable DOA performance. We initialize $\tau = (\mathbf{y}^{\mathrm{H}}\mathbf{y}/M - \nu)/(\rho L)$ (5) and $\rho = 0.5$ (6) [25, 26].

5. SIMULATION RESULTS

We examine a scenario with three sources with DOAs $[-3.5, 1.3, 65]^{\circ}$. Source magnitudes are complex zero-mean Gaussian with variance 1. We consider a ULA with $M = 16$ elements, half-wavelength spacing. The noise is modeled to achieve the signal-to-noise ratio (SNR), $\mathrm{SNR} = 20\log_{10}\big[\|\sum_{k=1}^{K} w_k\mathbf{a}(\theta_k)\|_2/\|\mathbf{e}\|_2\big]$. The performance of the DOA methods is compared in Fig. 2. We use metrics the root mean square error (RMSE), $\mathrm{RMSE} = \sqrt{\mathbb{E}\big[\frac{1}{K}\sum_{k=1}^{K}(\hat{\theta}_k - \theta_k)^2\big]}$, where $\hat{\theta}_k$ and θ_k represent estimated and true DOA of the kth source, and the Euclidean

Authorized licensed use limited to: Univ of Calif San Diego. Downloaded on February 21,2023 at 21:18:58 UTC from IEEE Xplore. Restrictions apply.

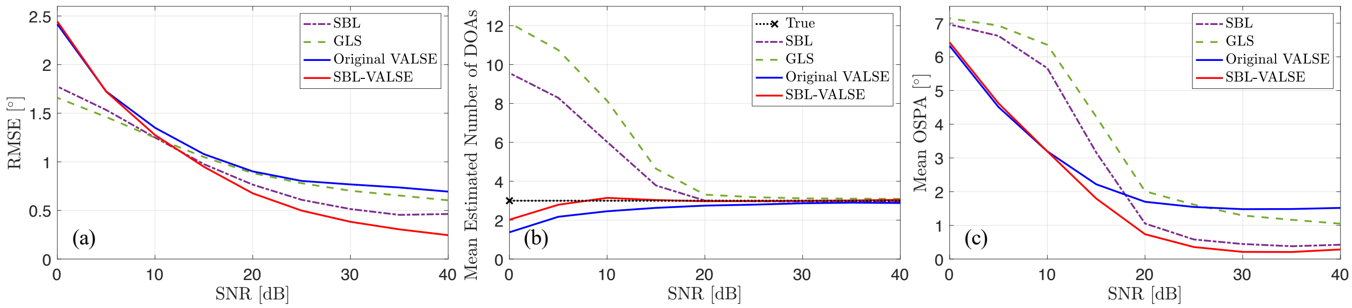

Fig. 2. DOA estimation by SBL, GLS, VALSE, and SBL-VALSE: (a) RMSE, (b) Mean estimated number of DOAs versus SNR and (c) mean OSPA error versus SNR. All results are based on 500 Monte Carlo simulations.

distance-based optimal subpattern assignment (OSPA) [33] with cutoff parameter 8°. The OSPA considers the estimated DOAs and the number of estimated DOAs.

We compare the suggested method with SBL [16], GLS [13], and the original VALSE [25]. For SBL, we use the steering vectors with the DOA-search grid $\bar{\boldsymbol{\theta}}=[-90{:}0.5{:}90]°$. 0.2° basis mismatch for DOA at 1.3° leads SBL to have a lower bound RMSE 0.12°. After performing SBL or GLS, we obtain final DOA estimates by using a threshold with ξ_{TH}-times the maximum source magnitude. We set $\xi_{\mathrm{TH}} = 0.01$ and $\xi_{\mathrm{TH}} = 0.1$ for SBL and GLS, showing favorable estimation.

SBL and GLS perform better with RMSE in low SNRs [Fig. 2(a)]. It comes from that they over-estimate the number of DOAs [Fig. 2(b)], which is unfavorable with OSPA [Fig. 2(c)]. SBL-VALSE improves the performance over the original VALSE. It estimates accurate DOAs due to its gridless nature and the correct number of sources at larger SNRs.

6. EXPERIMENTAL RESULTS

We validate the proposed method by using experimental data from the shallow water evaluation cell experiment 1996 Event S5 (SWellEx-96) [2, 19, 34]. Acoustic data was recorded by a 64-element vertical linear array with a uniform inter-sensor spacing of 1.875 m spanning water depths 94.125–212.25 m. We focus on a source towed at 54 m depth and its 166 Hz frequency component. The considered data was sampled at 1500 Hz. We used the record at 23:37–00:32 GMT divided into non-overlapping 300 snapshots. Each snapshot data is Fourier transformed with 2^{14} samples.

Single-snapshot DOA estimation versus time shows multiple DOAs in Fig. 3. At each time snapshot, the SBL-VALSE is initialized with SBL. Due to the underwater channel characteristics, the single source arrives at the array via multipath arrivals, corresponding to the multiple DOAs. To identify the DOA structure and variability over time, we simulate the acoustic field using the Kraken normal mode model [35]. CBF results with the simulated field [Fig. 3(a)] match well with those with the experimental data [Fig. 3(b)]. Discontinuity area [snapshot 90, 185, 200] is when the source was turned off, and ambient noise shows strong DOAs $[-8, 8]°$. SBL-VALSE [Fig. 3(c)] results in improved source tracking resolution by promoting sparsity, capturing snapshot-varying DOA structure, and reducing artifacts compared to the CBF.

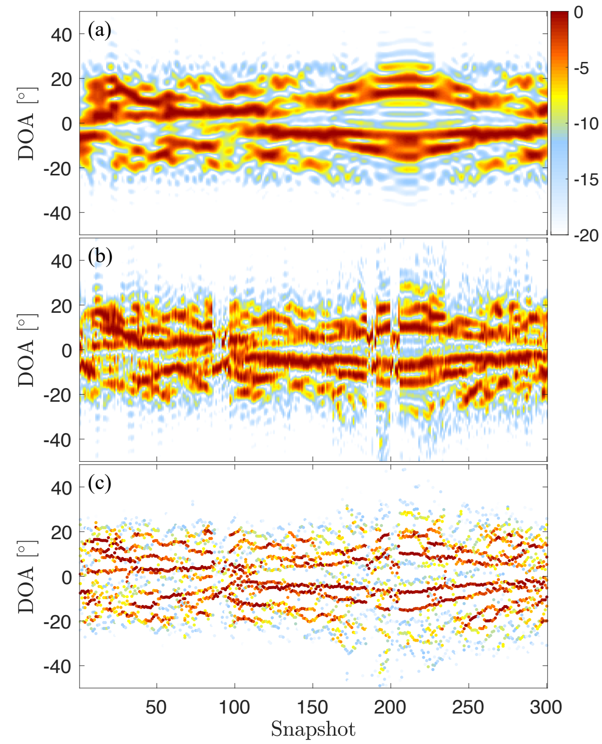

Fig. 3. DOA estimation versus snapshot for (a) CBF using simulated data from Kraken model and for (b) CBF and (c) SBL-VALSE using acoustic data from the SWellEx-96 experiment. Colored dots show the SBL-VALSE solution with source magnitudes.

7. CONCLUSION

We introduced a Bayesian method for estimating DOAs that promotes sparse solutions and is gridless. The approach is based on variational Bayesian line spectral estimation (VALSE), which is sensitive to an initial guess of the measurement noise variance and DOAs. To increase the robustness and accuracy of VALSE, we established an initialization based on sparse Bayesian learning. Performance evaluation based on simulated and real data from an ocean acoustics experiment showed good DOA estimation performance. A direction for future research is a combination of the proposed approach with sequential processing [19, 36] for time-varying DOAs and multiobject tracking [37–39].

Authorized licensed use limited to: Univ of Calif San Diego. Downloaded on February 21,2023 at 21:18:58 UTC from IEEE Xplore. Restrictions apply.

8. REFERENCES

[1] D. Malioutov, M. Cetin, and A. S. Willsky, “A sparse signal reconstruction perspective for source localization with sensor arrays,” *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[2] P. Gerstoft, A. Xenaki, and C. F. Mecklenbräuker, “Multiple and single snapshot compressive beamforming,” *J. Acoust. Soc. Am.*, vol. 138, no. 4, pp. 2003–2014, Oct. 2015.

[3] P. Gerstoft, C. F. Mecklenbräuker, A. Xenaki, and S. Nannuru, “Multisnapshot sparse Bayesian learning for DOA,” *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1469–1473, Oct. 2016.

[4] A. Xenaki, P. Gerstoft, and K. Mosegaard, “Compressive beamforming,” *J. Acoust. Soc. Am.*, vol. 136, no. 1, pp. 260–271, July 2014.

[5] A. Xenaki and P. Gerstoft, “Grid-free compressive beamforming,” *J. Acoust. Soc. Am.*, vol. 137, no. 4, pp. 1923–1935, Apr. 2015.

[6] R. R. Pote and B. D. Rao, “Robustness of sparse Bayesian learning in correlated environments,” in *Proc. IEEE ICASSP*, 2020, pp. 9100–9104.

[7] Z. Yang, J. Li, P. Stoica, and L. Xie, “Sparse methods for direction-of-arrival estimation,” in *Academic Press Library in Signal Processing: Array, Radar and Communications Engineering*, vol. 7, chapter 11, pp. 509–581. Academic Press, 2018.

[8] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, “Compressed sensing off the grid,” *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7465–7490, Nov. 2013.

[9] P. Chen, Z. Chen, Z. Cao, and X. Wang, “A new atomic norm for DOA estimation with gain-phase errors,” *IEEE Trans. Signal Process.*, vol. 68, pp. 4293–4306, July 2020.

[10] Y. Chi and M. Ferreira Da Costa, “Harnessing sparsity over the continuum: Atomic norm minimization for superresolution,” *IEEE Signal Process. Mag.*, vol. 37, no. 2, pp. 39–57, Mar. 2020.

[11] Y. Park, Y. Choo, and W. Seong, “Multiple snapshot grid free compressive beamforming,” *J. Acoust. Soc. Am.*, vol. 143, no. 6, pp. 3849–3859, June 2018.

[12] M. Wagner, Y. Park, and P. Gerstoft, “Gridless DOA estimation and root-MUSIC for non-uniform linear arrays,” *IEEE Trans. Signal Process.*, vol. 69, pp. 2144–2157, Mar. 2021.

[13] Z. Yang, L. Xie, and C. Zhang, “A discretization-free sparse and parametric approach for linear array signal processing,” *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 4959–4973, Oct. 2014.

[14] D. P. Wipf and B. D. Rao, “An empirical Bayesian strategy for solving the simultaneous sparse approximation problem,” *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3704–3716, July 2007.

[15] Z. M. Liu, Z. T. Huang, and Y. Y. Zhou, “An efficient maximum likelihood method for direction-of-arrival estimation via sparse Bayesian learning,” *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 1–11, Oct. 2012.

[16] S. Nannuru, A. Koochakzadeh, K. L. Gemba, P. Pal, and P. Gerstoft, “Sparse Bayesian learning for beamforming using sparse linear arrays,” *J. Acoust. Soc. Am.*, vol. 144, no. 5, pp. 2719–2729, Nov. 2018.

[17] Q. Wang, H. Yu, J. Li, F. Ji, and F. Chen, “Sparse Bayesian learning using generalized double Pareto prior for DOA estimation,” *IEEE Signal Process. Lett.*, vol. 28, pp. 1744–1748, Aug. 2021.

[18] J. Dai and H. C. So, “Real-valued sparse Bayesian learning for DOA estimation with arbitrary linear arrays,” *IEEE Trans. Signal Process.*, vol. 69, pp. 4977–4990, Aug. 2021.

[19] Y. Park, F. Meyer, and P. Gerstoft, “Sequential sparse Bayesian learning for time-varying direction of arrival,” *J. Acoust. Soc. Am.*, vol. 149, no. 3, pp. 2089–2099, Mar. 2021.

[20] Z. Yang, L. Xie, and C. Zhang, “Off-grid direction of arrival estimation using sparse Bayesian inference,” *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 38–43, Jan. 2013.

[21] X. Wu, W. Zhu, and J. Yan, “Direction of arrival estimation for off-grid signals based on sparse Bayesian learning,” *IEEE Sens. J.*, vol. 16, no. 7, pp. 2004–2016, Apr. 2016.

[22] Q. Wang, Z. Zhao, Z. Chen, and Z. Nie, “Grid evolution method for DOA estimation,” *IEEE Trans. Signal Process.*, vol. 66, no. 9, pp. 2374–2383, May 2018.

[23] R. Zheng, X. Xu, Z. Ye, T. H. Al Mahmud, J. Dai, and K. Shabir, “Sparse Bayesian learning for off-grid DOA estimation with Gaussian mixture priors when both circular and non-circular sources coexist,” *Signal Process.*, vol. 161, pp. 124–135, Aug. 2019.

[24] J. Yang and Y. Yang, “Sparse Bayesian DOA estimation using hierarchical synthesis lasso priors for off-grid signals,” *IEEE Trans. Signal Process.*, vol. 68, pp. 872–884, Jan. 2020.

[25] M. A. Badiu, T. L. Hansen, and B. H. Fleury, “Variational Bayesian inference of line spectra,” *IEEE Trans. Signal Process.*, vol. 65, no. 9, pp. 2247–2261, May 2017.

[26] 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. 161, pp. 155–164, Aug. 2019.

[27] J. Yang, Y. Yang, and J. Lu, “A variational Bayesian strategy for solving the DOA estimation problem in sparse array,” *Signal Process.*, vol. 90, pp. 28–35, July 2019.

[28] Y. Zhang, Y. Yang, and L. Yang, “Off-grid DOA estimation through variational Bayesian inference in colored noise environment,” *Digit. Signal Process.*, vol. 111, Apr. 2021.

[29] F. Meyer, Y. Park, and P. Gerstoft, “Variational Bayesian estimation of time-varying DOAs,” in *Proc. IEEE FUSION*, 2020, pp. 1–6.

[30] Q. Zhang, J. Zhu, Y. Gu, and Z. Xu, “Gridless variational direction-of-arrival estimation in heteroscedastic noise environment,” *IEEE J. Ocean. Eng.*, Apr. 2021.

[31] P. Stoica and Y. Selen, “Model-order selection: a review of information criterion rules,” *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, July 2004.

[32] K. P. Murphy, *Machine learning: a probabilistic perspective*, MIT Press, Cambridge, MA, 2006.

[33] D. Schuhmacher, B. T. Vo, and B. N. Vo, “A consistent metric for performance evaluation of multi-object filters,” *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3447–3457, Aug. 2008.

[34] F. Meyer and K. L. Gemba, “Probabilistic focalization for shallow water localization,” *J. Acoust. Soc. Am.*, vol. 150, no. 2, pp. 1057–1066, Oct. 2021.

[35] M. B. Porter, “The KRAKEN normal mode program,” Tech. Rep., SACLANT Undersea Research Centre, La Spezia, Italy, 1991.

[36] C. F. Mecklenbräuker, P. Gerstoft, A. Panahi, and M. Viberg, “Sequential Bayesian sparse signal reconstruction using array data,” *IEEE Trans. Signal Process.*, vol. 61, no. 24, pp. 6344–6354, Sept. 2013.

[37] F. Meyer, P. Braca, P. Willett, and F. Hlawatsch, “A scalable algorithm for tracking an unknown number of targets using multiple sensors,” *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3478–3493, July 2017.

[38] F. Meyer, T. Kropfreiter, J. L. Williams, R. A. Lau, F. Hlawatsch, P. Braca, and M. Z. Win, “Message passing algorithms for scalable multitarget tracking,” *Proc. IEEE*, vol. 106, no. 2, pp. 221–259, Feb. 2018.

[39] F. Meyer and J. L. Williams, “Scalable detection and tracking of geometric extended objects,” *IEEE Trans. Signal Process.*, vol. 69, pp. 6283–6298, Oct. 2021.

Authorized licensed use limited to: Univ of Calif San Diego. Downloaded on February 21,2023 at 21:18:58 UTC from IEEE Xplore. Restrictions apply.