Narrowband and Wideband Off-Grid Direction-of-Arrival Estimation via Sparse Bayesian Learning

Anup Das and Terrence J. Sejnowski, *Fellow, IEEE*

***Abstract*—The sparse Bayesian learning based relevance vector machine (SBLRVM) algorithm is a promising algorithm to estimate the directions-of-arrival (DOAs) of multiple narrowband signals. The parameters involved in the DOA estimation model are automatically estimated by the algorithm that makes it more attractive than the deterministic sparsity based DOA estimation algorithms in which fine-tuning of parameters is necessary. However, one limitation of the algorithm is that it assumes the DOAs of the signals to be exactly aligned with the angular grids, which may not be true in practice. In this paper, we first propose an off-grid version of the narrowband SBLRVM algorithm. Next, we propose an off-grid wideband SBLRVM algorithm. The algorithms assume that the true scenario DOAs of the signals are not exactly aligned with the angular grids and the parameters of the algorithms are automatically estimated by the expectation maximization approach. In the wideband DOA estimation algorithm, we estimate one spatial power spectrum by simultaneously exploiting sparsity from all frequency bins. We demonstrate the application of the proposed algorithms by analyzing data from the shallow water HF97 ocean acoustic experiment. The estimated DOAs of a narrowband tonal from the experiment by using our proposed narrowband DOA estimation algorithm are consistent with the nonadaptive conventional beamformer. Processing a wideband chirp from the experiment shows that estimating one spatial power spectrum by simultaneously exploiting sparsity from all frequency bins using the proposed wideband DOA estimation algorithm is a more valuable processor than an incoherent combination of the power spectra from the individual frequency bins estimated using the proposed narrowband DOA estimation algorithm. Moreover, since our proposed algorithms are off-grid algorithms, an empirical analysis for the choice of the discretization interval of the angular spread is not required as opposed to the on-grid DOA estimation algorithms. This results in a reduced computational complexity.**

***Index Terms*—Basis mismatch, beamforming, compressed sensing (CS), direction-of-arrival (DOA), expectation maximization (EM), likelihood, multiple measurement vector (MMV), off-grid model, sparse signal processing.**

Manuscript received May 22, 2016; revised January 2, 2017; accepted January 24, 2017. Date of publication February 21, 2017; date of current version January 11, 2018.

Associate Editor: L. Culver.

A. Das is with the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA 92093 USA and also with the Computational Neurobiology Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037 USA (e-mail: a1das@eng.ucsd.edu).

T. J. Sejnowski is with the Division of Biological Sciences and Institute of Neural Computation, University of California, San Diego, La Jolla, CA 92093 USA and also with the Howard Hughes Medical Institute, Salk Institute for Biological Studies, La Jolla, CA 92037 USA (e-mail: terry@salk.edu).

Digital Object Identifier 10.1109/JOE.2017.2660278

I. Introduction and Motivation

IDENTIFYING the directions-of-arrival (DOAs) of multiple signals using an array of sensors is one of the most fundamental problems in acoustic and geophysical sensing. The most commonly used methods for identifying the DOAs of multiple signals are the multiple signal classification (MUSIC) method [1], the matrix pencil method [2], [3], the estimating signal parameters via rotational invariance technique method [4], the minimum variance distortionless response (MVDR) method [5], and the deterministic maximum likelihood and stochastic maximum likelihood methods [6]. To reduce the computational complexity of these methods, low complexity learning-by-examples techniques based on artificial neural networks and support vector machines have also been proposed in the literature [7], [8].

In contrast to the aforementioned classical methods, estimating the DOAs by formulating the problem as a sparse signal recovery problem has received growing attention due to their higher resolution, robustness to noise, and better performance with a limited number of snapshots. Based on the theory of compressed sensing (CS) [9]–[12] and assuming the sparsity of signals in the spatial domain, sparse signal processing algorithms can estimate the DOAs of signals even when there is high spatial correlation between the signals or limited number of snapshots are present [13].

Several sparse signal recovery algorithms have been proposed in CS literature, which can be used to estimate the DOAs of multiple signals by exploiting the spatial sparsity. Matching pursuit methods [14]–[16] can be used to sequentially estimate the DOAs of the signals. Basis pursuit methods [15]–[17] use the ℓ_1-norm penalty to enforce sparsity in the spatial domain of the signals and can be used to identify the DOAs of the signals. Iterative reweighted methods [15] use an ℓ_p-norm penalty with $p \leq 1$ for enforcing sparsity and they give sparser solution as compared to matching pursuit and basis pursuit methods. Bayesian methods [18], [19] based on the sparse Bayesian learning principle [18] have even superior performance due to their use of data adaptive priors and capability of automatic regularization parameter selection and can be used to resolve the DOAs of multiple signals.

The above mentioned sparse signal recovery algorithms assume that the true scenario DOAs are exactly aligned with the angular grids. These algorithms discretize the continuous

angular domain, and hence, the estimated DOAs are always one of the discretized grids. However, in practice, the true DOAs may not be exactly aligned with the angular grids. Hence these algorithms suffer from grid bias. If the discretization of the angular spread is too coarse, then large error may be present in the estimated DOAs. This is known as basis mismatch [20], [21]. At the other extreme, if the discretization is too fine, then not only will we have a high computational complexity, but also the adjacent steering vectors will be heavily correlated and the spatial spectral power will spread over the adjacent steering vectors which also may result in grid-bias [22]. In practice, the grid interval is empirically determined for the on-grid DOA estimation methods, which significantly increases the computational complexity.

The researchers in [23] have proposed a Bayesian algorithm based on the sparse Bayesian learning principle [18] to estimate the DOAs of multiple signals. The algorithm first estimates the DOAs based on a predefined spatial discrete grid by using the sparse Bayesian learning formulation to obtain coarse location of signals. The algorithm then adopts a postprocessing step where a refined one-dimensional searching procedure is done to get a refined estimate of the DOAs one by one from the coarse DOA estimates to reduce the grid-bias. Similar multiresolution-based DOA refinement procedures have also been adopted by other researchers [24], [25]. But a limitation of such approaches is that the postprocessing procedure still increases the computational complexity (in terms of CPU time). Furthermore, the accuracy of the DOA estimates is limited by the resolution of the refining and the DOAs cannot be estimated to an arbitrary precision.

In this paper, we first propose an off-grid version of the narrowband sparse Bayesian learning based relevance vector machine algorithm (SBLRVM) algorithm (ON-SBLRVM algorithm). The SBLRVM algorithm is a promising algorithm to estimate the DOAs of multiple narrowband signals [13] since the parameters involved in the DOA estimation model are automatically estimated by the algorithm that makes it more attractive than the deterministic sparsity based DOA estimation algorithms. The novel algorithm directly incorporates an off-grid model in the DOA estimation model and simultaneously estimates the DOAs and the offsets in the DOAs during sparse Bayesian learning [18] and, thus, avoids computationally expensive empirical analyses or multiresolution postprocessing steps. The offsets in the DOAs can be estimated to an arbitrary precision and only limited by the machine precision. In our proposed algorithms, we directly incorporate the off-grid model from [26] in the DOA estimation model that takes into account the off-grid DOAs. All the hidden variables and parameters of the model are estimated in the sparse Bayesian inference step itself and, hence, no postprocessing steps are necessary. For efficient inference, we use the expectation maximization (EM) approach [27]. Next, we propose an off-grid wideband SBLRVM algorithm (OW-SBLRVM algorithm), which will be applicable to wideband signals.

We demonstrate the application of the proposed algorithms by analyzing data from the shallow water HF**97** ocean acoustic experiment [13], [28], [29]. The estimated DOAs of a narrowband tonal from the experiment by using our proposed narrowband DOA estimation algorithm were consistent with the nonadaptive conventional beamformer (CBF). Processing a wideband chirp from the experiment shows that estimating one spatial power spectrum by simultaneously exploiting sparsity from all frequency bins using the proposed wideband DOA estimation algorithm is a more valuable processor than an incoherent combination of the power spectra from the individual frequency bins estimated using the proposed narrowband DOA estimation algorithm. Moreover, since our proposed algorithms are off-grid algorithms, an empirical analysis for the choice of the discretization interval of the angular spread is not required as opposed to the on-grid DOA estimation algorithms. This also results in a reduced computational complexity.

The rest of this paper is organized as follows. In Section II, we describe the narrowband off-grid DOA estimation model and state the assumptions made on the statistics of signal and noise. In Section III, we describe our proposed narrowband off-grid sparse Bayesian learning algorithm based on the EM approach. In Section IV, we straightforwardly extend our proposed narrowband DOA estimation algorithm to the wideband case. A simulation study is done in Section V to demonstrate the advantage of an off-grid DOA estimation model. The application of the proposed algorithms for estimating the DOAs of narrowband and wideband multipath signals is also demonstrated in Section VI by analyzing data from the shallow water HF**97** ocean acoustic experiment. Finally, we draw conclusions in Section VII.

II. Narrowband Off-Grid DOA Estimation Model

A. DOA Estimation Model

We consider a uniform linear array (ULA) consisting of N identical sensors and receiving K far-field plane wave signals with arbitrary spatial correlation. We assume that we have preprocessed the sensor array data by taking fast Fourier transforms (FFTs) and, hence, the signals are narrowband. Let λ denote the wavelength corresponding to the frequency of the signals. These K signals arrive at the array from directions $\theta_1, \theta_2, \ldots, \theta_K$. Using complex signal representation, the measurement at the array at the jth instant can be represented by [6]

$$\mathbf{y}_{\cdot j} = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}_{\cdot j} + \mathbf{n}_{\cdot j} \tag{1}$$

where $\mathbf{y}_{\cdot j} \triangleq [y_{1j}, y_{2j}, \ldots, y_{Nj}]^T$ is the $N \times 1$ array output data vector of Fourier coefficients obtained via the FFT and $\mathbf{s}_{\cdot j} = [s_{1j}, s_{2j}, \ldots, s_{Kj}]^T$, $\mathbf{n}_{\cdot j} = [n_{1j}, n_{2j}, \ldots, n_{Nj}]^T$, and $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \ldots, \mathbf{a}(\theta_K)]$ (from now, the notation $\cdot j$ will be used to denote a column of a matrix and the notation $i\cdot$ will be used to denote a row of a matrix). Here, $\mathbf{a}(\theta_k)$ represents the direction vector associated with the kth signal and is given by

$$\mathbf{a}(\theta_k) = \frac{1}{\sqrt{N}}\left[e^{-\sqrt{-1}\frac{N-1}{2}\psi_k}, e^{-\sqrt{-1}\frac{N-3}{2}\psi_k}, \ldots, e^{\sqrt{-1}\frac{N-1}{2}\psi_k}\right]^T \tag{2}$$

where $\psi_k = 2\pi(d/\lambda)\cos\theta_k$. Here $(\cdot)^T$ denotes the transpose. Note that the direction vectors are l_2 normalized, i.e.,

$\mathbf{a}^H(\theta_k)\mathbf{a}(\theta_k) = 1$ for $k = 1, 2, \ldots, K$, where $(\cdot)^H$ denotes the complex conjugate transpose. The $N \times 1$ vector $\mathbf{n}_{\cdot j}$ is the additive noise at the array at the jth instant.

After preprocessing the sensor array data with FFTs, the signals are assumed to be zero-mean stationary complex Gaussian random processes [30]. Furthermore, the signals and the additive noise are assumed to be independent of each other. Each noise vector also is assumed to be a zero-mean stationary complex Gaussian random process. Furthermore, it is assumed that the noises are uncorrelated sensor-to-sensor and across measurements with common variance σ^2.

Assuming sparsity in the spatial domain, i.e., a very few number of signals are present, we formulate the problem of DOA estimation as a problem of sparse signal recovery in an overcomplete matrix. We discretize the angular spread $[0^\circ, 180^\circ]$ of the ULA to result in M steering vectors having the same formulation as the direction vectors given in (2). We construct the $N \times M$ matrix $\mathbf{A}$, which contains the M steering vectors as its columns with $N \ll M$. Note that matrix $\mathbf{A}(\boldsymbol{\theta})$ contains the direction vectors of the signals, whereas matrix $\mathbf{A}$ contains the steering vectors of the DOAs where a signal may or may not be present. Assuming L snapshots, i.e., $j = 1, 2, \ldots, L$ in (1), we construct the $M \times L$ matrix $\mathbf{X}$ where any particular row contains the complex amplitudes of a signal corresponding to the steering vector in $\mathbf{A}$ if a signal is present in that steering direction or zero, otherwise. We also assume that a very few number of signals are present, i.e., $\mathbf{X}$ is row sparse. We represent the array output vectors $\mathbf{y}_{\cdot j}$ as the columns of a matrix $\mathbf{Y}$ and the noise vectors $\mathbf{n}_{\cdot j}$ as the columns of a matrix $\boldsymbol{\mathcal{E}}$. Hence, the set of equations in (1), where $j = 1, 2, \ldots, L$, can be represented equivalently as

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \boldsymbol{\mathcal{E}}. \tag{3}$$

We note that in (3), the objective is to recover the row sparse matrix $\mathbf{X}$ given the observation matrix $\mathbf{Y}$ and the overcomplete matrix $\mathbf{A}$, giving rise to a noisy sparse signal recovery problem with multiple snapshots or multiple measurement vectors. As mentioned before, the true DOAs may not be exactly aligned with the steering vectors, and hence, the on-grid DOA estimation model in (3) may not be useful in practice. We next describe the off-grid DOA estimation model. This model is taken from [26].

B. Off-Grid Model

Let $\widetilde{\boldsymbol{\theta}} = \{\widetilde{\theta}_1, \widetilde{\theta}_2, \ldots, \widetilde{\theta}_M\}$ be a uniform discretization of the angular spread in $[0^\circ, 180^\circ]$. Assume that the true DOA set is $\{\theta_1, \theta_2, \ldots, \theta_K\}$. Suppose $\theta_k \notin \{\widetilde{\theta}_1, \widetilde{\theta}_2, \ldots, \widetilde{\theta}_M\}$ for some $k \in \{1, 2, \ldots, K\}$ and that $\widetilde{\theta}_{m_k},\ m_k \in \{1, 2, \ldots, M\}$, is the nearest grid point to θ_k. Using the Taylor series expansion, we make a linear approximation of the direction vector $\mathbf{a}(\theta_k)$ as

$$\mathbf{a}(\theta_k) \approx \mathbf{a}\left(\widetilde{\theta}_{m_k}\right) + \mathbf{b}\left(\widetilde{\theta}_{m_k}\right)\left(\theta_k - \widetilde{\theta}_{m_k}\right) \tag{4}$$

where $\mathbf{b}(\widetilde{\theta}_{m_k}) = \mathbf{a}'(\widetilde{\theta}_{m_k})$, i.e., $\mathbf{b}(\widetilde{\theta}_{m_k})$ is the derivative of $\mathbf{a}(\widetilde{\theta}_{m_k})$ with respect to $\widetilde{\theta}_{m_k}$ given by

$$\mathbf{b}\left(\widetilde{\theta}_{m_k}\right) = \begin{bmatrix} -\sqrt{-1}\frac{N-1}{2}2\pi\frac{d}{\lambda}(-\sin\widetilde{\theta}_{m_k})e^{-\sqrt{-1}\frac{N-1}{2}\widetilde{\psi}_{m_k}} \\ -\sqrt{-1}\frac{N-3}{2}2\pi\frac{d}{\lambda}(-\sin\widetilde{\theta}_{m_k})e^{-\sqrt{-1}\frac{N-3}{2}\widetilde{\psi}_{m_k}} \\ \vdots \\ \sqrt{-1}\frac{N-1}{2}2\pi\frac{d}{\lambda}(-\sin\widetilde{\theta}_{m_k})e^{\sqrt{-1}\frac{N-1}{2}\widetilde{\psi}_{m_k}} \end{bmatrix} \tag{5}$$

where $\widetilde{\psi}_{m_k} = 2\pi(\frac{d}{\lambda})\cos\widetilde{\theta}_{m_k}$. We have $\mathbf{A} = [\mathbf{a}(\widetilde{\theta}_1), \mathbf{a}(\widetilde{\theta}_2), \ldots, \mathbf{a}(\widetilde{\theta}_M)]$. Let us denote the derivative matrix as $\mathbf{B} = [\mathbf{b}(\widetilde{\theta}_1), \mathbf{b}(\widetilde{\theta}_2), \ldots, \mathbf{b}(\widetilde{\theta}_M)]$ and $\boldsymbol{\delta} = [\delta_1, \delta_2, \ldots, \delta_M]^T$ with

$$\delta_m = \begin{cases} \theta_k - \widetilde{\theta}_{m_k}, & \text{for } m = m_k \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where $m = 1, 2, \ldots, M$. Hence, the DOA estimation model in (3) can be written as

$$\mathbf{Y} = (\mathbf{A} + \mathbf{B}\boldsymbol{\Delta})\,\mathbf{X} + \boldsymbol{\mathcal{E}} \tag{7}$$

where $\boldsymbol{\Delta} = \mathrm{diag}(\boldsymbol{\delta})$. The model in (7) is the off-grid model used in this paper.

In Section III, we describe the EM approach to the off-grid DOA estimation problem.

III. Narrowband Off-Grid Bayesian Algorithm Using Expectation Maximization Approach

We first describe the hierarchical Bayesian modeling by assuming that all the observed and unknown variables are stochastic and their joint prior probability distribution is specified. For mathematical tractability, we also assume that this joint distribution can be factored into individual prior or conditional distributions of the variables under consideration. We then estimate all the parameters of the off-grid model by using the EM approach.

A. Stochastic Model

Denoting $\widetilde{\mathbf{A}} = \mathbf{A} + \mathbf{B}\boldsymbol{\Delta}$, the off-grid model in (7) becomes

$$\mathbf{Y} = \widetilde{\mathbf{A}}\mathbf{X} + \boldsymbol{\mathcal{E}}. \tag{8}$$

Since the additive noise $\boldsymbol{\mathcal{E}}$ has complex white Gaussian distribution with variance σ^2, $p(\mathbf{Y}|\mathbf{X}; \sigma^2)$ is also complex Gaussian. Thus, for each $\mathbf{y}_{\cdot j}, \mathbf{x}_{\cdot j}$ pair, we have the likelihood of the array output as

$$p(\mathbf{y}_{\cdot j}|\mathbf{x}_{\cdot j}; \sigma^2, \boldsymbol{\delta}) = (\pi\sigma^2)^{-N}\exp\left(-\frac{1}{\sigma^2}\left\|\mathbf{y}_{\cdot j} - \widetilde{\mathbf{A}}\mathbf{x}_{\cdot j}\right\|_2^2\right) \tag{9}$$

and hence

$$p(\mathbf{Y}|\mathbf{X}; \sigma^2, \boldsymbol{\delta}) = \prod_{j=1}^{L} p(\mathbf{y}_{\cdot j}|\mathbf{x}_{\cdot j}; \sigma^2, \boldsymbol{\delta}). \tag{10}$$

Following the sparse Bayesian learning principle [18], we assign the following M-dimensional complex Gaussian prior for each $\mathbf{x}_{\cdot j}$ as

$$p(\mathbf{x}_{\cdot j};\mathbf{\Gamma}) = \mathcal{CN}(\mathbf{x}_{\cdot j}|0,\mathbf{\Gamma}) \tag{11}$$

where $\mathbf{\Gamma}$ is the variance matrix given by $\mathbf{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma})$, where $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \ldots, \gamma_M]^T$. Such a prior can exploit the spatial sparsity of the signal [18]. Hence, we arrive at the full weight prior given by

$$p(\mathbf{X};\mathbf{\Gamma}) = \prod_{j=1}^{L} p(\mathbf{x}_{\cdot j};\mathbf{\Gamma}). \tag{12}$$

Combining the likelihood and prior distributions, we have the joint distribution as

$$p(\mathbf{Y},\mathbf{X};\sigma^2,\boldsymbol{\delta},\mathbf{\Gamma}) = p(\mathbf{Y}|\mathbf{X};\sigma^2,\boldsymbol{\delta})p(\mathbf{X};\mathbf{\Gamma}). \tag{13}$$

We next use the EM approach to estimate the hidden variables and parameters of the off-grid model.

B. Estimation of Hidden Variables and Parameters

The EM approach consists of iterative E-steps and M-steps. The E-step requires the computation of the first and second posterior moments of the hidden variable $\mathbf{X}$ and the M-step estimates the parameters σ^2, $\boldsymbol{\delta}$, and $\mathbf{\Gamma}$. This can be done as follows.

1) Inference of $\mathbf{X}$*:* From the posterior distribution $p(\mathbf{X}|\mathbf{Y};\sigma^2,\boldsymbol{\delta},\mathbf{\Gamma})$, it is straightforward to show that the covariance and mean of this distribution can be given respectively by

$$\mathbf{\Sigma} = \mathbf{\Gamma} - \mathbf{\Gamma}\widetilde{\mathbf{A}}^H\mathbf{\Sigma}_y^{-1}\widetilde{\mathbf{A}}\mathbf{\Gamma}, \qquad \langle\mathbf{X}\rangle = \mathbf{\Gamma}\widetilde{\mathbf{A}}^H\mathbf{\Sigma}_y^{-1}\mathbf{Y} \tag{14}$$

where $\mathbf{\Sigma}_y \triangleq \sigma^2\mathbf{I}_N + \widetilde{\mathbf{A}}\mathbf{\Gamma}\widetilde{\mathbf{A}}^H$.

2) Estimation of σ^2*,* $\boldsymbol{\delta}$*, and* $\mathbf{\Gamma}$*:* In the EM approach, we treat the weights $\mathbf{X}$ as hidden variables and maximize $E_{\mathbf{X}|\mathbf{Y};\sigma^2,\boldsymbol{\delta},\mathbf{\Gamma}}\left[\log p(\mathbf{Y}|\mathbf{X};\sigma^2,\boldsymbol{\delta})p(\mathbf{X};\mathbf{\Gamma})\right]$. For $\mathbf{\Gamma}$, ignoring the terms in the logarithm independent thereof, we equivalently maximize

$$E_{\mathbf{X}|\mathbf{Y};\sigma^2,\boldsymbol{\delta},\mathbf{\Gamma}}\left[\log p(\mathbf{X};\mathbf{\Gamma})\right] \tag{15}$$

which through differentiation gives the estimates for γ_i as

$$\widehat{\gamma_i} = \frac{1}{L}\|\langle\mathbf{x}_{i\cdot}\rangle\|_2^2 + \mathbf{\Sigma}_{ii} \quad \forall i = 1,\ldots,M. \tag{16}$$

Following the corresponding procedure for the noise level σ^2, we equivalently maximize

$$E_{\mathbf{X}|\mathbf{Y};\sigma^2,\boldsymbol{\delta},\mathbf{\Gamma}}\left[\log p(\mathbf{Y}|\mathbf{X};\sigma^2,\boldsymbol{\delta})\right] \tag{17}$$

which gives

$$\widehat{\sigma^2} = \frac{1}{NL}\|\mathbf{Y} - \widetilde{\mathbf{A}}\langle\mathbf{X}\rangle\|_F^2 + \frac{1}{N}\,\mathrm{Tr}\left(\widetilde{\mathbf{A}}^H\widetilde{\mathbf{A}}\mathbf{\Sigma}\right). \tag{18}$$

For the estimation of $\boldsymbol{\delta}$, we also equivalently maximize

$$E_{\mathbf{X}|\mathbf{Y};\sigma^2,\boldsymbol{\delta},\mathbf{\Gamma}}\left[\log p(\mathbf{Y}|\mathbf{X};\sigma^2,\boldsymbol{\delta})\right] \tag{19}$$

which leads to

$$\widehat{\boldsymbol{\delta}} = -\mathbf{P}^{-1}\mathbf{p} \tag{20}$$

where

$$\mathbf{P} = \mathrm{Re}\left[\overline{\mathbf{B}^H\mathbf{B}} \odot \left(\sum_{j=1}^{L}\langle\mathbf{x}_{\cdot j}\rangle\langle\mathbf{x}_{\cdot j}\rangle^H + L\mathbf{\Sigma}\right)\right] \tag{21}$$

and

$$\mathbf{p} = \mathrm{Re}\left[-\mathrm{diag}\left(\mathbf{B}^H\sum_{j=1}^{L}\left(\mathbf{y}_{\cdot j} - \mathbf{A}\langle\mathbf{x}_{\cdot j}\rangle\right)\langle\mathbf{x}_{\cdot j}\rangle^H\right). + L\,\mathrm{diag}\left(\mathbf{B}^H\mathbf{A}\mathbf{\Sigma}\right)\right] \tag{22}$$

where $\odot$ in (21) denotes the Hadamard (elementwise) product and $\mathrm{diag}(\cdot)$ in (22) denotes the column vector containing the diagonal of the matrix under brackets. Hence we have,

$$\widehat{\mathbf{\Delta}} = \mathrm{diag}\left(\widehat{\boldsymbol{\delta}}\right). \tag{23}$$

If $\mathbf{P}$ is not invertible, then we can update $\boldsymbol{\delta}$ elementwise as done in [26]. Equations (14), (16), (18), and (20) are the iterative update equations for the proposed off-grid Bayesian algorithm using the EM approach. To estimate the DOAs of the signals in the off-grid model, we first select the K largest peaks in the power spectrum of $\langle\mathbf{X}\rangle$. We denote the grid indices of these peaks as $\widehat{m}_k$ where $k = 1,2,\ldots,K$. The estimates of the DOAs of the K signals will be $\widehat{\theta}_k = \widetilde{\theta}_{\widehat{m}_k} + \widehat{\delta}_{\widehat{m}_k}$ for $k = 1,2,\ldots,K$.

Note that since $\delta_m \in [-r/2, r/2]$ for $m = 1,2,\ldots,K$ where r denotes the grid interval, if the estimate $\widehat{\delta}_m$ of δ_m falls outside this interval, then we set $\widehat{\delta}_m$ to be the corresponding upper limit of the interval.

The proposed off-grid Bayesian algorithm is summarized in Table I. We note that we have directly incorporated the off-grid model during sparse Bayesian learning as evident from (20). Hence any kind of postprocessing is unnecessary, thus reducing the computational complexity. We also note that since we can estimate the offset in the DOAs due to grid-bias in (20), we can adopt a coarse discretization of the angular domain thus reducing the computational complexity even further and at the same time achieving high accuracy in the DOA estimates of the signals.

Our proposed off-grid Bayesian algorithm is locally convergent due to the properties of the EM algorithm [27] and global convergence is not guaranteed. But, one of the major advantages of the ON-SBLRVM algorithm is that it is fairly robust to different initialization conditions since it is based on the sparse Bayesian learning principle [13], [18]. In [23], the researchers adopt the least squares initialization procedure, which we also follow for our simulations (see Section V) and experimental data analysis (see Section VI).

In Section IV, we extend the narrowband version of the proposed algorithm to the wideband case straightforwardly.

IV. Extension to the Wideband Case

Here, we extend the proposed narrowband off-grid Bayesian learning algorithm to the wideband case. For this, we assume that all the K incident signals are in the same spectral band

TABLE I
SUMMARY OF THE PROPOSED NARROWBAND OFF-GRID BAYESIAN ALGORITHM (ON-SBLRVM ALGORITHM)

Given the observation data $\mathbf{Y}$, the overcomplete matrix $\mathbf{A}$ containing the steering vectors, and the matrix $\mathbf{B}$ containing the derivatives of the steering vectors, the proposed narrowband off-grid Bayesian algorithm can be summarized as follows.
(i) Initialize $\mathbf{\Gamma}$, σ^2, and $\boldsymbol{\delta}$.
(ii) Compute the posterior moments $\mathbf{\Sigma}$ and $\langle \mathbf{X} \rangle$ using (14).
(iii) Update $\mathbf{\Gamma}$ using (16), update σ^2 using (18), and update $\boldsymbol{\delta}$ using (20).
(iv) Iterate (ii) and (iii) until convergence. Declare the algorithm to be converged when the change in $\mathbf{X}$ is less than some predefined threshold.
(v) After convergence, the value of $\langle \mathbf{X} \rangle$ is the estimate of the desired weight matrix $\mathbf{X}$. The algorithm forces the entries in the rows of $\mathbf{X}$ to be zero in which a signal component is not present. We then select the K largest peaks in the power spectrum of $\langle \mathbf{X} \rangle$. We denote the grid indices of these peaks as $\widehat{m}_k$ where $k = 1, 2, \ldots, K$. The estimates of the DOAs of the K signals will be $\widehat{\theta}_k = \widetilde{\theta}_{\widehat{m}_k} + \widehat{\delta}_{\widehat{m}_k}$ for $k = 1, 2, \ldots, K$.

(for a solution of the wideband DOA estimation problem of signals in different spectral bands, see [31]). Also assume that the array output has been decomposed into J frequency bins $f_1, \ldots, f_J$ within the band to obtain narrowband measurements $\mathbf{Y}_{f_1}, \ldots, \mathbf{Y}_{f_J}$. Hence, the wideband off-grid DOA estimation model in the sparse domain becomes

$$\mathbf{Y}_f = (\mathbf{A}_f + \mathbf{B}_f \mathbf{\Delta})\, \mathbf{X}_f + \boldsymbol{\mathcal{E}}_f \tag{24}$$

where $f \in \{f_1, \ldots, f_J\}$. In (24), $\mathbf{Y}_f$ is the measurement at frequency f, $\mathbf{A}_f$ is the matrix containing the frequency-dependent steering vectors corresponding to the frequency f, $\mathbf{B}_f$ is the matrix containing the derivatives of the frequency-dependent steering vectors, $\mathbf{X}_f$ is the row sparse matrix containing the complex amplitudes of the signals, and $\boldsymbol{\mathcal{E}}_f$ is the noise matrix where the noises are assumed to be uncorrelated temporally, spectrally, and sensor-to-sensor with common variance σ^2 and independent of the signals as before. It is also assumed that the signals received in the frequency bins $f_1, \ldots, f_J$ share the same DOAs and, hence, will have the same offset due to the grid-bias in the frequency bins $f_1, \ldots, f_J$. Thus, the offset $\mathbf{\Delta}$ in the DOAs is kept the same for all frequency bins $f_1, \ldots, f_J$.

To realize a joint sparse signal recovery problem, we assume that $\mathbf{X}_{f_1}, \ldots, \mathbf{X}_{f_J}$ have complex Gaussian distribution with the same variance matrix $\mathbf{\Gamma}$ as follows:

$$p(\mathbf{X}_f; \mathbf{\Gamma}) = \prod_{j=1}^{L} p(\mathbf{x}_{\cdot j_f}; \mathbf{\Gamma}), \quad f \in \{f_1, \ldots, f_J\} \tag{25}$$

where $\mathbf{x}_{\cdot j_f}$ is the jth column of $\mathbf{X}_f$ and

$$p(\mathbf{x}_{\cdot j_f}; \mathbf{\Gamma}) = \mathcal{CN}(\mathbf{x}_{\cdot j_f} | 0, \mathbf{\Gamma}). \tag{26}$$

Hence, the joint probability distribution can be written as

$$\begin{aligned} &p(\mathbf{Y}_{f_1}, \ldots, \mathbf{Y}_{f_J}, \mathbf{X}_{f_1}, \ldots, \mathbf{X}_{f_J}; \sigma^2, \boldsymbol{\delta}, \mathbf{\Gamma}) \\ &\quad = p(\mathbf{Y}_{f_1} | \mathbf{X}_{f_1}; \sigma^2, \boldsymbol{\delta}) \cdots p(\mathbf{Y}_{f_J} | \mathbf{X}_{f_J}; \sigma^2, \boldsymbol{\delta}) \\ &\qquad \times p(\mathbf{X}_{f_1}; \mathbf{\Gamma}) \cdots p(\mathbf{X}_{f_J}; \mathbf{\Gamma}). \end{aligned} \tag{27}$$

The hidden variables and parameters in this wideband model can be estimated by using the EM approach. Since the joint distribution for all frequency bins in (27) decomposes as the products of the likelihoods and priors for individual frequency bins, the joint posterior distribution $p(\mathbf{X}_{f_1}, \ldots, \mathbf{X}_{f_J} | \mathbf{Y}_{f_1}, \ldots, \mathbf{Y}_{f_J}; \sigma^2, \boldsymbol{\delta}, \mathbf{\Gamma})$ decomposes as the products of the posterior distributions for individual frequency bins. The covariance and mean of the posterior distributions at different frequency bins can be given by

$$\begin{aligned} \mathbf{\Sigma}_f &= \mathbf{\Gamma} - \mathbf{\Gamma} \widetilde{\mathbf{A}}_f^H \mathbf{\Sigma}_{y_f}^{-1} \widetilde{\mathbf{A}}_f \mathbf{\Gamma}, \quad f \in \{f_1, \ldots, f_J\} \\ \langle \mathbf{X}_f \rangle &= \mathbf{\Gamma} \widetilde{\mathbf{A}}_f^H \mathbf{\Sigma}_{y_f}^{-1} \mathbf{Y}_f, \quad f \in \{f_1, \ldots, f_J\} \end{aligned} \tag{28}$$

where $\mathbf{\Sigma}_{y_f} \triangleq \sigma^2 \mathbf{I}_N + \widetilde{\mathbf{A}}_f \mathbf{\Gamma} \widetilde{\mathbf{A}}_f^H$ and $\widetilde{\mathbf{A}}_f = \mathbf{A}_f + \mathbf{B}_f \mathbf{\Delta}$.

Applying the EM approach, we treat the weights $\mathbf{X}_{f_1}, \ldots, \mathbf{X}_{f_J}$ as hidden variables and maximize $E_{\mathbf{X}_{f_1}, \ldots, \mathbf{X}_{f_J} | \mathbf{Y}_{f_1}, \ldots, \mathbf{Y}_{f_J}; \sigma^2, \boldsymbol{\delta}, \mathbf{\Gamma}} [\log p(\mathbf{Y}_{f_1} | \mathbf{X}_{f_1}; \sigma^2, \boldsymbol{\delta}) \cdots p(\mathbf{Y}_{f_J} | \mathbf{X}_{f_J}; \sigma^2, \boldsymbol{\delta})\, p(\mathbf{X}_{f_1}; \mathbf{\Gamma}) \cdots p(\mathbf{X}_{f_J}; \mathbf{\Gamma})]$. For $\mathbf{\Gamma}$, ignoring the terms in the logarithm independent thereof, we equivalently maximize

$$E_{\mathbf{X}_{f_1}, \ldots, \mathbf{X}_{f_J} | \mathbf{Y}_{f_1}, \ldots, \mathbf{Y}_{f_J}; \sigma^2, \boldsymbol{\delta}, \mathbf{\Gamma}} [\log p(\mathbf{X}_{f_1}; \mathbf{\Gamma}) \cdots p(\mathbf{X}_{f_J}; \mathbf{\Gamma})] \tag{29}$$

which through differentiation gives the estimate for γ_i as

$$\widehat{\gamma_i} = \frac{1}{JL} \sum_{f \in \{f_1, \ldots, f_J\}} \|\langle \mathbf{x}_{i f \cdot} \rangle\|_2^2 + \frac{1}{J} \sum_{f \in \{f_1, \ldots, f_J\}} (\mathbf{\Sigma}_f)_{ii}. \tag{30}$$

Following the corresponding procedure for the noise level σ^2, we equivalently maximize

$$\begin{aligned} &E_{\mathbf{X}_{f_1}, \ldots, \mathbf{X}_{f_J} | \mathbf{Y}_{f_1}, \ldots, \mathbf{Y}_{f_J}; \sigma^2, \boldsymbol{\delta}, \mathbf{\Gamma}} \\ &\quad \left[\log p(\mathbf{Y}_{f_1} | \mathbf{X}_{f_1}; \sigma^2, \boldsymbol{\delta}) \cdots\ p(\mathbf{Y}_{f_J} | \mathbf{X}_{f_J}; \sigma^2, \boldsymbol{\delta})\right] \end{aligned} \tag{31}$$

which gives

$$\begin{aligned} \widehat{\sigma^2} = \frac{1}{J} \Bigg[\sum_{f \in \{f_1, \ldots, f_J\}} \Bigg\{ &\frac{1}{NL} \|\mathbf{Y}_f - \widetilde{\mathbf{A}}_f \langle \mathbf{X}_f \rangle\|_F^2 \\ &+ \frac{1}{N} \operatorname{Tr}\left(\widetilde{\mathbf{A}}_f^H \widetilde{\mathbf{A}}_f \mathbf{\Sigma}_f \right) \Bigg\} \Bigg]. \end{aligned} \tag{32}$$

For the estimation of $\boldsymbol{\delta}$, we also equivalently maximize

$$\begin{aligned} &E_{\mathbf{X}_{f_1}, \ldots, \mathbf{X}_{f_J} | \mathbf{Y}_{f_1}, \ldots, \mathbf{Y}_{f_J}; \sigma^2, \boldsymbol{\delta}, \mathbf{\Gamma}} \\ &\quad \left[\log p(\mathbf{Y}_{f_1} | \mathbf{X}_{f_1}; \sigma^2, \boldsymbol{\delta}) \cdots\ p(\mathbf{Y}_{f_J} | \mathbf{X}_{f_J}; \sigma^2, \boldsymbol{\delta})\right] \end{aligned} \tag{33}$$

which leads to

$$\widehat{\boldsymbol{\delta}} = -\mathbf{P}^{-1} \mathbf{p} \tag{34}$$

TABLE II
SUMMARY OF THE PROPOSED WIDEBAND OFF-GRID BAYESIAN ALGORITHM (OW-SBLRVM ALGORITHM)

Given the measurements $\mathbf{Y}_{f_1},\ldots,\mathbf{Y}_{f_J}$, the overcomplete matrices $\mathbf{A}_{f_1},\ldots,\mathbf{A}_{f_J}$ containing the frequency dependent steering vectors, and the matrices $\mathbf{B}_{f_1},\ldots,\mathbf{B}_{f_J}$ containing the derivatives of the frequency dependent steering vectors, the proposed wideband off-grid Bayesian algorithm can be summarized as follows.

(i) Initialize $\mathbf{\Gamma}$, σ^2, and $\boldsymbol{\delta}$.

(ii) Compute the posterior moments $\mathbf{\Sigma}_f,\ f\in\{f_1,\ldots,f_J\}$ and $\langle\mathbf{X}_f\rangle,\ f\in\{f_1,\ldots,f_J\}$ using (28).

(iii) Update $\mathbf{\Gamma}$ using (30), update σ^2 using (32), and update $\boldsymbol{\delta}$ using (34).

(iv) Iterate (ii) and (iii) until convergence. Declare the algorithm to be converged when the change in $\mathbf{\Gamma}$ is less than some predefined threshold.

(v) After convergence, the values of $\langle\mathbf{X}_{f_1}\rangle,\ldots,\langle\mathbf{X}_{f_J}\rangle$ are the estimates of the desired weight matrices $\mathbf{X}_{f_1},\ldots,\mathbf{X}_{f_J}$. The algorithm forces the entries in the rows of $\mathbf{X}_{f_1},\ldots,\mathbf{X}_{f_J}$ to be zero in which a signal component is not present. We then select the K largest values in the diagonal of estimated $\mathbf{\Gamma}$ which contains the power levels of the signals. We denote the corresponding grid indices of these signals as $\widehat{m}_k$ where $k=1,2,\ldots,K$. The estimates of the DOAs of the K signals will be $\widehat{\theta}_k=\widetilde{\theta}_{\widehat{m}_k}+\widehat{\delta}_{\widehat{m}_k}$ for $k=1,2,\ldots,K$.

where

$$\mathbf{P}=\mathrm{Re}\Bigg[\sum_{f\in\{f_1,\ldots,f_J\}}\Bigg\{\overline{\mathbf{B}_f^H\mathbf{B}_f}\odot\left(\sum_{j=1}^{L}\langle\mathbf{x}_{\cdot j_f}\rangle\langle\mathbf{x}_{\cdot j_f}\rangle^H+L\mathbf{\Sigma}_f\right)\Bigg\}\Bigg] \tag{35}$$

and

$$\mathbf{p}=\mathrm{Re}\Bigg[\sum_{f\in\{f_1,\ldots,f_J\}}\Bigg\{-\mathrm{diag}\left(\mathbf{B}_f^H\sum_{j=1}^{L}\left(\mathbf{y}_{\cdot j_f}-\mathbf{A}_f\langle\mathbf{x}_{\cdot j_f}\rangle\right)\times\langle\mathbf{x}_{\cdot j_f}\rangle^H\right)+L\,\mathrm{diag}\left(\mathbf{B}_f^H\mathbf{A}_f\mathbf{\Sigma}_f\right)\Bigg\}\Bigg]. \tag{36}$$

Hence we have,

$$\widehat{\mathbf{\Delta}}=\mathrm{diag}\left(\widehat{\boldsymbol{\delta}}\right). \tag{37}$$

Equations (28), (30), (32), and (34) are the iterative update equations for the proposed off-grid Bayesian algorithm for wideband signals. To estimate the DOAs of the signals, we first select the K largest values in the diagonal of estimated $\mathbf{\Gamma}$, which contains the power levels of the signals. We denote the corresponding grid indices of these signals as $\widehat{m}_k$, where $k=1,2,\ldots,K$. The estimates of the DOAs of the K signals will be $\widehat{\theta}_k=\widetilde{\theta}_{\widehat{m}_k}+\widehat{\delta}_{\widehat{m}_k}$ for $k=1,2,\ldots,K$. The algorithm for the wideband case is summarized in Table II.

Here, we clarify that even though the priors on the parameters in the ON-SBLRVM and OW-SBLRVM algorithms are assumed to be complex Gaussian, the algorithms are robust to the distribution of parameters since they are based on the sparse Bayesian learning principle [32]. This makes the proposed algorithms to be useful in practice in which the true distributions of the signals might be unknown.

V. SIMULATIONS AND DISCUSSION

We now carry out a simulation study and demonstrate the improvement in performance of our proposed ON-SBLRVM DOA estimation algorithm in comparison to the state-of-the-art algorithms. The significant advantage of an off-grid model in the proposed algorithms is that we can retain the DOA estimation accuracy even if the discretization of the angular spread is coarse. Thus, an empirical determination of the discretization interval is not necessary in contrast to on-grid DOA estimation algorithms. First, we assume that the true DOAs of the signals are exactly aligned with the steering vectors and study the root-mean-squared-error (RMSE) in the DOA estimates in terms of input signal-to-noise ratio (SNR) and number of snapshots. We compare the performance of the ON-SBLRVM algorithm with the state-of-the-art off-grid sparse Bayesian Inference (OGSBI) algorithm [26] and the on-grid complex SBLRVM algorithm [13] [note that the OGSBI algorithm in [26] uses a Gamma hyperprior to exploit the spatial sparsity of the signals, but our proposed algorithm does not use such a hyperprior, rather the sparsity is imposed via the sparse Bayesian learning principle [18] (for theoretical results, see [18])]. Next, we assume that the true DOAs of the signals are not exactly aligned with the steering vectors and analyze the RMSE in DOA estimates in terms of the grid interval and demonstrate the superior performance of the ON-SBLRVM algorithm. Similar advantages apply for the OW-SBLRVM algorithm.

For simulation, for each individual signal at each sensor, the input SNR is defined to be

$$\mathrm{SNR}=10\log_{10}\left(\frac{\sigma_s^2}{\sigma^2}\right) \tag{38}$$

where σ^2 is the variance of the ith sensor noise sequence defined as $\sigma^2\triangleq E[|n_{ij}|^2]$. The power of the kth signal is defined as $\sigma_s^2\triangleq E[|x_{kj}|^2]$.

A. Analysis of RMSE in the DOA Estimate

Here, we analyze the RMSE in the DOA estimate of OGSBI, ON-SBLRVM, and SBLRVM algorithms in terms of SNR. We consider a 12 element ULA with a half-wavelength spacing between the array elements. A 1° discretization on the angular spread and $L=50$ snapshots are also assumed. We consider two coherent multipath signals (i.e., the correlation coefficient is one between the two signals) of equal power levels impinging on the ULA from DOAs 60° and 68° so that the true DOAs are aligned with the steering vectors. Since the algorithms are based on CS theory, the performance does not change significantly by varying the correlation coefficient. For each input SNR, we consider a total of 100 trials and calculate the RMSE for the signal at 60° for all methods. The result is shown in Fig. 1. We note that the performances of all three algorithms are comparable to each other.

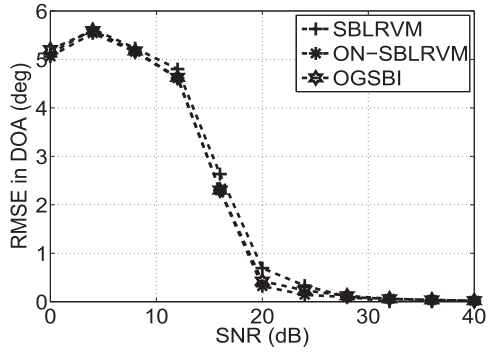

Fig. 1. Comparison of RMSE between OGSBI, ON-SBLRVM, and SBLRVM algorithms in terms of input SNR. Multipath arrivals of equal power levels at angles of $\theta_1 = 60^\circ$ and $\theta_2 = 68^\circ$ and $L = 50$ snapshots are considered.

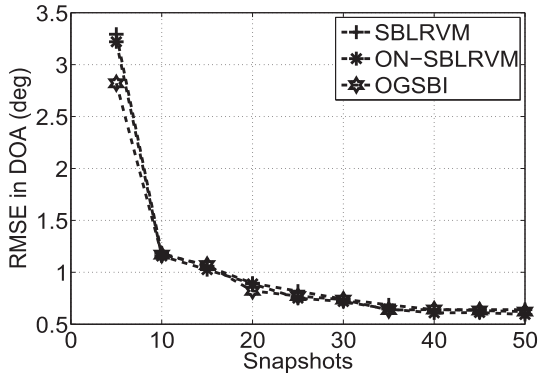

Fig. 2. Comparison of RMSE between OGSBI, ON-SBLRVM, and SBLRVM algorithms in terms of number of snapshots. Multipath arrivals of equal power levels at angles of $\theta_1 = 60^\circ$ and $\theta_2 = 75^\circ$ and SNR $= 5$ dB are considered.

B. Analysis in Terms of Number of Snapshots

Here, we analyze the RMSE in the DOA estimate of OGSBI, ON-SBLRVM, and SBLRVM algorithms in terms of number of snapshots. We consider a 12 element ULA with half-wavelength spacing between the array elements. A 1° discretization on the angular spread also is assumed. We consider two coherent multipath signals of equal power levels impinging on the ULA from DOAs 60° and 75° so that the true DOAs of the multipath signals are aligned with the steering vectors. The value of SNR is fixed at 5 dB. For each input SNR, we consider a total of 100 trials and calculate the RMSE for the signal at 60° for all the algorithms. The result is shown in Fig. 2. We note that the RMSE in the DOA estimate decreases as more snapshots are available for all the methods demonstrating that more snapshots improve the performance. Also note that the performances of the algorithms are comparable to each other.

C. Advantage of the Off-Grid Bayesian Algorithm

Here we analyze the effect of grid interval on the RMSE in DOA estimates of OGSBI, ON-SBLRVM, and SBLRVM algorithms. We consider a 12 element ULA with a half-wavelength spacing between the array elements. The grid interval values are chosen to be 0.5°, 1°, 2°, 3°, and 4°. We consider a total 100 trials. For each trial, we also consider two coherent signals

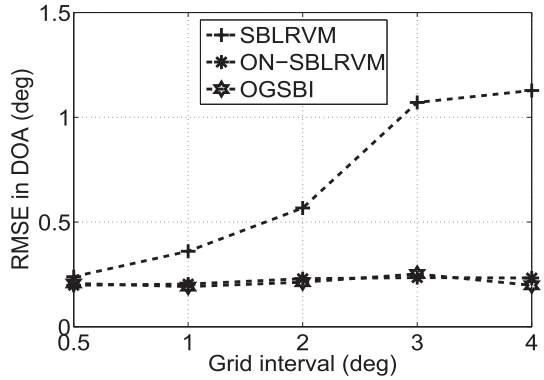

Fig. 3. Comparison of RMSE between OGSBI, ON-SBLRVM, and SBLRVM algorithms in terms of grid interval. SNR of 10 dB and $L = 50$ snapshots are considered.

of equal power levels whose DOAs are uniformly generated in the intervals $[57.5^\circ, 62.5^\circ]$ and $[82.5^\circ, 87.5^\circ]$, respectively, and calculate the RMSE in the DOA estimate of the first signal. The value of SNR is fixed at 10 dB and $L = 50$ snapshots are assumed. The analysis results are shown in Fig. 3. We note that even though the grid interval increases significantly, the RMSE in the DOA estimates of ON-SBLRVM and OGSBI algorithms does not increase much in contrast to the SBLRVM algorithm. Furthermore, the performances of the ON-SBLRVM and OGSBI algorithms are comparable to each other. Hence, we can adopt a relatively coarser sampling grid that will reduce the computational complexity and overcome the problems of finer sampling grid as mentioned before and at the same time retain the DOA estimation accuracy. In the on-grid DOA estimation methods such as SBLRVM, the discretization of the angular spread is determined empirically which significantly increases the computationally complexity. But in off-grid methods such as the ON-SBLRVM and OGSBI algorithms, we can adopt a fine enough discretization that will ensure high resolution without any empirical analysis since the offsets in the DOAs will be automatically estimated in such methods.

We next demonstrate the application of the proposed algorithms for estimating the DOAs of narrowband and wideband coherent multipath signals by analyzing data from the shallow water HF**97** ocean acoustic experiment [13], [28], [29].

VI. HF**97** Ocean Acoustic Experiment

The HF**97** experiment [28], [29] was carried out in shallow water off the coast of Point Loma, CA, USA, in October 1997. This experimental data have also been used in our previous work [13] where we demonstrated the resolution of the DOAs of coherent multipath signals of a narrowband tonal by using the on-grid complex SBLRVM algorithm. An overview of this experiment showing the source and receiver array (R/P FLIP) positions is shown in Fig. 4. The water depth was approximately 100 m and the source and the receiver were fixed to the bottom. The receiver consisted of a 64-element vertical linear array (VLA) and was deployed approximately 6 km away from the source. The interelement spacing of the VLA was $d = 0.1875$ m and the hydrophone elements of the VLA

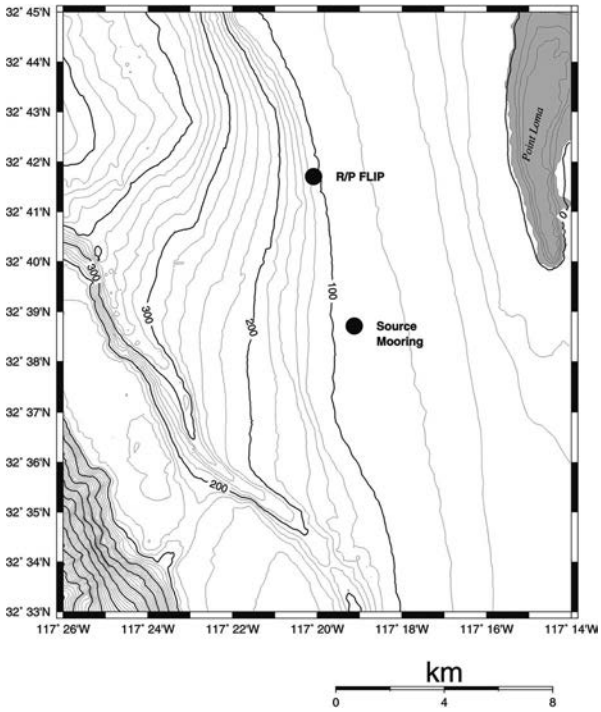

Fig. 4. Experiment overview showing the source mooring and receiving array locations in the HF**97** experiment.

were sampled at $f_s = 48$ kHz. The source transmitted several waveforms. Of interests here were a sinusoidal transmission at 3.1 kHz and a broadband chirp of central frequency 2.5 kHz and bandwidth 1 kHz. For the sinusoidal and broadband chirp transmissions, we note that there is no spatial aliasing since at these frequencies the interelement spacing is less than 0.5λ.

The start time of our data from the HF**97** experiment is Julian Day 301 2210 UTC. For the sinusoidal transmission, we first extract the 3.1 kHz tonal by using nonwindowed nonoverlapping FFTs of length $2^{12} = 4096$. For the tonal, we process a total of 400 snapshots from the start time, divide the 400 snapshots into two blocks of 200 snapshots each, and resolve the DOAs of the signals for each of the blocks of 200 snapshots using the ON-SBLRVM algorithm. We use all 64 sensor elements of the VLA for processing. For the algorithm, a 0.2° discretization on the angular spread is also assumed. Such a discretization was taken to resolve the DOAs of very closely spaced signals. For the broadband chirp transmission, we only consider the frequency range 2.5–3 kHz. We extract a total of 167 frequency bins by using nonwindowed nonoverlapping FFTs of length $2^{14} = 16\,384$. We process a total of 100 snapshots from the start time for each frequency bin. By combining all the frequency bins, we estimate the DOAs of the signals by the OW-SBLRVM algorithm.

Ray tracing results are shown in [28] using a CTD cast taken 2 h after the data discussed here. These results suggest that a number of arrivals are expected approximately $\pm 10^\circ$ of broadside (broadside corresponds to 90° in our case). Similar ray tracing results are shown in [29] including the observed arrival angle versus travel time structure of the channel impulse response at the same time the data discussed here were recorded. These also show a number of multipath arrivals approximately $\pm 10^\circ$ of broadside. Thus, the DOA estimates of the multipath signals shown here are consistent with the propagation physics and modeling. We also calculated the eigenvalue spectrum for each block of the 200 snapshots for the 3.1 kHz tonal and for the 100 snapshots in each frequency bin in the frequency range 2.5–3 kHz for the broadband chirp. All the eigenvalue spectra clearly showed only one dominant eigenvalue (at least 15 dB higher than the others) implying that the multipath signals are coherent. In such a scenario, the proposed algorithms in this paper are very attractive since they are based on CS theory and are not affected by the coherence between signals in contrast to conventional DOA estimation methods such as MVDR and MUSIC [6].

We first analyze the results for the 3.1 kHz tonal. Fig. 5 shows the result for the ON-SBLRVM algorithm. For the algorithm, we first estimated the spatial power spectrum and rejected all the peaks in the power spectrum whose power levels were more than 15 dB below the highest peak and then reestimated the power levels of the remaining peaks by using least squares to obtain more accurate estimates of the power levels. The least squares reestimation of the power levels was necessary since sparse signal recovery methods underestimate the true power levels of the signals [15], [33]–[35]. The total number of these remaining peaks was selected as the total number of signals. The choice of the value 15 dB was arbitrary.

There were no ground truth DOAs for this experiment. Hence we have included the estimated spatial power spectrum of the nonadaptive CBF in all the results. The CBF spatial power spectrum $\widehat{P}_{\text{CBF}}$ is defined as [6]

$$\widehat{P}_{\text{CBF}} = \mathbf{a}^H(\theta)\widehat{\mathbf{R}}\mathbf{a}(\theta) \tag{39}$$

where θ denotes an arbitrary arrival angle and $\widehat{\mathbf{R}}$ is the estimated output array covariance matrix. The nonadaptive CBF provides a reasonable (though not high resolution) representation of the acoustic field observed by the array and hence was used as an indication of the ground truth.

From Fig. 5, we note that the estimated DOAs for the proposed method are consistent with the nonadaptive CBF. Our algorithm estimated a multipath at approximately 92.0103° (result rounded up to four decimal places) in Fig. 5(a), which is indicated by a slightly broadened spectrum by CBF. Furthermore, our proposed algorithm estimated two DOAs at approximately 85.1031° and 86.8945° in Fig. 5(b). These two multipath signals were not resolved by CBF due to low resolution, but CBF gave an unusually broadened spectrum indicating the presence of more than one multipath signals. The rest of the DOAs estimated by our algorithm and the nonadaptive CBF were consistent with each other. This shows that our algorithm can estimate the DOAs of the multipath signals with higher resolution. The OGSBI algorithm also produced similar results, and hence, the results are not shown.

We next analyze the result for the broadband chirp. We only consider the 2.5–3-kHz bandwidth (see Fig. 6) for the chirp to demonstrate the application of the OW-SBLRVM algorithm. The DOAs for the 2–2.5-kHz bandwidth can be estimated in a similar manner. For the OW-SBLRVM algorithm, we first

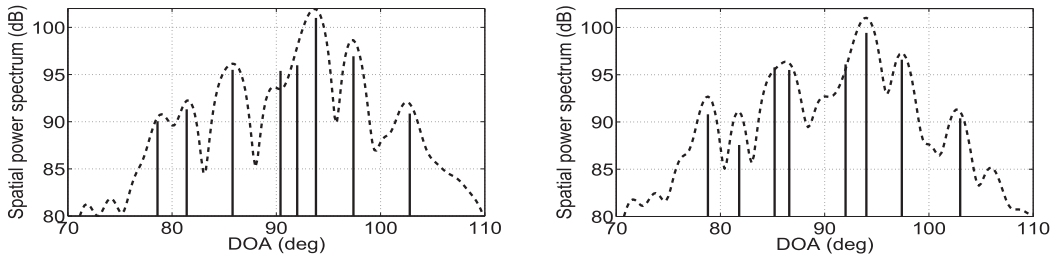

Fig. 5. DOA estimation of narrowband coherent multipath signals using the proposed narrowband DOA estimation algorithm for the 3.1 kHz frequency in the HF**97** experiment. Panels (a) and (b) correspond to the first and second blocks of 200 snapshots ($\sim$17 s each), respectively. Dotted line denotes the CBF spectrum and solid line denotes the proposed algorithm spectrum.

estimated $\mathbf{\Gamma}$ and rejected all the peaks in the power spectrum of $\mathbf{\Gamma}$ whose power levels were more than 20 dB below the highest peak. The DOAs corresponding to the remaining peaks were selected as the DOAs of the signals. To calculate the spatial power spectrum of our OW-SBLRVM algorithm, we reestimated the power levels in each frequency bin by using least squares by only considering the estimated DOAs and then took an average of the power levels across all frequency bins. We compared the wideband spatial power spectrum of our OW-SBLRVM algorithm with CBF results from two frequency bins (2625 and 2725 Hz) and the incoherent average of the power spectra of CBF from all frequency bins. The frequency bins 2625 and 2725 Hz were arbitrarily chosen. These results are shown in Fig. 6(a)–(c) respectively. From Fig. 6(c), we note that the estimated DOAs in our OW-SBLRVM method are consistent with the incoherently averaged CBF spectrum. We also observe that whenever CBF is not able to resolve the DOAs due to low resolution, our proposed OW-SBLRVM method is able to resolve them with higher resolution.

Our proposed algorithm shows the presence of two signals at approximately 93.7010° and 95.5100° which can also be seen in the incoherently averaged CBF spectrum in Fig. 6(c), but these two signals are not resolved by the CBF spectra from the individually considered frequency bins in Fig. 6(a) and (b). Our proposed OW-SBLRVM algorithm also shows the presence of two signals at approximately 88.7013° and 90.5005°. These two signals are clearly resolved in the CBF spectrum in the frequency bin in Fig. 6(b), but are not resolved by the CBF spectrum in the frequency bin in Fig. 6(a) and the incoherently averaged CBF spectrum in Fig. 6(c). Our algorithm indicates the presence of two signals at approximately 84.1215° and 85.2941°, which are resolved by the incoherently average CBF spectrum in Fig. 6(c). The CBF spectrum in Fig. 6(a) also indicates the presence of these two signals by an unusually broadened spectrum. But the CBF spectrum for the individual frequency bin in Fig. 6(b) is not able to resolve these two signals due to low resolution. Furthermore, we observe that our algorithm estimates the DOAs at approximately 79.2021° and 81.8001°, which are clearly estimated by the CBF spectrum in Fig. 6(b), but CBF is not able to resolve these two multipath signals in Fig. 6(a) and (c). We also note that the DOAs estimated by CBF and our algorithm are not exactly the same since our algorithm estimates the DOAs in a coherent manner, i.e., it estimates the DOAs by combining all the frequency bins which helps improve the resolution. Similar results were obtained for other frequency bins. It is not possible to exactly quantify the results in the experiment since there was no ground truth available.

To demonstrate the advantage of the OW-SBLRVM algorithm, which estimates the DOAs by coherently combining all frequency bins, over an incoherently averaged spectrum from all frequency bins by the ON-SBLRVM algorithm in the paper, we did the following. We first estimated the power spectrum for each frequency bin by using the proposed narrowband off-grid DOA estimation method in this paper. We then took an average of the power spectra from all 167 frequency bins. We rejected all the peaks in this average power spectrum whose power levels were more than 20 dB below the highest peak. The DOAs corresponding to the remaining peaks were selected as the DOAs of the signals. The estimated DOAs are shown in dotted black lines in Fig. 7. Comparing Figs. 6 and 7, we note that the incoherently averaged power spectrum by using the narrowband DOA estimation method in this paper for the individual frequency bins is not sparse and produces a large number of spurious DOAs (25 DOAs in total as opposed to 13 DOAs in the wideband DOA estimation algorithm). Since the DOAs of the multipath signals from different frequency bins are not exactly aligned with each other, simply taking an incoherent average of the power spectra from all frequency bins produces significantly larger number of spurious DOAs than the number of multipath signals indicated by the CBF spectrum. We also reestimated the power levels in each frequency bin using least squares by only considering the estimated DOAs and then took an average of the power levels across all frequency bins. But this did not give any useful estimate of the power levels, i.e., the power level estimates were significantly different from that estimated by the incoherently averaged CBF spectrum. This is due to the presence of the large number of spurious DOAs (note that this phenomenon can be easily demonstrated via simulation). Hence the power level information is not shown in Fig. 7. This shows that a coherent combination of all frequency bins by simultaneously exploiting sparsity is a more valuable processor for DOA estimation of wideband signals.

Since for both the narrowband and wideband cases, the signals were coherent, high-resolution DOA estimators such as MVDR and MUSIC spatial processors cannot be used [13]. Though spatial smoothing techniques produce improved results, they reduce the array aperture essentially reducing the effective resolution [13]. We have applied both MVDR and MUSIC (with spatial

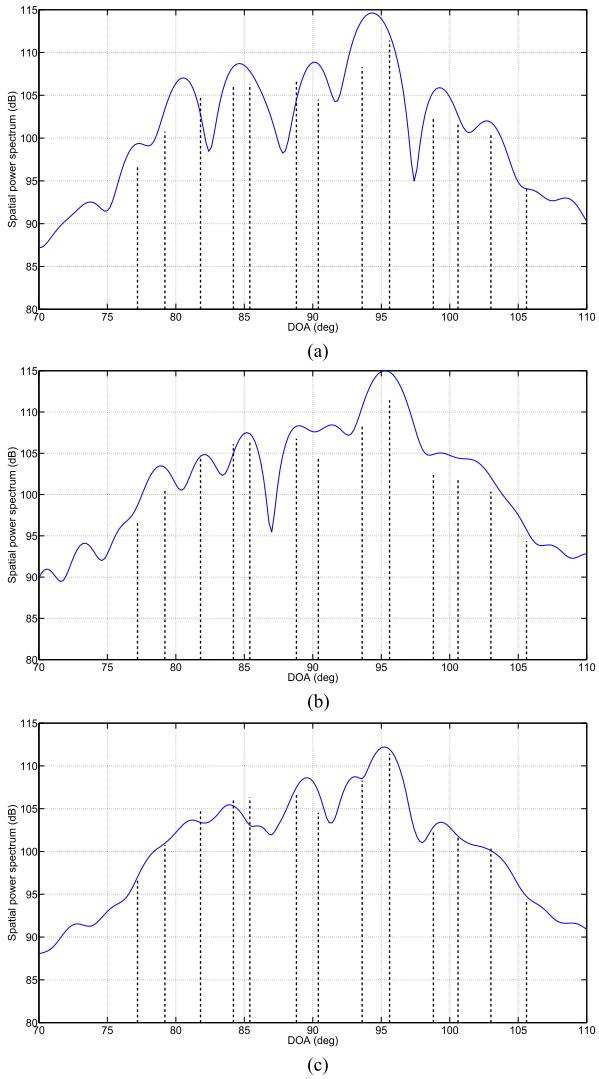

Fig. 6. Comparing the CBF spectra (denoted in blue) from two frequency bins [(a) 2625 Hz and (b) 2725 Hz] and the incoherent average of the CBF spectra from all frequency bins (c) with the estimated spectrum [denoted in dotted black in (a), (b), and (c)] by the proposed wideband DOA estimation algorithm.

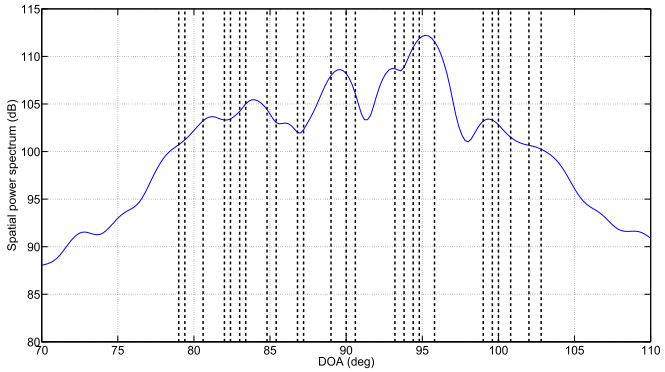

Fig. 7. Estimated DOAs (denoted by dotted black lines) by the incoherently averaged power spectrum using the narrowband DOA estimation method in our paper for the individual frequency bins. Also plotted (in blue) is the incoherently averaged CBF power spectrum.

smoothing) for our narrowband and wideband DOA estimation problem in the HF**97** experiment, but the resolution was even poorer than the CBF spatial processor.

Conventional wideband DOA estimation algorithms such as CSSM [36] and WAVES [37] are also available in array processing literature. To design a fully coherent processor by exploiting the properties of a chirp, methods such as [38]–[41] can also be used. However, these demonstrations are reserved for future research and are not objectives of this paper.

VII. Conclusion

We proposed a novel algorithm for the off-grid DOAs estimation problem using the sparse Bayesian learning principle and the EM approach. We derived algorithms for both narrowband and wideband cases. Since the algorithms directly model the DOA offsets, an empirical determination of the grid interval is not necessary in contrast to the on-grid DOA estimation algorithms, hence reducing the computational complexity. We also demonstrated the application of the proposed algorithms for estimating the DOAs of narrowband and wideband coherent multipath signals by analyzing data from the shallow water HF**97** ocean acoustic experiment. For the narrowband case, the experimental results from our algorithm were consistent with the nonadaptive CBF spatial processor and also showed higher resolution. For the wideband case, our algorithm showed higher resolution than the incoherently averaged CBF spectrum from all frequency bins. This shows that for wideband signals, simultaneously exploiting the sparsity from all frequency bins is a promising method for DOA estimation in contrast to an incoherently averaged power spectrum from all frequency bins using a narrowband sparse DOA estimation method for each frequency bin. Furthermore, since our algorithms are off-grid algorithms, we can adopt a coarse sampling grid, when required, which will relatively reduce the computational complexity and at the same time retain high accuracy in the DOA estimates.

Acknowledgment

The authors would like to thank Dr. W. S. Hodgkiss for providing the HF**97** experimental data and for providing feedback on an initial version of this manuscript.

References

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

[2] N. Yilmazer, J. Koh, and T. K. Sarkar, "Utilization of a unitary transform for efficient computation in the matrix pencil method to find the direction of arrival," *IEEE Trans. Antennas Propag.*, vol. 54, no. 1, pp. 175–181, Jan. 2006.

[3] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 5, pp. 814–824, May 1990.

[4] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul. 1989.

[5] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.

[6] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Optimum Array Processing*. Hoboken, NJ, USA: Wiley, 2004.

[7] A. H. El Zooghby, C. G. Christodoulou, and M. Georgiopoulos, "Performance of radial-basis function networks for direction of arrival estimation with antenna arrays," *IEEE Trans. Antennas Propag.*, vol. 45, no. 11, pp. 1611–1617, Nov. 1997.
[8] M. Pastorino and A. Randazzo, "A smart antenna system for direction of arrival estimation based on a support vector regression," *IEEE Trans. Antennas Propag.*, vol. 53, no. 7, pp. 2161–2168, Jul. 2005.
[9] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
[10] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
[11] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[12] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.
[13] A. Das, W. S. Hodgkiss, and P. Gerstoft, "Coherent multipath direction-of-arrival resolution using compressed sensing," *IEEE J. Ocean. Eng.*, 2016, DOI: 10.1109/JOE.2016.2576198.
[14] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, 2006.
[15] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, Jul. 2005.
[16] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634–4643, Dec. 2006.
[17] J. A. Tropp, "Algorithms for simultaneous sparse approximation. Part II: Convex relaxation," *Signal Process.*, vol. 86, no. 3, pp. 589–602, 2006.
[18] 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, Jul. 2007.
[19] M. Carlin, P. Rocca, G. Oliveri, F. Viani, and A. Massa, "Directions-of-arrival estimation through Bayesian compressive sensing strategies," *IEEE Trans. Antennas Propag.*, vol. 61, no. 7, pp. 3828–3838, Jul. 2013.
[20] M. A. Herman and T. Strohmer, "General deviants: An analysis of perturbations in compressed sensing," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 342–349, Apr. 2010.
[21] 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.
[22] P. Stoica and P. Babu, "Sparse estimation of spectral lines: Grid selection problems and their solutions," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 962–967, Feb. 2012.
[23] 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.
[24] 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.
[25] M. Donelli, F. Viani, P. Rocca, and A. Massa, "An innovative multiresolution approach for DOA estimation based on a support vector classification," *IEEE Trans. Antennas Propag.*, vol. 57, no. 8, pp. 2279–2292, Aug. 2009.
[26] 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.
[27] D. G. Tzikas, C. L. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 131–146, Nov. 2008.
[28] N. M. Carbone and W. S. Hodgkiss, "Effects of tidally driven temperature fluctuations on shallow-water acoustic communications at 18 kHz," *IEEE J. Ocean. Eng.*, vol. 25, no. 1, pp. 84–94, Jan. 2000.
[29] W. S. Hodgkiss, W. A. Kuperman, and D. E. Ensberg, "Channel impulse response fluctuations at 6 kHz in shallow water," in *Impact of Littoral Environmental Variability of Acoustic Predictions and Sonar Performance*. New York, NY, USA: Springer-Verlag, 2002, pp. 295–302.
[30] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.
[31] L. Wang, L. Zhao, G. Bi, C. Wan, L. Zhang, and H. Zhang, "Novel wideband DOA estimation based on sparse Bayesian learning with Dirichlet process priors," *IEEE Trans. Signal Process.*, vol. 64, no. 2, pp. 275–289, Jan. 2016.
[32] 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.
[33] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 586–597, Dec. 2007.
[34] C. F. Mecklenbrauker, 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.
[35] B. D. Rao, K. Engan, S. F. Cotter, J. Palmer, and K. Kreutz-Delgado, "Subset selection in noise based on diversity measure minimization," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 760–770, Mar. 2003.
[36] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 4, pp. 823–831, Aug. 1985.
[37] E. D. di Claudio and R. Parisi, "WAVES: Weighted average of signal subspaces for robust wideband direction finding," *IEEE Trans. Signal Process.*, vol. 49, no. 10, pp. 2179–2191, Oct. 2001.
[38] N. Ma and J. T. Joo Goh, "Ambiguity-function-based techniques to estimate DOA of broadband chirp signals," *IEEE Trans. Signal Process.*, vol. 54, no. 5, pp. 1826–1839, May 2006.
[39] X. G. Xia, "Discrete chirp-fourier transform and its application to chirp rate estimation," *IEEE Trans. Signal Process.*, vol. 48, no. 11, pp. 3122–3133, Nov. 2000.
[40] Y. Zhang, W. Ma, and M. G. Amin, "Subspace analysis of spatial time-frequency distribution matrices," *IEEE Trans. Signal Process.*, vol. 49, no. 4, pp. 747–759, Apr. 2001.
[41] A. Gershman and M. G. Amin, "Wideband direction-of-arrival estimation of multiple chirp signals using spatial time-frequency distributions," *IEEE Signal Process. Lett.*, vol. 7, no. 6, pp. 152–155, Jun. 2000.

Anup Das was born in Bhubaneswar, Orissa, India, on May 15, 1991. He received the B.Tech. degree in electronics and communication engineering from the Indian Institute of Technology, Guwahati, India, in 2013 and the M.S. degree in electrical engineering (signal and image processing) from the University of California San Diego, La Jolla, CA, USA, in 2015, where he is currently working toward the Ph.D. degree in electrical engineering (signal and image processing), under the supervision of Prof. T. Sejnowski.

His research interests include the general areas of statistical signal processing, sparse signal processing, and machine learning and Bayesian inference with applications to array signal processing and computational neuroscience.

Mr. Das received the University of California, San Diego Electrical and Computer Engineering Departmental Fellowship in 2013.

Terrence J. Sejnowski (F'00) received the Ph.D. degree in physics from Princeton University, Princeton, NJ, USA in 1978.

He is currently an Investigator with the Howard Hughes Medical Institute, Chevy Chase, MD, USA, a Distinguished Professor at the University of California San Diego, La Jolla, CA, USA, and holds the Francis Crick Chair at The Salk Institute, La Jolla, CA, USA. He is a pioneer in computational neuroscience and his goal is to understand the principles that link brain to behavior. His laboratory uses both experimental and modeling techniques to study the biophysical properties of synapses and neurons and the population dynamics of large networks of neurons. He has published more than 500 scientific papers and 14 books, including *The Computational Brain*.

Dr. Sejnowski is the President of the Neural Information Processing Systems Foundation, the premier conference on machine learning and neural computation. He is a member of the National Academy of Sciences, the National Academy of Medicine, and the National Academy of Engineering. He was instrumental in shaping the BRAIN Initiative that was announced by the White House in 2013 and served on the Working Group of the Advisory Committee to the Director of NIH for the BRAIN Initiative.