Signal Processing 159 (2019) 159–170

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Sparse Bayesian learning with multiple dictionaries

Santosh Nannuru[a,*], Kay L. Gemba[b], Peter Gerstoft[b], William S. Hodgkiss[b], Christoph F. Mecklenbräuker[c]

[a] Signal Processing and Communications Research Center, IIIT Hyderabad, India
[b] Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA
[c] Institute of Telecommunications, TU Wien, Vienna 1040, Austria

ARTICLE INFO

Article history:
Received 23 May 2018
Revised 7 January 2019
Accepted 4 February 2019
Available online 5 February 2019

Keywords:
Sparse Bayesian learning
Compressive sensing
Beamforming
DOA estimation
Multiple dictionaries
Multi frequency
Aliasing
Wide band
Heteroscedastic noise

ABSTRACT

Sparse Bayesian learning (SBL) has emerged as a fast and competitive method to perform sparse processing. The SBL algorithm, which is developed using a Bayesian framework, iteratively solves a non-convex optimization problem using fixed point updates. It provides comparable performance and is significantly faster than convex optimization techniques used in sparse processing. We propose a multi-dictionary SBL algorithm that simultaneously can process observations generated by different underlying dictionaries sharing the same sparsity profile. Two algorithms are proposed and corresponding fixed point update equations are derived. Noise variances are estimated using stochastic maximum likelihood. The multi-dictionary SBL has many practical applications. We demonstrate this using direction-of-arrival (DOA) estimation. The first example uses the proposed multi-dictionary SBL to process multi-frequency observations. We show how spatial aliasing can be avoided while processing multi-frequency observations using SBL. SWellEx-96 experimental data demonstrates qualitatively these advantages. In the second example we show how data corrupted with heteroscedastic noise can be processed using multi-dictionary SBL with data pre-whitening.

1. Introduction and motivation

Compressed sensing or sparse processing is the process of estimating sparse vectors using significantly fewer measurements. Mathematically, this corresponds to solving an underdetermined system of linear equations under the constraint that the solution is sparse. The exact solution has combinatorial complexity which is impractical to solve for high dimensional problems. The most popular, approximate and computationally feasible, sparse processing method is basis pursuit [1] implemented using the LASSO [2] algorithm. Basis pursuit relaxes the sparsity criteria and the solution is given by solving a convex optimization problem. Though feasible, solving the optimization problem for high dimensions is still computationally slow. One of the faster alternatives is the matching pursuit algorithm [3]. However, this approach is greedy and can lead to suboptimal solutions. Another alternative, which is not greedy and significantly faster than basis pursuit, is sparse Bayesian learning (SBL) [4–10].

In SBL, the sparse weight vectors in the underdetermined system of linear equations are treated as random vectors with a Gaussian prior density. Explicit sparsity constraints are not imposed on the weight vectors. Unlike traditional prior models, the parameters of the Gaussian prior are assumed unknown and are estimated by performing evidence maximization. The objective function for performing evidence maximization is non-convex and an approximate solution is obtained by formulating a fixed point update equation. The estimated prior parameters for the weights are sparse in practice.

SBL was introduced for regression and classification problems in the context of machine learning [4]. It has been used since in signal processing [5–9]. SBL can be viewed as a stochastic maximum likelihood approach and has similarities with the SPICE and LIKES [11] algorithms for parameter estimation.

SBL does not impose explicitly any sparsity constraints but determines sparsity automatically. Analytically, SBL solution can be obtained by solving an iterated reweighted LASSO problem and hence sparsity is expected [12,13]. Various sparse signal recovery algorithms including LASSO and SBL can be unified within the Bayesian framework [14]. Cramér–Rao bounds for SBL based parameter estimation are discussed in [15,16].

A significant advantage of SBL over basis pursuit is that it can determine sparsity automatically without any user input. Being a

* Corresponding author.
E-mail addresses: santosh.nannuru@iiit.ac.in (S. Nannuru), gemba@ucsd.edu (K.L. Gemba), gerstoft@ucsd.edu (P. Gerstoft), whodgkiss@ucsd.edu (W.S. Hodgkiss), cfm@ieee.org (C.F. Mecklenbräuker).
URL: http://noiselab.ucsd.edu (K.L. Gemba), http://noiselab.ucsd.edu (P. Gerstoft), http://noiselab.ucsd.edu (W.S. Hodgkiss)

https://doi.org/10.1016/j.sigpro.2019.02.003
0165-1684/

probabilistic approach, SBL computes the posterior distribution of the sparse weight vectors and hence provides estimates of their covariance along with the mean. Computationally, SBL can significantly outperform LASSO [10,17].

Most of the compressed sensing literature deals with processing observations which can be represented sparsely using a single dictionary. Though this is sufficient for most applications, in some cases the different observations can be represented using different dictionaries but with a shared sparsity profile. Traditionally such observations are processed independently but it is advantageous to process them together to make use of additional gain due to the increased number of observations.

In this paper we process simultaneously observations from multiple dictionaries. We derive two SBL algorithms for multi-dictionary processing to extract the common underlying sparsity profile. The first algorithm SBL-CC (SBL-common covariance) processes observations from all the dictionaries simultaneously giving a single update rule. The second algorithm SBL-MC (SBL-multiple covariance) processes observations from each dictionary separately and then combines the result in a post-processing step. The noise variance is estimated using the stochastic maximum likelihood approach which provides unbiased estimates.

The multi-dictionary SBL approach with common sparsity profile was used in Ref. [18] for image processing using a machine learning framework. Their update equations for the inverse variance parameters of the weights use approximations that sacrifice a rigorous maximum likelihood approach. *In contrast*, we develop update equations for the weight variance parameter and it requires no approximations. For each dictionary, multiple measurement snapshots can be combined in the same update equation. We apply our methods for direction-of-arrival estimation using simulations and experimental underwater acoustic data.

We demonstrate the usefulness of multi-dictionary SBL using practical applications discussed below. Our focus is on direction-of-arrival (DOA) estimation of multiple plane waves, also called beamforming.

1) **Multi-frequency SBL:** The different frequency (different dictionary) snapshots are generated by the same set of broadband sources and hence share the same sparsity. Combining multi-frequency observations using multi-dictionary SBL provides a processing gain especially at low SNR.
2) **Multi-frequency SBL and aliasing:** We show from simulations that SBL can reduce spatial aliasing when processing multiple frequencies. Multi-dictionary SBL is used to process data from the SWellEx-96 experiment demonstrating application to real data.
3) **Heteroscedastic noise:** Often the sensor array data is corrupted by noise whose statistics change spatially and temporally, i.e. heteroscedastic noise. Preprocessing of this data to whiten the noise gives rise to observations generated by a different dictionary for each snapshot. We process these whitened snapshots using multi-dictionary SBL to improve beamforming performance at low SNR.

The multi-dictionary SBL developed in this paper can be applied readily to process multi-frequency sensor array measurements. In fact, sparse processing of wide-band (i.e. multi-frequency) signals can be found in [9,17,19,20]. The wide-band SBL proposed in [9] for DOA estimation assumes the noise variance to be same across all the frequencies. Multi-frequency SBL is used in [17] for localizing a towed source from underwater acoustic array measurements.

The remainder of the paper is organized as follows. The signal model along with multi-dictionary priors and likelihoods are discussed in Section 2. The multi-dictionary SBL algorithms are derived in Section 3. The derived algorithms are studied using simulations and real data in Section 4. Conclusions are provided in Section 5. Some portions of this paper have been presented at a conference [21].

Notation: Scalar quantities are denoted by lowercase letters. A bold lowercase letter denotes a vector and a bold uppercase letter denotes a matrix. A vector or matrix of all zeros is denoted by $\mathbf{0}$ where appropriate dimensions are assumed. An identity matrix of dimension $N \times N$ is denoted $\mathbf{I}_N$. The notation $\mathbf{M}^H$ denotes the Hermitian (conjugate transpose). The transpose operation is denoted $\mathbf{M}^T$. The field of complex numbers is denoted $\mathbb{C}$.

2. Signal model

In this section, we discuss the signal model used and the assumptions made in this paper. The single dictionary signal model is discussed in Section 2.1 followed by a multiple dictionary signal model in Section 2.2.

2.1. Single dictionary

Let $\mathbf{y} \in \mathbb{C}^N$ be the complex signal which is expressed as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \tag{1}$$

where the noise $\mathbf{n} \in \mathbb{C}^N$ is zero mean circularly symmetric complex Gaussian with density $\mathcal{CN}(\mathbf{n}; \mathbf{0}, \sigma^2\mathbf{I}_N)$; $\mathbf{A} \in \mathbb{C}^{N\times M}$ is the dictionary or sensing matrix; $\mathbf{x} \in \mathbb{C}^M$ is the weight vector. In sparse problem formulations, $\mathbf{x}$ is assumed sparse with at most K non-zero entries where $K \ll M$. The sparsity level K is not required explicitly or modeled by SBL. The vector $\mathbf{x}$ acts as a selection operator identifying columns of $\mathbf{A}$ that best explain the signal $\mathbf{y}$. We assume $\mathbf{A}$ has the maximal column rank N.

We often process multiple observations (snapshots) simultaneously. Let $\mathbf{Y} = [\mathbf{y}_1 \dots \mathbf{y}_L] \in \mathbb{C}^{N\times L}$ denote L consecutive snapshots arranged column-wise in a matrix. The multi-snapshot analogue of (1) is

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{N} \tag{2}$$

where $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_L]$ and $\mathbf{N} = [\mathbf{n}_1 \dots \mathbf{n}_L]$. The following set of assumptions are made in our signal model:

Assumption 1

(1) For the lth snapshot, the weight vector $\mathbf{x}_l$ and the noise $\mathbf{n}_l$ are independent.
(2) The weights $\mathbf{x}_l$ and the noise $\mathbf{n}_l$ are assumed to be Gaussian, independent and identically distributed (i.i.d.) across snapshots for $l = 1, \dots, L$.

The above assumptions in turn imply that the observations $\mathbf{y}_l$ are independent across snapshots. Additionally, in SBL the $\mathbf{x}_l$ are assumed to be circularly symmetric complex Gaussian with zero mean and covariance $\boldsymbol{\Gamma}$,

$$p(\mathbf{X}) = \prod_{l=1}^{L} p(\mathbf{x}_l) = \prod_{l=1}^{L} \mathcal{CN}(\mathbf{x}_l; \mathbf{0}, \boldsymbol{\Gamma}), \tag{3}$$

where the unknown covariance matrix $\boldsymbol{\Gamma}$ is assumed diagonal, $\boldsymbol{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma})$, $\boldsymbol{\gamma} = [\gamma_1 \dots \gamma_M]$. The covariance $\boldsymbol{\Gamma}$ is estimated by SBL. Since the noise is i.i.d., the Type-I likelihood is given by

$$p(\mathbf{Y}|\mathbf{X}) = \prod_{l=1}^{L} p(\mathbf{y}_l|\mathbf{x}_l) = \prod_{l=1}^{L} \mathcal{CN}(\mathbf{y}_l; \mathbf{A}\mathbf{x}_l, \sigma^2\mathbf{I}_N). \tag{4}$$

2.2. Multiple dictionaries

We assume observations generated by a set of dictionaries are available simultaneously and a portion of the support is common for all the weights. We are interested in recovering this shared sparsity structure.

Let the observation vectors corresponding to the D dictionaries be denoted $\mathbf{Y}_{1:D} \equiv \{\mathbf{Y}_1 \dots \mathbf{Y}_D\}$. These observations are related to the corresponding sparse weights $\mathbf{X}_{1:D} \equiv \{\mathbf{X}_1 \dots \mathbf{X}_D\}$ by the linear model

$$\mathbf{Y}_d = \mathbf{A}_d \mathbf{X}_d + \mathbf{N}_d\,, \quad d = 1, \dots, D \tag{5}$$

where $\mathbf{A}_d$ are the dictionaries and $\mathbf{N}_d$ are noise contributions. The following set of assumptions are made for the multi-dictionary signal model:

Assumption 2

(1) For a given dictionary $\mathbf{A}_d$, the weights $\mathbf{X}_d$ and the noise $\mathbf{N}_d$ satisfy Assumption 1.
(2) The weights $\mathbf{X}_d$ and the noise $\mathbf{N}_d$ are assumed to be independent across dictionaries for $d = 1, \dots, D$.

From above assumptions the observations $\mathbf{Y}_d$ are independent across dictionaries.

2.3. Multi-dictionary prior models

There are two possibilities for the joint multi-dictionary prior over $\mathbf{X}_{1:\,D}$ as discussed next.

2.3.1. Common covariance (CC) prior

This model assumes that the prior for all dictionaries is governed by the same statistical distribution. Let $\boldsymbol{\Gamma}$ be the covariance of the weight vectors for all the dictionaries. Hence we have

$$p(\mathbf{X}_{1:D}) = \prod_{d=1}^{D} p(\mathbf{X}_d) = \prod_{d=1}^{D}\prod_{l=1}^{L} \mathcal{CN}(\mathbf{x}_{dl}; \mathbf{0}, \boldsymbol{\Gamma}). \tag{6}$$

This imposes identical sparsity constraints on all the weight vectors in $\mathbf{X}_{1:\,D}$ because of the shared covariance. A common covariance prior was used for multi-frequency beamforming in [9,17].

2.3.2. Multiple covariance (MC) prior

This model assumes that the statistics of the prior depend on the dictionary. Let $\boldsymbol{\Gamma}_d$ be the covariance of the weight vectors associated with dth dictionary. Hence the joint prior is given by

$$p(\mathbf{X}_{1:D}) = \prod_{d=1}^{D} p(\mathbf{X}_d) = \prod_{d=1}^{D}\prod_{l=1}^{L} \mathcal{CN}(\mathbf{x}_{dl}; \mathbf{0}, \boldsymbol{\Gamma}_d). \tag{7}$$

where $\boldsymbol{\Gamma}_d = \mathrm{diag}(\boldsymbol{\gamma}_d)$. Since the covariance depends on the dictionary, in general, the sparsity of vectors $\mathbf{x}_{dl}$ will depend on dictionary. To extract the common sparsity of interest, we post-process to obtain an average $\boldsymbol{\gamma}$ across dictionaries. This model has been used in the context of multi-frequency beamforming in [22].

2.4. Multi-dictionary likelihood

Let σ_d^2 be the variance of the noise associated with observations in the dth dictionary. The multi-dictionary likelihood can then be written as

$$p(\mathbf{Y}_{1:D}|\mathbf{X}_{1:D}) = \prod_{d=1}^{D} p(\mathbf{Y}_d|\mathbf{X}_d) \tag{8}$$

$$= \prod_{d=1}^{D}\prod_{l=1}^{L} \mathcal{CN}(\mathbf{y}_{dl}; \mathbf{A}_d\mathbf{x}_{dl}, \sigma_d^2 \mathbf{I}_N). \tag{9}$$

3. Multi-dictionary SBL

In this section we derive the multi-dictionary SBL algorithm. The evidence term is computed in Section 3.1 which on maximizing gives the fixed point update rules in Section 3.2 and Section 3.3. The noise update is discussed in Section 3.4.

3.1. Multi-dictionary evidence

In the SBL framework [4,6], the prior covariance parameter for weight vectors is assumed unknown and estimated using the observed signal $\mathbf{Y}_{1:\,D}$. It is estimated by maximizing the evidence (also called Type-II maximum likelihood). Since both the prior and the Type-I likelihood are Gaussian, the evidence term $p(\mathbf{Y}_{1:\,D})$ has a Gaussian form as well

$$p(\mathbf{Y}_{1:D}) = \prod_{d=1}^{D} p(\mathbf{Y}_d) = \prod_{d=1}^{D} \int p(\mathbf{Y}_d|\mathbf{X}_d)p(\mathbf{X}_d)d\mathbf{X}_d \tag{10}$$

$$= \prod_{d=1}^{D}\prod_{l=1}^{L} \int \mathcal{CN}(\mathbf{y}_{dl}; \mathbf{A}_d\mathbf{x}_{dl}, \sigma_d^2\mathbf{I}_N)\mathcal{CN}(\mathbf{x}_{dl}; \mathbf{0}, \boldsymbol{\Gamma}_d)d\mathbf{x}_{dl}$$

$$= \prod_{d=1}^{D}\prod_{l=1}^{L} \mathcal{CN}(\mathbf{y}_{dl}; \mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{y}_d})\,, \tag{11}$$

where we have used the multiple covariance (MC) prior above and can obtain the common covariance (CC) version by setting $\boldsymbol{\Gamma}_1 = \dots = \boldsymbol{\Gamma}_D = \boldsymbol{\Gamma}$. The data covariance $\boldsymbol{\Sigma}_{\mathbf{y}_d}$ is given as

$$\boldsymbol{\Sigma}_{\mathbf{y}_d} = \mathcal{E}(\mathbf{y}_{dl}\mathbf{y}_{dl}^H) = \mathcal{E}(\mathbf{A}_d\mathbf{x}_{dl}\mathbf{x}_{dl}^H\mathbf{A}^H) + \mathcal{E}(\mathbf{n}_{dl}\mathbf{n}_{dl}^H) \tag{12}$$

$$= \mathbf{A}_d\boldsymbol{\Gamma}_d\mathbf{A}_d^H + \sigma_d^2\mathbf{I}_N. \tag{13}$$

The logarithm of the evidence is

$$\log(p(\mathbf{Y}_{1:D})) = -\sum_{d=1}^{D}\sum_{l=1}^{L}\left(\log|\boldsymbol{\Sigma}_{\mathbf{y}_d}| + \mathbf{y}_{dl}^H\boldsymbol{\Sigma}_{\mathbf{y}_d}^{-1}\mathbf{y}_{dl}\right) + C. \tag{14}$$

Depending on the prior model used for the weights we have two SBL algorithms. The multi-dictionary SBL using CC prior model is discussed next.

3.2. SBL-Common Covariance

For this model we set $\boldsymbol{\Gamma}_1 = \dots = \boldsymbol{\Gamma}_D = \boldsymbol{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma})$. The unknown parameters $\boldsymbol{\gamma}$ are estimated by maximizing the evidence

$$\hat{\boldsymbol{\gamma}} = \arg\max_{\boldsymbol{\gamma}} \log\left(p(\mathbf{Y}_{1:D})\right) \tag{15}$$

$$= \arg\min_{\boldsymbol{\gamma}} \left\{\sum_{d=1}^{D} L\log|\boldsymbol{\Sigma}_{\mathbf{y}_d}| + \mathrm{Tr}(\mathbf{Y}_d^H\boldsymbol{\Sigma}_{\mathbf{y}_d}^{-1}\mathbf{Y}_d)\right\}. \tag{16}$$

One approach to solve this problem is to use the EM algorithm [23] but the resulting update equations have slow convergence [4,6]. We perform differentiation of the objective function (16) to obtain a local minimum. We have the following derivative relations for the data covariance $\boldsymbol{\Sigma}_{\mathbf{y}_d}$

$$\frac{\partial\,\log|\boldsymbol{\Sigma}_{\mathbf{y}_d}|}{\partial\gamma_m} = \mathrm{Tr}\left(\boldsymbol{\Sigma}_{\mathbf{y}_d}^{-1}\frac{\partial\boldsymbol{\Sigma}_{\mathbf{y}_d}}{\partial\gamma_m}\right), \tag{17}$$

$$\frac{\partial\,\boldsymbol{\Sigma}_{\mathbf{y}_d}^{-1}}{\partial\gamma_m} = -\boldsymbol{\Sigma}_{\mathbf{y}_d}^{-1}\frac{\partial\boldsymbol{\Sigma}_{\mathbf{y}_d}}{\partial\gamma_m}\boldsymbol{\Sigma}_{\mathbf{y}_d}^{-1}, \quad \frac{\partial\boldsymbol{\Sigma}_{\mathbf{y}_d}}{\partial\gamma_m} = \mathbf{a}_{dm}\mathbf{a}_{dm}^H. \tag{18}$$

Differentiating (16) with respect to the mth diagonal element γ_m

$$\frac{\partial}{\partial\gamma_m}\left\{\sum_{d=1}^{D} L\log|\boldsymbol{\Sigma}_{\mathbf{y}_d}| + \mathrm{Tr}(\mathbf{Y}_d^H\boldsymbol{\Sigma}_{\mathbf{y}_d}^{-1}\mathbf{Y}_d)\right\}$$

$$= \sum_{d=1}^{D} L\,\mathrm{Tr}\left(\boldsymbol{\Sigma}_{\mathbf{y}_d}^{-1}\mathbf{a}_{dm}\mathbf{a}_{dm}^H\right) - \mathrm{Tr}\left(\mathbf{Y}_d^H\boldsymbol{\Sigma}_{\mathbf{y}_d}^{-1}\mathbf{a}_{dm}\mathbf{a}_{dm}^H\boldsymbol{\Sigma}_{\mathbf{y}_d}^{-1}\mathbf{Y}_d\right).$$

Equating the derivative of the objective function to zero

$$1 = \frac{1}{L} \frac{\sum_{d=1}^{D} \mathrm{Tr}\left(\mathbf{Y}_d^H \boldsymbol{\Sigma}_{\mathbf{y}_d}^{-1} \mathbf{a}_{dm} \mathbf{a}_{dm}^H \boldsymbol{\Sigma}_{\mathbf{y}_d}^{-1} \mathbf{Y}_d \right)}{\sum_{d=1}^{D} \mathrm{Tr}\left(\boldsymbol{\Sigma}_{\mathbf{y}_d}^{-1} \mathbf{a}_{dm} \mathbf{a}_{dm}^H] \right)} \tag{19}$$

$$\frac{\gamma_m}{\gamma_m} = \left(\frac{1}{L} \frac{\sum_{d=1}^{D} \mathrm{Tr}\left(\mathbf{Y}_d^H \boldsymbol{\Sigma}_{\mathbf{y}_d}^{-1} \mathbf{a}_{dm} \mathbf{a}_{dm}^H \boldsymbol{\Sigma}_{\mathbf{y}_d}^{-1} \mathbf{Y}_d \right)}{\sum_{d=1}^{D} \mathrm{Tr}\left(\boldsymbol{\Sigma}_{\mathbf{y}_d}^{-1} \mathbf{a}_{dm} \mathbf{a}_{dm}^H] \right)} \right)^b \tag{20}$$

where we introduced γ_m terms to obtain an iterative update equation. Since the fixed point update is not unique, the exponent term b is introduced to include a broad range of update rules and to control the speed of convergence. Different update equations introduced in the literature can be obtained using different values of b. The update then is

$$\gamma_m^{\text{new}} = \gamma_m^{\text{old}} \left(\frac{\sum_{d=1}^{D} \mathrm{Tr}\left(\boldsymbol{\Sigma}_{\mathbf{y}_d}^{-1} \mathbf{a}_{dm} \mathbf{a}_{dm}^H \boldsymbol{\Sigma}_{\mathbf{y}_d}^{-1} \mathbf{S}_{\mathbf{y}_d} \right)}{\sum_{d=1}^{D} \mathrm{Tr}\left(\boldsymbol{\Sigma}_{\mathbf{y}_d}^{-1} \mathbf{a}_{dm} \mathbf{a}_{dm}^H] \right)} \right)^b . \tag{21}$$

where $\mathbf{S}_{\mathbf{y}_d}$ is the sample covariance matrix $\mathbf{S}_{\mathbf{y}_d} = \frac{1}{L} \mathbf{Y}_d \mathbf{Y}_d^H$, and $\boldsymbol{\Sigma}_{\mathbf{y}_d}$ is given by (13). In this multi-dictionary formulation, a unified update rule is obtained that combines all the observations together from different dictionaries. A single parameter vector $\boldsymbol{\gamma}$ is estimated using all the multi-dictionary observations. Since $\boldsymbol{\gamma}$ is the variance, even though it is forced to be same across dictionaries, the actual weight vector estimates $\mathbf{x}_{dl}$ could still be different from dictionary to dictionary. Single dictionary update rule can be obtained by setting $D = 1$ in (21).

Remark. There are multiple ways to formulate a fixed point update equation. Our formulation is inspired by some of the equations used in the literature [4,6,10] and convergence properties of the simulation results. It is not clear for what values of b, if any, convergence of (21) is guaranteed. For a single dictionary ($D = 1$), a value of $b = 1$ gives the update equation used in [4,6] and $b = 0.5$ gives the update equation in [10].

3.3. SBL-Multiple Covariance

For this model the unknown covariance for each dictionary has to be computed. Maximizing the evidence we have

$$\hat{\boldsymbol{\gamma}}_1, \ldots, \hat{\boldsymbol{\gamma}}_D = \arg\max_{\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_D} \log\left(p(\mathbf{Y}_{1:D}) \right) \tag{22}$$

$$= \arg\min_{\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_D} \left\{ \sum_{d=1}^{D} L \log |\boldsymbol{\Sigma}_{\mathbf{y}_d}| + \mathrm{Tr}(\mathbf{Y}_d^H \boldsymbol{\Sigma}_{\mathbf{y}_d}^{-1} \mathbf{Y}_d) \right\}.$$

Since the different dictionary components are decoupled, maximizing the joint evidence corresponds to maximizing the evidence for each dictionary individually. Thus the update rule for dth dictionary is obtained from (21) by setting $D = 1$ as

$$\gamma_{dm}^{\text{new}} = \gamma_{dm}^{\text{old}} \left(\frac{\mathrm{Tr}\left(\boldsymbol{\Sigma}_{\mathbf{y}_d}^{-1} \mathbf{a}_{dm} \mathbf{a}_{dm}^H \boldsymbol{\Sigma}_{\mathbf{y}_d}^{-1} \mathbf{S}_{\mathbf{y}_d} \right)}{\mathrm{Tr}\left(\boldsymbol{\Sigma}_{\mathbf{y}_d}^{-1} \mathbf{a}_{dm} \mathbf{a}_{dm}^H \right)} \right)^b . \tag{23}$$

In contrast to SBL-CC (21), multiple parameter vectors $\boldsymbol{\gamma}_d$ are estimated in SBL-MC (23) from multi-dictionary observations. We can combine $\boldsymbol{\gamma}_d$ to obtain a single multi-dictionary estimate by averaging as follows

$$\hat{\boldsymbol{\gamma}} = \frac{1}{D} \sum_{d=1}^{D} \hat{\boldsymbol{\gamma}}_d . \tag{24}$$

If the sparsity of $\hat{\boldsymbol{\gamma}}_d$ is the same across dictionaries, the averaging above could enhance the sparsity of the estimate $\hat{\boldsymbol{\gamma}}$ in the presence of noise. This is because the true sparse components which are present in all the dictionaries will get averaged, whereas the randomly distributed spurious components (due to sidelobes or noise) will get suppressed by the averaging process. The averaging in (24) is inspired by traditional multi-frequency processing in conventional beamforming where the beamformer outputs at each frequency are combined incoherently [17].

3.4. Noise estimate

Similar to γ_m, an update equation for σ_d^2 can be obtained using the derivative of the evidence with respect to σ_d^2. But this update is biased towards zero [6,9,10]. The noise also can be integrated out by modeling the noise precision parameter with a Gamma prior distribution as done in [18]. We use a stochastic maximum likelihood based method to estimate σ_d^2. Let $\mathbf{A}_{\mathcal{M}}$ be formed by K columns of $\mathbf{A}$ indexed by $\mathcal{M}$, where the set $\mathcal{M}$ indicates the location of non-zero entries of $\mathbf{x}$ with cardinality $|\mathcal{M}| = K$. We can estimate $\mathcal{M}$ using $\boldsymbol{\gamma}$ through thresholding or picking its highest entries. The noise variance estimate for dth dictionary is then [9,10,24]

$$\hat{\sigma}_d^2 = \frac{1}{N - K} \mathrm{Tr}\left((\mathbf{I}_N - \mathbf{A}_{d,\mathcal{M}} \mathbf{A}_{d,\mathcal{M}}^{+}) \mathbf{S}_{\mathbf{y}_d} \right), \tag{25}$$

where $\mathbf{A}_{\mathcal{M}}^{+}$ denotes the Moore–Penrose pseudo-inverse. In [9] a common noise estimate was used for all dictionaries (i.e. frequencies).

4. Simulations and experimental data

4.1. SBL implementation

This section discusses the algorithmic implementation of the SBL update rules developed in Sections 3.2 and 3.3. A pseudocode for the multi-dictionary SBL algorithm is given in Algorithm 1. The single dictionary algorithm is obtained by setting $D = 1$.

Algorithm 1 Multi-dictionary SBL algorithm.

Parameters: $\epsilon = 10^{-6}, N_t = 3000, b = 1$
Input: $\mathbf{S}_{\mathbf{y}_d}, \mathbf{A}_d^o\ \forall d$
Initialization: $\gamma_m^{\text{old}} = 1, \forall m, \quad \hat{\sigma}_d^2 = 0.1 \quad \forall d$
for $i = 1$ to N_t
  Compute: $\boldsymbol{\Sigma}_{\mathbf{y}_d} = \mathbf{A}_d \boldsymbol{\Gamma}^{\text{old}} \mathbf{A}_d^H + \hat{\sigma}_d^2 \mathbf{I} \quad \forall d$
  (if SBL-CC) $\quad \gamma_m^{\text{new}}$ update $\forall m$ using (21)
  (if SBL-MC) $\quad \gamma_{dm}^{\text{new}}$ update $\forall m, d$ using (23)-(24)
  $\hat{\sigma}_d^2$ estimate $\forall d$ using (25)
  If $\frac{\|\boldsymbol{\gamma}^{\text{new}} - \boldsymbol{\gamma}^{\text{old}}\|_1}{\|\boldsymbol{\gamma}^{\text{old}}\|_1} < \epsilon$, **break**
  $\boldsymbol{\gamma}^{\text{old}} = \boldsymbol{\gamma}^{\text{new}}$
end
Output : $\boldsymbol{\gamma}^{\text{new}}$

Parameters ϵ and N_t determine the error convergence criteria and the maximum number of iterations, respectively. Here we initialize $\boldsymbol{\gamma}$ and σ_d^2 to arbitrary constant values but a more informed initialization can be used if available (for example $\boldsymbol{\gamma}$ could be set to the conventional beamforming solution). In the examples we choose the power exponent in the update rules to be $b = 1$ as used in [4,6], but $b = 0.5$ also gives similar performance. For both values of b a good noise estimate (25) is essential for convergence as observed.

The inputs to the algorithm are the sample covariance matrices $\mathbf{S}_{\mathbf{y}_d}$ and the sensing matrices $\mathbf{A}_d$. The parameters to estimate,

γ_m and σ_d^2, are initialized to constant non-zero values. The γ_m are updated using (21) for SBL-CC and using (23) and (24) for SBL-MC algorithm. K strongest peak locations are identified from $\boldsymbol{\gamma}^{\text{new}}$ to construct $\mathbf{A}_{\mathcal{M}}$ and the dictionary-dependent noise estimate (25). Though we assume K to be known for estimating the noise, this can be avoided by using model order identification methods [9]. If there are fewer peaks $\hat{K} < K$ in $\boldsymbol{\gamma}^{\text{new}}$, then the algorithm uses $\hat{K}$ during its processing. Specifically, the set $\mathcal{M}$ now has the cardinality $|\mathcal{M}| = \hat{K}$ and $\mathbf{A}_{d,\mathcal{M}}$ has fewer columns which in turn affect the noise update (25). Though we did not encounter in our simulations, this could happen if the peaks are required to have a certain minimum amplitude.

For each iteration, the computation is dominated by the update rule for $\boldsymbol{\gamma}$. From the structure of the update rule, both algorithms have same complexity. Since $\boldsymbol{\Gamma}_d$ is diagonal, the complexity of (13) is $\mathcal{O}(MN^2)$ and computing the inverse of $\boldsymbol{\Sigma}_{\mathbf{y}_d}$ is $\mathcal{O}(N^3)$. In a vectorized implementation of (21) (or (23)), for each dictionary, the various matrix multiplications together require $\mathcal{O}(2MN^2 + MNL)$ computations. Thus the overall complexity of a single multi-dictionary iteration is of the order of $\mathcal{O}\big(D \times [N^3 + 3MN^2 + MNL]\big)$.

We use beamforming to demonstrate the benefits of the proposed SBL algorithms. Sparsity of SBL is measured by $\boldsymbol{\gamma}$. Since the beamforming dictionary has high coherence among neighboring columns, we only consider local peaks. A local peak is defined as an element which is larger than its adjacent elements, i.e. a peak is present at l if $\gamma_{l-1} < \gamma_l > \gamma_{l+1}$. Since $\boldsymbol{\gamma}$ corresponds to the source power, it is treated as the angular power spectrum.

4.2. Beamforming

For the beamforming application, the observed signal is a linear combination of plane waves. Since the number of sources (arrival angles) is small, finely dividing the angle space results in a sparse $\mathbf{x}$ of complex amplitudes. By formulating beamforming as an underdetermined linear problem, sparse processing algorithms can be used to recover the arrival angles [10,25–27]. We use SBL to recover the DOAs.

For a narrow-band signal of wavelength λ and uniform sensor array with separation d, the sensing matrix columns are

$$\mathbf{a}_m = [1, e^{j2\pi \frac{d}{\lambda}\sin(\theta_m)}, \dots, e^{j2\pi \frac{(N-1)d}{\lambda}\sin(\theta_m)}]^T, \quad (26)$$

for $m = 1 \dots M$, where θ_m is the mth discretized angle. In simulations, the angle space $[-90, 90]^\circ$ is discretized with 1° separation giving $M = 181$. We model a $N = 20$ sensor array. $L = 30\ (> N)$ snapshots are processed. The array SNR per snapshot is defined as

$$\text{SNR} = 10\log_{10}\frac{\mathcal{E}\{||\mathbf{Ax}||_2^2\}}{\mathcal{E}\{||\mathbf{n}||_2^2\}}. \quad (27)$$

4.3. Multi-frequency analysis using SBL

Performance gain can be obtained by processing multiple frequencies simultaneously. Consider three sources located at angles $[-20, -15, 75]^\circ$, their amplitudes are zero-mean complex Gaussians with standard deviation [4, 13, 10] respectively. The proximity of weaker source at -20° to the strongest source at -15° makes this challenging for DOA estimation. Let observations be recorded at F frequencies with the same source variance at each frequency for a given DOA. The SNR is the same for all snapshots and for all frequencies. DOA estimation is performed using the SBL-CC and SBL-MC algorithms. If θ_k^i and $\hat{\theta}_k^i$ are the true and estimated DOA of the kth source in the ith simulation, then the DOA RMSE is computed as follows

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N_{\text{iter}}}\sum_{k=1}^{K}(\theta_k^i - \hat{\theta}_k^i)^2}{K\,N_{\text{iter}}}}, \quad (28)$$

where N_{iter} is the number of Monte Carlo trials.

The multi-frequency gain is observed from the RMSE versus SNR plots in Fig. 1. The number of frequencies range from $F = 1, 2, 4, 8$. The frequencies used are {800} Hz, {600, 800} Hz, {400, 600, 800, 1000} Hz, and $\{400, 500, \dots, 1100\}$ Hz respectively. The sensor array separation is set to $d = \frac{\lambda_{800}}{2}$ where λ_{800} is the wavelength corresponding to 800 Hz signal. $N = 20$ sensors are used giving a constant aperture in all cases. The RMSE is computed over 1000 Monte Carlo trials.

For $F = 1$, SBL-CC and SBL-MC give identical RMSE in Fig. 1. We observe that for SBL-CC as more frequencies are used the RMSE reduces for a given SNR whereas for SBL-MC they increase. This

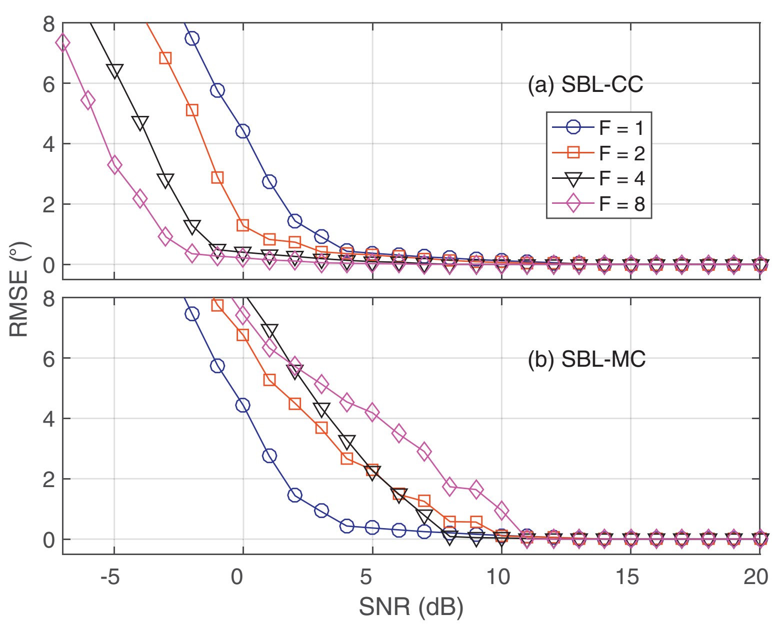


Fig. 1. RMSE vs SNR for (a) SBL-CC, and (b) SBL-MC with $N = 20$ sensors as number of frequencies is increased from $F = 1$ to $F = 8$.

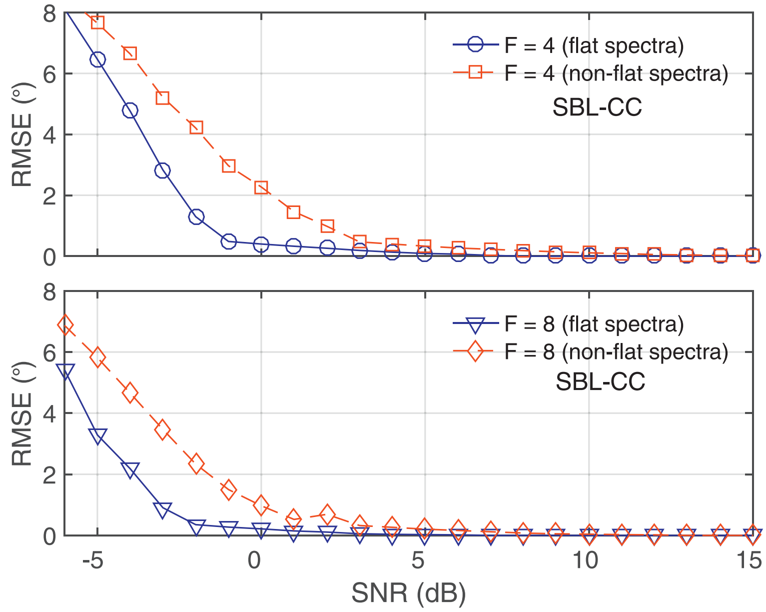


Fig. 2. RMSE vs SNR for SBL-CC with $N = 20$ sensors and $F = 4, 8$. The true source variance can either have a flat frequency spectra (solid line) or non-flat frequency spectra (dashed line).

is because for SBL-CC just one $\boldsymbol{\gamma}$ (21) is estimated from all observations making the estimation robust. For SBL-MC, since each frequency is processed independently (23) and later averaged (24), errors present in individual $\boldsymbol{\gamma}_d$ estimates, for example due to aliasing, cannot be fully overcome. Hence SBL-MC has higher error when compared to SBL-CC.

Though $\boldsymbol{\gamma}$ is assumed to be same for all frequencies, this may not be true in practice if the frequency spectrum is not flat. To demonstrate robustness of SBL-CC for DOA estimation in the presence of non-flat frequency spectra, we generate observations such that the source standard deviations are different across frequency. Let [4, 13, 10] be the nominal standard deviation for the sources as before. To simulate a non-flat spectrum for each source, its nominal standard deviation is multiplied by a different uniform random number in the range [0.5,1.5] for each frequency. The standard deviation is same across snapshots for a given frequency. The SBL-CC algorithm is applied to the generated observations and the RMSE for DOA estimation is shown in Fig. 2 for $F = 4, 8$. The presence of non-flat spectra has little impact on RMSE relative to the case when the spectra are flat. The performance is slightly better when more frequencies are used.

4.4. Aliasing suppression using multi-dictionary SBL

SBL can be used to process multi-frequency spatial data in the presence of aliasing. Each frequency has a different dictionary and the multi-dictionary analysis in Section 3 is used to process multi-frequency observations. Ref. [19] discusses aliasing suppression for wideband signals using basis pursuit and orthogonal matching pursuit. We demonstrate the aliasing suppression ability of SBL using both simulated and experimental data.

4.4.1. Simulation analysis

A large array aperture and hence a large sensor spacing is desirable to obtain high resolution beamforming. A drawback of large sensor spacing is that it limits the highest frequency that can be processed without encountering aliasing. This drawback partially can be overcome by multi-dictionary SBL.

The Gram matrix ($\mathbf{A}^H\mathbf{A}$) for two array spacings are shown in Fig. 3, $N = 20$. For a uniform linear array (ULA) spacing of $d = \frac{\lambda}{2}$ there is one main lobe for each angle. When the spacing is doubled, i.e. $d = \lambda$, grating (side) lobes appear due to aliasing.

We consider the three source example in Section 4.3 with sources located at $[-20, -15, 75]^\circ$ and source standard deviation [4, 13, 10]. Let f_1 and $f_2 = 2f_1$ be two frequencies with wavelengths λ_1 and $\lambda_2 = \frac{\lambda_1}{2}$. The frequency spectra are assumed to be flat in this section. The histograms of the top three peaks obtained from $\boldsymbol{\gamma}$ are shown in Fig. 4 when observations from each frequency is processed independently using SBL. Aliasing is absent in Fig. 4a since $d = \frac{\lambda_1}{2}$. Doubling the signal frequency with the same sensor spacing, Fig. 4b, gives aliased peaks. Higher frequency gives higher

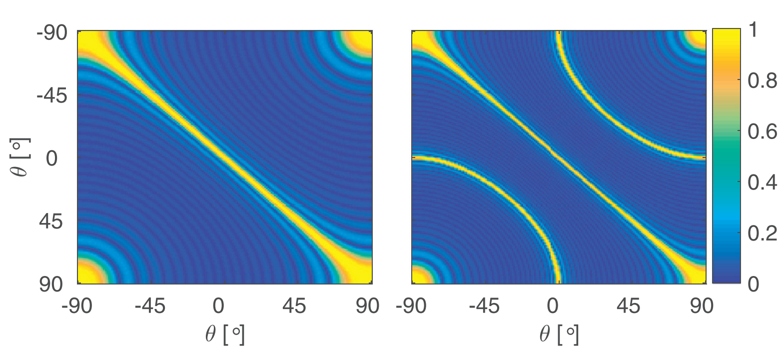


Fig. 3. Gram matrices for array spacings : $d = \frac{\lambda}{2}$ (left) and $d = \lambda$ (right). Number of sensors $N = 20$.

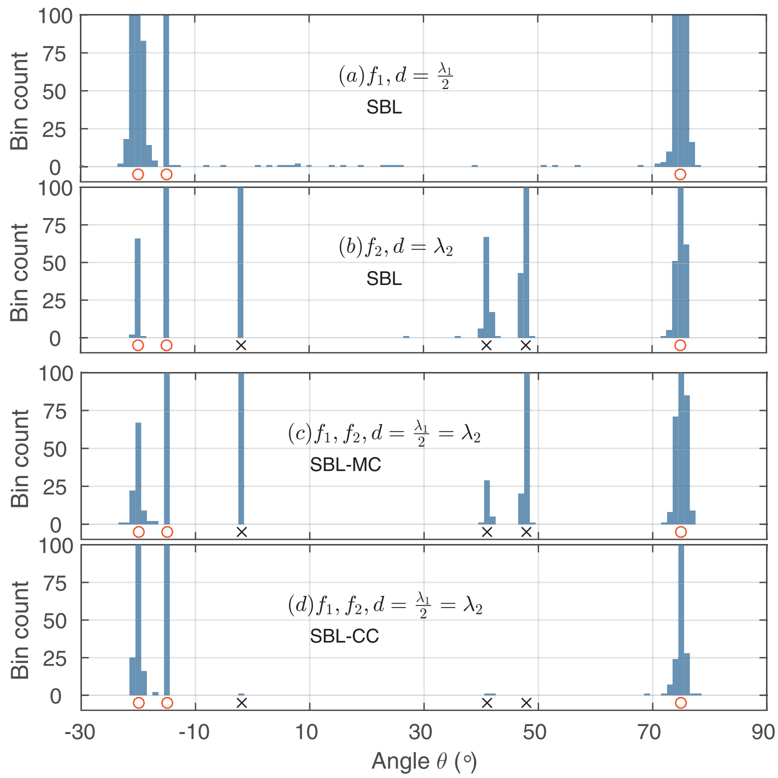


Fig. 4. Aliasing analysis using histograms of the top three peaks: Single frequency (a) SBL, half-wavelength spacing (b) SBL, full-wavelength spacing. Two frequencies (c) SBL-MC (d) SBL-CC. Number of sensors $N = 20$. Source (red circles) and aliased peak (black crosses) locations are indicated. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

resolution but with additional aliased peaks. Thus SBL has aliasing when only a single frequency is used.

We now combine the observations from the two frequencies using multi-dictionary SBL when the sensor spacing is fixed at $d = \frac{\lambda_1}{2} = \lambda_2$. The two multi-dictionary SBL formulations are discussed in Section 3. In SBL-MC, observations from each frequency are processed independently and the multi-frequency $\boldsymbol{\gamma}$ is obtained by summation (24). Fig. 4c shows the histogram when SBL-MC is used. The bin count is significant at aliased locations and hence SBL-MC cannot suppress aliasing. The SBL-CC enforces a common sparsity profile by requiring $\boldsymbol{\gamma}$ to be the same across frequencies. The histogram obtained using SBL-CC is shown in Fig. 4d. Since aliased peak locations are not shared across frequencies, they are suppressed by jointly processing multi-frequency observations using (21).

The aliasing suppression ability of SBL-CC also is demonstrated using a large number of frequencies, see Fig. 5. Specifically we consider measurements collected at 101 different frequencies which

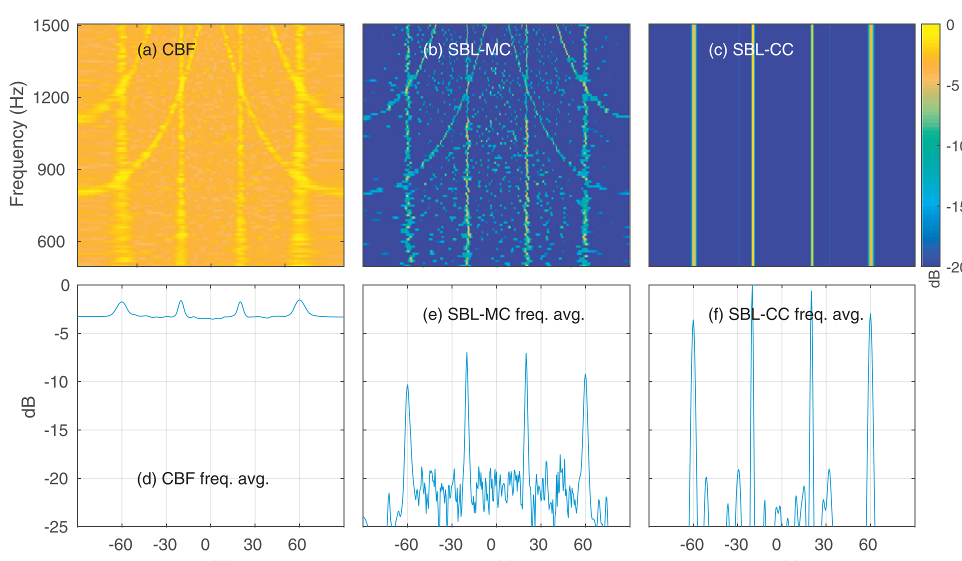


Fig. 5. Multi-frequency processing: The processor results as a function of frequency when 4 sources are present obtained using (a) CBF, (b) SBL-MC, and (c) SBL-CC. The bottom panel shows the plot of the average power across the frequencies. SNR $= -10$ dB and $L = 100$ snapshots per frequency. Number of sensors $N = 20$.

are $\{500, 510, \ldots, 1500\}$ Hz. The uniform sensor array has $N = 20$ sensors with spacing such that frequencies above 750 Hz experience aliasing. Four sources with equal (unit) power are located at $\{-60^\circ, -20^\circ, 20^\circ, 60^\circ\}$. Beamforming is performed using conventional beamforming (CBF) and SBL-MC for each frequency individually and using SBL-CC for all frequencies jointly. The results are shown in Fig. 5 for various frequencies at -10 dB SNR. CBF and SBL-MC experience aliasing at higher frequencies. The average power across frequencies is also shown. The CBF has the least resolution and dynamic range whereas the peaks in SBL-CC have the strongest ability to stand out even in the presence of a large amount of aliasing.

4.4.2. Experimental data analysis

The high-resolution performance of SBL compared to CBF is validated with experimental data in a complex multi-path, shallow-water environment. The aliasing suppression ability of multi-dictionary SBL is demonstrated by processing a subset of the array sensors. Though the true source spectrum is not flat for this experimental data, based on the analysis in Section 4.3, we still can apply multi-frequency SBL for accurate DOA estimation.

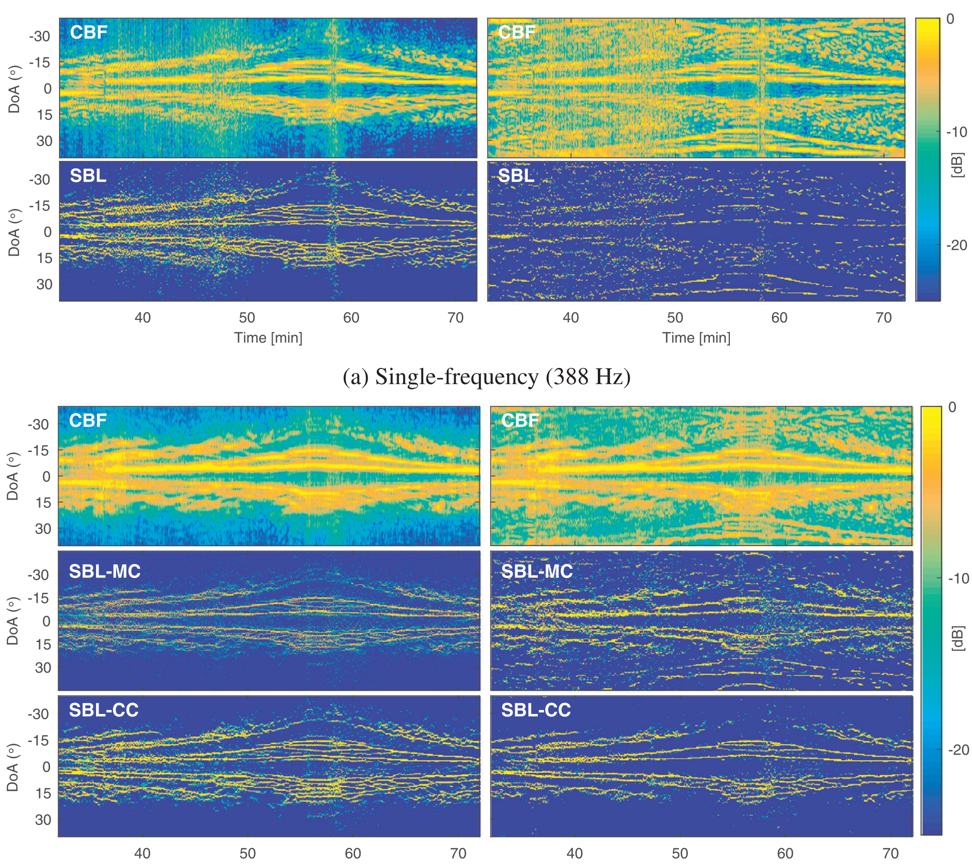
The data is from the Shallow Water evaluation cell Experiment 1996 (SWellEx-96) Event S5 [28] collected on a 64-element vertical line array. Element 43 is excluded from processing. The array spans the lower part of the 212 m watercolumn from 94 to 212 m with inter-sensor spacing $d = 1.875$ m. During the 77 min Event S5, a deep source submerged at 60 m was towed from 9 km southwest to 3 km northeast of the array at 5 km (2.5 m/s).

The source was transmitting a set of ten frequencies with constant source levels of which the three frequencies {166, 283, 388} Hz are processed. The data are split into 2257 overlapping segments, whereas a single segment is of 2.7 s duration. Snapshots are computed continuously from the data before being assigned to a segment. A FFT length of 2048 samples (1.35 s) with 50% overlap results in $L = 3$ snapshots for each segment with a FFT bin width of 0.75 Hz. To accommodate Doppler shift, we search two adjacent FFT bins and extract the bin with maximum power.

Both the full array (64 elements, Array-1) and a subset (21 elements, Array-2) are used for processing. Array-2 consists of every third element from Array-1 (Array-1 spacing d and Array-2 spacing $3d$). By design, Array-1 suffers no aliasing whereas Array-2 suffers aliasing above 133 Hz.

Single frequency (388 Hz) data is processed using both Array-1 and Array-2. Fig. 6a shows CBF output power (top row) and $\boldsymbol{\gamma}$ for SBL (bottom row) as the source moves over time. Array-1 processing does not suffer from aliasing (Fig. 6a, left) and multi-path arrivals can be seen. SBL provides finer angular resolution than CBF. Significant aliasing (Fig. 6a, right) is present in both the SBL and CBF outputs when Array-2 is used. This aliasing is due to insufficient spacial sampling. Significant power is redistributed into aliased locations causing ambiguities in DOA estimation.

Combining three frequencies {166, 283, 388} Hz and processing them from Array-1 and Array-2 is shown in Fig. 6b. Along with CBF output power (top row), the $\boldsymbol{\gamma}$ surfaces are shown for



(a) Single-frequency (388 Hz)

(b) Multi-frequency (166, 283, and 388 Hz)

Fig. 6. (a) Single-frequency (388 Hz) and (b) multi-frequency (166, 283, and 388 Hz) analysis of SWellEx-96 Event S5 data using 63 (left column) and 21 (right column) elements of the array. In (a) the top row is CBF and bottom row is single frequency SBL. In (b) the top row is CBF, middle row is SBL-MC, and the bottom row is SBL-CC. The columns of each of the panels are normalized.

SBL-MC (middle row) and SBL-CC (bottom row). Neither SBL nor CBF show any aliasing when Array-1 (Fig. 6b, left) data is processed. For Array-2 (Fig. 6b, right), CBF and SBL-MC both exhibit aliasing since the single frequency surfaces are averaged across frequencies. The relatively steep true arrivals around $\pm 20^\circ$ easily can get masked by the aliased arrivals causing DOA estimation errors. In comparison, SBL-CC shows no aliasing with Array-2 and the multi-path structure is preserved. We note that in general there are slightly fewer peaks identified, when compared to the corresponding Array-1 results, because of the reduced array gain of Array-2.

4.5. Multi-dictionary SBL for heteroscedastic noise

4.5.1. Heteroscedastic noise model

Consider the signal model (1) for the lth snapshot

$$\mathbf{y}_l = \mathbf{A}\mathbf{x}_l + \mathbf{n}_l, \tag{29}$$

where the statistics of the noise $\mathbf{n}_l$ are not stationary but change across sensors and snapshots. Specifically, the noise has zero mean and covariance $\boldsymbol{\Sigma}_{\mathbf{n}_l}$, $\mathbf{n}_l \sim \mathcal{CN}(\mathbf{n}_l; \mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{n}_l})$. We refer to this as a heteroscedastic noise model [29]. The noise is still assumed to be independent across snapshots and uncorrelated across sensors. Thus the noise covariance matrix is diagonal, $\boldsymbol{\Sigma}_{\mathbf{n}_l} = \mathrm{diag}([\sigma_{1l}^2, \ldots, \sigma_{Nl}^2])$ where σ_{nl}^2 is the variance of the nth sensor element at the lth snapshot.

4.5.2. Whitening

Direct application of single dictionary multi-snapshot SBL to the above set of observations would be sub-optimal because of the changing noise statistics. One way to process them is to whiten the noise. Multiplying (29) by $\mathbf{W}_l$ we have

$$\tilde{\mathbf{y}}_l = \tilde{\mathbf{A}}_l \mathbf{x}_l + \tilde{\mathbf{n}}_l, \tag{30}$$

$$\tilde{\mathbf{y}}_l = \mathbf{W}_l \mathbf{y}_l,\ \tilde{\mathbf{A}}_l = \mathbf{W}_l \mathbf{A},\ \tilde{\mathbf{n}}_l = \mathbf{W}_l \mathbf{n}_l, \tag{31}$$

where $\mathbf{W}_l = \mathrm{diag}([\sigma_{1l}^{-1}, \ldots, \sigma_{Nl}^{-1}])$ is the whitening matrix. The distribution of the modified noise is $\tilde{\mathbf{n}}_l \sim \mathcal{CN}(\tilde{\mathbf{n}}_l; \mathbf{0}, \mathbf{I})$ which is stationary across time and sensors. Though the resulting noise is

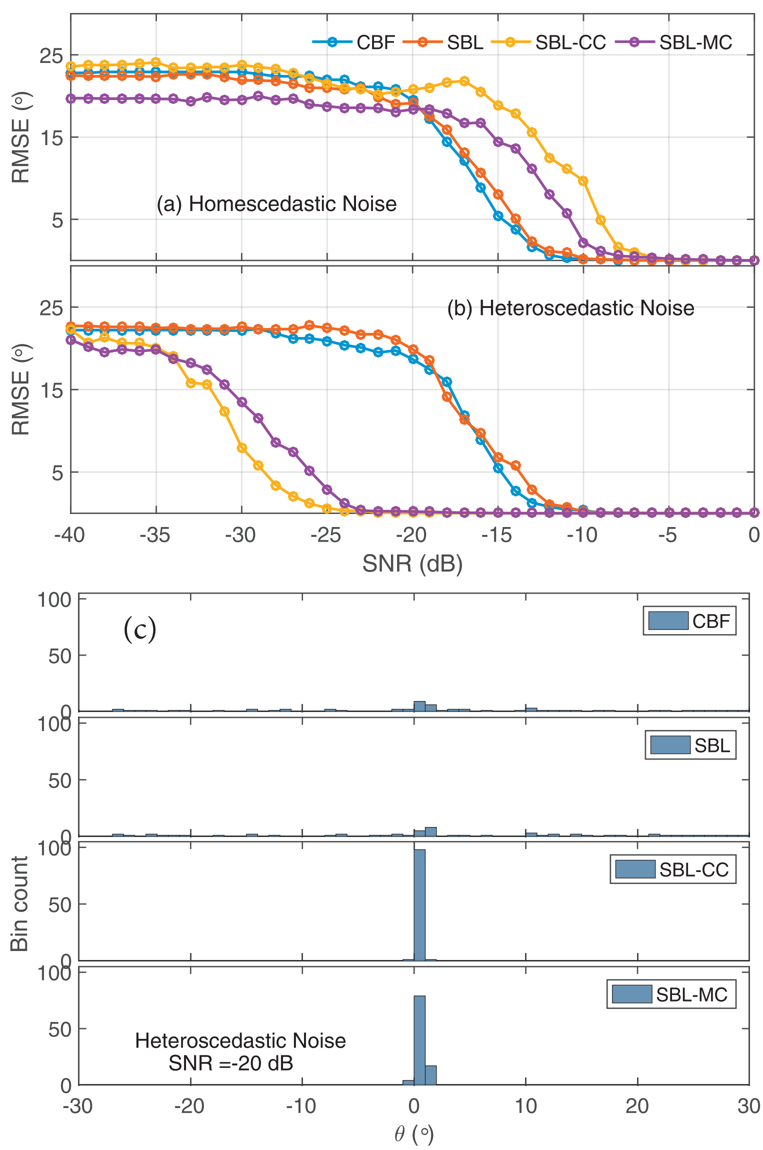


Fig. 7. Single source at 0°, RMSE vs. SNR for DOA estimation in (a) homoscedastic noise, (b) heteroscedastic noise. (c) Histogram of the location of the highest identified peak at an SNR of −20 dB.

whitened, we obtain a system of equations where the dictionary $\tilde{\mathbf{A}}_l$ is snapshot dependent. This system of equations can be solved using the multi-dictionary SBL developed in Section 3.

4.5.3. Multi-dictionary approach

From (30) the dictionary $\tilde{\mathbf{A}}_l$ changes for each snapshot. The source vector $\mathbf{x}_l$ is unaffected by this transformation and thus also $\boldsymbol{\gamma}$. After whitening, the array covariance matrix (13) becomes

$$\boldsymbol{\Sigma}_{\tilde{\mathbf{y}}_l} = \boldsymbol{\Sigma}_{\tilde{\mathbf{n}}_l} + \tilde{\mathbf{A}}_l \boldsymbol{\Gamma} \tilde{\mathbf{A}}_l^H. \tag{32}$$

Carrying this covariance as well as the $\tilde{\mathbf{y}}_l$ and $\tilde{\mathbf{A}}_l$ through the multi-dictionary SBL-CC derivation gives the following update for $\boldsymbol{\gamma}$:

$$\gamma_m^{\text{new}} = \gamma_m^{\text{old}} \left(\frac{\sum_{l=1}^{L} |\tilde{\mathbf{y}}_l^H \boldsymbol{\Sigma}_{\tilde{\mathbf{y}}_l}^{-1} \tilde{\mathbf{a}}_{lm}|^2}{\sum_{l=1}^{L} \tilde{\mathbf{a}}_{lm}^H \boldsymbol{\Sigma}_{\tilde{\mathbf{y}}_l}^{-1} \tilde{\mathbf{a}}_{lm}} \right)^b, \tag{33}$$

where the dictionary atoms $\tilde{\mathbf{a}}_{lm}$ change with snapshot. This SBL-CC algorithm is applied to observations with heteroscedastic noise after whitening. Similarly the SBL-MC approach can be applied to the whitened observations.

4.5.4. Noise estimate

For low SNR, we can assume the observed signal $\mathbf{Y}$ only contains noise and an analytic approximation is favored. Here we assume no knowledge of $\boldsymbol{\gamma}$ or its support. Assuming the diagonal entries of $\boldsymbol{\Sigma}_{\mathbf{y}_l}$ are much larger than the diagonal entries in $\mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^H$, we can approximate $\boldsymbol{\Sigma}_{\mathbf{n}_l}$ by the diagonal entries of $\boldsymbol{\Sigma}_{\mathbf{y}_l}$,

$$\boldsymbol{\Sigma}_{\mathbf{n}_l} \approx \operatorname{diag}(\operatorname{diag}(\boldsymbol{\Sigma}_{\mathbf{y}_l})). \tag{34}$$

Further, considering the fact we only have one snapshot available to estimate $\boldsymbol{\Sigma}_{\mathbf{n}_l}$, we can write

$$\boldsymbol{\Sigma}_{\mathbf{n}_l} \approx \operatorname{diag}(\operatorname{diag}(\mathbf{y}_l \mathbf{y}_l^H)) = \operatorname{diag}(|y_{1l}|^2, \ldots, |y_{Nl}|^2). \tag{35}$$

In other words, we consider

$$\hat{\sigma}_{nl} = |y_{nl}|. \tag{36}$$

This noise estimate is used for constructing the lth weighting matrix $\mathbf{W}_l$ for normalizing the observations $\mathbf{y}_l$ and the dictionary

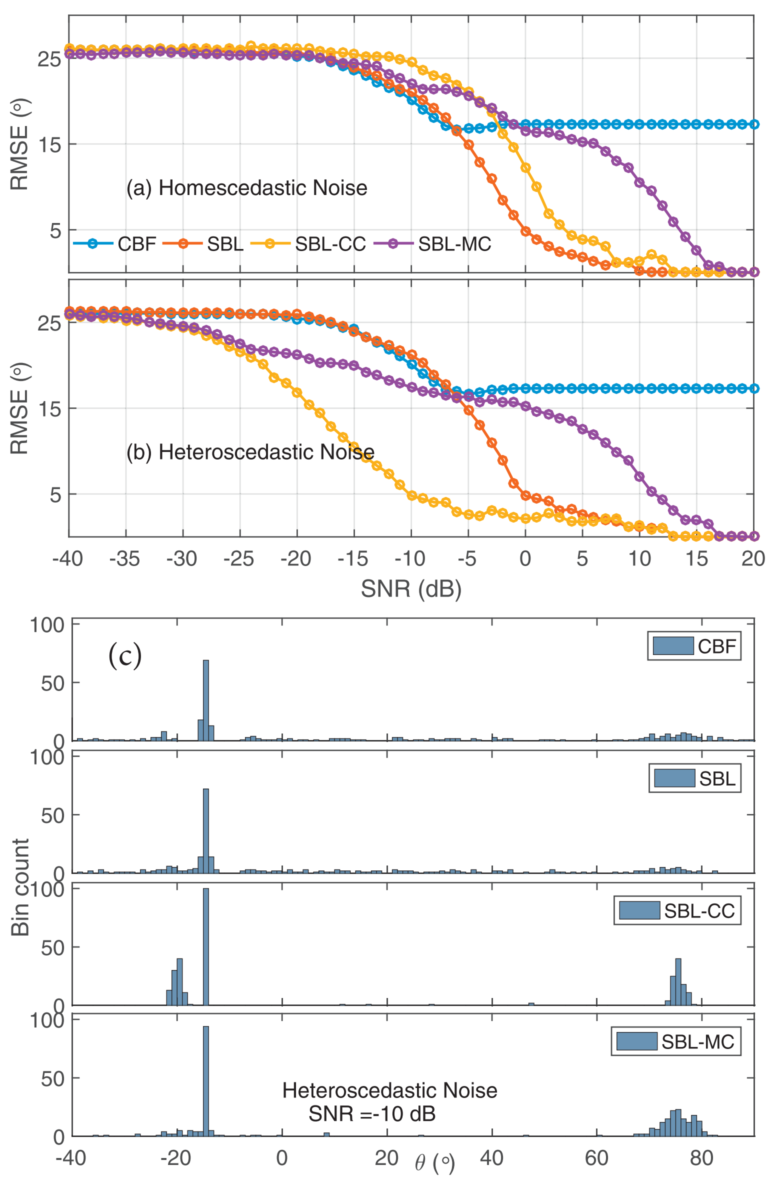


Fig. 8. Three sources at $\{-20, -15, 75\}^\circ$, RMSE vs. SNR for DOA estimation in (a) homoscedastic noise, (b) heteroscedastic noise. (c) Histogram of the three highest peak locations in $\boldsymbol{\gamma}$ at an SNR of -10 dB.

atoms $\mathbf{A}_l$, resulting in a different system of Eq. (30) for each snapshot. After whitening, the expected noise variance $\tilde{\sigma}_l^2$ of $\tilde{\mathbf{n}}_l$ approaches 1 for low SNR and approaches 0 for high SNR. Since the variance for each snapshot, $\tilde{\sigma}_l^2$, is expected to be similar, we compute the average noise variance

$$\tilde{\sigma}^2 = \frac{1}{L}\sum_{l=1}^{L}\tilde{\sigma}_l^2 \tag{37}$$

where $\tilde{\sigma}_l^2$ is given by (25).

The multi-dictionary SBL approach to process heteroscedastic data can be contrasted with the single dictionary SBL approach in [29]. The noise variance is estimated for each sensor per snapshot in [29]. Since the number of unknown parameters is equal to the size of the observed data, these estimates are less accurate. Whereas in the current method, the whitening step normalizes the noise variance across snapshots and sensors significantly reducing the number of unknown parameters to estimate.

4.5.5. Simulation - single DOA

Consider a single DOA located at 0° and observations recorded by an array with $N = 20$ sensors. Two noise cases are considered, homoscedastic noise and heteroscedastic noise. Homoscedastic noise has the same variance across sensors and snapshots. We simulate heteroscedastic noise by sampling the noise standard deviation from a log-normal distribution as $\log_{10}\sigma_{nl} \sim \mathcal{U}(-1, 1)$. The data is processed using various algorithms including CBF, SBL (single dictionary based processing assuming homoscedastic noise), SBL-CC, and SBL-MC. Both SBL-CC and SBL-MC process whitened observations as discussed in Section 4.5.3.

Plots of RMSE vs. SNR are shown in Fig. 7a and b for processing observations containing homoscedastic noise and heteroscedastic noise respectively. The performance of various algorithms is similar for homoscedastic noise as there is no advantage of using noise dependent processing such as SBL-CC and SBL-MC. Whereas both SBL-CC and SBL-MC perform significantly better in case of heteroscedastic noise, especially in the low SNR regime. For heteroscedastic noise case, a histogram of the highest peak (in $\boldsymbol{\gamma}$) is shown in Fig. 7c when SNR is −20 dB. The histograms for both SBL-CC and SBL-MC are concentrated around true source location of 0°. All figures are computed using 100 Monte Carlo simulations.

4.5.6. Simulation - three DOAs

We now consider an example with three DOAs located at $\{-20, -15, 75\}^\circ$ and having corresponding power of {10, 22, 20} dB. The RMSE vs. SNR is plotted in Fig. 8a for homoscedastic noise and in Fig. 8b for heteroscedastic noise. The SBL-MC performs very poorly in both noise cases because estimating $\boldsymbol{\gamma}_l$ per snapshot and averaging them is not effective at low SNR when sources are closely spaced. SBL-CC has the lowest RMSE when the noise is heteroscedastic. The histogram of the top three peaks is shown in Fig. 8c at −10 dB SNR. Only SBL-CC populates the histogram at all the three DOA locations.

5. Conclusions

We developed SBL to process observations from multiple dictionaries when a portion of the support is common for all of the weight vectors. The first multi-dictionary SBL shares the prior covariance across dictionaries and thus has fewer unknowns to estimate. A unified update rule is derived combining observations from all of the dictionaries. The second multi-dictionary SBL has dictionary-dependent prior parameters which are estimated independently for each dictionary. For the purpose of estimating a sparsity profile, a single covariance is obtained by averaging the dictionary-dependent estimates.

Beamforming simulations for DOA estimation are used to demonstrate the usefulness of multi-dictionary SBL. A lower RMSE for DOA estimates was obtained when multiple frequency observations were processed using SBL-CC. The SBL-CC algorithm also was able to correctly recover DOAs in the presence of spatial aliasing. This was demonstrated using both simulated and experimental data.

Whitening of the data with heteroscedastic noise resulted in observations generated by different dictionaries. We processed them using the multi-dictionary SBL algorithms to improve significantly the RMSE performance especially in the low SNR regime.

Acknowledgment

This work was supported by the Office of Naval Research Grant Nos. N00016-1-2341 and N00014-13-1-0632.

Conflicts of interest

None.

References

[1] S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, SIAM Rev. 43 (1) (2001) 129–159.
[2] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Ser. B (Methodological) 58 (1) (1996) 267–288.
[3] S.G. Mallat, Z. Zhang, Matching pursuits with time-frequency dictionaries, IEEE Trans. Signal Process. 41 (12) (1993) 3397–3415.
[4] M.E. Tipping, Sparse Bayesian learning and the relevance vector machine, J. Mach. Learn. Res. 1 (2001) 211–244.
[5] D.P. Wipf, B.D. Rao, Sparse Bayesian learning for basis selection, IEEE Trans. Signal Process. 52 (8) (2004) 2153–2164.
[6] D.P. Wipf, B.D. Rao, An empirical Bayesian strategy for solving the simultaneous sparse approximation problem, IEEE Trans. Signal Process. 55 (7) (2007) 3704–3716.
[7] S. Ji, Y. Xue, L. Carin, Bayesian compressive sensing, IEEE Trans. Signal Process. 56 (6) (2008) 2346–2356.
[8] Z. Zhang, B.D. Rao, Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning, IEEE J. Sel. Top. Signal Process. 5 (5) (2011) 912–926.
[9] Z.-M. Liu, Z.-T. Huang, Y.-Y. Zhou, An efficient maximum likelihood method for direction-of-arrival estimation via sparse Bayesian learning, IEEE Trans. Wireless Comm. 11 (10) (2012) 1–11.
[10] P. Gerstoft, C.F. Mecklenbräuker, A. Xenaki, S. Nannuru, Multi snapshot sparse Bayesian learning for DOA, IEEE Signal Process. Lett. 23 (10) (2016) 1469–1473.
[11] P. Stoica, P. Babu, SPICE And LIKES: two hyperparameter-free methods for sparse-parameter estimation, Signal Process. 92 (7) (2012) 1580–1590.
[12] D.P. Wipf, S.S. Nagarajan, A new view of automatic relevance determination, in: Adv. Neural Inf. Process. Systems, 20, 2008, pp. 1625–1632.
[13] D.P. Wipf, S.S. Nagarajan, Iterative reweighted L1 and L2 methods for finding sparse solutions, IEEE J. Sel. Top. Signal Process. 4 (2) (2010) 317–329.
[14] R. Giri, B. Rao, Type I and type II Bayesian methods for sparse signal recovery using scale mixtures, IEEE Trans. Signal Process. 64 (13) (2016) 3418–3428.
[15] R. Prasad, C.R. Murthy, Cramér-Rao-type bounds for sparse Bayesian learning, IEEE Trans. Signal Process. 61 (3) (2013) 622–632.
[16] A. Koochakzadeh, P. Pal, On saturation of the Cramér-Rao bound for sparse Bayesian learning, in: IEEE Int. Conf. Acoust., Speech Signal Process, New Orleans, USA, 2017.
[17] K.L. Gemba, S. Nannuru, P. Gerstoft, W.S. Hodgkiss, Multi-frequency sparse Bayesian learning for robust matched field processing, J. Acoust. Soc. Am. 141 (5) (2017) 3411–3420.
[18] S. Ji, D. Dunson, L. Carin, Multitask compressive sensing, IEEE Trans. Signal Process. 57 (1) (2009) 92–106.
[19] Z. Tang, G. Blacquiere, G. Leus, Aliasing-free wideband beamforming using sparse signal representation, IEEE Trans. Signal Process. 59 (7) (2011) 3464–3469.
[20] K.L. Gemba, J. Sarkar, B. Cornuelle, W.S. Hodgkiss, W.A. Kuperman, Estimating relative channel impulse responses from ships of opportunity in a shallow water environment, J. Acoust. Soc. Am. 144 (3) (2018) 1231–1244, doi:10.1121/1.5052259.
[21] S. Nannuru, K.L. Gemba, P. Gerstoft, Sparse Bayesian learning with multiple dictionaries, in: IEEE Global Conf. Signal Inf. Process, Montreal, Canada, 2017.
[22] P. Gerstoft, C.F. Mecklenbräuker, Wideband sparse Bayesian learning for DOA estimation from multiple snapshots, IEEE Sen. Array Multichannel Signal Process. Workshop, Rio de Janeiro, Brazil, 2016.

[23] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. R. Stat. Soc. Ser. B 39 (1) (1977) 1–38.
[24] P. Stoica, A. Nehorai, On the concentrated stochastic likelihood function in array signal processing, Circuits Syst. Signal Process. 14 (5) (1995) 669–674.
[25] D. Malioutov, M. Çetin, A.S. Willsky, A sparse signal reconstruction perspective for source localization with sensor arrays, IEEE Trans. Signal Process. 53 (8) (2005) 3010–3022.
[26] A. Xenaki, P. Gerstoft, K. Mosegaard, Compressive beamforming, J. Acoust. Soc. Am. 136 (1) (2014) 260–271.
[27] P. Gerstoft, A. Xenaki, C.F. Mecklenbräuker, Multiple and single snapshot compressive beamforming, J. Acoust. Soc. Am. 138 (4) (2015) 2003–2014.
[28] K.L. Gemba, W.S. Hodgkiss, P. Gerstoft, Adaptive and compressive matched field processing, J. Acoust. Soc. Am. 141 (1) (2017) 92–103.
[29] P. Gerstoft, S. Nannuru, C.F. Mecklenbräuker, G. Leus, Sparse Bayesian learning for DOA estimation in heteroscedastic noise, arXiv:1711.03847 (2018).