Graph-Based Estimation of Time-Varying DOAs

Yongsung Park, Member, IEEE, Florian Meyer, Member, IEEE, and Peter Gerstoft, Senior Member, IEEE

Abstract—This paper presents a graph-based estimation method for sequential direction finding. The proposed method estimates an unknown number of directions of arrivals (DOAs) by performing message passing on the factor graph that represents the statistical model of the estimation problem. At each time step, belief propagation predicts the number of DOAs and their DOAs based on a new state-transition model and utilizing posterior probability density functions from previous time steps. Mean field message passing updates the DOAs and their number iteratively. The method promotes sparse solutions through a Bernoulli-Gaussian amplitude model, is gridless, and provides marginal posterior probability density functions from which DOA estimates and their uncertainties can be extracted. To propagate source existence and DOA information across time steps, a Bernoulli-von Mises state transition model is introduced. Compared to non-sequential approaches, the method can reduce DOA estimation errors in scenarios involving multiple time steps and time-varying DOAs. Simulation results demonstrate performance improvements compared to state-of-the-art methods. We evaluate the proposed method using ocean acoustic experimental data.

Index Terms—Graphical model, message passing, variational Bayesian inference, factor graph, array processing, direction of arrival (DOA) estimation, sparsity.

I. INTRODUCTION

DIRECTION of arrival (DOA) estimation or direction-finding [1] is the task of localizing several sources from noisy measurements provided by an array of sensors. DOA estimation is a key problem in electromagnetic, acoustic, and seismic applications. In this paper, we consider a scenario where multiple snapshots (or time steps) of data are available and the number of sources and their positions is time-varying. We propose an estimation method based on belief propagation (BP) and variational Bayesian inference (also known as the mean field (MF) approximation) that processes the data of dynamically varying DOAs sequentially. The proposed method is derived by applying message passing operations to the factor graph describing the inference problem’s statistical model. BP is used to predict the number of sources and their DOAs based on a new state-transition model and by utilizing the previous time step’s posterior probability density functions (pdfs). In addition, variational message passing is used to update DOAs and their number in an iterative update step. The resulting combined message passing method provides approximate marginal posterior pdfs that can be used to estimate the number of sources and their DOAs. A further advantage of using the framework of factor graphs is that it provides a graphical description of all algorithmic operations and shows the evolution of information over time.

A. State of the Art

In most traditional DOA estimation methods (see [1] and reference therein), a grid of potential DOAs and corresponding “steering vectors” is created to avoid nonlinear estimation. Of particular interest in recent years have been sparse DOA estimation methods that have high-resolution capabilities. For example, compressed sensing (CS) [2]–[4] approaches are based on a convex optimization procedure that explicitly promotes sparse solutions. Contrary to eigenvalue-based traditional DOA methods [1], CS for DOA estimation [5], [6] also performs well in scenarios with coherent sources [2].

Another recently introduced approach for DOA estimation is sparse Bayesian learning (SBL) [7]–[13]. In SBL, a hierarchical prior model controls the prior variance of steering vector amplitudes and implicitly promotes sparse solutions [7]–[13]. SBL-based methods can provide high DOA estimation accuracy but are known to overestimate the model order, i.e., they typically provide spurious low-power sources [9]. An overview and discussion of grid-based CS and SBL type methods and insight into their differences are provided in [14]. The main limitations of grid-based DOA estimation methods are (i) basis mismatch [15]–[18], which is observed when look directions of steering vectors do not appropriately represent DOAs; and (ii) basis coherence [5], [13], which is caused by a dense grid of steering vectors and results in biased estimates.

In many scenarios, multiple snapshots of measurements are available, and the estimation of time-varying DOAs is performed across multiple time steps. Here, sequential processing [19], [20], where information from previous times is used to compute DOA estimates at the current time, can improve the overall estimation performance. Sequential processing of sparse signals has been considered in [21]–[25], and sequential sparse DOA estimation has been developed based on CS [26] and SBL [27]–[29]. The main idea of these methods is to determine the parameters of sparsity-promoting prior pdfs using statistical information from previous times.

Among emerging gridless sparse methods suitable for direction finding [15], [16], [30]–[33], the variational Bayesian line spectral estimation (VALSE) [32]–[35] explicitly promotes sparsity by means of a Bernoulli-Gaussian amplitude model. It has favorable properties that it (i) is guaranteed to converge, (ii) incorporates prior information, (iii) provides posterior pdfs of DOAs, and (iv) is gridless. Contrary to our approach, all existing methods are non-sequential processing.

Y. Park, F. Meyer, and P. Gerstoft are with the Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92093 USA (e-mail: yongsungpark@ucsd.edu; fmeyer@ucsd.edu; gerstoft@ucsd.edu).

F. Meyer and P. Gerstoft are also with the Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, CA 92093 USA.

Manuscript received January XX, 2022; revised August XX, 2022. This paper was presented in part at the 23rd International Conference on Information Fusion (FUSION) in July 2020. This work was supported by the Office of Naval Research, Grant No. N00014-21-1-2267. (Corresponding Author: Yongsung Park.)
B. Contribution and Notation

We propose a message passing method for sparse and gridless estimation of time-varying DOAs. Our approach relies on a Bernoulli-Gaussian amplitude model of the VALSE [32]–[35] to promote sparsity. Sequential Bayesian estimation is enabled by combining the amplitude model with a Bernoulli-von Mises state-transition function. The resulting statistical model is described by a factor graph that provides the basis for developing the proposed message-passing method for sequential variational Bayesian line spectral estimation (SVALSE).

The proposed method combines variational message passing (also referred to as the MF approximation [36]–[41]) and BP [42]–[47]. We use the theoretical framework in [48] to run MF on certain parts of the factor graph and BP on the others. MF approximates the joint posterior pdf with a factorization of marginal posterior pdfs. The marginal posterior pdfs are obtained by iteratively minimizing the Kullback-Leibler (KL) divergence of the joint posterior pdf with respect to the marginal posterior pdfs [36]–[41]. Iterative minimization is guaranteed to converge and can be interpreted as passing messages along the edges of the factor graph that represents the statistical model [36]–[41]. BP aims at computing marginal posterior pdfs from the joint posterior pdf in an efficient and scalable way. It operates by iteratively passing messages along the edges of a factor graph [42]–[47]. In case the factor graph is a tree, i.e., has no cycles, BP can provide the true joint posterior pdfs [42]–[47]. While BP yields an accurate approximation of marginal pdfs if the factor graph has no short cycles, MF always admits a convergent implementation and leads to closed-form messages. BP cannot provide closed-form approximation of marginal pdfs if the factor graph has no short cycles, MF always admits a convergent implementation and can achieve improved performance through a probabilistic message computation. The resulting sequential VALSE messages for DOA update. Thus, we use MF, which exploits the exponential models [48]. The theoretical background for combining BP and MF on a single factor graph is in [48].

We consider a Bernoulli-von Mises state transition model [28] and a sparsity promoting Bernoulli-Gaussian amplitude model [32] for sequential estimation of time-varying DOAs. To perform estimation efficiently, we leverage the benefits of BP and MF. In particular, we use BP for DOA prediction, where the proposed Bernoulli-von Mises state transition model leads to closed-form messages. BP cannot provide closed-form messages for DOA update. Thus, we use MF, which exploits the conjugate-exponential form of the amplitude model for efficient message computation. The resulting sequential VALSE method is an instance of the combined MF and BP introduced in [48]. Compared to non-sequential methods, our approach can achieve improved performance through a probabilistic information transfer between time steps. Contrary to existing sequential methods, the used factor graph formulation makes it possible to represent algorithmic operations and information propagation over time visually.

The key contributions of this paper are as follows:

1) We establish a sparsity-promoting statistical model for sequential Bayesian estimation of time-varying DOAs and present the corresponding factor graph.
2) We derive a Bayesian estimation method that computes approximate marginal pdfs efficiently by performing MF and BP message passing on the factor graph.
3) We compare the performance of the proposed MF-BP message passing method with the state-of-the-art and demonstrate performance improvements.
4) We validated our method using real acoustic measurements collected during the SWellEx96 underwater source localization experiment.

This paper advances the conference paper [28] by providing a detailed derivation of the proposed message passing method, discussing implementation aspects including a new initialization, and presenting additional simulation and data processing results.

Notation: Random variables are displayed in sans serif, upright fonts (e.g., x) and their realizations in serif, italic (e.g., x). Vectors are in bold lowercase (e.g., random x and realization x) and matrices are in uppercase bold. Further, f(x) denotes the probability density function (pdf) of continuous random vector x (short for f_x(x)), p(s) denotes the probability mass function (pmf) of discrete random vector s (short for p_s(s)). The expectation operator with respect to pdf f(x) is given by E_{f(x)}[·] = ∫ f(x)dx, similarly for pmf p(s). E_{p(s)}[·] = ∑s p(s). Further, δ(x) is the Dirac delta function of continuous variable x and δ(s) is the Kronecker delta function of discrete variable s. The complex multivariate Gaussian f_{CN}(·; μ, Σ) has mean μ and covariance Σ. The identity matrix of dimension M is I_M.

II. System Model and Problem Formulation

A. Measurement Model

We observe narrowband signals from K_l sources with frequency ω on an array of M sensors. Let c and d = 2π/ω be propagation speed and sensor spacing, respectively. At time t ∈ {0, 1, . . .}, the received time-sampled signal y_t = [y_{0,t}, . . ., y_{M−1,t}]^T ∈ C^M consists of elements

\[ y_{m,t} = \sum_{k=1}^{K_l} \alpha_{k,t} e^{jm \sin \beta_{k,t}} + v_{m,t}, \]

where β_{k,t} ∈ [−90°, 90°) and α_{k,t} ∈ C are the angle of arrival and the complex amplitude, respectively, of the signal component originated by source k, and v_{m,t} ∈ C is additive noise. Note that at each time t, the complex amplitudes α_{k,t} and angles β_{k,t}, k ∈ {1, . . ., K_l} as well as the number K_l of components are unknown. Estimating the number of components K_l is typically referred to as model-order selection [49].

Due to the unknown number of components K_l, we follow the approach in [32] and consider at most L potential sources (L > K_l). Each potential source corresponds to a random angle of arrival and weight. We formulate the underlying problem as line spectral estimation by introducing pseudo angles (PAs) θ_{l,t} = e^{j\pi/2} sin β_{l,t} ∈ II ∆= [−\pi, \pi) and model the measurement vector y_t as

\[ y_t = \sum_{l=1}^{L} w_{l,t} a(\theta_{l,t}) + u_t, \]

where \(a(\theta_{l,t}) \triangleq [e^{j\theta_{l,t}} . . . e^{j(M−1)\theta_{l,t}}]^T\) is the steering vector and \(u_t \in C^M\) is the measurement noise. For future reference, we introduce the vector \(w_t \triangleq [w_{1,t} . . ., w_{L,t}]^T \in C^L\), the ordered sequences \(w \triangleq (w_1, . . ., w_t)\), and \(y \triangleq (y_1, . . ., y_t)\).
B. Von Mises Probability Density Function

The von Mises pdf (VM) [50, p. 36] of angle \( \theta \in [-\pi, \pi) \) is defined as
\[
f_{\text{VM}}(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\theta - \mu)},
\]  
(3)
where \( \mu \) and \( \kappa \) are mean direction and concentration, respectively, and \( I_0(\cdot) \) is the modified Bessel function of the first kind and order \( p \). Alternatively, the VM (3) can also be parametrized by \( \eta = \kappa e^{j\mu} \) and then reads
\[
f_{\text{VM}}(\theta; \eta) = \frac{1}{2\pi I_0(|\eta|)} \exp\{\text{Re}\{\eta^* e^{j\theta}\}\}.
\]  
(4)
If \( \kappa > 0 \), the VM is symmetric around \( \mu \) and has a similar shape as the Gaussian pdf. If \( \kappa = 0 \), the VM is uniform, i.e., \( f(\theta) = \frac{1}{2\pi} \). For large \( \kappa \), the VM can be approximated accurately by a Gaussian with mean \( \mu \) and variance \( \sigma^2 = 1/\kappa \).

An important property of VMs is that they are closed under multiplication, i.e.,
\[
f_{\text{VM}}(\theta; \eta_1)f_{\text{VM}}(\theta; \eta_2) \propto f_{\text{VM}}(\theta; \eta)
\]  
(5)
with \( \eta = \eta_1 + \eta_2 \). Thus, the resulting VM has mean direction \( \text{arg}(\eta_1 + \eta_2) \) and concentration \( |\eta_1 + \eta_2| \).

C. Prior PDFs

At each time \( t \), only \( K_t \) of the \( L \) components have nonzero weights. We use a sparsity-promoting prior for the complex weights \( w_{l,t} \) and \( w_{l,t} \) are governed by independent Bernoulli variables \( s_{l,t} \in B \triangleq [0, 1] \), i.e.,
\[
f(w_{l,t}|s_{l,t}) = (1 - s_{l,t})\delta(w_{l,t}) + s_{l,t}f_{\text{CN}}(w_{l,t};0, \tau),
\]  
(6)
\[
p(s_{l,t}) = \rho_{l,t}^{s_{l,t}}(1 - \rho_{l,t})^{1-s_{l,t}}.
\]  
(7)
Thus the binary variable \( s_{l,t} \) “deactivates” the \( l \)th component, i.e., \( s_{l,t} = 0 \) implies that \( w_{l,t} \) is not used. If \( s_{l,t} = 1 \), the \( l \)th component is active and \( w_{l,t} \) is zero-mean Gaussian with variance \( \tau \). We set the probability \( \rho_{l,t} = \ldots = \rho_{L,t} = \rho \) when we do not have prior information on \( \rho_{l,t} \). We also introduce vectors \( \bm{\theta} = [\theta_{1,t}, \ldots, \theta_{L,t}]^T \in \mathbb{R}^L \) and \( \bm{s} = [s_{1,t}, \ldots, s_{L,t}]^T \in \mathbb{B}^L \) as well as ordered sequences \( \bm{s} \triangleq (s_1, \ldots, s_L) \) and \( \bm{\theta} \triangleq (\theta_1, \ldots, \theta_L) \).

It is assumed that given \( \bm{s} \), then \( \bm{w} \) is statistically independent of \( \bm{\theta} \) and that the entries of \( \bm{w} \) are statistically independent across \( l \) and \( t \), i.e.,
\[
f(\bm{w}|\bm{s}, \bm{\theta}) = f(\bm{w}|\bm{s}) = \prod_{t'=1}^T \prod_{l=1}^L f(w_{l,t'}|s_{l,t'})
\]  
(8)
For future reference, we introduce the set formed by the indexes of non-zero entries in \( \bm{s} \), i.e.,
\[
S = \{1 \leq l \leq L | s_l = 1 \}.
\]  
(9)
The PAs \( \bm{\theta} \) and Bernoulli variables \( \bm{s} \) are assumed to evolve independently and according to a first-order Markov model [19]. Further, at time \( t = 0 \), they are assumed statistically independent across \( l \) with prior pdf \( f(\theta_0, s_0) = \ldots \).

D. Likelihood Function and Joint Posterior PDF

The measurement model (2) implies that given \( \bm{w} \), the measurements \( \bm{y} \) are statistically independent of \( \bm{s} \). Further, the components of the noise \( \bm{u} \) (2) are assumed iid complex zero-mean Gaussian with variance \( \nu \) and statistically independent
where we specialize it to \( \theta \).

The factor for mating the DOAs as the blueprint for the development of a message passing posterior pdf is

\[
\ln f(y|\theta, w, s) = \prod_{t'=1}^{t} f(y_{t'}|\theta_{t'}, w_{t'})
\]

\[
f(y_{t'}|\theta_{t'}, w_{t'}) = f_{\text{CN}} \left( y_{t'}; \sum_{l=1}^{L} w_{l,t'} a_{l}(\theta_{l,t'}), \nu I_M \right).
\]

Joint pdfs: Using the chain rule, we obtain the joint pdf as

\[
f(\theta, w, s, y) = f(y|\theta, w, s)f(\theta, w, s)
\]

\[
= f(y|\theta, w, s)f(w|s, \theta)f(s, \theta).
\]

Inserting (12) for \( f(y|\theta, w, s) \), (8) for \( f(w|s, \theta) \), and (10) for \( f(s, \theta) \), yields the following factorization of the joint pdf

\[
f(\theta, w, s, y) = \prod_{l=1}^{t} f(w_{l,1}|s_{l,1}) f_{\text{VM}}(\theta_{l,1}) p(s_{l,1})
\]

\[
\times \prod_{t=2}^{t} f(y_{t'}|\theta_{t'}, w_{t'}) \prod_{l=1}^{L} f(w_{l,t'}|s_{l,t'})
\]

\[
\times f(\theta_{l,t'}, \theta_{l,t'-1}) p(s_{l,t'}|s_{l,t'-1}).
\]

Since the measurement \( y \) are observed and thus fixed, the joint posterior pdf is \( f(\theta, w, s|y) \propto f(\theta, w, s, y) \).

In Fig. 1, we illustrate the factor graph [42] for the joint posterior pdf (15) in short notation, which is given by

\[
f(\theta, w, s, y) = f(y_{t=1}|\theta_{0,\theta}) \prod_{l=1}^{L} f(w_{l,t}|s_{l,t}) f_{\theta_{l,t}}(\theta_{l,t}) \prod_{l=1}^{L} f(w_{l,t}|s_{l,t}) f_{\theta_{l,t}}(\theta_{l,t})
\]

\[
\times \prod_{t=2}^{t} f(y_{t'}|\theta_{t', w}) \prod_{l=1}^{L} f(w_{l,t'}|s_{l,t'}) f_{\theta_{l,t'}}(\theta_{l,t'}) \prod_{l=1}^{L} f(w_{l,t'}|s_{l,t'}) f_{\theta_{l,t'}}(\theta_{l,t'})
\]

\[
\times f(\theta_{l,t'}, \theta_{l,t'-1}) p(s_{l,t'}|s_{l,t'-1}).
\]

where we specialize it to \( t = 2 \) time steps as in Fig. 1,

\[
f(\theta, w, s, y) = f_{y_{t=1}} \prod_{l=1}^{L} f_{w_{l,2}}(w_{l,2})
\]

\[
\times \prod_{l=1}^{L} f_{\theta_{l,1}}(\theta_{l,1}) f_{\theta_{l,2}}(\theta_{l,2}) p(s_{l,2})
\]

\[
\times \prod_{l=1}^{L} f_{\theta_{l,2}}(\theta_{l,2}) f_{\theta_{l,1}}(\theta_{l,1}) p(s_{l,1})
\]

The factor for \( t = 1 \) correspond the joint posterior pdf of nonsequential VALUE [32, 33]. The factor graph in Fig. 1 serves as the blueprint for the development of a message passing algorithm [42, 48], [51] in Secs. III and V.

E. Problem Formulation

We consider the problem of the DOAs \( \theta_{k,l}, k \in \{1, \ldots, K_l\} \) among the PAs \( \theta_{i,l} \), \( l \in \{1, \ldots, L\} \) \((L > K_l)\) from all the observations \( y \) over time \( t \). This relies on the marginal posterior activation probability mass function (pmf) \( p(s_{l,t}|y) \) and the marginal posterior pdfs \( f(\theta_{l,t}|y) \). A PA \( \theta_{l,t} \) is declared active if \( p(s_{l,t} = 1|y) > P_{th} \), where \( P_{th} \in [0, 1] \) is a detection threshold [52, Ch. 2]. For PAs that are considered active, the PAs \( \theta_{l,t} \) are estimated by the MMSE estimator [52, Ch. 4] for angles

\[
\hat{\theta}_{l,t} = \arg \left( \mathbb{E}_{f(\theta_{l,t}|y)}[e^{j \theta_{l,t}}] \right).
\]

From the PAs \( \hat{\theta}_{l,t} \), the estimated DOAs are \( \hat{\beta}_{l,t} = \sin^{-1}(\frac{\omega}{2} \hat{\theta}_{l,t}) \).

Calculation of the pdfs \( f(\theta_{l,t}|y) \) and the pmfs \( p(s_{l,t}|y) \), needed for PA state detection and estimation, by direct marginalization from the joint posterior \( f(\theta, w, s|y) \) in (15) is infeasible. In our approach, for each \( l \in \{1, \ldots, L\} \) at time \( t \), approximations \( q(\theta_{l,t}|y) \) and \( q(s_{l,t}|y) \) of the marginal pdfs \( f(\theta_{l,t}|y) \) and \( p(s_{l,t}|y) \) are calculated sequentially by a prediction and update step. For the update, we consider a feasible approximate calculation by means of variational Bayesian estimation [36] based on VALUE [32]. Within this estimation approach, the resulting approximate marginal activation pmfs \( q(s_{l,t}|y) \) have the form \( q(s_{l,t}|y) = \delta(s_{l,t} - \hat{s}_{l,t}) \). Thus, the thresholding based on \( P_{th} \) is avoided, i.e., PA \( \theta_{l,t} \) is active at time \( t \) if \( \hat{s}_{l,t} = 1 \).

III. VARIATIONAL BAYESIAN APPROXIMATION

We here review the variational Bayesian approach [36, Ch. 10], [37, Ch. 21], [38, Ch. 10] based on the system model in Sec. II. We focus on one time step, i.e., we use the joint posterior pdf (15) at \( t = 1 \). For simplicity, the time index \( t \) is omitted. The joint pdf now reads,

\[
f(\theta, w, s, y) \propto f(y|\theta, w) \prod_{l=1}^{L} f(w_{l}|s_{l}) f_{\text{VM}}(\theta_{l}) p(s_{l}).
\]

A. Variational Bayes

The variational Bayesian approach aims to approximate \( f(\Phi|y) \) by a simpler pdf \( q(\Phi|y) \) that is in a family of tractable pdfs. The parameters are collectively denoted by \( \Phi = \{\theta, w, s\} \). Approximation is performed by minimizing the Kullback-Leibler (KL) divergence [36, p. 463], [37, p. 732], [53, p. 205] i.e.,

\[
\mathbb{KL}[q(\Phi|y)||f(\Phi|y)] = -\mathbb{E}_{q(\Phi|y)} \left[ \ln \frac{f(\Phi|y)}{q(\Phi|y)} \right].
\]

Note that the KL divergence is nonnegative [53, Ch. 6.2.4]. Using \( f(\Phi|y) = f(\Phi|y)/f(y) \) in (20), the log model evidence \( \ln f(y) \) is given by [36, Eq. (10.2)]

\[
\ln f(y) = \mathbb{KL}[q(\Phi|y)||f(\Phi|y)] + \mathcal{L}[q(\Phi|y)]
\]

\[
\mathcal{L}[q(\Phi|y)] \triangleq \mathbb{E}_{q(\Phi|y)} \left[ \ln f(\Phi|y) q(\Phi|y) \right].
\]

Since \( \ln f(y) \geq \mathcal{L}[q(\Phi|y)] \), due to the nonnegativity of the KL-divergence [53, Ch. 6.2.4], \( L \) is the evidence lower bound (ELBO) [38, Ch. 10.1.2], [54], [55]. For observed and thus fixed measurements \( y \in \mathbb{C}^{M} \), \( \ln f(y) \) is a constant \( 21 \). Minimizing the KL divergence (20) is thus equivalent to maximizing the ELBO (22).

For the development of the VALUE method [32], the ELBO \( \mathcal{L}[q(\Phi|y)] \) in (22) is maximized by assuming that \( q(\Phi|y) \) is in the following family of tractable pdfs

\[
q(\Phi|y) = q(\theta, w, s|y) = q(w, s|y) \prod_{l=1}^{L} q(\theta_{l}|y),
\]

where the factor \( q(w, s|y) = q(w|s, y) q(s|y) \) is further constrained by setting \( q(s|y) = \delta(s - s) \), i.e., the posterior pmf \( q(s|y) \) of the binary vector \( s \in \mathbb{B}^{L} \) has all its mass at
a single vector \( \hat{s} \in \mathcal{B}^{L} \). Thus, the final expression for the considered family of tractable pdfs reads

\[
q(\theta, w, s | y) = q(w | y, s) \delta(s - \hat{s}) \prod_{l=1}^{L} q(\theta_l | y). \tag{24}
\]

For future reference, \( \hat{S} \) is the set of indices of the non-zero entries in \( \hat{s} \) and, due to \( q(s | y) = \delta(s - \hat{s}) \), the equalities hold,

\[
q(w | y) = \sum_s q(w, s | y) = \sum_s q(w | y, s) \delta(s - \hat{s}) = q(w | y, \hat{s}). \tag{25}
\]

IV. VARIATIONAL LINE SPECTRAL ESTIMATION

A message passing algorithm for variational line spectral estimation is presented with closed-form expressions for the resulting message and approximate marginal pdfs.

Following the variational Bayesian estimation approach [36, Ch. 10.1.1], [37, Ch. 21.3], [38, Ch. 10.2] an alternating optimization is performed, since maximizing the ELBO \( \mathcal{L}[q(\theta, w, s | y)] \) for all approximate marginal pdfs \( q(\theta_l | y) \), \( l = 1, \ldots, L \) and \( q(w, s | y) \) simultaneously is infeasible. The ELBO \( \mathcal{L}[q(\theta, w, s | y)] \) is maximized in turns, over each of the approximate marginal pdfs \( q(w | y, s) \) and \( q(\theta_l | y) \), while the others are kept fixed. After initialization \( (p = 0) \), at each iteration \( p \in \{1, \ldots, P \} \) variational Bayesian estimation cycles through these approximate marginal pdfs and replaces them one by one with updated versions. Such an alternating optimization approach is guaranteed to converge to a local maximum of the ELBO \( \mathcal{L}[q(\theta, w, s | y)] \) [36, Ch. 10.1].

A. MF Message Passing

Let \( q^{(p-1)}(\theta_l | y) \) and \( q^{(p-1)}(w, s | y) \) be the approximate marginal pdfs updated at the previous iteration \( p - 1 \). At iteration \( p \), first the updated approximate marginal pdf \( q^{(p)}(\theta_l | y) \), \( l = \{1, \ldots, L\} \) are obtained as [36, Eq. (10.9)], [37, Eq. (21.25)]

\[
q^{(p)}(\theta_l | y) \propto \exp \left( \mathbb{E}_{q^{(p-1)}(\theta_l | y)} \left[ \ln f(y, \theta, w, s) \right] \right), \tag{26}
\]

where \( \mathbb{E}_{q^{(p-1)}(\theta_l | y)}[\cdot] \) is the expectation with respect to \( q^{(p)}(w, s | y) (\prod_{l'=1}^{L} q^{(p)}(\theta_{l'} | y)) (\prod_{l'=l+1}^{L} q^{(p-1)}(\theta_{l'} | y)) \). Eq. (26) says that the optimal factor \( q(\theta_l | y) \) is obtained considering the joint pdf over all variables \( f(y, \theta, w, s) \) (19) and then taking the expectation with respect to all of the other factors except \( q(\theta_l | y), \) i.e., \( q(w, s | y) \) and \( \{q(\theta_{l'} | y) \} \) for \( l' \neq l \) (23). Plugging \( f(y, \theta, w, s) \) in (19) into (26), we obtain

\[
q^{(p)}(\theta_l | y) \propto f_{VM}(\theta_l) m_l^{(p-1)}(\theta_l), \tag{27}
\]

where we introduce the messages

\[
m_l^{(p-1)}(\theta_l) \triangleq \exp \left( \mathbb{E}_{q^{(p-1)}(\theta_{l'} | y)} \left[ \ln f(y, \theta, w) \right] \right), \tag{28}
\]

where \( f(y, \theta, w) \) is the likelihood (12). At initialization \( (p = 0) \), (27) is initialized as \( q^{(0)}(\theta_l | y) \propto f_{VM}(\theta_l) \), which is obtained using (66).

Similarly as in (26), the approximate marginal pdfs \( q^{(p)}(w, s | y) \) are updated, i.e.,

\[
q^{(p)}(w, s | y) \propto \exp \left( \mathbb{E}_{q^{(p-1)}(w, s | y)} \left[ \ln f(y, \theta, w, s) \right] \right), \tag{29}
\]

Fig. 2: MF message passing procedure for variational inference. A portion of the factor graph in Fig. 1. The box (plate) denotes a set of nodes, and the connections are duplicated \( L \) times.

where \( \mathbb{E}_{q^{(p-1)}(w, s | y)}[\cdot] \) is the expectation with respect to \( q^{(p-1)}(\theta_l | y) \). Plugging \( f(y, \theta, w, s) \) in (19) into (29), we obtain

\[
q^{(p)}(w, s | y) \propto f(w | s) p(s) m_l^{(p-1)}(w), \tag{30}
\]

where \( f(w | s) \triangleq \prod_{l=1}^{L} f(w_l | s_l) \) (6), \( p(s) \triangleq \prod_{l=1}^{L} p(s_l) \), and the messages

\[
m_l^{(p-1)}(w) \triangleq \exp \left( \mathbb{E}_{q^{(p-1)}(w, s | y)} \left[ \ln f(y, \theta, w) \right] \right). \tag{31}
\]

At initialization \((p = 0)\), (30) is initialized with \( q^{(0)}(w, s | y) \) which is obtained using (51) based on \( q^{(0)}(\theta_l | y) \) (66). Eqs. (27) and (30) introduce a MF message passing algorithm [51]. MF messages passed along the edges of the factor graph in Figs. 1 and 2, shown in blue.

B. Computing \( q^{(p)}(\theta_l | y) \)

Let \( \hat{s} \) be the activation vector from previous iteration \( p - 1 \) and \( \hat{S} \) be the set of corresponding non-zeros entries in \( \hat{s} \), see (57). Computing the expectation \( \mathbb{E}_{q^{(p-1)}(\theta_l | y)}[\cdot] \) in (26), for \( l \in \hat{S} \), gives [32, Sec. III-A]

\[
q^{(p)}(\theta_l | y) \propto f_{VM}(\theta_l) \exp \{ \text{Re} (\eta_l^H a(\theta_l)) \}, \tag{32}
\]

where \( \eta_l \in \mathbb{C}^M \) is given by

\[
\eta_l = \frac{2}{\nu} \left( y - \sum_{l' \in \hat{S} \setminus \{l\}} \hat{w}_{l'} \hat{a}_{l'} \right) \hat{w}_{l}^H - \frac{2}{\nu} \sum_{l' \in \hat{S} \setminus \{l\}} \hat{C}_{l'l'} \hat{a}_{l'} . \tag{33}
\]

Here, steering vectors \( \hat{a}_{l'} \) are estimated from the approximate marginal pdf \( q^{(p)}(\theta_l | y) \) using the characteristic function [50, Ch. 3.3] of the VM pdf [50, Ch. 3.5.4],

\[
\hat{a}_{l'} \triangleq \begin{cases} 
E_{q^{(p)}(\theta_l, y)}[a(\theta_l)], & l' \in S \cap \{1, \ldots, l - 1\}, \\
E_{q^{(p-1)}(\theta_{l'}, y)}[a(\theta_{l'})], & l' \in S \cap \{l + 1, \ldots, L\}. 
\end{cases} \tag{34}
\]

Further, for \( l \in \hat{S} \) estimates of weights \( \hat{w}_l \) and corresponding variance \( \hat{C}_{l'l'} \) are computed from the posterior pdf \( q^{(p-1)}(w_l | y) \) in (25), as follows

\[
\hat{w}_l = E_{q^{(p-1)}(w_l | y)}[w_l], \tag{35}
\]

\[
\hat{C}_{l'l'} = E_{q^{(p-1)}(w_{l'} | y)}[w_l w_{l'}^H] - \hat{w}_l\hat{w}_l^H, \quad l' \in \hat{S}. \tag{36}
\]

For an efficient iterative optimization, \( E_{q^{(p)}(\eta_l | \theta_l, \eta_{l-1})}[a(\theta_l)] \) (34) is calculated in closed form. A mixture of VMs can approximate the marginal pdfs \( q(\theta_l | y) \) accurately and enable a closed form calculation [32, Sec. IV], as reviewed below. We use \( \eta_l = [\eta_0, l \ldots \eta_{M-1}, l] \). From (32), we obtain, cf. (4),

\[
q^{(p)}(\theta_l | y) \propto f_{VM}(\theta_l; \eta_{l-1}) \prod_{m=0}^{M-1} \exp \{ \text{Re} (\eta_{l+m}^H e^{j \theta_l}) \}.
\]
where we introduced the parameter $\eta_{m,l}$ of the prior VM pdf. The predicted VM based on the posterior pdfs of the previous time (62) is used as the prior VM. The $f_{\text{VM}}(m \theta_i; \eta_{m,l})$ is referred to as $m$-fold wrapped VM (37) with parameter $\eta_{m,l} = \kappa_{m,l} e^{\mu_{m,l}}$. The factor corresponding to $m = 0$ is a constant and thus omitted in (37).

An $m$-fold wrapped VM has $m$ modes and is approximated accurately by a mixture of $m$ VM [32, Eq. (30)], i.e.,

$$ f_{\text{VM}}(m \theta_i; \eta_{m,l}) \approx \frac{1}{m} \sum_{r=1}^{m} f_{\text{VM}}(\theta_i; \tilde{\eta}_{(m,r),l}) $$

with $\tilde{\eta}_{(m,r),l} = \tilde{\kappa}_{m,l} e^{\tilde{\mu}_{(m,r),l}}$. The $(m, r)$ component corresponds to the mode $r$ of the $m$-fold wrapped VM. The $m$ components of the mixture have equal concentrations but different directions. Matching the characteristic function of the wrapped VM and the mixture of VMs, the single concentration $\tilde{\kappa}_{m,l}$ is the solution of [32, Eq. (31)]

$$ I_m(\tilde{\kappa}_{m,l}) = I_0(\kappa_{m,l}) $$

and the means $\tilde{\mu}_{(m,r),l}, r = 1, \ldots, m$, are given by

$$ \tilde{\mu}_{(m,r),l} = \mu_{m,l} + 2\pi(r - 1) \frac{m}{m} $$

Plugging (38) into (37) and using that the VM is closed under multiplication (5), the approximate marginal pdf $q^{(p)}(\theta_i | y)$ reads [32, Eq. (33)]

$$ q^{(p)}(\theta_i | y) = \frac{1}{Z_{\theta}} \exp\{\Re(\xi_{\theta,i}(e^{\tilde{\theta}_i}))\} $$

$$ Z_{\theta} = 2\pi \sum_{r \in \mathcal{R}} I_0(|\xi_{r,i}|) $$

$$ \xi_{r,i} = \eta_{h_0} + \tilde{\eta}_{(1,r),i} + \ldots + \tilde{\eta}_{(M-1,rM-1),i} $$

where indexes $r$ are from the set $\mathcal{R}$, $r = (r_1, \ldots, r_{M-1}) \in \mathcal{R} \equiv \{1\} \times \{1, 2\} \times \ldots \times \{1, \ldots, M - 1\}$. For example, for $M = 4$ we have $\mathcal{R} = \{(1, 1, 1), (1, 1, 2), (1, 1, 3), (1, 2, 1), (1, 2, 3)\}$. When no prior information on $\theta$ is available, e.g., $t = 1$, we set $\eta_{h_0} = 0$.

Note that $q^{(p)}(\theta_i | y)$ has the form of a mixture of VMs,

$$ q^{(p)}(\theta_i | y) = \sum_{r \in \mathcal{R}} C_r f_{\text{VM}}(\theta_i; \xi_{r,i}) $$

with $C_r = 2\pi I_0(|\xi_{r,i}|) / Z_0$. Performing the variational Bayes optimization with all $|\mathcal{R}|$ components is intractable. Thus, a suboptimum pruning stage that keeps $D \ll |\mathcal{R}|$ most dominant components is employed,

$$ q^{(p)}(\theta_i | y) = \sum_{d=1}^{D} \tilde{C}_d f_{\text{VM}}(\theta_i; \tilde{\xi}_{d,i}) $$

where $\tilde{C}_d = 2\pi I_0(|\tilde{\xi}_{d,i}|) / \tilde{Z}_d$ and $\tilde{Z}_d = 2\pi \sum_{d=1}^{D} I_0(|\tilde{\xi}_{d,i}|)$. For each $d$, let $\tilde{a}_d$ be the subvector and submatrix of $\tilde{h}$ and $\tilde{J}$, that only consist of elements corresponding to active indexes in $s$. The approximate marginal pdf $q^{(p)}(w, s | y)$ in (30) is expressed as [32, Sec. III-B], [33, Eq. (22)]

$$ q^{(p)}(w, s | y) \propto \exp\{-|w_s - \tilde{w}_s|^2 / \tilde{C}_s \} $$

$$ \tilde{w}_s = \nu^{-1} \tilde{C}_s \tilde{h}_s $$

where $\nu = \sum_{d=1}^{D} \tilde{C}_d / \tilde{Z}_d$. For each $d$, let $\tilde{a}_d$ be the subvector and submatrix of $\tilde{h}$ and $\tilde{J}$, that only consist of elements corresponding to active indexes in $s$. The approximate marginal pdf $q^{(p)}(w, s | y)$ in (30) is expressed as [32, Sec. III-B], [33, Eq. (22)]

$$ q^{(p)}(w, s | y) \propto \exp\{-|w_s - \tilde{w}_s|^2 / \tilde{C}_s \} $$

where $\tilde{w}_s = \nu^{-1} \tilde{C}_s \tilde{h}_s$. Using $q^{(p)}(w, s | y)$ in (51) evaluated at the single vector $\tilde{s} \in \mathcal{B}^d$, we obtain the approximate marginal pdf

$$ q^{(p)}(w | y) \propto q^{(p)}(w, \tilde{s} | y) $$

Using (51) in (54), the final expression for the approximate marginal pdf $q^{(p)}(w | y)$ becomes

$$ q^{(p)}(w | y) = f_{\text{CS}}(w_s; \tilde{w}_s, \tilde{C}_s) \prod_{l \not\in \tilde{s}} \delta(w_l) $$

which is used for calculating $q^{(p)}(\theta_i | y)$, see (35) and (36).

Next, we discuss the calculation of the “best” activation $\tilde{s}$ that defines $\tilde{S}$ on the right hand side of (55). For the calculation of this activation $\tilde{s}$, the ELBO objective function is introduced

$$ O(s) = \mathcal{L}[q^{(p)}(\theta, w, s | y)] $$

$$ = \mathbb{E}_{q^{(p)}(\theta, w, s | y)} \left[ \ln \frac{f(y; \theta, w, s)}{q^{(p)}(\theta, w, s | y)} \right] $$

where $q^{(p)}(\theta, w, s | y)$ is from (24) replacing $q(\theta_i | y)$ and $q(w | y, s)$ by $q^{(p)}(\theta_i | y)$ and $q^{(p)}(w | y, s)$, respectively. The
The dynamic DOAs are modeled by a state transition model,
\[ \theta_{t,t} = \theta_{t,t-1} + r_{t,t}, \]  
(61)
where \( r_{t,t} \) is an independent and identically distributed (iid) sequence of von Mises distributed random variables with zero mean and \( \kappa_r \). For example, \( \kappa_r = 148 \) provides standard deviation of 1.499° for a half-wavelength spacing, i.e., \( \sigma_r = 0.0822 \) rad; \( \sigma_r = 1.499^\circ \). From the state transition model (61), we obtain the transition pdf \( f(\theta_{t,t}\vert\theta_{t-1}) = f_{\text{VM}}(\theta_{t,t}\vert\theta_{t-1};0,\kappa_r) \) (60).

The VMs \( f_{\text{VM}}(\theta_{t,t}\vert\theta_{t-1}) \) and \( f_{\text{VM}}(\theta_{t,t-1}) \) typically have large \( \kappa_r \), which is analogous to Gaussians having a small variance \( \sigma^2 = 1/\kappa_r \), i.e., these VMs are concentrated about their mean. An approximation of (60) is obtained as, cf. (46),
\[ q(\theta_{t,t}\vert y_{-}) = f_{\text{VM}}(\theta_{t,t};\tilde{\eta}_{t,t}'), \]  
(62)
where \( \tilde{\eta}_{t,t}' \) is obtained by approximating \( f_{\text{VM}}(\theta_{t,t}\vert\theta_{t-1}) \) by a zero-mean Gaussian with variance \( 1/\kappa_r \) and \( f_{\text{VM}}(\theta_{t-1}\vert\tilde{\eta}_{t-1}) \) by a \( \tilde{\eta}_{t-1} \)-mean Gaussian with variance \( 1/\tilde{\kappa}_{t-1} \). Performing closed-form integration gives a \( \tilde{\eta}_{t-1} \)-mean Gaussian with variance \( 1/\kappa_r + 1/\tilde{\kappa}_{t-1} \) using (61). Approximating the Gaussian by a VM, \( \tilde{\eta}_{t,t}' \) becomes
\[ \tilde{\eta}_{t,t}' = (1/\kappa_r + 1/\tilde{\kappa}_{t-1})^{-1}e^{\tilde{\eta}_{t-1}}. \]  
(63)
For prediction of \( s_{t,t} \), we compute
\[ q(s_{t,t}\vert y_{-}) = \sum_{s_{t,t}\in\{0,1\}} p(s_{t,t}\vert s_{t-1}) q(s_{t-1}\vert y_{-}) \]  
(64)
Using the transition pmf \( p(s_{t,t}\vert s_{t-1}) \) (11) and exploiting \( q(s_{t-1}\vert y_{-}) = \delta(s_{t-1} - \tilde{\eta}_{t-1}) \), (64) becomes
\[ q(s_{t,t}\vert y_{-}) = \begin{cases} p^s(1 - p^a)(1 - p^d) & s_{t,t} = 0, \\ p^a(1 - p^d) & s_{t,t} = 1. \end{cases} \]  
(65)
At time \( t = 1 \), \( q(\theta_{1,t-1}\vert y_{-}) \) in (60) and \( q(s_{1,t-1}\vert y_{-}) \) in (64) are replaced by the initial priors \( f_{\text{VM}}(\theta_{0}) \) and \( p(s_{0}) \), respectively. Note that according to (65), we have \( p_{t,t} = p^a \) if \( l \in \{1,\ldots,|\tilde{S}_{t-1}|\} \) and \( 1 - p^d \) otherwise, cf. (6). At time \( t = 1 \), we set \( p_{t,t} = \ldots = p_{L,t} = \rho \).

The approximate marginal pdfs and pmfs can be interpreted as BP messages passed along the edges of the factor graph in Figs. 1 and 3, shown in green.

**B. Update**

The prediction step is followed by an update step that computes \( \hat{s}_t \in \mathcal{B}^L \) and \( q(\theta_{t,t}\vert y_{-}), l \in \tilde{S}_t \). Iterative MF message
Algorithm 1 Sequential VALSE

Input: Measurements $y_t$.

Prior initialization (Sec. V-A)

For all $t = 1, \ldots, T$, and $l = 1, \ldots, L$:

1. Initialize $\hat{a}_l = \hat{a}^1, \nu, \tau$.

2. Compute $J(49)$ and $h(50)$ for $q(w, s|y)$ [Sec. IV-C]

3. Repeat

(a) Update step [Sec. V-B]

(b) Update $s$ (57), followed by $\bar{w}_\hat{S}$ (52)

4. Update $\nu$ and $\tau$ [Sec. IV-D]

5. For all $l \in \hat{S}$, update $q(\theta_l|y) = f_{\text{VM}}(\theta_l; \hat{y}_l)$ (46)

6. Incorporate the predicted information (t > 1)

7. Incorporate the predicted information (t > 1) for the prior VM pdfs, $\eta_{\nu} = \hat{y}_{\nu}^t$ (41) and (63) [Sec. IV-B]

8. Until stopping criterion

Output: Number of sources estimate $|\hat{S}|$, DOA estimates $\hat{\theta}_\hat{S}$, and $\bar{w}_{\hat{S}}$ at time $t$

To obtain $q(\theta_l|y)$ and $\hat{a}_l$, we utilize [32, Heuristic 2].

Noise variance $\nu$ initialization. The noise variance $\nu$ influences the sparsity of the solution. The solution becomes sparse for high $\nu$ as it admits more noise and fits the data with fewer components. High initialized $\nu$ excludes many potential DOAs and results in missing DOAs. Low initialized $\nu$ considers more DOAs and needs more computations for convergence.

The original VALSE [32] builds a Toeplitz estimate of $E[yH]|yH]$ by averaging the diagonal elements of $yH$. Then, $\nu$ is initialized with the average of the lower quarter of the Toeplitz matrix eigenvalues. This calculation provides random performance depending on the eigenvalues, e.g., it can initialize $\nu$ too high, which will exclude all potential DOAs.

At initialization ($p = 0$) for $\nu$, we assume a measurement-to-noise ratio 20 dB, i.e., $10\log_{10}(\|y\|^2/M\nu) = 20$ or $\nu = \|y\|^2/2M(10)^{20}/10$. At later iterations, $\nu$ is estimated from (58).

The variance $\tau$ of the Gaussian distributed amplitudes $w$ (6) is initialized with $\tau = \langle yH\rangle(M - \nu)/\rho L$ obtained from $yH\rho = \rho L\tau + \nu$ (2). The activation probability $p$ of the Bernoulli variable $s$ is initialized as $p = 0.5$ at time $t = 1$.

Number of DOA transitions for prediction. DOAs are estimated using $L$ PAs, and the sequential method propagates the PAs through the prediction and update steps. At time $t - 1$, after update, we obtain $L$ PAs with $q(\theta_l|y)_{l=1}^{L}$, among which nonzero entries in $\hat{s}_{t-1} \in R^L$ are active and their DOAs are estimated. The PAs from time $t - 1$ are transferred to the prediction at time $t$ and used for the update at iteration $p = 0$. For the sequential process, we propagate only the active PAs.

At time $t - 1$, during iterations $p \in \{1, \ldots, P\}$, we record the activations $s_{t-1}^{(p)}$. Only active components and components with the activation history $s_{t-1}^{(p)} = 1$ are propagated to the subsequent prediction using the corresponding VM $q(\theta_l|y)_{l=1}^{L}$. The rest of the components out of $L$ are propagated with no VM. At time $t$, these are re-initialized using the VM initialization scheme (68) based on the propagated
transmit at $\beta$ initial positions $\tau$. The variance of source amplitudes is fixed to deactivated sequentially is simulated in Figs. 5 and 6. The first two, second two, and third two DOAs from $90^\circ$ two out of six DOAs are deactivated and a scenario where Sources can suddenly appear and disappear. A scenario where three static DOAs located at $\beta$ with initial angles $\theta$ transmit at $\nu$ noise variance with deterministic amplitude of 10, see Figs. 7 and 9(b). The simulated three static DOAs located at $\beta$ and $\nu$ m/s. In Figs. 4–6, we simulated six DOAs ($K = 6$) and that $K \neq \tilde{K}$ when the true and estimated DOAs have different cardinalities, mismatch cost for false and missed DOAs and assigns the optimum sub-pattern assignment metric (GOSPA) \cite{56}. We evaluate and compare the DOA performance based on (65) and the generalization optimum sub-pattern assignment metric (GOSPA) \cite{56}.

GOSPA assesses the number of estimated DOAs (cardinality) and the estimation error in angle. We can encounter when the true and estimated DOAs have different cardinalities, $K \neq \tilde{K}$, and GOSPA considers false and missed DOAs. False DOAs are the estimated DOAs but with an estimation error above $\frac{\pi}{6}$ ($c$ is a user-chosen constant, here $10^\circ$), and missed DOAs are DOAs not estimated. GOSPA sets the cardinality mismatch cost for false and missed DOAs and assigns the estimated DOAs to the true DOAs with the minimum measure.

Let $\theta_a \in \{1, \ldots, K\} \times \{1, \ldots, \tilde{K}\}$ be assignment set between two sets with $K$ and $\tilde{K}$ elements, which has the property that $(i, j), (i, j') \in \theta_a$ implies $j = j'$ and that $(i, j), (i', j) \in \theta_a$ implies $i = i'$. GOSPA is a sum of errors for the DOAs, assigned to the true DOAs, and a penalty for missed and false DOAs. The GOSPA

$$\text{SNR} = 20\log_{10} \left\| \sum_{l=1}^{K} w_{l,t} a(\theta_{l,t}) \right\|_2 / \| u_l \|_2. \quad (69)$$

**B. Metrics for DOA performance evaluation**

We evaluate and compare the DOA performance based on the root-mean-square error (RMSE), and the generalized optimum sub-pattern assignment metric (GOSPA) \cite{56}.

- **SNR**
  - **Equation:** $\text{SNR} = 20\log_{10} \left\| \sum_{l=1}^{K} w_{l,t} a(\theta_{l,t}) \right\|_2 / \| u_l \|_2$. 
  - **Explanation:**
    - The proposed framework estimates the DOAs using a sequential VALSE algorithm.
    - The dynamic DOA performance is evaluated next.

**VII. SIMULATION**

A. Setup

We considered $L = 15$ sensors and $K = 15$ potential DOAs. The sensors form a uniform linear array with a sensor spacing of 3.75 m, half-wavelength spacing, and sound speed $c = 1500$ m/s. In Figs. 4–6, we simulated six DOAs ($K = 6$ for all $t$) with initial angles $\beta_0 = [-70^\circ, -55^\circ, -40^\circ, 35^\circ, 50^\circ, 65^\circ]^T$ that transmit at 200 Hz. The two “inner” DOAs $k \in \{3, 4\}$ with initial positions $\beta_{3,0} = -40^\circ$ and $\beta_{4,0} = 35^\circ$ are dynamic. Sources can suddenly appear and disappear. A scenario where two out of six DOAs are deactivated and a scenario where the first two, second two, and third two DOAs from $90^\circ$ are deactivated sequentially is simulated in Figs. 5 and 6. The variance of source amplitudes is fixed to $\tau = 1$. We also simulated three static DOAs located at $\beta_1 = [-3^\circ, 2^\circ, 60^\circ]^T$ with deterministic amplitude of 10, see Figs. 7 and 9(b). The noise variance $\nu$ is set to obtain a signal-to-noise ratio (SNR),
Fig. 7: (a) Total GOSPA error and its (b) distance error, (c) missed and (d) false DOA error contributions versus SNR of simulated methods in Scenario 1.

Fig. 8: As in Fig. 7, but in Scenario 2.

Fig. 9: RMSE versus SNR of simulated methods in (a) Scenario 1 and (b) Scenario 2.

We also compare the RMSE calculated as,
\[
\mathcal{D}_{RMSE} = \min_{\theta_x \in \Theta(\theta_{true})} \left[ \sum_{(i,j) \in \theta_x} (\theta_i - \hat{\theta}_j)^2 + c^2 (|\theta_{true} - |\theta_x|) \right]
\]  
(74)

where \( c \) is a maximum DOA error for missed or false DOAs that remain unassigned. The DOA assignment is done using the Hungarian method for optimal point assignment [57]. In (70) and (74), we used \( c = c' = 10^5 \).

C. Algorithms

In the implementation of the VALSE with sequential processing, we set \( p^a = 0.10 \) and \( p^d = 0.25 \). For all PAs, we consider a state transition function \( f_{VM} (\theta_{t, t-1}) \) (61), where \( \Sigma_{r} = 148 \) (this gives \( \Sigma_{r} = 0.0822 \) rad; \( \sin^{-1} \frac{\Sigma_{r}}{2\sigma^2} = 1.499^\circ \)). As reference methods, we use SBL [13], sequential SBL (SSBL) [29] and VALSE [32]. In SBL and SSBL, we use the steering vectors corresponding to potential DOAs \( \theta_{t, t-1} \in \{ -90^\circ, -89.5^\circ, \ldots, 90^\circ, 90^\circ \} \), \( l \in I \triangleq \{ 0, \ldots, 360 \} \). After performing SBL or SSBL processing, we obtain final DOA estimates by locating local maximum source magnitudes above threshold 1% of the maximum. Non-sequential VALSE is entirely parameter-free. Gridless sparse methods [15], [16], [30], [31] do not use statistical information for sequential processing and the VALSE methods [33]–[35] do not consider time-varying DOAs, thus are omitted.

D. Performance with dynamic DOAs

Fig. 4 shows a single simulation of conventional beamforming (CBF), non-sequential VALSE [32], [33], and the proposed SVALSE. The sequential methods can more reliably localize DOAs near to endfire of the array, i.e., at \(-70^\circ, -55^\circ, 50^\circ, \) and \( 65^\circ \). In sequential processing, DOA detection is supported by
prior information from previous steps. The proposed SValSE localizes DOAs near the endfire accurately.

Dynamic sources involve sources that suddenly appear and disappear. A scenario where two out of six DOAs are deactivated and a scenario where the first two, second two, and third two DOAs from 90° are deactivated sequentially is simulated in Figs. 5 and 6. Although the prediction step in sequential processing uses the previously estimated DOAs, that are suddenly deactivated in the current measurement, the update step incorporates the current measurement and filters the deactivated DOAs.

The proposed SValSE localizes dynamic DOAs accurately. Appearings in Figs. 4–6(d) are comparisons of VALSE and SValSE with GOSPA (70) versus time. The error of the proposed SValSE is lower than VALSE and more stable.

E. SNR performance

We show DOA performance results versus SNR, based on the GOSPA (70), Figs. 7 and 8, and the RMSE (74), Fig. 9. We set the noise variance $\nu$ such that SNRs $\{0, 5, \ldots, 40\}$ dB are obtained. “Scenario 1” in Figs. 7 and 9(a) has three static DOAs located at $\beta = \begin{bmatrix} -3° & 2° & 60° \end{bmatrix}^T$ that all have the same deterministic amplitude of 10, and “scenario 2” in Figs. 8 and 9(b) has six DOAs with the variance of amplitudes $\tau = 1$ as in Fig. 4. We considered $t = 50$ time steps as in Fig. 4 for a single run and 100 simulation runs.

Figs. 7 and 8 show the mean GOSPA $D$ (70) and its distance $D_{\text{dist}}$, missed DOAs $D_{\text{miss}}$, and false DOAs $D_{\text{false}}$ error contributions (averaged over time and simulation runs) versus SNR. It can be seen that (S)ValSE yields lower GOSPA errors than (S)BL at SNR values below 10 dB. This is because for low SNR values, (S)BL overestimates the number of DOAs, see Figs. 7(d) and 8(d). ValSE performs slightly worse than SBL for high SNR. Grid-based SBL has an advantage over the proposed gridless method in $D_{\text{dist}}$, Figs. 7(b) and 8(b), since static DOAs in both scenarios are just on the grid without mismatch, resulting in 0° errors.

The GOSPA of (S)ValSE shows low values at SNR below 10 dB (Figs. 7(b) and 8(b)) since $D_{\text{dist}}$ considers only successfully estimated-and-assigned DOAs for its computation. Further, (S)ValSE underestimates DOAs, and only successfully estimated DOAs go into $D_{\text{dist}}$, providing low errors. Due to the error contribution of the missed DOAs, Figs. 7(c) and 8(c), the results have a high GOSPA, Figs. 7(a) and 8(a).

Fig. 9 shows the mean RMSE (74) (averaged over time and simulation runs) versus SNR. For low SNR values, (S)BL provides lower RMSE errors than (S)ValSE. This is because for low SNR, (S)BL overestimates the number of DOAs [Figs. 7(d) and 8(d)], and it is likely to estimate DOAs lower than the maximum DOA error. For high SNR, SValSE outperforms (S)BL in Fig. 9(b), as SValSE is gridless and estimates the number of DOAs accurately.

The proposed SValSE outperforms all the other methods. Compared to (S)BL is due to SValSE being gridless and accurately estimating the number of DOAs. The proposed sequential processing and initialization enhance the accuracy and estimation of the number of DOAs relative to ValSE.

VIII. REAL DATA PERFORMANCE

The time-varying DOA performance of the sequential ValSE is compared with CBF using experimental data, see Figs. 10 and 11. The data is from the shallow water evaluation cell experiment 1996 (SWellEx-96) Event S5 recorded at 23:15–00:30 GMT on 10–11 May 1996, which was performed West of Point Loma, CA [29], [58]. A vertical uniform linear array recorded the acoustic data with $M = 64$ sensors with spacing $d = 1.875$ m and spanning a depth of 94.125–212.25 m (Element 43 was corrupted and excluded). A shallow source was towed from 9 km southwest to 3 km northeast of the array at 2.5 m/s, with the closest point of approach 900 m.

The shallow source transmitted nine tones at frequencies $\{109, 127, 145, 163, 198, 232, 280, 335, 385\}$ Hz. We focus on the signal component at 198 Hz. The data was sampled at 1500 Hz, and the record at 23:21–00:24 GMT is divided into non-overlapping 350-time steps, see Fig. 10(c). Each measurement vector is obtained using the discrete Fourier transform with $2^{14}$ samples.

To analyze the time-varying DOA structure, see Fig. 10; we simulate the acoustic field using underwater acoustic propagation models, the Kraken normal mode model, and the Bellhop ray-tracing model [59]. Both methods require environmental information, including sound speed profile, bathymetry, and geo-acoustic parameters, and use the information as in [60]. The characteristics of the acoustic environment, including the water column sound speed profile [Fig. 10(b)], cause waveguide multipaths [Fig. 10(a)]. As the source moves closer to the VLA (increasing time step), the absolute DOA increases. The critical angle for this environment is $\cos^{-1}(1488/1600) = 21.6°$, beyond which multipath ceases to exist.

The CBF has significant peaks for multipaths over time [Fig. 11(a)], and the simulated results with the Kraken and the Bellhop match visually well [Figs. 11(b) and 11(d)]. The ray-tracing Bellhop provides the eigen-rays with amplitudes and explains the corresponding multipath.

DOA with dominant strength $\sim -15°$ corresponds to a bottom-reflected (BR) and a surface-bottom-reflected (SBR) path. DOA with a direct path (DP) has the strongest amplitude, but DP and SR $12°$ are weaker than BR and SBR $-15°$; see time steps 250–350 in Figs. 10(a) and 11(d). As the towed source was near the ocean surface, a path and the path with a surface reflection near the source has a DOA difference of $\sim 1°$, thus difficult to distinguish. We omit notations for the surface reflected paths, e.g., BR for BR and SBR.

At time steps 220–250, the source is weak at the receiver via DP and SR, and BR, and BSR $15°$ and BSBR $-17°$ dominate. Similarly, multi-surface-bottom-reflected paths, e.g., BSBSR and BSBSBR, successively show dominant strength, see time steps 1–200 in Figs. 10(a) and 11(d).

Sparse signal processing, non-sequential and sequential VALSE [Figs. 11(c) and 11(e)], results in improved resolution relative to CBF. The CBF [Fig. 11(a)] does not give as high resolution as (non-)sequential VALSE. The sequential VALSE [Fig. 11(e)] achieves increased accuracy and distinguishes the multipaths; see time steps 1–100.
IX. CONCLUSION

We introduced a variational Bayesian message passing for sparse gridless DOA estimation and a BP message passing for sequential processing for tracking time-varying DOAs.

The variational Bayesian method is guaranteed to converge locally, is gridless, incorporates prior information, and provides posterior pdfs of DOAs. In the prediction step, the sequential processing estimates the current state variables from posterior pdfs of the previous time and incorporates the estimates as prior information in the update step. For the variational Bayesian method, the Bernoulli-Gaussian amplitude model was used to promote a sparse solution, and it estimates the number of DOAs. For sequential Bayesian estimation, a Bernoulli-von Mises state-transition model was used.

The suggested method is a joint message passing algorithm that combines mean field message passing for variational Bayesian estimation and belief propagation for sequential processing. The operations are presented on factor graphs. This could provide new methods with different statistical models.

Our numerical evaluation based on simulations indicated improved DOA estimation performance compared to state-of-the-art methods. The suggested method provided good estimation for DOAs near the endfire of the array and robust estimation for dynamic sources, including varying DOAs and suddenly activated/deactivated DOAs. The method was demonstrated on real data from an ocean acoustics experiment.

REFERENCES

[1] H. L. Van Trees, *Optimum Array Processing*. New York, NY: Wiley, 2002.
[2] 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.
[3] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[4] E. J. Candes and M. B. Wakin, “An introduction to compressive sampling,” *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, 2008.
[5] A. Xenaki, P. Gerstoft, and K. Mosegaard, “Compressive beamforming,” *J. Acoust. Soc. Am.*, vol. 136, no. 1, pp. 260–271, Jan. 2014.
[6] P. Gerstoft, A. Xenaki, and C. F. Mecklenbräuker, “Multiple and single snapshot compressive beamforming,” *J. Acoust. Soc. Am.*, vol. 136, no. 4, pp. 2003–2014, Apr. 2015.
[7] 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.
[8] M. E. Tipping, “Sparse Bayesian learning and the relevance vector machine,” *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.
