Variational Bayesian Estimation of Time-Varying DOAs

Florian Meyer, Yongsung Park, and Peter Gerstoft

Scripps Institution of Oceanography and Department of Electrical and Computer Engineering
University of California San Diego, La Jolla, CA
Email: {flmeyer, yop001, pgerstoft}@ucsd.edu

***Abstract*—We present a Bayesian method for sequential direction finding based on variational line spectral estimation (VALSE). The proposed method promotes sparse solutions by means of a Bernoulli-Gaussian amplitude model, is grid-less, and provides marginal posterior distributions from which DOA estimates and their uncertainties can be extracted. Simulation results demonstrate performance improvements in the considered scenario. We also evaluate the proposed method using acoustic data from an underwater source localization experiment.**

***Index Terms*—Array processing, direction of arrival (DOA) estimation, sparsity, factor graphs.**

I. Introduction

Direction of arrival (DOA) estimation or direction finding is the task of localizing several sources from noisy measurements provided by an array of sensors. It is a key problem in applications including radar [1], sonar [2], medical analytics [3], and seismic systems [4].

A. State of the Art Methods

For an overview of traditional DOA estimation methods, the reader is referred to [5]–[8] and references therein. In most methods, nonlinear estimation of angles is avoided by creating a grid of potential DOAs ("steering vectors"). In particular, the framework of compressive sensing (CS) [9]–[13], which is based on a convex optimization procedure that explicitly promotes sparse solutions, was recently employed for DOA estimation. CS for DOA [14], [15] has high-resolution estimation capabilities and, contrary to eigenvalue-based traditional DOA estimation methods [5]–[8], also performs well in scenarios with coherent sources [11]. However, CS-based DOA methods are limited by (i) basis mismatch [16], which occurs when DOAs do not coincide with the look directions of steering vectors; and (ii) basis coherence [17], which is caused by a dense grid of steering vectors and results in biased estimates.

Sparse Bayesian learning (SBL) [18], [19] is another recently introduced framework for DOA estimation. In SBL, a sparse source amplitude vector is implicitly promoted by a hierarchical prior model for the variance of its entries [18]–[22]. Estimation is performed on a grid potentially followed by subsequent refinement. SBL-based methods provide accurate DOA estimates but are known to overestimate the model order, i.e., they typically output additional low-power sources [20]. An overview and discussion of CS and SBL type methods and insight into their differences is provided in [23]. Sequential sparse DOA estimation methods based on the CS [24], [25] and SBL framework [26] have been introduced. These methods determine the parameters of sparsity-promoting prior distributions by using statistical information from previous time steps. All existing sequential approaches are grid-based and only provide point estimates of DOAs.

Among emerging gridless non-sequential methods suitable for direction finding [27]–[32], the variational line spectral estimation (VALSE) algorithm [31], [32], which explicitly promotes sparsity by means of a Bernoulli-Gaussian prior model, has the favorable properties that (i) it is guaranteed to converge, (ii) it can incorporate prior information, and (iii) it can provide posterior distributions of DOAs.

B. Contribution and Notation

We present a Bayesian method for sequential DOA estimation. To promote sparsity, we extend the Bernoulli-Gaussian amplitude model introduced in [31] by a state-transition model that enables sequential Bayesian estimation. The proposed method is based on the VALSE algorithm [31], [32]. In particular, at each time step an instance of VALSE that uses predicted posterior distributions of previous time steps as prior information is executed. In contrast to non-sequential approaches, our method can reduce DOA estimation errors in scenarios involving multiple time steps and time-varying DOAs. Contrary to existing sequential sparse DOA estimation methods [24]–[26], our approach is grid-less and provides marginal posterior distributions from which all DOA estimates and their uncertainties can be extracted. Numerical results demonstrate that, in scenarios with DOAs located at angles near endfire of the array, performance improvements are achieved compared to state of the art methods. Key contributions are as follows.

- We formulate sequential Bayesian DOA estimation where sparsity is promoted by a Bernoulli-Gaussian model.
- We compare the performance of sequential and non-sequential DOA estimation.
- We apply the presented method to measurement data from an underwater source localization experiment.

Notation: We will use the following basic notation. Random variables are displayed in sans serif, upright fonts; their realizations in serif, italic fonts. Vectors and matrices are denoted by bold lowercase and uppercase letters, respectively.

For example, x is a random variable and x is its realization, and $\mathbf{x}$ is a random vector and $\boldsymbol{x}$ is its realization. Furthermore, $f(\boldsymbol{x})$ denotes the probability density function (PDF) of continuous random vector $\mathbf{x}$ (this is a short notation for $f_{\mathbf{x}}(\boldsymbol{x})$), $p(\boldsymbol{s})$ denotes the PDF of discrete random vector $\mathbf{s}$ (this is a short notation for $p_{\mathbf{s}}(\boldsymbol{s})$), and $\propto$ indicates equality up to a normalization factor. $f_{\mathrm{CN}}(\cdot;\boldsymbol{\mu},\boldsymbol{\Sigma})$ denotes the complex multivariate Gaussian pdf with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$.

II. System Model

We observe narrowband signals from K_n sources with frequency ω on an array of L sensors. Let c and $d \leqslant \frac{c\pi}{\omega}$ be propagation speed and sensor spacing, respectively. At time step $n \in \{0,1,\ldots\}$, the received time-sampled signal $\mathbf{y}_n = [\mathsf{y}_{0,n} \ldots \mathsf{y}_{L-1,n}]^{\mathrm{T}} \in \mathbb{C}^L$ consists of elements

$$\mathsf{y}_{l,n} = \sum_{k=1}^{\mathsf{K}_n} \boldsymbol{\alpha}_{k,n} e^{jl\frac{\omega d}{c}\sin\beta_{k,n}} + \mathsf{v}_{l,n}$$

where $\beta_{k,n} \in [-90°, 90°)$ and $\boldsymbol{\alpha}_{k,n} \in \mathbb{C}$ are the angle of arrival and the complex amplitude, respectively, of the signal component originated by source k, and $\mathsf{v}_{l,n} \in \mathbb{C}$ is additive noise. Note that at each time n, the complex amplitudes $\boldsymbol{\alpha}_{k,n}$ and angles $\beta_{k,n}$, $k \in \{1,\ldots,\mathsf{K}_n\}$ as well as the number K_n of components itself is unknown. This is typically referred to as the model-order selection problem [33].

Due to the unknown number of components K_n, we follow the approach in [31] and consider at most K potential sources. Each potential source corresponds to a random angle of arrival and weight. We formulate the underlying problem as a line spectral estimation [8] problem by introducing pseudo angles (PAs) $\theta_{k,n} = \frac{\omega d}{c}\sin\beta_{k,n} \in \Pi \triangleq [-\pi,\pi)$ and model the measurements vector $\mathbf{y}_n$ as

$$\mathbf{y}_n = \sum_{k=1}^{K} \mathsf{w}_{k,n}\boldsymbol{a}(\theta_{k,n}) + \mathbf{u}_n \tag{1}$$

where $\boldsymbol{a}(\theta_{k,n}) \triangleq [1\ e^{j\theta_{k,n}} \ldots e^{j(L-1)\theta_{k,n}}]^{\mathrm{T}}$ and $\mathbf{u}_n \in \mathbb{C}^L$ is the measurement noise. For future reference, we introduce the vector $\mathbf{w}_n \triangleq [\mathsf{w}_{1,n} \ldots \mathsf{w}_{K,n}]^{\mathrm{T}} \in \mathbb{C}^K$ as well as the ordered sequences $\mathbf{w} \triangleq (\mathbf{w}_1,\ldots,\mathbf{w}_n)$ and $\mathbf{y} \triangleq (\mathbf{y}_1,\ldots,\mathbf{y}_n)$.

Sparsity-promoting prior PDF for $\mathbf{w}$: At each time n, only K_n of the K components have nonzero weights. For this reason, we use a sparsity-promoting prior for the weights. The complex weights $\mathsf{w}_{k,n}$ are governed by independent Bernoulli variables $\mathsf{s}_{k,n} \in \mathcal{B} \triangleq \{0,1\}$, i.e.,

$$f(w_{k,n}|s_{k,n}) = (1-s_{k,n})\delta(w_{k,n}) + s_{k,n} f_{\mathrm{CN}}(w_{k,n};0,\tau).$$

This means that the binary variable $s_{k,n}$ can "deactivate" the kth component, i.e., $\mathsf{s}_{k,n} = 0$ implies that $\mathsf{w}_{k,n} = 0$. If $\mathsf{s}_{k,n} = 1$, the kth component is active and $\mathsf{w}_{k,n}$ is zero-mean Gaussian with variance τ. We also introduce vectors $\boldsymbol{\theta}_n \triangleq [\theta_{1,n} \ldots \theta_{K,n}]^{\mathrm{T}} \in \Pi^K$ and $\mathbf{s}_n \triangleq [\mathsf{s}_{1,n} \ldots \mathsf{s}_{K,n}]^{\mathrm{T}} \in \mathcal{B}^K$ as well as ordered sequences $\mathbf{s} \triangleq (\mathbf{s}_1,\ldots,\mathbf{s}_n)$ and $\boldsymbol{\theta} \triangleq (\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_n)$.

It is assumed that given $\mathbf{s}$, $\mathbf{w}$ is statistically independent of $\boldsymbol{\theta}$ and that the entries of $\mathbf{w}$ are statistical independent across k and n, i.e.,

$$f(\boldsymbol{w}|\boldsymbol{s},\boldsymbol{\theta}) = f(\boldsymbol{w}|\boldsymbol{s}) = \prod_{n'=1}^{n}\prod_{k=1}^{K} f(w_{k,n'}|s_{k,n'}). \tag{2}$$

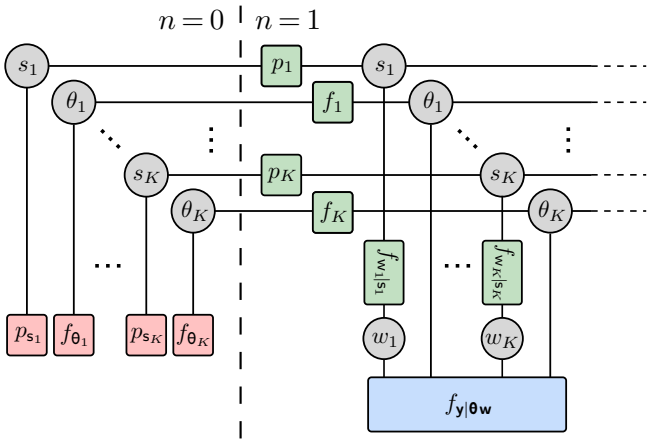


Fig. 1. Factor graph for variational Bayesian estimation of time-varying DOAs corresponding to the factorization (5). Factor nodes and variable nodes are depicted as boxes and circles, respectively. Boxes in red, green, and blue represent factors related to the initial prior PDFs, conditional prior PDFs, and likelihood functions, respectively. The time indexes n are omitted in the notations of the nodes, and the following short notations are used: $f_{\theta_l} \triangleq f(\theta_{l,n})$, $p_{\mathsf{s}_l} \triangleq p(s_{l,n})$, $f_l \triangleq f(\theta_{l,n}|\theta_{l,n-1})$, $p_l \triangleq p(s_{l,n}|s_{l,n-1})$, $f_{\mathsf{w}_l|\mathsf{s}_l} \triangleq f(w_{l,n}|s_{l,n})$, $f_{\mathbf{y}|\boldsymbol{\theta}\mathbf{w}} \triangleq f(\boldsymbol{y}_n|\boldsymbol{\theta}_n,\boldsymbol{w}_n)$, $\theta_l \triangleq \theta_{l,n}$, $w_l \triangleq w_{l,n}$, and $s_l \triangleq s_{l,n}$.

Prior PDF for $\boldsymbol{\theta}$ *and* $\mathbf{s}$*:* PAs $\boldsymbol{\theta}_n$ and Bernoulli variables $\mathbf{s}_n$ are assumed to evolve independently and according to a first-order Markov model [34], [35]. Furthermore, at time $n=0$, they are assumed statistically independent across k with prior PDF $f(\boldsymbol{\theta}_0,\boldsymbol{s}_0) = \prod_{k=1}^{K} f(\theta_{k,0})p(s_{k,0})$. Thus, the prior pdf of $\boldsymbol{\theta}$ and $\boldsymbol{s}$ reads

$$f(\boldsymbol{\theta},\boldsymbol{s}) = \left(\prod_{k=1}^{K} f(\theta_{k,0})p(s_{k,0})\right) \times \prod_{l=1}^{K}\prod_{n'=1}^{n} f_{\mathrm{VM}}(\theta_{l,n'}|\theta_{l,n'-1})p(s_{l,n'}|s_{l,n'-1}). \tag{3}$$

Here, $f_{\mathrm{VM}}(\theta_{k,n}|\theta_{k,n-1})$ is an arbitrary von Mises (VM) transition PDF and the transition probability mass function (PMF) $p(s_{k,n}|s_{k,n-1})$ is defined as follows. If component k is active at time $n-1$, i.e., $\mathsf{s}_{k,n-1}=1$, then the probability that it is inactive at time n, i.e., $\mathsf{s}_{k,n}=0$, is given by the deactivation probability p^{d}. Furthermore, if PA k was inactive at time $n-1$, i.e., $\mathsf{s}_{k,n-1}=0$, then the probability that it is active at time n, i.e., $\mathsf{s}_{k,n}=1$, is given by the activation probability p^{a}. At time $n=0$, the PAs have an arbitrary VM PDF and are assumed statistically independent, i.e., $f(\boldsymbol{\theta}_0) = \prod_{k=1}^{K} f_{\mathrm{VM}}(\theta_{k,0})$.

The VM PDF: The VM PDF [36] of angle $\theta \in [-\pi,\pi)$ is defined as

$$f_{\mathrm{VM}}(\theta) = \frac{1}{2\pi I(\kappa)} e^{\kappa\cos(\theta-\mu)}$$

where μ and κ are mean concentration and direction, respectively, and $I(\cdot)$ is the modified Bessel function of first kind

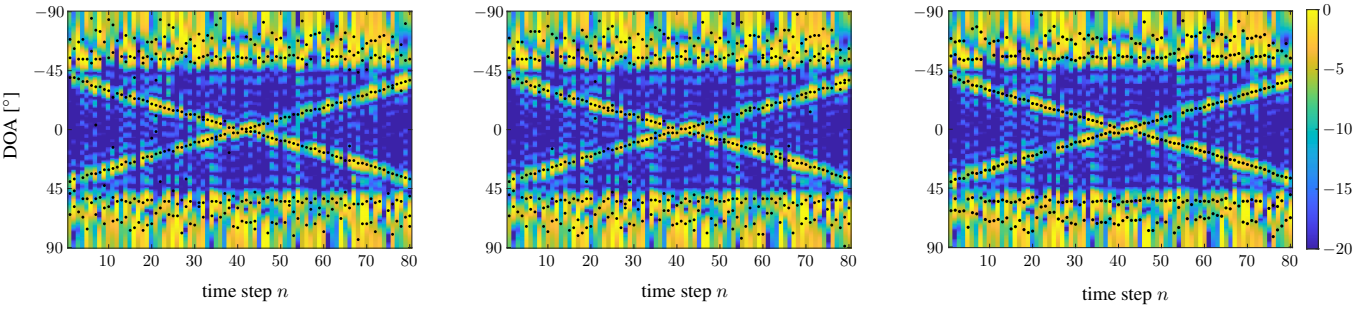


Fig. 2. DOA versus time steps for (a) SBL, (b) SCS, and (c) proposed SVALSE in the simulated scenario. DOA estimates are shown as black dots. The CBF solution is shown in the background.

and zero order. If $\kappa > 0$, the VM PDF is symmetric around μ and has a similar shape as the Gaussian distribution. If $\kappa = 0$, the VM PDF is uniform, i.e., $f(\theta) = \frac{1}{2\pi}$. For large κ, the VM PDF can be approximate accurately by a Gaussian distribution with mean μ and $\sigma^2 = 1/\kappa$.

Likelihood Function: The measurements model (1) implies that given $\mathbf{w}$, the measurements $\mathbf{y}$ are statistically independent of $\mathbf{s}$. Furthermore, it is assumed that the components of the noise $\mathbf{u}_n$ in (1) are iid complex zero-mean Gaussian with variance ν and that the noise is statistically independent across time steps n. This results in the following expression for joint likelihood function

$$f(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{w}, \boldsymbol{s}) = f(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{w}) = \prod_{n'=1}^{n} f(\boldsymbol{y}_{n'}|\boldsymbol{\theta}_{n'}, \boldsymbol{w}_{n'}) \quad (4)$$

where $f(\boldsymbol{y}_{n'}|\boldsymbol{\theta}_{n'}, \boldsymbol{w}_{n'}) = f_{\mathrm{CN}}\big(\boldsymbol{y}_{n'}; \sum_{k=1}^{K} w_{k,n'}\, \boldsymbol{a}(\theta_{k,n'}), \nu \boldsymbol{I}\big)$.

Joint Posterior PDF: In the following derivation of an expression for the joint posterior PDF $f(\boldsymbol{\theta}, \boldsymbol{w}, \boldsymbol{s}|\boldsymbol{y})$, measurements $\boldsymbol{y}$ are observed and thus fixed. Then, using Bayes' rule as well as the chain rule for PDFs, we obtain

$$\begin{aligned} f(\boldsymbol{\theta}, \boldsymbol{w}, \boldsymbol{s}|\boldsymbol{y}) &\propto f(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{w}, \boldsymbol{s}) f(\boldsymbol{\theta}, \boldsymbol{w}, \boldsymbol{s}) \\ &= f(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{w}, \boldsymbol{s}) f(\boldsymbol{w}|\boldsymbol{s}, \boldsymbol{\theta}) f(\boldsymbol{s}, \boldsymbol{\theta}). \end{aligned}$$

Inserting (4) for $f(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{w}, \boldsymbol{s})$, (2) for $f(\boldsymbol{w}|\boldsymbol{s}, \boldsymbol{\theta})$, and (3) for $f(\boldsymbol{s}, \boldsymbol{\theta})$, yields the following factorization of the joint posterior pdf

$$\begin{aligned} f(\boldsymbol{\theta}, \boldsymbol{w}, \boldsymbol{s}|\boldsymbol{y}) \propto \Bigg(\prod_{k=1}^{K} f(w_{k,0}|s_{k,0}) p(s_{k,0}) f(\theta_{k,0}) \Bigg) \\ \times \prod_{n'=1}^{n} f(\boldsymbol{y}_{n'}|\boldsymbol{\theta}_{n'}, \boldsymbol{w}_{n'}) \prod_{l=1}^{K} f\big(w_{l,n'}\big|s_{l,n'}\big) \\ \times f\big(\theta_{l,n'}\big|\theta_{l,n'-1}\big) p\big(s_{l,n'}\big|s_{l,n'-1}\big). \quad (5) \end{aligned}$$

The corresponding factor graph [37] is shown in Fig. 1.

III. Problem Statement and Proposed Method

We consider the problem of Bayesian detecting and estimating the PAs $\mathsf{\theta}_{k,n}$, $k \in \{1, \ldots, \mathsf{K}_n\}$ from all the observations $\boldsymbol{y}$ available at time n. This relies on the marginal posterior activation PMFs $p(s_{k,n}|\boldsymbol{y})$ and the marginal posterior PDFs $f(\theta_{k,n}|\boldsymbol{y})$. A PA k is declared active if $p(s_{k,n} = 1|\boldsymbol{y}) > P_{\mathrm{th}}$, where $P_{\mathrm{th}} \in [0, 1)$ is a detection threshold [38, Ch. 2]. For PAs that are considered active, an estimate of $\mathsf{\theta}_{k,n}$ is produced by the minimum mean square error (MMSE) estimator [38, Ch. 4]

$$\hat{\theta}_{k,n} \triangleq \int \theta_{k,n} f(\theta_{k,n}|\boldsymbol{y})\, \mathrm{d}\theta_{k,n}. \quad (6)$$

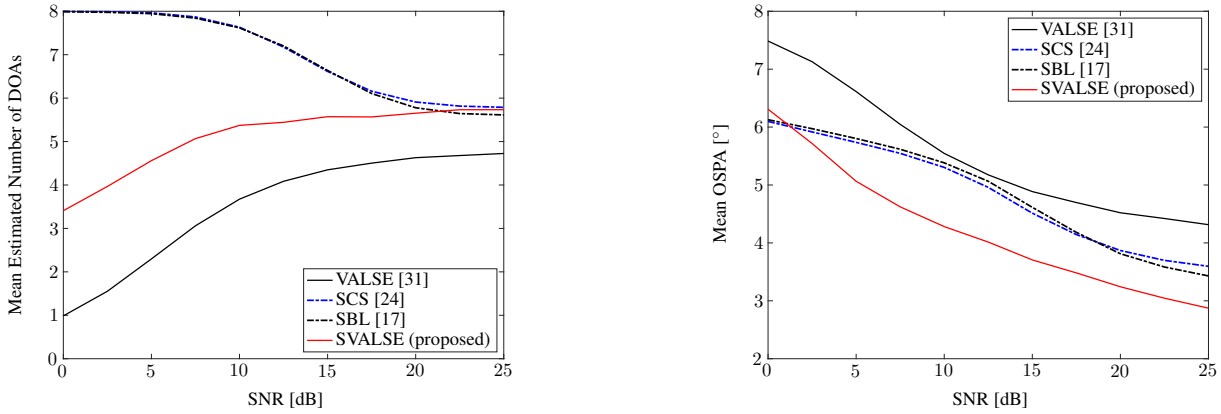
From estimated PAs $\hat{\theta}_{k,n}$, $k \in \{1, \ldots, \hat{K}_n\}$, we can directly get estimated DOAs as $\hat{\beta}_{k,n} = \sin^{-1}(\frac{c}{\omega d}\hat{\theta}_{k,n})$, $k \in \{1, \ldots, \hat{K}_n\}$.

Calculation of the PDFs $f(\theta_{k,n}|\boldsymbol{y})$ and the PMFs $p(s_{k,n}|\boldsymbol{y})$, needed for PA state detection and estimation, by direct marginalization from the joint posterior $f(\boldsymbol{\theta}, \boldsymbol{w}, \boldsymbol{s}|\boldsymbol{y})$ in (5) is infeasible. In our approach, for each $k \in \{1, \ldots, K\}$ and time step n, approximations $\tilde{f}(\theta_{k,n}|\boldsymbol{y})$ and $\tilde{p}(s_{k,n}|\boldsymbol{y})$ of the $f(\theta_{k,n}|\boldsymbol{y})$ and the $p(s_{k,n}|\boldsymbol{y})$ are calculated sequentially by performing a prediction and update step. For the update step, we consider a feasible approximate calculation by means of variational Bayesian estimation [39] by employing the single time step solution of the VALSE method [31]. Note that the VALSE method approximately represents marginal PDFs and PMFs, by VM PDFs and Kronecker delta functions, respectively, i.e., $\tilde{f}(\theta_{k,n}|\boldsymbol{y}) = f_{\mathrm{VM}}(\theta_{k,n}|\boldsymbol{y})$ and $\tilde{p}(s_{k,n}|\boldsymbol{y}) = \delta(s_{k,n} - \hat{s}_{k,n})$, $\hat{s}_{k,n} \in \mathcal{B}$.

Let us introduce the sequence of all measurements up to time $n-1$ as $\mathbf{y}_- \triangleq (\mathbf{y}_1, \ldots, \mathbf{y}_{n-1})$. At time step n, approximations $\tilde{f}(\theta_{k,n-1}|\boldsymbol{y}_-) = f_{\mathrm{VM}}(\theta_{k,n-1}|\boldsymbol{y}_-)$ and $\tilde{p}(s_{k,n-1}|\boldsymbol{y}_-) = \delta(s_{k,n-1} - \hat{s}_{k,n-1})$, $\hat{s}_{k,n-1} \in \mathcal{B}$ of $f(\theta_{k,n-1}|\boldsymbol{y}_-)$ and $p(s_{k,n-1}|\boldsymbol{y}_-)$, $k \in \{1, \ldots, K\}$ are available since they have been calculated at the previous time step $n-1$. First, a *prediction step* is performed for all $\mathsf{s}_{k,n}$ and $\mathsf{\theta}_{k,n}$, $k \in \{1, \ldots, K\}$. For prediction of PAs $\mathsf{\theta}_{k,n}$, we calculate

$$f_{\mathrm{VM}}(\theta_{k,n}|\boldsymbol{y}_-) \approx \int f_{\mathrm{VM}}(\theta_{k,n}|\theta_{k,n-1}) f_{\mathrm{VM}}(\theta_{k,n-1}|\boldsymbol{y}_-) \mathrm{d}\theta_{k,n-1} \quad (7)$$

by approximating both $f_{\mathrm{VM}}(\theta_{k,n}|\theta_{k,n-1})$ and $f_{\mathrm{VM}}(\theta_{k,n-1}|\boldsymbol{y}_-)$ by a Gaussian PDF, performing closed-form integration, and approximating the resulting Gaussian PDF by a VM PDF. Note that at time $n = 1$, $f_{\mathrm{VM}}(\theta_{k,n-1}|\boldsymbol{y}_-)$ in (7) is equal to the initial prior $f_{\mathrm{VM}}(\theta_{k,0})$.



(a) (b)

Fig. 3. Comparison of different DOA estimation methods: (a) Mean estimated number of DOAs versus SNR and (b) mean OSPA error versus SNR.

For prediction of the $\mathsf{s}_{k,n}$, we compute

$$\tilde{p}(s_{k,n}|\boldsymbol{y}_-) = \sum_{s_{k,n-1}\in\{0,1\}} p(s_{n,k}|s_{n-1,k})\,\tilde{p}(s_{k,n-1}|\boldsymbol{y}_-). \quad (8)$$

In particular, by using $p(s_{n,k}|s_{n-1,k})$ introduced in Section II and by exploiting $\tilde{p}(s_{k,n-1}|\boldsymbol{y}_-) = \delta(s_{k,n-1}-\hat{s}_{k,n-1})$, (8) can be expressed as

$$\tilde{p}(s_{k,n}|\boldsymbol{y}_-) = \begin{cases} p^{\mathrm{d}}\,\hat{s}_{k,n-1} + (1-p^{\mathrm{a}})(1-\hat{s}_{k,n-1}) & s_{k,n}=0 \\ p^{\mathrm{a}}(1-\hat{s}_{k,n-1}) + (1-p^{\mathrm{d}})\,\hat{s}_{k,n-1} & s_{k,n}=1. \end{cases}$$

Note that at time $n=1$, $\tilde{p}(s_{k,n-1}|\boldsymbol{y}_-)$ in (8) is replaced by the initial prior PDF $p(s_{k,0})$.

Next, approximations for predicted PDFs $f(\theta_{k,n}|\boldsymbol{y}_-)$ and predicted PMFs $p(s_{k,n}|\boldsymbol{y}_-)$, $k\in\{1,\ldots,K\}$ are used as prior information for the VALSE method [31], which is executed in the *update step* to obtain approximations of the PDFs $f(\theta_{k,n}|\boldsymbol{y})$ and PMFs $p(s_{k,n}|\boldsymbol{y})$ for PA detection and estimation according to (6).

IV. Simulation Results

We considered $L=15$ sensors and $K=15$ potential sources. The sensors form a uniform linear array with a sensor spacing of 7.5 m. The propagation speed of the medium is 1500 m/s. We simulated six sources ($K_n=6$ for all n) with initial angles $\boldsymbol{\beta}_0 = [-70^\circ\ -55^\circ\ -40^\circ\ 40^\circ\ 55^\circ\ 70^\circ]^{\mathrm{T}}$ that transmit at 100 Hz. The two "inner" sources $k\in\{3,4\}$ with initial positions $\beta_{3,0}=-40^\circ$ and $\beta_{4,0}=40^\circ$ move according to $\theta_{k,n}=\theta_{k,n-1}+v_{k,n}+\mathsf{u}_{k,n}$, where $v_{3,n}=0.05$ rad $\left(\sin^{-1}\frac{v_{3,n}c}{\omega d}=0.912^\circ\right)$, $v_{4,n}=-0.05$ rad, and $\mathsf{u}_{n,k}$ is an independent and identically distributed (iid) sequence of VM distributed random variables with zero mean and $\kappa_{\mathsf{u}}=40000$ $\left(\sigma_{\mathsf{u}}=0.005\text{ rad};\ \sin^{-1}\frac{\sigma_{\mathsf{u}}c}{\omega d}=0.091^\circ\right)$. We use the SNR definition in [15] and simulate 11 SNR values $(-5,-2.5,\ldots,10)$. Furthermore, we fix the variance of source amplitudes to $\tau=1$ and set the noise variance ν such that a specific SNR is obtained. We considered $n=80$ time steps and 100 simulation runs.

In the implementation of the proposed SVALSE method, we set $p^{\mathrm{a}}=0.15$ and $p^{\mathrm{d}}=0.2$. Furthermore, we use the same procedures for initialization and online learning of τ and ν as in VALSE [31]. For all PAs, we consider a state transition function $f_{\mathrm{VM}}(\theta_{k,n}|\theta_{k,n-1})$ that is given by the model $\theta_{k,n}=\theta_{k,n-1}+\mathsf{r}_{k,n}$, where $\mathsf{r}_{n,k}$ is an independent and identically distributed (iid) sequence of VM distributed random variables with zero mean and $\kappa_{\mathsf{r}}=156$ $\left(\sigma_{\mathsf{r}}=0.08\right.$ rad; $\left.\sin^{-1}\frac{\sigma_{\mathsf{r}}c}{\omega d}=1.459^\circ\right)$. As reference methods, we use CBF, SBL [17], sequential CS (SCS) [24], and VALSE [31]. In SBL and SCS we use the steering vectors corresponding to potential DOAs $\theta_{k,n}\in\{-90^\circ,-89^\circ,\ldots 89^\circ,90^\circ\}$, $k\in\mathcal{I}\triangleq\{0,\ldots,180\}$. After performing SBL or SCS processing, steering vectors with source amplitudes that are local maxima and above threshold $T\max_{k\in\mathcal{K}} w_{k,n}$ correspond to final DOA estimates. We set $T=0.1$ and $T=0.01$ for SBL and SCS, respectively, which led to favorable estimation errors. In SCS [24], we used the motion model in [40] with a finite lag of 8 and set $\alpha=0.05$ and $\lambda_0=0.3$. Nonsequential VALSE is completely parameters free. The performance of the various methods is measured by the Euclidean distance based OSPA [41] with cutoff parameter 8°. The OSPA performs optimal assignment of true to estimated DOAs and penalizes missing or excessive estimated DOAs by the cutoff parameter.

Figure 2 shows a single simulation run of SBL, CS, and the proposed variational method together with the CBF solution in the background. It can be seen that the sequential methods can more reliably localize DOAs near to endfire of the array, i.e., at -70°, -55°, 55°, and 70°. This can be explained by the fact that in sequential processing, DOA detection is supported by prior information from previous steps. The proposed SVALSE localizes DOAs near to endfire even more accurately compared to SCS. In addition, it outputs less spurious DOA estimates compared to SBL and SCS.

Figure 3(a) shows the mean estimated number of DOAs (averaged over time steps and simulation runs) versus SNR. It can be seen that for low SNR values, both VALSE and SVALSE underestimate, while SBL and SCS overestimate the

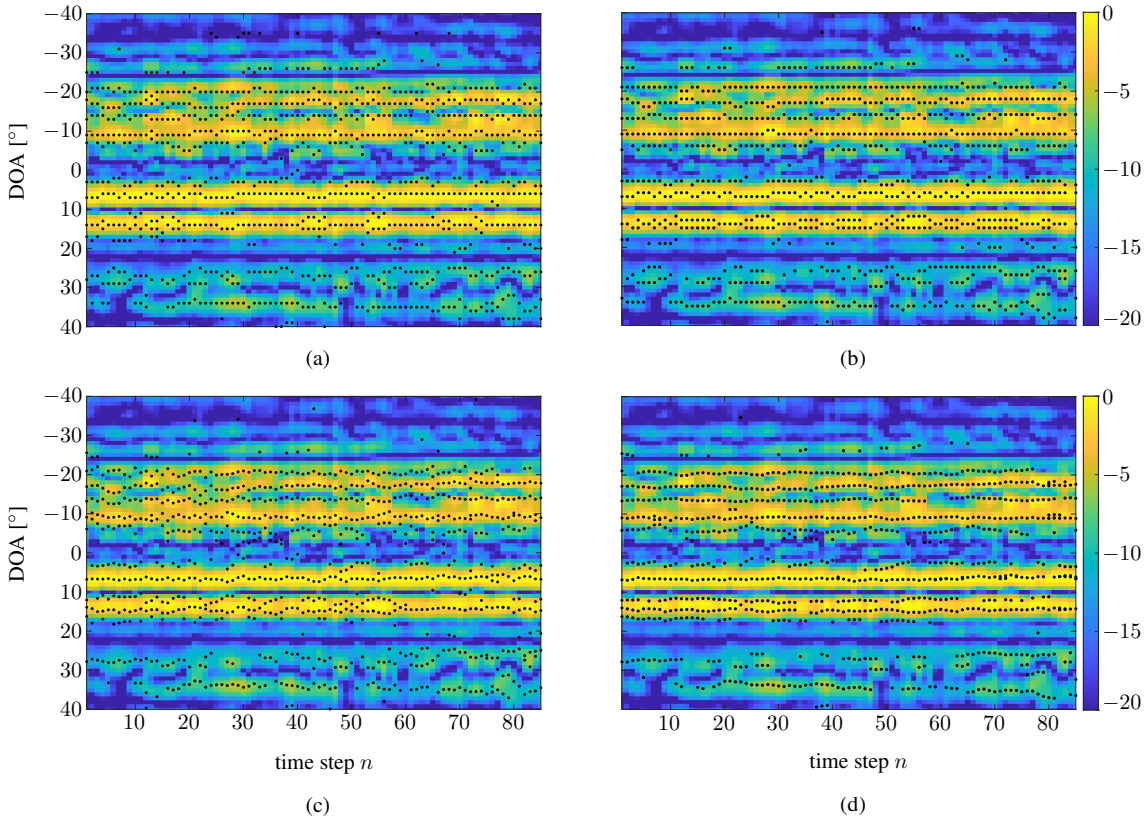


Fig. 4. Estimated DOAs versus time steps for (a) SBL, (b) SCS, and (c) VALSE, and (d) proposed SVALSE using acoustic data from the SWellEx-96 experiment. The CBF solution is shown in the background.

number of DOAs, respectively. At SNR values between 7.5 dB and 20 dB, SVALSE estimates the number DOAs more accurately compared to the all reference methods. Figure 3(b) shows the mean OSPA error (averaged over time steps and simulation runs) versus SNR. The following observations can be made. SBL and SCS perform similarly and outperform the VALSE. The large OSPA error of the VALSE is related to the fact that it strongly underestimates the number of DOAs. The proposed SVALSE outperforms the other methods for SNRs larger than 2.5 dB. At larger SNRs, this improved performance can be explained by the fact that SVALSE localizes DOAs near to endfire more accurately than the other methods. Note that the improved performance of the SVALSE comes at the cost of an increased runtime.

V. Experimental Results

We further validate the proposed SVALSE method and compare it with CBF, SBL, and SCS by using experimental data from a complex multi-path shallow-water environment. The considered dataset was collected during the shallow water evaluation cell experiment 1996 (SWellEx-96), which was performed West of Point Loma, CA [15], [42]. Acoustic data was recorded by a 64-element vertical linear array with a uniform inter-sensor spacing of 1.875 m that spanned water depths 94.125-212.25 m. Here, we are interested in the signal from a source towed at 54 m depth during it's closest point of approach (CPA) at a distance of 900 m and at 00:15 on 11 May 1996. The source sent a signal that consists of nine tones at frequencies $\{112, 130, 148, 166, 201, 235, 283, 338, 388\}$ Hz. The source level was approximately 158 dB re 1 μPa. We focus on the signal component at 201 Hz. The considered data has a duration of 1.5 min (covering 0.5 min before and 1 min after the CPA) and was sampled at 1500 Hz. Measurement vectors for 85 time steps are obtained by performing the discrete Fourier transforms with 2^{12} samples (2.7 s duration) and 63 % overlap. We used the same parameters as in the simulation results except that we set $K = 64$ and $\kappa_{\mathsf{r}} = 10000$ $\left(\sigma_{\mathsf{r}} = 0.01 \text{ rad}; \sin^{-1}\frac{\sigma_{\mathsf{r}} c}{\omega d} = 0.182^\circ\right)$ for the proposed method and $\lambda_0 = 0.2$ for SCS.

In Figure 4, we show the estimated DOAs versus time steps for the four considered methods. In particular, we show the 11 estimated DOAs with the highest amplitudes for SBL, SCS, VALSE, and SVALSE together with the CBF solution as an intensity plot in the background. Several multipath arrivals are present, which are almost stationary along all time steps. The CBF solution has significant peaks but is compromised by low resolution and spurious components due to sidelobes and noise. SBL produces many spurious estimates compared to the two sequential methods. Finally, it can be seen that due to it's grid-less nature, the proposed SVALSE can better adapt to small time variations of DOAs compared to SCS.

VI. Conclusion

We introduced a variational method for Bayesian estimation of time-varying directions of arrival that is grid-less, promotes sparse solutions, and provides marginal posterior PDFs of DOAs. Our numerical evaluation based on simulated measurements as well as acoustic data from an underwater localization experiment indicates a favorable performance compared to state of the art methods. Open work includes an investigation of the performance of VALSE and SVALSE in low SNR scenarios. A possible direction for future research is a combination of the considered approach with Bayesian methods for multiobject tracking [43].

VII. Acknowledgement

This work was supported by the University of California San Diego.

References

[1] M. I. Skolnik, *Introduction to Radar Systems*, 3rd ed. New York: Mcgraw-Hill, 2002.

[2] G. Ferri, A. Munafò, A. Tesei, P. Braca, F. Meyer, K. Pelekanakis, R. Petroccia, J. Alves, C. Strode, and K. LePage, "Cooperative robotic networks for underwater surveillance: an overview," *IET Radar Sonar Navig.*, vol. 11, no. 12, pp. 1740–1761, Dec. 2017.

[3] J. L. Semmlow, *Biosignal and Medical Image Processing*. Boca Raton, Fl: CRC Press, 2004.

[4] P. Goldstein and R. J. Archuleta, "Array analysis of seismic signals," *Geophys. Res.*, vol. 14, no. 1, pp. 13–16, 1987.

[5] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[6] H. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, Jul. 1996.

[7] H. L. Van Trees, *Optimum Array Processing*. New York, NY: Wiley, 2002.

[8] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ: Prentice-Hall, 2006.

[9] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. Roy. Statist. Soc. B*, vol. 58, no. 1, pp. 267–288, 1996.

[10] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, Mar. 1997.

[11] 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.

[12] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[13] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, 2008.

[14] A. Xenaki, P. Gerstoft, and K. Mosegaard, "Compressive beamforming," *J. Acoust. Soc. Am.*, vol. 136, no. 1, pp. 260–271, Jan. 2014.

[15] 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, Apr. 2015.

[16] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2182–2195, May 2011.

[17] 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.

[18] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.

[19] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.

[20] D. Shutin and B. H. Fleury, "Sparse variational Bayesian SAGE algorithm with application to the estimation of multipath wireless channels," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3609–3623, Aug. 2011.

[21] D. Shutin, S. R. Kulkarni, and H. V. Poor, "Incremental reformulated automatic relevance determination," *IEEE Trans. Signal Process.*, vol. 60, no. 9, Sep. 2012.

[22] D. Shutin, T. Buchgraber, S. R. Kulkarni, and H. V. Poor, "Fast variational sparse Bayesian learning with automatic relevance determination for superimposed signals," *IEEE Trans. Signal Process.*, vol. 59, no. 12, Dec. 2011.

[23] R. Giri and B. Rao, "Type I and Type II Bayesian methods for sparse signal recovery using scale mixtures," *IEEE Trans. Signal Process.*, vol. 64, no. 13, pp. 3418–3428, Jul. 2016.

[24] 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, Dec. 2013.

[25] A. S. Charles, A. Balavoine, and C. J. Rozell, "Dynamic filtering of time-varying sparse signals via ℓ_1 minimization," *IEEE Trans. Signal Process.*, vol. 64, no. 21, pp. 5644–5656, Nov. 2016.

[26] M. R. O' Shaughnessy, M. A. Davenport, and C. J. Rozell, "Sparse Bayesian learning with dynamic filtering for inference of time-varying sparse signals," *IEEE Trans. Signal Process.*, vol. 68, pp. 388–403, 2020.

[27] 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.

[28] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Comm. Pure Appl. Math.*, vol. 67, no. 6, pp. 906–956, 2014.

[29] Z. Yang and L. Xie, "On gridless sparse methods for line spectral estimation from complete and incomplete data," *IEEE Trans. Signal Process.*, vol. 63, no. 12, pp. 3139–3153, Jun. 2015.

[30] T. L. Hansen, B. H. Fleury, and B. D. Rao, "Superfast line spectral estimation," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2511–2526, May 2018.

[31] M.-A. Badiu, T. L. Hansen, and B. H. Fleury, "Variational Bayesian inference of line spectra," *IEEE Trans. Signal Process.*, vol. 65, no. 9, May 2017.

[32] J. Zhu, Q. Zhang, P. Gerstoft, M.-A. Badiu, and Z. Xu, "Grid-less variational Bayesian line spectral estimation with multiple measurement vectors," *Signal Processing*, vol. 161, pp. 155–164, 2019.

[33] 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, Jul. 2004.

[34] B. Ristic, M. S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Norwood, MA: Artech House, 2004.

[35] Y. Bar-Shalom, P. K. Willett, and X. Tian, *Tracking and Data Fusion: A Handbook of Algorithms*. Storrs, CT: Yaakov Bar-Shalom, 2011.

[36] K. V. Mardia and P. E. Jupp, *Directional Statistics*. Chichester, UK: Wiley, 2000.

[37] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[38] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York: Springer-Verlag, 1994.

[39] C. M. Bishop, *Pattern Recognition and Machine Learning*. Berlin, Heidelberg: Springer, 2006.

[40] A. Panahi and M. Viberg, "Fast LASSO based DOA tracking," in *Proc. IEEE CAMSAP-11*, San Juan, Puerto Rico, Dec. 2011, pp. 397–400.

[41] 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.

[42] G. L. D'Spain, J. J. Murray, W. S. Hodgkiss, N. O. Booth, and P. W. Schey, "Mirages in shallow water matched-field processing," *J. Acoust. Soc. Am.*, vol. 105, no. 6, pp. 3245–3265, 1999.

[43] 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.